THE NETWORK IS THE COMPUTER™

UltraSPARC T2™ Supplement

to the *UltraSPARC Architecture 2007*

Draft D1.4.3, 19 Sep 2007

Privilege Levels: *Hyperprivileged, Privileged, and Nonprivileged*

Distribution: Public

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054
U.S.A. 650-960-1300

Part No: 950-5556-02
Revision: Draft D1.4.3, 19 Sep 2007

Contents

CHAPTER 1

UltraSPARC T2 Basics

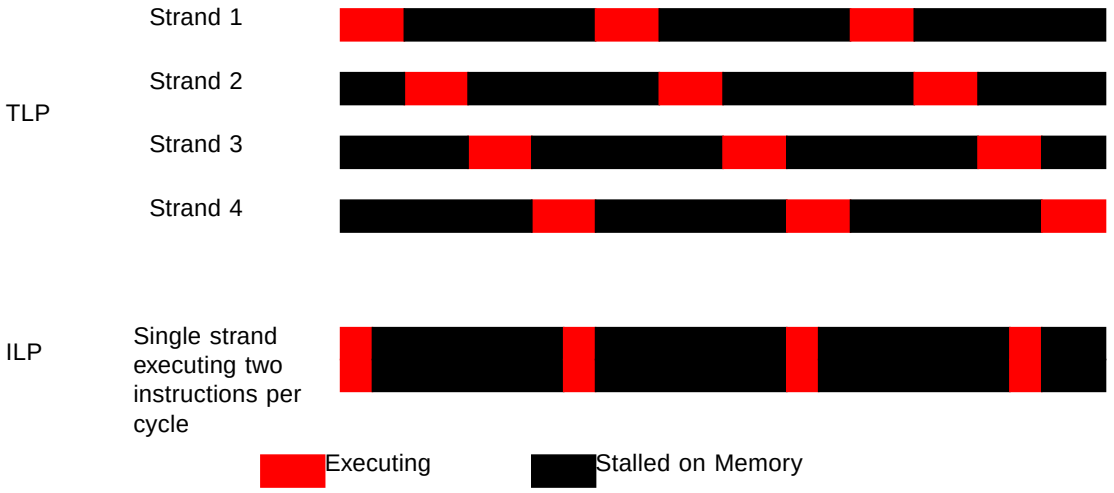
1.1 Background

UltraSPARC T2 is the follow-on chip multi-threaded (CMT) processor to the highly successful UltraSPARC T1 processor. The UltraSPARC T1 product line fully implements Sun's Throughput Computing initiative for the horizontal system space. Throughput Computing is a technique that takes advantage of the thread-level parallelism that is present in most commercial workloads. Unlike desktop workloads, which often have a small number of threads concurrently running, most commercial workloads achieve their scalability by employing large pools of concurrent threads.

Historically, microprocessors have been designed to target desktop workloads, and as a result have focused on running a single thread as quickly as possible. Single thread performance is achieved in these processors by a combination of extremely deep pipelines (over 20 stages in Pentium 4) and by executing multiple instructions in parallel (referred to as instruction-level parallelism or ILP). The basic tenet behind Throughput Computing is that exploiting ILP and deep pipelining has reached the point of diminishing returns, and as a result current microprocessors do not utilize their underlying hardware very efficiently. For many commercial workloads, the processor will be idle most of the time waiting on memory, and even when it is executing it will often be able to only utilize a small fraction of its wide execution width. So rather than building a large and complex ILP processor that sits idle most of the time, a number of small, single-issue processors that employ multithreading are built in the same chip area. Combining multiple processors on a single chip with

multiple strands per processor, allows very high performance for highly threaded commercial applications. This approach is called thread-level parallelism (TLP), and the difference between TLP and ILP is shown in the FIGURE 1-1.

FIGURE 1-1 Differences Between TLP and ILP



The memory stall time of one strand can often be overlapped with execution of other strands on the same processor, and multiple processors run their strands in parallel. In the ideal case, shown in FIGURE 1-1, memory latency can be completely overlapped with execution of other strands. In contrast, instruction-level parallelism simply shortens the time to execute instructions and does not help much in overlapping execution with memory latency.[1]

Given this ability to overlap execution with memory latency, why don't more processors utilize TLP? The answer is that designing processors is a mostly evolutionary process, and the ubiquitous deeply pipelined, wide ILP processors of today are the evolutionary outgrowth from a time when the processor was the bottleneck in delivering good performance. With processors capable of multiple GHz clocking, the performance bottleneck has shifted to the memory and I/O subsystems, and TLP has an obvious advantage over ILP for tolerating the large I/O and memory latency prevalent in commercial applications. Of course, every architectural technique has its advantages and disadvantages. The one disadvantage to employing TLP over ILP is that execution of a single thread will be slower on the TLP processor than an ILP processor. With processors running well over a GHz, a strand capable of executing only a single instruction per cycle is fully capable of completing tasks in the time required by the application, making this disadvantage a nonissue for nearly all commercial applications.

1. Processors that employ out-of-order ILP can overlap some memory latency with execution. However, this overlap is typically limited to shorter memory latency events such as L1 cache misses that hit in the L2 cache. Longer memory latency events such as main memory accesses are rarely overlapped to a significant degree with execution by an out-of-order processor.

1.2 UltraSPARC T2 Overview

UltraSPARC T2 is a single chip multi-threaded (CMT) processor. UltraSPARC T2 contains eight SPARC physical processor cores. Each SPARC physical processor core has full hardware support for eight strands, two integer execution pipelines, one floating-point execution pipeline, and one memory pipeline. The floating-point and memory pipelines are shared by all eight strands. The eight strands are hard-partitioned into two groups of four, and the four strands within a group share a single integer pipeline. While all eight strands run simultaneously, at any given time at most two strands will be active in the physical core, and those two strands will be issuing either a pair of integer pipeline operations, an integer operation and a floating-point operation, an integer operation and a memory operation, or a floating-point operation and a memory operation. Strands are switched on a cycle-by-cycle basis between the available strands within the hard-partitioned group of four using a least recently issued priority scheme. When a strand encounters a long-latency event, such as a cache miss, it is marked unavailable and instructions will not be issued from that strand until the long-latency event is resolved. Execution of the remaining available strands will continue while the long-latency event of the first strand is resolved.

Each SPARC physical core has a 16 KB, 8-way associative instruction cache (32-byte lines), 8 Kbytes, 4-way associative data cache (16-byte lines), 64-entry fully-associative instruction TLB, and 128-entry fully associative data TLB that are shared by the eight strands. The eight SPARC physical cores are connected through a crossbar to an on-chip unified 4 Mbyte, 16-way associative L2 cache (64-byte lines). The L2 cache is banked eight ways to provide sufficient bandwidth for the eight SPARC physical cores. The L2 cache connects to four on-chip DRAM controllers, which directly interface to a pair of fully buffered DIMM (FBD) channels. In addition, an on-chip PCI-EX controller, two 1-Gbit/10-Gbit Ethernet MACs, and several on-chip I/O-mapped control registers are accessible to the SPARC physical cores. Traffic from the PCI-EX port coherently interacts with the L2 cache.

A block diagram of the UltraSPARC T2 chip is shown in FIGURE 1-2.

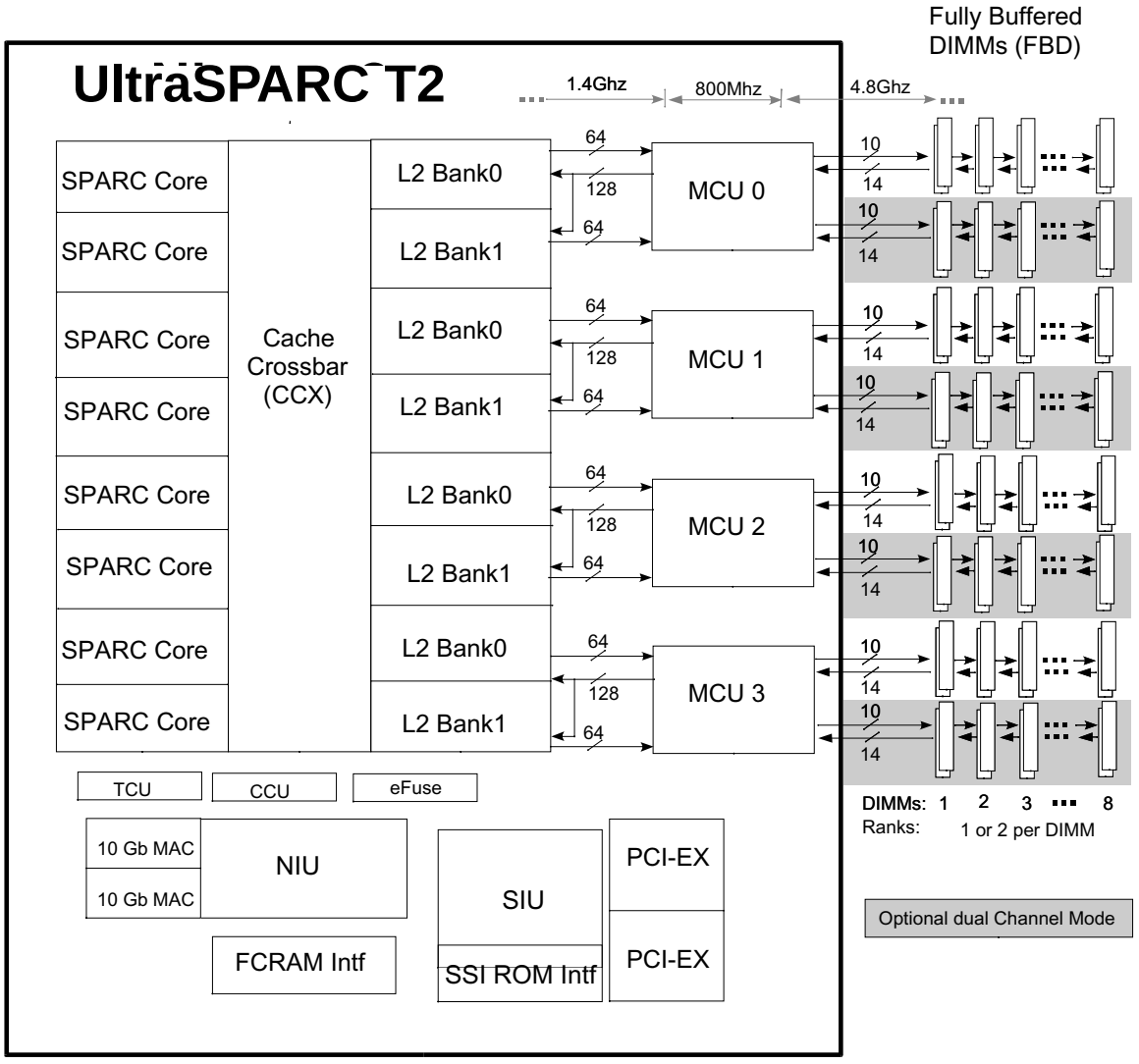


FIGURE 1-2 UltraSPARC T2 Chip Block Diagram

1.3 UltraSPARC T2 Components

This section describes each component in UltraSPARC T2.

1.3.1 SPARC Physical Core

Each SPARC physical core has hardware support for eight strands. This support consists of a full register file (with eight register windows) per strand, with most of the ASI, ASR, and privileged registers replicated per strand. The eight strands share the instruction and data caches and TLBs. An auto-demap feature is included with the TLBs to allow the multiple strands to update the TLB without locking.

Each SPARC physical core contains a floating-point unit, shared by all eight strands. The floating-point unit performs single- and double-precision floating-point operations, graphics operations, and integer multiply and divide operations.

1.3.2 L2 Cache

The L2 cache is banked eight ways. To provide for better partial-die recovery, UltraSPARC T2 can also be configured in 4-bank and 2-bank modes (with 1/2 and 1/4 the total cache size respectively). Bank selection based on physical address bits 8:6 for 8 banks, 7:6 for 4 banks, and 6 for 2 banks. The cache is 4 Mbytes, and 16-way set associative. The line size is 64 bytes. Unloaded access time is 26 cycles for an L1 data cache miss and 24 cycles for an L1 instruction cache miss.

1.3.3 Memory Controller Unit (MCU)

UltraSPARC T2 has four MCUs, one for each memory branch with a pair of L2 banks interacting with exactly one DRAM branch. The branches are interleaved based on physical address bits 7:6, and support 1–16 DDR2 DIMMs. Each memory branch is two FBD channels wide. A branch may use only one of the FBD channels in a reduced power configuration.

Each DRAM branch operates independently and can have a different memory size and a different kind of DIMM (for example, a different number of ranks or different CAS latency). Software should not use address space larger than four times the lowest memory capacity in a branch because the cache lines are interleaved across branches. The DRAM controller frequency is the same as that of the DDR (Double Data Rate) data buses, which is twice the DDR frequency. The FBDIMM links run at six times the frequency of the DDR data buses.

The UltraSPARC T2 MCU implements a DDR2 FBD design model that is based on various JEDEC-approved DDR2 SDRAM and FBDIMM standards. JEDEC has received information that certain patents or patent applications may be relevant to FBDIMM Advanced Memory Buffer standard (JESD82-20) as well as other standards related to FBDIMM technology (JESD206) (For more information, see http://www.jedec.org/download/search/FBDIMM/Patents.xls). Sun Microsystems does not provide any legal opinions as to the validity or relevancy of such patents or patent applications. Sun Microsystems encourages prospective users of the UltraSPARC T2 MCU design to review all information assembled by JEDEC and develop their own independent conclusion

1.3.4 Noncacheable Unit (NCU)

The NCU performs an address decode on I/O-addressable transactions and directs them to the appropriate block (for example, NIU, DMU, CCU). In addition, the NCU maintains the register status for external interrupts.

1.3.5 System Interface Unit (SIU)

The System Interface Unit connects the NIU, DMU and L2 Cache. SIU is the L2 Cache access point for the Network and PCI-Express subsystems. The SIU-L2 Cache interface is also the ordering point for PCI-Express ordering rule.

1.3.6 Data Management Unit (DMU)

The DMU manages Transaction Layer Packet (TLP) to/from the PEU and maintains the same ordering as from the PCI-Express Unit (PEU) and then to the SIU. For maintaining ordering between PEU and SIU, the DMU requires the policy that has PIO reads pulling DMA writes to completion. When the PEU issues complete TLP transactions to the DMU, the DMU segments the TLP packet into multiple cacheline-oriented SIU commands and issues them to the SIU. The DMU also queues the response cachelines from SIU, reassembly the multiple cachelines into one TLP packet with maximal payload size. Furthermore, the DMU accepts and queues the PIO transactions requests from NCU, and coordinates with the appropriate destination, to which the address and data will be sent.

The DMU encapsulates the functions necessary to resolve a virtual PCI-Express packet address into a L2 cacheline physical address which can be presented on the SIU interface. The DMU also encapsulates the functions necessary to interpret PCI-Express message signaled interrupts, emulated INTX interrupts and provides the functions to post interrupt events to queues managed by software in main memory and generates the Solaris Interrupt Mondo to notify software. The DMU decodes

INTACK and INTNACK from interrupt targets and conveys the information to the interrupt function so that it can move on to service the next interrupt if any (for INTACK) or replay the current interrupt (for INTNACK).

1.3.7 PCI-Express Unit (PEU)

The PEU implements the root complex behavior of the PCI-Express Base Spec 1.0A published by PCI-SIG. It interfaces to the Data Management Unit.

The PEU implements the three layers specified:

- Transaction Layer
- Data Link Layer
- Logical sub-block of the Physical Layer

The PEU supports x1, x2, x4 and x8 configuration at the data rate of 2.5 Gb/s. The PEU also supports the lane reversal feature.

1.3.8 Network Interface Unit (NIU)

The NIU connects a pair of on-chip 10 Gb/s Ethernet MACs to the rest of the system. The NIU also contains the registers to control Ethernet traffic.

1.3.9 SSI ROM Interface (SSI)

UltraSPARC T2 has a 50 Mb/s serial interface (SSI), which connects to an external field-programmable gate array (FPGA) that interfaces to the boot ROM. In addition, the SSI supports PIO accesses across the SSI, thus supporting optional Control and Status registers (CSRs) or other interfaces within the FPGA.

CHAPTER 2

Data Formats

Data formats supported by UltraSPARC T2 are described in the *UltraSPARC Architecture 2007* specification.

CHAPTER 3

Registers

3.1 Ancillary State Registers (ASRs)

This chapter discusses the UltraSPARC T2 ancillary state registers. TABLE 3-1 summarizes and defines these registers.

TABLE 3-1 Summary of UltraSPARC T2 Ancillary State Registers

Address	ASR Name	Access	priv	Replicated per-Strand	Description
00_{16}	Y	RW	N	Y	Y Register
01_{16}	*Reserved*	—			Any access causes a *illegal_instruction* trap
02_{16}	CCR	RW	N	Y	Condition Code register
03_{16}	ASI	RW	N	Y	ASI register
04_{16}	TICK	RW	Y[1]	Partially	TICK register
05_{16}	PC	RO[2]	N	Y	Program counter
06_{16}	FPRS	RW	N	Y	Floating-Point Registers Status register
07_{16}–$0E_{16}$	*Reserved*	-		—	Any access causes an *illegal_instruction* trap
$0F_{16}$	(MEMBAR, STBAR, SIR)	—	N	—	Instruction opcodes only, not an actual ASR.
10_{16}	PCR	RW	Y[3]	Y	Performance counter control register
11_{16}	PIC	RW	Y[4]	Y	Performance instrumentation counter
12_{16}	*Reserved*	—			Any access causes an *illegal_instruction* trap
13_{16}	GSR	RW	N	Y	General Status register

TABLE 3-1 Summary of UltraSPARC T2 Ancillary State Registers *(Continued)*

Address	ASR Name	Access	priv	Replicated per-Strand	Description
14_{16}	SOFTINT_SET	W	Y[5]	Y	Set bit in Soft Interrupt register
15_{16}	SOFTINT_CLR	W	Y[5]	Y	Clear bit in Soft Interrupt register
16_{16}	SOFTINT	RW	Y[3]	Y	Soft Interrupt register
17_{16}	TICK_CMPR	RW	Y[3]	Y	TICK Compare register
18_{16}	STICK	RW	Y[6]	Partially	System Tick register
19_{16}	STICK_CMPR	RW	Y[3]	Y	System TICK Compare register
$1A_{16}$–$1F_{16}$	*Reserved*	—		—	Any access causes an *illegal_instruction* trap

Notes:

1. Nonprivileged software may read this register if the npt bit is 0. An attempt to read this register by nonprivileged software with npt = 1 causes a *privileged_action* trap. An attempted write by privileged software causes an *illegal_instruction* trap. An attempted write by nonprivileged software causes a *privileged_opcode* trap.
2. An attempted write to this register causes an *illegal_instruction* trap.
3. An attempted access in nonprivilged mode causes a *privileged_opcode* trap.
4. An attempted access in nonprivilged mode with PCR.priv = 1 causes a *privileged_action* trap.
5. Read accesses cause an *illegal_instruction* trap. An attempted write access in nonprivilged mode causes a *privileged_opcode* trap.
6. Nonprivileged software may read this register if the npt bit is 0. An attempt to read this register by nonprivileged software with npt = 1 causes a *privileged_action* trap. A write by privileged or user software causes an *illegal_instruction* trap.

3.1.1 Tick Register (TICK)

The TICK register contains two fields: npt and counter. The npt field is replicated per strand, while the counter field is shared by the eight strands on a physical core. Hyperprivileged software on any strand can write the TICK register. A write of the TICK register will update both the shared counter as well as the writing strand's npt field (the npt fields for other strands will be unaffected). The counter increments each processor core clock that `ASI_CMT_TICK_ENABLE.tick_enable` is set to 1. See *ASI_CMT_TICK_ENABLE* on page 189 for more details. On a warm reset, npt is set to

1, but counter continues counting (if ASI_CMT_TICK_ENABLE.tick_enable is 1) or remains unchanged (if ASI_CMT_TICK_ENABLE.tick_enable is 0). On all other resets, TICK does not change (other than the normal counting of counter).

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.2 Program Counter (PC)

Each strand has a read-only program counter register. The PC contains a 48-bit virtual address and VA{63:48} is sign-extended from VA{47}. The format of this register is shown in TABLE 3-2.

TABLE 3-2 Program Counter – PC (ASR 05_{16})

Bit	Field	Initial (POR) Value	R/W	Description
63:48	va_high	$FFFF_{16}$[1]	RO	Sign-extended from VA{47}.
47:2	va	$3FFFFC000008_{16}$[1]	RO	Virtual address contained in the program counter.
1:0	—	0	RO	The lower 2 bits of the program counter always read as 0.

1. Initial value listed is when ASI_RST_VEC_MASK.VEC_MASK = 0. If ASI_RST_VEC_MASK.VEC_MASK = 1, the initial value for the register is 20_{16}. See Section 20.1.4, *ASI_RST_VEC_MASK*, on page 448 for more details.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.3 Floating-Point State Register (FSR)

Each virtual processor has a Floating-Point State register. This register follows the UltraSPARC Architecture 2007 specification, with the ver field permanently set to 0 and the qne field permanently set to 0 (UltraSPARC T2 does not support a FQ).

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.4 General Status Register (GSR)

Each virtual processor has a nonprivileged general status register (GSR). When PSTATE.pef or FPRS.fef is zero, accesses to this register cause an *fp_disabled* trap.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.5 Software Interrupt Register (SOFTINT)

Each virtual processor has a privileged software interrupt register. Nonprivileged accesses to this register cause a *privileged_opcode* trap. The TICK_CMPR register contains three fields: sm, int_level, and tm. Note that while setting the sm (bit 16), tm (bit 0), and SOFTINT{14} bits all generate *interrupt_level_14*, these bits are considered completely independent of each other. Thus a STICK compare will only set bit 16 and generate *interrupt_level_14*, not also set bit 14.

TABLE 3-3 specifies how *interrupt_level_14* will be shared between SOFTINT writes, STICK compares, and TICK compares.

TABLE 3-3 Sharing of *interrupt_level_14*

Event	tm	SOFTINT{14}	sm	Action
STICK compare when sm = 0	Unchanged	Unchanged	1	*interrupt_level_14* if PSTATE.ie = 1 and PIL < 14
Set sm = 1 when sm = 0	Unchanged	Unchanged	1	*interrupt_level_14* if PSTATE.ie = 1 and PIL < 4
Set SOFTINT{14} = 1 when SOFTINT{14} = 0.	Unchanged	1	Unchanged	*interrupt_level_14* if PSTATE.ie = 1 and PIL < 4
TICK compare when tm = 0	1	Unchanged	Unchanged	*interrupt_level_14* if PSTATE.ie = 1 and PIL < 4
Set tm=1 when tm = 0	1	Unchanged	Unchanged	*interrupt_level_14* if PSTATE.ie = 1 and PIL < 4

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.6 Tick Compare Register (TICK_CMPR)

Each virtual processor has a privileged Tick compare register. Nonprivileged accesses to this register cause a *privileged_opcode* trap. The TICK_CMPR register contains two fields: int_dis and tick_cmpr. A full 63-bit tick_cmpr field is implemented in the register, but the bottom seven bits are ignored when comparing against the TICK counter field. The int_dis bit controls whether a TICK *interrupt_level_14* interrupt is posted in the SOFTINT register when tick_cmpr bits 62:7 match TICK bits 62:7.

Caution To reliably create *interrupt_level_14* interrupts using the tick compare register, software should ensure that the value written to bits 62:7 of the Tick Compare Register is larger than the value subsequently read from bits 62:7 of the TICK Register.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.7 System Tick Register (STICK)

STICK and TICK are derived from the same register. Writes to STICK affect TICK and vice versa.

Writes by user-level code to TICK generate a *privileged_opcode* trap, while writes by user-level code to STICK generate an *illegal_instruction* trap.

Reads of STICK.counter{6:0} are tied to $7F_{16}$. This prevents software from setting the System Tick Compare Register or Hyperprivileged System Tick Compare Register to a value that should cause a subsequent interrupt but that would not be detected due to the System Tick Compare Register and Hyperprivileged System Tick Compare implementation. The compare registers are not continuously compared to STICK, but are compared periodically (at least once every 128 cycles).

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.1.8 System Tick Compare Register (STICK_CMPR)

Each virtual processor has a privileged System Tick Compare (STICK_CMPR register. Nonprivileged accesses to this register cause a *privileged_opcode* trap. STICK_CMPR contains two fields: int_dis and stick_cmpr. A full 63-bit stick_cmpr field is implemented in the register, but the bottom seven bits are ignored when comparing against the STICK counter field. To assist software in reliably creating interrupt_level_14 interrupts using the system tick compare register, UltraSPARC T2 always returns reads of the system tick register with bits 6:0 set to 7'h7F. This ensures that if software writes a value to the system tick compare register that is greater than the value subsequently read from the system tick register that a match will occur in the future.

The int_dis bit controls whether a STICK *interrupt_level_14* interrupt is posted in the SOFTINT register when stick_cmpr bits 62:7 match STICK bits 62:7.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.2 Privileged PR State Registers

TABLE 3-4 lists the privileged registers.

TABLE 3-4 Privileged Registers

Register	Register Name	Access	Replicated per-Strand	Description
00_{16}	TPC	RW	Y	Trap PC[1]
01_{16}	TNPC	RW	Y	Trap Next PC[1]
02_{16}	TSTATE	RW	Y	Trap State
03_{16}	TT	RW	Y	Trap Type
04_{16}	TICK	RW	Partially	Tick
05_{16}	TBA	RW	Y	Trap Base Address[1]
06_{16}	PSTATE	RW	Y	Process State
07_{16}	TL	RW	Y	Trap Level
08_{16}	PIL	RW	Y	Processor Interrupt Level
09_{16}	CWP	RW	Y	Current Window Pointer
$0A_{16}$	CANSAVE	RW	Y	Savable Windows
$0B_{16}$	CANRESTORE	RW	Y	Restorable Windows
$0C_{16}$	CLEANWIN	RW	Y	Clean Windows
$0D_{16}$	OTHERWIN	RW	Y	Other Windows
$0E_{16}$	WSTATE	RW	Y	Window State
10_{16}	GL	RW	Y	Global Level

1. UltraSPARC T2 only implements bits 47:0 of the TPC, TNPC, and TBA registers. Bits 63:48 are always sign-extended from bit 47.

3.2.1 Trap State Register (TSTATE)

Each virtual processor has *MAXTL* (6) Trap State registers. These registers hold the state values from the previous trap level. The format of one element the TSTATE register array (corresponding to one trap level) is shown in TABLE 3-5.

TABLE 3-5 Trap State Register

Bit	Field	Initial (POR) Value	R/W	Description
63:42	—	0	RO	*Reserved*.
41:40	gl	0	RW	Global level at previous trap level
39:32	ccr	0	RW	CCR at previous trap level
31:24	asi	0	RW	ASI at previous trap level
23:21	—	0	RO	*Reserved*
20	pstate tct	0	RW	PSTATE.tct at previous trap level
19:18	—	0	RO	*Reserved* (corresponds to bits 11:10 of PSTATE)
17	pstate cle	0	RW	PSTATE.cle at previous trap level
16	pstate tle	0	RW	PSTATE.tle at previous trap level
15:13	—	0	RO	*Reserved* (corresponds to bits 7:5 of PSTATE)
12	pstate pef	0	RW	PSTATE.pef at previous trap level
11	pstate am	0	RW	PSTATE.am at previous trap level
10	pstate priv	0	RW	PSTATE.priv at previous trap level
9	pstate ie	0	RW	PSTATE.ie at previous trap level
8	—	0	RO	*Reserved* (corresponds to bit 0 of PSTATE)
7:3	—	0	RO	*Reserved*
2:0	cwp	0	RW	CWP from previous trap level

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.2.2 Processor State Register (PSTATE)

Each virtual processor has a Processor State register. More details on PSTATE can be found in the UltraSPARC Architecture 2007 specification. The format of this register is shown in TABLE 3-6; note that the memory model selection field (mm) mentioned in UltraSPARC Architecture 2007 is not implemented in UltraSPARC T2.

TABLE 3-6 Processor State Register

Bit	Field	Initial (POR) Value	R/W	Description
63:13	—	0	RO	*Reserved*
12	tct	0	RW	Trap on control transfer
11:10	—	0	RO	*Reserved*
9	cle	0	RW	Current little endian
8	tle	0	RW	Trap little endian

TABLE 3-6 Processor State Register

Bit	Field	Initial (POR) Value	R/W	Description
7:6	—	0	RO	*Reserved* (mm; not implemented in UltraSPARC T2)
5	—	0	RO	*Reserved* (was red)
4	pef	1	RW	Enable floating-point
3	am	0	RW	Address mask
2	priv	1	RW	Privileged mode
1	ie	0	RW	Interrupt enable
0	—	0	RO	*Reserved* (was ag)

Implementation Note Traps to hyperprivileged space will set PSTATE.priv to 0. PSTATE.priv could be set to either a 0 or 1 for this case, as HPSTATE.hpriv being a 1 overrides the setting in PSTATE.priv.

Programming Note Hyperprivileged changes to translation in delay slots of delayed control transfer instructions should be avoided; see Section 12.3.2, *Real-to-Physical Address Mapping and Speculative Instruction Fetch*, on page 115.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.2.3 Trap Level Register (TL)

Each virtual processor has a Trap Level register. Writes to this register saturate at *MAXPTL* (2) when in privileged mode and at *MAXTL* (6) in hyperprivileged mode. This saturation is based on bits 2:0 of the write data; bits 63:3 of the write data are ignored.

Note Hyperprivileged software can set TL to greater than *MAXPTL* for user or supervisor code by writing to TSTATE followed by a DONE/RETRY, doing a JMPL/WRHPR pair, etc. Operation of the UltraSPARC T2 chip when HPSTATE.hpriv = 0 and TL > *MAXPTL* follows UltraSPARC Architecture, and while in this state all traps destined for privileged level will instead be delivered to hyperprivileged level using the guest watchdog vector.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.2.4 Current Window Pointer (CWP) Register

Since *N_REG_WINDOWS* = 8 on UltraSPARC T2, the CWP register in each virtual processor is implemented as a 3-bit register.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.2.5 Global Level Register (GL)

Each virtual processor has a Global Level register, which controls which set of four global register windows is in use. The maximum global level (*MAXGL*) for UltraSPARC T2 is 3, so GL is implemented as a 2-bit register on UltraSPARC T2. GL is restricted to be less than or equal to *MAXPTL* (2) for privileged and nonprivileged code. On a trap, GL is set to min(GL + 1, *MAXGL*) for traps to hyperprivileged mode and to **min**(GL + 1,*MAXPTL*) for traps to privileged mode. On a DONE or RETRY, if executed with HTSTATE[TL].HPSTATE.hpriv = 1 (so that the DONE or RETRY places the virtual processor in hyperprivileged mode), the value of GL is restored from TSTATE[TL].gl.

Writes to the GL register saturate at *MAXPTL* when in privileged mode, and *MAXGL* in hyperprivileged mode. This saturation is based on bits 3:0 of the write data; bits 63:4 of the write data are ignored.

The format of the GL register is shown in TABLE 3-7.

TABLE 3-7 Global Level Register

Bit	Field	Initial (POR) Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	gl	3	RW	Global level.

Note | Hyperprivileged software can still set GL to greater than *MAXPTL* for nonprivileged or privileged code (although this is ***not*** recommended, except in diagnostic code) by doing a JMPL/WRHPR pair when GL > *MAXPTL*. The UltraSPARC T2 chip allows software normal access to the global registers when HPSTATE.hpriv = 0 and GL > *MAXPTL*.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.3 Hyperprivileged Registers

TABLE 3-8 shows the format of the UltraSPARC T2 hyperprivileged registers.

TABLE 3-8 Hyperprivileged Registers

Register	Register Name	Access	Replicated by Strand	Description
00_{16}	HPSTATE	RW	Y	Hypervisor Processor State register
01_{16}	HTSTATE	RW	Y	Hypervisor Trap State register
03_{16}	HINTP	RW	Y	Hypervisor Interrupt Pending register
05_{16}	HTBA	RW	Y	Hypervisor Trap Base Address register[1]
06_{16}	HVER	RO	N	Version register
$1E_{16}$	HALT	RW	Y	Halt instruction
$1F_{16}$	HSTICK_CMPR	RW	Y	Hypervisor System Tick Compare register

1. UltraSPARC T2 only implements bits 47:14 of the tba field. Bits 63:48 are always sign-extended from bit 47.

3.3.1 Hypervisor Processor State Register (HPSTATE)

Each virtual processor has a Hypervisor Processor State register, HPSTATE.

Full documentation on the Hypervisor Processor State register can be found in the UltraSPARC Architecture 2007 specification.

Note The tlz bit retains its current value when a trap is taken, which is different from the UltraSPARC Architecture specification, which specifies it is cleared when any trap is taken.

Programming Note Hyperprivileged changes to translation in delay slots of delayed control transfer instructions should be avoided; see Section 12.3.2, *Real-to-Physical Address Mapping and Speculative Instruction Fetch*, on page 115.

3.3.2 Hypervisor Trap State Register (HTSTATE)

Each virtual processor has a set of Hypervisor Trap State registers, one per trap level. These registers hold the hyperprivileged state values from the previous trap level. Full documentation on this register can be found in the UltraSPARC Architecture specification.

3.3.3 Hypervisor Interrupt Pending Register (HINTP)

Each virtual processor has a Hypervisor Interrupt Pending register. Full documentation on this register can be found in the UltraSPARC Architecture specification.

3.3.4 Hypervisor Trap Base Address Register (HTBA)

Each virtual processor has a Hypervisor Trap Base Address Register. Full documentation on this register can be found in the UltraSPARC Architecture specification.

Note | UltraSPARC T2 only implements bits 47:14 of the tba field of HTBA. Bits 63:48 are always sign-extended from bit 47.

3.3.5 Hyperprivileged Version Register (HVER)

The strands on a physical core share a read-only Version register. Writes to this register generate an *illegal_instruction* trap.

3.3.6 Hyperprivileged System Tick Compare Register (HSTICK_CMPR)

Each virtual processor has a Hyperprivileged System Tick Compare register. HSTICK_CMPR register contains two fields: int_dis and hstick_cmpr.

In the UltraSPARC T2 implementation, a full 63-bit hstick_cmpr field is implemented in the register but the bottom seven bits are ignored when comparing to the STICK counter field. To assist software in reliably creating *hstick_match* traps using the hyperprivileged system tick compare register, UltraSPARC T2 always returns reads of the system tick register with bits 6:0 set to $7F_{16}$ (all ones). This ensures that if software writes a value to HSTICK_CMPR that is greater than the value subsequently read from the system tick register, a match will occur in the future.

For more information on this register, see the UltraSPARC Architecture 2007 specification.

3.3.7 Halt

UltraSPARC T2 provides an iimplementation-specific “halt” pseudo-instruction that can place the virtual processor (strand) executing it into the `halt` state. The “halt” pseudo-instruction is encoded as a write (via WRHPR) to HPR $1E_{16}$ (the “halt”

psuedo-register) The virtual strand enters the `halt` state when:
(a) hyperprivileged software writes to HPR $1E_{16}$ and
(b) there are no pending interrupts or modes (as described below) that would prevent entering the `halt` state.

A RDHPR of HPR $1E_{16}$ returns zero.

The format of the "halt" pseudo-register is shown in Table 3-9 .

TABLE 3-9 "Halt" Pseudo-Register

Bit	Field	Initial (POR) Value	R/W	Description
63:0	—	0	RW	*Reserved.Reads return zero and write data is ignored.*

The operation of the Halt pseudo-instruction is as follows. The virtual processor can be parked, disabled, running, or halted. A virtual processor can be enabled or disabled by writing to ASI_CORE_ENABLE (see 14.1.3 on page 188). A virtual processor, when enabled, can either be parked or unparked, by writing to the ASI_CORE_RUNNING_RW register (see 14.1.7 on page 190). When enabled and unparked, the virtual processor is normally running. A running virtual core may be halted by writing to the Halt register. Once halted, the virtual processor remains halted until an interrupt arrives. When the interrupt arrives, the processor transitions back to running. It resumes execution at the NPC of the Halt instruction.

When halted, the virtual processor consumes no execution resource. It is similar to the parked state except that it awakens upon an interrupt.

If an interrupt arrives coincident with the execution of the Halt instruction, the virtual processor remains running and takes the interrupt.

The following interrupts transition a halted virtual processor back to the running state:

1. The virtual processor receives an interrupt from another virtual processor via the Interrupt Vector Dispatch Register(see 7.3.3 on page 60).
2. The virtual processor receives an XIR.
3. The virtual processor receives any disrupting or deferred error leading to a *software_recoverable_error* trap, *hardware_corrected_error* trap, or a deferred *store_error* trap. Note: ASI_SETER masking is not applied, so even if the virtual processors ASI_SETER masks the error trap, the virtual processor transitions to the running state.
4. The virtual processor sets softint[16] (softint.sm). Only stick_match can do this while in halted state. If softint[16] is 1 when the Halt instruction executes, the virtual processor remains in running state. Masking via PIL is not applied, so even if PIL masks the exception, the virtual processor transitions to running state.

5. The virtual processor sets softint[0] (softint.tm). Only tick_match can do this while in halted state. If softint[0] is 1 when the Halt instruction executes, the virtual processor remains in running state. Masking via PIL is not applied, so even if PIL masks the exception, the strand transitions to running state

6. The virtual processor receives an *hstick_match_interrupt*. If hintp is 1 when the Halt instruction executes, the strand remains in running state.

7. The virtual processor receives an *interrupt_vector* exception.

8. The virtual processor receives a park request, as a result of another virtual processor writing to the ASI_CORE_RUNNING_RW or ASI_CORE_RUNNING_W1C registers. In this case, the virtual processor transitions from halted to running to parked.

9. Entry to Single Step mode

10. Entry to Disable Overlap mode

Note: If the virtual processor is executing in Single Step or Disable Overlap mode and executes a Halt instruction, the virtual processor remains running in that mode. It does not enter halt state.

CHAPTER 4

Instruction Format

Instruction formats are described in the UltraSPARC Architecture 2006 specification.

CHAPTER 5

Instruction Definitions

5.1 Instruction Set Summary

The UltraSPARC T2 CPU implements the UltraSPARC Architecture 2007*UltraSPARC Architecture 2007* instruction set.

TABLE 5-1 lists the complete UltraSPARC T2 instruction set supported in hardware. All instructions are documented in the *UltraSPARC Architecture 2007* specification.

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(1 of 6)*

Opcode	Description
ADD (ADDcc)	Add (and modify condition codes)
ADDC (ADDCcc)	Add with carry (and modify condition codes)
ALIGNADDRESS	Calculate address for misaligned data access
ALIGNADDRESSL	Calculate address for misaligned data access (little-endian)
ALLCLEAN	Mark all windows as clean
AND (ANDcc)	And (and modify condition codes)
ANDN (ANDNcc)	And not (and modify condition codes)
ARRAY{8,16,32}	3-D address to blocked byte address conversion
Bicc	Branch on integer condition codes
BMASK	Writes the GSR.mask field
BPcc	Branch on integer condition codes with prediction
BPr	Branch on contents of integer register with prediction
BSHUFFLE	Permutes bytes as specified by the GSR.mask field
CALL[1]	Call and link
CASA	Compare and swap word in alternate space
CASXA	Compare and swap doubleword in alternate space
DONE	Return from trap
EDGE{8,16,32}{L}{N}	Edge boundary processing {little-endian} {non-condition-code altering}

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(2 of 6)*

Opcode	Description
FABS(s,d)	Floating-point absolute value
FADD(s,d)	Floating-point add
FALIGNDATA	Perform data alignment for misaligned data
FANDNOT1{s}	Negated *src1* **and** *src2* (single precision)
FANDNOT2{s}	*Src1* **and** negated *src2* (single precision)
FAND{s}	Logical **and** (single precision)
FBPfcc	Branch on floating-point condition codes with prediction
FBfcc	Branch on floating-point condition codes
FCMP(s,d)	Floating-point compare
FCMPE(s,d)	Floating-point compare (exception if unordered)
FCMPEQ{16,32}	Four 16-bit / two 32-bit compare: set integer dest if *src1* = *src2*
FCMPGT{16,32}	Four 16-bit / two 32-bit compare: set integer dest if *src1* > *src2*
FCMPLE{16,32}	Four 16-bit / two 32-bit compare: set integer dest if *src1* ≤ *src2*
FCMPNE{16,32}	Four 16-bit / two 32-bit compare: set integer dest if *src1* ≠ *src2*
FDIV(s,d)	Floating-point divide
FEXPAND	Four 8-bit to 16-bit expand
FiTO(s,d)	Convert integer to floating-point
FLUSH	Flush instruction memory
FLUSHW	Flush register windows
FMOV(s,d)	Floating-point move
FMOV(s,d)cc	Move floating-point register if condition is satisfied
FMOV(s,d)R	Move floating-point register if integer register contents satisfy condition
FMUL(s,d)	Floating-point multiply
FMUL8SUX16	Signed upper 8- x 16-bit partitioned product of corresponding components
FMUL8ULX16	Unsigned lower 8- x 16-bit partitioned product of corresponding components
FMUL8X16	8- x 16-bit partitioned product of corresponding components
FMUL8X16AL	Signed lower 8- x 16-bit lower α partitioned product of four components
FMUL8X16AU	Signed upper 8- x 16-bit lower α partitioned product of four components
FMULD8SUX16	Signed upper 8- x 16-bit multiply → 32-bit partitioned product of components
FMULD8ULX16	Unsigned lower 8- x 16-bit multiply → 32-bit partitioned product of components
FNAND{s}	Logical **nand** (single precision)
FNEG(s,d)	Floating-point negate
FNOR{s}	Logical **nor** (single precision)
FNOT1{s}	Negate (1's complement) *src1* (single precision)
FNOT2{s}	Negate (1's complement) *src2* (single precision)
FONE{s}	One fill (single precision)
FORNOT1{s}	Negated *src1* **or** *src2* (single precision)

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(3 of 6)*

Opcode	Description
FORNOT2{s}	*src1* **or** negated *src2* (single precision)
FOR{s}	Logical **or** (single precision)
FPACKFIX	Two 32-bit to 16-bit fixed pack
FPACK{16,32}	Four 16-bit/two 32-bit pixel pack
FPADD{16,32}{s}	Four 16-bit/two 32-bit partitioned add (single precision)
fPMERGE	Two 32-bit to 64-bit fixed merge
FPSUB{16,32}{s}	Four 16-bit/two 32-bit partitioned subtract (single precision)
FsMULd	Floating-point multiply single to double
FSQRT(s,d)	Floating-point square root
FSRC1{s}	Copy *src1* (single precision)
FSRC2{s}	Copy *src2* (single precision)
F(s,d)TO(s,d)	Convert between floating-point formats
F(s,d)TOi	Convert floating point to integer
F(s,d)TOx	Convert floating point to 64-bit integer
FSUB(s,d)	Floating-point subtract
FXNOR{s}	Logical **xnor** (single precision)
FXOR{s}	Logical **xor** (single precision)
FxTO(s,d)	Convert 64-bit integer to floating-point
FZERO{s}	Zero fill (single precision)
ILLTRAP	Illegal instruction
INVALW	Mark all windows as CANSAVE
JMPL	Jump and link
LDBLOCKF	64-byte block load
LDDF	Load double floating-point
LDDFA	Load double floating-point from alternate space
LDF	Load floating-point
LDFA	Load floating-point from alternate space
LDFSR	Load floating-point state register lower
LDSB	Load signed byte
LDSBA	Load signed byte from alternate space
LDSH	Load signed halfword
LDSHA	Load signed halfword from alternate space
LDSTUB	Load-store unsigned byte
LDSTUBA	Load-store unsigned byte in alternate space
LDSW	Load signed word
LDSWA	Load signed word from alternate space
LDTW	Load twin words

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(4 of 6)*

Opcode	Description
LDTWA	Load twin words from alternate space
LDUB	Load unsigned byte
LDUBA	Load unsigned byte from alternate space
LDUH	Load unsigned halfword
LDUHA	Load unsigned halfword from alternate space
LDUW	Load unsigned word
LDUWA	Load unsigned word from alternate space
LDX	Load extended
LDXA	Load extended from alternate space
LDXFSR	Load extended floating-point state register
MEMBAR	Memory barrier
MOVcc	Move integer register if condition is satisfied
MOVr	Move integer register on contents of integer register
MULScc	Multiply step (and modify condition codes)
MULX	Multiply 64-bit integers
NOP	No operation
NORMALW	Mark other windows as restorable
OR (ORcc)	Inclusive-or (and modify condition codes)
ORN (ORNcc)	Inclusive-or not (and modify condition codes)
OTHERW	Mark restorable windows as other
PDIST	Distance between 8 8-bit components
POPC	Population count
PREFETCH	Prefetch data
PREFETCHA	Prefetch data from alternate space
PST	Eight 8-bit/4 16-bit/2 32-bit partial stores
RDASI	Read ASI register
RDASR	Read ancillary state register
RDCCR	Read condition codes register
RDFPRS	Read floating-point registers state register
RDHPR	Read hyperprivileged register
RDPC	Read program counter
RDPR	Read privileged register
RDTICK	Read TICK register
RDY	Read Y register
RESTORE	Restore caller's window
RESTORED	Window has been restored
RETRY	Return from trap and retry

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(5 of 6)*

Opcode	Description
RETURN	Return
SAVE	Save caller's window
SAVED	Window has been saved
SDIV (SDIVcc)	32-bit signed integer divide (and modify condition codes)
SDIVX	64-bit signed integer divide
SETHI	Set high 22 bits of low word of integer register
SIAM	Set interval arithmetic mode
SIR	Software-initiated reset
SLL	Shift left logical
SLLX	Shift left logical, extended
SMUL (SMULcc)	Signed integer multiply (and modify condition codes)
SRA	Shift right arithmetic
SRAX	Shift right arithmetic, extended
SRL	Shift right logical
SRLX	Shift right logical, extended
STB	Store byte
STBA	Store byte into alternate space
STBAR	Store barrier
STBLOCKF	64-byte block store
STDF	Store double floating-point
STDFA	Store double floating-point into alternate space
STF	Store floating-point
STFA	Store floating-point into alternate space
STFSR	Store floating-point state register
STH	Store halfword
STHA	Store halfword into alternate space
STTW	Store twin words
STTWA	Store twin words into alternate space
STW	Store word
STWA	Store word into alternate space
STX	Store extended
STXA	Store extended into alternate space
STXFSR	Store extended floating-point state register
SUB (SUBcc)	Subtract (and modify condition codes)
SUBC (SUBCcc)	Subtract with carry (and modify condition codes)
SWAP	Swap integer register with memory
SWAPA	Swap integer register with memory in alternate space

TABLE 5-1 Complete UltraSPARC T2 Hardware-Supported Instruction Set *(6 of 6)*

Opcode	Description
TADDcc (TADDccTV)	Tagged add and modify condition codes (trap on overflow)
TSUBcc (TSUBccTV)	Tagged subtract and modify condition codes (trap on overflow)
Tcc	Trap on integer condition codes (with 8-bit sw_trap_number, if bit 7 is set trap to hyperprivileged)
UDIV (UDIVcc)	Unsigned integer divide (and modify condition codes)
UDIVX	64-bit unsigned integer divide
UMUL (UMULcc)	Unsigned integer multiply (and modify condition codes)
WRASI	Write ASI register
WRASR	Write ancillary state register
WRCCR	Write condition codes register
WRFPRS	Write floating-point registers state register
WRHPR	Write hyperprivileged register
WRPR	Write privileged register
WRY	Write Y register
XNOR (XNORcc)	Exclusive-nor (and modify condition codes)
XOR (XORcc)	Exclusive-or (and modify condition codes)

1. The PC format saved by the CALL instruction is the same as the format of the PC register specified in Section 3.1.2, *Program Counter (pc)*, on page 13.

TABLE 5-2 lists the SPARC V9 and sun4v instructions that are not directly implemented in hardware by UltraSPARC T2, and the exception that occurs when an attempt is made to execute it.

TABLE 5-2 UltraSPARC Architecture 2007 Instructions Not Directly Implemented by UltraSPARC T2 Hardware *(1 of 2)*

Opcode	Description	Exception
FABSq	Floating-point absolute value quad	*illegal_instruction*
FADDq	Floating-point add quad	*illegal_instruction*
FCMPq	Floating-point compare quad	*illegal_instruction*
FCMPEq	Floating-point compare quad (exception if unordered)	*illegal_instruction*
FDIVq	Floating-point divide quad	*illegal_instruction*
FdMULq	Floating-point multiply double to quad	*illegal_instruction*
FiTOq	Convert integer to quad floating-point	*illegal_instruction*
FMOVq	Floating-point move quad	*illegal_instruction*
FMOVqcc	Move quad floating-point register if condition is satisfied	*illegal_instruction*

TABLE 5-2 UltraSPARC Architecture 2007 Instructions Not Directly Implemented by UltraSPARC T2 Hardware *(2 of 2)*

Opcode	Description	Exception
FMOVqr	Move quad floating-point register if integer register contents satisfy condition	*illegal_instruction*
FMULq	Floating-point multiply quad	*illegal_instruction*
FNEGq	Floating-point negate quad	*illegal_instruction*
FSQRTq	Floating-point square root quad	*illegal_instruction*
F(s,d,q)TO(q)	Convert between floating-point formats to quad	*illegal_instruction*
FQTOI	Convert quad floating point to integer	*illegal_instruction*
FQTOX	Convert quad floating point to 64-bit integer	*illegal_instruction*
FSUBq	Floating-point subtract quad	*illegal_instruction*
FxTOq	Convert 64-bit integer to floating-point	*illegal_instruction*
IMPDEP1 (not listed in TABLE 5-1)	Implementation-dependent instruction	*illegal_instruction*
IMPDEP2 (not listed in TABLE 5-1)	Implementation-dependent instruction	*illegal_instruction*
LDQF	Load quad floating-point	*illegal_instruction*
LDQFA	Load quad floating-point into alternate space	*illegal_instruction*
STQF	Store quad floating-point	*illegal_instruction*
STQFA	Store quad floating-point into alternate space	*illegal_instruction*

5.2 UltraSPARC T2-Specific Instructions

5.3 Block Load and Store Instructions

See the LDBLOCKF and STBLOCKF instruction descriptions in the *UltraSPARC Architecture 2007* specification for the standard definitions of these instructions.

Block loads are not allowed to IO space (which is indicated on UltraSPARC T2 by PA{39} = 1).

A block load to IO space generates a *DAE_nc_page* trap.

Block stores to IO space are permitted.

Block store commits in UltraSPARC T2 do NOT force the data to be written to memory as specified in the *UltraSPARC Architecture 2007* specification. Block store commits are implemented the same as block stores in UltraSPARC T2. As with all stores, block stores and block store commits will maintain coherency with all I-caches, but will not flush any modified instructions executing down a pipeline. Flushing those instructions requires the pipeline to execute a FLUSH instruction.

Notes

If LDBLOCKF is used with an `ASI_BLK_COMMIT_{P,S}` and a destination register number rd is specified which is not a multiple of 8 (a misaligned `rd`), UltraSPARC T2 generates an *illegal_instruction* exception (impl. dep. #255-U3-Cs10).

If LDBLOCKF is used with an `ASI_BLK_COMMIT_{P,S}` and a memory address is specified with less than 64-byte alignment, UltraSPARC T2 generates a *mem_address_not_aligned* exception (impl. dep. #256-U3)

These instructions are used for transferring large blocks of data (more than 256 bytes); for example, `bcopy()` and `bfill()`. On UltraSPARC T2, a block load forces a miss in the primary cache and will not allocate a line in the primary cache, but does allocate in L2.

UltraSPARC T2 treats block loads as interlocked with respect to following instructions. That is, all floating-point registers are updated before any subsequent instruction issues.

STBLOCKF source data registers are interlocked against completion of previous instructions, including block load instructions.

LDBLOCKF does not follow memory model ordering with respect to stores. In particular, read-after-write hazards to overlapping addresses are not detected. The side-effect bit associated with the access is ignored (see *Translation Table Entry (TTE)* on page 105). If ordering with respect to earlier stores is important (for example, a block load that overlaps previous stores), then there must be an intervening MEMBAR #`StoreLoad` or stronger MEMBAR. If the LDBLOCKF overlaps a previous store and there is no intervening MEMBAR or data reference, the LDBLOCKF may return data from before or after the store.

Compatibility Note

Prior UltraSPARC implementations may have provided the first two registers at the same time. If code depends upon this unsupported behavior it must be modified for UltraSPARC T2.

STBLOCKF does not follow memory model ordering with respect to loads, previous block stores, or subsequent stores. (UltraSPARC T2 orders block stores with respect to previous nonblock stores). In particular, read-after-write hazards to overlapping addresses are not detected. The side-effects bit associated with the access is ignored.

If ordering with respect to later loads is important then there must be an intervening MEMBAR instruction. If the STBLOCKF overlaps a later load and there is no intervening MEMBAR `#StoreLoad` instruction, the contents of the block are undefined.

Compatibility Notes

Block load and store operations do not obey the ordering restrictions of the currently selected processor memory model (TSO, PSO, or RMO); block operations always execute under an RMO memory ordering model. In general, explicit MEMBAR instructions are required to order block operations among themselves or with respect to normal loads and stores. In addition, block operations do not generally conform to dependence order on the issuing virtual processor; that is, no read-after-write or write-after-read checking occurs between block loads and stores. Explicit MEMBARs are required to enforce dependence ordering between block operations that reference the same address. However, UltraSPARC T2 partially orders some block operations.

TABLE 5-3 describes the synchronization primitives required in UltraSPARC T2, if any, to guarantee TSO ordering between various sequences of memory reference operations. The first column contains the reference type of the first or earlier instruction; the second column contains the reference type of the second or the later instruction. UltraSPARC T2 orders loads and block loads against all subsequent instructions.

TABLE 5-3 UltraSPARC T2 Synchronization Requirements for Memory Reference Operations

First reference	Second reference	Synchronization Required
Load	Load	—
	Block load	—
	Store	—
	Block store	—
Block load	Load	—
	Block load	—
	Store	—
	Block store	—

TABLE 5-3 UltraSPARC T2 Synchronization Requirements for Memory Reference Operations

First reference	Second reference	Synchronization Required
Store	Load	—
	Block load	MEMBAR #StoreLoad or #Sync
	Store	—
	Block store	—
Block store	Load	MEMBAR #StoreLoad or #Sync
	Block load	MEMBAR #StoreLoad or #Sync
	Store	MEMBAR #Sync
	Block store	MEMBAR #Sync

Block Initializing Store ASIs

Instruction	imm_asi	ASI Value	Operation
ST[B,H,W,TW,X]A	ASI_ST_BLKINIT_AS_IF_USER_PRIMARY (ASI_STBI_AIUP)	22_{16}	64-byte block initialing store to primary address space, user privilege
	ASI_ST_BLKINIT_AS_IF_USER_SECONDARY (ASI_STBI_AIUS)	23_{16}	64-byte block initialing store to secondary address space, user privilege
	ASI_ST_BLKINIT_NUCLEUS (ASI_STBI_N)	27_{16}	64-byte block initialing store to nucleus address space
	ASI_ST_BLKINIT_AS_IF_USER_PRIMARY_LITTLE (ASI_STBI_AIUPL)	$2A_{16}$	64-byte block initialing store to primary address space, user privilege, little-endian
	ASI_ST_BLKINIT_AS_IF_USER_SECONDARY_LITTLE (ASI_STBI_AIUS_L)	$2B_{16}$	64-byte block initialing store to secondary address space, user privilege, little-endian
	ASI_ST_BLKINIT_NUCLEUS_LITTLE (ASI_STBI_NL)	$2F_{16}$	64-byte block initialing store to nucleus address space, little-endian
	ASI_ST_BLKINIT_PRIMARY (ASI_STBI_P)	$E2_{16}$	64-byte block initialing store to primary address space

Instruction	imm_asi	ASI Value	Operation
	ASI_ST_BLKINIT_SECONDARY (ASI_STBI_S)	$E3_{16}$	64-byte block initialing store to secondary address space
	ASI_ST_BLKINIT_PRIMARY_LITTLE (ASI_STBI_PL)	EA_{16}	64-byte block initialing store to primary address space, little-endian
	ASI_ST_BLKINIT_SECONDARY_LITTLE (ASI_STBI_SL)	EB_{16}	64-byte block initialing store to secondary address space, little-endian

Description

Block initializing store instructions are selected by using one of the block initializing store ASIs with integer store instructions. These ASIs allow block initializing stores to be performed to the same address spaces as normal stores. Little-endian ASIs access data in little-endian format, otherwise the access is assumed to be big-endian.

Integer stores of all sizes (to alternate space) are allowed to use these ASIs

All stores to these ASIs operate under relaxed memory ordering (RMO), regardless of the PSTATE.mm setting, and software must follow a sequence of these stores with a MEMBAR #Sync to ensure ordering with respect to subsequent loads and stores. Stores to these ASIs where the least-significant 6 bits of the address are non-zero (that is, not the first word in the cache line) behave the same as a normal RMO store. A store to these ASIs where the least-significant 6 bits are zero will load a cache line in the L2 cache with either all zeros or the existing data, and then update that line with the new store data. This special store will make sure the line maintains coherency when it is loaded into the cache, but will not generally fetch the line from memory (initializing it with zeros instead). Stores using these ASIs to a noncacheable address (PA{39} = 1) will behave the same as a normal store.

Note These instructions are used for transferring large blocks of data (more than 256 bytes); for example, `bcopy()` and `bfill()`. On UltraSPARC T2, a quad load forces a miss in the primary cache and will not allocate a line in the primary cache, but does allocate in L2.

Access to these ASIs by a floating-point store (STFA, STDFA) will result in a *DAE_invalid_ASI* trap (or *mem_address_not_aligned* trap if not properly aligned for the store size).

The following pseudocode shows how these ASIs can be used to do a quadword aligned (on both source and destination) copy of N quadwords from *A* to *B* (where N > 3). Note that the final 64 bytes of the copy is performed using normal stores, guaranteeing that all initial zeros in a cache line are overwritten with copy data.

```
%l0 ← [A]
%l1 ← [B]
prefetch [%l0]
for (i = 0; i < N-4; i++) {
```

```
    if (!(i % 4)) {
       prefetch [%l0+64]
    }
    ldtxa [%l0] #ASI_TWINX_P, %l2
    add %l0, 16, %l0
    stxa %l2, [%l1] #ASI_ST_BLKINIT_PRIMARY
    add %l1, 8, %l1
    stxa %l3, [%l1] #ASI_ST_BLKINIT_PRIMARY
    add %l1, 8, %l1
}
for (i = 0; i < 4; i++) {
    ldtxa [%l0] #ASI_TWINX_P, %l2
    add %l0, 16, %l0
    stx %l2, [%l1]
    stx %l3,d [%l1+8]
    add %l1, 16, %l1
}
membar #Sync
```

Programming Notes

These ASIs are specific to UltraSPARC T2 to provide a high-performance bcopy alternative to block load and store (which fetch the lined stored to from memory to the L2 cache, requiring three memory operations for `bcopy()` and two memory operations for a `bfill()`). These ASIs are of Class "N" and are only allowed in dynamically linked, platform-specific, OS-enabled libraries.

These ASIs provide a higher-performance `bcopy()` or `bfill()` than LDBLOCKF and STBLOCKF, due to their ability to avoid the unnecessary fetch from memory of the data that is overwritten by the store.

5.3.1 Load Twin Extended Word

Load Twin Extended Word instructions are not allowed to access IO space (indicated by PA{39} = 1 on UltraSPARC T2). An LDTXA to IO space generates a *DAE_nc_page* trap.

CHAPTER 6

Traps

6.1 Trap Levels

Each virtual processor supports six trap levels (*MAXTL* = 6). Traps to privileged mode (supervisor software) while in privileged mode when TL = *MAXPTL* (2) will trap instead to the hyperprivileged mode (hypervisor software), using the guest watchdog vector in the hyperprivileged trap table. TL will be incremented, but the processor will not enter `RED_state` and the trap type will be set to that of the trap that caused the event, not the watchdog trap type.

6.2 Trap Behavior

TABLE 6-1 specifies the codes used in the tables below.

TABLE 6-1 Table Codes

Code	Meaning
H	Trap is taken via the Hyperprivileged trap table, in Hyperprivileged mode (HSTATE.hpriv = 1)
P	Trap is taken via the Privileged trap table, in Privileged mode (PSTATE.priv = 1)

TABLE 6-1 Table Codes

Code	Meaning
H^U	Trap is taken via the Hyperprivileged trap table, in Hyperprivileged mode (HSTATE.hpriv = 1). However, the trap is *unexpected*. While hardware can legitimately generate this trap, it should not do so unless there is a programming error or some other error. Therefore, occurrence of this trap indicates an actual error to hyperprivileged software.
-x-	Not possible. Hardware cannot generate this trap in the indicated running mode. For example, all privileged instructions can be executed in both privileged and hyperprivileged modes, therefore a *privileged_opcode* trap cannot occur in privileged or hyperprivileged mode.
—	This trap can only legitimately be generated by hyperprivileged software, not by the CPU hardware. So, for the purposes of sun4v, the trap vector has to be correct, but for a hardware CPU implementation these trap types are not generated by the hardware, therefore the resultant running mode is irrelevant.

For example, trap 1 ("power on reset") in TABLE 6-2, if delivered in any running mode, results in a delivery directly to the hypervisor mode.

TABLE 6-2 Trap Behavior *(1 of 3)*

			From privilege level:		
TT #	Hardware Trap Name	Priority	Nonprivileged	Privileged	Hyperprivileged
0_{16}	*Reserved*	—	—	—	—
1_{16}	*power_on_reset*	0	H	H	H
2_{16}	*watchdog_reset*	variable[2]	H	H	H
	guest watchdog	variable[3]	H	H	—
3_{16}	*externally_initiated_reset*	1.1	H	H	H
4_{16}	*software_initiated_reset*	1.3	-x-	-x-	H
5_{16}	*RED_state_exception*	1.4	H	H	H
6_{16}	*Reserved*	—	—	—	—
7_{16}	*store_error*	2.1	H	H	H
8_{16}	*IAE_privilege_violation*	3.1	H	-x-	-x-
9_{16}	*instruction_access_MMU_miss*	2.8	H	H	-x-[9]
A_{16}	*instruction_access_error*	4	H	H	H
B_{16}	*IAE_unauth_access*	2.9[4]	H	H	H^U
C_{16}	*IAE_nfo_page*	3.3	H	H	H^U
D_{16}	*instruction_address_range*	2.6	H	H	H^U
E_{16}	*instruction_real_range*	2.6	H	H	H^U
F_{16}	*Reserved*	—	—	—	—
10_{16}	*illegal_instruction*	6.1[5]	H	H	H
11_{16}	*privileged_opcode*	7	P	-x-	-x-
12_{16}	*unimplemented_LDTW*	—	—	—	—
13_{16}	*unimplemented_STTW*	—	—	—	—

TABLE 6-2 Trap Behavior *(2 of 3)*

TT #	Hardware Trap Name	Priority	From privilege level: Nonprivileged	Privileged	Hyperprivileged
14_{16}	*DAE_invalid_asi*	12.1	H	H	H^U
15_{16}	*DAE_privilege_violation*	12.4	H	H	H^U
16_{16}	*DAE_nc_page*	12.5	H	H	H^U
17_{16}	*DAE_nfo_page*	12.6	H	H	H^U
18_{16}–$1F_{16}$	*Reserved*	—	—	—	—
20_{16}	*fp_disabled*	8.1	P	P	H^U
21_{16}	*fp_exception_ieee_754*	11.1	P	P	H^U
22_{16}	*fp_exception_other*	11.1	P	P	H^U
23_{16}	*tag_overflow*	14	P	P	H^U
24_{16}–27_{16}	*clean_window*	10.1	P	P	H^U
28_{16}	*division_by_zero*	15	P	P	H^U
29_{16}	*internal_processor_error*	8.2 or 12.10[6]	H	H	H
$2A_{16}$	*instruction_invalid_TSB_entry*	2.10[7]	H	H	-x-
$2B_{16}$	*data_invalid_TSB_entry*	12.3	H	H	H
$2C_{16}$	*Reserved*	—	—	—	—
$2D_{16}$	*mem_real_range*	11.3	H	H	H^U
$2E_{16}$	*mem_address_range*	11.3	H	H	H^U
$2F_{16}$	*Reserved*	—	—	—	—
30_{16}	*DAE_so_page*	12.6	H	H	H^U
31_{16}	*data_access_MMU_miss*	12.3	H	H	H
32_{16}	*data_access_error*	12.9	H	H	H
33_{16}	*data_access_protection*	—	—	—	—
34_{16}	*mem_address_not_aligned*	10.2	H	H	H^U
35_{16}	*LDDF_mem_address_not_aligned*	10.1	H	H	H^U
36_{16}	*STDF_mem_address_not_aligned*	10.1	H	H	H^U
37_{16}	*privileged_action*	11.1	H	H	-x-
38_{16}	*LDQF_mem_address_not_aligned*	—	—	—	—
39_{16}	*STQF_mem_addess_not_aligned*	—	—	—	—
$3A_{16}$	*Reserved*	—	—	—	—
$3B_{16}$	*unsupported_page_size*	13	H	H	H^U
$3C_{16}$	*control_word_queue_interrupt*	16.5	H	H	H
$3D_{16}$	*modular_arithmetic_interrupt*	16.4	H	H	H
$3E_{16}$	*inst_real_translation_miss*	2.8	H	H	N[9]
$3F_{16}$	*data_real_translation_miss*	12.3	H	H	H
40	*sw_recoverable_error*	33.1	H	H	H
41_{16}–$4F_{16}$	*interrupt_level_n*	$32 - n$	P	P	-x-

TABLE 6-2 Trap Behavior *(3 of 3)*

TT #	Hardware Trap Name	Priority	From privilege level: Nonprivileged	Privileged	Hyperprivileged
$4F_{16}$	*pic_overflow* (*interrupt_level_15*)	16.0, variable[8]	P	P	-x-
50_{16}–$5D_{16}$	*Reserved*	—	—	—	—
$5E_{16}$	*hstick_match*	16.1	H	H	H
$5F_{16}$	*trap_level_zero*	2.2	H	H	-x-
60_{16}	*interrupt_vector_trap*	16.3	H	H	H
61_{16}	*PA_watchpoint*	12.8	H	H	H
62_{16}	*VA_watchpoint*	11.2	H	H	-x-
63_{16}	*hw_corrected_error*	33.2	H	H	H
64_{16}[1]	*fast_instruction_access_MMU_miss*	2.8	H	H	-x-
68_{16}[1]	*fast_data_access_MMU_miss*	12.3	H	H	H
$6C_{16}$[1]	*fast_data_access_protection*	12.7	H	H	H
70_{16}	*Reserved*	—	—	—	—
71_{16}	*instruction_access_MMU_error*	2.7	H	H	-x-
72_{16}	*data_access_MMU_error*	12.2	H	H	H
73_{16}	*Reserved*	—	—	—	—
74_{16}	*control_transfer_instruction*	11.1	P	P	H
75_{16}	*instruction_VA_watchpoint*	2.5	H	H	-x-
76_{16}	*instruction_breakpoint*	6.2[5]	H	H	H
77_{16}–$7B_{16}$	*Reserved*	—	—	—	—
$7C_{16}$	*cpu_mondo_trap*	16.6	P	P	-x-
$7D_{16}$	*dev_mondo_trap*	16.7	P	P	-x-
$7E_{16}$	*resumable_error*	33.3	P	P	-x-
$7F_{16}$	*nonresumable_error* (generated by software only)	—	—	—	—
80_{16}–$9C_{16}$[1]	*spill_n_normal* (n = 0–7)	9	P	P	H^U
A0–BC_{16}[1]	*spill_n_other* (n = 0–7)	9	P	P	H^U
$C0_{16}$–DC_{16}[1]	*fill_n_normal* (n = 0–7)	9	P	P	H^U
$E0_{16}$–FC_{16}[1]	*fill_n_other* (n = 0–7)	9	P	P	H^U
100_{16}–$17F_{16}$	*trap_instruction*	16.2	P	P	H
180_{16}–$1FF_{16}$	*htrap_instruction*	16.2	-x-	H	H^U

1. Trap extends across four TT #s to allow trap handler to contain 32 inline instructions instead of the standard 8 inline instructions.
2. The *watchdog_reset* priority is inherited from the underlying exception.
3. The *guest_watchdog* priority is inherited from the underlying exception.
4. UltraSPARC T2 deviates from the 3.2 priority in UltraSPARC Architecture 2007 for *IAE_unauth_access*.

5. UltraSPARC T2 deviates from UltraSPARC Architecture 2007 and swaps the priority of *illegal_instruction* (6.2 in UltraSPARC Architecture 2007) and *instruction_breakpoint* (6.1 in UltraSPARC Architecture 2007)
6. IRF/FRF errors that are encountered on the second or subsequent passes of a multicycle operation (partial store, compare and swap, block store) generate an *internal_processor_error* trap at priority 12.10. All other *internal_processor_error* traps are priority 8.2. See *Error Trap Vectors* on page 212.
7. UltraSPARC T2 deviates from the UltraSPARC Architecture 2007 priority of 2.8 for *instruction_invalid_TSB_entry*.
8. To make the *pic_overflow* trap the highest-priority disrupting trap, *pic_overflow* has an elevated priority over a normal *interrupt_level_15* trap (such as would be generated by writing 1 to SOFTINT{15}). Per Section 10.2, *SPARC Performance Instrumentation Counter*, on page 90, if the *pic_overflow* trap is taken on the instruction that caused the overflow, then the effective priority of the *pic_overflow* inherits from the condition that caused the overflow.
9. Hyperprivileged instruction access always bypasses translation. See *Translation* on page 128.

6.3 Trap Masking

TABLE 6-3 specifies the codes used inTABLE 6-3.

TABLE 6-3 Codes

Code	Meaning
(nm)	Never Masked — when the condition occurs in this running mode, it is never masked out and the trap is always taken.
(ie)	When the outstanding disrupting trap condition occurs in this privilege mode, it may be conditioned (masked out) by PSTATE.ie = 0 (but remains pending).
PIL	Masked by PSTATE.ie and PIL
tct	Masked by PSTATE.tct
ibe	Masked by HPSTATE.ibe
tlz	Masked by HPSTATE.tlz

TABLE 6-3 Codes *(Continued)*

Code	Meaning
M	Always masked
—	This trap can only legitimately be generated by hyperprivileged software, not by the CPU hardware. So, for the purposes of sun4v, the trap vector has to be correct, but for a hardware CPU implementation these trap types are not generated by the hardware, therefore the resultant running mode is irrelevant.

For example, trap $7C_{16}$ ("cpu mondo") in TABLE 6-4 is masked by PSTATE.ie in nonprivileged and privileged mode and is always masked in hyperprivileged mode.

TABLE 6-4 lists the trap mask behavior.

TABLE 6-4 Trap Mask Behavior *(1 of 3)*

			From privilege level:		
TT #	**Hardware Trap Name**	**Type**	**Nonprivileged**	**Privileged**	**Hyperprivileged**
0_{16}	*Reserved*	Reset	(nm)	(nm)	(nm)
1_{16}	*power_on_reset*	Reset	(nm)	(nm)	(nm)
2_{16}	*watchdog_reset*	Reset	(nm)	(nm)	(nm)
	guest watchdog	Reset	(nm)	(nm)	—
3_{16}	*externally_initiated_reset*	Reset	(nm)	(nm)	(nm)
4_{16}	*software_initiated_reset*	Reset	(nm)	(nm)	(nm)
5_{16}	*RED_state_exception*	Reset	(nm)	(nm)	(nm)
6_{16}	*Reserved*	—	—	—	—
7_{16}	*store_error*	Deferred	(nm)	(nm)	(nm)
8_{16}	*IAE_privilege_violation*	Precise	(nm)	(nm)	(nm)
9_{16}	*instruction_access_MMU_miss*	Precise	(nm)	(nm)	—
A_{16}	*instruction_access_error*	Precise	(nm)	(nm)	(nm)
B_{16}	*IAE_unauth_access*	Precise	(nm)	(nm)	(nm)
C_{16}	*IAE_nfo_page*	Precise	(nm)	(nm)	(nm)
D_{16}	*instruction_address_range*	Precise	(nm)	(nm)	(nm)
E_{16}	*instruction_real_range*	Precise	(nm)	(nm)	(nm)
F_{16}	*Reserved*	—	—	—	—
10_{16}	*illegal_instruction*	Precise	(nm)	(nm)	(nm)
11_{16}	*privileged_opcode*	Precise	(nm)	—	—
12_{16}	*unimplemented_LDTW*	—	—	—	—
13_{16}	*unimplemented_STTW*	—	—	—	—
14_{16}	*DAE_invalid_asi*	Precise	(nm)	(nm)	(nm)
15_{16}	*DAE_privilege_violation*	Precise	(nm)	(nm)	(nm)
16_{16}	*DAE_nc_page*	Precise	(nm)	(nm)	(nm)

TABLE 6-4 Trap Mask Behavior *(2 of 3)*

TT #	Hardware Trap Name	Type	From privilege level:		
			Nonprivileged	Privileged	Hyperprivileged
17_{16}	*DAE_nfo_page*	Precise	(nm)	(nm)	(nm)
18_{16}–$1F_{16}$	*Reserved*	—	—	—	—
20_{16}	*fp_disabled*	Precise	(nm)	(nm)	(nm)
21_{16}	*fp_exception_ieee_754*	Precise	(nm)	(nm)	(nm)
22_{16}	*fp_exception_other*	Precise	(nm)	(nm)	(nm)
23_{16}	*tag_overflow*	Precise	(nm)	(nm)	(nm)
24_{16}–27_{16}	*clean_window*	Precise	(nm)	(nm)	(nm)
28_{16}	*division_by_zero*	Precise	(nm)	(nm)	(nm)
29_{16}	*internal_processor_error*	Precise	(nm)	(nm)	(nm)
$2A_{16}$	*instruction_invalid_TSB_entry*	Precise	(nm)	(nm)	—
$2B_{16}$	*data_invalid_TSB_entry*	Precise	(nm)	(nm)	(nm)
$2C_{16}$	*Reserved*	—	—	—	—
$2D_{16}$	*mem_real_range*	Precise	(nm)	(nm)	(nm)
$2E_{16}$	*mem_address_range*	Precise	(nm)	(nm)	(nm)
$2F_{16}$	*Reserved*	—	—	—	—
30_{16}	*DAE_so_page*	Precise	(nm)	(nm)	(nm)
31_{16}	*data_access_MMU_miss*	Precise	(nm)	(nm)	(nm)
32_{16}	*data_access_error*	Precise	(nm)	(nm)	(nm)
33_{16}	*data_access_protection*	—	—	—	—
34_{16}	*mem_address_not_aligned*	Precise	(nm)	(nm)	(nm)
35_{16}	*LDDF_mem_address_not_aligned*	Precise	(nm)	(nm)	(nm)
36_{16}	*STDF_mem_address_not_aligned*	Precise	(nm)	(nm)	(nm)
37_{16}	*privileged_action*	Precise	(nm)	—	—
38_{16}	*LDQF_mem_address_not_aligned*	—	—	—	—
39_{16}	*STQF_mem_addess_not_aligned*	—	—	—	—
$3A_{16}$	*Reserved*	—	—	—	—
$3B_{16}$	*unsupported_page_size*	Precise	(nm)	(nm)	(nm)
$3C_{16}$	*control_word_queue_interrupt*	Disrupting	(nm)	(nm)	(ie)
$3D_{16}$	*modular_arithmetic_interrupt*	Disrupting	(nm)	(nm)	(ie)
$3E_{16}$	*inst_real_translation_miss*	Precise	(nm)	(nm)	—
$3F_{16}$	*data_real_translation_miss*	Precise	(nm)	(nm)	(nm)
40	*sw_recoverable_error*	Disrupting	(nm)	(nm)	(ie)
41_{16}–$4F_{16}$	*interrupt_level_n*	Disrupting	PIL	PIL	M
$4F_{16}$	*pic_overflow* (*interrupt_level_15*)	Disrupting	PIL	PIL	M
50_{16}–$5D_{16}$	*Reserved*	—	—	—	—
$5E_{16}$	*hstick_match*	Disrupting	(nm)	(nm)	(ie)

TABLE 6-4 Trap Mask Behavior *(3 of 3)*

TT #	Hardware Trap Name	Type	From privilege level: Nonprivileged	Privileged	Hyperprivileged
$5F_{16}$	*trap_level_zero*	Disrupting	tlz	tlz	—
60_{16}	*interrupt_vector_trap*	Disrupting	(nm)	(nm)	(ie)
61_{16}	*PA_watchpoint*	Precise	(nm)	(nm)	(nm)
62_{16}	*VA_watchpoint*	Precise	(nm)	(nm)	—
63_{16}	*hw_corrected_error*	Disrupting	(nm)	(nm)	(ie)
64_{16}[1]	*fast_instruction_access_MMU_miss*	Precise	(nm)	(nm)	—
68_{16}[1]	*fast_data_access_MMU_miss*	Precise	(nm)	(nm)	(nm)
$6C_{16}$[1]	*fast_data_access_protection*	Precise	(nm)	(nm)	(nm)
70_{16}	*Reserved*	—	—	—	—
71_{16}	*instruction_access_MMU_error*	Precise	(nm)	(nm)	—
72_{16}	*data_access_MMU_error*	Precise	(nm)	(nm)	(nm)
73_{16}	*Reserved*	—	—	—	—
74_{16}	*control_transfer_instruction*	Precise	tct	tct	tct
75_{16}	*instruction_VA_watchpoint*	Precise	(nm)	(nm)	—
76_{16}	*instruction_breakpoint*	Precise	ibe	ibe	ibe
77_{16}–$7B_{16}$	*Reserved*	—	—	—	—
$7C_{16}$	*cpu_mondo_trap*	Disrupting	(ie)	(ie)	M
$7D_{16}$	*dev_mondo_trap*	Disrupting	(ie)	(ie)	M
$7E_{16}$	*resumable_error*	Disrupting	(ie)	(ie)	M
$7F_{16}$	*nonresumable_error* (generated by software only)	—	—	—	—
80_{16}–$9C_{16}$[1]	*spill_n_normal* (n = 0–7)	Precise	(nm)	(nm)	(nm)
A0–BC_{16}[1]	*spill_n_other* (n = 0–7)	Precise	(nm)	(nm)	(nm)
$C0_{16}$–DC_{16}[1]	*fill_n_normal* (n =0–7)	Precise	(nm)	(nm)	(nm)
$E0_{16}$–FC_{16}[1]	*fill_n_other* (n = 0–7)	Precise	(nm)	(nm)	(nm)
100_{16}–$17F_{16}$	*trap_instruction*	Precise	(nm)	(nm)	(nm)
180_{16}–$1FF_{16}$	*htrap_instruction*	Precise	—	(nm)	(nm)

1. Trap extends across four TT #s to allow trap handler to contain 32 inline instructions instead of the standard 8 inline instructions.

The UltraSPARC T2 implmentation deviates from UltraSPARC Architecture 2007 in following way: when an SIR or XIR reset occurs, the virtual processor enters `RED_state` **regardless** of the value of TL (that is, when TL = *MAXTL*, as well as when TL < *MAXTL*.). In particular, when TL = *MAXTL* and an XIR occurs, UltraSPARC

Architecture 2007 specifies that the virtual processor enters `error_state` to generate a WDR reset, but an UltraSPARC T2 virtual processor instead directly takes an XIR reset.

CHAPTER 7

Interrupt Handling

The chapter describes the hardware interrupt delivery mechanism for the UltraSPARC T2 chip. I/O and CPU cross-call interrupts are delivered to hyperprivileged code on each virtual processor using an interrupt vector trap as described in *Interrupt Flow* on page 50. Error interrupts are delivered to hyperprivileged code on each virtual processor using the *sw_recoverable_error* and *hw_corrected_error* traps. These error interrupts are described in Chapter 16, *Error Handling*.

Hyperprivileged code notifies privileged code about *interrupt_vector* traps, *sw_recoverable_error* traps *hw_corrected_error* traps (and precise error traps) through the *cpu_mondo*, *dev_mondo*, and *resumable_error* traps as described in *Interrupt Queue Registers* on page 57. Software interrupts are delivered to each virtual processor using the *interrupt_level_n* traps. Software interrupts are described in the UltraSPARC Architecture 2006 Specification. The *pic_overflow* trap, generated by the performance counters, is described in Chapter 10, *Performance Instrumentation*. The *hstick_match* and *trap_level_zero* interrupts are described in the UltraSPARC Architecture 2006 specification.

Interrupt vector traps have a corresponding 64-bit `ASI_INTR_RECEIVE` register. I/O devices and inter-CPU cross-call interrupts contain a 6-bit identifier, which determines which interrupt vector (level) in the `ASI_INTR_RECEIVE` register the interrupt will target. Each virtual processor's `ASI_INTR_RECEIVE` register can queue up to 64 outstanding interrupts, one for each interrupt vector. Interrupt vectors are implicitly prioritized, with vector 63 being the highest priority and vector 0 being the lowest priority.

Two types of I/O interrupts are supported. "Internal" I/O interrupts, such as those generated by the Network Interface Unit, are generated by I/O devices on the UltraSPARC T2 processor. Unlike "mondo" interrupts, these interrupts do not contain any additional data payload. Each internal I/O interrupt source has a hardwired interrupt number, which is used to index a table of interrupt vector information (INT_MAN) in the NCU. Generally, each I/O interrupt source will be assigned a unique virtual processor target and vector level. This association is defined by software programming of the interrupt vector and strand fields in the INT_MAN table in the NCU. Software must maintain the association between interrupt vector and hardware interrupt number to index the appropriate entry in the INT_MAN table.

The second type of interrupts are external “mondo” interrupts, such as those generated by PCI-Express. These interrupts follow the standard mondo interrupt ACK/NACK flow control. Only the first two 64-bit words of mondo data are supported by UltraSPARC T2.

7.1 Interrupt Flow

7.1.1 Sources

CPU cross-call interrupts can be generated by writing the Interrupt Vector Dispatch register described in *Interrupt Vector Dispatch Register* on page 60. Dispatching inter-CPU interrupts is described in *Dispatching* on page 50.

TAP interrupts can be generated by writing the NCU Interrupt Vector/Trap Dispatch Register described in *Mondo Data Tables* on page 55.

SSI error interrupts (device ID = 1) are caused by SSI detected errors, as described in *Boot ROM Interface (SSI)* on page 315.

SSI interrupts (device ID = 2) are caused by an assertion (edge trigger) on the EXT_INT_L pin.

The network interface unit generates interrupts with device IDs 64–127.

PCI-Express interrupts can either come from INTx emulation or message signaled interrupts (MSI) as described in *Interrupt Model* on page 512 and errors described in *Error Register Overview* on page 524.

7.1.2 Dispatching

CPU cross-call interrupts can be generated by writing the Interrupt Vector Dispatch register described in *Interrupt Vector Dispatch Register* on page 60. Unlike mondo interrupts, interrupts are always received by the destination, and stores to this register will follow the TSO memory model (no MEMBAR #Sync is required). The store data supplies the destination virtual processor and vector. The bit corresponding to the specified vector is set in the Interrupt Receive register of the destination virtual processor.

Note An interrupt that is sent to a virtual processor that is not enabled (has its bit clear in ASI_CORE_ENABLE_STATUS described in *ASI_CORE_ENABLE_STATUS* on page 187) will be lost. An interrupt that is sent to a virtual processor that is parked (has its bit clear in ASI_CORE_RUNNING_STATUS described in *ASI_CORE_RUNNING_STATUS* on page 191) will result in the bit in the interrupt receive register being set, and the interrupt will be taken once the virtual processor is unparked and the interrupt is enabled.

7.1.3 States

Each bit in the Interrupt Receive register can be in one of two states: set or cleared. If an incoming interrupt attempts to set an already set bit, the additional incoming interrupt will be lost (there is no overflow indication on the interrupt bits). Writes to the Interrupt Receive register will clear any bit in which the corresponding write data is 0, and reads to the Incoming Vector register will clear the bit of the highest-priority pending interrupt as a side effect. If another interrupt attempts to set a bit on the same cycle as the bit is being cleared by an Interrupt Receive register write or Incoming Vector register read, the additional interrupt will take precedence over the clear and the bit will remain set.

Mondo interrupts have two states in the MONDO_INT_BUSY table, namely, BUSY and IDLE (not BUSY). When a mondo interrupt transaction is received, if the current state is IDLE, the transaction is accepted (and ACKed to the requestor), state is changed to BUSY, the mondo data is stored in the MONDO_INT_DATA0/1 table, and the bit specified by MONDO_INT_VEC is set in the Interrupt Receive register of the virtual processor specified in the INT transaction. While in the BUSY state, any INT transactions to the same virtual processor will be rejected and NACKed back to the requestor, and the requestor is responsible for preserving that interrupt in a pending state. When software has adequately serviced the interrupt, it explicitly clears the busy bit in the MONDO_INT_BUSY register to return to the IDLE state.

For "internal" I/O interrupts, such as an SSI, network interface unit, or MCU error interrupt, the bit specified by INT_MAN is set in the Interrupt Receive register of the virtual processor specified in INT_MAN. Software can use the Incoming Vector register described in *Incoming Vector Register* on page 60 to atomically clear and return the vector of the highest status bit. Since there is no ACK/NACK flow control on the internal interrupts and no indication of interrupt bit overflow, for sources capable of generating multiple interrupts between interrupt servicing, software will need to properly handle the multiple interrupt case.

7.1.4 Prioritizing

Interrupt vector traps are implicitly prioritized by the Incoming Vector register described in Section 7.3.4 from bit 63 (highest) to bit 0 (lowest).

The priority of I/O interrupts is done by specifying the vector value in the INT_MAN table and in MONDO_INT_VEC (see Section 7.2.1).

7.1.5 Initialization

The interrupt vector traps are initialized by writing 0_{16} to the Interrupt Receive register described in *Interrupt Receive Register* on page 59.

I/O interrupt handling is initialized by

a. Initializing the PIU event queues as described in *Interrupt Model* on page 512 and setting up the NIU logical device groups as described in *System Interrupts* on page 667.

b. Specifying the strand/vector pair to receive "internal" I/O interrupts, such as those generated by the network interface unit, by programming the INT_MAN table, described in *SSI/NIU Interrupt Management Registers* on page 54.

c. Specifying the interrupt vector to receive external "mondo" interrupts, such as those generated by PCI-Express, by programming the MONDO_INT_VEC register, described in *SSI/NIU Interrupt Management Registers* on page 54.

d. Clearing the busy bits in the MONDO_INT_BUSY table, described in *Mondo Interrupt Busy Table* on page 56.

7.1.6 Servicing

Interrupt vector traps are typically serviced by reading the Incoming Vector register described in *Incoming Vector Register* on page 60. When this register is read by software, the 6-bit vector corresponding to the highest priority pending interrupt in the Interrupt Receive register is returned. The pending interrupt bit for that vector is cleared.

If the incoming interrupt matches the virtual processor and vector for SSI error interrupts (device ID = 1), the handler should read the SSI error logs, described in *SSI Error Registers* on page 316, to determine the cause of the interrupt. It should service the interrupt appropriately, checking for multiple interrupts that could have occurred on the same bit in the Interrupt Receive register.

If the incoming interrupt matches the virtual processor and vector for SSI interrupts (device ID = 2), the handler should service the interrupt appropriately, checking for multiple interrupts that could have occurred on the same bit in the Interrupt Receive register.

If the incoming interrupt matches the virtual processor and vector for NIU interrupts (device ID = 64-127), the handler should service the interrupt appropriately, checking for multiple interrupts that could have occurred on the same bit in the Interrupt Receive register.

If the incoming interrupt matches the vector for "mondo" interrupts, the handler should then read the mondo source and data, from the MONDO_INT_ADATA0/1 registers (described in *Mondo Data Tables*) and MONDO_INT_ABUSY registers (described in *Mondo Interrupt Busy Table*). After reading these registers, it should enable receiving the next mondo interrupt to this virtual processor by clearing the busy bit in the MONDO_INT_ABUSY register.

7.2 NCU Interrupt Registers

The following registers are defined for interrupt and reset management. The base address is defined below.

The NCU handles two types of interrupts: those generated on chip by the NIU and those generated externally through the SSI EXT_INT_L pin.

TABLE 7-1 lists the device ID assignment for interrupts.

TABLE 7-1 Device ID Assignments

Device	ID Range	Comment
Reserved	0	
SSI Errors	1	SSI parity or timeout error.
SSI Interrupt	2	SSI interrupt from EXT_INT_L pin.
Reserved	3–63	
NIU	64–127	NIU interrupts

On-chip interrupt hardware contains an SSI/NIU Interrupt Management table. Each internal "Device ID" in the I/O subsystem has an entry in the SSI/NIU Interrupt Management Table described in *SSI/NIU Interrupt Management Registers*.

Device ID 0 is used internally by hardware but is architecturally reserved.

Device ID 1 is used to report SSI Errors. Software will have to poll the SSI error registers to determine the error type.

Device ID 2 is the interrupt from EXT_INT_L pin, of the SSI interface, which is intended for use as a console interrupt.

Device IDs 3-63 are reserved.

Device IDs 64-127 are used for NIU interrupts

Note | The 64 NIU device IDs (64–127) correspond to the Logical Device Group Numbers, described in *System Interrupts* on page 667, where 64 has been added to the Logical Device Group Number.

7.2.1 SSI/NIU Interrupt Management Registers

The SSI/NIU Interrupt Management registers specify the CPU ID to send the interrupt and the interrupt vector associated with the interrupt issued by the NCU on behalf of the device.

Each device will send its device ID to the NCU. The device ID is used to index into the SSI/NIU Interrupt Management table. Before changing the SSI/NIU Interrupt Management register, software must disable all incoming interrupts. Note that the offset address of the corresponding device can be calculated by multiplying the device ID by 8 for INT_MAN.

Software or boot code must program the INT_MAN table before any of the non-mondo type interrupt is generated. Reading of the INT_MAN table without first initializing it by software or boot code will result in false parity errors in NCU.

TABLE 7-2 shows the format of the SSI/NIU Interrupt Management register.

TABLE 7-2 SSI/NIU Interrupt Management – INT_MAN Register ($80\ 0000\ 0000_{16}$)

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	cpu	X	RW	ID of virtual processor to manage the device.
7:6	—	0	RO	*Reserved*
5:0	vector	0	RW	Interrupt vector (encodes bit set in ASI_INTR_RECEIVE).

7.2.2 Mondo Interrupt Vector Register

The Mondo Interrupt Vector register specifies the interrupt vector for PCI-Express mondo interrupts. Since the virtual strand ID is specified in the mondo interrupt, this register is shared among the 64 virtual processors.

TABLE 7-3 shows the format of the Mondo Interrupt Vector register.

TABLE 7-3 Mondo Interrupt Vector Register – MONDO_INT_VEC ($80\ 0000\ 0A00_{16}$)

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	vector	X	RW	Interrupt vector for mondo interrupts (encodes bit set in ASI_INTR_RECEIVE).

MONDO_INT_VEC performs the identical function for PCI-Express Mondo interrupts that INT_MAN performs for the SSI/NIU interrupts, except that the virtual strand ID is specified in the mondo interrupt transaction.

7.2.3 Mondo Data Tables

The following registers manage the receipt from mondo interrupts.

The base address of the mondo interrupt registers is defined below.

There are two Mondo Interrupt Data tables. The tables are read-only by software, and the entries are updated by an incoming mondo interrupts provided that the interrupt is not busy. The NCU will ACK the interrupt if it is not busy; otherwise, the NCU will NACK it.

TABLE 7-4 shows the format of the Mondo Interrupt Data 0 table.

TABLE 7-4 Mondo Interrupt Data 1 – MONDO_INT_DATA0 ($80\ 0004\ 0000_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	data0	X	RO	First 64 bits of mondo interrupt data.

TABLE 7-5 shows the format of the Mondo Interrupt Data 1 table.

TABLE 7-5 Mondo Interrupt Data 1 – MONDO_INT_DATA1 ($80\ 0004\ 0200_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	data1	X	RO	Second 64 bits of mondo interrupt data.

TABLE 7-6 shows the format of the Mondo Interrupt Alias Data 0 table.

This register address is actually an alias for MONDO_INT_DATA0[*My CPUID*], so each virtual processor can read its own interrupt payload without having to do an address calculation based on strand_id. This address should never be accessed by the TAP (since it does not have a strand_id).

TABLE 7-6 Mondo Interrupt Alias Data 0 – MONDO_INT_ADATA0 ($80\ 0004\ 0400_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	data0	X	RO	First 64 bits of mondo interrupt data.

TABLE 7-7 shows the format of the Mondo Interrupt Alias Data 1 Table.

This register address is actually an alias for MONDO_INT_DATA1[*My CPUID*], so each virtual processor can read its own interrupt payload, without having to do an address calculation based on strand_id. This address should never be accessed by the TAP (since it does not have a strand_id).

TABLE 7-7 Mondo Interrupt Alias Data 1 – MONDO_INT_ADATA1 ($80\ 0004\ 0600_{16}$)

Bit	Field	Initial Value	R/W	Initial Value
63:0	data1	X	RO	Second 64 bits of mondo interrupt data.

7.2.4 Mondo Interrupt Busy Table

When the NCU receives a mondo interrupt, it sets the busy bit to 1 and ACKs the interrupt. When the busy bit is set, it implies an interrupt is waiting to be serviced or is being serviced. Software will reset the busy bit when it completes servicing the interrupt. If the busy bit is already set when an interrupt is received, a NACK will be sent to the interrupt source. The busy bit is set after a reset and software has to clear it to begin receiving interrupts.

TABLE 7-8 shows the format of the Mondo Interrupt Busy table.

TABLE 7-8 Mondo Interrupt Busy – MONDO_INT_BUSY $(80\ 0004\ 0800)_{16}$

Bit	Field	Initial Value	R/W	Description
63:7	—	0	RO	*Reserved*
6	busy	1	RW	Hardware sets busy to 1 when an interrupt is received. Hardware NACKs an incoming interrupt if busy is set.
5:0	—	0	RO	*Reserved*

TABLE 7-9 shows the format of the Mondo Interrupt Alias Busy table.

This register address is actually an alias for MONDO_INT_BUSY[*My CPUID*], so each virtual processor can update its own mondo interrupt busy bit without having to do an address calculation based on strand_id. This address should never be accessed by the TAP (since it does not have a strand_id).

TABLE 7-9 Mondo Interrupt Alias Busy – MONDO_INT_ABUSY (80 0004 $0A00_{16}$)

Bit	Field	Initial Value	R/W	Description
63:7	—	0	RO	*Reserved*
6	busy	1	RW	Hardware sets busy to 1 when an interrupt is received. Hardware NACKs an incoming interrupt if busy is set.
5:0	—	0	RO	*Reserved*

7.3 CPU Interrupt Registers

7.3.1 Interrupt Queue Registers

Each virtual processor has eight ASI_QUEUE registers at ASI = 25_{16}, VA{63:0} = $3C0_{16}$-$3F8_{16}$ that are used for communicating interrupts to the operating system. These registers contain the head and tail pointers for four supervisor interrupt queues: *cpu_mondo*, *dev_mondo*, *resumable_error*, *nonresumable_error*. The tail registers are read-only by supervisor, and read/write by hypervisor. Writes to the tail registers by the supervisor generate a *DAE_invalid_ASI* trap. The head registers are read/write by both supervisor and hypervisor.

Whenever the CPU_MONDO_HEAD register does not equal the CPU_MONDO_TAIL register, a *cpu_mondo* trap is generated. Whenever the DEV_MONDO_HEAD register does not equal the DEV_MONDO_TAIL register, a *dev_mondo* trap is generated. Whenever the RESUMABLE_ERROR_HEAD register does not equal the RESUMABLE_ERROR_TAIL register, a *resumable_error* trap is generated. Unlike the other queue register pairs, the *nonresumable_error* trap is *not* automatically generated whenever the NONRESUMABLE_ERROR_HEAD register does not equal the NONRESUMABLE_ERROR_TAIL register; instead, the hypervisor will need to generate the *nonresumable_error* trap.

TABLE 7-10 through TABLE 7-17 define the format of the eight ASI_QUEUE registers.

TABLE 7-10 CPU Mondo Head Pointer – ASI_QUEUE_CPU_MONDO_HEAD (ASI 25_{16}, VA $3C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	head	X	RW	Head pointer for CPU mondo interrupt queue.
5:0	—	0	RO	*Reserved*

TABLE 7-11 CPU Mondo Tail Pointer – `ASI_QUEUE_CPU_MONDO_TAIL` (ASI 25_{16}, VA $3C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	tail	X	RW (hyperpriv) RO (priv)	Tail pointer for CPU mondo interrupt queue.
5:0	—	0	RO	*Reserved*

TABLE 7-12 Device Mondo Head Pointer – `ASI_QUEUE_DEV_MONDO_HEAD` (ASI 25_{16}, VA $3D0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	head	X	RW	Head pointer for device mondo interrupt queue.
5:0	—	0	RO	*Reserved*

TABLE 7-13 Device Mondo Tail Pointer – `ASI_QUEUE_DEV_MONDO_TAIL` (ASI 25_{16}, VA $3D8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	tail	X	RW (hyperpriv) RO (priv)	Tail pointer for device mondo interrupt queue.
5:0	—	0	RO	*Reserved*

TABLE 7-14 Resumable Error Head Pointer – `ASI_QUEUE_RESUMABLE_HEAD` (ASI 25_{16}, VA $3E0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	Reserved.
17:6	head	X	RW	Head pointer for resumable error queue.
5:0	—	0	RO	*Reserved*

TABLE 7-15 Resumable Error Tail Pointer – `ASI_QUEUE_RESUMABLE_TAIL` (ASI 25_{16}, VA $3E8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	tail	X	RW (hyperpriv) RO (priv)	Tail pointer for resumable error queue.
5:0	—	0	RO	*Reserved*

TABLE 7-16 Nonresumable Error Head Pointer – `ASI_QUEUE_NONRESUMABLE_HEAD` (ASI 25_{16}, VA $3F0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	head	X	RW	Head pointer for nonresumable error queue.
5:0	—	0	RO	*Reserved*

TABLE 7-17 Nonresumable Error Tail Pointer – `ASI_QUEUE_NONRESUMABLE_TAIL` (ASI 25_{16}, VA $3F8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:6	tail	X	RW (hyperpriv) RO (priv)	Tail pointer for nonresumable error queue.
5:0	—	0	RO	*Reserved*

7.3.2 Interrupt Receive Register

Each virtual processor has a hyperprivileged `ASI_INTR_RECEIVE` register at ASI = 72_{16}, VA{63:0} = 0. Each time an interrupt transaction arrives for that virtual processor, the bit corresponding to the interrupt vector will be set. Bit zero of the register corresponds to interrupt vector number zero and so on. Interrupt vectors are implicitly prioritized with vector number 63 being the highest priority and vector number 0 being the lowest priority. Software writes to this register are **and**ed with the register contents to allow the software to selectively clear register bits, although normally the incoming vector register described in Section 7.3.4 will be used to clear the bit corresponding to the pending interrupt. When an interrupt arrives at the same time as a register write, the interrupt will take precedence over the write and the bit will be set. Software can read this register to determine all pending interrupts, although normally the incoming vector register will be used to get the highest priority pending interrupt. Nonprivileged or supervisor access to this register causes a *privileged_action* trap.

TABLE 7-18 defines the format of the `ASI_INTR_RECEIVE` register.

TABLE 7-18 Interrupt Receive Register – `ASI_INTR_RECEIVE` (ASI 72_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:0	pending	X	RW	Pending interrupts.

7.3.3 Interrupt Vector Dispatch Register

Each virtual processor has a hyperprivileged write-only ASI_INTR_W register at ASI = 73_{16}, VA{63:0} = 0 that is used to send CPU cross-call interrupts to other virtual processors. Unlike mondo interrupts, interrupts cannot be NACKed by the destination, and multiple interrupts that set the same Interrupt Receive register bit before it has been cleared will only generate a single interrupt. Interrupts generated by stores to this register will follow the TSO memory model (no MEMBAR #Sync is required). The store data supplies the destination virtual processor and vector. The bit corresponding to the specified vector is set in the Interrupt Receive register of the destination virtual processor.

Programming Note After an interrupt vector trap is taken by the destination virtual processor, it is the responsibility of the interrupt handler to clear the highest-priority pending bit in the interrupt register, usually by a read to the Incoming Vector register as described in *Incoming Vector Register*.

The format of the register is shown in TABLE 7-19.

TABLE 7-19 Interrupt Vector Dispatch Register – ASI_INTR_W (ASI 73_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	strand	X	W	Destination virtual processor
7:6	—	0	RO	*Reserved*.
5:0	vector	X	W	Interrupt Vector (encodes bit set in ASI_INTR_RECEIVE)

Nonprivileged or supervisor access to this register causes a *privileged_action* trap. A read from this ASI causes a *DAE_invalid_asi* trap.

Implementation Note This register is actually implemented in the NCU and is also available at address 90_{16}-$01CC_{16}$-0000_{16} as described in *Interrupt Vector Dispatch Register (ASI 7316 VA 016)* on page 207.

7.3.4 Incoming Vector Register

Each virtual processor has a hyperprivileged read-only ASI_INTR_R register at ASI = 74_{16}, VA{63:} = 0_{16}. When this register is read by software, the 6-bit vector corresponding to the highest priority pending interrupt in the Interrupt Receive register is returned. The pending interrupt bit for that vector is cleared. If no interrupt bits are set, a read of this register will return all zeros. When an interrupt arrives at the same time as the register is read, the interrupt will take precedence over the clearing and the bit will remain set. Nonprivileged or supervisor access to this register causes a *privileged_action* trap. A store to this register will result in a *DAE_invalid_ASI* trap.

Programming Note The interrupt handler will normally use the Incoming Vector register to determine the highest-priority interrupt that is pending while atomically clearing the bit corresponding to that highest priority interrupt.

TABLE 7-20 defines the format of the ASI_INTR_R register.

TABLE 7-20 Incoming Vector Register – ASI_INTR_R (ASI 74_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	vector	X	RO	Interrupt vector.

CHAPTER 8

Memory Models

SPARC V9 defines the semantics of memory operations for three memory models. From strongest to weakest, they are Total Store Order (TSO), Partial Store Order (PSO), and Relaxed Memory Order (RMO). The differences in these models lie in the freedom an implementation is allowed in order to obtain higher performance during program execution. The purpose of the memory models is to specify any constraints placed on the ordering of memory operations in uniprocessor and shared-memory multiprocessor environments. UltraSPARC T2 supports only TSO, with the exception that certain ASI accesses (such as block loads and stores) may operate under RMO.

Although a program written for a weaker memory model potentially benefits from higher execution rates, it may require explicit memory synchronization instructions to function correctly if data is shared. MEMBAR is a SPARC V9 memory synchronization primitive that enables a programmer to control explicitly the ordering in a sequence of memory operations. Processor consistency is guaranteed in all memory models.

The current memory model is indicated in the PSTATE.mm field. It is unaffected by normal traps, but is set to TSO (PSTATE.mm = 0) when the virtual processor enters `RED_state`. UltraSPARC T2 ignores the value set in this field and always operates under TSO.

A memory location is identified by an 8-bit address space identifier (ASI) and a 64-bit virtual address. The 8-bit ASI may be obtained from a ASI register or included in a memory access instruction. The ASI is used to distinguish between and provide an attribute for different 64-bit address spaces. For example, the ASI is used by the UltraSPARC T2 MMU and memory access hardware to control virtual-to-physical address translations, access to implementation-dependent control and data registers, and for access protection. Attempts by nonprivileged software (PSTATE.priv = 0) to access restricted ASIs (ASI{7} = 0) cause a *privileged_action* trap.

Memory is logically divided into real memory (cached) and I/O memory (noncached with and without side effects) spaces, based on bit 39 of the physical address (0 for real memory, 1 for I/O memory). Real memory spaces can be accessed without side effects. For example, a read from real memory space returns the information most recently written. In addition, an access to real memory space does

not result in program-visible side effects. In contrast, a read from I/O space may not return the most recently written information and may result in program-visible side effects.

8.1 Supported Memory Models

The following sections contain brief descriptions of the two memory models supported by UltraSPARC T2. These definitions are for general illustration. Detailed definitions of these models can be found in *The SPARC Architecture Manual-Version 9*. The definitions in the following sections apply to system behavior as seen by the programmer.

Notes Stores to UltraSPARC T2 internal ASIs, block loads, and block stores and block initializing stores are outside the memory model; that is, they need MEMBARs to control ordering.

Atomic load-stores are treated as both a load and a store and can only be applied to cacheable address spaces.

8.1.1 TSO

UltraSPARC T2 implements the following programmer-visible properties in Total Store Order (TSO) mode:

- Loads are processed in program order; that is, there is an implicit MEMBAR `#LoadLoad` between them.
- Loads may bypass earlier stores. Any such load that bypasses such earlier stores must check (snoop) the store buffer for the most recent store to that address. A MEMBAR `#Lookaside` is not needed between a store and a subsequent load at the same noncacheable address.
- A MEMBAR `#StoreLoad` must be used to prevent a load from bypassing a prior store if Strong Sequential Order is desired.
- Stores are processed in program order.
- Stores cannot bypass earlier loads.
- Accesses with PA{39} set (that is, to I/O space) are all strongly ordered with respect to each other.

Compatibility Note | Prior UltraSPARC implementations strongly order accesses based on the e bit being set. The e bit is ignored by UltraSPARC T2 for the purposes of strong ordering; only PA{39} is used for determining strong ordering.

- An L2 cache update is delayed on a store hit until all outstanding stores reach global visibility. For example, a cacheable store following a noncacheable store is not globally visible until the noncacheable store has reached global visibility; there is an implicit MEMBAR #MemIssue between them.

8.1.2 RMO

UltraSPARC T2 implements the following programmer-visible properties for special ASI accesses that operate under Relaxed Memory Order (RMO) mode:

- There is no implicit order between any two memory references, either cacheable or noncacheable, except that noncacheable accesses with PA{39} set (that is, to I/O space) are all strongly ordered with respect to each other.

Compatibility Note | Prior UltraSPARC implementations strongly order accesses based on the e bit being set. The e bit is ignored by UltraSPARC T2 for the purposes of strong ordering, only PA{39} is used for determining strong ordering.

- A MEMBAR must be used between cacheable memory references if stronger order is desired. A MEMBAR #MemIssue is needed for ordering of cacheable after noncacheable accesses. A MEMBAR #Lookaside should be used between a store and a subsequent load at the same noncacheable address.

CHAPTER 9

Address Spaces and ASIs

9.1 Physical Address Spaces

UltraSPARC T2 supports a 48-bit virtual address space and a 40-bit physical address space. The 40-bit physical address space is further broken into two sections, based on bit 39. If bit 39 is a 0, the address maps to a memory location. If bit 39 is a 1, the address maps to an I/O location.

9.1.1 Access to Nonexistent Physical Memory Addresses

Access to nonexistent physical memory addresses is described in *Access to Nonexistent Memory* on page 369.

9.1.2 Access to Nonexistent I/O Addresses

A load access from a nonexistent memory or I/O location will cause a *data_access_error* exception. An instruction fetch from a nonexistent memory or I/O location will cause an *instruction_access_error* exception. A store access to a nonexistent memory or I/O location will be silently discarded by the system.

9.1.3 Instruction Fetching from I/O

Instruction fetching from I/O addresses is only permitted from the SSI (FF 0000 0000_{16}–FF FFFF $FFFC_{16}$) and L2CSR spaces (A0 0000 0000_{16}–BF FFFF $FFFC_{16}$). Instruction fetches from I/O addresses outside the SSI and L2CSR spaces will take an *instruction_access_error* trap.

Warning | Instruction fetching from the L2CSR space can cause undefined behavior. Software needs to prevent instruction fetches from accessing the L2CSR space.

9.1.4 Supported vs. Unsupported Access Sizes to I/O

All I/O locations internal to UltraSPARC T2 are 64-bit locations, and only support 8-byte (64-bit) loads and stores. Accesses in other sizes may cause traps or have other unexpected results. In particular, non-8-byte aligned load accesses to internal UltraSPARC T2 I/O addresses (except internal PCI-Express or SSI locations) will result in *data_access_error* trap. Non-8-byte-aligned store accesses to internal UltraSPARC T2 I/O addresses (except internal PCI-Express or SSI locations) will be silently discarded by the system. Non-8-byte-aligned load accesses from internal PCI-Express or SSI locations are treated internally as 8-byte loads, with potentially undefined results. Non-8-byte-aligned store accesses to internal PCI-Express or SSI locations are treated internally as 8-byte stores, also with potentially undefined results.

UltraSPARC T2 supports 1-byte, 2-byte, 4-byte, and 8-byte loads and stores via the SSI bus (Boot ROM port). 8-byte stores under mask (generated by STDFA to `ASI_PST*`) are undefined. 16-byte loads (generated by an LDDA to `ASI_TWINX*`), block loads, and block stores generate a *DAE_nc_page* exception. (UltraSPARC T2 cannot generate a 16-byte store.)

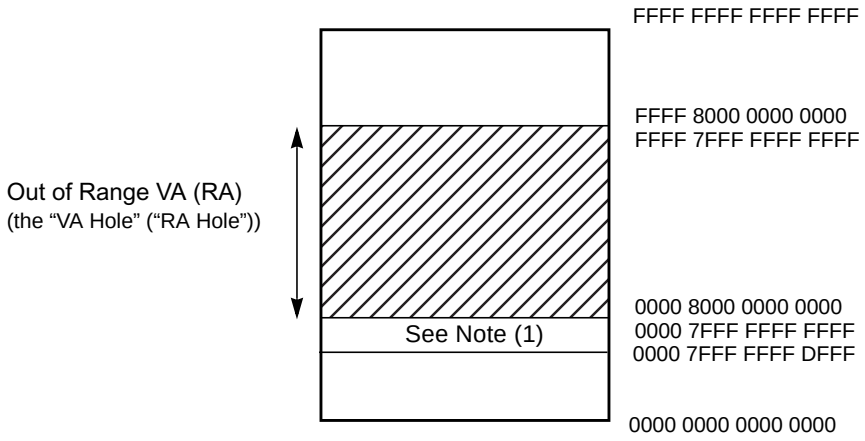
UltraSPARC T2 supports 1-byte, 2-byte, 4-byte, and 8-byte loads and stores, plus 8-byte stores under mask (generated by STDFA to `ASI_PST*`) and block stores, to PCI-Express (for external PCI-Express locations). 16-byte loads (generated by an LDDA to `ASI_TWINX*`) and block loads are not supported to PCI-Express and generate a *DAE_nc_page* exception.

9.1.5 48-bit Virtual and Real Address Spaces

UltraSPARC T2 supports a 48-bit subset of the full 64-bit virtual and real address spaces. Although the full 64 bits are generated and stored in integer registers, legal addresses are restricted to two equal halves at the extreme lower and upper portions of the full virtual (real) address space. Virtual (real) addresses between $0000\ 8000\ 0000\ 0000_{16}$ and $FFFF\ 7FFF\ FFFF\ FFFF_{16}$ inclusive lie within a "VA hole" ("RA hole"), are termed "out-of-range"[1], and are illegal. Prior UltraSPARC implementations introduced the additional restriction on software to not use pages within 4 Gbytes of the VA (RA) hole as instruction pages to avoid problems with prefetching into the VA (RA) hole. UltraSPARC T2 implements a hardware check for instruction fetching near the VA (RA) hole and generates an *instruction_address_range* or *instruction_real_range* trap when instructions are

[1] Another way to view an out-of-range address is as any address where bits {63:48} are not all equal to bit {47}.

executed from a location in the address range 0000 7FFF FFFF $FFE0_{16}$ to 0000 7FFF FFFF $FFFF_{16}$, inclusive. However, even though UltraSPARC T2 provides this hardware checking, it is still recommended that software should *not* use the 8-Kbyte page before the VA (RA) hole for instructions. Address translation and MMU related descriptions can be found in *Translation* on page 128.



Note (1): Use of this region restricted to data only.

FIGURE 9-1 UltraSPARC T2's 48-bit Virtual and Real Address Spaces, With Hole

Throughout this document, when virtual (real) address fields are specified as 64-bit quantities, they are assumed to be sign-extended based on VA{47} (RA{47}).

A number of state registers are affected by the reduced virtual and real address spaces. The PC, I/D-TLB Tag Access, instruction and data watchpoint registers are 48 bits, sign-extended to 64-bits on read accesses. DMMU SFAR, TBA, TPC, and TNPC, registers are 48-bits and their values are *not* sign-extended when read. No checks are done when these registers are written by software. It is the responsibility of privileged (or hyperprivileged) software to properly update these registers.

An out-of-range virtual (real) address during an instruction access, caused by execution into the VA (RA) hole or into 0000 7FFF FFFF $FFE0_{16}$ to 0000 7FFF FFFF $FFFF_{16}$ inclusive, results in an *instruction_address_range* (*instruction_real_range*) trap if PSTATE.am = 0. In addition, UltraSPARC T2 hardware detects when a branch target or the target of a DONE or RETRY instruction is in the VA (RA) hole, and PSTATE.am changes from being set ('1') for the branch, DONE, or RETRY to being cleared ('0') for the target instruction (via the branch delay slot instruction or TSTATE) and generates an *instruction_address_range* (*instruction_real_range*) exception.

Note The *instruction_address_rangne* and *instruction_real_range* exceptions occur on a fetch that enters the VA hole or that is in the cache line immediately before the VA hole (or when the target of a branch or DONE/RETRY instruction is in the VA hole and VA hole detection is enabled only for the target). Hardware does not store state to create an *instruction_address_range* or *instruction_real_range* exception when the strand is executing from the VA hole in a state that cannot detect VA hole exceptions (for example, when PSTATE.am = 1), and then software transitions the strand to be able to detect the VA hole (for example, by setting PSTATE.am = 0) from within the VA hole itself.

If the target virtual (real) address of a JMPL, RETURN, branch, or CALL instruction is an out-of-range address and PSTATE.am = 0, a *mem_address_range* (*mem_real_range*) trap is generated with TPC equal to the address of the JMPL, RETURN, branch, or CALL instruction. The target address is loaded into the D-MMU SFAR. Because the D-MMU SFAR contains only 48 bits, the trap handler must decode the load or store instruction if the full 64-bit virtual address is needed. See also *I-/D-MMU Synchronous Fault Address Registers (SFAR)* on page 139.

An out-of-range virtual (real) address during a data access results in a *mem_address_range* (*mem_real_range*) trap if PSTATE.am = 0. Because the D-MMU SFAR contains only 48 bits, the trap handler must decode the load or store instruction if the full 64-bit virtual address is needed. See also *I-/D-MMU Synchronous Fault Address Registers (SFAR)* on page 139.

9.1.6 I/O Address Spaces

I/O addresses are distinguished from memory addresses via their high-order physical address bit (bit 39). If bit 39 is 0, the address is a memory address. If bit 39 is 1, the address is an I/O address.

TABLE 9-1 summarizes the UltraSPARC T2 address space, which is broken down into sections based on the eight most significant bits of the physical address.

TABLE 9-1 UltraSPARC T2 Address Space

Address Range (PA{39:32})	Block Name	Comment
00_{16} – $7F_{16}$	DRAM	Main memory
80_{16}	NCU	Noncacheable Unit
81_{16}	NIU	Network Interface Unit
82_{16}	—	*Reserved*
83_{16}	CCU	Clock Unit.

TABLE 9-1 UltraSPARC T2 Address Space *(Continued)*

Address Range (PA{39:32})	Block Name	Comment
84_{16}	MCU	Memory Control Unit, address bits {13:12} select one of the four MCUs.
85_{16}	TCU	JTAG/TAP unit.
86_{16}	DBG	Debug.
87_{16}	—1	*Reserved*
88_{16}	DMU	DMU CSRs.
89_{16}	RST	Reset.
$8A_{16}$–$8F_{16}$	—2	*Reserved*
90_{16}	ASI	CPU shared registers (directly accessible only by JTAG/TAP unit).
91_{16}–$9F_{16}$	—2	*Reserved*
$A0_{16}$–BF_{16}	L2CSR	L2 control and status registers.
$C0_{16}$–CF_{16}	PCIE	PCI-Express (64 GB) and DMU PIO.
$D0_{16}$–FE_{16}	—3	*Reserved*
FF_{16}	SSI	Boot ROM.

Programming Note Address space 90_{16} provides an alias for the shared CPU registers. This alias is directly accessible only by the JTAG/TAP unit. Access to these shared CPU registers by the strands should be done directly through the ASIs listed in *Alternate Address Spaces* on page 71.

9.2 Alternate Address Spaces

TABLE 9-2 summarizes the ASI usage in UltraSPARC T2. The Section/Page column contains a reference to the detailed explanation of the ASI (the page number refers to this chapter). For internal ASIs, the legal VAs are listed (or the field contains “Any” if all VAs are legal). Only bits 47:0 are checked when determining the legal VA range. An access outside the legal VA range will generate a *DAE_invalid_asi* trap.

Notes | All internal, nontranslating ASIs in UltraSPARC T2 can only be accessed using LDXA and STXA.

ASIs 80_{16}–FF_{16} are unrestricted (access allowed in all modes -- nonprivileged, privileged, and hyperprivileged). ASIs 00_{16}–$2F_{16}$ are restricted to privileged and hyperprivileged modes, while ASIs 30_{16}–$7F_{16}$ are restricted to hyperprivileged mode only. Attempted access by nonprivileged or privileged code to a hyperprivileged ASI will result in a *privileged_action* trap.

TABLE 9-2 UltraSPARC T2 ASI Usage *(1 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
00_{16}–03_{16}			Any	—	*DAE_invalid_asi*	
04_{16}	ASI_NUCLEUS	RW	Any	—	Implicit address space, nucleus context, TL > 0	(See UA-2007)
05_{16}–$0B_{16}$			Any	—	*DAE_invalid_asi*	
$0C_{16}$	ASI NUCLEUS_LITTLE	RW	Any	—	Implicit address space, nucleus context, TL > 0 (LE)	(See UA-2007)
$0D_{16}$–$0F_{16}$			Any	—	*DAE_invalid_asi*	
10_{16}	ASI_AS_IF_USER_PRIMARY	RW	Any	—	Primary address space, user privilege	(See UA-2007)
11_{16}	ASI_AS_IF_USER_SECONDARY	RW	Any	—	Secondary address space, user privilege	(See UA-2007)
12_{16}–13_{16}			Any	—		
14_{16}	ASI_REAL	RW	Any	—	Real address (normally used as cacheable)	page 82
15_{16}	ASI_REAL_IO	RW	Any	—	Real address (normally used as noncacheable, with side effect)	page 82
16_{16}	ASI_BLOCK_AS_IF_USER_PRIMARY	RW	Any	—	64-byte block load/store, primary address space, user privilege	5.3
17_{16}	ASI_BLOCK_AS_IF_USER_SECONDARY	RW	Any	—	64-byte block load/store, secondary address space, user privilege	5.3
18_{16}	ASI_AS_IF_USER_PRIMARY_LITTLE	RW	Any	—	Primary address space, user privilege (LE)	(See UA-2007)
19_{16}	ASI_AS_IF_USER_SECONDARY_LITTLE	RW	Any	—	Secondary address space, user privilege (LE)	(See UA-2007)
$1A_{16}$–$1B_{16}$			Any	—		

TABLE 9-2 UltraSPARC T2 ASI Usage *(2 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
$1C_{16}$	ASI_REAL_LITTLE	RW	Any	—	Real address (normally used as cacheable) (LE)	page 82
$1D_{16}$	ASI_REAL_IO_LITTLE	RW	Any	—	Real address (normally used as noncacheable, with side effect) (LE)	page 82
$1E_{16}$	ASI_BLOCK_AS_IF_USER_PRIMARY_LITTLE	RW	Any	—	64-byte block load/store, primary address space, user privilege (LE)	5.3
$1F_{16}$	ASI_BLOCK_AS_IF_USER_SECONDARY_LITTLE	RW	Any	—	64-byte block load/store, secondary address space, user privilege (LE)	5.3
20_{16}	ASI_SCRATCHPAD	RW	0_{16}–18_{16}	Y	Scratchpad registers	page 82
20_{16}	ASI_SCRATCHPAD	RW	20_{16}–28_{16}	—		page 82
20_{16}	ASI_SCRATCHPAD	RW	30_{16}–38_{16}	Y	Scratchpad registers	page 82
21_{16}	ASI_MMU	RW	8_{16}	Y	I/DMMU Primary Context register 0	12.10.2
21_{16}	ASI_MMU	RW	10_{16}	Y	DMMU Secondary Context register 0	12.10.2
21_{16}	ASI_MMU	RW	108_{16}	Y	I/DMMU Primary Context register 1	12.10.2
21_{16}	ASI_MMU	RW	110_{16}	Y	DMMU Secondary Context register 1	12.10.2
22_{16}	ASI_TWINX_AIUP, ASI_STBI_AIUP	RW	Any	—	Load: 128-bit atomic load twin extended word, primary address space, user privilege Store: Block initializing store, primary address space, user privilege	5.7.4
23_{16}	ASI_TWINX_AIUS, ASI_STBI_AIUS	RW	Any	—	Load: 128-bit atomic load twin extended word, secondary address space, user privilege Store: Block initializing store	(See UA-2007)
24_{16}	ASI_TWINX	RO	Any	—	128-bit atomic load twin extended word	(See UA-2007)
25_{16}	ASI_QUEUE	RW	$3C0_{16}$	Y	CPU Mondo Queue head pointer	7.3.1

TABLE 9-2 UltraSPARC T2 ASI Usage *(3 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
25_{16}	ASI_QUEUE	RW (hyperpriv) RO (priv)	3C8	Y	CPU Mondo Queue tail pointer	7.3.1
25_{16}	ASI_QUEUE	RW	$3D0_{16}$	Y	Device Mondo Queue head pointer	7.3.1
25_{16}	ASI_QUEUE	RW (hyperpriv) RO (priv)	$3D8_{16}$	Y	Device Mondo Queue tail pointer	7.3.1
25_{16}	ASI_QUEUE	RW	$3E0_{16}$	Y	Resumable Error Queue head pointer	7.3.1
25_{16}	ASI_QUEUE	RW (hyperpriv) RO (priv)	$3E8_{16}$	Y	Resumable Error Queue tail pointer	7.3.1
25_{16}	ASI_QUEUE	RW	$3F0_{16}$	Y	Nonresumable Error Queue head pointer	7.3.1
25_{16}	ASI_QUEUE	RW (hyper-priv) RO (priv)	$3F8_{16}$	Y	Nonresumable Error Queue tail pointer	7.3.1
26_{16}	ASI_TWINX_REAL	R	Any	—	128-bit atomic LDDA, real address	(See UA-2007)
27_{16}	ASI_TWINX_NUCLEUS, ASI_STBI_N	RW	Any	—	Load: 128-bit atomic load twin extended word from nucleus context Store: Block initializing store from nucleus context	(See UA-2007)
28_{16}–29_{16}			Any	—	*DAE_invalid_asi*	
$2A_{16}$	ASI_TWINX_AIUPL, ASI_STBI_AIUPL	RW	Any	—	Load: 128-bit atomic load twin extended word, primary address space, user privilege, little endian Store: Block initializing store, primary address space, user privilege, little endian	(See UA-2007)

TABLE 9-2 UltraSPARC T2 ASI Usage *(4 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
$2B_{16}$	ASI_TWINX_AIUSL, ASI_STBI_AIUSL	RW	Any	—	Load: 128-bit atomic load twin extended word, secondary address space, user privilege, little endian Store: Block initializing store, secondary address space, user privilege, little endian	(See UA-2007)
$2C_{16}$	ASI_TWINX_LITTLE	RO	Any	—	128-bit atomic load twin extended word, little endian	(See UA-2007)
$2D_{16}$			Any	—	*DAE_invalid_asi*	
$2E_{16}$	ASI_TWINX_REAL_LITTLE	RO	Any	—	128-bit atomic LDDA, real address (LE)	(See UA-2007)
$2F_{16}$	ASI_TWINX_NL, ASI_STBI_NL	RW	Any	—	Load: 128-bit atomic load twin extended word from nucleus context, little endian Store: Block initializing store from nucleus context, little endian	(See UA-2007)
30_{16}	ASI_AS_IF_PRIV_PRIMARY	RW	Any	—	Primary address space, privilege access	(See UA-2007)
31_{16}	ASI_AS_IF_PRIV_SECONDARY	RW	Any	—	Secondary address space, privilege access	(See UA-2007)
32_{16}–35_{16}			Any	—	*DAE_invalid_asi*	
36_{16}	ASI_AS_IF_PRIV_NUCLEUS	RW	Any	—	Nucleus address space, privilege access	(See UA-2007)
37_{16}			Any	—	*DAE_invalid_asi*	
38_{16}	ASI_AS_IF_PRIV_PRIMARY_LITTLE	RW	Any	—	Primary address space, privileged access, little endian	(See UA-2007)
39_{16}	ASI_AS_IF_PRIV_SECONDARY_LITTLE	RW	Any	—	Secondary address space, privilege access, little endian	(See UA-2007)
$3A_{16}$–$3D_{16}$			Any	—	*DAE_invalid_asi*	
$3E_{16}$	ASI_AS_IF_PRIV_NUCLEUS_LITTLE	RW	Any	—	Nucleus address space, privilege access, little endian	(See UA-2007)
$3F_{16}$			Any	—	*DAE_invalid_asi*	

TABLE 9-2 UltraSPARC T2 ASI Usage *(5 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
41_{16}	ASI_CMT	RO	0	S	Strand available	14.1.1
41_{16}	ASI_CMT	RO	10_{16}	S	Strand enable status	14.1.2
41_{16}	ASI_CMT	RW	20_{16}	S	Strand enable	14.1.3
41_{16}	ASI_CMT	RW	30_{16}	S	XIR steering	14.1.4
41_{16}	ASI_CMT	RW	38_{16}	S	Tick_Enable	14.1.5
41_{16}	ASI_CMT	RW	50_{16}	S	Running_RW	14.1.7
41_{16}	ASI_CMT	RO	58_{16}	S	Running Status	14.1.8
41_{16}	ASI_CMT	WO	60_{16}	S	Running_W1S	14.1.9
41_{16}	ASI_CMT	WO	68_{16}	S	Running_W1C	14.1.10
42_{16}	ASI_INST_MASK_REG	RW	8_{16}	N	SPARC Instruction Mask register	19.3.1
42_{16}	ASI_LSU_DIAG_REG	RW	10_{16}	N	Load/Store Unit Diagnostic register	19.4.1
43_{16}	ASI_ERROR_INJECT_REG	RW	0	N	ASI_ERROR_INJECT_REG	16.8.9
44_{16}			Any		*DAE_invalid_asi*	
45_{16}	ASI_LSU_CONTROL_REG	RW	0	Y	Load/Store Unit Control register	19.1
45_{16}	ASI_DECR	RW	8_{16}	N	Trap unit	16.7.10
45_{16}	ASI_RST_VEC_MASK	RW	18_{16}	S	SOC	20.1.4
46_{16}	ASI_DCACHE_DATA	RW	Any	Y	Dcache data array diagnostic access	19.6.1
47_{16}	ASI_DCACHE_TAG	RW	Any	Y	Dcache tag and valid bit diagnostic access	19.6.2
48_{16}	ASI_IRF_ECC_REG	RO	Any	Y	IRF ECC diagnostic access	19.7.1
49_{16}	ASI_FRF_ECC_REG	RO	Any	Y	FRF ECC diagnostic access	19.8.1
$4A_{16}$	ASI_STB_ACCESS	RO	Any	Y	Store buffer diagnostic access	19.9
$4B_{16}$			Any	—	*DAE_invalid_asi*	
$4C_{16}$	ASI_DESR	RO	0	Y	Disrupting Error Status register (DESR)	16.8.5
$4C_{16}$	ASI_DFESR	RO	8_{16}	Y	Deferred Error Status register	16.8.6
$4C_{16}$	ASI_CERER	RW	10_{16}	N	ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER	16.8.1

TABLE 9-2 UltraSPARC T2 ASI Usage *(6 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
$4C_{16}$	ASI_SETER	RW	18_{16}	Y	ASI_STRAND_ERROR_TRAP_ENABLE_REGISTER	16.8.2
$4C_{16}$	ASI_CLESR	RO	20_{16}	Y	ASI_CLESR	16.8.7
$4C_{16}$	ASI_CLFESR	RO	28_{16}	Y	ASI_CLFESR	16.8.8
$4D_{16}$					*DAE_invalid_asi*	
$4E_{16}$	ASI_SPARC_PWR_MGMT	RW	0	N	SPARC power management register	18.1
$4F_{16}$	ASI_HYP_SCRATCHPAD	RW	0_{16}–38_{16}	Y	Hypervisor scratchpad	page 83
50_{16}	ASI_ITSB_TAG_TARGET	RO	0	Y	IMMU Tag Target register	12.10.3
50_{16}	ASI_ISFSR	RW	18	Y	IMMU Synchronous Fault Status register	16.8.3
50_{16}	ASI_ITLB_TAG_ACCESS	RW	30	Y	IMMU TLB Tag Access register	12.10.5
50_{16}	ASI_IMMU_VA_WATCHPOINT	RW	38	Y	IMMU Watchpoint register	19.2.2
51_{16}	ASI_MRA_ACCESS	RO		Y	HWTW MRA access	19.13
52_{16}	ASI_MMU_REAL_RANGE_0	RW	108_{16}	Y	MMU TSB real range 0	12.10.9
52_{16}	ASI_MMU_REAL_RANGE_1	RW	110_{16}	Y	MMU TSB real range 1	12.10.9
52_{16}	ASI_MMU_REAL_RANGE_2	RW	118_{16}	Y	MMU TSB real range 2	12.10.9
52_{16}	ASI_MMU_REAL_RANGE_3	RW	120_{16}	Y	MMU TSB real range 3	12.10.9
52_{16}	ASI_MMU_PHYSICAL_OFFSET_0	RW	208_{16}	Y	MMU TSB physical offset 0	12.10.10
52_{16}	ASI_MMU_PHYSICAL_OFFSET_1	RW	210_{16}	Y	MMU TSB physical offset 1	12.10.10
52_{16}	ASI_MMU_PHYSICAL_OFFSET_2	RW	218_{16}	Y	MMU TSB physical offset 2	12.10.10
52_{16}	ASI_MMU_PHYSICAL_OFFSET_3	RW	220_{16}	Y	MMU TSB physical offset 3	12.10.10
53_{16}	ASI_ITLB_PROBE	RO		Y	ITLB Probe	12.10.8
54_{16}	ASI_ITLB_DATA_IN_REG	WO	0, 400_{16}	Y	IMMU data in register	12.10.15
54_{16}	ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_0	RW	10_{16}	Y	Context zero TSB Config 0	12.10.11
54_{16}	ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_1	RW	18_{16}	Y	Context zero TSB Config 1	12.10.11
54_{16}	ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_2	RW	20_{16}	Y	Context zero TSB Config 2	12.10.11
54_{16}	ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_3	RW	28_{16}	Y	Context zero TSB Config 3	12.10.11

TABLE 9-2 UltraSPARC T2 ASI Usage *(7 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
54_{16}	ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_0	RW	30_{16}	Y	Context nonzero TSB Config 0	12.10.11
54_{16}	ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_1	RW	38_{16}	Y	Context nonzero TSB Config 1	12.10.11
54_{16}	ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_2	RW	40_{16}	Y	Context nonzero TSB Config 2	12.10.11
54_{16}	ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_3	RW	48_{16}	Y	Context nonzero TSB Config 3	12.10.11
54_{16}	ASI_MMU_ITSB_PTR_0	RO	50_{16}	Y	I-ITSB Pointer 0	12.10.12
54_{16}	ASI_MMU_ITSB_PTR_1	RO	58_{16}	Y	I-TSB Pointer 1	12.10.12
54_{16}	ASI_MMU_ITSB_PTR_2	RO	60_{16}	Y	I-TSB Pointer 2	12.10.12
54_{16}	ASI_MMU_ITSB_PTR_3	RO	68_{16}	Y	I-TSB Pointer 3	12.10.12
54_{16}	ASI_MMU_DTSB_PTR_0	RO	70_{16}	Y	D-TSB Pointer 0	12.10.12
54_{16}	ASI_MMU_DTSB_PTR_1	RO	78_{16}	Y	D-TSB Pointer 1	12.10.12
54_{16}	ASI_MMU_DTSB_PTR_2	RO	80_{16}	Y	D-TSB Pointer 2	12.10.12
54_{16}	ASI_MMU_DTSB_PTR_3	RO	88_{16}	Y	D-TSB Pointer 3	12.10.12
54_{16}	ASI_PENDING_TABLEWALK_CONTROL	RW	90_{16}	Y	Pending tablewalk control	12.10.13
54_{16}	ASI_PENDING_TABLEWALK_STATUS	RO	98_{16}	N	Pending Tablewalk status	12.10.14
55_{16}	ASI_ITLB_DATA_ACCESS_REG	RW	0–$1F8_{16}$, 400_{16}–5F8	Y	IMMU TLB Data Access register	12.10.15
56_{16}	ASI_ITLB_TAG_READ_REG	RO	0–$1F8_{16}$, 400_{16}–$5F8_{16}$	Y	IMMU TLB Tag Read register	12.10.15
57_{16}	ASI_IMMU_DEMAP	WO	Any	Y	IMMU TLB demap	12.11.1
58_{16}	ASI_DTSB_TAG_TARGET	RO	0	Y	DMMU Tag Target register	12.10.3
58_{16}	ASI_DSFSR	RW	18_{16}	Y	DMMU Synchronous Fault Status register	16.8.4.1
58_{16}	ASI_DSFAR	RO	20_{16}	Y	DMMU Synchronous Fault Address register	16.8.4.2
58_{16}	ASI_DTLB_TAG_ACCESS	RW	30_{16}	Y	DMMU TLB Tag Access register	12.10.5
58_{16}	ASI_DMMU_WATCHPOINT	RW	38_{16}	Y	DMMU Watchpoint register	19.2.1
58_{16}	ASI_HWTW_CONFIG	RW	40_{16}	Y	I/DMMU Hardware Tablewalk Config register	12.10.7
58_{16}	ASI_PARTITION_ID	RW	80_{16}	Y	I/DMMU Partition ID	12.10.6

TABLE 9-2 UltraSPARC T2 ASI Usage *(8 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
59_{16}	ASI_SCRATCHPAD_ ACCESS	RO	Any	Y	Scratchpad register diagnostic access	19.10
$5A_{16}$	ASI_TICK_ACCESS	RO	$0–8_{16}$, 10_{16}, $20_{16}–30_{16}$	Y	TICK register diagnostic access	19.11
$5B_{16}$	ASI_TSA_ACCESS	RO	Any	Y	TSA diagnostic access	19.12
$5C_{16}$	ASI_DTLB_DATA_IN_REG	WO	$0,400_{16}$	Y	DMMU data-in register	12.10.15
$5D_{16}$	ASI_DTLB_DATA_ACCESS_R EG	RW	$0–7F8_{16}$	Y	DMMU TLB Data Access register	12.10.15
$5E_{16}$	ASI_DTLB_TAG_READ_REG	RO	$0–7F8_{16}$	Y	DMMU TLB Tag Read register	12.10.15
$5F_{16}$	ASI_DMMU_DEMAP	WO	Any	Y	DMMU TLB demap	12.11
$60_{16}–62_{16}$			Any		*DAE_invalid_asi*	
63_{16}	ASI_CMT_CORE_INTR_ID	RO	0	Y	Strand interrupt ID	14.2.1
63_{16}	ASI_CMT_STRAND_ID	RO	10_{16}	Y	Strand ID	14.2.2
$64_{16}–65_{16}$			Any		*DAE_invalid_asi*	
66_{16}	ASI_ICACHE_INSTR	RW	Any	Y	Icache data array diagnostics access	19.5.1
67_{16}	ASI_ICACHE_TAG	RW	Any	Y	Icache tag and valid bit diagnostics access	19.5.2
$68_{16}–71_{16}$			Any		*DAE_invalid_asi*	
72_{16}	ASI_INTR_RECEIVE	RW	0	Y	Interrupt Receive register	7.3.2
73_{16}	ASI_INTR_W	WO	0	Y	Interrupt Vector Dispatch register	7.3.3
74_{16}	ASI_INTR_R	RO	0	Y	Incoming Vector register	7.3.4
$75_{16}–7F_{16}$			Any	—	*DAE_invalid_asi*	
80_{16}	ASI_PRIMARY	RW	Any	—	Implicit primary address space	(See UA-2007)
81_{16}	ASI_SECONDARY	RW	Any	—	Implicit secondary address space	(See UA-2007)
82_{16}	ASI_PRIMARY_NO_FAULT	RO	Any	—	Primary address space, no fault	(See UA-2007)
83_{16}	ASI_SECONDARY_NO_ FAULT	RO	Any	—	Secondary address space, no fault	(See UA-2007)
$84_{16}–87_{16}$			Any	—	*DAE_invalid_asi*	
88_{16}	ASI_PRIMARY_LITTLE	RW	Any	—	Implicit primary address space (LE)	(See UA-2007)
89_{16}	ASI_SECONDARY_LITTLE	RW	Any	—	Implicit secondary address space (LE)	(See UA-2007)

TABLE 9-2 UltraSPARC T2 ASI Usage *(9 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
$8A_{16}$	ASI_PRIMARY_NO_FAULT_LITTLE	RO	Any	—	Primary address space, no fault (LE)	(See UA-2007)
$8B_{16}$	ASI_SECONDARY_NO_FAULT_LITTLE	RO	Any	—	Secondary address space, no fault (LE)	(See UA-2007)
$8C_{16}$–BF_{16}			Any	—	*DAE_invalid_asi*	
$C0_{16}$	ASI_PST8_P		Any	—	Eight 8-bit conditional stores, primary address	(See UA-2007)
$C1_{16}$	ASI_PST8_S		Any	—	Eight 8-bit conditional stores, secondary address	(See UA-2007)
$C2_{16}$	ASI_PST16_P		Any	—	Four 16-bit conditional stores, primary address	(See UA-2007)
$C3_{16}$	ASI_PST16_S		Any	—	Four 16-bit conditional stores, secondary address	(See UA-2007)
$C4_{16}$	ASI_PST32_P		Any	—	Two 32-bit conditional stores, primary address	(See UA-2007)
$C5_{16}$	ASI_PST32_S		Any	—	Two 32-bit conditional stores, secondary address	(See UA-2007)
$C6_{16}$–$C7_{16}$			Any	—	*DAE_invalid_asi*	
$C8_{16}$	ASI_PST8_PL		Any	—	Eight 8-bit conditional stores, primary address, little endian	(See UA-2007)
$C9_{16}$	ASI_PST8_SL		Any	—	Eight 8-bit conditional stores, secondary address, little endian	(See UA-2007)
CA_{16}	ASI_PST16_PL		Any	—	Four 16-bit conditional stores, primary address, little endian	(See UA-2007)
CB_{16}	ASI_PST16_SL		Any	—	Four 16-bit conditional stores, secondary address, little endian	(See UA-2007)
CC_{16}	ASI_PST32_PL		Any	—	Two 32-bit conditional stores, primary address, little endian	(See UA-2007)
CD_{16}	ASI_PST32_SL		Any	—	Two 32-bit conditional stores, secondary address, little endian	(See UA-2007)
CE_{16}–CF_{16}			Any	—	*DAE_invalid_asi*	
$D0_{16}$	ASI_FL8_P		Any	—	8-bit load/store, primary address	(See UA-2007)
$D1_{16}$	ASI_FL8_S		Any	—	8-bit load/store, secondary address	(See UA-2007)

TABLE 9-2 UltraSPARC T2 ASI Usage *(10 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
$D2_{16}$	ASI_FL16_P		Any	—	16-bit load/store, primary address	(See UA-2007)
$D3_{16}$	ASI_FL16_S		Any	—	16-bit load/store, secondary address	(See UA-2007)
$D4_{16}$–$D7_{16}$			Any	—	*DAE_invalid_asi*	
$D8_{16}$	ASI_FL8_PL		Any	—	8-bit load/store, primary address, little endian	(See UA-2007)
$D9_{16}$	ASI_FL8_SL		Any	—	8-bit load/store, secondary address, little endian	(See UA-2007)
DA_{16}	ASI_FL16_PL		Any	—	16-bit load/store, primary address, little endian	(See UA-2007)
DB_{16}	ASI_FL16_SL		Any	—	16-bit load/store, secondary address, little endian	(See UA-2007)
DC_{16}–DF_{16}			Any	—	*DAE_invalid_asi*	
$E0_{16}$	ASI_BLK_COMMIT_PRIMARY	RW	Any	—	64-byte block commit store, primary address	5.3
$E1_{16}$	ASI_BLK_COMMIT_SECONDARY	RW	Any	—	64-byte block commit store, secondary address	5.3
$E2_{16}$	ASI_TWINX_P, ASI_STBI_P	RW	Any	—	Load: 128-bit atomic load twin extended word, primary address space Store: Block initializing store, primary address space	(See UA-2007)
$E3_{16}$	ASI_TWINX_S, ASI_STBI_S	RW	Any	—	Load: 128-bit atomic load twin extended word, sedondary address space Store: Block initializing store, sedondary address space	(See UA-2007)
$E4_{16}$–$E9_{16}$			Any	—	*DAE_invalid_ASI*	
EA_{16}	ASI_TWINX_PL, ASI_STBI_PL	RW	Any	—	Load: 128-bit atomic load twin extended word, primary address space, little endian Store: Block initializing store, primary address space, little endian	(See UA-2007)

TABLE 9-2 UltraSPARC T2 ASI Usage *(11 of 11)*

ASI	ASI Name	R/W	VA	Copy per Strand	Description	Section/Page
EB_{16}	ASI_TWINX_PL, ASI_STBI_PL	RW	Any	—	Load: 128-bit atomic load twin extended word, sedondary address space, little endian Store: Block initializing store, sedondary address space, little endian	(See UA-2007)
EC_{16}–EF_{16}			Any	—	*DAE_invalid_asi*	
$F0_{16}$	ASI_BLK_P	RW	Any	—	64-byte block load/store, primary address	5.3
$F1_{16}$	ASI_BLK_S	RW	Any	—	64-byte block load/store, secondary address	5.3
$F2_{16}$–$F7_{16}$			Any	—	*DAE_invalid_asi*	
$F8_{16}$	ASI_BLK_PL	RW	Any	—	64-byte block load/store, primary address (LE)	5.3
$F9_{16}$	ASI_BLK_SL	RW	Any	—	64-byte block load/store, secondary address (LE)	5.3
FA_{16}–FF_{16}			Any	—	*DAE_invalid_asi*	

9.2.1 ASI_REAL, ASI_REAL_LITTLE, ASI_REAL_IO, and ASI_REAL_IO_LITTLE

These ASIs are used to bypass the VA-to-RA translation. For these ASIs, the real address is set equal to the truncated virtual address (that is, RA{39:0} ← VA{39:0}), and the attributes used are those present in the matching TTE. The hypervisor will normally set the TTE attributes for ASI_REAL and ASI_REAL_LITTLE to cacheable (cp = 1) and for ASI_REAL_IO and ASI_REAL_IO_LITTLE to noncacheable, with side effect (cp = 0, e = 1).

9.2.2 ASI_SCRATCHPAD

Each virtual processor has a set of six privileged ASI_SCRATCHPAD registers at ASI 20_{16} with VA{63:} = 0_{16}–18_{16}, 30_{16}–38_{16}. These registers are for scratchpad use by privileged software.

UltraSPARC T2 Implementation Note Standard support of the ASI_SCRATCHPAD is eight registers, so accesses to VA 20_{16} and 28_{16} cause a *DAE_invalid_asi* trap to allow hyperprivileged software to emulate the additional registers.

UltraSPARC T2 Implementation Note

There is only a single set of eight scratchpad registers, which are accessible via both ASI_SCRATCHPAD and ASI_HYP_SCRATCHPAD. ASI_SCRATCHPAD is intended to used primarily by privileged code, and only has access to the first four and last two registers of the eight entry scratchpad array. ASI_HYP_SCRATCHPAD can only be accessed when hyperprivileged and has full access to all eight scratchpad registers. Note that the registers at VA 20_{16} and 28_{16} are exclusively (directly) accessible via ASI_HYP_SCRATCHPAD.

9.2.3 ASI_HYP_SCRATCHPAD

Each virtual processor has a set of eight hyperprivileged ASI_HYP_SCRATCHPAD registers at ASI $4F_{16}$, VA{63:0} = 0_{16}–$3_{16}8$. These registers are for scratchpad use by the hypervisor and for aliased access to the supervisor scratchpad registers.

UltraSPARC T2 Implementation Note

There is only a single set of eight scratchpad registers, which are accessible via both ASI_SCRATCHPAD and ASI_HYP_SCRATCHPAD. ASI_SCRATCHPAD is intended to used primarily by privileged code and only has access to the first four and last two registers of the eight entry scratchpad array. ASI_HYP_SCRATCHPAD can only be accessed when hyperprivileged, and has full access to all eight scratchpad registers. Note that the registers at VA 20_{16} and 28_{16} are exclusively (directly) accessible via ASI_HYP_SCRATCHPAD.

CHAPTER **10**

Performance Instrumentation

10.1 SPARC Performance Control Register

Each virtual processor has a privileged Performance Control register. Nonprivileged accesses to this register cause a *privileged_opcode* trap. The Performance Control register contains thirteen fields: hold_ov1, hold_ov0, ov1, sl1, mask1, ov0, sl0, mask0, toe, ht, ut, st, and priv. hold_ov1 and hold_ov0 read as 0 and control whether ov1 and ov0, respectively, are updated on a write. All bits except ov1 and ov0 are always updated on a Performance Control register write. ov1 and ov0 are state bits associated with the PIC.h and PIC.l overflow traps and are provided to allow software to determine which PIC counter has overflowed. sl1 and sl0 controls which events are counted in PIC.h and PIC.l, respectively. mask1 (mask0) is used in conjunction with sl1 (sl0) in determining which set of subevents are counted in PIC.h (PIC.l). toe controls whether a trap is generated when the PIC counter overflows. ut controls whether user-level events are counted. st controls whether supervisor-level events are counted. ht controls whether hypervisor level events are counted. priv controls whether the PIC register can be read or written by nonprivileged software. The format of this register is shown in TABLE 10-1. Note that changing the fields in PCR does not affect the PIC values. To change the events monitored, software needs to disable counting via PCR, reset the PIC, and then enable the new event via the PCR.

Note | As the ht bit controls the counting of hyperprivileged events, writes to this bit while privileged are ignored.

TABLE 10-1 Performance Control Register – PCR (ASR 10_{16})

Bit	Field	Initial Value	R/W	Description
63	hold_ov1	0	Write to 0 or 1, reads as 0	If set to 0 on a write, update ov1 from bit 31 of the write data; else, don't update ov1. In this case ov1 holds its previous value.
62	hold_ov0	0	Write to 0 or 1, reads as 0	If set to 0 on a write, update ov0 from bit 18 of the write data; else, don't update ov0. In this case ov0 holds its previous value.
61:32	—	0	RO	*Reserved*
31	ov1	0	RW	Set to 1 when PIC.h wraps from 2^{32} –1 to 0, or when PIC.h is within --16..-1 inclusively, and an event occurs which causes PIC.h to increment. Once set, ov1 remains set until reset by software.
30:27	sl1	0	RW	Selects 1 of 16 events to be counted for PIC.h as per the following table.
26:19	mask1	0	RW	Mask event for PIC.h as listed in TABLE 10-2.
18	ov0	0	RW	Set to 1 when PIC.l wraps from 2^{32} –1 to 0, or when PIC.l is within --16..-1 inclusively, and an event occurs which causes PIC.l to increment. Once set, ov0 remains set until reset by software.
17:14	sl0	0	RW	Selects one of sixteen events to be counted for PIC.l as per the following table.
13:6	mask0	0	RW	Mask event for PIC.l as listed in TABLE 10-2.
5:4	toe	0	RW	Trap-on-Event: This field controls whether a disrupting trap to hyperprivileged software (*pic_overflow*) will occur if the corresponding counter overflows. toe{1} corresponds to ov1, and toe{0} to ov0. Hardware will **and** the value of toe{*i*} with ov{*i*} to produce a trap. Events in event groups 2) and 3) are "precisely" trapped, assuming that PCR.toe = 1 — TPC will contain the address of an instruction that generated a count event. If PCR.toe = 0 when the counter overflows, TPC will contain the address of the instruction to be executed next when the trap is eventually taken. Events in other event groups are not directly related to the instruction stream; therefore, the TPC may be some number of instructions later than when the overflow event occurred.
3	ht	0	RO (priv) RW (hyperpriv)	If ht = 1, count events in hyperprivileged mode; otherwise, ignore hyperprivileged mode events.
2	ut	0	RW	If ut = 1, count events in user mode; otherwise, ignore user mode events.
1	st	0	RW	If st = 1, count events in privileged mode; otherwise, ignore privileged mode events.
0	priv	0	RW	If priv = 1, prevent access to PIC by user-level code. If priv = 0, allow access to PIC by user-level code.

Note that hold_ov1 and hold_ov0 control whether ov1 and ov0, respectively, are updated when a write occurs. All of the 4 combinations of the hold_ov1 and hold_ov0 fields are supported: both, either, or none of the ov1 and ov0 bits can be updated independently. This allows software to avoid a race condition that may occur, for example, if ov1 is set when the trap handler is entered, then ov0 is set by a counter overflow between the time software reads PCR and resets ov1 prior to leaving the trap handler.

TABLE 10-2 describes the settings of the sl0 and sl1 fields. Note that with the exception of sl = 0, all events correspond to a given strand. Most sl fields have a mask associated with them. Setting multiple mask bits at the same time can lead to multiple events being counted as one event. More details are described in TABLE 10-2.

TABLE 10-2 sl Field Settings *(1 of 3)*

sl	mask	Event	Description
0	—	All strands idle	Count cycles when no strand can be picked for the physical core on which the monitoring strand resides.[2]
1	—	—	*Reserved*
2	01_1	Completed branches	
	02_{16}	Taken branches	Taken branches are always mispredicted[3]
	04_{16}	FGU arithmetic instructions	All FADD, FSUB, FCMP, convert, FMUL, FDIV, FNEG, FABS, FSQRT, FMOV, FPADD, FPSUB, FPACK, FEXPAND, FPMERGE, FMUL8, FMULD8, FALIGNDATA, BSHUFFLE, FZERO, FONE, FSRC, FNOT1, FNOT2, FOR, FNOR, FAND, FNAND, FXOR, FXNOR, FORNOT1, FORNOT2, FANDNOT1, FANDNOT2, PDIST, SIAM.
	08_{16}	Load instructions	
	10_{16}	Store instructions	
	20_{16}	`sethi %hi(fc000`$_{16}$`), %g0`	Software count instructions.
	40_{16}	Other instructions	
	80_{16}	Atomics	Atomics are LDSTUB/A, CASA/XA, SWAP/A
	Any other value 03_{16}–FF_{16}	Any subset of instructions	Count instruction types identified by a 1 in the corresponding mask register bit; e.g., FD_{16} counts all instructions. Certain instructions (e.g., LDSTUB, CAS, SWAP) are decoded as both Load and Store instructions.

TABLE 10-2 sl Field Settings *(2 of 3)*

sl	mask	Event	Description
3	01_{16}	Icache misses	**Note:** This counts only primary instruction cache misses, and does not count duplicate instruction cache misses.[4] Also, only "true" misses are counted. If a thread encounters an I$ miss, but the thread is redirected (due to a branch misprediction or trap, for example) before the line returns from L2 and is loaded into the I$, then the miss is not counted.
	02_{16}	Dcache misses	Note: This counts both primary and duplicate data cache misses.[4]
	04_{16}	—	Undefined operation.
	08_{16}	—	Undefined operation.
	10_{16}	L2 cache instruction misses	
	20_{16}	L2 cache load misses	**Note:** Block loads are treated as one L2 miss event. In reality, each individual load can hit or miss in the L2 since the block load is not atomic.
	03_{16}, 11_{16}, 12_{16}, 13_{16}, 21_{16}, 22_{16}, 23_{16}, 30_{16}, 31_{16}, 32_{16}, 33_{16}	Subset of misses	Count subset of misses identified by a '1' in corresponding mask bit; e.g., 23_{16} counts I-cache, D-cache, and L2 load misses; this counter can advance at most 1 per cycle. **Note:** Instructions that get both an I-Cache miss (or an L2 cache instruction miss) and a D-Cache miss (or L2 cache load miss) count as one event.
	Any other value	—	*Reserved*, Undefined operation.
4	01_{16}	—	*Reserved*
	02_{16}	—	*Reserved*
	04_{16}	ITLB references to L2	For each ITLB miss with hardware tablewalk enabled, count each access the ITLB hardware tablewalk makes to L2.
	08_{16}	DTLB references to L2	For each DTLB miss with hardware tablewalk enabled, count each access the DTLB hardware tablewalk makes to L2.
	10_{16}	ITLB references to L2 which miss in L2	For each ITLB miss with hardware tablewalk enabled, count each access the ITLB hardware tablewalk makes to L2 which misses in L2. **Note:** Depending upon the hardware table walk configuration, each ITLB miss may issue from 1 to 4 requests to L2 to search TSBs.
	20_{16}	DTLB references to L2 which miss in L2	For each DTLB miss with hardware tablewalk enabled, count each access the DTLB hardware tablewalk makes to L2 which misses in L2. **Note:** Depending upon the hardware tablewalk configuration, each DTLB miss may issue from 1 to 4 requests to L2 to search TSBs.
	C_{16}, 14_{16}, 18_{16}, $1C_{16}$, 24_{16}, 28_{16}, $2C_{16}$, 34_{16}, 38_{16} $3C_{16}$	Subset of above events	Count subset of misses identified by a 1 in corresponding mask bit; e.g., 14_{16} counts ITLB and DTLB hardware tablewalk references to L2; this counter can advance at most 1 per cycle. Certain combinations (14_{16}, 28_{16}, 34_{16}, 38_{16}, $3C_{16}$) are likely not useful.
	Any other value	—	*Reserved*. Undefined operation.

TABLE 10-2 sl Field Settings *(3 of 3)*

sl	mask	Event	Description
5			
	00_{16}–01_{16}	—	*Reserved*
	04_{16}	CPU Load to PCX	Count CPU loads to L2.
	08_{16}	CPU I-fetch to PCX	Count I-fetches to L2.
	10_{16}	CPU Store to PCX	Count CPU stores to L2.
	20_{16}	MMU Load to PCX	Count MMU loads to L2.
	Any other value 03_{16}–$3F_{16}$	Subset of PCX requests	Count subset of PCX requests identified by a '1' in corresponding mask bit; e.g., $3F_{16}$counts all PCX requests; this counter increments at most one per cycle.
	40_{16}–FF_{16}	—	*Reserved*
6[1]			
	00_{16}–$3F_{16}$	—	
	40_{16}–FF_{16}	—	*Reserved*
7[1]			
	00_{16}–$3F_{16}$	—	
	40_{16}–FF_{16}	—	*Reserved*
8-10	—	—	*Reserved*
11	04_{16}	ITLB misses	Includes all misses (successful and unsuccessful tablewalks).
	08_{16}	DTLB misses	Includes all misses (successful and unsuccessful tablewalks).
	$0C_{16}$	TLB misses	Count both ITLB and DTLB misses, including successful and unsuccessful tablewalks.
	Any other value	—	*Reserved.* Undefined operation.
12-15	—	—	*Reserved*

1. PCR.UT, PCR.HT, and PCR.ST must all be set in order to properly count events in groups 6 and 7.
2. Unrestricted access to performance events for sl field setting 0 may have security implications since they contain information about other strands. Privileged software can protect against unrestricted access by setting the PCR.priv bit. Hyperprivileged software can protect against unrestricted access by not having partitions span an eight-strand boundary.
3. In conjunction with the completed branch count, the taken branch count can be used to compute not-taken prediction accuracy. Also it can be used to sum idle cycles in single-strand mode by assuming a fixed number of pipeline bubble cycles per mispredicted branch.
4. A duplicate miss is a miss for which another thread has already missed in the cache for the line, and the cache fill is pending. UltraSPARC {N2} does not count duplicate I-cache misses but does count duplicate D-cache misses.

10.2 SPARC Performance Instrumentation Counter

Each virtual processor has a Performance Instrumentation Counter register. Access privilege is controlled by the setting of PCR.priv. When PCR.priv = 1 an attempt to access this register in nonprivileged mode causes a *privileged_action* trap.

The PIC counter contains two fields: h and l. The h field counts the event select by PCR.sl1. The l field counts the event selected by PCR.sl0. The ut, st, and ht fields for PCR control which combination of user, supervisor, and/or hypervisor events are counted.

For the setting sl0 (sl1) = 2 and sl0 (sl1) = 3, when a counter overflow occurs for the event, hardware generates a disrupting *pic_overflow* trap that is guaranteed to occur immediately before an instruction that generated a count event. This instruction causing the counter to be within epsilon[1] of overflow will not have been executed, and the PC and NPC of the instruction will be stored on the trap stack, assuming the *pic_overflow* trap is enabled and is the highest priority trap when the counter overflows[2]. The corresponding PIC counter will be incremented. In addition, the ov0 or ov1 bit (depending on which counter overflowed) will be set to help software determine which counter overflowed. ov0 and ov1 can be cleared independently by a write that sets the bit to 0 (see TABLE 10-1 on page 86 above).

For other settings of sl0 (sl1), the trap will not be “precise” to the instruction causing the counter overflow. The amount of skid possible is TBD.

Counter overflow is recorded in the ov0 or ov1 bit of the counter as well as in bit 15 of the SOFTINT register. The overflow causes a disrupting *pic_overflow* exception. The strand takes a *pic_overflow* trap if PSTATE.ie is set and the value in the Processor Interrupt Level (PIL) is less than 15. The *pic_overflow* priority of 16.0 is higher than the *interrupt_level_15* trap priority of 17.

The format of the PIC register is shown in TABLE 10-3.

1. The definition of epsilon is –16 to -1, inclusive, instructions generating the event being counted before/after the overflow. The PC is guaranteed to point to an instruction that generated the event being counted, if the trap is taken when the counter "overflows".

2. In certain corner cases, the counter will not be incremented nor will the corresponding ov0/1 bit be set. These cases are: [a) the counter is in range and the instruction will cause the counter to increment, or, b) the OV bit is already set and the instruction will cause the counter to increment,] and c) one of the following four pending disrupting conditions are present: i) "disrupting single step completion" exception (this only occurs for special ’replay’ conditions that occur in Single Step mode, which is a debug mode controlled via JTAG), or ii) XIR request, or iii) store_error trap request, or iv) SIR (i.e., this instruction is an SIR instruction).

TABLE 10-3 Performance Instrumentation Counter Register – PIC (ASR 11_{16})

Bit	Field	Initial Value	R/W	Description
63:32	h	0	RW	Programmable event counter, event controlled by PCR.sl1.
31:0	l	0	RW	Programmable event counter, event controlled by PCR.sl0.

10.3 DRAM Performance Counter

Each DRAM channel has a pair of performance counters, packed into a single register, plus a register to control what is counted. The counters count all events for that particular DRAM channel, which corresponds to traffic from a pair of L2 banks.

TABLE 10-4 DRAM Performance Control Register – DRAM_PERF_CTL_REG (84_{16}-0000_{16}-0400_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	X	RO	*Reserved*
7:4	sel0	0	RW	Select code for performance counter 0.
3:0	sel1	0	RW	Select code for performance counter 1.

TABLE 10-5 DRAM Performance Counter Register – DRAM_PERF_COUNT_REG (84_{16}-0000_{16}-0408_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63	sticky0	0	RW	Sticky overflow for counter 0.
62:32	counter0	0	RW	Performance counter 0
31	sticky1	0	RW	Sticky overflow for counter 1.
30:0	counter1	0	RW	Performance counter 1.

TABLE 10-6 DRAM Performance Counter Select Codes

Select	Description
0000	Read transactions.
0001	Write transactions.
0010	Read + write transactions.
0011	Bank busy stalls; incremented by 1 each cycle there are requests in the queue, but none can issue because of bank conflicts

TABLE 10-6 DRAM Performance Counter Select Codes

Select	Description
0100	Read queue latency; incremented by *n* each cycle, where *n* is the number of read transactions in the queue.
0101	Write queue latency; incremented by *n* each cycle, where *n* is the number of write transactions in the queue
0110	(Read + Write) queue latency; incremented by *n* each cycle, where *n* is the number of transactions in the queue
0111	Writeback buffer hits; incremented by 1 each time a read transaction is deferred because it conflicts with a queued write transaction.
1xxx	*Reserved*

10.4 PCI-EX Performance Counters

The PCI-Express performance counters are documented in *Performance Register Overview* on page 526

10.5 Ethernet Performance Counters

Ethernet performance counters are described in *DRR Performance Monitoring* on page 766, *Port Scheduler* on page 703, and *Receive Performance Management and Discard Statistics* on page 725.

CHAPTER 11

Implementation Dependencies

11.1 SPARC V9 General Information

11.1.1 Level-2 Compliance (Impdep #1)

UltraSPARC T2 is designed to meet Level-2 SPARC V9 compliance. It

- Correctly interprets all nonprivileged operations, and
- Correctly interprets all privileged elements of the architecture.

Note | System emulation routines (for example, quad-precision floating-point operations) shipped with UltraSPARC T2 also must be Level-2 compliant.

11.1.2 Unimplemented Opcodes, ASIs, and ILLTRAP

SPARC V9 unimplemented, *reserved*, ILLTRAP opcodes, and instructions with invalid values in *reserved* fields (other than *reserved* FPops) encountered during execution cause an *illegal_instruction* trap. Unimplemented and *reserved* ASI values cause a *DAE_invalid_ASI* trap.

11.1.3 Trap Levels (Impdep #37, 38, 39, 40, 114, 115)

UltraSPARC T2 supports two privileged trap levels and six hyperprivileged trap levels; that is, *MAXPTL* = 2 and at *MAXTL* = 6. Normal execution is at TL = 0. Traps at *MAXTL* − 1 cause the virtual processor to enter `RED_state`. If a trap is generated while the virtual processor is operating at TL = *MAXTL*, the virtual processor will pass through `error_state` and generate a watchdog reset (WDR). Window traps that cause a watchdog reset trap still update CWP if they would have done so with no watchdog trap being generated.

A virtual processor normally executes at trap level 0 (`execute_state`, TL = 0). Per SPARC V9, a trap causes the virtual processor to enter the next higher trap level, which is a very fast and efficient process because there is one set of trap state registers for each trap level. After saving the most important machine states (PC, NPC, PSTATE) on the trap stack at this level, the trap (or error) condition is processed.

For a complete description of traps and `RED_state` handling, see *Machine State After Reset and in RED_State* on page 171.

Note | The `RED_state` trap vector address (*RSTVADDR*) is 256 Mbytes below the top of the virtual address space; this is, at virtual address FFFF FFFF F000 0000_{16}, which is passed through to physical address FF F000 0000_{16} in `RED_state`.

11.1.4 Trap Handling (Impdep #16, 32, 33, 35, 36, 44)

UltraSPARC T2 supports precise trap handling for all operations except for deferred and disrupting traps from hardware failures and interrupts. UltraSPARC T2 implements precise traps, interrupts, and exceptions for all instructions, including long-latency floating-point operations. Multiple traps levels are supported, allowing graceful recovery from faults. Three of the trap levels (zero through two) are provided for application and OS use. The remaining three levels are provided for hyperprivileged and `RED_state` use. UltraSPARC T2 can efficiently execute kernel code even in the event of multiple nested traps, promoting strand efficiency while dramatically reducing the system overhead needed for trap handling.

Four sets of global registers are provided, for use by TL0, TL1, TL2, and TL3-5. This further increases OS performance, providing fast trap execution by avoiding the need to save and restore registers while processing exceptions.

All traps supported in UltraSPARC T2 are listed in TABLE 6-2 on page 40.

11.1.5 SIR Support (Impdep #116)

UltraSPARC T2 initiates a software-initiated reset (SIR) by executing a SIR instruction while in hyperprivileged mode. When executed in privileged or user mode, SIR generates an *illegal_instruction* trap. See also *Watchdog Reset (WDR) and error_state* on page 170.

11.1.6 Secure Software

To establish an enhanced security environment, it may be necessary to initialize certain virtual processor states between contexts. Examples of such states are the contents of integer and floating-point register files, condition codes, and state registers. See also *Clean Window Handling (Impdep #102)*.

11.1.7 Operation in Nonprivileged Mode with TL > 0

Operation with HPSTATE.hpriv = 0, PSTATE.priv = 0, and TL > 0 is invalid and will result in an *IAE_privilege_violation* trap on UltraSPARC T2.

11.1.8 Address Masking (Impdep #125)

UltraSPARC T2 follows UltraSPARC Architecture 2007 for PSTATE.am masking. In addition to the masking required by UltraSPARC Architecture 2007, addresses to non-translating ASIs and *REAL* ASIs are masked if PSTATE.am = 1. Translating accesses that bypass translation are also masked if PSTATE.am = 1.

11.2 SPARC V9 Integer Operations

11.2.1 Integer Register File and Window Control Registers (Impdep #2)

UltraSPARC T2 implements an eight-window 64-bit integer register file; that is, *N_REG_WINDOWS* = 8. UltraSPARC T2 truncates values stored in the CWP, CANSAVE, CANRESTORE, CLEANWIN, and OTHERWIN registers to three bits. This includes implicit updates to these registers by SAVE, SAVED, RESTORE, and RESTORED instructions. The most significant two bits of these registers read as zero.

11.2.2 Clean Window Handling (Impdep #102)

SPARC V9 introduced the concept of “clean window” to enhance security and integrity during program execution. A clean window is defined to be a register window that contains either all zeroes or addresses and data that belong to the current context. The CLEANWIN register records the number of available clean windows.

When a SAVE instruction requests a window and there are no more clean windows, a *clean_window* trap is generated. System software needs to clean one or more windows before returning to the requesting context.

11.2.3 Integer Multiply and Divide

Integer multiplications (MULScc, SMUL{cc}, MULX) and divisions (SDIV{cc}, UDIV{cc}, UDIVX) are executed directly in hardware.

11.2.4 MULScc

SPARC V9 does not define the value of xcc and rd{63:32] for MULScc. UltraSPARC T2 sets xcc.n to 0, xcc.z to 1 if rd{63:0} is zero and to 0 if rd{63:0} is not zero, xcc.v to 0, and xcc.c to 0. UltraSPARC T2 sets rd{63:33} to zeros, and sets rd{32} to icc.c (that is, rd{32} is set if there is a carry-out of rd{31}; otherwise, it is cleared).

11.2.5 Version Register (Impdep #2, 13, 101, 104)

Consult the product data sheet for the contents of the Version register for a specific UltraSPARC T2 implementation. The format of the Version register is described in *Hyperprivileged Version Register (hver)* on page 21.

11.3 SPARC V9 Floating-Point Operations

11.3.1 Subnormal Operands and Results; Nonstandard Operation

UltraSPARC T2 handles some cases of subnormal operands or results directly in hardware and traps on the rest. In the trapping cases, an *fp_exception_other* [fft = unfinished_FPop] trap is signaled and these operations are handled in system software.

Because trapping on subnormal operands and results can be quite costly, UltraSPARC T2 supports the nonstandard result option of the SPARC-V9 architecture. When the FSR.ns bit is set, subnormal operands or results encountered in trapping cases are flushed to zero and the unfinished_FPop floating-point trap is not taken.

11.3.2 Overflow, Underflow, and Inexact Traps (Impdep #3, 55)

UltraSPARC T2 implements precise floating-point exception handling. Underflow is detected before rounding. Prediction of overflow, underflow, and inexact traps for operations as well as prediction of invalid operation is used to simplify the hardware.

> Significant performance degradation may be observed while running with the inexact exception enabled.

11.3.3 Quad-Precision Floating-Point Operations (Impdep #3)

All quad-precision floating-point instructions, listed in TABLE 11-1, cause an *illegal_instruction* trap. These operations are then emulated by system software.

TABLE 11-1 Unimplemented Quad-Precision Floating-Point Instructions

Instruction	Description
F<s|d>TOq	Convert single-/double- to quad-precision floating-point.
F<i|x>TOq	Convert 32-/64-bit integer to quad-precision floating-point.
FqTO<s|d>	Convert quad- to single-/double-precision floating-point.
FqTO<i|x>	Convert quad-precision floating-point to 32-/64-bit integer.
FCMP<E>q	Quad-precision floating-point compares.
FMOVq	Quad-precision floating-point move.
FMOVqcc	Quad-precision floating-point move if condition is satisfied.
FMOVqr	Quad-precision floating-point move if register match condition.
FABSq	Quad-precision floating-point absolute value.
FADDq	Quad-precision floating-point addition.
FDIVq	Quad-precision floating-point division.
FdMULq	Double- to quad-precision floating-point multiply.
FMULq	Quad-precision floating-point multiply.
FNEGq	Quad-precision floating-point negation.
FSQRTq	Quad-precision floating-point square root.
FSUBq	Quad-precision floating-point subtraction.

11.3.4 Floating-Point Upper and Lower Dirty Bits in FPRS Register

The FPRS_dirty_upper (du) and FPRS_dirty_lower (dl) bits in the Floating-Point Registers State (FPRS) register are set when an instruction that modifies the corresponding upper or lower half of the floating-point register file is issued. Floating-point register file modifying instructions include floating-point operate, graphics, floating-point loads and block load instructions.

While SPARC V9 allows FPRS.du and FPRS.dl to be set pessimistically, UltraSPARC T2 only sets FPRS.du or FPRS.dl when an instruction that updates the floating-point register file successfully completes. This implies that floating-point instructions that do not update a floating-point register (for example, an FMOVcc that does not meet the condition or a floating-point operate instruction that takes a trap) leave FPRS.du and FPRS.dl unchanged.

11.3.5 Floating-Point Status Register (FSR) (Impdep #13, 19, 22, 23, 24)

UltraSPARC T2 supports precise-traps and implements all three exception fields (tem, cexc, and aexc) conforming to IEEE Standard 754-1985.

UltraSPARC T2 implements the FSR register according to the definition in UltraSPARC Architecture 2007, with the following implementation-specific clarifications:

- UltraSPARC T2 does not contain an FQ, therefore FSR.qne always reads as 0 and an attempt to read the FQ with an RDPR instruction causes an *illegal_instruction* trap.
- UltraSPARC T2 does not detect the unimplemented_FPop, sequence_error, hardware_error or invalid_fp_register floating-point trap types directly in hardware, therefore does not generate a trap when those conditions occur.

11.4 SPARC V9 Memory-Related Operations

11.4.1 Load/Store Alternate Address Space (Impdep #5, 29, 30)

Supported ASI accesses are listed in *Alternate Address Spaces* on page 71.

11.4.2 Read/Write ASR (Impdep #6, 7, 8, 9, 47, 48)

Supported ASRs are listed in Chapter 3, *Registers*.

11.4.3 MMU Implementation (Impdep #41)

UltraSPARC T2 memory management is based on Hardware Tablewalk-managed (or software-managed if Hardware Tablewalk is disabled) instruction and data Translation Lookaside Buffers (TLBs) and in-memory Translation Storage Buffers (TSBs) backed by a Software Translation Table. See Chapter 12, *Memory Management Unit* for more details.

11.4.4 FLUSH and Self-Modifying Code (Impdep #122)

FLUSH is needed to synchronize code and data spaces after code space is modified during program execution. FLUSH is described in *Memory Synchronization: MEMBAR and FLUSH* on page 933. On UltraSPARC T2, the FLUSH effective address is ignored, and as a result, FLUSH cannot cause a *DAE_invalid_ASI* or a *data_access_MMU_miss* trap.

Note | SPARC V9 specifies that the FLUSH instruction has no latency on the issuing virtual processor. In other words, a store to instruction space prior to the FLUSH instruction is visible immediately after the completion of FLUSH. When a flush is performed, UltraSPARC T2 guarantees that earlier code modifications will be visible across the whole system.

11.4.5 PREFETCH{A} (Impdep #103, 117)

For UltraSPARC T2, PREFETCH{A} instructions follow TABLE 11-2 based on the fcn value. All prefetches in UltraSPARC T2 are of the "weak" variety (that is, on an MMU miss, the prefetch is dropped) so the only trap generated by prefetch is *illegal_instruction* (for fcn = 5_{16}–F_{16}).

TABLE 11-2 PREFETCH{A} Variants in UltraSPARC T2

fcn	Prefetch Function	Action
0_{16}	Weak prefetch for several reads	Weak prefetch into Level 2 cache.
1_{16}	Weak prefetch for one read	
2_{16}	Weak prefetch for several writes	
3_{16}	Weak prefetch for one write	

TABLE 11-2 PREFETCH{A} Variants in UltraSPARC T2

fcn	Prefetch Function	Action
4_{16}	Prefetch Page	No operation.
5_{16}–F_{16}	—	*Illegal_instruction* trap.
10_{16}	Invalidate read-once prefetch	Weak prefetch into Level 2 cache.
11_{16}	Prefetch for read to nearest unified cache	Weak prefetch into Level 2 cache.
12_{16}–13_{16}	Strong prefetches	Weak prefetch into Level 2 cache.
14_{16}	Strong prefetch for several reads	Weak prefetch into Level 2 cache.
15_{16}	Strong prefetch for one read	
16_{16}	Strong prefetch for several writes	
17_{16}	Strong prefetch for one write	
18_{16}	Invalidate cache entry	No operation for PREFETCHA. For Prefetch, if executed in user or privileged mode, no operation. If executed while hyperprivileged, invalidate cache line from Level 2 cache (writing back to memory if dirty) leaving Level 2 cache line invalid.
19_{16}–$1F_{16}$	—	No operation

11.4.6 LDD/STD Handling (Impdep #107, 108)

LDD and STD instructions are directly executed in hardware.

Note | LDD/STD are deprecated in SPARC V9. In UltraSPARC T2 it is more efficient to use LDX/STX for accessing 64-bit data. LDD/STD take longer to execute than two 32- or 64-bit loads/stores.

11.4.7 FP mem_address_not_aligned (Impdep #109, 110, 111, 112)

LDDF{A}/STDF{A} cause an *LDDF_/STDF_ mem_address_not_aligned* trap if the effective address is 32-bit aligned but not 64-bit (doubleword) aligned.

LDQF{A}/STQF{A} are not directly executed in hardware; they cause an *illegal_instruction* trap.

11.4.8 Supported Memory Models (Impdep #113, 121)

UltraSPARC T2 supports only the TSO memory model, although certain specific operations such as block loads and stores operate under the RMO memory model. See Chapter 8, Section 8.2. Supported Memory Models.”.

11.4.9 I/O Operations (Impdep #118, 123)

I/O spaces and their accesses are specified in *I/O Address Spaces* on page 70.

11.4.10 Implicit ASI When TL > 0 (Impdep #124)

UltraSPARC T2 matches all UltraSPARC Architecture implementations and makes the implicit ASI for instruction fetching `ASI_NUCLEUS` when TL > 0, while the implicit ASI for loads and stores when TL > 0 is `ASI_NUCLEUS` if PSTATE.cle=0 or `ASI_NUCLEUS_LITTLE` if PSTATE.cle=1.

11.5 Non-SPARC V9 Extensions

11.5.1 Cache Subsystem

UltraSPARC T2 contains one or more levels of cache. The cache subsystem architecture is described in Appendix D, *Caches and Cache Coherency*.

11.5.2 Memory Management Unit

UltraSPARC T2 implements a multi-level memory management scheme. The MMU architecture is described in Chapter 12, *Memory Management Unit*.

11.5.3 Error Handling

UltraSPARC T2 implements a set of programmer-visible error and exception registers. These registers and their usage are described in Chapter 16, *Error Handling*.

11.5.4 Block Memory Operations

UltraSPARC T2 supports 64-byte block memory operations utilizing a block of eight double-precision floating point registers as a temporary buffer. See *Block Load and Store Instructions* on page 33.

11.5.5 Partial Stores

UltraSPARC T2 supports 8-/16-/32-bit partial stores to memory. See *Block Load and Store Instructions* on page 33.

11.5.6 Short Floating-Point Loads and Stores

UltraSPARC T2 supports 8-/16-bit loads and stores to the floating-point registers.

11.5.7 Load Twin Extended Word

UltraSPARC T2 supports 128-bit atomic load operations to a pair of integer registers. See *Load Twin Extended Word* on page 38.

11.5.8 Interrupt Vector Handling

CPUs and I/O devices can interrupt a selected virtual processor by assembling and sending an interrupt packet. This allows hardware interrupts and cross-calls to have the same hardware mechanism and to share a common software interface for processing. Interrupt vectors are described in Chapter 7, *Interrupt Handling*.

11.5.9 Power-Down Support

UltraSPARC T2 supports the ability to power down virtual processors and I/O devices to reduce power requirements during idle periods.

11.5.10 UltraSPARC T2 Instruction Set Extensions (Impdep #106)

The UltraSPARC T2 processor supports VIS 2.0. VIS instructions are designed to enhance graphics functionality and improve the efficiency of memory accesses.

Unimplemented IMPDEP1 and IMPDEP2 opcodes encountered during execution cause an *illegal_instruction* trap.

11.5.11 Performance Instrumentation

UltraSPARC T2 performance instrumentation is described in Chapter 10, *Performance Instrumentation*.

11.5.12 Debug and Diagnostics Support

UltraSPARC T2 support for debug and diagnostics is described in Chapter 19, *Configuration and Diagnostics Support*.

CHAPTER 12

Memory Management Unit

This chapter provides detailed information about the UltraSPARC T2 Memory Management Unit. It describes the internal architecture of the MMU and how to program it.

12.1 Translation Table Entry (TTE)

The Translation Table Entry holds information for a single page mapping. The TTE is broken into two 64-bit words, representing the tag and data of the translation. Just as in a hardware cache, the tag is used to determine whether there is a hit in the TSB.

. TABLE 12-1 shows the sun4v TTE tag format.

TABLE 12-1 TTE Tag Format

Bit	Field	Description
63:61	—	*Reserved*
60:48	context	The 13-bit context identifier associated with the TTE.
47:42	—	*Reserved*
41:0	va	Virtual Address Tag{63:22}. The virtual page number. Bits 21 through 13 are not maintained in the tag, since these bits are used to index the smallest TSB (512 entries). **NOTE:** Hardware only supports a 48-bit VA.

The sun4v TTE data format is shown in TABLE 12-2.

TABLE 12-2 TTE Data Format

Bit	Field	Description
63	v	Valid. If the Valid bit is set, the remaining fields of the TTE are meaningful.
62	nfo	No-fault-only. If this bit is set, loads with ASI_PRIMARY_NO_FAULT{_LITTLE}, ASI_SECONDARY_NO_FAULT{_LITTLE} are translated. Any other DMMU access will trap with a *DAE_nfo_page* trap. For the IMMU, if the nfo bit is set, an *iae_nfo_page* trap will be taken.
61:56	soft2	soft2 and soft are software-defined fields, provided for use by the operating system. Software fields are not implemented in the UltraSPARC T2 TLB. soft and soft2 fields may be written with any value; they read from the TLB as zero, with the exception of soft{61}, which contains the TLB data parity bit.
55:13	ra	The real page[1] number. For UltraSPARC T2, a 40-bit real address range is supported by the hardware tablewalker, and bits {55:40} should always be zero. **NOTE:** UltraSPARC T2 TLBs store physical addresses, not real addresses. Hyperprivileged code is responsible for translation between real and physical addresses. The UltraSPARC T2 TLBs store PA{39:13}.
12	ie	Invert endianess. If this bit is set, accesses to the associated page are processed with inverse endianness from what is specified by the instruction (big-for-little and little-for-big). See Section 12.6 on page 126 for details. For the IMMU, the ie bit in the TTE is written into the ITLB but ignored during ITLB operation. The value of the ie bit written into the ITLB will be read out on an ITLB Data Access read. **Note:** This bit is intended to be set primarily for noncacheable accesses.
11	e	Side effect. If this bit is set, noncacheable memory accesses other than block loads and stores are strongly ordered against other e bit accesses, and noncacheable stores are not merged. This bit should be set for pages that map I/O devices having side effects. Note, however, that the e bit does not prevent normal instruction prefetching. For the IMMU, the e bit in the TTE is written into the ITLB, but ignored during ITLB operation. The value of the e bit written into the ITLB will be read out on an ITLB Data Access read. **NOTE:** The e bit does not force an uncacheable access. It is expected, but not required, that the cp and cv bits will be set to zero when the e bit is set.
10:9	cp, cv	The cacheable-in-physically-indexed-cache and cacheable-in-virtually-indexed-cache (cp, cv) bits determine the placement of data in UltraSPARC T2 caches, according to TABLE 12-3. The MMU does not operate on the cacheable bits, but merely passes them through to the cache subsystem. The cv bit is ignored by UltraSPARC T2, and is not written into the TLBs and returns zero on a Data Access read.

TABLE 12-3 Cacheable Field Encoding (from TSB)

Cacheable (cp:cv)	Meaning of TTE When Placed in: iTLB (I-cache PA-Indexed)	Meaning of TTE When Placed in: dTLB (D-cache PA-Indexed)
0x	Cacheable L2 cache only	Cacheable L2 cache only
1x	Cacheable L2 cache, I-cache	Cacheable L2 cache, D-cache

TABLE 12-2 TTE Data Format *(Continued)*

Bit	Field	Description
8	p	Privileged. If the p bit is set, only privileged software can access the page mapped by the TTE. If the p bit is set and an access to the page is attempted when PSTATE.priv = 0, the MMU will signal an *IAE_privilege_violation* or *DAE_privilege_violation* trap.
7	ep	Executable. If the ep bit is set, the page mapped by this TTE has execute permission granted. Otherwise, execute permission is not granted and the hardware table-walker will not load the ITLB with a TTE with ep = 0. For the IMMU and DMMU, the ep bit in the TTE is not written into the TLB, and returns zero on a Data Access read.
6	w	Writable. If the w bit is set, the page mapped by this TTE has write permission granted. Otherwise, write permission is not granted and the MMU will cause a *fast_data_access_protection* trap if a write is attempted. For the IMMU, the w bit in the TTE is written into the ITLB, but ignored during ITLB operation. The value of the w bit written into the ITLB will be read out on an ITLB Data Access read.
5:4	soft	(see soft2, above)
3:0	size	The page size of this entry, encoded as shown in TABLE 12-4.

TABLE 12-4 Size Field Encoding (from TTE)

Size{2:0}	Page Size
0000	8 KB
0001	64 KB
0010	Reserved
0011	4 MB
0100	Reserved
0101	256 MB
0110-1111	Reserved

1. sun4v supports translation from virtual addresses (VA) to real addresses (RA) to physical addresses (PA). Privileged code manages the VA-to-RA translations, while hyperprivileged code manages the RA-to-PA translations. The TLBs contain VA-to-PA translations or RA-to-PA translations (the latter are distinguished from the former by a Real bit in the TLB).

12.2 Translation Storage Buffer (TSB)

A TSB is an array of TTEs managed entirely by software. It serves as a cache of the Software Translation table, used to quickly reload the TLB in the event of a TLB miss. The discussion in this section assumes the use of the hardware support for TSB access described in Section 12.3.1, although the operating system is not required to make use of this support hardware.

Inclusion of the TLB entries in a TSB is not required; that is, translation information may exist in the TLB that is not present in the TSB.

A TSB is arranged as a direct-mapped cache of TTEs. The UltraSPARC T2 MMU provides hardware tablewalk support and precomputed pointers into the TSB(s) for both zero and nonzero contexts for four different TSB, as specified in the following registers:

- ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_0
- ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_1
- ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_2
- ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_3
- ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_0
- ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_1
- ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_2
- ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_3

In each case, the *n* least significant bits of the respective virtual page number are used as the offset from the TSB base address, with *n* equal to log base 2 of the number of TTEs in the TSB.

Hardware TSB indexing support is provided for TTEs in the following registers:

- ASI_MMU_ITSB_PTR_0
- ASI_MMU_ITSB_PTR_1
- ASI_MMU_ITSB_PTR_2
- ASI_MMU_ITSB_PTR_3
- ASI_MMU_DTSB_PTR_0, ASI_MMU_DTSB_PTR_1
- ASI_MMU_DTSB_PTR_2
- ASI_MMU_DTSB_PTR_3

While the hardware tablewalk uses the TSB configuration generated by these pointers, the hardware tablewalk can be disabled and a full software implementation for TLB miss handling can be used. Under a full software implementation, simple modifications to the index pointers provided by the hardware allow formation of an M-way set-associative TSB, multiple TSBs per page size, multiple page sizes per TSB, and multiple TSBs per process.

The TSB exists as a normal data structure in memory and therefore may be cached. Indeed, the speed of the TLB miss handler relies on the TSB accesses hitting the level-2 cache at a substantial rate. This policy may result in some conflicts with normal instruction and data accesses, but the dynamic sharing of the level-2 cache resource should provide a better overall solution than that provided by a fixed partitioning.

FIGURE 12-1 shows the TSB organization. The constant *N* is determined by the size field in the TSB register; it may range from 512 entries to 16 M entries.

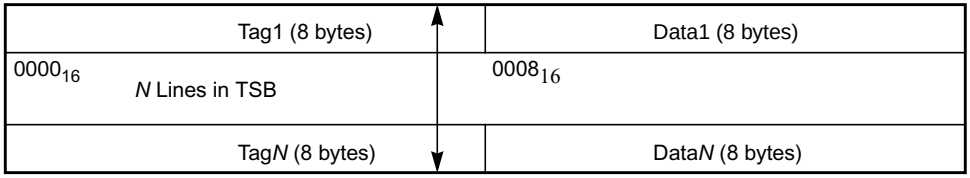


FIGURE 12-1 TSB Organization

12.3 Hardware Support for Hypervisor

To support hypervisor, a number of additions to the MMU are included.

First, a 3-bit pid (partition ID) field is included in each TLB entry to allow multiple guest OSs to share the MMU. This field is loaded with the value of the Partition Identifier register when a TLB entry is loaded. In addition, the PID entry of a TLB is compared against the Partition Identifier register to determine if a TLB hit occurs.

Second, the MMU is designed to support both virtual-to-physical and real-to-physical translations, using a single r (real translation) bit included in the TLB entry. This field is loaded with bit 10 from the VA used by the store to the I-/D-TLB Data In register or the I-/D-TLB Data Access register. The real bit distinguishes between VA → PA translations (r = 0) and RA → PA translations (r = 1). If the real bit is 1, the context ID is ignored when determining a TLB hit. TLB misses on real to physical translations generate a *data_real_translation_miss* or *inst_real_translation_miss* trap instead of the *fast_data_access_MMU_miss* and *fast_instruction_access_MMU_miss* traps respectively.

Finally, the translation operation performed depends on the state of HPSTATE.hpriv, PSTATE.priv, the MMU enables, and PSTATE.red (for IMMU), as described in *Translation* on page 128.

When the MMU is bypassed, TABLE 12-5 specifies the default physical page attribute bits. When bypassed, all LDXA and STXA operations to internal registers are correctly performed, and traps based on the page attribute bits are signaled just as if the MMU were not bypassed.

TABLE 12-5 Default Physical Page Attribute Bits

Physical Address{39}	Physical Page Attribute Bits							
	cp	ie	cv	e	p	ep	w	nfo
0	1	0	0	0	0	1	1	0
1	0	0	0	1	0	1	1	0

12.3.1 Hardware Support for TSB Access

The MMU hardware provides services to allow the TLB miss handler to efficiently reload a missing TLB entry. These services include:

- Hardware reload of missing TTE entry (hardware tablewalk).
- Formation of TSB Pointers based on the missing virtual address.
- Formation of the TTE Tag Target used for the TSB tag comparison.
- Efficient atomic write of a TLB entry with a single store ASI operation.

12.3.1.1 Hardware Tablewalk

Hardware Tablewalk is a hardware state machine that services reload requests from the TLBs. It accesses the TSBs to find TTEs that match the VA and one of the contexts of the request. Hardware Tablewalk can access up to four separate TSBs for each request.

Note | If any of a strand's TSB Config Registers has the Enable bit sit, hardware tablewalk is considered to be enabled for the strand.

Hardware Tablewalk also provides a real page number (RPN) to physical page number (PPN) translation mechanism. The supervisor controls the TTE, but the supervisor cannot access or control physical memory, so its TTEs contain RPNs, not PPNs The hypervisor programs the RPN-to-PPN translation within Hardware Tablewalk to permit Hardware Tablewalk to load supervisor-controlled TTEs into the TLBs that can translate VAs into PAs.

Hardware Tablewalk does not translate real requests. In the event that a Real Address misses the TLB, the TLU signals a *inst_real_translation_miss* or *data_real_translation_miss* trap, and software loads the TLB as described in *Software TLB Reload* on page 114.

Hardware Tablewalk is stranded and pipelined; up to four TSB accesses for each of the eight strands can be in the pending at one time. The basic dataflow is pipelined, so that a single instance of the dataflow supports all eight strands.

A typical TLB miss and refill sequence when hardware tablewalk is enabled is as follows:

1. Hardware Tablewalk uses the TSB Configuration registers and the VA of the access to calculate the address of the TTE to examine. The TSB Configuration register provides the base address of the TSB as well as the number of TTEs in the TSB and the size of the pages translated by the TTEs.[1] Hardware Tablewalk uses a Nonzero Context TSB Configuration register if the context of the request is nonzero; otherwise, it uses a Zero Context TSB Configuration register. The context of the request is assumed to be the content of Context register 0 (in the event of a TLB miss on a Primary or Secondary Context access). Hardware Tablewalk uses the page size from the TSB Configuration register to calculate the presumed VPN for the given VA.[2] The VPN is generated with VA{63:48} sign-extended from VA{47} to allow TTE entries pointing into the VA hole to mismatch in the hardware tablewalk VPN comparison. Hardware Tablewalk then uses the number of TTE entries and the presumed VPN to generate an index into the TSB. This index is concatenated with the upper bits of the base address to generate the TTE address, using the formula specified in *MMU I-/D-TSB Pointer Registers* on page 146.

2. Hardware Tablewalk forwards a quadword load request for the TTE address to the L2 cache. At some later time, the L2 returns the TTE to Hardware Tablewalk.

3. Hardware Tablewalk compares the VPN (masked using the page size of the TTE) and context of the request and the page size from the configuration register to that from the TTE, and also examines the v bit, and for an ITLB miss, the ep bit of the TTE. If the v bit is set, reserved fields in the TTE Tag ({63:61} and {47:42}) are zero, the page size of the TTE is a supported page size that is not smaller than the page size of the configuration register, and the VPN and context match (and for an ITLB miss, the TTE ep bit is 1), Hardware Tablewalk forwards the TTE to the TLB with the RPN translated into a PPN (see *Real Page Number To Physical Page Number Translation* below). For an ITLB miss, if the v bit is set, reserved fields in the TTE Tag are zero, the page size of the TTE is supported and not smaller than the page size of the configuration register, and the VPN and context match, but the ep bit of the TTE is 0, an *IAE_unauth_access* trap is generated. If the v bit is clear, reserved fields in the TTE Tag are not all zero, the page size of the TTE is unsupported or smaller than the page size of the configuration, or the VPN or context do not match, Hardware Tablewalk waits for the rest of the enabled TSBs to return TTEs; Hardware Tablewalk supports four TSBs per strand for zero contexts and four for nonzero contexts. In some configurations, Hardware Tablewalk ignores the context match; see *Multiple Contexts* below.

1. Hardware tablewalk will only be able to refill the TLB when the desired TTE in the TSB is the same size or larger than that specified in the TSB Configuration register.

2. If pages of size larger than that specified in the TSB Configuration register are also cached in the TSB, this implies that the TTE entry for the larger pages must be replicated in the TSB (e.g., a 64-Kbyte page in a TSB configured for 8-Kbyte pages must occupy eight consecutive TSB entries).

4. If none of the TTE entries from the four TSBs meet the v bit, reserved TTE Tag fields, page size, and matching VPN and context requirements, hardware generates an *instruction_access_MMU_miss* or *data_access_MMU_miss* trap.

Multiple Contexts. Multiple primary and secondary contexts permit different processes to share TTEs within the TLBs. The use_context_0 and use_context_1 bits in the TSB Configuration register disable the context match for Hardware Tablewalk. Hardware Tablewalk ignores the contexts in the TSB TTEs if either of these bits is active for requests with nonzero contexts. If either bit is 1 and the TTE v bit is set, reserved fields in the TTE Tag are zero, the page size is supported by UltraSPARC T2 and not smaller than the page size in the configuration register, and the VPN matches, Hardware Tablewalk signals the TLB to write either context 0 or context 1 (depending on which bit is set) as the context of the TTE when it is loaded (instead of the context in the TTE itself). Hardware Tablewalk ignores these bits for requests with a zero (nucleus) context value.

Real Page Number To Physical Page Number Translation. When Hardware Tablewalk fetches a TTE from a TSB, it can treat the ra field as either an RA or a PA under control of the ra_not_pa field of the TSB config register. If the ra_not_pa bit is set, the hardware tablewalker will translate the Real Page Number in the TTE into a Physical Page Number. The TLBs store this physical page number. The TLBs use this PPN to translate VAs into PAs. The hypervisor controls the RPN to PPN translation mechanism.

The RPN-to-PPN translation mechanism provides both range checking as well as mapping of address ranges from one location to another. The first check is that the RPN does not contain a non-zero bits 55:40. If RPN{55:40} is nonzero, then an *instruction_invalid_TSB_entry* or *data_invalid_TSB_entry* trap is generated to the strand that initiated the Hardware Tablewalk. Otherwise, the translation mechanism uses the RPN and page size in the TTE and calculates the starting and ending addresses for the specified real page. It then checks that these addresses lie in one of four ranges specified by the Real Range registers. If the real page lies completely inside one of the ranges (and the range is enabled), then the translation mechanism adds the RPN in the TTE to the corresponding field in the Physical Offset register to create the Physical Page Number.[1] If the real page does not lie completely within either range, then an *instruction_invalid_TSB_entry* or *data_invalid_TSB_entry* trap is generated to strand that initiated the Hardware Tablewalk. Each strand has four dedicated ranges with corresponding physical offsets. The RPN-to-PPN translation does not depend on the context value being zero or nonzero.

Note | When the TSB Config register has ra_not_pa = 0, no range checking is provided for PPNs.

1. Strictly speaking, this is not a real page number but a real address, as the bits below the page size boundary in the RPN may not be zeroed. They are zeroed when written into the TLB so the value that is written into the TLB is a true RPN.

The following is pseudocode for the hardware tablewalk sequence (translation miss address is in VA, miss context in `ctxt`, instruction miss is in `inst_translate`, context type selection is in primary):

```
tsb_config = (ctxt) ? TSB_NONZERO_CONFIG[i] : TSB_ZERO_CONFIG[i];
for (i = 0; i < 4, i++) {
  vpn = generate_vpn(VA, tsb_config.page_size);
  ignore_context = (ctxt == 0) || tsb_config.use_context_0 ||
    tsb_config.use_context_1;
  tte_ptr = generate_tte_ptr(VA, tsb_config);
  tte = *tte_ptr;
  tte_vpn = generate_tte_vpn(tte.va, tsb_config.page_size)
  if ((tte.v == 1) && (tte.rsvd0 == 0) && (tte.rsvd1 == 0) &&
       ((tte.size < 2) || (tte.size == 3) || (tte.size == 5)) &&
       (tte.size >= tsb_config.page_size) &&
       (ignore_context || (tte.context == ctxt)) &&
       (tte_vpn == vpn)) {
    if (inst_translate && (tte.ep == 0) raise IAE_UNAUTH_ACCESS;
    break;
  }
}
if (i == 4) {
  // no matching TTE found
  raise (inst_translate) ?
    INST_ACCESS_MMU_MISS : DATA_ACCESS_MMU_MISS;
}
if (ctxt == 0) {
  tte.context = 0;
} else if (tsb_config.use_context_0) {
  tte.context = (primary_context) ? ASI_PRIMARY_CONTEXT_0 :
    ASI_SECONDARY_CONTEXT_0;
} else if (tsb_config.use_context_1) {
  tte.context = (primary_context) ? ASI_PRIMARY_CONTEXT_1 :
    ASI_SECONDARY_CONTEXT_1;
}
if (tsb_config.RA_not_PA == 0) {
  load_tlb(tte);
} else {
  if (tte.ra && RA_55_40_MASK) {
    raise (inst_translate) ?
      INST_INVALID_TSB_ENTRY : DATA_INVALID_TSB_ENTRY;
  }
  mask = (1 << (tte.size*3)) - 1;
  rpn_low = tte.ra & ~mask;
  rpn_high = tte.ra | mask;
  for (i = 0; i < 4; i++ ) {
    if ((rpn_low >= MMU_REAL_RANGE[i].rpn_low) &&
        (rpn_high <= MMU_REAL_RANGE[i].rpn_high)) {
      tte.ra += MMU_PHYSICAL_OFFSET[i].ppn;
```

```
      load_tlb(tte);
      break;
    }
  }
  if (i == 4) {
    // ra does not lie entirely inside any real range
    raise (inst_translate) ?
      INST_INVALID_TSB_ENTRY : DATA_INVALID_TSB_ENTRY;
  }
}
```

12.3.1.2 Software TLB Reload

TLB misses can be either for virtual-to-physical translations or for real-to-physical translations.

Virtual to Physical Address Translation. For a virtual address, a typical TLB miss and refill sequence when hardware tablewalk is disabled is as follows:

1. A TLB miss causes either a *fast_instruction_access_MMU_miss* or a *fast_data_access_MMU_miss* exception.
2. The appropriate TLB miss handler loads the TSB Pointers and the TTE Tag Target with loads from the MMU alternate space.
3. Using this information, the TLB miss handler checks to see if the desired TTE exists in the TSB. If so, the TTE data is stored into the TLB Data In register (with the Real bit in the virtual address clear) to initiate an atomic write of the TLB entry chosen by the replacement algorithm.
4. If the TTE does not exist in the TSB, the TLB miss handler jumps to a more sophisticated (and slower) TSB miss handler.

The virtual address used in the formation of the pointer addresses comes from the Tag Access register (described in *I-/D-TLB Tag Access Registers* on page 140), which holds the virtual address oF the load or store responsible for the MMU exception. (Note that there are no separate physical registers in UltraSPARC T2 hardware for the Pointer registers, but rather they are implemented through a dynamic re-ordering of the data stored in the Tag Access and the TSB registers.)

Pointers are provided by hardware in the TSB Pointer registers for all four TSBs. These pointers give the physical addresses where the TTEs for that VA and context combination would be stored if it is present in the TSB.

The TSB Tag Target register (described in *I-/D-TSB Tag Target Registers* on page 138) is formed by aligning the missing access VA (from the Tag Access register) and the current context to positions found in the description of the TTE tag. This allows an XOR instruction for TSB hit detection.

Real-to-Physical Address Translation. For a real address, a typical TLB miss and refill sequence when hardware tablewalk is disabled is as follows:

1. A TLB miss causes either a *instruction_real_translation_miss* or a *data_real_translation_miss* exception.
2. The appropriate real miss handler determines whether and how to create a real to physical translation.
3. If a real-to-physical translation is created it is inserted into the TLB with the r bit set and the instruction is retried.

The real address used in the software formation of the pointer addresses comes from the Tag Access register (described in Section 12.10.5), which holds the real address and context of the load or store responsible for the MMU exception.

No pointers are provided by hardware for real to physical translations.

The TSB Tag Target register (described in Section 12.10.3) is formed by aligning the missing access RA (from the Tag Access register) and the current context to positions found in the description of the TTE tag. This allows an XOR instruction for TSB hit detection.

12.3.2 Real-to-Physical Address Mapping and Speculative Instruction Fetch

UltraSPARC T2 speculatively fetches instructions. Under certain conditions, this can cause the memory controller to receive an unsupported physical address. Consider the following instruction sequence which exits hyperprivileged mode and returns to user or privileged mode (executed with HPSTATE.hpriv initially set to 1):

```
jmpl %g3 + %g0, %g0
wrhpr   %g0, %g0, %hpstate
```

This will cause the IMMU to go from bypass (during which VA{39:0} is passed directly to PA{39:0}) into either RA → PA or VA → PA translation. However, since the fetch of the target of the jmpl is fetched speculatively, the memory controller may see VA{39:0} of the target of the jmpl as a physical address. This address may not be supported, in which case a disrupting *software_recoverable_error* trap could result, even though no real error has occurred.

To avoid this disrupting trap, hypervisor should avoid changing translation in the delay slot of delayed control transfer instructions. For example, the sequence above could be replaced with the following code:

```
mov    %tl, %g5
add    %g5, 1, %g5
mov    %g5, %tl
```

```
mov    %g3, %tnpc
mov    0, %htstate
done
```

Although the example refers to changes in HPSTATE, any instruction that can potentially change translation should avoid being placed in the delay slot of delayed control transfer instructions. These include writes to PSTATE, I-/D-TLB Data-In/ Data-Access registers, I-/D-MMU Demap registers, and the ASI_LSU_CONTROL_REG register.

12.4 MMU-Related Faults and Traps

TABLE 12-6 lists the traps recorded by the MMU.

TABLE 12-6 MMU Traps

		Register Update		
Trap Name	**Trap Cause**	**I-Tag Access**	**D-SFAR**	**D-Tag Access**
fast_instruction_access_MMU_miss	iTLB miss with hardware tablewalk disabled	x		
instruction_access_MMU_miss	iTLB miss with hardware tablewalk enabled	x		
instruction_real_translation_miss	iTLB miss	x		
instruction_invalid_TSB_entry	RA out of range on iTLB hardware tablewalk reload	x		
IAE_privilege_violation	Privilege violation	x		
IAE_unauth_access	Hardware tablewalk attempts to load IMMU with page with ep clear	x		
IAE_NFO_page	Instruction fetch from nfo page	x		
instruction_address_range	Fetch address out of range	x		
instruction_real_range	Fetch address out of range	x		
fast_data_access_MMU_miss	dTLB miss with hardware tablewalk disabled			x
data_access_MMU_miss	dTLB miss with hardware tablewalk enabled			x
data_invalid_TSB_entry	RA out of range on dTLB hardware tablewalk reload			x
data_real_translation_miss	dTLB miss			x
DAE_invalid_asi	Invalid ASI, size, etc.		x	

TABLE 12-6 MMU Traps *(Continued)*

Trap Name	Trap Cause	Register Update		
		I-Tag Access	D-SFAR	D-Tag Access
DAE_privilege_violation	Privilege violation		x	x
DAE_nc_page	Atomic to noncacheable		x	x
DAE_nfo_page	Access to nfo page by non no-faulting		x	x
DAE_side_effect_page	Access to page with e =1 by nonfaulting load		x	x
mem_address_range	va out of valid range		x	
mem_real_range	ra out of valid range		x	
fast_data_access_protection	Protection violation		x	x
privileged_action	Use of privileged ASI			
pa_watchpoint, va_watchpoint	Data watchpoint hit		x	
instruction_va_watchpoint	Instruction watchpoint hit			
**_mem_address_not_aligned*	Misaligned memory op		x	
unsupported_page_size	TLB or TSB register loaded with illegal page size			

Note The *fast_data_access_protection* trap is generated instead of the *data_access_protection* trap.

12.4.1 *fast_instruction_access_MMU_miss* Trap

This trap occurs when Hardware Tablewalk is disabled and the I-MMU is unable to find a translation for an instruction access that is executing using a virtual address.

Real-to-physical translations (for example, when LSU_CONTROL.im = 0) that miss in the MMU generate an *instruction_real_translation_miss* trap instead.

12.4.2 *instruction_access_MMU_miss* Trap

This trap occurs when Hardware Tablewalk is enabled and the I-MMU is unable to find a translation for an instruction access that is executing using a virtual address.

Implementation Note This trap is taken when the appropriate TTE is not present in the iTLB with the r bit cleared and Hardware Tablewalk is unable to find the appropriate TTE in any of up to four TSBs.

Note | Real-to-physical translations (for example, when LSU_CONTROL.im=0) that miss in the MMU generate an *instruction_real_translation_miss* trap instead.

12.4.3 *instruction_real_translation_miss* Trap

This trap occurs when the I-MMU is unable to find a translation for an instruction access that is executing using a real address.

Note | Hardware Tablewalk does not load real to physical translations, and therefore *instruction_real_translation_miss* is taken regardless of whether Hardware Tablewalk is enabled or disabled.

Programming Note | The MMU Real Range and Physical Offset registers are used in the real-to-physical translation portion of a Hardware Tablewalk's virtual-to-physical translation and have no effect on whether an *instruction_real_translation_miss* trap is taken.

12.4.4 *instruction_invalid_TSB_entry* Trap

This trap occurs when the Hardware Tablewalk is loading the I-MMU with RA-to-PA translation enabled and is unable to complete the RA-to-PA portion of the translation due to the real address not lying completely in the range specified by any of the valid MMU Range registers. It also occurs on real to physical translations where bits 55:40 of the real address are non-zero.

12.4.5 *IAE_privilege_violation* Trap

The I-MMU detects a privilege violation for an instruction fetch; that is, an attempted access to a privileged page when PSTATE.priv = 0.

12.4.6 *IAE_unauth_access* Trap

The I-MMU detects an access to a page marked with the ep (execute privilege) bit clear during hardware tablewalk. There is no ep bit in the I-MMU, so on software loads of the I-MMU there is no hardware check of the ep bit.

12.4.7 *IAE_nfo_page* Trap

The I-MMU detects an access to a page marked with the nfo (no-fault-only) bit.

Implementation Note | The nfo bit is only checked on I-MMU translations. It is not checked on hardware tablewalk.

12.4.8 *instruction_address_range* Trap

The *instruction_address_range* occurs when the virtual address out of range and PSTATE.am = 0 (see *48-bit Virtual and Real Address Spaces* on page 68). It also occurs whenever UltraSPARC T2 is fetching from the address range 0000 7FFF FFFF $FFE0_{16}$ to FFFF 7FFF FFFF $FFFF_{16}$ inclusive, the IMMU is in VA → PA translation mode, and PSTATE.am = 0. The *instruction_address_range* exception also occurs whenever a branch target or the target of a DONE or RETRY instruction is in the range 0000 800 0000 000_{16} to FFFF 7FFF FFFF $FFFF_{16}$ inclusive, the IMMU is in VA → PA translation mode, and PSTATE.am = 0.

12.4.9 *instruction_real_range* Trap

The *instruction_real_range* trap occurs when the Real address out of range (see *48-bit Virtual and Real Address Spaces* on page 68). It also occurs whenever UltraSPARC T2 is fetching from the address range 0000 7FFF FFFF $FFE0_{16}$ to FFFF 7FFF FFFF $FFFF_{16}$ inclusive, the IMMU is in RA → PA translation mode, and PSTATE.am = 0. The *instruction_real_range* exception also occurs whenever a branch target or the target of a DONE or RETRY instruction is in the range 0000 8000 0000 0000_{16} to FFFF 7FFF FFFF $FFFF_{16}$ inclusive, the IMMU is in RA → PA translation mode, and PSTATE.am = 0.

12.4.10 *fast_data_access_MMU_miss* Trap

This trap occurs when Hardware Tablewalk is disabled and the MMU is unable to find a translation for a data access that is using a virtual-to-physical translation.

Note | Real-to-physical translations (for example, through `ASI_*REAL*`) that miss in the MMU generate a *data_real_translation_miss* trap instead.

12.4.11 *data_access_MMU_miss* Trap

This trap occurs when Hardware Tablewalk is enabled and the MMU is unable to find a translation for a data access that is using a virtual-to-physical translation.

Note | Real-to-physical translations (for example, through `ASI_*REAL*`) that miss in the MMU generate a *data_real_translation_miss* trap instead.

12.4.12 *data_invalid_TSB_entry* Trap

This trap occurs when Hardware Tablewalk is loading the D-MMU with RA-to-PA translation enabled and is unable to complete the RA-to-PA portion of the translation due to the real address not lying completely in the range specified by any of the valid MMU Range registers. It also occurs on real-to-physical translations where bits 55:40 of the real address are non-zero.

12.4.13 *data_real_translation_miss* Trap

This trap occurs when the MMU is unable to find a translation for a data access that is using a real-to-physical translation.

Note | Hardware Tablewalk does not load real-to-physical translations, and therefore *data_real_translation_miss* is taken regardless of whether hardware tablewalk is enabled or disabled.

Programming Note | The MMU Real Range and Physical Offset registers are used in the real-to-physical translation portion of a Hardware Tablewalk's virtual-to-physical translation and have no effect on whether a *data_real_translation_miss* trap is taken.

12.4.14 *DAE_privilege_violation* Trap

The D-MMU detects a privilege violation for a data access; that is, an attempted access to a privileged page when PSTATE.priv = 0.

12.4.15 *DAE_side_effect_page* Trap

A speculative (nonfaulting) load instruction issued to a page marked with the side-effect (e) bit = 1.

12.4.16 *DAE_nc_page* Trap

An atomic instruction (including 128-bit atomic load) issued to a memory address marked uncacheable in a physical cache; that is, with cp = 0 or with PA{39} = 1.

Implementation Note For UltraSPARC T2, cp only controls cacheability in the primary cache, not the shared secondary, and thus the hardware supports the ability to complete an atomic operation for pages with the cp bit = 0 as long as the secondary cache is enabled. However, to keep UltraSPARC T2 compliant with the UltraSPARC Architecture 2006 specification, the *DAE_nc_page* trap is generated when an atomic is issued to a memory address marked with cp = 0.

12.4.17 *DAE_invalid_asi* Trap

An invalid LDA/STA ASI value, invalid virtual address, read to write-only register, or write to read-only register, but not for an attempted user access to a restricted ASI (see the *privileged_action* trap described below).

12.4.18 *DAE_nfo_page* Trap

An access with an ASI other than ASI_{PRIMARY,SECONDARY}_NO_FAULT{_LITTLE} to a page marked with the nfo (no-fault-only) bit.

12.4.19 *mem_address_range* Trap

Virtual address out of range and PSTATE.am = 0 for data access, JMPL/RETURN, or branch/CALL. See *48-bit Virtual and Real Address Spaces* on page 68.

12.4.20 *mem_real_range* Trap

Real address out of range for data access, JMPL/RETURN, or branch/CALL. See *48-bit Virtual and Real Address Spaces* on page 68.

12.4.21 *fast_data_access_protection* Trap

This trap occurs when the MMU detects a protection violation for a data access. A protection violation is defined to be an attempted store to a page without write permission.

Note Protection violations are checked for both virtual-to-physical translations and real-to-physical translations.

12.4.22 *privileged_action* Trap

This trap occurs when an access is attempted using a *restricted* ASI while in non-privileged mode (PSTATE.priv = 0).

12.4.23 *instruction_VA_watchpoint* Trap

This trap occurs when instruction virtual watchpoints are enabled and the I-MMU detects a instruction execution at the virtual address specified by the VA Instruction Watchpoint register. See *Watchpoint Support* on page 409.

Programming Note | *instruction_VA_watchpoint* is never generated when HPSTATE.red = 1 or HPSTATE.hpriv = 1. In addition, *instruction_VA_watchpoint* traps are only generated when a virtual-to-physical translation is performed. Real accesses do not generate *instruction_VA_watchpoint* traps.

12.4.24 *VA_watchpoint* Trap

This trap occurs when virtual watchpoints are enabled and the D-MMU detects a load or store to the virtual address specified by the VA Data Watchpoint register. See *Watchpoint Support* on page 409.

Programming Note | *VA_watchpoint* is never generated when HPSTATE.hpriv = 1. In addition, *VA_watchpoint* traps are only generated when a virtual-to-physical translation is performed. Real accesses (for example, through ASI_*REAL*) do not generate *VA_watchpoint* traps.

12.4.25 *PA_watchpoint* Trap

This trap occurs when physical watchpoints are enabled and the D-MMU detects a load or store to the physical address specified by the PA Data Watchpoint register. See *Watchpoint Support* on page 409.

12.4.26 **_mem_address_not_aligned* Traps

The *lddf_mem_address_not_aligned*, *stdf_mem_address_not_aligned*, and *mem_address_not_aligned* traps occur when a load, store, atomic, or JMPL/RETURN instruction with a misaligned address is executed. The LSU signals this trap, but the D-MMU records the fault information in the DSFAR.

12.4.27 *Unsupported_page_size* Trap

This trap occurs when the IMMU or DMMU is loaded with an illegal page size, or a TSB register is programmed with an illegal page size.

12.5 MMU Operation Summary

TABLE 12-9 summarizes the behavior of the D-MMU for noninternal ASIs using tabulated abbreviations. TABLE 12-10 summarizes the behavior of the I-MMU. In each case, and for all conditions, the behavior of the MMU is given by one of the abbreviations in TABLE 12-7. TABLE 12-8 lists abbreviations for ASI types.

TABLE 12-7 Abbreviations for MMU Behavior

Abbreviation	Meaning
ok	Normal translation
dmiss	*fast_data_access_MMU_miss* or *data_access_MMU_miss* trap
dasi	*DAE_invalid_asi* trap
dreal	*data_real_translation_miss* trap
dpriv	*DAE_privilege_violation* trap
dse	*DAE_side_effect_page* trap
dprot	*fast_data_access_protection* trap
imiss	*fast_instruction_access_MMU_miss* or *instruction_access_MMU_miss* trap
ireal	*instruction_real_translation_miss* trap
iexc	*IAE_privilege_violation* trap

TABLE 12-8 Abbreviations for ASI Types

Abbreviation	Meaning
NUC	`ASI_NUCLEUS*`
PRIM	Any ASI with PRIMARY translation, except `*NO_FAULT`
SEC	Any ASI with SECONDARY translation, except `*NO_FAULT`
PRIM_NF	`ASI_PRIMARY_NO_FAULT*`
SEC_NF	`ASI_SECONDARY_NO_FAULT*`
U_PRIM	`ASI_*_AS_IF_USER_PRIMARY*`
U_SEC	`ASI_*_AS_IF_USER_SECONDARY*`
U_PRIV	`ASI_*_AS_IF_PRIV_*`
REAL	`ASI_*REAL*`

Note | The *_LITTLE versions of the ASIs behave the same as the big-endian versions with regard to the MMU table of operations.

Other abbreviations include “w” for the writable bit, “e” for the side-effect bit, and “p” for the privileged bit.

TABLE 12-9 and TABLE 12-10 do not cover the following cases:

- Invalid ASIs, ASIs that have no meaning for the opcodes listed, or nonexistent ASIs; for example, ASI_PRIMARY_NO_FAULT for a store or atomic; also, access to UltraSPARC T2 internal registers other than LDXA, LDFA, STDFA or STXA; the MMU signals a *DAE_invalid_asi* trap for this case.
- Attempted access using a restricted ASI in nonprivileged mode; the MMU signals a *privileged_action* trap for this case. Attempted use of a hyperprivileged ASI in privileged mode; the MMU also signals *privileged_action* trap for this case.
- An atomic instruction (including 128-bit atomic load) issued to a memory address marked uncacheable in a physical cache (that is, with cp = 0 or pa{39} = 1); the MMU signals a *DAE_nc_page* trap for this case.
- A data access with an ASI other than ASI_{PRIMARY,SECONDARY}_NO_FAULT{_LITTLE} or an instruction access to a page marked with the nfo (no-fault-only) bit; the MMU signals a *DAE_nfo_page* or *IAE_nfo_page* trap for this case.
- An instruction fetch to a memory address marked non-executable (ep = 0). This is checked when Hardware Tablewalk attempts to load the I-MMU, and an *IAE_unauth_access* trap is taken instead.
- Real address out of range; the MMU signals an *instruction_real_range* or *mem_real_range* trap for this case.
- Virtual address out of range and PSTATE.am is not set; the MMU signals an *instruction_address_range* or *mem_address_range* trap for this case.

TABLE 12-9 D-MMU Operations for Normal ASIs

Condition				Behavior				
Opcode	priv Mode	ASI	w	TLB Miss	e = 0 p = 0	e = 0 p = 1	e = 1 p = 0	e = 1 p = 1
Load	non-privileged	PRIM, SEC	—	dmiss	ok	dpriv	ok	dpriv
		PRIM_NF, SEC_NF	—	dmiss	ok	dpriv	dse	dpriv
	privileged	PRIM, SEC, NUC	—	dmiss	ok			
		PRIM_NF, SEC_NF	—	dmiss	ok		dse	
		U_PRIM, U_SEC	—	dmiss	ok	dpriv	ok	dpriv
		REAL	—	dreal	ok			
	hyper-privileged	PRIM, SEC, NUC[1]	1	—	ok	—	ok	—
		PRIM_NF, SEC_NF[1]	1	—	ok	—	dse	—
		U_PRIM, U_SEC	—	dmiss	ok	dpriv	ok	dpriv
		U_PRIV	—	dmiss	ok			
		REAL	—	dreal	ok			
FLUSH	non-privileged		—	ok				
	privileged		—	ok				
	hyper-privileged		—	ok				
Store or Atomic	non-privileged	PRIM, SEC	0	dmiss	dprot	dpriv	dprot	dpriv
			1	dmiss	ok	dpriv	ok	dpriv
	privileged	PRIM, SEC, NUC	0	dmiss	dprot			
			1	dmiss	ok			
		U_PRIM, U_SEC	0	dmiss	dprot	dpriv	dprot	dpriv
			1	dmiss	ok	dpriv	ok	dpriv
		REAL	0	dreal	dprot			
			1	dreal	ok			
	hyper-privileged	PRIM, SEC, NUC[1]	1	—	ok	—	ok	—
		U_PRIM, U_SEC	0	dmiss	dprot	dpriv	dprot	dpriv
			1	dmiss	ok	dpriv	ok	dpriv
		U_PRIV	0	dmiss	dprot	ok	dprot	ok
			1	dmiss	ok			
		REAL	0	dreal	dprot			
			1	dreal	ok			

1. When hyperprivileged, these context values use the default physical page attributes from TABLE 12-5, which specify that w = 1, e = PA{39}, and p = 0.

When LSU_CONTROL_REG.dm = 0, the table above applies, but *dmiss* entries in the TLB Miss column change to *dreal*.

TABLE 12-10 I-MMU Operations

Condition	Behavior		
priv Mode	TLB Miss	P = 0	P =1
nonprivileged	imiss	ok	iexc
privileged	imiss	ok	
hyperprivileged	—	ok	

When LSU_CONTROL_REG.im = 0, the table above applies, but *imiss* entries in the TLB Miss column change to *ireal*. When HPSTATE.red = 1, the ITLB is bypassed and has the same behavior as the hyperprivileged mode row in the table.

See *Alternate Address Spaces* on page 71 for a summary of the UltraSPARC T2 ASI map.

12.6 ASI Value, Context, and Endianness Selection for Translation

The MMU uses a two-step process to select the context for a translation:

1. The ASI is determined (conceptually by the Integer Unit) from the instruction, trap level, and the virtual processor endian mode
2. The context register is determined directly from the ASI.

The ASI value and endianness (little or big) are determined for the I-MMU and D-MMU respectively according to TABLE 12-11 and TABLE 12-12 on page 127.

Notes

The secondary context is never used to fetch instructions.

The endianness of a data access is specified by three conditions: the ASI specified in the opcode or ASI register, the PSTATE current little endian bit, and the D-MMU invert endianness bit.

The D-MMU invert endianness (ie) bit inverts the endianness for all accesses to translating ASIs, including LD/st/Atomic alternates that have specified an ASI. That is, LDXA [%g1]ASI_PRIMARY_LITTLE will be big-endian if the ie bit is on. Accesses to nontranslating ASIs are not affected by the D-MMU's ie bit. See *Alternate Address Spaces* on page 71 or information about nontranslating ASIs.

TABLE 12-11 ASI Mapping for Instruction Accesses

Condition for Instruction Access	Resulting Action	
PSTATE.tl	Endianness	ASI Value (in SFSR)
0	Big	ASI_PRIMARY
> 0	Big	ASI_NUCLEUS

TABLE 12-12 ASI Mapping for Data Accesses

Condition for Data Access				Access Processed with:	
Opcode	PSTATE.tl	PSTATE.cle	TTE.ie	Endianness	ASI Value (Recorded in SFSR)
LD/ST/Atomic/FLUSH (Using Default ASI)	0	0	0	Big	ASI_PRIMARY
			1	Little	
		1	0	Little	ASI_PRIMARY_LITTLE
			1	Big	
	> 0	0	0	Big	ASI_NUCLEUS
			1	Little	
		1	0	Little	ASI_NUCLEUS_LITTLE
			1	Big	
LD/st/Atomic Alternate with specified ASI *not* ending in "_LITTLE"	Don't Care	Don't Care	0	Big[1]	Specified ASI value from immediate field in opcode or ASI register
			1	Little[1]	
LD/st/Atomic Alternate with specified ASI ending in '_LITTLE"	Don't Care	Don't Care	0	Little	Specified ASI value from immediate field in opcode or ASI register
			1	Big	

[1] Accesses to nontranslating ASIs are always made in "big endian" mode, regardless of the setting of the various ie bits. See *Alternate Address Spaces* on page 71 for information about nontranslating ASIs.

The context registers used by the data and instruction MMUs is determined from TABLE 12-13. A comprehensive list of ASI values can be found in the ASI map in *Alternate Address Spaces* on page 71. The context register selection is not affected by the endianness of the access

.

TABLE 12-13 I-MMU and D-MMU Context Register Usage

ASI Value	Context Register
ASI_*NUCLEUS*[1]	Nucleus (0000_{16} hard-wired)
ASI_*PRIMARY*[2]	Primary 0 and Primary 1
ASI_*SECONDARY*[3]	Secondary 0 and Secondary 1
All other ASI values	(Not applicable, no translation)

1. Any ASI name containing the string "NUCLEUS".
2. Any ASI name containing the string "PRIMARY".

3. Any ASI name containing the string “SECONDARY”.

12.7 Translation

The translation operation of MMUs is determined by the LSU_CONTROL_REG and the HPSTATE registers.

12.7.1 Instruction Translation

TABLE 12-14 describes the operation of the I-MMU.

TABLE 12-14 IMMU Translation

Control State			
LSU.im	**HPSTATE.hpriv**	**HPSTATE.red**	**IMMU Translation**
Don’t Care	Don’t Care	1	Bypass[1]
Don’t Care	1	0	Bypass[1]
0	0	0	RA $\rightarrow$ PA[2]
1	0	0	VA $\rightarrow$ PA[3]

1. VA{39:0} passed directly to PA{39:0}
2. VA{47:0} passed directly to RA{47:0}, RA{47:0} translated via the IMMU.
3. VA{47:0} translated via the IMMU.

12.7.1.1 Instruction Prefetching

UltraSPARC T2 fetches instructions sequentially (including delay slots). UltraSPARC T2 fetches delay slots before the branch is resolved (before whether the delay slot will be annulled is known). UltraSPARC T2 also fetches the target of a DCTI before the delay slot executes. For both these cases, UltraSPARC T2 may fetch from a nonexistent PA (in the case of a fetch from memory) or from an I/O address with side effects. Hypervisor should protect against this for virtual- and real-to-physical translations by maintaining valid mappings of sequential and target addresses at all times. Hypervisor should protect against this for bypassing translations by ensuring that all sequential and target addresses are backed by memory.

Certain instructions change the translation mode (writes to HPSTATE, stores to ASI_LSU_CONTROL_REG). Both of these translation mode changes can only be made while hyperprivileged. Since ASI_LSU_CONTROL_REG.im has no effect on translation while hyperprivileged, there are no issues with the hypervisor storing to ASI_LSU_CONTROL_REG.

However, a write to HPSTATE may change both the translation and privilege level at the same time. This translation change may result in the sequential instruction or the branch target being fetched in both the old translation mode (bypass) and then refetched in the new translation mode (either real to physical or virtual to physical) when UltraSPARC T2 realizes that translation has changed. For this case, it is desirable that the hypervisor enforce that valid translations exist for fetches in both the old translation mode (bypass) and the new translation mode (real to physical or virtual to physical). If the hypervisor is only able to enforce translations for the real or virtual to physical case, it is possible that an error may be encountered on the bypassing instruction fetch. Any precise error trap associated with the error will be suppressed and not be presented to the strand; however, the error status registers in the L2 cache, memory controller, and/or IO subsystem will still record the error (and possibly generate a disrupting trap in addition). Therefore, if the hypervisor is unable to enforce translation for both the bypass and real-/virtual-to-physical translations, it must be able to handle the potential spurious error logging in the L2 cache, memory controller, and/or IO subsystem.

12.7.2 Data Translation

TABLE 12-15 describes the operation of the D-MMU.

TABLE 12-15 DMMU Translation

Control State		DMMU Translation
LSU.dm	HPSTATE.hpriv	
0	0	RA → PA[1]
0	1	Follows
1	See	Follows

1. VA{63:0} passed directly to RA{63:0}, RA{63:0} translated via the DMMU.

TABLE 12-16 DMMU Translation When LSU_CONTROL_REG.dm = 1 or HPSTATE.hpriv = 1 *(1 of 5)*

ASI Value (hex)	ASI NAME	Translation		
		Nonprivileged	Privileged	Hypervisor
00_{16}–03_{16}	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
04_{16}	ASI_NUCLEUS	*privileged_action*	VA → PA	bypass
05_{16}–$0B_{16}$	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
$0C_{16}$	ASI_NUCLEUS_LITTLE	*privileged_action*	VA → PA	bypass
$0D_{16}$–$0F_{16}$	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
10_{16}	ASI_AS_IF_USER_PRIMARY	*privileged_action*	VA → PA	VA → PA
11_{16}	ASI_AS_IF_USER_SECONDARY	*privileged_action*	VA → PA	VA → PA
12_{16}–13_{16}	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
14_{16}	ASI_REAL	*privileged_action*	RA → PA	RA → PA
15_{16}	ASI_REAL_IO	*privileged_action*	RA → PA	RA → PA
16_{16}	ASI_BLOCK_AS_IF_USER_PRIMARY	*privileged_action*	VA → PA	VA → PA
17_{16}	ASI_BLOCK_AS_IF_USER_SECONDARY	*privileged_action*	VA → PA	VA → PA
18_{16}	ASI_AS_IF_USER_PRIMARY_LITTLE	*privileged_action*	VA → PA	VA → PA
19_{16}	ASI_AS_IF_USER_SECONDARY_LITTLE	*privileged_action*	VA → PA	VA → PA
$1A_{16}$–$1B_{16}$	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
$1C_{161}$	ASI_REAL_LITTLE	*privileged_action*	RA → PA	RA → PA
$1D_{16}$	ASI_REAL_IO_LITTLE	*privileged_action*	RA → PA	RA → PA
$1E_{16}$	ASI_BLOCK_AS_IF_USER_PRIMARY_LITTLE	*privileged_action*	VA → PA	VA → PA
$1F_{16}$	ASI_BLOCK_AS_IF_USER_SECONDARY_LITTLE	*privileged_action*	VA → PA	VA → PA
20_{16}	ASI_SCRATCHPAD	*privileged_action*	nontranslating	nontranslating
21_{16}	ASI_MMU	*privileged_action*	nontranslating	nontranslating
22_{16}	ASI_TWINX_AIUP, ASI_STBI_AIUP	*privileged_action*	VA → PA	VA → PA
23_{16}	ASI_TWINX_AIUS, ASI_STBI_AIUS	*privileged_action*	VA → PA	VA → PA
24_{16}	ASI_TWINX	*privileged_action*	VA → PA	bypass
25_{16}	ASI_QUEUE	*privileged_action*	nontranslating	nontranslating
26_{16}	ASI_TWINX_REAL	*privileged_action*	RA → PA	RA → PA

TABLE 12-16 DMMU Translation When LSU_CONTROL_REG.dm = 1 or HPSTATE.hpriv = 1 *(2 of 5)*

ASI Value (hex)	ASI NAME	Translation		
		Nonprivileged	Privileged	Hypervisor
27_{16}	ASI_TWINX_NUCLEUS, ASI_STBI_N	*privileged_action*	VA → PA	bypass
28_{16}–29_{16}	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
$2A_{16}$	ASI_TWINX_AIUPL, ASI_STBI_AIUPL	*privileged_action*	VA → PA	VA → PA
$2B_{16}$	ASI_TWINX_AIUSL, ASI_STBI_AIUSL	*privileged_action*	VA → PA	VA → PA
$2C_{16}$	ASI_TWINX_LITTLE	*privileged_action*	VA → PA	bypass
$2D_{16}$	*Reserved*	*privileged_action*	*DAE_invalid_asi*	*DAE_invalid_asi*
$2E_{16}$	ASI_TWINX_REAL_LITTLE	*privileged_action*	RA → PA	RA → PA
$2F_{16}$	ASI_TWINX_NL, ASI_STBI_NL	*privileged_action*	VA → PA	bypass
30_{16}	ASI_AS_IF_PRIV_PRIMARY	*privileged_action*	*privileged_action*	VA → PA
31_{16}	ASI_AS_IF_PRIV_SECONDARY	*privileged_action*	*privileged_action*	VA → PA
32_{16}–35_{16}	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
36_{16}	ASI_AS_IF_PRIV_NUCLEUS	*privileged_action*	*privileged_action*	VA → PA
37_{16}	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
38_{16}	ASI_AS_IF_PRIV_PRIMARY_LITTLE	*privileged_action*	*privileged_action*	VA → PA
39_{16}	ASI_AS_IF_PRIV_SECONDARY_LITTLE	*privileged_action*	*privileged_action*	VA → PA
$3A_{16}$–$3D_{16}$	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
$3E_{16}$	ASI_AS_IF_PRIV_NUCLEUS_LITTLE	*privileged_action*	*privileged_action*	VA → PA
$3F_{16}$	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
40_{16}	ASI_STREAM	*privileged_action*	*privileged_action*	nontranslating
41_{16}	ASI_CMP	*privileged_action*	*privileged_action*	nontranslating
42_{16}	ASI_INST_MASK_REG/ ASI_LSU_DIAG_REG	*privileged_action*	*privileged_action*	nontranslating
43_{16}	ASI_ERROR_INJECT_REG	*privileged_action*	*privileged_action*	nontranslating
44_{16}	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
45_{16}	ASI_LSU_CONTROL_REG, ASI_DECR, ASI_RST_VEC_MASK	*privileged_action*	*privileged_action*	nontranslating
46_{16}	ASI_DCACHE_DATA	*privileged_action*	*privileged_action*	nontranslating
47_{16}	ASI_DCACHE_TAG	*privileged_action*	*privileged_action*	nontranslating
48_{16}	ASI_IRF_ECC_REG	*privileged_action*	*privileged_action*	nontranslating
49_{16}	ASI_FRF_ECC_REG	*privileged_action*	*privileged_action*	nontranslating

TABLE 12-16 DMMU Translation When LSU_CONTROL_REG.dm = 1 or HPSTATE.hpriv = 1 *(3 of 5)*

ASI Value (hex)	ASI NAME	Translation		
		Nonprivileged	Privileged	Hypervisor
$4A_{16}$	ASI_STB_ACCESS	*privileged_action*	*privileged_action*	nontranslating
$4B_{16}$	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
$4C_{16}$	ASI_DESR/ASI_DFESR/ASI_CERER/ ASI_SETER/ASI_CLESR/ASI_ CLFESR	*privileged_action*	*privileged_action*	nontranslating
$4D_{16}$	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
$4E_{16}$	ASI_SPARC_PWR_MGMT	*privileged_action*	*privileged_action*	nontranslating
$4F_{16}$	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
50_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
51_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
52_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
53_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
54_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
55_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
56_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
57_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
58_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
59_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
$5A_{16}$	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
$5B_{16}$	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
$5C_{16}$	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
$5D_{16}$	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
$5E_{16}$	ASI_IMMU	*privileged_action*	*privileged_action*	nontranslating
$5F_{16}$	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
60_{16}–62_{16}	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
63_{16}	ASI_HYP_SCRATCHPAD	*privileged_action*	*privileged_action*	nontranslating
64_{16}–65_{16}	ASI_IMMU	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
66_{16}	ASI_ICACHE_INSTR	*privileged_action*	*privileged_action*	nontranslating
67_{16}	ASI_ICACHE_TAG	*privileged_action*	*privileged_action*	nontranslating
64_{16}–71_{16}	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
72_{16}	ASI_INTR_RECEIVE	*privileged_action*	*privileged_action*	nontranslating
73_{16}	ASI_INTR_W	*privileged_action*	*privileged_action*	nontranslating
74_{16}	ASI_INTR_R	*privileged_action*	*privileged_action*	nontranslating

TABLE 12-16 DMMU Translation When LSU_CONTROL_REG.dm = 1 or HPSTATE.hpriv = 1 *(4 of 5)*

ASI Value (hex)	ASI NAME	Translation		
		Nonprivileged	Privileged	Hypervisor
75_{16}–$7F_{16}$	*Reserved*	*privileged_action*	*privileged_action*	*DAE_invalid_asi*
80_{16}	ASI_PRIMARY	VA → PA	VA → PA	bypass
81_{16}	ASI_SECONDARY	VA → PA	VA → PA	bypass
82_{16}	ASI_PRIMARY_NO_FAULT	VA → PA	VA → PA	bypass
83_{16}	ASI_SECONDARY_NO_FAULT	VA → PA	VA → PA	bypass
84_{16}–87_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
88_{16}	ASI_PRIMARY_LITTLE	VA → PA	VA → PA	bypass
89_{16}	ASI_SECONDARY_LITTLE	VA → PA	VA → PA	bypass
$8A_{16}$	ASI_PRIMARY_NO_FAULT_LITTLE	VA → PA	VA → PA	bypass
$8B_{16}$	ASI_SECONDARY_NO_FAULT_ LITTLE	VA → PA	VA → PA	bypass
$8C_{16}$–BF_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$C0_{16}$	ASI_PST8_P	VA → PA	VA → PA	bypass
$C1_{16}$	ASI_PST8_S	VA → PA	VA → PA	bypass
$C2_{16}$	ASI_PST16_P	VA → PA	VA → PA	bypass
$C3_{16}$	ASI_PST16_S	VA → PA	VA → PA	bypass
$C4_{16}$	ASI_PST32_P	VA → PA	VA → PA	bypass
$C5_{16}$	ASI_PST32_S	VA → PA	VA → PA	bypass
$C6_{16}$–$C7_{16}$	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$C8_{16}$	ASI_PST8_PL	VA → PA	VA → PA	bypass
$C9_{16}$	ASI_PST8_SL	VA → PA	VA → PA	bypass
CA_{16}	ASI_PST16_PL	VA → PA	VA → PA	bypass
CB_{16}	ASI_PST16_SL	VA → PA	VA → PA	bypass
CC_{16}	ASI_PST32_PL	VA → PA	VA → PA	bypass
CD_{16}	ASI_PST32_SL	VA → PA	VA → PA	bypass
CE_{16}–CF_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$D0_{16}$	ASI_FL8_P	VA → PA	VA → PA	bypass
$D1_{16}$	ASI_FL8_S	VA → PA	VA → PA	bypass
$D2_{16}$	ASI_FL16_P	VA → PA	VA → PA	bypass
$D3_{16}$	ASI_FL16_S	VA → PA	VA → PA	bypass
$D4_{16}$–$D7_{16}$		*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$D8_{16}$	ASI_FL8_PL	VA → PA	VA → PA	bypass

TABLE 12-16 DMMU Translation When LSU_CONTROL_REG.dm = 1 or HPSTATE.hpriv = 1 *(5 of 5)*

ASI Value (hex)	ASI NAME	Translation		
		Nonprivileged	Privileged	Hypervisor
$D9_{16}$	ASI_FL8_SL	VA → PA	VA → PA	bypass
DA_{16}	ASI_FL16_PL	VA → PA	VA → PA	bypass
DB_{16}	ASI_FL16_SL	VA → PA	VA → PA	bypass
DC_{16}–DF_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$E0_{16}$	ASI_BLK_COMMIT_PRIMARY	VA → PA	VA → PA	bypass
$E1_{16}$	ASI_BLK_COMMIT_SECONDARY	VA → PA	VA → PA	bypass
$E2_{16}$	ASI_TWINX_P, ASI_STBI_P	VA → PA	VA → PA	bypass
$E3_{16}$	ASI_TWINX_S, ASI_STBI_S	VA → PA	VA → PA	bypass
$E4_{16}$–$E9_{16}$	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
EA_{16}	ASI_TWINX_PL, ASI_STBI_PL	VA → PA	VA → PA	bypass
EB_{16}	ASI_TWINX_PL, ASI_STBI_PL	VA → PA	VA → PA	bypass
EC_{16}-EF_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$F0_{16}$	ASI_BLK_PRIMARY	VA → PA	VA → PA	bypass
$F1_{16}$	ASI_BLK_SECONDARY	VA → PA	VA → PA	bypass
$F2_{16}$-$F7_{16}$	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*
$F8_{16}$	ASI_BLK_PRIMARY_LITTLE	VA → PA	VA → PA	bypass
$F9_{16}$	ASI_BLK_SECONDARY_LITTLE	VA → PA	VA → PA	bypass
FA_{16}-FF_{16}	*Reserved*	*DAE_invalid_asi*	*DAE_invalid_asi*	*DAE_invalid_asi*

12.8 MMU Behavior During Reset and Upon Entering RED_state

MMU Reset and RED_state behavior is described in *Machine State After Reset and in RED_State* on page 171.

12.9 Compliance With the SPARC V9 Annex F

The UltraSPARC T2 MMU complies completely with the SPARC V9 MMU Requirements described in Annex F of the *The SPARC Architecture Manual, Version 9*. TABLE 12-17 shows how various protection modes can be achieved, if necessary, through the presence or absence of a translation in the I- or D-MMU. Note that this behavior requires specialized TLB miss handler code to guarantee these conditions.

TABLE 12-17 MMU Compliance With SPARC V9 Annex F Protection Mode

Condition			
TTE in D-MMU	**TTE in I-MMU**	**Writable Attribute Bit**	**Resultant Protection Mode**
Yes	No	0	Read-only
No	Yes	Don't Care	Execute-only
Yes	No	1	Read/Write
Yes	Yes	0	Read-only/Execute
Yes	Yes	1	Read/Write/Execute

12.10 MMU Internal Registers and ASI Operations

12.10.1 Accessing MMU Registers

All internal MMU registers can be accessed directly by the virtual processor through ASIs defined by UltraSPARC T2.

See Section 12.7 for details on the behavior of the MMU during all other UltraSPARC T2 ASI accesses.

Note | STXA to an MMU register *does not* require any subsequent instructions such as a MEMBAR #Sync, FLUSH, DONE, or RETRY before the register effect will be visible to load / store / atomic accesses. UltraSPARC T2 resolves all MMU register hazards via an automatic synchronization on all MMU register writes.

If the low order three bits of the VA are non-zero in an LDXA/STXA to/from these registers, a *mem_address_not_aligned* trap occurs. Writes to read-only, reads to write-only, illegal ASI values, or illegal VA for a given ASI may cause a *DAE_invalid_asi* trap.

Caution | UltraSPARC T2 does not check for out-of-range virtual addresses during an STXA to any internal register; it simply sign-extends the virtual address based on VA{47}. Software must guarantee that the VA is within range.

Writes to the TSB register, Tag Access register, and Instruction and Data Watchpoint Address registers are not checked for out-of-range VA. No matter what is written to the register, VA{63:47} will always be identical on a read.

TABLE 12-18 UltraSPARC T2 MMU Internal Registers and ASI Operations

I-MMU ASI	D-MMU ASI	VA{63:0}	Access	Register or Operation Name
21_{16}		8_{16}	Read/Write	Primary Context 0 register
—	21_{16}	10_{16}	Read/Write	Secondary Context 0 register
21_{16}		108_{16}	Read/Write	Primary Context 1 register
—	21_{16}	110_{16}	Read/Write	Secondary Context 1 register
50_{16}	58_{16}	0_{16}	Read-only	I-/D-TSB Tag Target registers
—	58_{16}	20_{16}	Read-only	D-TLB Synchronous Fault Address register
50_{16}	58_{16}	30_{16}	Read/Write	I-/D-TLB Tag Access registers
50_{16}	58_{16}	38_{16}	Read/Write	Watchpoint address
58_{16}		40_{16}	Read/Write	Hardware Tablewalk Config register
58_{16}		80_{16}	Read/Write	Partition identifier
52_{16}		108_{16}	Read/Write	MMU Real Range 0 register
52_{16}		110_{16}	Read/Write	MMU Real Range 1 register
52_{16}		118_{16}	Read/Write	MMU Real Range 2 register
52_{16}		120_{16}	Read/Write	MMU Real Range 3 register
52_{16}		208_{16}	Read/Write	MMU Physical Offset 0 register
52_{16}		210_{16}	Read/Write	MMU Physical Offset 1 register
52_{16}		218_{16}	Read/Write	MMU Physical Offset 2 register
52_{16}		220_{16}	Read/Write	MMU Physical Offset 3 register
54_{16}		10_{16}	Read/Write	MMU Context Zero TSB Config 0 register
54_{16}		18_{16}	Read/Write	MMU Context Zero TSB Config 1 register
54_{16}		20_{16}	Read/Write	MMU Context Zero TSB Config 2 register
54_{16}		28_{16}	Read/Write	MMU Context Zero TSB Config 3 register
54_{16}		30_{16}	Read/Write	MMU Context Nonzero TSB Config 0 register
54_{16}		38_{16}	Read/Write	MMU Context Nonzero TSB Config 1 register
54_{16}		40_{16}	Read/Write	MMU Context Nonzero TSB Config 2 register

TABLE 12-18 UltraSPARC T2 MMU Internal Registers and ASI Operations *(Continued)*

I-MMU ASI	D-MMU ASI	VA{63:0}	Access	Register or Operation Name
54_{16}		48_{16}	Read/Write	MMU Context Nonzero TSB Config 3 register
54_{16}		50_{16}	Read-only	MMU I-TSB Pointer 0 register
54_{16}		58_{16}	Read-only	MMU I-TSB Pointer 1 register
54_{16}		60_{16}	Read-only	MMU I-TSB Pointer 2 register
54_{16}		68_{16}	Read-only	MMU I-TSB Pointer 3 register
54_{16}		70_{16}	Read-only	MMU D-TSB Pointer 0 register
54_{16}		78_{16}	Read-only	MMU D-TSB Pointer 1 register
54_{16}		80_{16}	Read-only	MMU D-TSB Pointer 2 register
54_{16}		88_{16}	Read-only	MMU D-TSB Pointer 3 register
54_{16}		90_{16}	Read/Write	MMU Tablewalk Pending Control register
54_{16}		98_{16}	Read-only	MMU Tablewalk Pending Status register
54_{16}	$5C_{16}$	See Section 12.10.15	Write-only	I-/D-TLB Data In registers
55_{16}	$5D_{16}$	See Section 12.10.15	Read/Write	I-/D-TLB Data Access registers
56_{16}	$5E_{16}$	See Section 12.10.15	Read-only	I-/D-TLB Tag Read register
57_{16}	$5F_{16}$	See Section 12.11.1	Write-only	I-/D-MMU demap operation

12.10.2 Context Registers

UltraSPARC T2 supports a pair of primary and a pair of secondary context registers per strand, which are shared by the I- and D-MMUs. Primary Context 0 and Primary Context 1 are the primary context registers, and a TLB entry for a translating primary ASI can match the context field with either Primary Context 0 or Primary Context 1 to produce a TLB hit. Secondary Context 0 and Secondary Context 1 are the secondary context registers, and a TLB entry for a translating secondary ASI can match the context field with either Secondary Context 0 or Secondary Context 1 to produce a TLB hit.

Compatibility Note | To maintain backward compatibility with software designed for a single primary and single secondary context register, writes to Primary (Secondary) Context 0 Register also update Primary (Secondary) Context 1 Register.

The Primary Context 0 and Primary Context 1 registers are defined as shown in FIGURE 12-2, where pcontext is the context value for the primary address space.

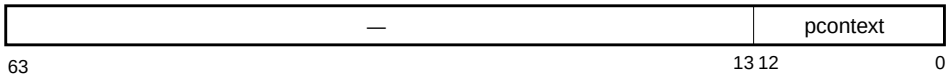


FIGURE 12-2 Primary Context 0/1 register

The Secondary Context 0 and Secondary Context 1 Registers are defined in FIGURE 12-3, where scontext is the context value for the secondary address space.

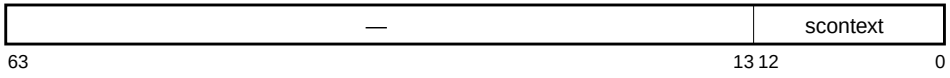


FIGURE 12-3 Secondary Context 0/1 Register

The contents of the Nucleus Context register are hardwired to the value zero:

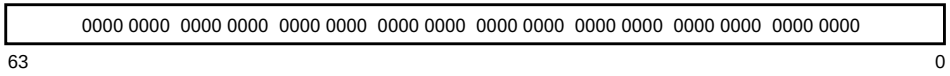


FIGURE 12-4 Nucleus Context Register

12.10.3 I-/D-TSB Tag Target Registers

The I- and D-TSB Tag Target registers are simply respective bit-shifted versions of the data stored in the I- and D-Tag Access registers. Since the I- or D-Tag Access registers are updated on I- or D-TLB misses, respectively, the I- and D-Tag Target registers appear to software to be updated on an I- or D-TLB miss. A write to this register results in a *DAE_invalid_asi* trap being taken. The registers are illustrated in FIGURE 12-5 and described in the table below the figure.

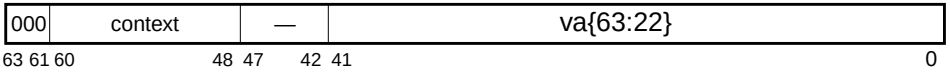


FIGURE 12-5 MMU Tag Target Registers (Two Registers)

Bit	Field	Description
60:48	context	I/D context{12:0}: The context associated with the missing virtual address. For real translations, the context value is set to zero.
41:0	va	I/D context{12:0}: The context associated with the missing virtual address. For real translations, the context value is set to zero.

Notes When PSTATE.am = 1, the upper 32 bits of the VA captured in this register will be zero.

For a 256-Mbyte page, VA{27:22} contain bits 27:22 of the virtual address and are not zeroed by hardware.

12.10.4 I-/D-MMU Synchronous Fault Address Registers (SFAR)

12.10.4.1 I-MMU Fault Address

There is no I-MMU Synchronous Fault Address register. Instead, software must read the TPC register appropriately as discussed here.

For *instruction_access_MMU_miss* traps, TPC contains the virtual address that was not found in the I-MMU TLB.

For *IAE_privilege_violation*, *IAE_unauth_access*, and *IAE_nfo_page* traps, TPC contains the virtual address of the instruction in the privileged page that caused the exception.

For *instruction_address_range* and *instruction_real_range* traps, note that the TPC in these cases contains only a 48-bit virtual (real) address, which is sign-extended based on bit VA{47} (RA{47}) for read. Thus, the TPC contains only the lower 48 bits of the virtual (real) address that is out of range.

12.10.4.2 D-MMU Fault Address

The Synchronous Fault Address register contains the virtual memory address of the access that caused the following exceptions:

- *DAE_invalid_asi*
- *DAE_privilege_violation*
- *DAE_nc_page*
- *DAE_nfo_page*
- *DAE_side_effect_page*
- *mem_address_range*
- *mem_real_range*
- *asi_data_access_protection*
- *privileged_action*
- *VA_watchpoint*
- *PA_watchpoint*
- *mem_address_not_aligned*
- *LDDF_mem_address_not_aligned*

- *STDF_mem_address_not_aligned*

This register is read-only, a write to this register results in a *DAE_invalid_asi* trap being taken.

FIGURE 12-6 illustrates the D-SFAR; the va field is described below the table.

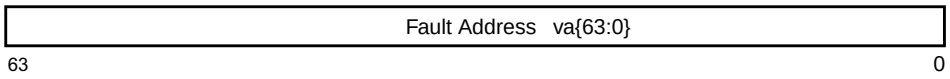


FIGURE 12-6 D-MMU Synchronous Fault Address Register (SFAR) Format

Bit	Field	Description
63:0	va	Fault Address: The virtual (real) address associated with the translation fault. This field is sign-extended based on VA{47} (RA{47}), so bits VA{63:48} (RA{63:48}) do not correspond to the virtual (real) address used in the translation for the case of a VA (RA) out-of-range *mem_address_range* (mem_real_range) trap (for this case, software must disassemble the trapping instruction).

Notes When PSTATE.am = 1, the upper 32 bits of the VA captured in this register will be zero.

The DSFAR is shared for precise error handling, and contains the error address as described in *DMMU Synchronous Fault Address Register* on page 247 following an *internal_processor_error*, *data_access_MMU_error*, or *data_access_error*.

12.10.5 I-/D-TLB Tag Access Registers

In each MMU the Tag Access register is used as a temporary buffer for writing the TLB Entry tag information. The Tag Access register may be updated during any of the following operations:

1. When the MMU signals a trap due to a miss, exception, or protection. The MMU hardware automatically writes the missing VA and the appropriate context (`ASI_PRIMARY_CONTEXT_0` for primary context accesses, `ASI_SECONDARY_CONTEXT_0` for secondary context accesses, `ASI_NUCLEUS_CONTEXT` for other accesses) into the Tag Access register to facilitate formation of the TSB Tag Target register. See TABLE 12-6 for the Tag Access register update policy.

2. An ASI write to the Tag Access register. Before an ASI store to the TLB Data Access registers, the operating system must set the Tag Access register to the values desired in the TLB Entry. Note that an ASI store to the TLB Data-In register for automatic replacement also uses the Tag Access register, but typically the value written into the Tag Access register by the MMU hardware is appropriate.

3. An I-/D-MMU demap operation. For an I-/D-MMU demap operation, the corresponding Tag Access register va field is loaded with the matching VA bits from the demap store address. If the Context ID field of the demap store address (see Section 12.11.1) is 00, then the context field of the Tag Access register is loaded with the context of `ASI_PRIMARY_CONTEXT_0`. Otherwise, the context field of the Tag Access register is loaded with all zeros.

4. An I-/D-TLB load by the Hardware Tablewalker.

Note | Any update to the Tag Access registers immediately affects the data that is returned from subsequent reads of the Tag Target and TSB Pointer registers.

Compatibility Note | The updating of Tag Access on a demap operation and hardware tablewalk is specific to UltraSPARC T2. No previous UltraSPARC processors updated Tag Access on demap, and no previous UltraSPARC processors supported Hardware Tablewalk.

The TLB Tag Access registers are defined in FIGURE 12-7; register bits are described in the table below the figure.

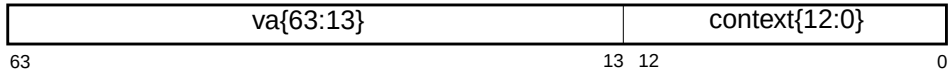


FIGURE 12-7 I/D MMU TLB Tag Access Registers

Bit	Field	Description
63:13	va	The 51-bit virtual page number. Note that writes to this field are not checked for out-of-range violation, but sign extended based on VA{47}. **NOTE:** When PSTATE.am = 1, the upper 32 bits of the VA captured in this register will be zero.
12:0	context	The 13-bit context identifier. This field reads zero when there is no associated context with the access, such as for an internal ASI or a real to physical translation.

Caution – Stores to the Tag Access registers are not checked for out-of-range violations. Reads from these registers are sign-extended based on VA{47}.

12.10.6 Partition Identifier

A partition identifier register is provided per strand to allow multiple OSs to share the same TLB. The partition identifier register contents are compared in all TLB operations such as demaps and translations, and are loaded into the pid field of the TLB tag during insertions.

The Partition Identifier register is defined in FIGURE 12-8, where pid is the 3-bit partition identifier.

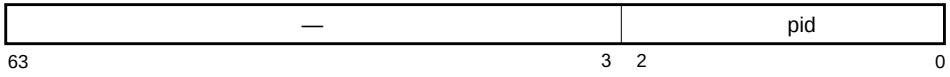


FIGURE 12-8 Partition Identifier Register

12.10.7 Hardware Tablewalk Configuration Register

Each strand has a Hardware Tablewalk Configuration register that controls operation of the Hardware Tablewalk unit.

The Hardware Tablewalk Configuration register is defined in FIGURE 12-9; the register bits are described in the table below the figure.

.

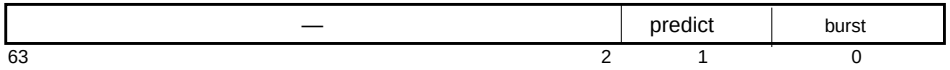


FIGURE 12-9 Hardware Tablewalk Config Register

Bit	Field	Description
1	predict	If burst is set to 0, predict controls whether hardware prediction is used to order the TSB reads. If set to 0, the order of TSBs is always 0, 1, 2, 3. If set to 1, a hardware predictor is used to order the TSB reads as either 0, 1, 2, 3 or 1, 0, 2, 3. Initial value is 0.
0	burst	If set to 1, TSB reads are issued to all four TSBs in parallel. If set to 0, TSB reads are issued sequentially, stopping when a TSB hit is detected. Initial value is 0.

12.10.8 ITLB Probe

A read-only ITLB probe ASI assists software with determining the physical address assigned to a virtual address of an instruction. The virtual address used with the ASI is presented to the ITLB, and if a translation exists for the specified address and the primary contexts in the ITLB, the result data will contain the physical page address and have the valid bit set. If a translation does not exist for the specified address and

the primary contexts in the ITLB, the result data will have the valid bit clear. The ITLB probe does not trap on a probe that specifies a virtual address in the VA hole; bits 63:48 of the address are ignored.

Programming Note The ITLB probe always uses the primary context registers to determine an ITLB hit. If software desires to check for a nucleus translation, it must first zero the primary context before issuing a load to the ITLB probe ASI.

The format of the ITLB probe virtual address is shown in FIGURE 12-10; the fields are described in the table below the figure

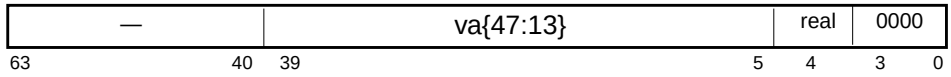


FIGURE 12-10 ITLB Probe Address Format.

Bit	Field	Description
39:13	va	The 35-bit virtual page number (real page number if the real bit is 1).
4	real	If set, the ITLB is checked for real-to-physical translations. If cleared, the ITLB is checked for virtual-to-physical translations.

The ITLB probe data format is defined in FIGURE 12-11; the fields are described in TABLE 12-19.

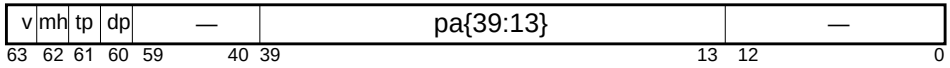


FIGURE 12-11 ITLB Probe Data Format.

TABLE 12-19 Format of ITLB Probe Data Fields

Bit	Field	Description
63	v	Valid bit for the physical page number. Set if there was a match in the TLB.
62	mh	Multiple hit. Valid only if v is 1. Set if there were multiple matches in the TLB.
61	tp	Tag parity error. Valid only if v is 1 and mh is 0. Set if there was a tag parity error in the matching TLB entry.
60	dp	Data parity error. Valid only if v is 1, mh is 0, and tp is 0. Set if there was a data parity error in the matching TLB entry.
39:13	pa	The 27-bit physical page number.

12.10.9 MMU Real Range Registers

There are four Real Range registers per strand. The RPN-to-PPN translation associates each Real Range register with its corresponding Physical Offset register. The RA-to-PA translation applies to TTEs from TSBs with the ra_not_pa bit set in the TSB Config register, regardless of zero or non-zero context, as described in *Hardware Tablewalk* on page 110.

If the enable field is 0, then this range and offset pair are not used. If all range and offset pairs are disabled, any hit in a TSB with the ra_not_pa bit set in the TSB Config register results in an *instruction_invalid_TSB_entry* or *data_invalid_TSB_entry* trap.

TABLE 12-20 lists the fields of the MMU Real Range registers.

TABLE 12-20 MMU Real Range Register Format

Bit	Field	Description
63	enable	Enables range and offset pair.
62:54	—	*Reserved*
53:27	rpn_high	RA{39:13} of the upper limit of the RPN range (bounds).
26:0	rpn_low	RA{39:13} of the lower limit of the RPN range (base).

12.10.10 MMU Physical Offset Registers

There are four Physical Offset registers per strand. The RPN-to-PPN translation associates each Real Range register with its corresponding Physical Offset register. The RA-to-PA translation applies to TTEs from TSBs with the ra_not_pa bit set in the TSB Config register, regardless of zero or nonzero context, as described in Section 12.3.1.1.

Programming Note For proper operation at all page sizes, the value programmed into the ppn field must be aligned to the size of the largest page that will use the Physical Offset register for RA-to-PA translation.

TABLE 12-21 lists the fields of the MMU Physical Offset registers.

TABLE 12-21 MMU Physical Offset Register Format

Bit	Field	Description
63:40	—	*Reserved*
39:13	ppn	Added to RA{39:13} of the request to generate PA{39:13}.
12:0	—	*Reserved*

12.10.11 MMU TSB Config Registers

The TSB Config registers (MMU_{NON}ZERO_CONTEXT_TSB_CONFIG_<0,1,2,3>) to provide information for Hardware Tablewalk and for the hardware formation of TSB pointers and tag targets to assist software in handling TLB misses quickly. If the TSB concept is not employed in the software memory management strategy, and therefore the hardware tablewalk, pointer, and tag access registers are not used, then the TSB Config registers need not contain valid data other than having the enable bit set to 0. Each strand has four separate TSB pointers for both the zero and non-zero contexts.

TABLE 12-22 describes the fields of the TSB Config registers.

TABLE 12-22 TSB Config Register Format

Bit	Field	Description
63	enable	If set to 1, Hardware Tablewalk will search this TSB on TLB misses. **NOTE**: If any of a strand's TSB Config registers has the enable bit set, Hardware Tablewalk is considered to be enabled for the strand.
62	use_context_0	With use_context_1, controls whether Hardware Tablewalk checks the context value in the TTE from this TSB and what context value is written into the TTE in the TLB. If both bits are 0, then Hardware Tablewalk compares the context in the TTE from the TSB to the context of the request and stores that context into the TLB if the TTE matches. If either bit is 1, Hardware Tablewalk ignores the context of the TTE from the TSB. If use_context_0 is 1, Hardware Tablewalk writes the value of Context Register 0 to the TLB; otherwise, if use_context_1 is 1, Hardware Tablewalk writes the value of Context Register 1 to the TLB. **NOTE:** When the requesting context is zero (nucleus), Hardware Tablewalk ignores these bits.
61	use_context_1	See use_context_0 above.
60:40	—	*Reserved*
39:13	tsb_base	PA{39:13} of the base of the TSB table
12:9	—	*Reserved*
8	ra_not_pa	If set, enables RPN-to-VPN translation in Hardware Tablewalk. **CAUTION!** When using Hardware Tablewalk for a TSB, the TSB may contain either RAs or PAs, but not both. The ra_not_pa bit should be set when the TSB contains RAs.
7:4	page_size	Contains the size of the pages mapped by the TTEs in the TSB. This page size is used to generate the TSB pointer. If a reserved page size value is attempted to be stored to this field, an *unsupported_page_size* trap is taken instead.

TABLE 12-22 TSB Config Register Format *(Continued)*

Bit	Field	Description
3:0	tsb_size	The size field provides the size of the TSB according to the following: • Number of entries in the TSB = 512×2^{tsb_size}. • Number of entries in the TSB ranges from 512 entries at tsb_size = 0 (8-Kbyte TSB), to 16 M entries at tsb_size =1 5 (256-Mbyte TSB). **NOTE:** When the page size for a TSB Base is set to 5 (256-Mbyte pages), setting tsb_size to a value greater than 11 is a programming error that creates a TSB that maps a larger than 48-bit VA range. UltraSPARC T2 forces the TSB pointer bits generated by VA bits above VA{47} to be 0 for this case. **NOTE:** Any update to the TSB Config register immediately affects the data that is returned from later reads of the corresponding TSB Pointer registers.

12.10.12 MMU I-/D-TSB Pointer Registers

The per-strand TSB Pointer registers (MMU_ITSB_POINTER_<0,1,2,3>, MMU_DTSB_POINTER_<0,1,2,3>) are provided to allow software to location of a missing TTE in a software-maintained TSB.

The TSB Pointer registers are implemented as a reorder of the current data stored in the Tag Access register and the appropriate TSB Config register. If the Tag Access register or the TSB Config register is updated through a direct software write (via an STXA instruction), then the Pointer registers' values will be updated as well.

The I-/D-TSB Pointer registers are defined in FIGURE 12-12. pa{39:0} is the full physical address of the TTE in the TSB, as determined by the MMU hardware. The formula to generate this field is as follows:

```
PA{39:0} = TSB_Base{39:13+N} || VA{21+N+3*PS:13+3*PS} || 0000
```

where *N* is defined to be the tsb_size field of the TSB Config register; it ranges from 0 to 15. TSB_Base refers to the tsb_base field of the TSB Config register. PS refers to the page_size field of the TSB Config register.

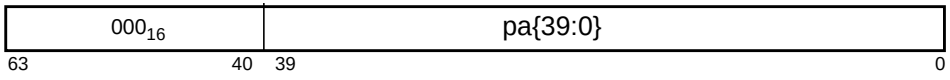


FIGURE 12-12 I-/D-TSB Pointer Registers

12.10.13 MMU Tablewalk Pending Control Register

Each strand has a MMU Tablewalk Pending Control register. This register can be used by software to indicate the status of a software tablewalk. Minimally, software should write a 1 to the stp bit before it fetches a TTE from a TSB and should write a 0 to the STP bit after it has written the TTE to the TLB or has determined that the TTE will not be written to the TLB.

Note | This register is completely maintained by software.

TABLE 12-24 lists the fields of the MMU Tablewalk Pending Control register.

TABLE 12-23 MMU Tablewalk Pending Control Register Format

Bit	Field	Description
63:1	—	*Reserved*
0	stp	Indicates whether a software tablewalk is in progress.

12.10.14 MMU Tablewalk Pending Status Register

Each physical core has a read-only MMU Tablewalk Pending Status register. This register allows software to identify when in-progress tablewalks have completed. Software can invalidate an entry in a TSB and then poll this register to identify tablewalks that may be temporarily caching the entry that has been invalidated. The bits that are 1 on the initial poll indicate pending tablewalks. A bit that initially sampled as 1 but later samples as 0 indicates an in-progress tablewalk has completed. Once each of the bits that initially were sampled as 1 have been sampled as 0, all tablewalks that were in progress when the initial poll was taken have been completed.

Programming Note Because successive hardware tablewalks can set the htp bits again, it is possible for software to undersample. That is, polling software can miss a 1 to 0 transition if hardware clears and sets the bit between adjacent polls. Polling software should be structured to minimize the possibility of undersampling.

TABLE 12-24 lists the fields of the MMU Tablewalk Pending Control register.

TABLE 12-24 MMU Tablewalk Pending Control Register Format

Bit	Field	Description
63:40	—	*Reserved*
39:32	htp	Indicates whether a hardware tablewalk is in progress for strands 7:0.
31:8	—	*Reserved*
7:0	stp	Indicates whether a software tablewalk is in progress for strands 7:0.

12.10.15 I-/D-TLB Data-In/Data-Access/Tag-Read Registers

Access to the TLB is complicated due to the need to provide an atomic write of a TLB entry data item (tag and data) that is larger than 64 bits, the need to replace entries automatically through the TLB entry replacement algorithm as well as provide direct diagnostic access, the need to allow that multiple strands on the

physical core that share the TLB to do a lock-free TLB update, and the need for hardware assist in the TLB miss handler. TABLE 12-25 shows the effect of loads and stores on the Tag Access register and the TLB.

TABLE 12-25 Effect of Loads and Stores on MMU Registers

Software Operation		Effect on MMU Physical Registers		
Load/Store	**Register**	**TLB tag**	**TLB data**	**Tag Access Register**
Load	Tag Read	No effect. Contents returned	No effect	No effect
	Tag Access	No effect	No effect	No effect. Contents returned
	Data In	Trap with *DAE_invalid_asi*		
	Data Access	No effect	No effect. Contents returned	No effect
Store	Tag Read	Trap with *DAE_invalid_asi*		
	Tag Access	No effect	No effect	Written with store data
	Data In	TLB entry determined by replacement policy written with contents of Tag Access Register	TLB entry determined by replacement policy written with store data	No effect
	Data Access	TLB entry specified by STXA address written with contents of Tag Access Register	TLB entry specified by STXA address written with store data	No effect
TLB miss		No effect	No effect	Written with VA and context of access

The Data In and Data Access registers are the means of reading and writing the TLB for all operations. The TLB Data In register is used for TLB miss and TSB-miss handler automatic replacement writes; the TLB Data Access register is used for operating system and diagnostic directed writes (writes to a specific TLB entry). The real bit of the TLB is under the control of bit 10 of the VA. If this bit is set, the real bit of the TLB entry is set; otherwise, the real bit of the TLB entry is cleared.

Notes When a real-to-physical translation is loaded into the TLB, the context value loaded into the TLB is always 3_{16}.

Hardware Tablewalk updates the corresponding I- or D-Data In/Data Access registers (in addition to the Tag Access register) whenever it loads a translation into the TLB.

The hardware supports an autodemap function to handle the case where two strands sharing a TLB try to enter the same translation into the TLB (for example, due to near-simultaneous TLB misses on the same page). A TLB replacement that attempts to add an already existing translation will cause the existing translation to be removed from the TLB.

Notes

Autodemapping of existing translations will always remove an existing page of the same size or larger than the one being added to the TLB. For example, an insertion of a 8-Kbyte page that sits inside the virtual address range of a 64-Kbyte page will cause the 64-Kbyte page to be autodemapped. Smaller pages that sit inside a page being added to the TLB may not be autodemapped. For example, an insertion of a 4-Mbyte page that overlaps the virtual address of one or more 64 KB pages may not autodemap the overlapping 64-Kbyte pages.[1] A subsequent multiple-hit error in the TLB could be generated as the result of a programming error that inserted a larger page in the TLB that overlapped smaller pages present in the TLB. A multiple-hit error occurs when a translation request matches more than one TTE in the TSB, and so a multiple-hit error only occurs for accesses of the region common to the overlapping pages.

The pids and real bits on the pages must match for autodemap to take place. If the real bit is 0, the context IDs must match as well.

IF a TLB replacement is attempted using a reserved page size value, an *unsupported_page_size* trap will be taken instead.

If a TLB replacement is attempted with the value of the valid bit (v) equal to 0, the MMU will treat that the same as if the valid bit was 1 for purposes of allocating and overwriting a TLB entry and autodemapping matching pages, and the entry will be written into the TLB with the v bit set to 0.

1. Whether the smaller pages are autodemapped depends on the actual demap address used and the position of the smaller page within the larger page.

The format of the TLB Data-In register virtual address is shown in FIGURE 12-13, where real is written to the real bit of the TLB entry.

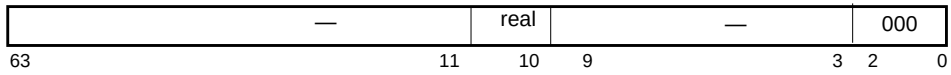


FIGURE 12-13 MMU TLB Data-In Virtual Address Format

The format of the TLB Data Access register virtual address is shown in FIGURE 12-14 and described in the table below the figure.

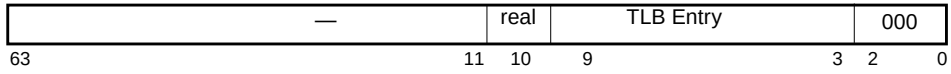


FIGURE 12-14 MMU TLB Data Access Virtual Address Format

Bit	Field	Description
10	real	Written to the Real bit of the TLB entry.
9:3	TLB Entry	The TLB Entry number to be accessed, in the range 0..63 for the ITLB, 0..127 for the DTLB.

The data format for TLB Data In and TLB Data Access registers is shown in TABLE 12-26. Reserved fields ignore writes and return all zeros on reads.

TABLE 12-26 I-/D-MMU TLB Data In and Data Access Registers

Bit	Field	Description
63	v	Valid.
62	nfo	No-fault-only.
61	parity	Parity for the TLB Data Entry.[1]
60:40	—	*Reserved*
39:13	pa	PA{39:13}.
12	ie	Invert endianess.
11	e	Side-effect.
10	cp	Cacheable in physically-indexed cache.
9	cv	*Reserved*
8	p	Privileged.
7	ep	*Reserved*
6	w	Writable.
5:4	soft	*Reserved*
3:0	size	Size.

1. Data parity is generated across pa{39:13}, nfo, ie, cp, e, p, w, and an encoded version of the page size. The page size is encoded in three bits as follows: 111– 256 Mbytes; 011 – 4 Mbytes; 001 – 64 Kbytes;
000 – 8 Kbytes. For the encoded page sizes, 256 Mbyte changes from the size{2:0} value of 101 to 111, while the encoding for all other pages match size{2:0}.

The format of the Tag Read register virtual address is shown in FIGURE 12-15 and described in the table below the figure.

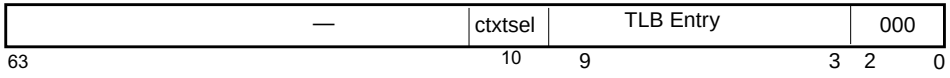


FIGURE 12-15 MMU Tag Read Virtual Address Format

Bit	Field	Description
10	ctxtsel	If 0, context A is read out in the context field. If 1, context B is read out in the context field. **NOTE:** The TLBs store duplicate copies of the context field for error detection, and ctxtsel allows software to examine both copies.
9:3	TLB Entry	The TLB Entry number to be accessed, in the range 0..63 for the ITLB, 0..127 for the DTLB

The data format for the Tag Read register is shown in FIGURE 12-16 and described in TABLE 12-27.

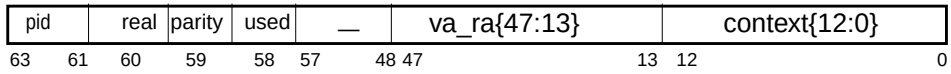


FIGURE 12-16 I-/D-MMU TLB Tag Read Registers

TABLE 12-27 Data Format for I-/D-MMU TLB Tag Read Registers

Bit	Field	Description
63:61	pid	3-bit partition identifier.
69	real	If set, identifies an RA-to-PA translation instead of a VA-to-PA.
59	parity	Parity for the tag entry. Parity is generated across pid, real, va{47:13}, and the context value that was written into the pair of context fields.
58	used	Used bit for replacement algorithm.
47:13	va_ra	If the r bit is 0, contains the lower bits of the 51-bit virtual page number (VA{47:13}). If the r bit is 1, contains the lower bits of a 51-bit real page number (RA{47:13}). Page offset bits for page sizes larger than 8 KB are stored as zeros in the TLB and returned for a Tag Read register read; that is, va_ra{15:13}, va_ra{21:13}, and va_ra{27:13} are zeroed for 64-Kbyte, 4-Mbyte, and 256-Mbyte pages, respectively. **PROGRAMMING NOTE:** Software needs to sign-extend the va_ra field based on va_ra{47}. **PROGRAMMING NOTE:** If the ra bit is 1, it is up to software to ensure that va_ra{47:40} are all zeros.
12:0	context	13-bit context identifier. The copy of context loaded from the TLB Tag entry is selected by the ctxtsel bit of the address. If ctxtsel is 0, this field contains context A If ctxtsel is 1, this field contains context B.

An ASI store to the TLB Data Access register initiates an internal atomic write to the specified TLB Entry. The TLB entry data is obtained from the store data, and the TLB entry tag is obtained from the current contents of the TLB Tag Access register.

An ASI store to the TLB Data-In register initiates an automatic atomic replacement of the TLB Entry pointed to by the replacement index generated internally by the TLB. The TLB data and tag are formed as in the case of an ASI store to the TLB Data Access register described above.

An ASI load from the TLB Data Access register initiates an internal read of the data portion of the specified TLB entry.

An ASI load from the TLB Tag Read register initiates an internal read of the tag portion of the specified TLB entry.

ASI loads from the TLB Data-In register are not supported and generate a *DAE_invalid_asi* trap.

12.11 I/D-MMU Demap

12.11.1 I-/D-MMU Demap

Demap is an MMU operation, as opposed to a register operation as described above. The purpose of demap is to remove zero, one, or more entries in the TLB. Four types of demap operation are provided: Demap Page, Demap Context, Demap All, and Demap All Pages. All demap operations only demap those pages whose PID matches the PID specified in the Partition Identifier register. Demap Page removes zero or one[1] TLB entry that matches exactly the specified virtual page number and real bit. Demap Context removes zero, one, or many TLB entries that match the specified context identifier and have the real bit cleared. Demap Context will never demap a real translation (r = 1). Demap All Pages removes all pages that either have their real bit set (if the r bit in the demap address is set) or their real bit clear (if the r bit in the demap address is clear), regardless of their context. Demap All removes all pages, regardless of their context or real bit.

The Demap Page operation has two forms: demap page and demap real page. Address bit 10 controls which form of demap page is used. If address bit 10 is a 1, only a real translation entry in the TLB (r = 1) where the RA matches the VA portion of the demap address will be demapped (the context is ignored on demap real translation). If address bit 10 is a 0, only a virtual-to-physical translation entry in the TLB (r = 0) where the VA of the entry matches the VA of portion of the demap address and the context matches the specified context will be demapped.

Demap causes the associated tag access register to be updated; see Section 12.10.5, *I-/D-TLB Tag Access Registers*, on page 140.

Demap is initiated by an STXA with ASI = 57_{16} for I-MMU demap or $5F_{16}$ for D-MMU demap. FIGURE 12-17 shows the Demap format; TABLE 12-28 describes the fields.

1. Demap Page may in fact remove more than one TLB entry for the error case where multiple TLB entries match the virtual page number and real bit. For this multiple match error case, Demap Page will remove all matching TLB entries.

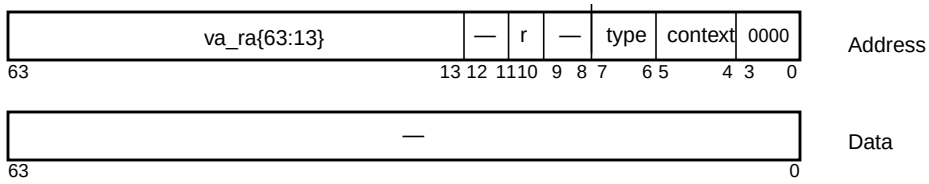


FIGURE 12-17 MMU Demap Operation Format

TABLE 12-28 Field Description for MMU Demap Operation Format

Bit	Field	Description
63:13	va_ra	The virtual page number of the TTE to be removed from the TLB; This field is not used by the MMU for the Demap Context, Demap All, or Demap All Pages operations. **NOTE**: The virtual address for demap is *not* checked for out-of-range violations; instead, va{63:48} is ignored.
10	r	Valid for Demap Page and Demap All Pages only, selects between demapping real translation(s) (r = 1) or virtual translation(s) (r = 0).
7:6	type	The type of demap operation, as described in TABLE 12-29.
5:4	context	Context ID: Context register selection, as described in TABLE 12-30; Use of the reserved value causes the demap to be ignored for demap page and demap context, but is a valid value for a Demap All Pages or Demap All operation.

TABLE 12-29 MMU Demap Operation Type Field Description

Type Field	Demap Operation
00	Demap Page
01	Demap Context
10	Demap All
11	Demap All Pages

TABLE 12-30 MMU Demap Operation Context Field Description

Context ID Field	Context Used in Demap
00	Primary 0
01	Secondary 0
10	Nucleus
11	*Reserved*

Note | Address bits 12:11, 9:8 and 3 are ignored during a demap and may be any value.

A demap operation does not invalidate the TSB in memory. It is the responsibility of the software to modify the appropriate TTEs in the TSB before initiating any Demap operation.

The demap operation produces no output.

12.11.2 I-/D-Demap Page (type = 0)

Demap Page removes the TTE from the specified TLB matching the specified virtual page number, real bit, partition identifier register, and context register.

Virtual page offset bits {15:13}, {21:13}, and {27:13}, for 64-Kbyte, 4-Mbyte, and 256-Mbyte page TLB entries, respectively, are stored in the TLB, but are always set to zero and do not participate in the match for that entry. This is the same condition as for a translation match.

Note | For the IMMU, the Demap Page operation does not support the Secondary Context encoding, and using it will cause the demap to be ignored.

12.11.3 I-/D-Demap Context (type = 1)

Demap Context removes all TTEs from the specified TLB having the specified context, a real bit of 0, and matching the partition identifier register.

Note | For the IMMU, the Demap Context operation does not support the Secondary Context encoding, and using it will cause the demap to be ignored.

12.11.4 I-/D-Demap All (type = 2)

Demap All removes all TTEs from the specified TLB matching the partition identifier register.

12.11.5 I-/D-Demap All Pages (type = 3)

Demap Real removes all TTEs from the specified TLB matching the specified real bit and the partition identifier register.

12.12 TLB Hardware

12.12.1 TLB Operations

The TLB supports exactly one of the following operations per clock cycle:

- **Translation.** The TLB receives a virtual address or real address, a partition identifier and context identifier as input and produces a physical address and page attributes as output.
- **Demap operation.** The TLB receives a virtual address and a context identifier as input and sets the Valid bit to zero for any entry matching the demap page or demap context criteria. This operation produces no output.
- **Read operation**. The TLB reads either the CAM or RAM portion of the specified entry. (Since the TLB entry is greater than 64 bits, the CAM and RAM portions must be returned in separate reads. See *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147 for details.)
- **Write operation**. The TLB simultaneously writes the CAM and RAM portion of the specified entry, or the entry given by the replacement policy described in Section 12.12.2.
- **No operation.** The TLB performs no operation.

12.12.2 TLB Replacement Policy

UltraSPARC T2 uses a used bit scheme to generate a replacement index. Each TLB entry has an associated valid and used bit. An entry's used bit is set on each TLB translation hit and also on the write of an entry. When setting the used bit for a translation or TLB write would result in all used bits being set, the used bits for all TLB entries are cleared instead.

On an automatic write to the TLB initiated through an ASI store to the TLB Data-In register, the TLB replaces the first invalid or unused entry.

CHAPTER 13

Clocks, Reset, RED_state, and Initialization

13.1 Clock Unit

The clock unit block contains the control registers for chipwide clocking.

UltraSPARC T2 has three synchronous clock domains and two asynchronous clock domains. The synchronous clock domains consist of:

- CMP (physical processors, crossbar and L2 cache) clock domain (target 1.4 GHz)
- IO clock domain (target 350 MHz)
- Memory (DR) clock domain (target 333 MHz)

In the synchronous clock domains, all the clocks are derived from the same reference clock. There is one PLL to generate CMP, IO, and memory clocks.

The two asynchronous clock domains are

- PCI-Express clock domain 250 MHz
- Ethernet MAC clock domain 312.5 MHz

The PCI-Express clock domain derives its clock from a PLL in the Tx SerDes which is driven by an external clock. In mission mode, the external clock is asynchronous to the CMP, I/O, and memory clock domains. The Ethernet MAC also has its own clock domain asynchronous to the rest of the chip. It comes from its SerDes.

Controls for the PLL are found in the PLL Control register, whose format is shown in TABLE 13-1.

TABLE 13-1 PLL Control Register – PLL_CTL ($83\ 0000\ 0000_{16}$)

Bit	Field	Initial Value	WMR Protected	R/W	Description
63:37	—	0	—	RO	*Reserved.*
36	pll_clamp_ fltr	0	Yes	RW	PLL Clamp Filter Setting
35:34	st_delay_dr	0_{16}	Yes	RW	DR Stretch Delay Setting (40 ps intervals). [00, 01, 10, 11] → [40, 80, 120, 160] ps
33	pll_char_in	0	Yes	RW	PLL characterization test input.
32	change	1	Yes	RW	If 1, change frequency on next warm reset.
31:30	align_shift	0	Yes	RW	Shift align detect point by [-1:1] CMP cycle. Affects dr_sync pulse generation. All other sync pulses unchanged. 00 : No shift, 01 : +1 cycle, 10 : –1 cycle, 11 : No shift.
29	serdes_dtm2	0	Yes	RW	Mode 2 - IO/IO2x set to DR rate; used for observing debug data on MIO at (up to) CMP rate with DR sync en
28	serdes_dtm1	0	Yes	RW	Mode 1 - IO/IO2x set to DR rate; used for observing PEU TX data and MCU TX CRC bits on MIO at DR rate
27:26	st_delay_ cmp	0	Yes	RW	CMP Stretch delay setting (40 ps intervals). [00, 01, 10, 11]→ [40, 80, 120, 160] ps
25	st_phase_hi	0	Yes	RW	If 1, stretch high phase of clock. If 0, stretch low phase of clock.
24:18	pll_div4	8	Yes	RW	PLL VCO divisor (D4) for DR.
17:12	pll_div3	1	Yes	RW	PLL VCO divisor (D3) for CMP.
11:6	pll_div2	7	Yes	RW	PLL feedback divisor (D2).
5:0	pll_div1	1	Yes	RW	PLL prescalar (D1).

CAUTION! The values of pll_div1, pll_div2, pll_div3, and pll_div4 in the PLL_CTL register are interdependent and must be changed together in a coherent fashion. Illegal values exist that will inhibit correct functional operation. In addition, valid values should not be reverse engineered by experimentation. Values exist that may work at some process, voltage, and temperature points, but do not allow sufficient margin for correct electrical operation across PVT combinations.

The recommended procedure for doing a frequency change with warm reset is listed below:

1. Write the PLL_CTL register to the values needed for the new frequency point. This write should have the change bit set to a 1 and must be a 64-bit write (doubleword store).
2. Generate a warm reset.

The PLL is programmed through a combination of registers (csr fields), direct chip-level pin control and combinational logic. External pin-level control is applicable typically in test mode.

Dividers D1, D2, and D3 perform integer division. D4 has fractional divide capability in discrete increments of 0.5 by using both phases of the VCO clock. The divider configurations allow cmp_pll_clk to run at different multiples of pll_sys_clk, but dr_pll_clk is always twice as fast as pll_sys_clk. The DR clock output may not have 50/50 duty cycle, but should be within 10%.This is not an issue within UltraSPARC T2 since there is no operation on the low phase.

The divider values are summarized in the table below with information on both effective and actual bits.

TABLE 13-2 PLL Divider Programming for Mission Mode

DIV	Bits	(Effective) Valid Range	Binary Encoded Values	Comments
D1	6	2	00_0001	Binary value = Effective value – 1
D2	6	8–21	00_0111 – 01_0100	Binary value = Effective value – 1
D3	6	2	00_0001	Binary value = Effective value – 1
D4	7	4.0–10.5	00_0100_0 – 00_1010_1	Binary value {6:1} = Effective value; bit 0 = 0 for integer effective, and 1 for effective x.5

Even though all four dividers can be programmed via CSR writes, there is a subset of values that are valid. D3, for example, needs to be set to divide by 2. Putting a divide by 3 or higher will result in a non 50/50 duty cycle CMP clock. dr_pll_clk may not be produced correctly since it uses both phases of the VCO clock. Acceptable values for normal operating or mission mode with corresponding clock frequencies are given in Tables13-3 and 13-4.

The clock frequency multiplication equations with respect to the frequency (fsys) of the sys_clk input pin are shown.

fvco ← (D2 × D3 ÷ D1) fsys

fcmp ← (1 /D3) fvco ← (D2 ÷ D1) fsys

fdr ← (1 /D4) fvco ← (D2 × D3) ÷ (D1 × D4) fsys

fio ← 1/4 fcmp ← (D2 ÷ 4D1) fsys

fio2x ← 1/2 fcmp ← (D2 ÷ 2D1) fsys

The first row in any of the three sets in the table below holds the default divider ratio during power-on-reset. The rows in blue (14 and 10) of the two sets refer to the targeted operating frequencies. Red sections are beyond the scope of expected operation, even though within UltraSPARC T2 there is no check for these configurations.

TABLE 13-3 Div Ratios for sys_clk = 133.33 MHz

No	Sys_clk (MHz)	Effec-tive D1	Effec-tive D2	Effec-tive D3	Effec-tive D4	D2 * D3	vco (MHz)	cmp_ clk (MHz)	io_ clk (MHz)	io2x_ clk (MHz)	dr_ clk (MHz)	cmp : dr ratio
1	133.33	2	8	2	4	16	1066.67	533.33	133.33	266.67	266.67	2.0
2	133.33	2	9	2	4.5	18	1200	600	150	300	266.67	2.25
3	133.33	2	10	2	5.0	20	1333.33	666.67	166.67	333.33	266.67	2.5
4	133.33	2	11	2	5.5	22	1466.67	733.33	183.33	367.67	266.67	2.75
5	133.33	2	12	2	6.0	24	1600	800	200	400	266.67	3.00
6	133.33	2	13	2	6.5	26	1733.33	866.67	216.67	433.33	266.67	3.25
7	133.33	2	14	2	7.0	28	1866.67	933.33	233.33	466.67	266.67	3.5
8	133.33	2	15	2	7.5	30	2000	1000	250	500	266.67	3.75
9	133.33	2	16	2	8.0	32	2133.33	1066.67	266.67	533.33	266.67	4.0
10	133.33	2	17	2	8.5	34	2266.67	1133.33	283.33	566.67	266.67	4.25
11	133.33	2	18	2	9.0	36	2400	1200	300	600	266.67	4.5
12	133.33	2	19	2	9.5	38	2533.33	1266.67	316.67	633.33	266.67	4.75
13	133.33	2	20	2	10.0	40	2666.67	1333.33	333.33	666.67	266.67	5.0
14	133.33	2	21	2	10.5	42	2800	1400	350	700	266.67	5.25

TABLE 13-4 Div Ratios for sys_clk = 166.67 MHz

No	Sys_clk (MHz)	Effec-tive D1	Effec-tive D2	Effec-tive D3	Effec-tive D4	D2 * D3	vco (MHz)	cmp_clk (MHz)	io_ clk (MHz)	io2x_clk (MHz)	dr_ clk (MHz)	cmp : dr ratio
1	166.67	2	8	2	4	16	1333.33	666.67	166.67	333.33	333.33	2.0
2	166.67	2	9	2	4.5	18	1500	750	187.5	375	333.33	2.25
3	166.67	2	10	2	5.0	20	1666.67	833.33	208.33	416.67	333.33	2.5
4	166.67	2	11	2	5.5	22	1833.33	916.67	229.17	458.33	333.33	2.75
5	166.67	2	12	2	6.0	24	2000	1000	250	500	333.33	3.0
6	166.67	2	13	2	6.5	26	2166.67	1083.33	270.83	541.67	333.33	3.25
7	166.67	2	14	2	7.0	28	2333.33	1166.67	291.67	583.33	333.33	3.5
8	166.67	2	15	2	7.5	30	2500	1250	312.5	625	333.33	3.75
9	166.67	2	16	2	8.0	32	2666.67	1333.33	333.33	666.67	333.33	4.0
10	166.67	2	17	2	8.5	34	2833.33	1416.67	354.17	708.33	333.33	4.25

TABLE 13-4 Div Ratios for sys_clk = 166.67 MHz

No	Sys_clk (MHz)	Effec-tive D1	Effec-tive D2	Effec-tive D3	Effec-tive D4	D2 * D3	vco (MHz)	cmp_clk (MHz)	io_ clk (MHz)	io2x_clk (MHz)	dr_ clk (MHz)	cmp : dr ratio
11	166.67	2	18	2	9.0	36	3000	1500	375	750	333.33	4.5
12	166.67	2	19	2	9.5	38	3166.67	1583.33	395.83	791.67	333.33	4.75
13	166.67	2	20	2	10.0	40	3333.33	1666.67	416.67	833.33	333.33	5.0
14	166.67	2	21	2	10.5	42	3500	1750	437.5	875	333.33	5.25

13.1.1 Other Clock Unit Registers

The clock unit also contains the random number generator registers.

Note | Data is shifted serially, so doing more than 1 read in 64 cycles will not produce truly random data in diagnostic mode (rng_ctl = 1).

TABLE 13-5 RNG_CTL Register (83 -0000 0020_{16})

Bit	Field	Initial Value	WMR Protected	R/W	Description
63:25	*Reserved*	0_{16}	—	R	*Reserved*
24:9	rng_wait_ cnt	$003E_{16}$	No	RW	Minimum wait time before successive rng data is sent.
8	rng_bypass	0_{16}	No	RW	Controls VCO voltage source. 0 = sets noise cell VCO control voltage = output of feedback amplifier. 1 sets noise cell VCO control voltage = output of bias generator
7:6	rng_vcoctrl_sel	0_{16}	No	RW	Pmos diode D/A setting bus. Controls VCO rate for each noise cell.
5:4	rng_anlg_ sel	0_{16}	No	RW	Analog mux select for characterization.
3	rng_ctl4	1_{16}	No	RW	Enables using LFSR or plain shift register. Set to LFSR mode by default.
2	rng_ctl3	1_{16}	No	RW	Control for using noise cell 3.
1	rng_ctl2	1_{16}	No	RW	Control for using noise cell 2.
0	rng_ctl1	1_{16}	No	RW	Control for using noise cell 1.

TABLE 13-6 RNG_DATA Register ($83\ 0000\ 0030_{16}$)

Bit	Field	Initial Value	WMR Protected	R/W	Description
63:0	rng_data	X	—	R	Random-number-generator data.

The random number generator (rng) generates random numbers from three noise cells. There is one rng block and one LFSR (Linear Feedback Shift Register) to be shared among the eight processor cores. Only one of the cells may be active at a time, all three may be active, or none of them may be active. Any other combination defaults to selecting all three noise cells. The following encoding applies:

TABLE 13-7 Encoding for Noise Cell Selection

rng_ctl3	rng_ctl3	rng_ctl3	Effect
0	0	0	Deselect all noise cells (feeds 0 into LFSR)
0	0	1	Select noise cell 1
0	1	0	Select noise cell 2
1	0	0	Select noise cell 3
011,101,110,111			Select all 3 noise cells

The raw generators will serially output 1 data bit into a 64-bit register. Under functional mode, the register generates data by implementing the CRC polynomial.

P(x) = x64 + x61 + x57 + x56 + x52 + x51 + x50 + x48 + x47 + x46 + x43 + x42 + x41 + x39 + x38 + x37 + x35 + x32 + x28 + x25 + x22 + x21 + x17 + x15 + x13 + x12 + x11 + x7 + x5 + x + 1

After each read request, it is important to not maintain any correlation with the past generated values, so the LFSR will be flushed after every read acknowledge. The register will be flushed with a non-zero state $FFFF\_FFFF\_FFFF\_FFFF_{16}$. Also, multiple requests for rng_data are automatically separated by $n + 2$ cycles, where n can be programmed by writing to the 16-bit field rng_wait_cnt in the RNG_CTL register.

In diagnostic mode (ctl4 = 0), the LFSR acts as a simple shift register capturing the noise cell output directly, determined independently by ctl1, ctl2, and ctl3 as per encoding. The additional constraint in this mode is that successive read requests for the rng_data will be delayed by 64 iol2clk cycles. Also, flushing the LFSR after every read will be disabled in this mode.

The nominal frequency of the oscillator in each noise cell that generates a serial output can also be set independently by programming the rng_vcoctrl_sel{1:0} field. There are four settings that correspond to four different frequencies; however, each

cell must be programmed one at a time. As an example, consider the following configuration: noise cell1 → 00 setting, cell2 → 10 setting, cell 3 → 01 setting, and observe all 3 cells. One would proceed as follows:

1. Set CTL3,CTL2,CTL1 = 001 and set RNG_VCO_CTRL = 00

2. Set CTL3,CTL2,CTL1 = 010 and set RNG_VCO_CTRL = 10

3. Set CTL3,CTL2,CTL1 = 100 and set RNG_VCO_CTRL = 01

13.2 Reset Unit

13.2.1 Reset Generation

The Reset Generation register, shown in TABLE 13-8, allows software to generate an external (XIR) resets to all processors specified in the `ASI_XIR_STEERING` register or a chipwide warm reset (WMR) or debug reset (DBR).

TABLE 13-8 Reset Generation Register – RESET_GEN ($89\ 0000\ 0808_{16}$)

Bit	Field	Initial Value	R/W	Description
63:3	—0	0	RO	*Reserved*
3	dbr_gen	0	RW	Set to 1 to generate a DBR. Value is automatically cleared once the DBR is complete.
2	—	0	RO	*Reserved* (was por_gen on Fire).
1	xir_gen	0	RW	Set to 1 to generate an XIR. Value is automatically cleared once the XIR is complete.
0	wmr_gen	0	RW	Set to 1 to generate a WMR. Value is automatically cleared once the WMR is complete.

Programming Note Software may only write a 1 to one of the reset generation bits at a time. Behavior of UltraSPARC T2 is undefined if software writes 1's to multiple reset generation bits.

13.2.2 Reset Source

The Reset Source register, shown in TABLE 13-9 allows software to identify the source of a reset. The bits in this register are write-one to clear.

TABLE 13-9 Reset Source Register – RESET_SOURCE ($89\ 0000\ 0818_{16}$)

Bit	Field	Initial Value	R/W	Description
63:16	—0	0	RO	*Reserved*
15	l2t7_fatal	0	RW1C	Bank 7 of the L2 cache detected a fatal error.
14	l2t6_fatal	0	RW1C	Bank 6 of the L2 cache detected a fatal error.
13	l2t5_fatal	0	RW1C	Bank 5 of the L2 cache detected a fatal error.
12	l2t4_fatal	0	RW1C	Bank 4 of the L2 cache detected a fatal error.
11	l2t3_fatal	0	RW1C	Bank 3 of the L2 cache detected a fatal error.
10	l2t2_fatal	0	RW1C	Bank 2 of the L2 cache detected a fatal error.
9	l2t1_fatal	0	RW1C	Bank 1 of the L2 cache detected a fatal error.
8	l2t0_fatal	0	RW1C	Bank 0 of the L2 cache detected a fatal error.
7	ncu_fatal	0	RW1C	The NCU or a block interfacing to it detected a fatal error.
6	pb_xir	0	RW1C	The user asserted the BUTTON_XIR_ input pin.
5	pb_rst	0	RW1C	The user asserted the PB_RST_L input pin.
4	pwron_rst	1	RW1C	The system processor asserted the PWRON_RST_L input pin.
3	dbr_gen	0	RW1C	Software wrote a 1 to the dbr_gen field of the RESET_GEN register.
2	—	0	RO	*Reserved* (In Fire was, software wrote a 1 to the por_gen field of the RESET_GEN register).
1	xir_gen	0	RW1C	Software wrote a 1 to the xir_gen field of the RESET_GEN register.
0	wmr_gen	0	R/W1C	Software wrote a 1 to the wmr_gen field of the RESET_GEN register.

13.2.3 Reset Fatal Error Enable

The Reset Fatal Error Enable register, shown in TABLE 13-10 allows software to control whether L2 fatal errors generate a warm reset.

TABLE 13-10 Reset Fatal Error Enable Register – RESET_FEE ($89\ 0000\ 0820_{16}$)

Bit	Field	Initial Value	R/W	Description
63:16	—0	0	RO	*Reserved*
15	l2t7_fee	0	RW	Allow Bank 7 of the L2 cache to generate a fatal error.
14	l2t6_fee	0	RW	Allow Bank 6 of the L2 cache to generate a fatal error.
13	l2t5_fee	0	RW	Allow Bank 5 of the L2 cache to generate a fatal error.
12	l2t4_fee	0	RW	Allow Bank 4 of the L2 cache to generate a fatal error.
11	l2t3_fee	0	RW	Allow Bank 3 of the L2 cache to generate a fatal error.
10	l2t2_fee	0	RW	Allow Bank 2 of the L2 cache to generate a fatal error.

TABLE 13-10 Reset Fatal Error Enable Register – RESET_FEE ($89\ 0000\ 0820_{16}$)

Bit	Field	Initial Value	R/W	Description
9	l2t1_fee	0	RW	Allow Bank 1 of the L2 cache to generate a fatal error.
8	l2t0_fee	0	RW	Allow Bank 0 of the L2 cache to generate a fatal error.
7:0	—	0	RO	*Reserved*

13.2.4 Subsystem Reset

The Subsystem Reset Generation register, shown in TABLE 13-11, allows software to reset selected I/O subsystems.

TABLE 13-11 Subsystem Reset Generation Register – SSYS_RESET ($89\ 0000\ 0838_{16}$)

Bit	Field	Initial Value	R/W	Description
63:7	—0	0	RO	*Reserved*
6	mac_protect	0	RW	Set to 1 to suppress the reset of the NIU on warm reset.
5	mcu_selfrsh	0	RW	Set to 1 to have the MCUs put the DRAM into self-refresh.
4	—	0	RW	*Reserved* (was set to 1 to protect the FBDIMM interfaces of each MCU from being reset by either warm reset or debug reset).
3:2	—2	0	RO	*Reserved*
1	dmu_peu	0	RW	Set to 1 to generate a warm reset to the PCI Express Unit (PIU), both ingress and egress. Value is automatically cleared once the WMR is complete.
0	niu	0	RW	Set to 1 to generate a warm reset to the Ethernet subsystem, both ingress and egress. Value is automatically cleared once the WMR is complete.

13.2.5 Reset Status

The chip Reset Status register, shown in TABLE 13-12, is maintained for all chipwide reset and power management commands. The reset source bits in this register are writable to allow software to clear them after the chip reset sequence is complete, in order for virtual processor warm resets to be distinguished from chip resets. Hardware will copy the current reset status into a shadow status whenever a warm reset occurs.

TABLE 13-12 Reset Status Register – RESET_STAT $(89\ 0000\ 0810)_{16}$

Bit	Field	Initial Value	R/W	Description
63:12	—0	0	RO	*Reserved*
11	freq_s	0	RO	Shadow status of FREQ
10	por_s	0	RO	Shadow status of POR

TABLE 13-12 Reset Status Register – RESET_STAT $(89\ 0000\ 0810)_{16}$

Bit	Field	Initial Value	R/W	Description
9	wmr_s	0	RO	Shadow status of WMR
8:5	—	0	RO	*Reserved*
4	—2	0	RW	*Reserved*
3	freq	0	RW	Set to 1 if the reset is a warm reset that changed frequency.
2	por	1	RW	Set to 1 if the reset is from PWRON_RST_L pin.
1	wmr	0	RW	Set to 1 if the reset is from the PB_RST_L pin.
0	—	0	RO	*Reserved*

13.2.6 Lock Time

The Lock Time register determines the length of time the Reset Unit in UltraSPARC T2 waits for all the PLLs in UltraSPARC T2 to lock. The initial value is an estimated time only that software can reprogram during the warm reset sequence. Moreover, software can enable the pre-WMR boot code to perform warm reset with the same PLL configuration register values, obviating the need to wait for the PLLs to relock.

TABLE 13-13 Lock Time Register – LOCK_TIME $(89\ 0000\ 0870_{16})$

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	lock_time	1400_{16}	RW	The length of time the Reset Unit in UltraSPARC T2 waits for all the PLLs in UltraSPARC T2 to lock.

13.2.7 Propagation Time

The Propagation Time register indicates how long it takes for the longest scan chain to flush. The register initializes to the estimated longest time needed (assuming the highest planned reference clock frequency), that software can reprogram during warm reset sequence.

TABLE 13-14 Propagation Time Register – PROP_TIME $(89\ 0000\ 0880_{16})$

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	prop_time	$C00_{16}$	RW	Time taken for longest scan chain to flush.

13.2.8 NIU Time

The NIU Time register indicates how long it takes for initial values to shift throughout the NIU.The initial value is an estimate only that software can reprogram during warm reset sequence.

TABLE 13-15 NIU Time Register – NIU_TIME ($89\ 0000\ 0890_{16}$)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	prop_time	640_{16}	RW	Time taken by initial values to shift throughout the NIU.

13.3 Reset Overview

A reset is anything that causes an entry to `RED_state`. Two classes of resets exist: chipwide and virtual processor.

Chipwide resets are power-on reset (POR), warm reset (WMR), and debug reset (DBR). POR affects all subsystems in UltraSPARC T2, while WMR affects all subsystems other than optionally NIU (through SSYS_RESET.mac_protect bit) and DBR affects all subsystems except NIU and PIU. Chipwide resets are generated from the PWRON_RST_L input pin (POR), PB_RST_L pin (WMR), by software writing a 1 to the wmr_gen field of the RESET_GEN register (WMR) or to the dbr_gen field of the RESET_GEN register (DBR) and from fatal errors in the processor (WMR).

Virtual processor resets are XIR, WDR, SIR and are generated by the BUTTON_XIR_ pin (XIR), software resets (SIR), software writing a 1 to the xir_gen field of the RESET_GEN register (XIR), and error conditions (WDR), and only affect the operation of a single virtual processor (or for the case of XIR, only the virtual processors specified in XIR_STEERING as described in *ASI_XIR_STEERING* on page 189). In addition to forcing entry to `RED_state`, various resets cause different effects in initializing processor state, as discussed in the following sections. Reset priorities from highest to lowest are: POR, WMR, DBR, XIR, WDR, SIR. Resets are not maskable (that is, resets ignore PSTATE.ie).

Programming Note Chipwide resets cause the virtual processors to enter `RED_state` and initialize some, but not all, of the processor state. Significant programming effort is required to take the UltraSPARC T2 chip from a chipwide reset to the point where the operating system can be loaded.

13.4 Chipwide Resets

Chipwide resets affect all virtual processors in a chip, as well as all I/O, cache, and DRAM subsystems and are categorized as power-on or warm reset. Power-on reset is used when the chip power and clock inputs are outside their operating specifications. Warm reset is used when the power and clock inputs are stable. Warm reset is typically used to modify clock frequencies or ratios, or to reinitialize the chip after an unrecoverable hardware or software failure. Warm reset resets all subsystems in UltraSPARC T2 other than optionally NIU (through the SSYS_RESET.mac_protect bit).

13.4.1 Power-on Reset (POR)

A power-on reset occurs when the PWRON_RST_L pin is asserted and then deasserted. The PWRON_RST_L pin must be asserted until 1 clock after the CPU voltages and input clocks reach their operating specifications. When the PWRON_RST_L pin is asserted, all other resets and traps are ignored. Power-on reset has a trap type of 001_{16} at physical address offset 20_{16}. Since POR and warm reset share the same trap type and trap vector, the RSET_STAT register described in Section 13.2.5 has separate POR and warm reset bits to allow software to distinguish between POR and warm resets. All pending transactions are cancelled. Strand 0 of the first available physical core begins executing at the `RED_State_Trap_Vector` base plus POR offset, while the remaining strands start out inactive. BIST testing may optionally be initiated by software as part of the chip initialization sequence.

After a power-on reset, software must initialize values specified as *unknown* in *Machine State After Reset and in RED_State* on page 171.

Note | Each unknown register must be initialized before it is used. Failure to initialize registers or states properly before use may result in unpredictable or incorrect results.

13.4.2 Warm Reset (WMR)

A warm reset occurs when the PB_RST_L pin is asserted and then deasserted, when software writes a 1 to the wmr_gen field of the RESET_GEN register, or when a fatal error is detected in the processor. When a warm reset is received, all other resets and traps except POR are ignored. Warm reset has the same trap type and vector as power-on reset: a trap type of 001_{16} at physical address offset 20_{16}. Software can distinguish between POR and the various sources of warm reset by checking the RSET_STAT register. The memory controller places DRAM in self-refresh mode prior to warm reset only if SSYS_RESET.mcu_selfrsh bit is set by software prior to

the warm reset. Otherwise, on a warm reset, the memory controller does not put the DRAM in self-refresh.Warm reset can be programmed to do BIST testing. After warm reset, strand 0 of the first available physical core begins executing at the RED_State_Trap_Vector base plus POR offset, while the remaining strands start out inactive.

After a warm reset, software must initialize values specified as unknown in *Machine State After Reset and in RED_State* on page 171. If there was a clean shutdown, the primary instruction, primary data, L2 caches, and main memory are still valid. Otherwise, I-cache tags, D-cache tags, and L2 cache tags should be initialized before enabling the caches. The iTLB and dTLB also must be initialized before enabling memory management.

Note that if a warm reset is received without software first placing the chip in a quiescent state, the hardware will still maintain the state of the primary instruction, primary data, L2 caches, main memory, and all error registers/logs. However, the caches and main memory may no longer be completely coherent after the warm reset, because any transactions in flight when the warm reset was received will have been lost. In particular, dirty lines in the process of being written back to main memory may have been dropped.

13.4.3 Debug Reset (DBR)

A debug reset occurs when software writes a 1 to the dbr_gen field of the RESET_GEN register. DBR behaves the same as warm reset, except that the NIU and PIUs are not reset.

13.5 Virtual Processor Resets

Virtual processors can receive reset traps. Virtual processor reset traps do not set any bits in the RSET_STAT register.

13.5.1 Externally Initiated Reset (XIR)

An externally initiated reset can be generated either from the BUTTON_XIR_ pin or by writing a 1 to the xir_gen field of the RESET_GEN register. When either of these events occurs, an XIR reset trap is sent to all virtual processors specified in the XIR_STEERING register. This trap causes a SPARC V9 XIR, which has a trap type of 003_{16} at physical address offset 60_{16}. It has higher priority than all other virtual processor core resets. XIR is used for system debug.

13.5.2 Watchdog Reset (WDR) and error_state

A SPARC V9 WDR is generated when a virtual processor encounters a trap when TL = *MAXTL*, it passes through error_state and signals itself internally to take a WDR trap. Window traps that cause watchdog traps still update CWP if they would have done so with no watchdog trap being generated.

13.5.3 Software-Initiated Reset (SIR)

A SPARC V9 SIR interrupt can be generated on a virtual processor by issuing a SIR instruction while operating in hyperprivileged mode. This virtual processor reset has a trap type of 004_{16} at physical address offset 80_{16}.

13.6 RED_state

RED_state is an acronym for Reset, Error, and Debug State. RED_state is described in the UltraSPARC Architecture 2007 specification.

13.7 RED_state Trap Vector

When a SPARC V9 virtual processor processes a reset or trap that enters RED_state, it takes a trap at an offset relative to the RED_state_trap_vector base address (*RSTVADDR*). The trap offset depends on the type of red mode trap and takes the values:

- POR, WMR, or DBR 20_{16}
- XIR 60_{16}
- WDR 40_{16}
- SIR 80_{16}
- other $A0_{16}$

In UltraSPARC T2 the RSTV base address is FFFF FFFF F000 0000_{16} if ASI_RST_VEC_MASK.vec_mask = 0. If ASI_RST_VEC_MASK.vec_mask = 1, RSTV base address is 0000 0000 0000 0000_{16}. See *ASI_RST_VEC_MASK* on page 448 for more details on the ASI_RST_VEC_MASK register.

13.8 Machine State After Reset and in RED_State

TABLE 13-16 shows CPU state created as a result of any reset, or after entering RED_state.

TABLE 13-16 CPU State After Reset and in RED_state *(1 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
Integer Registers		0	Unchanged				
Floating Point Registers		0	Unchanged				
iTLB/dTLB	Mappings	All invalid	Unchanged				
PSTATE	tct	0 (Trap on control transfer)					
	mm	0 (TSO)					
	red	0 (RED_state bit is in HPSTATE register)					
	pef	1 (FPU on)					
	am	0 (Full 64-bit addresses)					
	priv	1					
	ie	0 (Disable interrupts)					
	ag	0 (Alternate globals always 0)					
	cle	0 (Current not little endian)					
	tle	0 (Trap not little endian)	Unchanged				
	ig	0 (Interrupt globals always 0)					
	mg	0 (MMU globals always 0)					
HPSTATE	ibe	0 (Instruction breakpoint disabled)					
	red	1 (RED_state)					
	hpriv	1 (Hyperprivileged mode)					
	tlz	0 (tlz traps disabled)					
TBA{63:15}		0	Unchanged				
HTBA{63:15}		0	Unchanged				
Y		0	Unchanged				
PIL		0	Unchanged				
CWP		0	Unchanged (except for window traps)				
PC		RSTV|20_{16}	RSTV|20_{16}	RSTV|40_{16}	RSTV|60_{16}	RSTV80_{16}	RSTV|$A0_{16}$
NPC		RSTV|24_{16}	RSTV|24_{16}	RSTV|44_{16}	RSTV|64_{16}	RSTV|$_{16}84$	RSTV|$A4_{16}$
TT[TL]		1	1	2 or Trap type	3	4	Trap type
TPC[TL]			PC				

TABLE 13-16 CPU State After Reset and in `RED_state` *(2 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
TNPC[TL]		0	NPC				
Store Buffer		Empty		Unchanged			Empty
CCR		0	Unchanged				
ASI		0	Unchanged				
TL		*MAXTL*		**min**(TL+1,*MAXTL*)			
GL		*MAXGL*		**min**(GL+1,*MAXGL*)			
TSTATE[TL]	GL	0	pre-WMR[1]	GL			
	CCR	0	pre-WMR[1]	CCR			
	ASI	0	pre-WMR[1]	ASI			
	PSTATE	0	0_{16}	PSTATE			
	CWP	0	pre-WMR[1]	CWP			
HTSTATE[TL]	ibe	0	0	HPSTATE.ibe			
	red	0	0	HPSTATE.red			
	hpriv	0	0	HPSTATE.hpriv			
	tlz	0	0	HPSTATE.tlz			
TICK	npt	1	1	Unchanged			
	counter	0	Count	Count			
CANSAVE		6_{16}	Unchanged				
CANRESTORE		0_{16}	Unchanged				
OTHERWIN		0_{16}	Unchanged				
CLEANWIN		7_{16}	Unchanged				
WSTATE	other	0_{16}	Unchanged				
	normal	0_{16}	Unchanged				
VER	manuf	$003E_{16}$					
	impl	0024_{16}					
	mask	Mask-dependent (4 bits major, 4 bits minor)					
	maxtl	6					
	maxgl	3					
	maxwin	7					
FSR	all	0	Unchanged				
FPRS	all	4_{16}	Unchanged				
GSR	all	0	Unchanged				
PERF_CONTROL (PCR)	all	0 (off)	Unchanged				
PIC		0	Unchanged				
TICK_CMPR	int_dis	1	Unchanged				
	tick_cmpr	0	Unchanged				

TABLE 13-16 CPU State After Reset and in RED_state *(3 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
STICK_CMPR	int_dis	1	Unchanged				
	stick_cmpr	0	Unchanged				
HSTICK_CMPR	int_dis	1	Unchanged				
	hstick_cmpr	0	Unchanged				
HINTP		0	Unchanged				
SOFTINT		0	Unchanged				
ASI_SCRATCHPAD_0_REG		0	Unchanged				
ASI_SCRATCHPAD_1_REG		0	Unchanged				
ASI_SCRATCHPAD_2_REG		0	Unchanged				
ASI_SCRATCHPAD_3_REG		0	Unchanged				
ASI_SCRATCHPAD_6_REG		0	Unchanged				
ASI_SCRATCHPAD_7_REG		0	Unchanged				
ASI_PRIMARY_CONTEXT_0		0	Unchanged				
ASI_SECONDARY_CONTEXT_0		0	Unchanged				
ASI_PRIMARY_CONTEXT_1		0	Unchanged				
ASI_SECONDARY_CONTEXT_1		0	Unchanged				
ASI_CPU_MONDO_QUEUE_HEAD		0	Unchanged				
ASI_CPU_MONDO_QUEUE_TAIL		0	Unchanged				
ASI_DEVICE_QUEUE_HEAD		0	Unchanged				
ASI_DEVICE_QUEUE_TAIL		0	Unchanged				
ASI_RES_ERROR_QUEUE_HEAD		0	Unchanged				
ASI_RES_ERROR_QUEUE_TAIL		0	Unchanged				
ASI_NONRES_ERROR_QUEUE_HEAD		0	Unchanged				
ASI_NONRES_ERROR_QUEUE_TAIL		0	Unchanged				
ASI_CORE_AVAILABLE		FFFFFFFFFFFFFFFF$_{16}$ (if all cores available)	Unchanged				
ASI_CORE_ENABLE_STATUS		ASI_CORE_AVAILABLE	ASI_CORE_ENABLE	Unchanged			
ASI_CORE_ENABLE		ASI_CORE_AVAILABLE	Unchanged				
ASI_XIR_STEERING		ASI_CORE_AVAILABLE	ASI_CORE_ENABLE	Unchanged			
ASI_CMT_TICK_ENABLE		0	Unchanged				
ASI_CORE_RUNNING_RW		1_{16} (or lowest enabled strand)		Unchanged			

TABLE 13-16 CPU State After Reset and in RED_state *(4 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
ASI_CORE_RUNNING_STATUS		1_{16} (or lowest enabled strand)		Unchanged			
ASI_INST_MASK_REG		0		Unchanged			
ASI_LSU_DIAG_REG		0		Unchanged			
ASI_ERROR_INJECT_REG		0		Unchanged			
ASI_LSU_CONTROL_REG		0					
ASI_DECR		0		Unchanged			
ASI_RST_VEC_MASK		0	Unchanged				
ASI_DESR		0	Unchanged				
ASI_DFESR		0	Unchanged				
ASI_CERER		0		Unchanged			
ASI_CETER		0		Unchanged			
ASI_clesr		0	Unchanged				
ASI_CLFESR		0	Unchanged				
ASI_SPARC_PWR_MGMT		0		Unchanged			
ASI_HYP_SCRATCHPAD_0_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_1_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_2_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_3_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_4_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_5_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_6_REG		0	Unchanged				
ASI_HYP_SCRATCHPAD_7_REG		0	Unchanged				
ASI_IMMU_TAG_TARGET		0	Unchanged				
ASI_IMMU_SFSR		0	Unchanged				
ASI_IMMU_TAG_ACCESS		0	Unchanged				
ASI_IMMU_VA_WATCHPOINT		0	Unchanged				
ASI_MMU_REAL_RANGE_0		0	Unchanged				
ASI_MMU_REAL_RANGE_1		0	Unchanged				
ASI_MMU_REAL_RANGE_2		0	Unchanged				
ASI_MMU_REAL_RANGE_3		0	Unchanged				
ASI_MMU_PHYSICAL_OFFSET_0		0	Unchanged				
ASI_MMU_PHYSICAL_OFFSET_1		0	Unchanged				
ASI_MMU_PHYSICAL_OFFSET_2		0	Unchanged				
ASI_MMU_PHYSICAL_OFFSET_3		0	Unchanged				
ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_0		0	Unchanged				

TABLE 13-16 CPU State After Reset and in RED_state *(5 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_1		0			Unchanged		
ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_2		0			Unchanged		
ASI_MMU_ZERO_CONTEXT_TSB_CONFIG_3		0			Unchanged		
ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_0		0			Unchanged		
ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_1		0			Unchanged		
ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_2		0			Unchanged		
ASI_MMU_NONZERO_CONTEXT_TSB_CONFIG_3		0			Unchanged		
ASI_MMU_ITSB_PTR_0		0			Unchanged		
ASI_MMU_ITSB_PTR_1		0			Unchanged		
ASI_MMU_ITSB_PTR_2		0			Unchanged		
ASI_MMU_ITSB_PTR_3		0			Unchanged		
ASI_MMU_DTSB_PTR_0		0			Unchanged		
ASI_MMU_DTSB_PTR_1		0			Unchanged		
ASI_MMU_DTSB_PTR_2		0			Unchanged		
ASI_MMU_DTSB_PTR_3		0			Unchanged		
ASI_PENDING_TABLEWALK_CONTROL		0			Unchanged		
ASI_PENDING_TABLEWALK_STATUS		0			Unchanged		
ASI_DMMU_TAG_TARGET		0			Unchanged		
ASI_DMMU_SFSR		0			Unchanged		
ASI_DMMU_SFAR		0			Unchanged		
ASI_DMMU_TAG_ACCESS		0			Unchanged		
ASI_DMMU_WATCHPOINT		0			Unchanged		
ASI_HWTW_CONFIG		0			Unchanged		
ASI_PARTITION_ID		0			Unchanged		
ASI_CMT_CORE_INTR_ID		COREID					
ASI_CMT_CORE_INTR_ID		$7003F0000_{16}$ | COREID					
ASI_INTR_RECEIVE		0			Unchanged		
PLL_CTL		$1000204E1_{16}$			Unchanged		
I/D cache tags		All invalid		Unchanged if BISI not run, else invalid			
L2 tags and data		Unknown		Unchanged if BISI not run, else invalid			

TABLE 13-16 CPU State After Reset and in `RED_state` *(6 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
L2 directory		All invalid	Unchanged if BISI not run, else invalid				
L2 Error En Reg	all	0 (reporting disabled)		Unchanged			
L2 Error Status Reg	synd	Unknown	Unchanged				
	Other fields	0	Unchanged				
L2 Error Address		Unknown	Unchanged				
L2 NotData Error	vcid	Unknown	Unchanged				
	rw	Unknown	Unchanged				
	address	Unknown	Unchanged				
	Other fields	0	Unchanged				
L2 Error Inject		0	Unchanged				
L2 Mask Reg		0	Unchanged				
L2 Address Compare Reg		0	Unchanged				
L2 Bank Available		FF_{16} (if all banks available)	Unchanged				
L2 Bank Enable		FF_{16} (if all banks available)	Unchanged				
L2 Bank Enable Status		$F00_{16}$ (if all banks available)	$F0F_{16}$ (if all banks available)	Unchanged			
L2 Index Hash Enable		0_{16}	Unchanged				
L2 Control Reg		1_{16}	Unchanged				
DRAM Refresh Counter		0	Unchanged				
DRAM Error Status Register	synd	Unknown	Unchanged				
	Other fields	0	Unchanged				
DRAM Error Address		Unknown	Unchanged				
DRAM Error Inject		0	Unchanged				
DRAM Error Counter		0	Unchanged				
DRAM FBD Error Syndrome		0	Unchanged				
DRAM Error Location		Unknown	Unchanged				
DRAM Debug Enable Trigger	dbg_en	0	Unchanged				
	Other fields	1_{16}	Unchanged				
DRAM CAS Address Width		B_{16}	Unchanged				
DRAM RAS Address Width		F_{16}	Unchanged				
DRAM CAS Latency		3_{16}	Unchanged				
DRAM Scrub Frequency		FFF_{16}	Unchanged				
DRAM Refresh Frequency		514_{16}	Unchanged				
DRAM Open Bank Max		0	Unchanged				

TABLE 13-16 CPU State After Reset and in `RED_state` *(7 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
DRAM Scrub Enable		0			Unchanged		
DRAM Programmable Time Counter		0			Unchanged		
DRAM RAS to RAS Different Banks Delay		0			Unchanged		
DRAM RAS to RAS Same Bank Delay		C_{16}			Unchanged		
DRAM RAS to CAS Delay		3			Unchanged		
DRAM Write to Read CAS Delay		0			Unchanged		
DRAM Read to Write CAS Delay		0			Unchanged		
DRAM Internal Read to Precharge Delay		2_{16}			Unchanged		
DRAM Active to Precharge Delay		9_{16}			Unchanged		
DRAM Precharge Command Period		3_{16}			Unchanged		
DRAM Write Recover Period		3_{16}			Unchanged		
DRAM Autorefresh to Active Period		27_{16}			Unchanged		
DRAM Mode Register Set Command Period		2_{16}			Unchanged		
DRAM Four Activate Window		2_{16}			Unchanged		
DRAM Internal Write to Read Command Delay		2_{16}			Unchanged		
DRAM DIMM Stacked		0			Unchanged		
DRAM Extended Mode (2)		0			Unchanged		
DRAM Extended Mode (1)		18_{16}			Unchanged		
DRAM Extended Mode (3)		0			Unchanged		
DRAM 8 Bank Mode		1_{16}			Unchanged		
DRAM Branch Disabled		0			Unchanged		
DRAM Select Low Order Address Bits		0			Unchanged		
DRAM Single Channel Mode		0			Unchanged		
DRAM DIMMs Present		1_{16}			Unchanged		
DRAM Fail-Over Status		0			Unchanged		
DRAM Fail-Over Mask		0			Unchanged		
FBD Channel State		0			Unchanged		
FBD Fast Reset Flag		0			Unchanged		
FBD Channel Reset		0			Unchanged		
TS1 Southbound to Northbound Mapping		0			Unchanged		

TABLE 13-16 CPU State After Reset and in `RED_state` *(8 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
TS1 Test Parameter		0	Unchanged				
TS3 Failover Configuration		$FFFF_{16}$	Unchanged				
Disable State Period		$3F_{16}$	Unchanged				
Calibrate State Period		0	Unchanged				
Training State Minimum Time		FF_{16}	Unchanged				
Training State Timeout		FF_{16}	Unchanged				
Testing State Timeout		FF_{16}	Unchanged				
Polling State Timeout		FF_{16}	Unchanged				
DRAM Per-Rank CKE		$FFFF_{16}$	Unchanged				
L0s Duration		$2A_{16}$	Unchanged				
Channel Sync Frame Frequency		$2A_{16}$	Unchanged				
SerDes Configuration Bus		0	Unchanged				
SerDes Transmitter and Receiver Differential Pair Inversion		0	Unchanged				
SerDes Test Configuration Bus		$C000_{16}$	Unchanged				
DRAM FBD Injected Error Source		0	Unchanged				
DRAM FBR Count		0	Unchanged				
IMU Error Log Enable Register		$7FFF_{16}$	Unchanged				
IMU Error Status Clear Register		0_{16}	Unchanged				
IMU Error Status Set Register		0_{16}	Unchanged				
IMU RDS Error Log Register		0_{16}	Unchanged if any Primary Error bit in IMU Error Status Clear Register in RDS Group is set				
IMU SCS Error Log Register		0_{16}	Unchanged if any Primary Error bit in IMU Error Status Clear Register in SCS Group is set				
IMU EQS Error Log Register		0_{16}	Unchanged if any Primary Error bit in IMU Error Status Clear Register in EQS Group is set				
MMU Error Log Enable Register		$1FFFFF_{16}$	Unchanged				
MMU Error Status Clear Register		0_{16}	Unchanged				
MMU Translation Fault Address Register		0_{16}	Unchangedonly if any Primary Error bit is set inMMU Error Status Clear Register				
MMU Translation Fault Status Register		0_{16}	Unchangedonly if any Primary Error bit is set inMMU Error Status Clear Register				
MMU TTE Cache Data Registers		0_{16}	Unchanged				
MMU DEV2IOTSB Registers		0_{16}	Unchanged				
MMU IOTSBDESC Registers		0_{16}	Unchanged				
ILU Error Log Enable Register		$F0_{16}$	Unchanged				
ILU Error Status Clear Register		0_{16}	Unchanged				
ILU Error Status Set Register		0_{16}	Unchanged				
DMU ILU Diagnostic Register		0_{16}	Unchanged (only bits 3:2)				

TABLE 13-16 CPU State After Reset and in `RED_state` *(9 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
PEU Control Register		1_{16}	Unchanged				
PEU Ingress Credits Initial Register		$10000200C0_{16}$	Unchanged				
PEU Other Event Log Enable Register		$FFFFFF_{16}$	Unchanged				
PEU Other Event Status Clear Register		0_{16}	Unchanged				
PEU Other Event Status Set Register		0_{16}	Unchanged				
PEU Receive Other Event Header1 Log Register		0_{16}	Unchanged only if any of Ingress completion header error, Memory read capture, Write unsuccessful completion status, Read unsuccessful completion status, Configuration request retry completion status Primary Error bits in PEU Other Event Status Clear Register is set				
PEU Receive Other Event Header2 Log Register		0_{16}	Unchanged only if any of : Ingress completion header error, Memory read capture, Write unsuccessful completion status, Read unsuccessful completion status, Configuration request retry completion status Primary Error bits in PEU Other Event Status Clear Register is set				
PEU Transmit Other Event Header1 Log Register		0_{16}	Unchanged only if any of Completion Timeout Error, Write Unsuccessful Completion Status, Read Unsuccessful Completion Status, Configuration Request Retry Completion Status Primary error bits in PEU Other Event Status Clear Register is set				
PEU Transmit Other Event Header2 Log Register		0_{16}	Unchanged only if any of Completion Timeout Error, Write Unsuccessful Completion Status, Read Unsuccessful Completion Status, Configuration Request Retry Completion Status Primary error bits in PEU Other Event Status Clear Register is set				
PEU Uncorrectable Error Log Enable Register		$17F011_{16}$	Unchanged				
PEU Uncorrectable Error Status Clear Register		0_{16}	Unchanged				
PEU Uncorrectable Error Status Set Register		0_{16}	Unchanged				
PEU Receive Uncorrectable Error Header1 Log Register		0_{16}	Unchangedonly when a primary error bit is set in the PEU Uncorrectable Error Status Clear register				
PEU Receive Uncorrectable Error Header2 Log Register		0_{16}	Unchangedonly when a primary error bit is set in the PEU Uncorrectable Error Status Clear register				
PEU Transmit Uncorrectable Error Header1 Log Register		0_{16}	Unchangedonly when a primary error bit is set in the PEU Uncorrectable Error Status Clear register				
PEU Transmit Uncorrectable Error Header2 Log Register		0_{16}	Unchangedonly when a primary error bit is set in the PEU Uncorrectable Error Status Clear register				
PEU Correctable Error Log Enable Register		$11C1_{16}$	Unchanged				

TABLE 13-16 CPU State After Reset and in `RED_state` *(10 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
PEU Correctable Error Status Clear Register		0_{16}	Unchanged				
PEU Correctable Error Status Set Register		0_{16}	Unchanged				
PEU CXPL/SerDes Revision Register		0_{16}	Unchanged				
PEU CXPL Event/Error Log Enable Register		31:24: F_{16} 17:0: $3FFFF_{16}$	Unchanged				
PEU CXPL Event/Error Status Clear Register		0_{16}	Unchanged				
PEU CXPL Event/Error Status Set Register		0_{16}	Unchanged				
PEU SerDes PLL Control/Status Register		1_{16}	Unchanged				
PEU SerDes Receiver Lane 0–7 Control Register		444_{16}	Unchanged				
PEU SerDes Receiver Lane 0–7 Status Register		0_{16}	bits 3,0 Unchanged bit 1 can change to either 0_2 or 1_2 Please refer to TABLE 21-168 on page 619				
PEU SerDes Transmitter Lane 0–7 Control Register		$1F8_{16}$	Unchanged				
PEU SerDes Transmitter Lane 0–7 Status Register		0_{16}	Unchanged				
PEU SerDes MACRO 0–1 Test Configuration Register		03_{16}	Unchanged				
IBIST Registers		0	Unchanged				
NCU Debug Trigger Enable		0	Unchanged				
SOC DECR		0	Unchanged				
SOC Error Status		0	Unchanged				
SOC Pending Error Status		0	Unchanged				
SOC SII Error Syndrome		0	Unchanged				
SOC NCU Error Syndrome		0	Unchanged				
Debug Port Configuration		0	Unchanged				
IO Quiesce Control	niu_stall	0	Unchanged				
	dmu_stall	0	Unchanged				
	Other fields	Unknown		Unchanged			
Serial Number		Unique Value					
EFUSE Status		FFFF FFFF FFFF $FFFF_{16}$					
OVERLAP_MODE		0	Unchanged				
CLK_SCKSEL		0	Unchanged				

TABLE 13-16 CPU State After Reset and in `RED_state` *(11 of 11)*

Name	Fields	POR	WMR|DBR	WDR	XIR	SIR	RED_State[2]
SSI_LOG		0_{16}			Unchanged		
MBIST_MODE	Bits{3:0}	0			Unchanged		
MBIST_BYPASS	Bits {47:0}	0			Unchanged		
MBIST_RESULT	Bits {1:0}	0			Unchanged		
MBIST_DONE	Bits {47:0}	0			Unchanged		
MBIST_FAIL	Bits {47:0}	0			Unchanged		
LBIST_MODE	Bits {1:0}	0			Unchanged		
LBIST_BYPASS	Bits {7:0}	0			Unchanged		
LBIST_DONE	Bits {7:0}	0			Unchanged		
DCR	Bits {2:0}	0			Unchanged		
TRIGOUT		0			Unchanged		
PEU_TESTCONFIG_EN		0			Unchanged		
CYCLE_COUNTER	Bits {63:0}	0			Unchanged		
CLOCK STOP DELAY counter	Bits{6:0}	0			Unchanged		
MBIST_START		0			Unchanged		
MBIST_ABORT		0			Unchanged		
MBIST_START_WMR		0			Unchanged		
LBIST_START		0			Unchanged		
LOCK TIME		1400_{16}			Unchanged		
PROP TIME		$C00_{16}$			Unchanged		
NIU TIME		640_{16}			Unchanged		

1. The TSTATE fields are sampled from the relevant registers on WMR. Since these registers are preserved through WMR, the value in the TSTATE fields after a WMR are the pre-WMR values from the relevant registers.
2. This column applies to all other traps that take the processor into `RED_state` (i.e., traps while at *MAXTL* − 1).

13.9 Boot Sequence

A high-level overview of the typical UltraSPARC T2 power-up reset sequence is as follows:

1. On power-up of the system, the system controller asserts PWRON_RST_L and RST asserts all other reset signals. This causes (1) all UltraSPARC T2 internal states to reset, including all control registers and memory refresh state machines, (2) causes IO outputs to reset, and (3) protects the internal tristate muxes. The main CPU PLL will be locking to the default clock ratio during this time. The other PLLs, in the NIU SerDes and the PIU SerDes, also will be locking to their frequencies during this time.

2. Once power is up in the system, the system controller then deasserts PWRON_RST_L and the CPU fetches reset configuration programming code from the boot PROM where configuration registers (clock ratios, etc.) are programmed. RST must deassert all reset signals simultaneously and synchronously to their respective clocks.

3. A second, warm reset caused by writing to wmr_gen bit is used to reset most of UltraSPARC T2 (everything except items reset only by PWRON_RST_L above), relock the CPU PLL, and restart instruction fetch of boot code running at the reprogrammed clock ratio. The main PLL is locking during this time if the ratio is updated.

4. Subsequent warm resets may take place later via writing to wmr_gen bit, which do not disturb states which are reset only by PWRON_RST_L. Warm reset may also be caused by assertion of the PB_RST_L pin.

13.9.1 Assumed POR Software Initialization Sequence

This is the sequence we envision a machine in normal use would follow. During debug, an engineer may wish to forego some steps, such as the warm reset.

Guaranteed by hardware:

- L2 Tag, Data, and VUAD arrays, when BISTed to zeros, are initialized to empty with good parity and good ECC.
- L2 Directory (of L1 tags) is marked invalid on reset.
- L1 I-cache, L1 D-cache, when BISTed to zeros, initialized to good parity

Assumptions:

- Main Memory – Fast ECC initialization with Bzero ASI.
- SPD (Serial Presence Detect) – SPD is launched by software, and can take up to ~1/12 of a second to complete.

Sequence:

1. Service processor asserts TRST_L and PWRON_RST_L.

2. Power ramps up.

3. PLLs start up, and clocks start toggling.

4. Service processor asserts TMS and applies one TCK pulse.
5. Service processor deasserts TRST_L.
6. Service processor issues 5 TCK clock pulses - TMS still asserted.
   At this point the JTAG logic is reset.

Once TRST_L is deasserted, registers in the JTag portion of the TCU may be accessed via the JTag TAP using TCK while the PLLs in UltraSPARC T2 have not locked yet. To do this, the user needs to execute TAP_JTPOR_ACCESS after deasserting TRST_L but before deasserting PWRON_RST_L. The status of TCU can then be checked with TAP_JTPOR_STATUS; a status of 1 indicates that the TCU is paused and the JTAG programming window is active. JTAG instructions can be executed during this window. To continue with POR, the user should execute TAP_JTPOR_CLEAR after deasserting PWRON_RST_L, which will cause TCU to continue with the POR sequence.

7. Service processor deasserts PWRON_RST_L. Note that it must deassert PWRON_RST_L after deasserting TRST_L.
8. PLLs lock. NIU PLL must lock before the NIU starts.
9. The lowest-numbered available virtual processor begins fetching and executing instructions at $RSTV_{ADDR} \mid\mid 20_{16}$. The MMUs are turned off, in bypass mode, with default mapping. At first, only PROM working. Software has to enable everything else.
10. Read RSET_STAT register, which indicates POR.
11. Initialize CLK_DIV register with desired ratios.
12. Write 1 to wmr_gen bit of RESET_GEN, to initiate warm reset.
13. The lowest-numbered available virtual processor begins fetching and executing instructions at $RSTV_{ADDR} \mid\mid 20_{16}$. The MMUs are turned off, in bypass mode, with default mapping. At first, only PROM working. Software has to enable everything else.
14. Read RSET_STAT register, which indicates warm reset, with clock change.
15. Enable error detection on L1 and L2 caches.
16. Enable L1 and L2 caches.
17. Copy bootstrap into L2 cache, using ASI_BLK_INIT_ST_*.
18. Branch to bootstrap (now executing from cache).
19. Copy/decode code segments from PROM space to cacheable space, using ASI_BLK_INIT_ST_*.
20. Initialize DRAM interface blocks.

21. Force refresh asynchronicity, by zeroing out the refresh counters on each DRAM controller, at precise intervals ($n/4$ of the refresh interval value).

22. Initialize main memory using `ASI_BLK_INIT_ST_*`.

23. Initialize rest of blocks on the chips.

24. Slowly unpark the other enabled virtual processors via the `ASI_CORE_RUNNING_W1S` or `ASI_CORE_RUNNING_RW` registers. Additional virtual processors should be unparked one at a time, with at least 20 microseconds elapsing between successive unpark operations to avoid surges in power consumption.

25. All virtual processors initialize strand-specific state.

26. Strand 0 in each available physical core initializes physical core state (such as enable L1$).

27. Jump into hypervisor.

13.9.2 Assumed Warm Reset Software Initialization Sequence

Assumptions:

- Need to check whether error reset or software-generated reset.

Sequence:

1. Read RSET_STAT register, which indicates warm reset.

2. Check local error logs.

3. If errors, go to crash-dump handling.

4. If no errors, initialize/clear L1 caches.

5. Turn on caches.

6. Initialize strand-specific state.

7. Slowly unpark the other enabled virtual processors via the `ASI_CORE_RUNNING_W1S` or `ASI_CORE_RUNNING_RW` registers. Additional virtual processors should be unparked one at a time, with at least 20 microseconds elapsing between successive unpark operations to avoid surges in power consumption.

8. All virtual processors initialize strand-specific state.

9. Strand 0 in each available physical core initializes physical core state (such as clear & enable L1$).

10. Continue, as desired by SW.

13.9.3 Reset Sequence for NIU

1. Software makes sure that all outstanding transactions are complete.

2. Software writes to the niu bit of the SSYS_RESET register.

3. Software waits until the niu bit of the SSYS_RESET register is cleared.

CHAPTER **14**

CMT

UltraSPARC T2 implements the *Sun Microsystems Standard CMT* Programming Model Specification **version 2.2.1** (as incorporated into the CMT chapter of the *UltraSPARC Architecture 2007* specification), with the exception that the `ASI_CMT_ERROR_STEERING` register is not supported. Please refer to that document for details of the operation of the CMT registers listed in this chapter.

14.1 CMT Registers

14.1.1 ASI_CORE_AVAILABLE

All virtual processors share a single `ASI_CORE_AVAILABLE` register at ASI 41_{16}, VA{63:0} = 0_{16}.

TABLE 14-1 defines the format of this register.

TABLE 14-1 Strand Available – `ASI_CORE_AVAILABLE` (ASI 41_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:0	core_avail	FFF FFFF FFFF FFFF$_{16}$[1]	RO	Bits are set to 1 if the virtual processor is available, 0 if unavailable. Each physical core in UltraSPARC T2 (represented by a byte) will either be all 0s or all 1s.

1. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores unavailable may contain 00_{16} bytes in the initial value.

14.1.2 ASI_CORE_ENABLE_STATUS

All virtual processors share a single `ASI_CORE_ENABLE_STATUS` register at ASI 41_{16}, VA{63:0} = 10_{16}.

TABLE 14-2 defines the format of this register.

TABLE 14-2 Strand Enable Status – ASI_CORE_ENABLE_STATUS (ASI 41_{16}, VA 10_{16})

Bit	Field	Initial Value	R/W	Description
63:0	enable_status	FFF FFFF FFFF FFFF$_{16}$[1]	RO	Bits are set to 1 if the virtual processor is enabled, 0 if disabled. Loaded with the contents of ASI_CORE_AVAILABLE upon completion of a power-on reset. This register is loaded with the contents of ASI_CORE_ENABLE upon completion of a warm reset. Each physical core in UltraSPARC T2 (represented by a byte) will either be all 0s or all 1s.

1. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores unavailable may contain 00_{16} bytes in the initial value.

Notes The CMT spec uses CORE_ENABLE_STATUS instead of ASI_CORE_ENABLE_STATUS. ASI_CORE_ENABLE_STATUS is used in this document to avoid confusion between the two very similar register names in the CMT spec.

If ASI_CORE_ENABLE is all zeros, the lowest available physical core will remain enabled.
An interrupt sent to a disabled virtual processor will be ignored by the disabled virtual processor and will not have any effect on the virtual processor sending the interrupt.

14.1.3 ASI_CORE_ENABLE

All virtual processors share a single ASI_CORE_ENABLE register at ASI 41_{16}, VA{63:0} = 20_{16}.

TABLE 14-3 defines the format of this register.

TABLE 14-3 Strand Enable – ASI_CORE_ENABLE (ASI 41_{16}, VA 20_{16})

Bit	Field	Initial Value	R/W	Description
63:0	enable	FFF FFFF FFFF FFFF$_{16}$[1]	RW	Set bit to 1 to enable the virtual processor on the next warm reset. Set bit to 0 to disable the virtual processor on the next warm reset. Loaded with the contents of ASI_CORE_AVAILABLE upon completion of a power-on reset. Each physical core in UltraSPARC T2 (represented by a byte) will either be all zeros or all ones. When written to, if any bit within a byte is a zero, the whole byte will be set to 00_{16}.

1. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores unavailable may contain 00_{16} bytes in the initial value.

14.1.4 ASI_XIR_STEERING

All virtual processors share a single ASI_XIR_STEERING register at ASI 41_{16}, VA{63:0} = 30_{16}.

TABLE 14-4 defines the format of this register.

TABLE 14-4 XIR Steering – ASI_XIR_STEERING (ASI 41_{16}, VA 30_{16})

Bit	Field	Initial Value	R/W	Description
63:0	xirsteering	FFF FFFF FFFF FFFF$_{16}$[1]	RW	Bits are set to 1 if the virtual processor will receive an XIR when the external XIR pin is asserted. Loaded with the contents of ASI_CORE_AVAILABLE upon completion of a power-on reset. This register is loaded with the contents of ASI_CORE_ENABLE upon completion of a warm reset.

1. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores unavailable may contain 00_{16} bytes in the initial value.

Note | UltraSPARC T2 allows the ASI_XIR_STEERING register to be set to 0_{16}. When set to 0_{16}, assertion of the external XIR pin will have no effect.

14.1.5 ASI_CMT_TICK_ENABLE

All virtual processors share a single ASI_CMT_TICK_ENABLE register at ASI 41_{16}, VA{63:0} = 38_{16}. This register is preserved across warm reset.

TABLE 14-5 defines the format of this register.

TABLE 14-5 Tick Enable – ASI_CMT_TICK_ENABLE (ASI 41_{16}, VA 38_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0_{16}	RO	*Reserved*
0	tick_enable	0_{16}	RW	Set to 1 to enable incrementing of TICK register in all available, enabled physical cores

The ASI_CMT_TICK_ENABLE register synchronizes the tick registers in all physical cores of UltraSPARC T2. Each physical core contains one tick register. A physical core's TICK register increments only when it is 1) available, 2) enabled, and 3) ASI_CMT_TICK_ENABLE is set to 1. The output of the ASI_CMT_TICK_ENABLE register is distributed synchronously and with the same delay to all physical cores. Thus, when ASI_CMT_TICK_ENABLE changes, all available, enabled physical cores start or stop incrementing their tick register at the same system clock cycle.

Programming Note Hyperprivileged software can synchronize the tick registers across all available, enabled physical cores as follows. First, one strand writes 0 to `ASI_CMT_TICK_ENABLE`. Then it initializes a mutex-protected counter (`c`) to the number of available, enabled physical cores. Then, one strand on each available, enabled physical core writes the desired tick value to its core's tick register and decrements `c`. Each strand checks the value of the counter. If `c` is 0, that strand writes a 1 to the `ASI_CMT_TICK_ENABLE` register.

14.1.6 ASI_CMT_ERROR_STEERING

UltraSPARC T2 does not implement the `ASI_CMT_ERROR_STEERING` (ASI 41_{16}, VA 40_{16}) register.

14.1.7 ASI_CORE_RUNNING_RW

All virtual processors share a single `ASI_CORE_RUNNING_RW` register at ASI 41_{16}, VA{63:0} = 50_{16}.

TABLE 14-6 defines the format of this register.

TABLE 14-6 Strand Running RW – `ASI_CORE_RUNNING_RW` (ASI 41_{16}, VA 50_{16})

Bit	Field	Initial Value	R/W	Description
63:0	running_rw	1_{16}[1,2]	RW	Bits are set to 0 to park a virtual processor and set to 1 to unpark a virtual processor.

1. Initial value listed is that seen by software. After a power-on reset, the register contains 0_{16} until reset is complete and the initial strand is unparked, and an external agent viewing this register (through the tap controller) may see the zero value.
2. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores enabled will have a single bit corresponding to the lowest enabled virtual processor set.

Notes As per the *CMT Programming Model Specification*, UltraSPARC T2 prevents software from parking all strands and forces one hardware strand (the one performing the parking) to keep running when an attempt is made to park all strands. However, software must allow for a change in `ASI_CORE_RUNNING_RW` to propagate by waiting for the value of `ASI_CORE_RUNNING_STATUS` to match the value written to `ASI_CORE_RUNNING_RW` before making another update, otherwise hardware operation is unpredictable. In particular, a strand or all strands may become parked or unparked and remain unresponsive to further unpark or park commands, until a warm reset or POR is performed.

If a strand parks itself, the strand is guaranteed to not execute any instructions beyond the instruction that parked it (i.e. there is no skid following the parking instruction).

WARNINGS! Software must not attempt to park a strand that is not completely unparked (that is, the strand's bit in `ASI_CORE_RUNNING_STATUS` must be 1 before clearing the strand's bit in `ASI_CORE_RUNNING_RW`). Operation of UltraSPARC T2 is undefined when a strand that is not unparked has its bit in `ASI_CORE_RUNNING_RW` cleared. In particular, the strand may become unparked and remain unresponsive to further park commands, until a warm reset or POR is performed.
Software must not attempt to unpark a strand that is not completely parked (that is, the strand's bit in `ASI_CORE_RUNNING_STATUS` must be 0 before setting the strand's bit in `ASI_CORE_RUNNING_RW`). Operation of UltraSPARC T2 is undefined when a strand that is not parked has its bit in `ASI_CORE_RUNNING_RW` set. In particular, the strand may become parked and remain unresponsive to further unpark commands, until a warm reset or POR is performed.

14.1.8 ASI_CORE_RUNNING_STATUS

All virtual processors share a single `ASI_CORE_RUNNING_STATUS` register at ASI 41_{16}, VA{63:0} = 58_{16}.

TABLE 14-7 defines the format of this register.

TABLE 14-7 Strand Running Status – ASI_CORE_RUNNING_STATUS (ASI 41_{16}, VA 58_{16})

Bit	Field	Initial Value	R/W	Description
63:0	running_status	1_{16}[1,2]	RO	Bits are set to 0 if the virtual processor is currently parked and set to 1 if the virtual processor is currently running.

1. Initial value listed is that seen by software. After a power-on reset, the register contains 0_{16} until reset is complete and the initial strand is unparked, and an external agent viewing this register (through the tap controller) may see the zero value.
2. Initial value listed is for a fully available UltraSPARC T2. An UltraSPARC T2 with some physical cores enabled will have a single bit corresponding to the lowest enabled virtual processor set.

Note | While a virtual processor is parked, interrupt and XIR events targeting the virtual processor will be held pending and will be taken once the virtual processor is unparked.

14.1.9 ASI_CORE_RUNNING_W1S

All virtual processors share a single ASI_CORE_RUNNING_W1S register at ASI 41_{16}, VA{63:0} = 60_{16}.

TABLE 14-8 defines the format of this register.

TABLE 14-8 Strand Running W1S – ASI_CORE_RUNNING_W1S (ASI 41_{16}, VA 60_{16})

Bit	Field	Initial Value	R/W	Description
63:0	running_w1s	0_{16}	WO	Writing 1 to a bit will set the corresponding bit in ASI_CORE_RUNNING_RW. Writing 0 to a bit will leave the corresponding bit in ASI_CORE_RUNNING_RW unchanged.

Software must not attempt to unpark a strand that is not completely parked (that is, the strand's bit in ASI_CORE_RUNNING_STATUS must be 0 before using ASI_CORE_RUNNING_W1S to set the strand's bit in ASI_CORE_RUNNING_RW). Operation of UltraSPARC T2 is undefined when a strand that is not parked has its bit in ASI_CORE_RUNNING_RW set. In particular, the strand may become unparked and remain unresponsive to further park commands, until a warm reset or POR is performed.

14.1.10 ASI_CORE_RUNNING_W1C

All virtual processors share a single ASI_CORE_RUNNING_W1C register at ASI 41_{16}, VA{63:0} = 68_{16}.

TABLE 14-9 defines the format of this register.

TABLE 14-9 Strand Running W1C – ASI_CORE_RUNNING_W1C (ASI 41_{16}, VA 68_{16})

Bit	Field	Initial Value	R/W	Description
63:0b	running_w1c	0_{16}	WO	Writing 1 to a bit will clear the corresponding bit in ASI_CORE_RUNNING_RW. Writing a zero to a bit will leave the corresponding bit in ASI_CORE_RUNNING_RW unchanged.

Notes As per the *CMT Programming Model Specification*, UltraSPARC T2 prevents software from parking all strands and forces one hardware strand (the one performing the parking) to keep running when an attempt is made to park all strands. However, software must allow for a change in ASI_CORE_RUNNING_W1C to propagate by waiting for the value of ASI_CORE_RUNNING_STATUS to match the value written to ASI_CORE_RUNNING_W1C before making another update, otherwise hardware operation is unpredictable. In particular, the strand may become parked and remain unresponsive to further unpark commands, until a warm reset or POR is performed. If a strand parks itself, the strand is guaranteed to not execute any instructions beyond the instruction that parked it (i.e. there is no skid following the parking instruction).

Software must not attempt to park a strand that is not completely unparked (that is, the strand's bit in ASI_CORE_RUNNING_STATUS must be 1 before using ASI_CORE_RUNNING_W1C to clear the strand's bit in ASI_CORE_RUNNING_RW). Operation of UltraSPARC T2 is undefined when a strand that is not unparked has its bit in ASI_CORE_RUNNING_RW cleared. In particular, the strand may become parked and remain unresponsive to further unpark commands, until a warm reset or POR is performed.

14.2 ASI_CMT_CORE Registers

14.2.1 ASI_CMT_CORE_INTR_ID

Each virtual processor has a read-only ASI_CMT_CORE_INTR_ID register at ASI 63_{16}, VA{63:0} = 0_{16}.

TABLE 14-10 defines the format of this register.

TABLE 14-10 Strand Interrupt ID – `ASI_CMT_CORE_INTR_ID` (ASI 63_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15:6	intr_id_hi	0_{16}	RO	Upper bits of Interrupt ID are all 0.
5:0	intr_id_lo	coreid	RO	Matches `ASI_CMT_STRAND_ID` bits 5:0.

14.2.2 `ASI_CMT_STRAND_ID`

Each virtual processor has a read-only `ASI_CMT_STRAND_ID` register at ASI 63_{16}, VA{63:0} = 10_{16}.

TABLE 14-11 defines the format of this register.

TABLE 14-11 Strand ID – `ASI_CMT_STRAND_ID` (ASI 63_{16}, VA 10_{16})

Bit	Field	Initial Value	R/W	Description
63:38	—	0_{16}	RO	*Reserved*
37:32	max_strand_id	7_{16}	RO	Each physical core on UltraSPARC T2 consists of 8 strands.
31:22	—	0_{16}	RO	*Reserved*
21:16	max_strand_id	$3F_{16}$	RO	UltraSPARC T2 contains 64 virtual processors
15:6	—	0_{16}	RO	*Reserved*
5:0	strand_id	coreid	RO	Physical strand ID in 5:3, strand ID in 2:0.

Note The strand ID in UltraSPARC T2 is fixed based on physical core and strand numbers. This implies that an UltraSPARC T2 with unavailable cores will have holes in the strand ID space (for example, if physical core 1 is unavailable, there will be no strand_id 8_{16}–F_{16}).

CHAPTER 15

Noncacheable Unit (NCU) and Boot ROM Interfaces

15.1 Noncacheable Unit (NCU)

The main functions of the NCU are to route PIO accesses from the CMP virtual processors to the I/O subsystem and to vector interrupts from the I/O subsystem to the CMP virtual processors. The NCU provides CSRs for NCU management, configuration of the PCI Express (PCIE) address space, and mondo interrupt management.

The NCU decodes the I/O physical address space. UltraSPARC T2 supports 40-bit physical addresses, where the MSB (bit 39) is 0 for cacheable accesses (memory system) and 1 for noncacheable accesses (I/O subsystem).

NCU determines the destination of a PIO access by examining the 8 MSB (bit 39:32) of the physical address. All accesses received by NCU have bit 39 of the physical address set to 1. The address range of each IO subsystem block can be found in the following table.

TABLE 15-1 Global Physical Address Assignments

MSB Address Range{39:32}	Assignment
00_{16} ~ $7F_{16}$	Not supported by NCU (memory)
80_{16}	NCU
81_{16}	NIU
82_{16}	*Reserved*
83_{16}	CCU
84_{16}	MCUs 13:12 = 00_2 for MCU0, 13:12 = 01_2 for MCU1 13:12 = 10_2 for MCU2, 13:12 = 11_2 for MCU3
85_{16}	TCU (JTAG / TAP unit)

TABLE 15-1 Global Physical Address Assignments

MSB Address Range{39:32}	Assignment
86_{16}	DBG
87_{16}	*Reserved*
88_{16}	DMU
89_{16}	RST
$8A_{16}$ ~ $8F_{16}$	*Reserved*
90_{16}	ASI CPU shared registers (directly accessible only by JTAG/TAP unit)
91_{16} ~ $9F_{16}$	*Reserved*
$A0_{16}$ ~ BF_{16}	Not supported by NCU (L2 control and status registers)
$C0_{16}$ ~ CF_{16}	PCIE (64 Gbytes) / DMUPIO
$D0_{16}$ ~ FE_{16}	*Reserved*
FF_{16}	SSI (boot ROM)

15.2 NCU Management Registers

The NCU provides a unique serial number for each UltraSPARC T2 chip. In addition, the NCU contains registers showing the eFuse, Sparc core, and L2 bank status.

15.2.1 Serial Number

The serial number register format is show in TABLE 15-2.

TABLE 15-2 Processor Serial Number – SER_NUM ($80\ 0000\ 1000_{16}$)

Bit	Name	Initial Value	R/W	Description
63:60	delta_vdd	X	RO	Delta_Vdd[3] is a sign bit (increase or decrease with respect to nominal). Bits [2:0] specify one of 8 increments.
59:50	delta_t	X	RO	Indicates the temperature offset for the thermal diode, in increments of 20 mV.
49	reserved	0	RO	reserved for testinfo
48:46	fab	X	RO	
45:41	reserved	0	RO	reserved for testinfo
40	bin	X	RO	

TABLE 15-2 Processor Serial Number – SER_NUM ($80\ 0000\ 1000_{16}$)

39:16	lot	X	RO
15:10	wafer	X	RO
9:5	column	X	RO
4:0	row	X	RO

15.2.2 eFuse Status

The eFuse Status register format is show in TABLE 15-3.

TABLE 15-3 eFuse Status – EFU_STAT ($80\ 0000\ 1008_{16}$)

Bit	Name	Initial Value	R/W	Description
63:0	efu_status	FFFF FFFF FFFF $FFFF_{16}$	RO	eFuse status programmed by eFuse block

15.2.3 Strand Available

The Strand Available register format is show in TABLE 15-4. This register is an alias for the ASI_CORE_AVAILABLE register described in 14.1.1 ON PAGE 187.

TABLE 15-4 Strand Available – CORE_AVAIL ($80\ 0000\ 1010_{16}$)

Bit	Name	Initial Value	R/W	Description
63:0	avail	FFFF FFFF FFFF $FFFF_{16}$	RO	Strand available programmed by eFuse. Note that all strands within a physical core will be programmed to the same value (all available or all unavailable).

15.2.4 L2 Configuration Control and Status Registers

The NCU contains several L2 Configuration Control and Status registers.

- The L2 Bank Available register is described in Section 19.14.2, *L2 Bank Available*, on page 433.
- The L2 Bank Enable register is described in Section 19.14.3, *L2 Bank Enable*, on page 433.
- The L2 Bank Enable Status Register is described in Section 19.14.4, *L2 Bank Enable Status*, on page 435.
- The L2 Index Hash Enable register is described in Section 19.14.5, *L2 Index Hash Enable*, on page 436.
- The L2 Index Hash Enable Status register is described in Section 19.14.6, *L2 Index Hash Enable Status*, on page 437.

15.3 NCU PCIE Address Mapping Registers

Note | NCU adopted the Fire ASIC's offset base and offset mask for PCIE PIO mapping. This information is derived from the Fire Programmer's Reference Manual (PRM), Section 1.2.2.2. The Fire PRM is available from either
http://wikis.sun.com/display/FOSSdocs/Home
or
http://www.sun.com/processors/documentation.html

15.3.1 Physical Address Partitioning

The 64-Gbyte region of noncacheable physical memory, from C0 0000 0000_{16} to CF FFFF $FFFF_{16}$, is mapped by is reserved for mapping NCU to access PCIE address space. This 64-Gbyte noncacheable space is split into three subregions: the PCI Express configuration and I/O subregion; the 32-bit addressable PCI Express memory subregion; and the 64-bit addressable PCI Express memory subregion. These are associated with the one PCI Express link, which is called PCIE-A, supported by UltraSPARC T2.

The base address and size of these subregions is programmed during chip initialization using the Address Mask/Match registers defined in Section 15.3.6, *NCU PCIE Registers*, on page 203. Accesses to PCIE's 6-Gbyte noncacheable region **outside** the subregions mapped by the Address Mask/Match registers are errors (reads return an error and writes are silently dropped).

The subregions within the 64-Gbyte noncacheable region are summarized in TABLE 15-5.

TABLE 15-5 64 GByte Noncacheable Region Partitioning

Subregion	Size	Supported PIO accesses
PCIE-A CFG/IO	512 Mbyte	noncacheable read (4 byte max) noncacheable write (4 byte max)
PCIE-A Mem32	16Mbyte–2Gbyte	noncacheable read (8 byte) noncacheable write (8 byte)
PCIE-A Mem64	16Mbyte–32Gbyte	noncacheable read (8 byte) noncacheable write (8 byte)

15.3.2 Offset Base and Offset Mask Register Behavior

PCIE uses a common implementation for all its address match and mask registers. In some cases this leads to extra bits that are not strictly necessary as they allow PCIE to map a subregion larger than what is implemented by the corresponding subregion. PCIE uses these registers to determine where the various subregions live within the 64-Gbyte noncacheable region.

NCU first checks PA bits 39:36 = C_{16} to decode PIO accesses to the PCIE space. If the address is found to be in this memory range, bits 35:24 are masked with (bitwise **and**) Offset Mask register bits 35:24 and compared to the Offset Base register bits 35:24.

WARNING! Changing the Offset Mask or Base register values, while allowed, will cause the location of PCIE's address mappings to move in physical address space.

The following sequence may be used to update the Mask and Base registers (assuming that loads are blocking):

1. Prevent new PIO accesses to mapped space (that is, quiesce).
2. Write new values to Mask and Base registers.
3. Perform a read of Mask or Base register (guarantees write has occurred).
4. Reenable PIO accesses to mapped space at new location.

Note that it is software's job to prevent the PIO accesses from occurring in the first step and to reenable them (allow them) in the last.

The address filtering is performed as described in FIGURE 15-1.

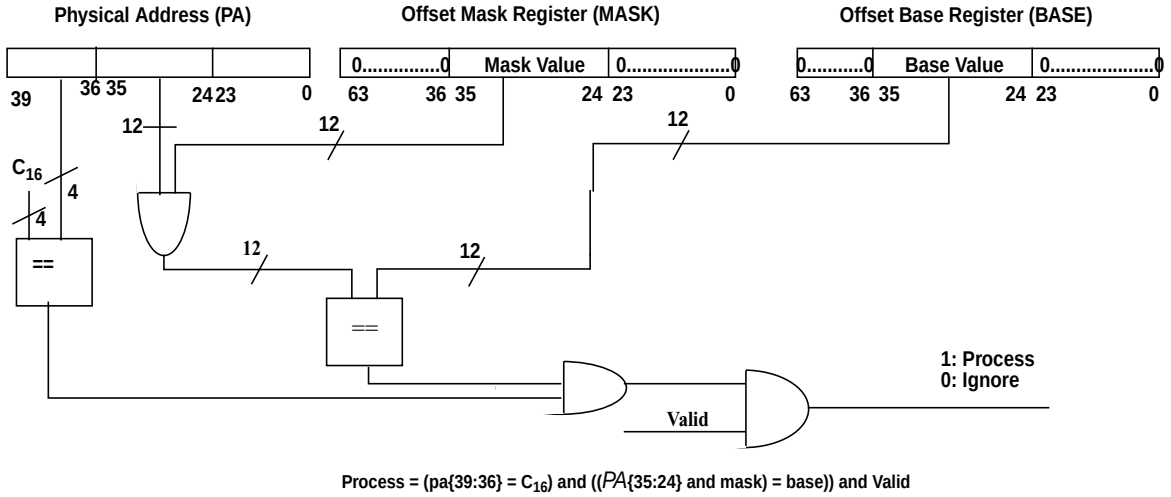


FIGURE 15-1 Address Filtering in PCIE

15.3.3 PCI Express Configuration and I/O Subregion

There are two PCI Express configuration and I/O address subregions. A PCI Express configuration and I/O address subregion are a fixed size of 512 Mbytes. The offset within PCIE's 64-Gbyte noncacheable region and subregion size are set using the PCIE-A Cfg/IO Offset Base register and the PCIE-A Cfg/IO Offset Mask register. The offset base and offset mask registers for this subregion must be initialized for a 512-Mbyte subregion. Failure to initialize them will result in undefined behavior. The first 256 Mbytes of the PCI Express configuration and I/O address subregion map to PCI Express configuration space. The last 256 Mbytes of this subregion map to PCI Express IO space. All configuration and IO packets are issued with the virtual channel ID equal to zero (the default virtual channel).

The mapping from the software-relative (that is, byte addressing, no byte masks) physical address 39:00 to the PCI Express Configuration and I/O subregion is as follows:

- 39:36 — Set to C_{16}
- 35:29 — Configuration and I/O Address match. These bits determine if the access to PCIE maps into one of the configuration and I/O subregions.
- 28 — Configuration or I/O operation. 0_{16} (0_2) signifies a configuration operation. 1_{16} (1_2) signifies an I/O operation.
- 27:00 — Configuration- or I/O space-specific addressing information. See *Accessing PCI Express Configuration Space* and *Accessing PCI Express I/O Space* for more information on how these address bits are used.

15.3.3.1 Accessing PCI Express Configuration Space

The mapping from the 28-bit physical address offset within the PCI Express Configuration space to the Configuration Addressing information is as follows:

- 27:20 (8) — PCI Express bus number
- 19:15 (5) — Device number
- 14:12 (3) — Function number
- 11:00 (12) — PCI Express Configuration register address. Address bits 01:00 determine the byte enables on PCI Express.

Note The Configuration Address Space per Device Function has been extended in PCI Express from 256 bytes to 4096 bytes. Configuration accesses through PCIE are different from previous generation Sun Bridges. The Configuration register address has been extended to 12 bits (from 8) and the bus number, device number, and function number have been shifted left 4 bits. Therefore, software that accesses configuration registers for PCI Express devices and PCI devices sitting behind a PCI Express to PCI bridge will be different on PCIE-based systems.

If the Bus Number for a configuration access matches the Bus Number in the DMC PCI Express Configuration register,[1] a Type 0 configuration packet is generated. The bus number in DMC PCI Express Configuration register is programmable by software. If the bus number is not equal to the bus number in DMC PCI Express Configuration register, a Type 1 configuration packet is generated.

A PCI special cycle can be generated on a PCI bus attached (directly or indirectly) to a PCI Express to PCI bridge by sending a Type 1 configuration packet with the appropriate bus number, all 1's for the device number ($1F_{16}$) and function number (7_{16}), and all zeros for the configuration address; for example, PartialAddress{19:00} = $FF000_{16}$.

PCIE does not implement any registers within PCI Express configuration space.

See the PCI and PCI Express specifications for more detail on accessing configuration space.

15.3.3.2 Accessing PCI Express I/O Space

The mapping from the 28-bit physical address offset within the PCI Express I/O space to the PCI Express I/O address is as follows:

- 27:0 -— PCI Express I/O Address. Address bits 31:28 are always set to zero.

1. This register is defined in the PCIE chapter.

15.3.4 PCI Express 32-bit Addressing Memory Subregion

The PCI Express 32-bit addressing memory subregion has a programmable size ranging from 16 Mbytes to 2 Gbytes. The offset within PCIE's 64-Gbyte noncacheable region and subregion size are set using the PCI Express Bus A Mem32 Offset Base register and the PCI Express Bus A Mem32 Offset Mask register. The offset base and offset mask registers for this subregion must be initialized for a subregion between 16 Mbytes and 2 Gbytes (inclusive). Failure to initialize these registers will result in undefined behavior.

The mapping from the software-relative (that is, byte addressing, no byte masks) physical address {39:00} to the PCI Express 32-bit addressing memory subregion is as follows:

- 39:36 — Set to C_{16}.
- 35:(*X* + 1) — Address match. These bits determine if the access to PCIE maps into one of the 32-bit addressing memory subregions. *X* is determined by the size of the Mem32 subregion, which is set by the PCI Express Bus A Mem32 Offset Mask register. See Section 15.3.2, *Offset Base and Offset Mask Register Behavior*, on page 199 for more information.
- X:00 — PCI Express 32-bit memory address. The PCI Express Mem32 address bits 31:(X+1) are set to 0. X is determined by the size of the Mem32 subregion which is set by the PCI Express Bus A Mem32 Offset Mask register. See Section 15.3.2, *Offset Base and Offset Mask Register Behavior*, on page 199 for more information.

15.3.5 PCI Express 64-bit Addressing Memory Subregion

The PCI Express 64-bit addressing memory subregion has a programmable size ranging from 16 Mbytes to 32 Gbytes. The offset within PCIE's 64-Gbyte noncacheable region and subregion size are set by the PCI Express Bus A Mem64 Offset Base register and the PCI Express Bus A Mem64 Offset Mask register. The offset base and offset mask registers for this subregion must be initialized for a subregion between 16 Mbytes and 32 Gbytes (inclusive). Failure to initialize them will result in undefined behavior. All downbound packets to this subregion are issued with the virtual channel ID equal to zero (the default virtual channel).

The mapping from the software-relative (that is, byte addressing, no byte masks) physical address 42:00 to the PCI Express 64-bit addressing memory subregion is as follows:

- 39:36 — Set to C_{16}.
- 35:(*X* + 1) — Address match. These bits determine if the access to PCIE maps into one of the 64-bit addressing memory subregions. *X* is determined by the size of the Mem64 subregion which is set by the PCI Express Bus A Mem64 Offset Mask register. See Section 15.3.2, *Offset Base and Offset Mask Register Behavior*, on page 199 for more information.
- X:00 — Offset within PCI Express 64-bit memory address.

The PCI Express Mem64 PIO address is formed from

{offset{63:36} :: (offset{35:24} | phys addr[{35:24}) :: phys addr{23:2}}

where phys addr{35:0} = {(pa{35:24} & (~mask{35:24})) :: pa{23:0}} and offset{63:24} is from the Mem64 PCIE Offset register.[1]

15.3.6 NCU PCIE Registers

The following registers define the base/mask for the PCIE subregions as described in the previous sections. The table captions specify the offset of each register relative to the base address of the NCU management registers ($80\ 0000\ 0000_{16}$).

These registers are not protected by warm reset , they get reset to their POR state on warm reset.

TABLE 15-6 PCIE LinkA Mem32 Addr Offset Base – PCIE_A_MEM32_OFFSET_BASE (2000_{16})

Bit	Name	Initial Value	R/W	Description
63	mem32_enable	0	RW	Mem32 region enable.
62:36	—	0	RO	*Reserved*
35:24	mem32_base	0	RW	Mem32 offset base.
23:0	—	0	RO	*Reserved*

TABLE 15-7 PCIE LinkA Mem32 Address Mask – PCIE_A_MEM32_OFFSET_MASK (2008_{16})

Bit	Name	Initial Value	R/W	Description
63:40	—	0	RO	*Reserved*
39:36	mem32_mask_hi	F_{16}	RO	Position mask of PA{39:36}, always 1111_2.
35:24	mem32_mask	0	RW	Position mask of PA{35:24}.
23:0	—	0	RO	*Reserved*

TABLE 15-8 PCIE LinkA Memory64 Domain Address Base – PCIE_A_MEM64_OFFSET_BASE (2010_{16})

Bit	Name	Initial Value	R/W	Description
63	mem64_enable	0	RW	Mem64 region enable.
62:36	—	0	RO	*Reserved*
35:24	mem64_base	0	RW	Mem64 offset base.
23:0	—	0	RO	*Reserved*

1. This is a register defined in the PCIE chapter.

TABLE 15-9 PCIE LinkA Memory64 Domain Address Mask – PCIE_A_MEM64_OFFSET_MASK (2018_{16})

Bit	Name	Initial Value	R/W	Description
63:40	—	0	RO	*Reserved*
39:36	mem64_mask_hi	F_{16}	RO	Position mask of PA{39:36}, always 1111_2.
35:24	mem64_mask	0	RW	Position mask of PA{35:24}.
23:0	—	0	RO	*Reserved*

TABLE 15-10 PCIE LinkA IOConfig Domain Address Base – PCIE_A_IOCON_OFFSET_BASE (2020_{16})

Bit	Name	Initial Value	R/W	Description
63	iocon_enable	0	RW	IOconfig region enable.
62:36	—	0	RO	*Reserved*
35:24	iocon_base	0	RW	IOconfig offset base.
23:0	—	0	RO	*Reserved*

TABLE 15-11 PCIE LinkA IOConfig Domain Address Mask – PCIE_A_IOCON_OFFSET_MASK (2028_{16})

Bit	Name	Initial Value	R/W	Description
63:40	—	0	RO	*Reserved*
39:36	iocon_mask_hi	F_{16}	RO	Position mask of PA{39:36}, always 1111_2.
35:24	iocon_mask	0	RW	Position mask of PA{35:24}.
23:0	—	0	RO	*Reserved*

TABLE 15-12 PCIE LinkA FLUSH – PCIE_A_FSH (2030_{16})

Bit	Name	Initial Value	R/W	Description
63:0	mmu_flush	0	RW	Each write into this register will cause one cycle of ncu_dmu_mmu_addr_vld to DMU.

15.4 NCU ASI Registers

Several CMP and Interrupt registers are located in the NCU and made accessible to the JTAG/TAP controller, as listed below. In addition, all ASI registers are made accessible to the JTAG/TAP controller through the addressing shown in TABLE 15-13.

TABLE 15-13 NCU ASI Register Physical Address Map

Bit	Field	Value	Description
39:32	ncu	90_{16}	Identifies NCU space.
31:29	core	0_{16}–7_{16}	Identifies physical core being targeted (set to 0_{16} for a shared ASI register).
28:26	strand	0_{16}–7_{16}	Identifies strand being targeted (set to 0_{16} for a shared ASI register).
25:18	asi	0_{16}–FF_{16}	Identifies ASI being targeted.
17:3	va	0_{16}–$7FFFF_{16}$	Identifies va{17:13} being targeted.
2:0	—	0_{16}	Always zero for 64-bit access.

Note | For a shared ASI register, different virtual processor, thread may access the same ASI VA register in NCU. For those registers, NCU ignores PA{31:26} when PA{39:32} = 90_{16}.

15.4.1 Strand Available Register (ASI 41_{16} VA 00_{16})

This register is described in Section 14.1.1, *ASI_CORE_AVAILABLE*, on page 187. It is available to the JTAG/TAP controller at address 90 0104 0000_{16}.

15.4.2 Strand Enable Status Register (ASI 41_{16} VA 10_{16})

This register is described in Section 14.1.2, *ASI_CORE_ENABLE_STATUS*, on page 187. It is available to the JTAG/TAP controller at address 90 0104 0010_{16}.

15.4.3 Strand Enable Register (ASI 41_{16} VA 20_{16})

This register is described in Section 14.1.3, *ASI_CORE_ENABLE*, on page 188. It is available to the JTAG/TAP controller at address 90 0104 0020_{16}.

15.4.4 XIR Steering Register (ASI 41_{16} VA 30_{16})

This register is described in Section 14.1.4, *ASI_XIR_STEERING*, on page 189. It is available to the JTAG/TAP controller at address 90 0104 0030_{16}.

15.4.5 CMP Tick Enable Register (ASI 41_{16} VA 38_{16})

This register is described in Section 14.1.5, *ASI_CMT_TICK_ENABLE*, on page 189. It is available to the JTAG/TAP controller at address 90 0104 0038_{16}.

15.4.6 Strand Running RW Register (ASI 41_{16} VA 50_{16})

This register is described in Section 14.1.7, *ASI_CORE_RUNNING_RW*, on page 190. It is available to the JTAG/TAP controller at address 90 0104 0050_{16}.

15.4.7 Strand Running Status Register (ASI 41_{16} VA 58_{16})

This register is described in Section 14.1.8, *ASI_CORE_RUNNING_STATUS*, on page 191. It is available to the JTAG/TAP controller at address 90 0104 0058_{16}.

15.4.8 Strand Running W1S Register (ASI 41_{16} VA 60_{16})

This register is described in Section 14.1.9, *ASI_CORE_RUNNING_W1S*, on page 192. It is available to the JTAG/TAP controller at address 90 0104 0060_{16}.

15.4.9 Strand Running W1C Register (ASI 41_{16} VA 68_{16})

This register is described in Section 14.1.10, *ASI_CORE_RUNNING_W1C*, on page 192. It is available to the JTAG/TAP controller at address 90 0104 0068_{16}.

15.4.10 SOC Error Steering Register (90 0104 1000_{16})

This register is described in Section 16.23.4, *SOC Error Steering Register*, on page 345. It is available to the JTAG/TAP controller at address 90 0104 1000_{16}.

15.4.11 Warm Reset Vector Mask Register (ASI 45_{16} VA 18_{16})

This register is described in Section 20.1.4, *ASI_RST_VEC_MASK*, on page 448. It is available to the JTAG/TAP controller at address 90 0114 0018_{16}.

15.4.12 Interrupt Vector Dispatch Register (ASI 73_{16} VA 0_{16})

This register is described in Section 7.3.3, *Interrupt Vector Dispatch Register*, on page 60. It is available to the JTAG/TAP controller at address 90 01CC 0000_{16}.

15.5 Boot ROM Address Region

The format of the Boot ROM Range register is defined in TABLE 15-14.

TABLE 15-14 Address Range Definition Boot ROM Range (FF FXXX $XXXX_{16}$)

Bit	Field	Value	Description
39:32	bootspace	FF_{16}	Identifies BOOT ROM space
31:28	bootrom	F_{16}	Identifies BOOT ROM range
27:0	romaddr		Byte address sent to Boot ROM (256 Mbyte available)

Addresses within the Boot ROM address range (FF F000 0000_{16} to FF FFFF $FFFF_{16}$) are issued to the Boot ROM, aliasing the Boot ROM to cover the top 16 Mbytes of the available space. The only transactions that are supported directly to the Boot ROM are

- 1, 2, 4, 8 byte-aligned reads
- 1, 2, 4, 8 byte-aligned writes

Since the Boot ROM is predominantly used for instructions, which are explicitly always big-endian, all accesses to the Boot ROM are treated as big-endian.

15.5.1 Boot ROM Interface Registers

All of the BOOT ROM Interface registers other than SSI Clock Select register are defined elsewhere, specifically in Chapter 16, *Error Handling*.

15.5.1.1 SSI Clock Select Register

The SSI Clock Select register controls the SSI clock frequency as a fraction of the core clock frequency. The format of the SSI Clock Select register is as shown in TABLE 15-15.

TABLE 15-15 SSI Clock Select Register – SSI_CLKSEL ($80\ 0000\ 3040_{16}$)

Bit	Field	Initial Value	R/W	Description
63:2	*Reserved*	0_{16}	RO	*Reserved*
1:0	ssi_scksel	0_{16}	RW	0_{16}: SSI Clock = iol2clk/8 1_{16}: SSI Clock = iol2clk/4 2_{16}: SSI Clock = iol2clk/8 3_{16}: SSI Clock = iol2clk/8 Preserved on warm reset. Value programmed takes effect on next warm reset.

CHAPTER 16

Error Handling

16.1 Error Classes

Errors on UltraSPARC T2 fall into three main classes: fatal (FE), hardware uncorrected (UE), and hardware corrected (CE). Hardware uncorrected errors have two subclasses: software recoverable and software unrecoverable. Hardware corrected errors have two subclasses: hardware corrected and cleared, and hardware corrected but not cleared.

Fatal errors are generated whenever the hardware detects a condition where an error has occurred and the extent to which the error may have propagated is unbounded. An example of a fatal error is an uncorrectable VUAD corruption in the L2 cache; this case implies that global cache coherence has been lost. Since a fatal error may have corrupted key operating system or hypervisor data structures, fatal errors generate an immediate warm reset to the UltraSPARC T2 chip.

Uncorrected errors are errors for which the hardware does not take corrective action. Uncorrected errors fall into two classes: software recoverable and software unrecoverable. For software unrecoverable errors, the extent to which the error may have propagated is tightly bounded. Examples of uncorrected errors include a data parity error on the DTLB, and a double-bit ECC error in the L2 cache data. The data parity error in the DTLB may be recoverable in software by forcing the TLB entry with the bad parity to be invalidated, which will then be reloaded over with a new TTE entry and good parity on the subsequent TLB miss. A double-bit L2 ECC error is unrecoverable as the data cannot be corrected by software, but the extent to which the error could have propagated is limited to the address space of any process that has access to that memory location, and software may be able to keep the system running by killing all processes that could be affected by the error and then scrubbing the bad memory location. Uncorrected errors are reported through several traps, including precise, deferred, and disrupting traps.

Deferred errors consist solely of uncorrectable store buffer errors in the SPARC virtual processor. They cause a nonmaskable, deferred *store_error* trap. NotData errors (NDE) are a specific case of uncorrected errors, where the virtual processor encountering the NotData error is not the first virtual processor to encounter the error.

Corrected errors are errors that are corrected by UltraSPARC T2 in hardware. Corrected errors can either be cleared by hardware or require software to be cleared. Examples of correctable errors that are cleared by hardware are a data parity error in the instruction cache or a tag parity error in the data cache. Corrected errors that are cleared by hardware generate a disrupting *hw_corrected_error* trap. Examples of corrected errors that require clearing by software are a single-bit error in the L2 cache access for an instruction fetch or data fetch. Corrected errors that are not cleared by hardware generate a disrupting *sw_recoverable_error* trap.

16.2 NotData Overview

NotData is used to flag a data value that has an uncorrectable error, after an uncorrectable error trap has been signaled to one and only one virtual processor. Subsequent accesses by virtual processors to NotData also take an uncorrectable error trap, but their Error Status register flags the trap as having come from access to NotData. Support for NotData prevents the problem of multiple virtual processors attempting to fix the same error, and also prevents the problem of a single uncorrectable error appearing as multiple errors as it enters different subsystems or is accessed multiple times.

In UltraSPARC T2, NotData is implemented for items as they enter into the L2 cache only. This handles the most significant sources of uncorrectable errors, namely, UEs in main memory, UEs generated by the processor, or UEs coming in from the I/O subsystem through the SIU. UEs arising in the L2 itself are kept to a very low probability through hardware scrubbing.

On a cache fill, UltraSPARC T2 signals a UE trap to one and only one virtual processor and loads each 16-byte chunk with a UE into the L2 with the NotData indication (inverted check bits on the four 4-byte chunks). Likewise, on a processor request with a UE, each 4-byte chunk of data in error is loaded into the L2 with the NotData indication. The UE trap for processor requests has already been generated by the processor subsystem.

The UltraSPARC T2 implementation of NotData protects for the most frequent UEs but does not provide full NotData protection. The following very low probability cases are not protected by NotData and can result in multiple UE traps:

- Writeback of a line with a UE or NotData is written back to main memory with an ECC encoding that is not distinguishable from a normal memory UE. Thus, a refetch of the line from memory will generate a second UE trap.
- A multiple-bit error in the L2 data array. The UE is *not* converted to NotData on an access, scrubber read, etc., because this was too large a change from the UltraSPARC T1 implementation and provided negligible impact on FIT rate. Multiple accesses of the UE before software has a chance to clean up the line will result in multiple UE traps.

16.3 CMP Error Overview

Errors detected in the CMP and memory subsystem are reported in three sets of error registers: Core, L2 cache, and DRAM. Precise and deferred errors are reported in the Core register set in program order. The Core Error Recording Enable register (CERER) controls whether errors are recorded in the virtual processor. The Strand Error Trap Enable register (SETER) controls whether the strand takes a trap as a result of any reported errors. Errors are reported in the L2 cache and DRAM error sets in the order the errors occur. The L2 cache Error Enable register controls whether errors associated with the L2 cache and DRAM are reported back to the initiator (logging of the errors in the L2 cache and DRAM registers is always performed). If the L2 cache error enable bit (ceen or nceen, depending on whether the error is correctable or uncorrectable/NotData) is not set, error information is not reported back to the initiator. For the uncorrectable/NotData case (nceen clear), this means that bad data will be executed/used, and thus the nceen bit is only intended to be cleared during heavily controlled phases of diagnostic operation.

The UltraSPARC T2 core records errors in five error status registers (ESRs). All of the ESRs provide one copy per strand (virtual processor). UltraSPARC T2 records instruction-fetch-related errors in the I-SFSR (Instruction Synchronous Fault Status register) and precise data-access-related errors in the D-SFSR (Data Synchronous Fault Status register) and D-SFAR (Data Synchronous Fault Address register). The D-SFAR is shared by the UltraSPARC T2 implementation between errors and MMU processing to avoid the need for a separate address register for precise error logging. UltraSPARC T2 also has a disrupting ESR (DESR) to log disrupting errors. Finally, the UltraSPARC T2 core detects store buffer errors. Some of these are handled as nonmaskable, deferred traps and are recorded in the deferred ESR (DFESR).

The L2 error registers can log detailed information for a single error. A dedicated error register is provided for NotData errors, while correctable, uncorrectable, and fatal errors share a single register. The L2 error registers have bits to indicate if multiple errors have occurred.

The DRAM error registers can log detailed information for a single error. A single error register is provided for the two different error classes: uncorrectable and correctable. The DRAM error registers have bits to indicate if multiple errors have occurred.

For the L2 and DRAM registers, fatal and uncorrectable errors overwrite earlier correctable error information. The error registers have bits to indicate if multiple errors have occurred. There are two bits for multiple errors: meu (multiple uncorrectable errors) and mec (multiple correctable errors). TABLE 16-1 lists the multiple error logging and overwrite behavior of the error registers (a CE for the main logged error with the meu bit set is not possible due to FE and UE being higher priority than CE in logging).

TABLE 16-1 Multiple Error Logging in the L2 and DRAM Registers

Main Logged Error	meu	mec	Description
FE	0	0	Single FE encountered and logged.
FE	0	1	Single FE encountered and logged. One or more correctable errors encountered but details on the correctable errors not logged.
FE	1	0	Single FE encountered and logged. One or more uncorrectable errors encountered but details on the uncorrectable errors not logged.
FE	1	1	Single FE encountered and logged. One or more uncorrectable errors encountered but details on the uncorrectable errors not logged. One or more correctable errors encountered but details on the correctable errors not logged.
UE	0	0	Single UE encountered and logged.
UE	0	1	Single UE encountered and logged. One or more correctable errors encountered but details on the correctable errors not logged.
UE	1	0	Two or more UEs encountered, details on the first logged.
UE	1	1	Two or more UEs encountered, details on the first logged. One or more corrected errors encountered but details on the corrected errors not logged.
CE	0	0	Single CE encountered and logged.
CE	0	1	Two or more CEs encountered; details on the first logged.

16.4 Error Trap Vectors

The following table describes the trap vectors used in UltraSPARC T2 to handle hardware errors.

TABLE 16-2 UltraSPARC T2 Error Traps

Trap Vector	Trap Type	Trap Class	UltraSPARC T2-specific	Trap Priority Level	Remarks
power_on_reset (warm reset)	01_{16}	Reset	N	0	Used for fatal errors. Error state captured in L2 ESR/EAR.
store_error	07_{16}	Deferred	Y	2.1	Used for store buffer errors. Error state captured in DFESR.
instruction_access_error	$0A_{16}$	Precise	N	4	Used for instruction access errors. Error state captured in ISFSR or L2 ESR/EAR/NDER.

TABLE 16-2 UltraSPARC T2 Error Traps

Trap Vector	Trap Type	Trap Class	UltraSPARC T2-specific	Trap Priority Level	Remarks
internal_processor_error	29_{16}	Precise	N	8.2	Used for all IRF/FRF errors (IRFU/IRFC/FRFU/FRFC) except those on the second or subsequent passes of a multicycle operation (partial store, compare and swap, block store). Used for store buffer bypass errors (SBDLU/SBDLC) and register array errors (MRAU/TSAU/TSAC/SCAU/SCAC/tcup/TCCP). Error state captured in DSFSR/DSFAR.
internal_processor_error	29_{16}	Precise	N	12.10	Used for IRF/FRF errors (IRFU/IRFC/FRFU/FRFC) on the second or subsequent passes of a multicycle operation (partial store, compare and swap, block store). Error state captured in DSFSR/DSFAR.
data_access_error	32_{16}	Precise	N	12.9	Used for data access errors. Error state captured in DSFSR or L2 ESR/EAR/NDER.
sw_recoverable_error	40_{16}	Disrupting	N	33.1	Used for disrupting errors not corrected and not cleared by hardware. Error state captured in DESR or L2 ESR/EAR/NDER.
hw_corrected_error	63_{16}	Disrupting	N	33.2	Used for disrupting errors corrected and cleared by hardware. Error state captured in DESR or L2 ESR/EAR/NDER.
instruction_access_MMU_error	71_{16}	Precise	N	2.7	Used for IMMU errors. Error state captured in ISFSR or L2 ESR/EAR/NDER.
data_access_MMU_error	72_{16}	Precise	N	12.2	Used for DMMU errors. Error state captured in DSFSR/DSFAR or L2 ESR/EAR/NDER.

To conform to the latest RAS specification, UltraSPARC T2 uses a different trap vector for each type of hardware error (disrupting, precise, or deferred). There is no unique trap vector for NotData errors. Instead NotData is indicated by an I-SFSR, D-SFSR, or DESR encoding.

16.5 Error Barrier

A MEMBAR #Sync acts as an error barrier on UltraSPARC T2. Before a MEMBAR #Sync completes, all previous, enabled precise and deferred error traps will be taken by the strand. Disrupting errors related to the instruction execution stream also treat MEMBAR #Sync as a memory barrier (with a possible skid of one instruction as described below). Disrupting errors unrelated to the instruction stream do not have a memory barrier operation on UltraSPARC T2.

The following lists the behavior for errors that treat MEMBAR #Sync as an error barrier:

- Any precise error that occurs (and is not masked) will cause a trap before the MEMBAR #Sync executes, and TPC will be the PC of the instruction with the precise error.
- For deferred errors, if the error is detected before the MEMBAR #Sync executes, the trap will be taken before the MEMBAR #Sync executes, and the TPC will be the PC of the MEMBAR #Sync or the PC of some earlier instruction. If the deferred error is detected during the execution of the MEMBAR #Sync (a MEMBAR #Sync causes the store buffer to drain), then the trap will be taken after execution of the MEMBAR #Sync, and the TPC will be the NPC of the MEMBAR #Sync.
- The disrupting errors listed below are related to the instruction stream. If the error is detected before the MEMBAR #Sync executes and the trap is not masked, the trap will be taken either before the MEMBAR #Sync executes, or on the first instruction after the MEMBAR #Sync. The TPC will be the PC of the MEMBAR #Sync, the PC of some earlier instruction, or the PC of the instruction after the MEMBAR #Sync. If the error is detected during the execution of the MEMBAR #Sync and the trap is not masked, then the trap will be taken after executing the instruction following the MEMBAR #Sync, and the TPC will be the NPC of the instruction following the MEMBAR #Sync.
  - Instruction Cache Valid bit Parity
  - Instruction Cache Tag Parity
  - Instruction Cache Tag Multiple
  - Instruction Cache Data Parity
  - Data Cache Valid bit Parity
  - Data Cache Tag Parity
  - Data Cache Tag Multiple
  - Data Cache Data Parity
  - Store Buffer Data PCX read Correctable ECC
  - Store Buffer Data PCX read Uncorrectable ECC
  - IT L2 Correctable
  - IC L2 Correctable
  - DT L2 Correctable
  - DC L2 Correctable

The following disrupting errors are not related to the instruction stream. So MEMBAR #Sync has no relationship to these errors and therefore doesn't act as a barrier for these errors. If the error is detected before the MEMBAR #Sync executes and the trap is not masked, the trap will be taken either before the MEMBAR #Sync executes or on the first instruction after the MEMBAR #Sync. The TPC will be the PC of the MEMBAR #Sync, the PC of some earlier instruction, or the PC of the instruction after the MEMBAR #Sync. If the error is detected during the execution of the MEMBAR #Sync and the trap is not masked, then the trap will be taken after executing the instruction following the MEMBAR #Sync, and the TPC will be the

NPC of the instruction following the MEMBAR #Sync. If the error is detected after the MEMBAR #Sync completes, the trap (if not masked) will be taken on some later instruction.

- Tick_compare correctable disrupting
- Tick_compare uncorrectable disrupting
- L2 correctable ECC error
- L2 uncorrectable ECC error
- L2 NotData error
- SOC correctable
- SOC uncorrectable

16.6 Virtual Processor Error Handling Overview

The core RAS architecture has two objectives:

1. Provide a FIT rate sufficient for the target market at minimal core cost.
2. Conform to the architecture described by the SPARC SWG RAS & Error Handling Working Group, with implementation suggestions from UltraSPARC Architecture 2007 specification.

To meet both these objectives, UltraSPARC T2 generally does not correct or retry correctable errors detected by hardware. Any hardware-detected error that can be attributed to an instruction's execution results in a precise trap if enabled. Software at the trap handler diagnoses the error. If the error is correctable, it attempts recovery, for example by correcting a single-bit error in a register file. Software then retries the instruction. If another failure occurs, software can determine the appropriate action to take. This approach minimizes the hardware dedicated to error-specific handling flows in UltraSPARC T2. Software is provided with sufficient information to enable error recovery.

16.6.1 Error Status Registers

The UltraSPARC T2 core records errors in five per-strand error status registers (ESRs): I-SFSR (Instruction Synchronous Fault Status register), D-SFSR (Data Synchronous Fault Status register), D-SFAR (Data Synchronous Fault Address register), DESR (disrupting ESR), and DFESR (deferred ESR).

UltraSPARC T2 cores detect errors as they occur but pipeline precise error indications along with the instruction until commit time. At this time a precise error is recorded in either the I-SFSR or D-SFSR and D-SFAR. The error is only recorded if the appropriate CERER bit is set. In addition, for certain precise errors, the SETER.pscce bit must also be set for the error to be recorded.

Disrupting errors are recorded only if the appropriate CERER bit is set.

Deferred errors are always recorded.

16.6.2 Error Summary

TABLE 16-3 lists the errors the UltraSPARC T2 core detects and also lists salient information about each error. Columns in TABLE 16-3 are explained in the following table.

Column	Terminology
Error Type	Specifies the type of error, as follows: **CE:** hardware-corrected error **UEr:** hardware-uncorrected error that is software recoverable **UEU**: hardware-uncorrected error that is software unrecoverable (**Note:** Software unrecoverable does not imply that the error is fatal).
Trap Type	Specifies the type of trap that is caused, as follows: **P:** precise—logged in either the I-SFSR or the D-SFSR **D:** disrupting—logged in the DESR **Df:** deferred error—results in a *store_error* trap to the virtual processor that detected the error.
Trap Vector	Describes the trap vector to which the processor will be directed when an error occurs. **IAME:** *instruction_access_MMU_error* **IA:** *instruction_access_error* **IPE:** *internal_processor_error* **DAM:** *data_access_MMU_error* **DAE:** *data_access_error* **HCE:** *hw_corrected_error* **SRE:** *sw_recoverable_error* **SE:** *store_error*. UltraSPARC T2 directs disrupting traps and precise traps to different vectors. Software can then inspect the I-SFSR, D-SFSR, or DESR to determine more information regarding the failure.
Maskable	Defines which errors can be masked by SETER.pscce{62}, de{61} or dhcce{60} bits, as follows: **N:** Errors are not maskable by SETER bits. **Number** (62, 61, or 60)**:** Specifies which bit of SETER masks the error trap.

TABLE 16-3 UltraSPARC T2 Processor Error Types *(1 of 5)*

Error	Access Type or Unit	Error Type	Trap Type	Trap Vector	Maskable	ESR ErrorAddr Info	HW Corrected	HW Cleared	Notes
IT tag multiple hit (ITTM)	IFetch	UER	P	IAME	N	ISFSR.ittm	N	N	HW does not invalidate entry; SW should demap all; cannot distinguish between same or different contexts.
ITLB tag parity (ITTP)	IFetch	UER	P	IAME	N	ISFSR.ittp	N	N	HW does not invalidate entry; SW should log tag and data for all entries and demap page.
ITLB data parity (ITDP)	IFetch	UER	P	IAME	N	ISFSR.itdp	N	N	HW does not invalidate entry; SW should demap page.
ITLB MRA uncorrectable (ITMU)	IFetch	UER	P	IAME	N	MRA index{2:0}	N	N	SW should correct MRA by reloading.
ITLB L2 correctable (ITL2C)	IFetch	UER	D	SRE	61	Recorded in L2 ESR	Y	N	SW should correct L2 error.
ITLB L2 uncorrectable (ITL2U)	IFetch	UEU	P	IAME	N	Recorded in L2 ESR	N	N	SW should correct L2 error if possible.
ITLB L2 NotData (ITL2ND)	IFetch	UEU	P	IAME	N	Recorded in L2 ESR	N	N	NotData should have been preceded by a UE.
Icache valid bit (ICVP)	IFetch	CE	D	HCE	60	Index{5:0}, way{2:0}	Y	Y	HW invalidates all ways of I$ index; HW refetches.
Ic tag multiple hit (ICTM)	IFetch	CE	D	HCE	60	Index{5:0}, way{2:0}	Y	Y	HW invalidates all ways of I$ index; HW refetches.
Icache tag parity (ICTP)	IFetch	CE	D	HCE	60	Index{5:0}, way{2:0}	Y	Y	HW invalidates all ways of I$ index; HW refetches.
ICache data parity (ICDP)	IFetch	CE	D	HCE	60	Index{5:0}	Y	Y	HW invalidates all ways of I$ index; HW refetches.
Icache L2 correctable (ICL2C)	IFetch	UER	D	SRE	61	Recorded in L2 ESR	Y	N	SW should correct L2 error.
Icache L2 uncorrectable (ICL2U)	IFetch	UEU	P	IAE	62	Recorded in L2 ESR	N	N	SW should correct L2 error if possible.

TABLE 16-3 UltraSPARC T2 Processor Error Types *(2 of 5)*

Error	Access Type or Unit	Error Type	Trap Type	Trap Vector	Maskable	ESR ErrorAddr Info	HW Corrected	HW Cleared	Notes
Icache L2 NotData (ICL2ND)	IFetch	UEU	P	IAE	62	Recorded in L2 ESR	N	N	NotData should have been preceded by a UE.
IRF correctable ECC error (IRFC)	EXU	UER	P	IPE	62	GL{1:0}, Index{4:0}, ECC Syndrome{7:0}	N	N	SW should correct and retry.
IRF uncorrectable ECC error (IRFU)	EXU	UEU	P	IPE	62	GL{1:0}, Index{4:0}, ECC Syndrome{7:0}	N	N	
FRF correctable ECC error (FRFC)	FGU	UER	P	IPE	62	Index{5:0}, (2) ECC Syndrome{6:0}	N	N	SW should correct and retry.
FRF uncorrectable ECC error (FRFU)	FGU	UEU	P	IPE	62	Index{5:0}, (2) ECC Syndrome{6:0}	N	N	
DTLB tag parity (DTTP)	Load	UER	P	DAME	N	VA{47:0} in D-SFAR	N	N	HW does not invalidate entry; SW should demap page
DT tag multiple hit (DTTM)	Load	UER	P	DAME	N	VA{47:0} in D-SFAR	N	N	HW does not invalidate all entries, SW should demap all; can not distinguish between same or different contexts
DTLB data parity (DTDP)	Load	UER	P	DAME	N	VA{47:0} in D-SFAR	N	N	HW does not invalidate entry; SW should demap page
DTLB MRA uncorrectable (DTMU)	Load	UER	P	DAME	N	MRA index{2:0}	N	N	SW can reload MRA
DTLB L2 correctable (DTL2C)	Load	UER	D	SRE	61	Recorded in L2 ESR	Y	N	SW should correct L2 error
DTLB L2 uncorrectable (DTL2U)	Load	UEU	P	DAME	N	Recorded in L2 ESR	N	N	SW should correct L2 error if possible
DTLB L2 NotData (DTL2ND)	Load	UEU	P	DAME	N	Recorded in L2 ESR	N	N	NotData should have been preceded by a UE
Dcache valid bit (DCVP)	Load	CE	D	HCE	60	Index{6:0}, way{1:0}	Y	Y	HW invalidates all ways of D$ index; HW refetches data; if trap taken, SW can log

TABLE 16-3 UltraSPARC T2 Processor Error Types *(3 of 5)*

Error	Access Type or Unit	Error Type	Trap Type	Trap Vector	Maskable	ESR ErrorAddr Info	HW Corrected	HW Cleared	Notes
Dcache tag parity (DCTP)	Load	CE	D	HCE	60	Index{6:0}, way{1:0}	Y	Y	HW invalidates all ways of D$ index; HW refetches data; if trap taken, SW can log
Dcache tag multiple hit (DCTM)	Load	CE	D	HCE	60	Index{6:0}, way{1:0}	Y	Y	HW invalidates all ways of D$ index; HW refetches data; if trap taken, SW can log
Dcache data parity (DCDP)	Load	CE	D	HCE	60	Index{6:0}, way{1:0}	Y	Y	HW invalidates all ways of D$ index; HW refetches data; if trap taken, SW can log
Dcache L2 Correctable (DCL2C)	Load	UER	D	SRE	61	Recorded in L2 ESR	Y	N	SW should correct L2 error.
Dcache L2 Uncorrectable (DCL2U)	Load	UEU	P	DAE	62	Recorded in L2 ESR	N	N	SW should correct L2 error if possible.
Dcache L2 NotData (DCL2ND)	Load	UEU	P	DAE	62	Recorded in L2 ESR	N	N	NotData should have been preceded by a UE.
Store buffer data load hit correctable ECC (SBDLC)	RAW (Read-After-Write)	UER	P	IPE	62	STB index{2:0}	N	N	SW should MEMBAR #Sync, then retry.
Store buffer data load hit uncorrectable ECC (SBDLU)	RAW	UEU	P	IPE	62	STB index{2:0}	N	N	SW should MEMBAR #Sync, then retry (likely a nonrecoverable error).
Store buffer data PCX read or ASI store correctable (SBDPC)	PCX	CE	D	HCE	60	STB index{2:0}	Y	Y	SW to log.
Store buffer data PCX read (SBDPU)	PCX	UEU	D	SRE	61	STB index{2:0}	N	N	HW generates NotData.
Store buffer address PCX read or ASI store parity (SBAPP)	PCX	UEU	DF	SE	N	DFESR bit 61, privilege level, STB index{2:0}	N	N	Nonmaskable, causes *store_error* trap.

TABLE 16-3 UltraSPARC T2 Processor Error Types *(4 of 5)*

Error	Access Type or Unit	Error Type	Trap Type	Trap Vector	Maskable	ESR ErrorAddr Info	HW Corrected	HW Cleared	Notes
Store buffer data I/O read and ASI store uncorrectable (SBDIOU)	PCX	UEU	DF	SE	N	DFESR bit 60, privilege level, STB index{2:0}	N	N	Nonmaskable, causes *store_error* trap.
TSA correctable (TSAC)	TLU	UER	P	IPE	62	TSA index{2:0}, syndrome	N	N	SW can correct or reload TSA.
TSA uncorrectable (TSAU)	TLU	UER	P	IPE	62	TSA index{2:0}, syndrome	N	N	SW can correct or reload TSA.
MRA uncorrectable (MRAU)	MMU	UER	P	IPE (ASI read or read-modify-write)	62	MRA index{2:0}	N	N	SW can correct or reload MRA.
SCA correctable (SCAC)	MMU	UER	P	IPE	62	SCA index{2:0}, syndrome	N	N	SW can correct SCA.
SCA uncorrectable (SCAU)	MMU	UEU	P	IPE	62	SCA index{2:0}, syndrome	N	N	
TICK_CMPR correctable precise (TCCP)	TLU	UER	P	IPE	62	TCA index{1:0}, syndrome	N	N	SW can correct or reload TICK_CMPR.
TICK_CMPR correctable disrupting (TCCD)	TLU	UER	D	SRE	61	TCA index{1:0}, syndrome	N	N	SW can correct or reload TICK_CMPR.
TICK_CMPR uncorrectable precise (TCUP)	TLU	UER	P	IPE	62	TCA index{1:0}, syndrome	N	N	SW can reload TICK_CMPR.
TICK_CMPR uncorrectable disrupting (TCUD)	TLU	UER	D	SRE	61	TCA index{1:0}, syndrome	N	N	SW can reload TICK_CMPR.
L2C	L2	CE	D	HCE	61	Recorded in L2 ESR	Y	N	See Section 16.9.
L2U	L2	UEU	D	SRE	61	Recorded in L2 ESR	N	N	See Section 16.9.
L2ND	L2	UEU	D	SRE	61	Recorded in L2 ESR	N	N	See Section 16.9.

TABLE 16-3 UltraSPARC T2 Processor Error Types *(5 of 5)*

Error	Access Type or Unit	Error Type	Trap Type	Trap Vector	Maskable	ESR ErrorAddr Info	HW Corrected	HW Cleared	Notes
SOC Correctable (SOCC)	SOC	CE	D	HCE	61	Recorded in SOC Error Status Register	Y	Y	Software must read SOC ESRs to determine details of the error.
SOC Uncorrectable (SOCU)	SOC	UEU	D	SRE	61	Recorded in SOC Error Status Register	N	N	Software must read SOC ESRs to determine details of the error.

16.7 SPARC Error Descriptions

16.7.1 ITLB Errors

The priority of ITLB errors is ITTM → ITTP → ITDP → ITMU.

16.7.1.1 ITLB Tag Multiple Hit Error (ITTM)

The UltraSPARC T2 ITLB checks for multiple tag hits on each access if CERER.ittm is set. A multiple tag hit error has higher priority than a tag parity error, and parity is not factored into the tag hit determination. A multiple hit can occur as a result of a hardware failure or a software error. Each ITLB tag entry is a tuple consisting of partition ID, real address indicator, context, and VA, adjusted for page size. A hardware error in any of these fields can cause a multiple tag hit. A multiple hit can also occur if software maps the same virtual address using different contexts[1], loads the ITLB with those mappings simultaneously, and sets each context register to point to one of the two contexts. A multiple hit can also occur if software loads pages with differing pages sizes that do not cause the first page to be autodemapped1.

When a multiple hit error is detected, hardware records the error in the I-SFSR, and a precise *instruction_access_MMU_error* trap is taken. The VA of the instruction fetch is recorded in TPC[TL].

1. The autodemap feature of the TLB, described in *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147, prevents translations with identical page sizes, VA, and context from existing in the TLB simultaneously. However, software can generate multiple matches by inserting overlapping translations of differing page sizes, or by inserting translations that differ only in context, and then programming the context 0 and context 1 registers to match the pair of translations.

Programming Notes

To handle an ITTM error, software at the trap handler logs the error, and issues either a demap_all or a sequence of demap_pages to all active contexts. Then software issues a RETRY instruction. UltraSPARC T2 refetches the instruction and reaccesses the ITLB. This time either a hit will occur (if the translation was reloaded by another strand) or a miss will occur. If an ITLB miss occurs, hardware retranslates the address and reloads the ITLB.

UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.1.2 ITLB Tag Parity Error (ITTP)

Each ITLB tag entry is protected with parity. The tag entry bits covered by parity are listed in *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147.

ITLB tag parity is checked with each instruction translation if CERER.ittp is set. ITLB tag parity is not checked on loads to `ASI_ITLB_TAG_READ_REG`. When a parity error is detected, the error is logged in the I-SFSR and the strand takes a precise *instruction_access_MMU_error* trap. The VA of the instruction fetch is recorded in TPC[TL]. Software at the trap handler logs the error and issues a demap_page to the TPC[TL]. Then software issues a RETRY instruction. UltraSPARC T2 refetches the instruction and reaccesses the ITLB. This time either a hit will occur (if the translation was reloaded by another strand), or a miss will occur. If an ITLB miss occurs, hardware retranslates the address and reloads the ITLB.

Implementation Notes

Parity is only checked for translations that hit in the TLB. This implies that a page that would otherwise hit in the TLB may experience a TLB miss instead (for example., if a VA bit has flipped). Normal replacement will eventually remove such a TLB entry, or a page that would otherwise miss in the TLB may experience an ITTP error instead, due to a VA bit flip.

If a parity error occurs in the page size stored in the TTE data, a TLB tag parity error may result. This can occur since an error in the page size will compute incorrect parity for the tag entry. In this case, both tag and data parity errors are detected, but since tag parity errors have higher priority than data parity errors, only the tag parity error exception is taken and recorded.

Programming Note UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.1.3 ITLB Data Parity Error (ITDP)

Each ITLB data entry is protected with parity. The data entry bits covered by parity are listed in *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147.

ITLB data parity is checked with each instruction translation if CERER.itdp is set. ITLB data parity is not checked on loads to ASI_ITLB_DATA_ACCESS_REG. When a parity error is detected, hardware records the error in the I-SFSR, and takes a precise *instruction_access_MMU_error* trap. The VA of the instruction fetch is recorded in TPC[TL]. Software at the trap handler logs the error and issues a demap_page to the TPC[TL]. Then software issues a RETRY instruction. UltraSPARC T2 refetches the instruction and reaccesses the ITLB. This time either a hit will occur (if the translation was reloaded by another strand) or a miss will occur. If an ITLB miss occurs, hardware retranslates the address and reloads the ITLB.

Programming Note UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.1.4 ITLB MMU Register Array Uncorrectable Error (ITMU)

The MMU register array (MRA) is protected by parity. When hardware tablewalk is enabled, MRA parity is checked for all translations that do not have an entry present in the ITLB during the ITLB reload process. If CERER.hwtwmu is set and a parity error is encountered, hardware records the error in the I-SFSR and takes a precise *instruction_access_MMU_error* trap. The VA of the instruction fetch is recorded in TPC[TL], while the failing MRA index is recorded in the D-SFAR. To recover from an ITMU error, software can reload the MRA register with a "clean" copy of the MRA data which it has stored elsewhere.

16.7.2 DTLB Errors

The priority of DTLB errors is DTTM → DTTP → DTDP → DTMU.

16.7.2.1 DTLB Tag Multiple Hit Error (DTTM)

The UltraSPARC T2 DTLB checks for multiple tag hits on each access (including prefetch instructions) if CERER.dttm is set. A multiple tag hit error has higher priority than a tag parity error, and parity is not factored into the tag hit determination. A multiple hit can occur as a result of a hardware failure or a software error. Each DTLB tag entry is a tuple consisting of partition ID, real address indicator, context, and VA, adjusted for page size. A hardware error in any of these fields can cause a multiple tag hit. A multiple hit can also occur if software maps the same virtual address using different contexts,[1] loads the DTLB with those mappings simultaneously, and sets each context register to point to one of the two contexts. A multiple hit can also occur if software loads pages with differing pages sizes that do not cause the first page to be autodemapped.[1]

When a multiple hit error is detected, hardware records the error in the D-SFSR, and takes a precise *data_access_MMU_error* trap. The VA of the data access is recorded in D-SFAR.

Programming Notes

To handle a DTTM error, software at the trap handler logs the error, and issues either a demap_all or a sequence of demap_pages to all active contexts. Then software issues a retry instruction. UltraSPARC T2 refetches the instruction and reaccesses the DTLB. This time either a hit will occur (if the translation was reloaded by another strand) or a miss will occur. If a DTLB miss occurs, hardware retranslates the address and reloads the DTLB.

UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.2.2 DTLB Tag Parity Error (DTTP)

Each DTLB tag entry is protected with parity. The tag entry bits covered by parity are listed in *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147.

DTLB tag parity is checked with each data translation (including prefetch instructions) if CERER.dttp is set. DTLB tag parity is not checked on loads to ASI_DTLB_TAG_READ_REG. When a parity error is detected, hardware records the

1. The autodemap feature of the TLB, described in Section 12.10.15, prevents translations with identical page sizes, VA, and context from existing in the TLB simultaneously. However, software can generate multiple matches by inserting overlapping translations of differing page sizes, or by inserting translations that differ only in context and then programming the context 0 and context 1 registers to match the pair of translations.

error in the D-SFSR and takes a precise *data_access_MMU_error* trap. The VA of the data access is recorded in the D-SFAR. Software at the trap handler logs the error and issues a demap_page to the D-SFAR. Then software issues a RETRY instruction. UltraSPARC T2 refetches the instruction and reaccesses the DTLB. This time either a hit will occur (if the translation was reloaded by another strand) or a miss will occur. If a DTLB miss occurs, hardware retranslates the address and reloads the DTLB.

Implementation Notes

Parity is only checked for data translations that hit in the TLB. This implies that a page that would otherwise hit in the TLB may experience a TLB miss instead (for example, if a VA bit has flipped). Normal replacement will eventually remove such a TLB entry, or a page that would otherwise miss in the TLB may experience a DTTP error instead due to a VA bit flip.

If a parity error occurs in the page size stored in the TTE data, a TLB tag parity error may result. This can occur since an error in the page size will compute incorrect parity for the tag entry. In this case, both tag and data parity errors are detected, but since tag parity errors have higher priority than data parity errors, only the tag parity error exception is taken and recorded.

Programming Note

UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.2.3 DTLB Data Parity Error (DTDP)

Each DTLB data entry is protected with parity. The data entry bits covered by parity are listed in *I-/D-TLB Data-In/Data-Access/Tag-Read Registers* on page 147.

DTLB data parity is checked with each data translation (including prefetch instructions) if CERER.dtdp is set. DTLB data parity is not checked on loads to ASI_DTLB_DATA_ACCESS_REG. When a parity error is detected, hardware records the error in the D-SFSR, and takes a precise *data_access_MMU_error* trap. The VA of the data access is recorded in D-SFAR. Software at the trap handler logs the error, and issues a demap_page to the D-SFAR. Then software issues a retry instruction. UltraSPARC T2 refetches the instruction and reaccesses the DTLB. This time either a hit will occur (if the translation was reloaded by another strand), or a miss will occur. If a DTLB miss occurs, hardware retranslates the address and reloads the DTLB.

Programming Note UltraSPARC T2 provides no special hardware to deal with permanent single-bit failures in a TLB entry. To work around a permanent single-bit failure, software must identify the bad TLB entry, then disable hardware tablewalking and manage the TLB in software to avoid the bad TLB entry.

16.7.2.4 DTLB MMU Register Array Uncorrectable Error (DTMU)

The MMU register array (MRA) is protected by parity. When hardware tablewalk is enabled, MRA parity is checked for all translations that do not have an entry present in the DTLB during the DTLB reload process. If CERER.hwtwmu is set and a parity error is encountered, hardware records the error in the D-SFSR, and takes a precise *data_access_MMU_error* trap. The failing MRA index is recorded in the D-SFAR. To recover from an DTMU error, software can reload the MRA register with a "clean" copy of the MRA data which it has stored elsewhere.

16.7.3 Icache Errors

The priority of Icache errors is ICVP > ICTP > ICTM > ICDP.

16.7.3.1 Icache Valid Parity Error (ICVP)

The Icache tag valid bits are protected by duplication: there is a *master* copy and a *slave* copy. The valid bits are checked for each instruction cache access for all ways in the accessed set if CERER.icvp is set.

If CERER.icvp is not set, the *master* copy of the valid bit simply determines whether a cache hit or miss occurred. If a valid bit mismatch occurs and CERER.icvp is set, hardware invalidates all ways in the accessed Icache set.

If the DESR.f bit is clear, hardware records the index and the way[1] with the error in the DESR ErrorAddress field, and sets the DESR.f and DESR.icvp bits. If the DESR.f bit is set, hardware sets the DESR.me bit but does not record the index and way in the DESR.

A disrupting *hw_corrected_error* trap is generated to the requesting virtual core when SETER.dhcce is set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap. The earliest instruction on which the disrupting trap can be taken is the instruction after the one that encountered the error.

1. The way information captured may not be correct for an instruction in a control transfer delay slot.

Implementation Note Due to an implementation bug, the normal forwarding of the instruction around the Icache (used to guarantee forward progress with multiple strands) does **not** guarantee forward progress for the case where the valid bit has a permanent error.

16.7.3.2 Icache Tag Parity Error (ICTP)

The Icache tag is protected with parity. The Icache tag parity is checked with each instruction fetch if CERER.ictp is set. The parity is checked for each of the possible valid ways in the set. If a parity error is found in any of the ways, hardware encodes ictp and captures the way[1] and set information in the DESR register. Hardware invalidates all ways in the accessed Icache set.

A disrupting *hw_corrected_error* trap is generated to the requesting virtual core when SETER.dhcce is set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap. The earliest instruction on which the disrupting trap can be taken is the instruction after the one that encountered the error.

Implementation Note The normal forwarding of the instruction around the Icache (used to guarantee forward progress with multiple strands) will also guarantee forward progress for the case where the tag has a permanent error.

16.7.3.3 Icache Tag Multiple Hit Error (ICTM)

Multiple errors in the tags can lead to multiple hits within a set. If CERER.ictm is set, each instruction fetch is checked for multiple hits. If multiple hits are detected on an instruction fetch, hardware encodes ictm and captures one of the ways that hit[2] and the set information in the DESR register. Hardware invalidates all ways in the accessed Icache set.

A disrupting *hw_corrected_error* trap is generated to the requesting virtual core when SETER.dhcce is set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap. The earliest instruction on which the disrupting trap can be taken is the instruction after the one that encountered the error.

Implementation Note The normal forwarding of the instruction around the Icache (used to guarantee forward progress with multiple strands) will also guarantee forward progress for the case where the tag has a permanent error.

---

1. The way information captured may not be correct for an instruction in a control transfer delay slot.
2. The way information captured may not be correct for an instruction in a control transfer delay slot.

16.7.3.4 Icache Data Parity Error (ICDP)

Each 32-bit instruction in the Icache data array and the instruction buffers is protected with parity. The instruction data parity is checked with each instruction decode if CERER.icdp is set. Hardware can not distinguish between a parity error that occurs in the instruction cache data array and a parity error that occurs in the instruction buffer.

When a parity error is detected, hardware encodes icdp and captures the set information in the DESR register. Hardware invalidates all ways in the accessed Icache set. If the parity error actually occurred in the instruction buffer and not in the data array, the set information in the DESR is irrelevant.

A disrupting *hw_corrected_error* trap is generated to the requesting virtual core when SETER.dhcce is set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap. The earliest instruction on which the disrupting trap can be taken is the instruction after the one that encountered the error.

Implementation Notes The normal forwarding of the instruction around the Icache (used to guarantee forward progress with multiple strands) will also guarantee forward progress for the case where the data has a permanent error.

16.7.4 Dcache Errors

The priority of Dcache errors is DCVP > DCTP > DCTM > DCDP.

Programming Note If the D$ is in direct-mapped mode, and one of the following parity errors occurs, the way replaced will be the way which experienced the parity error, not the way specified by address bits [12:11].

16.7.4.1 Dcache Valid Parity Error (DCVP)

The Dcache tag valid bits are protected by duplication: there is a *master* copy and a *slave* copy. The valid bits are checked for each data load for all ways in the accessed set if CERER.dcvp is set. If CERER.dcvp is not set, the *master* copy of the valid bit simply determines whether a cache hit or miss occurred. If a valid bit mismatch occurs and CERER.dcvp is set, hardware invalidates all ways in the accessed set. If the DESR.f bit is clear, hardware records the index and the way with the error in the DESR ErrorAddress field, and sets the DESR.f and DESR.dcvp bits. If the DESR.f bit is set, hardware sets the DESR.me bit but does not record the index and way in the DESR. In addition, if the SETER.dhcce bit is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *hw_corrected_error* trap is generated to the

requesting virtual processor for logging of the error. If the *hw_corrected_error* trap is not taken, the error is remembered via DESR.f and a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor when SETER.dhcce is later set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap.

Implementation Notes

The normal forwarding of the data around the Dcache guarantees forward progress for the case where the valid bit has a permanent error.

Because stores do not read the Dcache they do not check the Dcache valid parity. Stores do not read the Dcache since coherency is maintained by the L2 directory.

16.7.4.2 Dcache Tag Parity Error (DCTP)

The Dcache tag is protected with parity. The Dcache tag parity is checked with each data load if CERER.dctp is set. If CERER.dctp is not set, the error is ignored. The parity is checked for each of the possible valid ways in the set. If a parity error is found in any of the ways, hardware invalidates all ways in the accessed set, encodes dctp and captures the way and set information in the DESR register. In addition, if the SETER.dhcce bit is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor for logging of the error. If the *hw_corrected_error* trap is not taken, the error is remembered via DESR.f and a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor when SETER.dhcce is later set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher-priority trap.

Implementation Notes

The normal forwarding of the data around the Dcache guarantees forward progress for the case where the tag has a permanent error.

Because stores do not read the Dcache they do not check the Dcache tag parity. Stores do not read the Dcache since coherency is maintained by the L2 directory.

16.7.4.3 Dcache Tag Multiple Hit Error (DCTM)

Multiple errors in the tags can lead to multiple hits within a set. If CERER.dctm is set, each data load checks for multiple hits. If multiple hits are detected on an data load, hardware invalidates all ways in the accessed set, encodes dctm and captures one of the ways that hit and the set information in the DESR register. In addition, if the SETER.dhcce bit is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor for logging of the error. If the *hw_corrected_error* trap is not taken, the error is

remembered via DESR.f and a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor when SETER.dhcce is later set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher priority trap.

Implementation Notes

The normal forwarding of the data around the Dcache guarantees forward progress for the case where the tag has a permanent error.

Because stores do not read the Dcache, they do not check the Dcache for multiple tag hits. Stores do not read the Dcache since coherency is maintained by the L2 directory.

16.7.4.4 Dcache Data Parity Error (DCDP)

Each byte in the Dcache data array is protected with parity. The Dcache data parity is checked for each of the valid ways with each data load if CERER.dcdp is set. When a parity error is detected, hardware invalidates all ways in the accessed set, encodes dcdp and captures the set information in the DESR register. In addition, if the SETER.dhcce bit is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor for logging of the error. If the *hw_corrected_error* trap is not taken, the error is remembered via DESR.f and a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor when SETER.dhcce is later set (and PSTATE.ie is set or HPSTATE.hpriv is clear) and there is no higher-priority trap.

Implementation Notes

The normal forwarding of the data around the Dcache guarantees forward progress for the case where the data has a permanent error.

Because stores do not read the Dcache, they do not check Dcache data parity. Stores do not read the Dcache since coherency is maintained by the L2 directory.

16.7.5 IRF ECC Error (IRFC and IRFU)

Each IRF entry is protected by SEC/DED ECC. Up to three operands can be read from the IRF at a time. Hardware checks each operand's ECC independently. Hardware prioritizes operand errors in the order rs1 > rs2 > rs3. This means that an uncorrectable error which occurs on a lower-priority operand (e.g., rs2) simultaneously with a correctable error on a higher-priority operand (e.g., rs1) will not be reported (until the correctable error is cleared by software and the instruction is retried).

Hardware can only detect an IRF ECC error if CERER.irf is set. If hardware detects either a correctable or uncorrectable error for any valid operand, what happens depends upon the setting of SETER.pscce.

If SETER.pscce is set, hardware records the error type in the D-SFSR by encoding IRFC or IRFU as appropriate, records the IRF index and the ECC syndrome in the D-SFAR (see Table 16-11 on page 248), and generates a precise *internal_processor_error* trap.

Programming Note

Software should correct an IRFC error before issuing a retry instruction. Software can correct the error as follows.

- It reads the D-SFAR contents and decodes the failing address location in the IRF, turns off the SETER.pscce bit, and reads the data. If desired for logging, software can also read the ECC bits by using an LDXA to the ASI_IRF_ECC_REG register (See *ASI_IRF_ECC_REG* on page 419).
- It decodes the ECC syndrome. If the error is in the data bits, it **xor**s the correction mask with the data read from the IRF.
- It then writes the corrected data into the IRF using a normal integer instruction (hardware generates the proper ECC prior to writing the IRF entry). In the process of reading the failing location, another IRF ECC error will occur, but will not be recorded or trapped as SETER.pscce is clear.

After correcting the data, software should turn the SETER.pscce bit back on. Then software can retry the original failing instruction.

An IRFU error is generally not recoverable.

If the SETER.pscce bit is not set, the error is not recorded, and hardware continues executing, using the uncorrected data read from the IRF. This will lead to data corruption.

Since up to three operands can be read for each instruction, software may define an appropriate error threshold for the instruction before considering the IRF broken.

16.7.6 FRF ECC Error (FRFC and FRFU)

Each FRF entry is protected by SEC/DED ECC. Up to two operands can be read from the FRF at a time. Hardware checks each operand's ECC independently. Note: PDIST reads 3 operands over 2 cycles, so hardware still prioritizes three operands for reporting errors. Hardware prioritizes operand errors in the order rs1 > rs2 > rs3. This means that an uncorrectable error which occurs on a lower-priority operand (e.g., rs2) simultaneously with a correctable error on a higher priority operand (e.g., rs1) will not be reported (until the correctable error is cleared by software and the instruction is retried).

Hardware can only detect an FRF ECC error if CERER.frf is set. If hardware detects either a correctable or uncorrectable error for any valid operand, what happens depends upon the setting of SETER.pscce.

If SETER.pscce is set, hardware records the error type by encoding FRFC or FRFU in the D-SFSR as appropriate, records the failing FRF index and ECC syndrome in the D-SFAR (see Table 16-11 on page 248), and generates a precise *internal_processor_error* trap.

Programming Note

Software should correct an FRFC error before issuing a retry instruction. Handling of an FRFC error is similar to an IRFC error. The additional complication is that each FRF entry contains two ECC words (due to the single-precision FP registers). So two corrections may have to be performed. Software can correct a correctable error as follows:

- It reads the failing FRF index and ECC syndrome from the D-SFAR. It decodes the failing address location in the FRF, turns off the SETER.pscce bit, and reads the data. If desired for logging, software can also read the ECC bits by using an LDXA to the `ASI_FRF_ECC_REG` register (See *ASI_FRF_ECC_REG* on page 420).
- Hardware reports 2 syndromes in the D-SFAR, even for single-precision FP operations (the syndromes correspond to the 2 SP registers of an even/odd SP pair). For a double-precision operation, both syndromes are pertinent. Hardware does not indicate which of the syndromes are pertinent for an SP operation. Both syndromes may in fact indicate an ECC error. It is also possible for one of the syndromes (the one corresponding to the even or odd SP register which was not read) to be incorrect.[1] Software should inspect the instruction at %tpc. If it was a double-precision operation, it should decode both syndromes. If it was a single-precision operation, it decodes the register numbers of the sources accordingly. That identifies which of the syndromes is pertinent.
- If an error occurred in the data bits, software **xor**s the correction mask with the data.
- It then writes the corrected data into the FRF using a normal FP operation (hardware generates the proper ECC before writing the FRF). In the process of reading the failing FRF location, another FRF ECC error will occur, but will not be recorded or trapped as SETER.pscce is clear.

After correcting the data, software should turn the SETER.pscce bit back on. Then software can retry the original failing instruction.

1. Hardware captures the syndromes in the D-SFAR at the time an instruction reads its operands from the FRF, without regard to previous single-precision operations in the pipeline which may update the value of one of the SP registers of an even/odd pair. If both the even and odd SP registers of an even-odd pair have ECC errors, hardware reports an error on the register being read, and the other syndrome is unneeded. Consider the following example. If the odd SP register of an even/odd pair is being overwritten by a single-precision instruction still in the pipeline when the instruction reading from the even register reads the FRF, hardware takes a trap on the instruction reading the even register. Since at the time of the trap and the capture of the syndromes in the D-SFAR, the instruction writing the odd register has not updated the register value, it still shows as having an ECC error.

An FRFU error is generally not recoverable.

If SETER.pscce is not set, the error is not recorded, and hardware continues executing, using the uncorrected data read from the FRF. This will lead to data corruption.

Software may define an appropriate error threshold for the instruction before considering the FRF broken.

16.7.7 Store Buffer

The Store Buffer (STB) is organized as a CAM which contains the tag portion of the address and a RAM which contains the data and status bits. The status bits consist of the privilege level of the store. UltraSPARC T2 implements ECC in the data array for data bits, and the Cam bits are protected by a single parity bit. The store buffer is accessed on data loads (to check for RAW (Read-After-Write) hits) and on PCX reads[1] and ASI ring stores.[2] It can also be accessed with diagnostic reads, but these accesses do not cause parity or ECC errors.

16.7.7.1 Correctable Data ECC Error on a Load (SBDLC)

Hardware detects this error only if CERER.sbdlc is set. If a load which results in a full RAW hit in the STB gets a single-bit data error, hardware does not correct the load data.

---

1. A PCX read occurs when the store is sent to the L2 cache or NCU. PCX stands for Processor to Cache Xbar.

2. Stores to ASI space which go over the ASI ring internal to the processor are referred to as ASI ring stores.

If SETER.pscce is set, hardware records the error in the D-SFSR by encoding sbdlc, writes the store buffer index of the entry in error in the D-SFAR, and generates a precise *internal_processor_error* trap. Since hardware will correct the data before writing the store data to memory, this error is likely recoverable; software can issue a retry to reexecute the load.

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

16.7.7.2 Uncorrectable Data ECC Error on a Load (SBDLU)

Hardware detects this error only if CERER.sbdlu is set. If a load which results in a full RAW hit in the STB gets an uncorrectable data ECC error, the following flow occurs.

If SETER.pscce is set, the error is recorded in the D-SFSR by encoding sbdlu, and writing the store buffer index of the entry in error to the D-SFAR, and generates a precise *internal_processor_error* trap. Another uncorrectable error will likely occur when hardware reads the store buffer entry to write the store data to memory. (See *Uncorrectable Data ECC Error on a PCX Read to Memory (SBDPU)* below.)

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

16.7.7.3 STB Address Parity Error on a Load

UltraSPARC T2 does not check CAM parity for load accesses.

16.7.7.4 Correctable Data ECC Error on a PCX Read to Memory or I/O or Read for an ASI Ring Store (SBDPC)

If CERER.sbdpc is set, on a PCX read to memory or I/O space or a read for an ASI ring store which results in a single bit ECC error, hardware corrects the error before forwarding the data to the crossbar or the ASI ring. Hardware encodes sbdpc and writes the failing store buffer index to the DESR.

If SETER.dhcce is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware presents a disrupting *hw_corrected_error* trap to the core.

If SETER.dhcce is not set, hardware continues executing. Assuming software has not reset DESR.f, a disrupting trap will be presented to the core when software sets SETER.dhcce (and PSTATE.ie is set or HPSTATE.hpriv is clear).

16.7.7.5 Uncorrectable Data ECC Error on a PCX Read to Memory (SBDPU)

If CERER.sbdpu_sbdiou is set, on a PCX read to memory which results in an uncorrectable ECC error, hardware generates NotData before forwarding the data to the crossbar. The error is recorded in the DESR by encoding sbdpu, and writing the store buffer index to the DESR.

If SETER.de is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware presents a disrupting *sw_recoverable_error* trap to the core.

If SETER.de is not set, hardware continues executing. Note that if SETER.de is not set, hardware has performed a bad store which will not be detected until the bad value is loaded. When software sets SETER.de (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware will present a disrupting *sw_recoverable_error* trap to the core.

16.7.7.6 Uncorrectable Data ECC Error on a PCX Read to I/O or Read for an ASI Ring Store (SBDIOU)

If CERER.sbdpu_sbdiou is set, on a PCX read to I/O space or read for ASI Ring Store which results in an uncorrectable ECC error, hardware suppresses the store and all subsequent stores then in the store buffer for that strand. The error is recorded in the DFESR by encoding sbdiou, and writing the store buffer index and privilege level to the DFESR. The privilege level recorded is the highest privilege level for any store in the store buffer at the time of the error.

Hardware presents a deferred *store_error* trap to the core. Software can decide what termination action is appropriate. Software at the trap handler should read the store buffer using store buffer diagnostic reads (see *Store Buffer — ASI_STB_ACCESS* on page 421) before issuing any stores which will overwrite the store buffer.

16.7.7.7 Address Bit Parity Error on a PCX read or Read for an ASI Ring Store (SBAPP)

If CERER.sbapp is set, on a PCX read or an ASI ring store read which exposes a parity error on the address bits, hardware suppresses the store, and logs the error in the DFESR by setting sbapp. Hardware suppresses other (younger, subsequent) stores in the store buffer. The highest privilege level for any store in the store buffer at the time of the error is also recorded.

Hardware presents a deferred *store_error* trap to the core.

If a user process was running, software may be able to avoid taking down the entire chip; similarly if an operating system was running, software may be able to kill only the partition the operating system was running in.

16.7.8 Scratchpad Array (SCAC and SCAU)

The Scratchpad array contains the scratchpad registers. It can be accessed only via normal ASI loads and stores or diagnostic ASI loads. The array is protected via SEC/DED ECC.

Hardware detects correctable Scratchpad errors only if CERER.scac is set, and uncorrectable errors only if CERER.scau is set.

ECC is not checked for diagnostic reads of the array, so a diagnostic read can not result in an error.

If a normal ASI read of the array results in a correctable ECC error, hardware corrects neither the returned data nor the error in the array.

If the SETER.pscce bit is set, hardware records the error in the D-SFSR by encoding scac, and records the array index with the error and syndrome in the D-SFAR. Hardware signals a precise *internal_processor_error* trap to the core. When software takes the trap, it can correct the data in the array. It decodes the syndrome, issues a diagnostic ASI read to read the data and ECC check bits, computes the correct data, and writes the corrected data back using a normal ASI store. Hardware generates the proper ECC before writing to the array.

If the SETER.pscce bit is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

If a normal ASI read of the array results in an uncorrectable ECC error, and SETER.pscce is set, hardware records the error in the D-SFSR by encoding SCAU. The array index with the error is stored in the D-SFAR. Hardware signals a precise *internal_processor_error* trap to the core.

If the SETER.pscce bit is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

16.7.9 Tick_compare (TCCP, TCUP, TCCD, TCUD)

The Tick_compare arrays are also protected via SEC/DED ECC. The Tick_compare array stores TICK_CMPR, STICK_CMPR, and HSTICK_CMPR. The arrays have three access means. The first is via ASR reads and writes. The second is via diagnostic ASI loads. The third, compare access, is implicit as hardware cycles through the entries to compare the TICK/STICK register with the TICK_CMPR registers.

Hardware detects correctable Tick_compare precise/disrupting errors only if CERER.tccp/tccd is set, and uncorrectable precise/disrupting errors only if CERER.tcup/tcud is set.

ECC is checked for a normal ASR read. If a correctable error occurs, hardware corrects neither the returned data nor the array location.

If SETER.pscce is set, hardware records the error in the D-SFSR by encoding tccp and the failing array index and syndrome is stored in the D-SFAR. Hardware generates a precise *internal_processor_error* trap to the core. For a correctable error, software at the trap handler can rewrite the array location and retry the failing instruction. It decodes the syndrome, issues a diagnostic ASI read to read the data and check bits, computes the correct data, and writes the corrected data back using a normal ASR write. Hardware generates the proper ECC before writing the data to the array.

If the SETER.pscce bit is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

If an uncorrectable error occurs on a normal ASR read, and SETER.pscce is set, hardware records the error in the D-SFSR by encoding tcup and writes the failing index and syndrome to the D-SFAR. Hardware takes a precise *internal_processor_error* trap. Software may be able to recover from this error by picking a reasonable value to load the TICK_CMPR register with, and retrying the ASR read.

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

ECC is not checked for a diagnostic ASI load, so no error is recorded and no trap can occur for this access type.

ECC is checked for a compare access. If a correctable or uncorrectable error occurs, hardware does not correct the data in the array, and suppresses any compare operation. Hardware records the error by encoding either tccd or tcud, and writing the failing array index in the DESR. The DESR is updated irrespective of the setting of SETER.de.

If SETER.de is set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware presents a disrupting *sw_recoverable_error* trap to the core.

Programming Note For a TCCD error, software can attempt recovery by using diagnostic array ASI accesses to correct the data as described for TCCP processing above. For a TCUD error, software may be able to recover from the error by setting the appropriate softened bit, and reloading the TICK_CMPR register after processing of the trap completes.

If SETER.de is not set, hardware continues executing without regard to the error. This can result in HW taking an *hstick_match* or *interrupt_level_n* trap based upon a faulty comparison. When software sets SETER.de (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware will present a disrupting *sw_recoverable_error* trap to the core.

16.7.10 Trap Stack Array (TSAC and TSAU)

The TSA array is protected via SEC/DED ECC. It contains the Trap Stack array and the mondo interrupt queue registers. It can be accessed via privileged register reads and writes or diagnostic ASI accesses. Privileged register writes require a read-modify-write operation, so privileged writes can generate an ECC error. The TSA is also accessed during Done and Retry instructions. Diagnostic accesses to the TSA do not generate ECC errors.

Hardware detects correctable TSA errors only if CERER.tsac is set. It detects uncorrectable TSA errors only if CERER.tsau is set.

If hardware detects a correctable error during an ASR read or write, or a DONE or RETRY instruction, hardware corrects neither the data returned by the read nor the array location. If the access was an ASI write, hardware suppresses the array write.

If SETER.pscce is set, hardware records the error in the D-SFSR by encoding tsac, and writes the failing TSA index and syndrome in the D-SFAR. Hardware presents the core with a precise *internal_processor_error* trap. Software can attempt recovery as follows. First note that each TSA array location contains several logical registers (see *Trap Stack Array (TSA)* on page 424) for details). Also, certain bits in each array location are unused (and assumed to be 0 during ECC generation). First, it turns off TSA error reporting by setting CERER.tsac to 0. The decoded syndrome from the D-SFAR indicates whether the error was in one of the unused bits in the array. If not, based upon the decoded syndrome, it reads the architected register which spans the failing bit using a RDPR/HPR instruction. After flipping the erroneous bit, it writes the new value back to the architected register using a WRPR/HPR instruction. If the decoded syndrome pointed to an unused bit, then software can read any register in the failing location using a RDPR/HPR instruction, and write the same value back using a WRPR/HPR instruction. Hardware will regenerate the correct ECC. Software then sets CERER.tsac to 1 to reenable TSAC error reporting.

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

If hardware detects an uncorrectable error during the read access for a privileged register read or write, or a DONE or RETRY instruction, and SETER.pscce is set, it records the uncorrectable error to the D-SFSR by encoding tsau, and writes the failing array index to the D-SFAR. It then presents the core with a precise *internal_processor_error* trap.

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

16.7.11 MMU Register Array (MRAU)

The MRA (MMU Register Array) contains various pointers and configuration registers used by hardware tablewalk and the MMU. Software can maintain a clean copy of the MRA contents; hardware does not update or store any MRA contents without software participation. (See *MMU Register Array (MRA)* on page 428 for details of the information stored in the MRA).

Each MRA location is protected by parity. The MRA is accessed by normal ASI reads and writes, diagnostic ASI reads and writes, and for hardware tablewalks. It is also read-modify-write for ASI writes. Diagnostic accesses to the MRA do not generate parity errors.

Hardware detects MRA parity errors only if CERER.mrau is set.

If hardware detects a parity error during a normal ASI access, hardware signals an MRAU error to the trap unit. If the ASI access was a normal ASI write, hardware suppresses the array update.

If SETER.pscce is set, hardware records the error in the D-SFSR by encoding mrau, and writes the failing array index to the D-SFAR. Hardware presents a precise *internal_processor_error* to the strand. Software can attempt recovery from the error by reloading the entry from its clean copy of the MRA contents and retrying the ASI access.

If SETER.pscce is not set, hardware continues executing using the uncorrected, and possibly erroneous, data. The error is not recorded.

Parity is not checked for diagnostic ASI reads and writes.

If an MRA location gets a parity error during a hardware tablewalk, the MRA error results in a precise *instruction_access_MMU_error* or *data_access_MMU_error* trap to the strand (see previous error handling sections). Software can attempt recovery from an error as above for an ASI access.

16.8 SPARC Error Registers

CERER (Core Error Recording Enable) register enables recording of an individual hardware error in either the I-SFSR, D-SFSR and D-SFAR, DESR, or DFESR. Bits in SETER (Core Error Trap Enable) register further control whether or not a trap occurs for a recorded error. A trap can never occur for an error that is not recorded; thus, a trap will be taken only if both CERER and associated SETER enable bits are set. Additionally, precise traps which are masked off (either by the appropriate CERER bit or SETER.pscce) will not update the I-SFSR or D-SFSR and D-SFAR. Hardware only updates the I-SFSR or D-SFSR and D-SFAR if a precise trap is taken.

16.8.1 ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER

The strands on a physical core share a hyperprivileged ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER (CERER) at ASI $4C_{16}$, VA{63:0} = 10_{16}. This register controls the reporting of errors to the virtual processor, and is intended only for use during debug. Software may also use this register to selectively disable error recording and trap generation in special circumstances. The format of the CERER register is shown in TABLE 16-4.

TABLE 16-4 CERER – ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER (ASI $4C_{16}$, VA 10_{16}) *(1 of 3)*

Bit	Field	Initial Value	R/W	Description
63	ittp	0	RW	If set to 0, mask ITTP errors. If set to 1, enable ITTP errors.
62	itdp	0	RW	If set to 0, mask ITDP errors. If set to 1, enable ITDP errors.
61	ittm	0	RW	If set to 0, mask ITTM errors. If set to 1, enable ITTM errors.
60	—	0	RO	*Reserved*
59	hwtwmu	0	RW	If set to 0, mask all uncorrectable MRA errors on hwtw. If set to 1, enable all uncorrectable MRA errors on hwtw.
58	hwtwl2	0	RW	If set to 0, mask all L2 errors on hwtw. If set to 1, enable all L2 errors on hwtw.
57	—	0	RO	*Reserved*
56	—	0	RO	*Reserved*
55	icl2c	0	RW	If set to 0, mask ICL2C errors. If set to 1, enable ICL2C errors.
54	icl2u	0	RW	If set to 0, mask ICL2U errors. If set to 1, enable ICL2U errors.
53	icl2nd	0	RW	If set to 0, mask ICL2ND errors. If set to 1, enable ICL2ND errors.
52	irf	0	RW	If set to 0, mask all IRF errors. If set to 1, enable all IRF errors.
51	—	0	RO	*Reserved*
50	frf	0	RW	If set to 0, mask all FRF errors. If set to 1, enable all FRF errors.
49	—	0	RO	*Reserved*
48	dttp	0	RW	If set to 0, mask DTTP errors. If set to 1, enable DTTP errors.
47	dttm	0	RW	If set to 0, mask DTTM errors. If set to 1, enable DTTM errors.
46	dtdp	0	RW	If set to 0, mask DTDP errors. If set to 1, enable DTDP errors.
45	—	0	RO	*Reserved*
44	—	0	RO	*Reserved*
43	—	0	RO	*Reserved*
42	—	0	RO	*Reserved*
41	—	0	RO	*Reserved*
40	dcl2c	0	RW	If set to 0, mask DCL2C errors. If set to 1, enable DCL2C errors.
39	dcl2u	0	RW	If set to 0, mask DCL2U errors. If set to 1, enable DCL2U errors.

TABLE 16-4 CERER – ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER (ASI $4C_{16}$, VA 10_{16}) *(2 of 3)*

Bit	Field	Initial Value	R/W	Description
38	dcl2nd	0	RW	If set to 0, mask DCL2ND errors. If set to 1, enable DCL2ND errors.
37	sbdlc	0	RW	If set to 0, mask SBDLC errors. If set to 1, enable SBDLC errors.
36	sbdlu	0	RW	If set to 0, mask SBDLU errors. If set to 1, enable SBDLU errors.
35	—	0	RO	*Reserved*
34	—	0	RO	*Reserved*
33	mrau	0	RW	If set to 0, mask MRAU errors. If set to 1, enable MRAU errors.
32	tsac	0	RW	If set to 0, mask TSAC errors. If set to 1, enable TSAC errors.
31	tsau	0	RW	If set to 0, mask TSAU errors. If set to 1, enable TSAU errors.
30	scac	0	RW	If set to 0, mask SCAC errors. If set to 1, enable SCAC errors.
29	scau	0	RW	If set to 0, mask SCAU errors. If set to 1, enable SCAU errors.
28	tccp	0	RW	If set to 0, mask TCCP errors. If set to 1, enable TCCP errors.
27	tcup	0	RW	If set to 0, mask TCUP errors. If set to 1, enable TCUP errors.
26:24	—0	0	RO	*Reserved*
23	sbapp	0	RW	If set to 0, mask SBAPP errors. If set to 1, enable SBAPP errors.
22	—	0	RO	*Reserved*
21	l2c_socc	0	RW	If set to 0, mask L2C and SOCC errors. If set to 1, enable L2C and SOCC errors.
20	l2u_socu	0	RW	If set to 0, mask L2U and SOCU errors. If set to 1, enable L2U and SOCU errors.
19	l2nd	0	RW	If set to 0, mask L2ND errors. If set to 1, enable L2ND errors.
18	icvp	0	RW	If set to 0, mask ICVP errors. If set to 1, enable ICVP errors.
17	ictp	0	RW	If set to 0, mask ICTP errors. If set to 1, enable ICTP errors.
16	ictm	0	RW	If set to 0, mask ICTM errors. If set to 1, enable ICTM errors.
15	icdp	0	RW	If set to 0, mask ICDP errors. If set to 1, enable ICDP errors.
14	dcvp	0	RW	If set to 0, mask DCVP errors. If set to 1, enable DCVP errors.
13	dctp	0	RW	If set to 0, mask DCTP errors. If set to 1, enable DCTP errors.
12	dctm	0	RW	If set to 0, mask DCMH errors. If set to 1, enable DCMH errors.
11	dcdp	0	RW	If set to 0, mask DCDP errors. If set to 1, enable dcDP errors.
10	sbdpc	0	RW	If set to 0, mask SBDPC errors. If set to 1, enable SBDPC errors.
9	sbdpu_ sbdiou	0	RW	If set to 0, mask SBDPU and SBDIOU errors. If set to 1, enable SBDPU and SBDIOU errors.
8	mamu	0	RW	If set to 0, mask MAMU errors. If set to 1, enable MAMU errors.
7	tccd	0	RW	If set to 0, mask TCCD errors. If set to 1, enable TCCD errors.
6	tcud	0	RW	If set to 0, mask tcud errors. If set to 1, enable tcud errors.
5	mal2c	0	RW	If set to 0, mask MAL2C errors. If set to 1, enable MAL2C errors.
4	mal2u	0	RW	If set to 0, mask MAL2U errors. If set to 1, enable MAL2U errors.
3	mal2nd	0	RW	If set to 0, mask MAL2ND errors. If set to 1, enable MAL2ND errors.

TABLE 16-4 CERER – ASI_CORE_ERROR_RECORDING_ENABLE_REGISTER (ASI $4C_{16}$, VA 10_{16}) *(3 of 3)*

Bit	Field	Initial Value	R/W	Description
2	—	0	RW	—
1	—	0	RW	—
0	—	0	RW	—

Each bit of the enable fields for precise errors (CERER bits 63:61, 59:58, 54:53, 52, 50, 48:46, 39:36, 33:27) controls whether a corresponding precise error is recorded in the I-SFSR or D-SFSR. A 1 in the CERER bit position enables the corresponding error to be recorded if that error occurs. Otherwise, that error is not recorded. Similarly, each bit of the DESR enable field For disrupting errors (CERER bits 55, 40, 21:10, 8:0) control whether the corresponding disrupting error is recorded in the DESR. Bits 23 and 9 oF the CERER control whether the associated error is logged in the DFESR.

Logically the CERER bit is **and**ed with error sources before an error is encoded and recorded in the I-SFSR, D-SFSR, DESR, or DFESR. A masked enable in the CERER Precise Error enable field cannot cause the I-SFSR or D-SFSR to be updated if the associated error occurs. Similarly a masked enable in the CERER DESR enable field cannot cause the DESR.f or the DESR.me bits to be set, nor can a masked enable in the DFESR enable field cause a DFESR bit to be set.

16.8.2 ASI_STRAND_ERROR_TRAP_ENABLE_REGISTER

Each virtual processor has a hyperprivileged ASI_STRAND_ERROR_TRAP_ENABLE_REGISTER (SETER) at ASI $4C_{16}$, VA{63:0} = 18_{16}. This register controls the generation of traps for errors that are reported to the virtual processor.

Each of the SETER bits apply to a particular class of maskable errors. Nonmaskable errors always result in traps regardless of the setting of the SETER bits (assuming that the CERER bit for the error is set).

Bit 62, pscce, controls whether a trap will be taken if a maskable, precise software correctable error is detected during the execution of an instruction. Since hardware performs neither correction nor clearing of the error, software must take action to avoid unpredictable execution and possible data loss. This bit should always be set for normal operation. An example of this type of error is a correctable IRF ECC error. Furthermore, if pscce is 0, no maskable, precise software correctable errors will be recorded in the I-SFSR or the D-SFSR and D-SFAR.

Bit 61, de, controls whether a *sw_recoverable_error* disrupting trap will be taken if the DESR.f is set for a maskable disrupting error which is not hardware corrected and cleared. (Note that a disrupting trap is also conditioned by the setting of the PSTATE.ie bit when HPSTATE.hpriv is set; PSTATE.ie must be set in this case to take a disrupting trap).

Bit 60, dhcce, controls whether a *hw_corrected_error* disrupting trap will be taken for a maskable disrupting hardware error that was corrected and cleared. In this case, software should log the error. (Note that a disrupting trap is also conditioned by the setting of the PSTATE.ie bit when HPSTATE.hpriv is set; PSTATE.ie must be set in this case to take a disrupting trap).

Note | In normal operation, pscce and de should *always* be set. Otherwise, the UltraSPARC T2 core will continue to execute in the face of hardware errors, leading to unpredictable behavior.

The format of the SETER register is shown in TABLE 16-5.

TABLE 16-5 SETER – ASI_STRAND_ERROR_TRAP_ENABLE_REGISTER (ASI $4C_{16}$, VA 18_{16})

Bit	Field	Initial Value	R/W	Description
63	—	0	RO	*Reserved*
62	pscce	0	RW	If set to 1, trap on maskable, precise software corrected and cleared errors
61	de	0	RW	If set to 1, trap on maskable, disrupting errors which are not hardware corrected and cleared
60	dhcce	0	RW	If set to 1, trap on maskable, disrupting hardware corrected and cleared errors.
59:0	—1	0	RO	*Reserved*

16.8.3 IMMU Synchronous Fault Status Register

UltraSPARC T2 uses the I-SFSR to record precise hardware errors encountered during the instruction fetch process. For most instruction fetch errors, the TPC[TL] records the relevant VA. For an IT MRA[1]correctable or uncorrectable error, the D-SFAR records the failing MRA index. Using the D-SFAR obviates the need for an ISFAR. If the error occurred for a fetch to L2 (either during an hwtw or instruction fetch), an L2 ESR contains more information about the error.

1. MRA stands for the MMU Register Array; it contains various configuration registers required for the MMU and hardware tablewalk.

Each virtual processor has one hyperprivileged, read/write, ASI_IMMU_SFSR register located at ASI 50_{16}, VA 18_{16}. The format of the ISFSR register is shown in TABLE 16-6.

TABLE 16-6 ISFSR – ASI_IMMU_SYNCHRONOUS_FAULT_STATUS_REGISTER (ASI 50_{16}, VA 18_{16})

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2:0	errtype	0	RW	Error type information, format defined in TABLE 16-7.

Note | ISFSR is preserved across warm resets to allow software to determine the cause of an error that required a warm reset.

TABLE 16-7 describes the hardware errors that are recorded in the ISFSR register. If multiple error conditions occur for the same instruction, only the highest priority is logged. If multiple exceptions occur for the same instruction, the ISFSR will only be updated if the error is the highest-priority exception for the instruction. For example, if an access to a privileged page in user mode occurs in conjunction with an Icache tag parity error, the I-SFSR will not be updated. Since ISFSR only logs precise errors, a write clearing this register cannot be simultaneous to an error that would set the I-SFSR.

Note | There is no “latching” of the first error. If another error occurs before the I-SFSR has been examined, the new error information is captured in the I-SFSR, overwriting the old information.

TABLE 16-7 Hardware Errors Recorded in the I-SFSR Register

Trap Type	Error Name	Error Description	Relative Priority	Contents of I-SFSR ErrorType Field	Contents of D-SFAR (see TABLE 16-11)
instruction_access_MMU_error	ITTM	IT Tag Multiple hit	1	1	TPC[TL] contains VA
	ITTP	IT Tag Parity	2	2	TPC[TL] contains VA
	ITDP	IT Data Parity	3	3	TPC[TL] contains VA
	ITMU	IT MRA Uncorrectable	4	4	2
	ITL2U	IT L2 Uncorrectable	5	5	TPC[TL] contains VA
	ITL2ND	IT L2 NotData	5	6	TPC[TL] contains VA
instruction_access_error	ICL2U	IC L2 Uncorrectable	6	1	TPC[TL] contains VA
	ICL2ND	IC L2 NotData	6	2	TPC[TL] contains VA

The Trap Type column describes the trap type which will result if the error occurs and the trap is enabled (see SETER in *ASI_STRAND_ERROR_TRAP_ENABLE_REGISTER* on page 242). The Error Name column names the hardware error, and the Error Description column provides a brief description of the error. If multiple errors occur for the same instruction, the error that is recorded in the I-SFSR is determined by the Relative Priority column, with 1 being the highest priority. The error information recorded in the I-SFSR is indicated by the encoding specified in the errortype field while the format of the address information recorded in the D-SFAR is described by the entry in the last column for ITMU errors, with reference to TABLE 16-11 on page 248.

16.8.4 DMMU Synchronous Fault Status and Address Registers

UltraSPARC T2 records data-related precise hardware errors in the D-SFSR and D-SFAR. Since a hardware error has higher priority than a normal program error, UltraSPARC T2 saves hardware by using the D-SFSR and D-SFAR for precise data hardware errors.

The error types recorded in the D-SFSR and D-SFAR are listed in TABLE 16-8. TSA stands for Trap Stack Array; the TSA contains the V9 trap stack and mondo queue interrupt registers. SCA stands for Scratchpad Array; the SCA contains the

scratchpad registers. If multiple error conditions occur for the same access, only the highest priority is logged. If multiple exceptions occur for the same instruction, the D-SFSR will only be updated if the error is the highest priority exception for the instruction. For example, if an FRF ECC error occurs in conjunction with an privileged opcode exception, the D-SFSR will not be updated. Since the D-SFSR and D-SFAR log precise errors, a write clearing the D-SFSR cannot be simultaneous to an error that would set the D-SFSR and D-SFAR.

Notes

On a block store instruction which gets multiple FRF ECC errors, only the first error is logged in the D-SFSR and D-SFAR.

There is no "latching" of the first error in the D-SFSR and D-SFAR. If another error occurs before the D-SFSR and D-SFAR have been examined, the new error information is captured in the D-SFSR and D-SFAR, overwriting the old information.

TABLE 16-8 Hardware Errors Recorded in the D-SFSR and D-SFAR Registers

Trap Type	Error Name	Error Description	Relative Priority	Contents of D-SFSR ErrorType Field	Contents of D-SFAR (see TABLE 16-11
internal_processor_error	IRFU	Integer register file uncorrectable	1	1	8
	IRFC	Integer register file correctable	2	2	8
	FRFU	Floating-point register file uncorrectable	3	3	9
	FRFC	Floating-point register file correctable	4	4	9
	SBDLC	Store buffer data load hit correctable	10	5	6
	SBDLU	Store buffer data load hit uncorrectable	10	6	6
	MRAU	MRA uncorrectable	11	7	10
	TSAC	TSA correctable	11	8	11
	TSAU	TSA uncorrectable	11	9	11
	SCAC	SCA correctable	11	10	12
	SCAU	SCA uncorrectable	11	11	12
	TCCP	Tick compare correctable precise	11	12	13
	tcup	Tick compare uncorrectable precise	11	13	13

TABLE 16-8 Hardware Errors Recorded in the D-SFSR and D-SFAR Registers *(Continued)*

Trap Type	Error Name	Error Description	Relative Priority	Contents of D-SFSR ErrorType Field	Contents of D-SFAR (see TABLE 16-11
data_access_MMU_error	DTTM	DT tag multiple hit	5	1	1
	DTTP	DT tag parity	6	2	1
	DTDP	DT data parity	7	3	1
	DTMU	DT MRA uncorrectable	8	4	2
	DTL2U	DT L2 uncorrectable	9	5	3
	DTL2ND	DT L2 NotData	9	6	3
data_access_error	DCL2U	dc L2 uncorrectable	11	1	3
	DCL2ND	dc L2 NotData	11	2	3
	SOCU	SOC uncorrectable	12	4	See page 317

16.8.4.1 DMMU Synchronous Fault Status Register

Each virtual processor has one hyperprivileged, read/write, ASI_DMMU_SFSR register located at ASI 58_{16}, VA 18_{16}. The format of the DSFSR register is shown in TABLE 16-9.

TABLE 16-9 DSFSR – ASI_DMMU_SYNCHRONOUS_FAULT_STATUS_REGISTER (ASI 58_{16}, VA 18_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	errtype	0	RW	Error type information, format defined in TABLE 16-8.

Note | DSFSR is preserved across warm resets to allow software to determine the cause of an error that required a warm reset.

16.8.4.2 DMMU Synchronous Fault Address Register

Each virtual processor has one hyperprivileged, read-only, ASI_DMMU_SFAR register located at ASI 58_{16}, VA 20_{16}. The format of the DSFAR register is shown in TABLE 16-10.

TABLE 16-10 DSFAR – ASI_DMMU_SYNCHRONOUS_FAULT_ADDRESS_REGISTER (ASI 58_{16}, VA 20_{16})

Bit	Field	Initial Value	R/W	Description
63:48	—	0	RO	*Reserved*
47:0	erraddr	0	RO	Error address information, format defined in TABLE 16-11.

Note | DSFAR is preserved across warm resets to allow software to determine the cause of an error that required a warm reset.

TABLE 16-11 further describes the contents of the DSFAR or DESR registers, which contain information about either the address or syndrome of the error, for each type of error which can occur.

TABLE 16-11 Contents of D-SFAR or DESR Error Address Field for Various Error Types

Error Class in I-SFSR, D-SFSR, or DESR	Bit Index in D-SFAR or DESR	ErrorAddr Information	Remarks
1 (ITLB/DTLB)	47:0	VA{47:0} for D-SFAR	Undefined for ITLBE: VA of ITLB error is recorded in TPC[TL]; VA of DTLB error is recorded in D-SFAR.
2 (MRA access for HWTW)	47:3	—	Undefined.
	2:0	MRA index{2:0}	
3 (Memory access)	47:0	—	Undefined; L2 ESR contains more information about the error.
4 (IC access)	47:9	—	Undefined.
	8:6	Way{2:0}	One of the ways if a multiple-way hit occurred; not recorded for Icache data parity errors.
	5:0	IC index{5:0}	
5 (DC access)	47:9	—	Undefined.
	8:7	Way{1:0}	One of the ways if a multiple-way hit occurred.
	6:0	DC Index{6:0}	
6 (Store Buffer Data ECC)	47:3	—	Undefined.
	2:0	STB Index{2:0}	

TABLE 16-11 Contents of D-SFAR or DESR Error Address Field for Various Error Types *(Continued)*

Error Class in I-SFSR, D-SFSR, or DESR	Bit Index in D-SFAR or DESR	ErrorAddr Information	Remarks
8 (IRF ECC)	47:15	—	Undefined.
	14:7	Syndrome{7:0}	ECC Syndrome; see Appendix G, *ECC Codes* for ECC syndrome decode.
	6:5	GL{1:0}	GL field when error occurred.
	4:0	IRF Index{4:0}	Physical array index where the error occurred. For indices which point to %in or %out registers, the interpretation depends upon whether %cwp is pointing to an even or an odd window, as follows: For an even window: index 0 --> %g0 index 8 --> %o0 index 16 --> %l0 index 24 --> %i0 For an odd window, the indices which map to %in and %out registers swap: index 0 --> %g0 index 24 --> %o0 index 16 --> %l0 index 8 --> %i0
9 (FRF ECC)	47:20	—	Undefined.
	19:13	Even Syndrome{6:0}	Syndrome of lower 32 bits (bits 31:0); see Appendix G, *ECC Codes* for ECC syndrome decode.
	12:6	Odd Syndrome{6:0}	Syndrome of upper 32 bits (bits 63:32); see Appendix G, *ECC Codes* for ECC syndrome decode.
	5:0	FRF Index{5:0}	
10 (MRA)	47:3	—	Undefined. (No syndrome since parity is stored in MRA.)
	2:0	MRA Index{2:0}	
11 (TSA)	47:19	—	Undefined.
	18:11	Odd Syndrome{7:0}	
	10:3	Even Syndrome{7:0}	
	2:0	TSA Index{2:0}	
12 (Scratchpad)	47:3	—	Undefined.
	10:3	Syndrome{7:0}	
	2:0	SCPD Index{2:0}	

TABLE 16-11 Contents of D-SFAR or DESR Error Address Field for Various Error Types *(Continued)*

Error Class in I-SFSR, D-SFSR, or DESR	Bit Index in D-SFAR or DESR	ErrorAddr Information	Remarks
13 (Tick_compare)	47:2	—	Undefined.
	9:2	Syndrome{7:0}	
	1:0	Tick_compare Index{1:0}	

16.8.5 Disrupting Error Status Register (DESR)

Each virtual processor has a hyperprivileged, read-only DESR located at ASI $4C_{16}$, VA 0_{16}. The DESR records all disrupting errors. A read of DESR clears all fields. The format of the DESR register is shown in TABLE 16-12.

TABLE 16-12 DESR – ASI_DISRUPTING_ERROR_STATUS_REGISTER (ASI $4C_{16}$, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63	f	0	RC	Full, register contains valid data.
62	me	0	RC	Multiple errors detected.
61	s	0	RC	Error trap type. If 1, a *sw_recoverable_error* was logged. If 0, a *hw_corrected_error* was logged.
60:56	errtype	0	RC	Error type, format defined in TABLE 16-13.
55:11	—	0	RC	*Reserved*
10:0	erraddr	0	RC	Error address information, format defined in TABLE 16-11.

Note | DESR is preserved across warm resets to allow software to determine the cause of an error that required a warm reset.

DESR captures information on both *sw_recoverable_error* and *hw_corrected_error* errors. The s bit denotes the type of error whose information is logged in the errortype and erroraddr fields. If set to 1, the errortype and erroraddr fields contain information regarding a *sw_recoverable_error*; else, the errortype and erroraddr fields contain information regarding a *hw_corrected_error.* The s bit is required since UltraSPARC T2 can detect both *hw_corrected_error* and *sw_recoverable_error* simultaneously. Furthermore, in the event DESR contains error information regarding a *hw_corrected_error* and a *sw_recoverable_error* occurs, the *sw_recoverable_error* must take precedence, to allow software the opportunity to recover from the error.

Hardware sets the f bit when a disrupting error occurs and the corresponding CERER bit is set; the f bit signifies that the register contains valid data (for example, is full). The s bit is set when a *sw_recoverable_error* occurs and the s bit is not

already set. The me bit is set when an error occurs and the f bit is already set, since disrupting errors can occur while the f bit is set from a previous error. If the f bit was set from a previous *hw_corrected_error* and the current error is a *sw_recoverable_error*, the error information is captured for the *sw_recoverable_error*. Otherwise, no additional information is captured about an error that sets the me bit.

If the f bit is set and the s bit is cleared (which implies only hardware corrected and cleared errors have occurred), a *hw_corrected_error* trap will be generated if SETER.dhcce is set and either PSTATE.ie is set or HPSTATE.hpriv is clear. If not enabled, the trap remains pending while the f bit is set and the s bit is cleared until SETER.dhcce is set and either PSTATE.ie is set or HPSTATE.hpriv is clear.

If the f bit is set and the s bit is set (which implies an error that was not hardware corrected and cleared has occurred), a *sw_recoverable_error* trap will be generated if SETER.de is set and either PSTATE.ie is set or HPSTATE.hpriv is clear. If not enabled, the trap remains pending while the f bit is set and the s bit is set until SETER.de is set and either PSTATE.ie is set or HPSTATE.hpriv is clear.

A more detailed description of the DESR semantics is as follows:

1. If the f bit is clear and a *sw_recoverable_error* occurs, hardware sets the s and f bits to 1 and captures the *sw_recoverable_error* information in the errortype and erroraddr fields. If another *hw_corrected_error* or *sw_recoverable_error* occurs before software reads the DESR and clears the s and f bits, hardware sets the me bit and leaves the errortype and erroraddr fields unchanged.

2. If the f bit is clear and a *hw_corrected_error* occurs and there is no simultaneous *sw_recoverable_error*, hardware sets DESR.s to 0, DESR.f to 1, and captures the error information about the *hw_corrected_error*.

3. If the f bit is set and a *sw_recoverable_error* occurs:

   a. If the s bit is set, hardware sets the me bit, and leaves all other DESR fields unchanged.

   b. If the s bit is clear, hardware overwrites the DESR as follows. It sets the s, f, and me bits to 1 and overwrites the contents of the errortype and erroraddr fields with the information about the *sw_recoverable_error*. In this case, software could find that hardware took a *hw_corrected_error* trap and upon reading the DESR, realize that a *sw_recoverable_error* occurred which overwrote the *hw_corrected_error*. All detailed information about the *hw_corrected_error* is lost.

4. If the f bit is set, and a *hw_corrected_error* occurs, hardware sets the me bit and leaves the other DESR fields unchanged.

5. If a *hw_corrected_error* and a *sw_recoverable_error* occur simultaneously, the *sw_recoverable_error* takes precedence. Hardware sets the f, s, and me bits to 1 and records the *sw_recoverable_error* in the DESR errortype and erroraddr fields. The *sw_recoverable_error* exception becomes pending. The pending *hw_corrected_error* exception is cleared.

6. If software performs an ASI read of the DESR simultaneously with hardware reporting a new disrupting exception:

   a. Hardware returns the value previously stored in the DESR to the ASI read.

   b. Hardware stores the new disrupting exception in the DESR as if the f bit were clear.

Programming Notes

ASI reads of the DESR clear the DESR. Consequently, hardware clears the corresponding f bit in the CLESR as well.

The DESR is read-only. To test error recovery code, software must inject an error (using the `ASI_CORE_ERROR_INJECT` register), then cause the erroneous data to be used, thereby taking a *hw_corrected_error* or *sw_recoverable_error* trap.

The DESR errortype field contains information that describes the error and is detailed in TABLE 16-13 below. If multiple disrupting errors occur at the same time, hardware prioritizes the errors and stores the syndrome for the highest-priority error in the errortype field, according to the column labeled Relative Priority, with 1 being the highest priority.

TABLE 16-13 DESR Error Type Bits

Trap Type	Error Name	Description	Relative Priority	Contents of DESR errortype Field (bits 60:56)	Contents of DESR erroraddr Field (see TABLE 16-11)
hw_corrected_error	ICVP	Instruction cache valid bit parity.	11.1	1	4
	ICTP	Instruction cache tag parity.	11.2	2	4
	ICTM	Instruction cache tag multiple.	11.3	3	4
	ICDP	Instruction cache data parity.	11.4	4	4
	DCVP	Data cache valid bit parity.	12.1	5	5
	DCTP	Data cache tag parity.	12.2	6	5
	DCTM	Data cache tag multiple.	12.3	7	5
	DCDP	Data cache data parity.	12.4	8	5
	L2C	L2 cache correctable.	13	9	3
	SBDPC	Store buffer data PCX read correctable ECC.	14	10	6
	SOCC	SOC correctable.	15	11	(see SOC)

TABLE 16-13 DESR Error Type Bits *(Continued)*

Trap Type	Error Name	Description	Relative Priority	Contents of DESR errortype Field (bits 60:56)	Contents of DESR erroraddr Field (see TABLE 16-11)
sw_recoverable_error	SBDPU	Store buffer data PCX read uncorrectable ECC.	1	6	6
	TCCD	Tick_compare correctable disrupting.	2	14	13
	TCUD	Tick_compare uncorrectable disrupting.	2	15	13
	L2C	L2 correctable ECC error.	5	20	3
	L2U	L2 uncorrectable ECC error.	5	16	3
	L2ND	L2 NotData error.	5	17	3
	ITL2C	IT L2 correctable.	6	1	3
	ICL2C	ic L2 correctable.	6	2	3
	DTL2C	DT L2 correctable.	6	3	3
	DCL2C	dc L2 correctable.	6	4	3
	SOCU	SOC uncorrectable.	7	19	See Section 16.17

16.8.6 Deferred Error Status Register (DFESR)

Each virtual processor has a hyperprivileged, read-only-and-clear DFESR at ASI $4C_{16}$, VA 8_{16}. A read of DFESR clears all fields to 0. DFESR records deferred errors detected by the UltraSPARC T2 core. The format of the Deferred Error Status register is shown in TABLE 16-14.

TABLE 16-14 DFESR – ASI_DEFERRED_ERROR_STATUS_REGISTER (ASI $4C_{16}$, VA 8_{16})

Bit	Field	Initial Value	R/W	Description
63:62	—	0	RO	*Reserved*
61:60	type	0	RC	Error type, format defined in TABLE 16-15.
59:58	priv	0	RC	Privilege level, format defined in TABLE 16-16.
57:55	stbindex	0	RC	Store buffer index.
54:0	—	0	RO	*Reserved*

Note | DFESR is preserved across warm resets to allow software to determine the cause of an error that required a warm reset.

The content of the error type field is as follows:

TABLE 16-15 DFESR ErrorType

Bit Index in FESR	Field Name	Error	Remarks
61	sba	SBAPP	Store buffer address parity error.
60	sbdio	SBDIOU	Store buffer data I/O uncorrectable error.

The content of the privilege-level field is as follows:

TABLE 16-16 Privilege Level Field Contents

Bit Index in FESR	Field Name	Interpretation	Remarks
59:58	priv	Highest privilege level of any store in the store buffer	00 - User (hpriv = priv = 0). 01 - Supervisor (hpriv = 0, priv = 1). 10 - Hpriv (hpriv = 1). 11 - Error in privilege level field.

Using the privilege level field, software can determine an appropriate response to a fatal thread error. For example, if a user-level store received the error, software may kill only the user process, instead of taking down the entire chip or partition. An encoding of 11 in the privilege level field means that there was an error in this field in the store buffer contents, so the privilege level of the store cannot be determined.

Bits 57:55 of the DFESR contain the index in the store buffer which had the error.

Programming Note Hardware cannot support simultaneous reads of the DFESR and updates of the DFESR contents (which can occur due to a previously issued store to any address). Therefore, software must not attempt a store in the store error trap handler until it has read the DFESR (and implicitly cleared the DFESR).

16.8.7 ASI_CLESR

Each physical core has a shared, read-only, hyperprivileged Core Local Error Status register located at ASI $4C_{16}$, VA 20_{16}.

TABLE 16-17 CLESR Contents

T7	T6	T5	T4	T3	T2	T1	T0	*Reserved*
63:62	61:60	59:58	57:56	55:54	53:52	51:50	49:48	47:0

Each of the eight fields T7–T0 contains 2 bits. The high-order bit is a copy of the DESR.f bit for that strand; the low-order bit is the **or** of the two DFESR.errortype bits for that strand. For example, CLESR bit 54 is the **or** of the DFESR.errortype bits for strand 3.

16.8.8 ASI_CLFESR

Each physical core has a shared, read-only, hyperprivileged Core Local First Error Status register located at ASI $4C_{16}$, VA 28_{16}. CLFESR logs the first error recorded by a physical core's strands DESRs and DFESRs in CLESR. Thus, there is a one-to-one correspondence between the bits in the CLESR and CLFESR. When CLESR is clear, any error that is recorded in a strand's DESR or DFESR causes the error to be logged in CLESR and CLFESR. If multiple strands log an error in their respective DESR or DFESR simultaneously, then multiple bits can be set in the CLFESR.

Once the CLFESR has recorded an error, it will not log any subsequent error until software clears all f bits in all the physical core's strands' DESRs and errortype bits in the DFESRs. At this point the CLESR will be all zeroes, and hardware clears the CLFESR.

TABLE 16-18 CLFESR contents

T7	T6	T5	T4	T3	T2	T1	T0	*Reserved*
63:62	61:60	59:58	57:56	55:54	53:52	51:50	49:48	47:0

16.8.9 ASI_ERROR_INJECT_REG

Each physical core has a hyperprivileged ASI_ERROR_INJECT_REG register at ASI: 43_{16}, VA{63:0} = 0. The ASI_ERROR_INJECT register enables software to inject errors into a subset of the UltraSPARC T2 core's error protected arrays. The subset includes all error-protected arrays except for the Icache and Dcache valid bit, tag, and data arrays. The format of the Error Injection register is shown in TABLE 16-19.

TABLE 16-19 Error Injection Register – ASI_ERROR_INJECT_REGISTER (ASI 43_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	enb_hp	0	RW	If 1, enable error injection; else disable error injection.
30	—	0	RO	*Reserved*
29	imdu	0	RW	Inject ITLB data array parity error on an ITLB update.
28	imtu	0	RW	Inject ITLB CAM parity error on an ITLB update.
27	dmdu	0	RW	Inject DTLB data array parity error on a DTLB update.

TABLE 16-19 Error Injection Register – ASI_ERROR_INJECT_REGISTER (ASI 43_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
26	dmtu	0	RW	Inject DTLB CAM parity error on a DTLB update.
25	ircu	0	RW	Inject an IRF ECC error on any IRF write.
24	frcu	0	RW	Inject an FRF ECC error on any FRF write.
23	scau	0	RW	Inject an ECC error into the Scratchpad array.
22	tcup	0	RW	Inject an ECC error into the TICK_CMPR array.
21	tsau	0	RW	Inject an ECC error into the TSA array.
20	mrau	0	RW	Inject a parity error into the MRA array.
19	stau	0	RW	Inject a parity error into the store buffer CAM.
18	—	0	RO	*Reserved*
17	stdu	0	RW	Inject an ECC error into the Store Buffer Data array.
16:8	—	0	RO	*Reserved*
7:0	eccmask	0	RW	Mask of bits to be **xor**ed with ECC bits for the IRF, FRF, SCA, TCA, TSA, or STB data arrays.

Setting more than one error source bit active at a time produces undefined results. In order to inject an error, both the enb_hp bit and one of the error source bits (bits 29:17) must be set.

Only one strand should be active at a time when injecting errors, since hardware may schedule a strand at any time and that strand may inject errors or receive the effect of errors generated by another strand, complicating diagnosis.

Writes to ASI_ERROR_INJECT are actually post-synchronizing, so no MEMBAR #Sync is required after the write. However, there are cases where writes to ASI_ERROR_INJECT are not presynchronizing (that is, an instruction *before* the write could see the effect). To guard against this, software needs to put a MEMBAR #Sync *before* a write to ASI_ERROR_INJECT.

16.9 L2 Cache Error Descriptions

The L2 cache protects its tag with SEC ECC. Each 32-bit data subline is protected with SECDED ECC. In the following L2 cache behavior descriptions, a partial store refers to a store that updates less than the full 32 bits of any 32-bit data subline.

Errors detected in the L2 cache and also the DRAMs are reported back to the requesting cores using two packet types. L2_Load_Return packets are sent by the L2 cache to the requesting core for synchronous requests such as loads and prefetches, and Error_Indication_(L2) packets for asynchronous errors such as store, writeback, and scrub errors.

The flow of errors detected in the L2 cache or detected in the DRAM and passed to the L2 cache is shown in FIGURE 16-1 for synchronous (load) requests to the L2 cache from the requesting core. Load request can be made from the SPARC core for load, instruction fetches and TTE requests. The errors are presented in the 2-bit err field of the L2_Load_Return packet, indicating correctable (01), uncorrectable (10) and NotData (11).

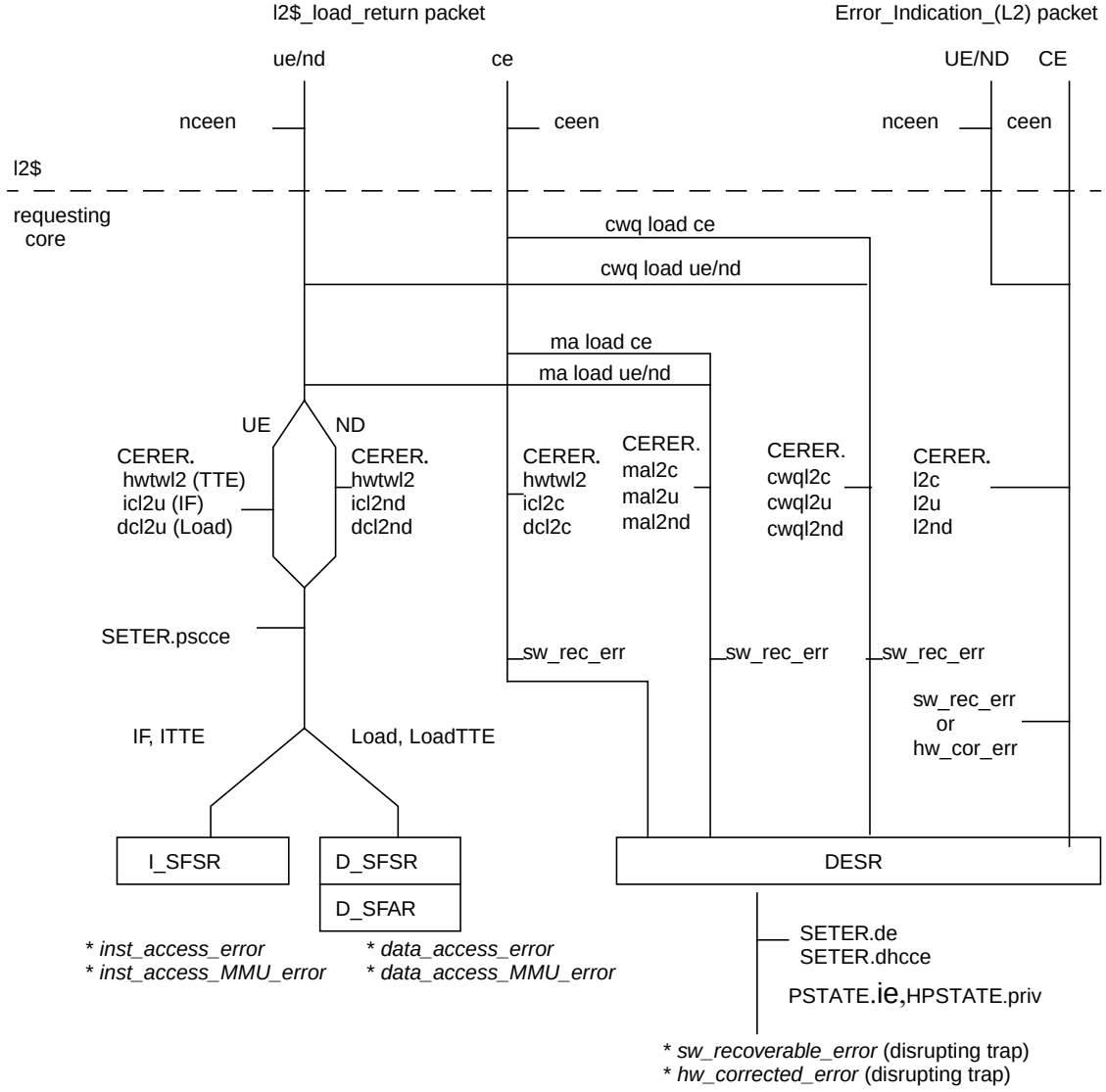


FIGURE 16-1 L2$ Error Flow for Errors Detected by L2$/DRAM

The flow of errors detected in the L2 cache or detected in the DRAM and passed to the L2 cache is shown in FIGURE 16-1 for asynchronous errors from the L2 cache to the requesting core or the L2_CSR_REG.errorsteer core. These errors include store writeback, scrub, DMA access, and prefetch errors. The errors are presented in the 2-bit err field of the Error_Indication_(L2) packet, indicating correctable (01), uncorrectable (10) and NotData (11).

Uncorrectable and NotData errors presented in the L2_Load_Return packet for loads requested by the core result in precise traps. Correctable errors on core loads result in *sw_recoverable_error* disrupting traps.

Uncorrectable errors and some correctable errors (DMA Read and Prefetch) presented in the Error_Indication_(L2) packet result in *sw_recoverable_error* disrupting traps. The other correctable errors presented in the Error_Indication_(L2) packet result in *hw_corrected_error* disrupting traps.

This section details the errors that occur from L2 hits. DRAM accesses are dealt with in Section 16.11 on page 287.

The L2 cache error detection and reporting mechanism is functional whether the L2 cache is enabled or disabled. The only difference is that L2 cache data and tag array errors are not detected since the arrays are disabled.

16.9.1 L2 Cache Data Correctable ECC Error for Access (LDAC)

Correctable data errors are detected by the L2 cache, captured in the L2 Cache Error Status register and the L2 Cache Error Address register, and then reported to the requesting virtual processor. The following control bits are used:

L2 Cache Error Enable ceen bit — When set, causes the error to be sent to the requesting virtual processor; when reset, error is ignored but still captured in the L2 Cache Error Status register and the L2 Cache Error Address register.

Associated CERER bit — When set causes L2$ error (assuming ceen = 1) to be captured in DESR register, and conditions reporting of disrupting trap condition; when disabled, L2$ error is ignored, and desr is not changed.

SETER.de bit — When set (along with ceen and associated CERER bit), causes disrupting *sw_recoverable_error* trap condition to be presented to the virtual processor for load requests that use the L2_Load_Return packet. The disrupting trap is further conditioned by HPSTATE.hpriv and PSTATE.ie. When SETER.de/dhcce is reset, the disrupting trap condition remains pending. When SETER.de/dhcce is later set, the disrupting trap is taken, conditioned by HPSTATE.hpriv and PSTATE.ie.

SETER.dhcce bit — When set (along with ceen and associated CERER bit), causes disrupting *hw_corrected_error* trap condition to be presented to the virtual processor for store requests that use the Error Indication L2) packet. The disrupting trap is further conditioned by HPSTATE.hpriv and PSTATE.ie. When SETER.de/dhcce is reset, the disrupting trap condition remains pending. When SETER.de/dhcce is later set, the disrupting trap is taken, conditioned by HPSTATE.hpriv and PSTATE.ie.

The error information stored in the L2 Cache Error Status register is shown in TABLE 16-21 on page 277, and the physical address captured in the L2 Cache Error Address register is shown in TABLE 16-25 on page 284. They are recorded as LDAC errors.

16.9.1.1 TTE Request for ITLB (ITL2C)

When a correctable ECC error is detected, the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen, and SPARC CERER.hwtwl2 and SETER.de are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware generates a disrupting *sw_recoverable_error* trap to the requesting virtual processor. Assuming CERER.hwtwl2 is set and the DESR.f bit is clear, hardware encodes itl2c in the DESR.errortype field. The DESR.erroraddr field is undefined. Hardware corrects the error on the data being returned from the L2 cache but does not correct the L2 cache data itself. Software can correct the L2 error by invalidating the L2 line as appropriate.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.2 TTE Request for DTLB (DTL2C)

When a correctable ECC error is detected the error information (ldac, vcid, synd) is captured in the L2 Cache Error Status and PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen, and SPARC CERER.hwtwl2 and SETER.de are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware generates a disrupting *sw_recoverable_error* trap to the requesting virtual processor. Assuming CERER.hwtwl2 is set, and the DESR.f bit is clear, hardware encodes dtl2c in the DESR.errortype field. The DESR.erroraddr field is undefined. Hardware corrects the error on the data being returned from the L2 cache but does not correct the L2 cache data itself. Software can correct the L2 error by invalidating the L2 line as appropriate.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.3 Instruction Fetch Hit (ICL2C)

When a correctable ECC error is detected the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen, and SPARC CERER.icl2c and SETER.de bits are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware generates a disrupting *sw_recoverable_error* trap to the requesting virtual processor. Assuming CERER.icl2c is set, and the DESR.f bit is clear, hardware encodes icl2c in the DESR.errortype field. The DESR.erroraddr field is undefined. Hardware corrects the error on the data being returned from the L2 cache but does not correct the L2 cache data itself. Software can correct the L2 error by invalidating the L2 line as appropriate.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.4 Load Hit (DCL2C)

When a correctable ECC error is detected the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen, and SPARC CERER.dcl2c and SETER.de bits are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware generates a disrupting *sw_recoverable_error* trap to the requesting virtual processor. Assuming CERER.dcl2c is set and the DESR.f bit is clear, hardware encodes dcl2c in the DESR.errortype field. The DESR.erroraddr field is undefined. Hardware corrects the error on the data being returned from the L2 cache, but does not correct the L2 cache data itself. Software can correct the L2 error by invalidating the L2 line as appropriate.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.5 Prefetch Hit (L2C)

When a correctable ECC data error is detected the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error

Enable ceen and SPARC CERER.l2c_socc and SETER.de bits are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), hardware generates a disrupting *sw_recoverable_error* trap to the requesting virtual processor. Assuming the DESR.f bit is clear, hardware encodes l2c in the DESR.errortype field. The contents of the DESR.erroraddr field are undefined.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.6 Partial Store Hit (L2C)

When a correctable ECC error is detected the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status and L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen and SPARC CERER.l2c_socc and SETER.dhcce bits are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *hw_corrected_error* trap is generated to the requesting virtual processor. Assuming CERER.l2c_socc is set and the DESR.f bit is clear, hardware encodes L2C in the DESR.errortype field. The DESR.erroraddr field is undefined. Hardware corrects the error in the L2 cache and returns the corrected data.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.1.7 Atomic Hit (DCL2C)

When a correctable ECC error is detected on the read portion of the read-modify-write operation, the error information (ldac, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen and SPARC CERER.dcl2c and SETER.de bits are set (and PSTATE.ie is set or HPSTATE.hpriv is clear), a disrupting *sw_recoverable_error* trap is generated to the requesting virtual processor. Assuming CERER.dcl2c is set, and the DESR.f bit is clear, hardware encodes dcl2c in the DESR.errortype field. The DESR.erroraddr field is undefined.

Hardware may or may not have corrected the error, depending upon whether the error was completely contained in the data being updated as a result of a successful atomic operation. For example, if a CASA operation succeeds, and a bit in the 4 bytes updated by CASA had the correctable error, then hardware corrected the error simply by overwriting the data in error. The same is true for a successful LDSTUB (on the byte in error being overwritten) or SWAP (on the 4 bytes in error being overwritten) or CASXA (on the 8 bytes being overwritten).

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.2 L2 Cache Data Correctable ECC Error for Writeback (LDWC)

When a correctable ECC error is detected the error information (ldwc) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. The synd field is not captured in the L2 Cache Error Status register. Hardware corrects the error on the data being written to memory. In addition, if the L2 Cache Error Enable ceen bit is set, an L2C error is reported to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set, and the DESR.f bit is clear, hardware encodes L2C in the DESR.errortype field. The DESR.erroraddr field is undefined. If SETER.dhcce is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *hw_corrected_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

> If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.2.1 DMA Read

When a correctable ECC error is detected, the error information (ldrc, rw) is captured in the L2 Cache Error Status and the PA{39:6} is captured in the L2 Cache Error Address register. Hardware corrects the error on the data being returned from the L2 cache but does not correct the L2 cache data itself. In addition, if the L2 Cache Error Enable ceen bit is set, an L2C error is reported to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set and if the DESR.f bit is clear, hardware records the error in the DESR by encoding L2C. The contents of the DESR.erroraddr field are undefined. If SETER.de is also set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware takes a disrupting *sw_recoverable_error* trap.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

> If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.2.2 DMA Write Partial

When a correctable ECC error is detected on the read portion of the read-modify-write operation, the error information (ldrc, rw, synd) is captured in the L2 Cache Error Status register and the PA{39:4} is captured in the L2 Cache Error Address register. Hardware corrects the error in the L2 cache line. In addition, if the L2 Cache Error Enable ceen bit is set, an L2C error is reported to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set and the DESR.f bit is clear, hardware records the error in the DESR by encoding l2c. The contents of the DESR.erroraddr field are undefined. If SETER.dhcce is also set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *hw_corrected_error* trap is generated.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

L2 Data ECC is only checked for partial DMA stores. Aligned 4-byte and 8-byte (and larger) DMA writes overwrite previous contents and therefore never check data ECC.

> If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.3 L2 Cache Data Correctable ECC Error for Scrub (LDSC)

When a correctable ECC error is detected on a data array scrub, the error information (ldsc) is captured in the L2 Cache Error Status register and the {way[3:0],Index[8:0]} is captured in the L2 Cache Error Address register. Hardware corrects the error in the L2 cache line by writing back the corrected data and parity (this rewrite will be unable to correct a permanently failed bit). In addition, if the L2

Cache Error Enable ceen bit is set, an L2C error is reported to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set and the DESR.f bit is clear, hardware records the error in the DESR by encoding l2c. The contents of the DESR.erroraddr field are undefined. If SETER.dhcce is also set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *hw_corrected_error* trap is generated.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note | If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.4 L2 Cache Tag Correctable ECC Error (LTC)

UltraSPARC T2 provides SEC ECC on the L2 cache tags.

On every L2 access, ECC is checked for all 16 tags in the set. When a correctable ECC error is detected the error information (LTC) is captured in the L2 Cache Error Status register and the PA{21:6} is captured in the L2 Cache Error Address register. Note that the Syndrome is *not* captured for a L2 cache tag ECC error. Hardware corrects all errors in all the tags in the set. In addition, if the L2 Cache Error Enable ceen bit is set, hardware generates an L2C error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set and the DESR.f bit is clear, hardware records the error in the DESR by setting L2C. The contents of the DESR.erroraddr field are undefined. If SETER.dhcce is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *hw_corrected_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Implementation Notes | Hardware generates the error report to the core on detection of the error .However the error would get corrected by hardware only if the access that detected the error was a miss in L2. Since the error can be reported multiple times on L2 tag hits and not get corrected, software should force the correction by issuing Prefetch ICE instruction to that index . If the error is due to a hard failure in the tag, hardware or software will not be able to complete the correction and no further accesses will be able to be processed by the L2 (that is, the L2 bank is hung.)

Note | If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.5 L2 Cache VUAD Correctable ECC Error (LVC)

On every L2 access, ECC is checked for all 32 vd (valid, dirty) bits and 32 ua (used, allocate) bits in the set.[1] When a correctable ECC single bit error is detected, the error information (LVC, SYND) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. Note that the Syndrome is captured for a L2 cache VUAD ECC error. Hardware corrects the single bit error in all the vd and ua bits in the set. In addition, if the L2 Cache Error Enable ceen bit is set, hardware generates an L2C error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2c_socc bit is set, and the DESR.f bit is clear, hardware records the error in the DESR by setting l2c. The contents of the DESR.erroraddr field are undefined. If SETER.dhcce is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *hw_corrected_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note | If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

Implementation Note | For instructions that require multiple passes through the L2 pipeline, VUAD ECC is only checked on the first pass.

1. Even though the used bits need not be ECC protected (since their value is noncritical: any error in the used bits will cause potentially different replacement order, but still functionally correct operation), since vd and ua arrays are implemented out of the same Register File array, UltraSPARC T2 protects both the used bits and the allocate bits with ECC.

16.9.6 L2 Cache Data Uncorrectable ECC Error for Access (LDAU)

Uncorrectable data errors are detected by the L2 Cache, captured in the L2 Cache Error Status register and the L2 Cache Error Address register, and then reported to the requesting virtual processor. The following control bits are used:

L2 Cache Error Enable nceen bit — When set, causes the error to be sent to the requesting virtual processor; when reset, error is ignored, but still captured in the L2 Cache Error Status register and the L2 Cache Error Address register.

Associated CERER bit — When set (along with nceen and SETER.pscce bit) causes L2$ error to be captured in I/DSFSR register, and conditions reporting the precise error trap; when disabled, L2$ error is ignored, I/DSFSR is not changed, the precise error trap is lost, and processing continues with bad data.

SETER.pscce bit — When set (along with nceen and associated CERER bit), causes L2$ error to be captured in I/DSFSR register, and conditions reporting the precise error trap to be presented to virtual processor. When SETER.pscce is reset, the precise error trap is lost, and processing continues with bad data.

The error information stored in the L2Cache Error Status register is shown in TABLE 16-21, and the physical address captured in the L2 Cache Error Address register is shown in TABLE 16-25. They are recorded as LDAU errors.

16.9.6.1 TTE Request for ITLB (ITL2U)

When an uncorrectable ECC data error is detected, the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.hwtwl2 bits are set, a precise *instruction_access_MMU_error* trap is generated to the requesting virtual processor. The ISFSR will have the itl2u encoding set. The VA of the instruction fetch is available in TPC[TL]. Software can attempt recovery.

Note | If nceen or CERER.hwtwl2 are not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.6.2 TTE Request for DTLB (DTL2U)

When an uncorrectable ECC data error is detected, the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable ceen and SPARC CERER.hwtwl2 bits are set, a precise *data_access_MMU_error* trap is generated to the requesting virtual processor. The

DSFSR will have the dtl2u encoding set. The VA of the data access is *not* logged in the D-SFAR. Bits {47:13} of the VA are available in the tag access register. Software must search the L2 ESRs to determine the failing physical address.

Note | If nceen or CERER.hwtwl2 are not set, the error is not reported and no trap is generated. Hardware continues executing with the (invalid) data read from L2.

16.9.6.3 Instruction Fetch Hit (ICL2U)

When an uncorrectable ECC data error is detected, the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. If the L2 Cache Error Enable nceen and SPARC CERER.icl2u bits are set, the line is loaded into the L1 instruction cache with bad parity. In addition, if the SPARC SETER.pscce bit is set, a precise *instruction_access_error* trap is generated to the requesting virtual processor. The ISFSR will contain icl2u, and the VA of the instruction fetch is logged in TPC[TL]. Software can attempt recovery.

Note | If nceen, CERER.icl2u, or SETER.pscce are not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.6.4 Load Hit (DCL2U)

When an uncorrectable ECC data error is detected, the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. If the L2 Cache Error Enable nceen and SPARC CERER.dcl2u bits are set, the line is loaded into the L1 data cache with bad parity. Additionally, if the SPARC SETER.pscce bit is set, a precise *data_access_error* trap is generated to the requesting virtual processor. The D-SFSR will contain dcl2u. The VA of the data access is *not* logged in the D-SFAR. Software must search the L2 ESRs to determine the failing physical address.

Note | If nceen, CERER.dcl2u, or SETER.pscce is not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.6.5 Atomic Hit (DCL2U)

When an uncorrectable ECC data error is detected, the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable nceen, and SPARC CERER.dcl2u and SETER.pscce bits are set, hardware records the error in the DSFSR by encoding dcl2u. The VA of the data access is *not*

logged in the D-SFAR. Software must search the L2 ESRs to determine the failing physical address. Hardware generates a precise *data_access_error* trap to the requesting virtual processor. The atomic operation will not complete its update, leaving the original data and the bad ECC unchanged. Software can attempt recovery.

Note | If nceen, CERER.dcl2u, or SETER.pscce is not set, hardware neither records the error in the DSFSR nor takes a trap.

16.9.6.6 Prefetch Hit (L2U)

When an uncorrectable ECC data error is detected the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.l2u_socu are set and the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the requesting virtual processor.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.6.7 Partial Store Hit (L2U)

When an uncorrectable ECC data error is detected on the read portion of the read-modify-write operation (assuming NotData is not set), the error information (ldau, rw, vcid, moda, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.l2u_socu bits are set, and if the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the requesting virtual processor. The partial store does not complete its update, leaving the original data and the bad ECC unchanged. NotData will *not* be set.

If SETER.dhcce is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note | If either the L2 Cache Error Enable nceen or CERER.l2u_socu bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.7 L2 Cache Data Uncorrectable ECC Error for Writeback (LDWU)

When an uncorrectable ECC error is detected the error information (ldwu) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the L2 Cache Error Address register. The synd field is not captured in the L2 Cache Error Status register. Hardware indicates the error on the data being written to memory, and the DRAM controller writes the data back with signaling ECC. The specific signaling ECC used is described in Appendix G, *ECC Codes*. In addition, if the L2 Cache Error Enable nceen is set, hardware generates an L2U error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2u_socu bit is set, and if the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note | If either the L2 Cache Error Enable nceen or CERER.l2u_socu bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.8 L2 Cache Data Uncorrectable ECC Error for DMA (LDRU)

16.9.8.1 DMA Read

When an uncorrectable ECC error is detected, the error information (ldru, rw) is captured in the L2 Cache Error Status register and the PA{39:4} is captured in the L2 Cache Error Address register. The synd field is not captured for DMA reads. Hardware returns the data with a UE error indicator back to the DMA requestor. In addition, if the L2 Cache Error Enable nceen bit is set, hardware generates an L2U error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC

CERER.l2u_socu bit is set, and the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note | If either the L2 Cache Error Enable nceen or CERER.l2u_socu bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.8.2 DMA Write Partial

When an uncorrectable ECC error is detected on the read portion of the read-modify-write operation, the error information (ldru, rw, synd) is captured in the L2 Cache Error Status register and the PA{39:4} is captured in the L2 Cache Error Address register. If the L2 Cache Error Enable nceen bit is set, hardware generates an L2U error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

The DMA write partial will not complete its update, leaving the original data and the bad ECC unchanged. NotData is *not* set.

Note | If either the L2 Cache Error Enable nceen or CERER.l2u_socu bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.9 L2 Cache Data Uncorrectable ECC Error for Scrub (LDSU)

When an uncorrectable data ECC error is detected on a data array scrub, the error information (ldsu) is captured in the L2 Cache Error Status register and the {way[3:0],Index[8:0]} is captured in the L2 Cache Error Address register. In addition, if the L2 Cache Error Enable nceen bit is set, hardware generates an L2U error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2u_socu bit is set, and the DESR.f bit is clear, hardware records the error in the DESR by setting l2u. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

NotData is *not* set.

Note | If either the L2 Cache Error Enable nceen or CERER.l2u_socu bit is reset, the error is ignored, but the error information is still captured in the L2 Cache Error Status and L2 Cache Error Address registers.

16.9.10 L2 Cache Tag Uncorrectable ECC Error

UltraSPARC T2 provides SEC ECC on the L2 cache tags, thus multiple bit ECC errors are not detected (unless they happen to generate an illegal single-bit error syndrome). An L2 Cache scrub mechanism is provided to detect single bit errors and to correct the single bit errors before they can become double bit errors. However, the scrubber only works on accessed indices. If an index is not accessed for a long time, it is susceptible to double-bit errors. The probability of this is very low.

Note | To ensure that all lines are routinely accessed, software can provide a “software scrub” of each index.

16.9.11 L2 Cache VUAD Uncorrectable ECC Error (LVF)

On every L2 access, ECC is checked for all 32 vd bits and 32 ua bits in the set. When an uncorrectable ECC multiple bit error is detected, the error information (lvf, synd) is captured in the L2 Cache Error Status register and the PA{39:6} is captured in the

L2 Cache Error Address register. Note that the Syndrome is captured for a L2 cache VUAD ECC error. In addition, this fatal error generates a *warm_reset* trap to the entire chip.

Implementation Note | For instructions that require multiple passes through the L2 pipeline, VUAD ECC is only checked on the first pass.

16.9.12 L2 Cache Directory Uncorrectable Parity Error (LRF)

During every L2 store operation, parity is checked for the directory entry. When an uncorrectable parity error is detected, the error information LRF is captured in the L2 Cache Error Status register. The L2 Cache Error Address register captures the following directory index information in the address field:

- bits{15:12} — panel
- bits{11:9} — strandid
- bits{8:7} — l1way{1:0}
- bit{6} — icachedir

If icachedir = 1, panel = PA{10, 9, 5, l1way2{2}}; else panel = PA{10, 9, 5, 4}

In addition, this fatal error generates a *warm_reset* trap to the entire chip.

16.9.13 L2 Cache Data NotData Error for Processor Access (NDSP)

NotData errors are detected by the L2 Cache, captured in the L2 NotData Error register and then reported to the requesting virtual processor. The following control bits are used:

L2 Cache Error Enable nceen bit — When set, causes the error to be sent to the requesting virtual processor; when reset, error is ignored, but still captured in the L2 NotData Error register.

Associated CERER bit — When set (along with nceen and SETER.pscce bit) causes L2$ error to be captured in I/DSFSR register, and conditions reporting the precise error trap; when disabled, L2$ error is ignored, I/DSFSR is not changed, the precise error trap is lost, and processing continues with bad data.

SETER.pscce bit — When set (along with nceen and associated CERER bit), causes L2$ error to be captured in I/DSFSR register, and conditions reporting the precise error trap to be presented to the virtual processor. When SETER.pscce is reset, the precise error trap is lost, and processing continues with bad data.

The error information stored in the L2 NotData Error register is shown in TABLE 16-27 on page 285. The errors are recorded as NDSP errors.

16.9.13.1 TTE Request for ITLB (ITL2ND)

When a NotData error is detected on a TTE access for an instruction fetch, the error information (ndsp, rw, vcid, PA{39:4}) is captured in the L2 NotData Error register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.hwtwl2 bits are set, a precise *instruction_access_MMU_error* trap is generated to the requesting virtual processor. The ISFSR will have the itl2nd encoding set. The VA of the instruction fetch is available in TPC[TL]. Software can attempt recovery.

Note | If nceen or CERER.hwtwl2 are not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.13.2 TTE Request for DTLB (DTL2ND)

When a NotData error is detected on a TTE access for a load or store, the error information (ndsp, rw, vcid, PA{39:4}) is captured in the L2 NotData Error register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.hwtwl2 bits are set, a precise *data_access_MMU_error* trap is generated to the requesting virtual processor. The DSFSR will have the dtl2nd encoding set. The VA of the data access is *not* logged in the D-SFAR. Bits 47:13 of the VA are available in the tag access register. Software must search the L2 ESRs to determine the failing physical address.

Note | If nceen or CERER.hwtwl2 are not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.13.3 Instruction Fetch (ICL2ND)

When a NotData error is detected on an instruction fetch access, the error information (ndsp, rw, vcid, PA{39:4}) is captured in the L2 NotData Error register. If the L2 Cache Error Enable nceen and SPARC CERER.icl2nd bits are set, the line is loaded into the L1 instruction cache with bad parity. In addition, if the SPARC SETER.pscce bit is set, a precise *instruction_access_error* trap is generated to the requesting virtual processor. The ISFSR will contain icl2nd, and the VA of the instruction fetch is logged in TPC[TL].

Note | If nceen, CERER.icl2nd, or SETER.pscce are not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.13.4 Load Hit (DCL2ND)

When a NotData error is detected on a load access, the error information (ndsp, rw, vcid, PA{39:4} is captured in the L2 NotData Error register. If the L2 Cache Error Enable nceen and SPARC CERER.dcl2nd bits are set, the line is loaded into the L1 data cache with bad parity. In addition, if the SPARC SETER.pscce bit is set, a precise *data_access_error* trap is generated to the requesting virtual processor. The DSFSR will contain dcl2nd. The VA of the data access is *not* logged in the D-SFAR. Software must search the L2 ESRs to determine the failing physical address.

Note | If nceen, CERER.dcl2nd, or SETER.pscce is not set, the error is not reported and no trap is generated. Hardware continues executing using the (invalid) data read from L2.

16.9.13.5 Atomic Hit (DCL2ND)

When a NotData error is detected on an atomic access, the error information (ndsp, rw, vcid, PA{39:4}) is captured in the L2 NotData Error register. In addition, if the L2 Cache Error Enable nceen, and SPARC CERER.dcl2nd and SETER.pscce bits are set, hardware records the error in the DSFSR by encoding dcl2nd. The VA of the data access is *not* logged in the D-SFAR. Software must search the L2 ESRs to determine the failing physical address. Hardware generates a precise *data_access_error* trap to the requesting virtual processor. The atomic operation will not complete its update, leaving the original data unchanged and marked with NotData.

If nceen, CERER.dcl2nd, or SETER.pscce is not set, hardware neither records the error in the DSFSR nor takes a trap. It continues executing using the (invalid) data read from L2.

16.9.13.6 Prefetch Hit (L2ND)

When a NotData error is detected on a prefetch access, the error information (ndsp, rw, vcid, PA{39:4}) is captured in the L2 NotData Error register. In addition, if the L2 Cache Error Enable nceen and SPARC CERER.l2nd bits are set and the DESR.f bit is clear, hardware records the error in the DESR by encoding l2nd. The contents of the DESR.erroraddr field are undefined. If SETER.de is set and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the requesting virtual processor.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.13.7 Partial Store Hit (L2ND)

When a NotData error is detected on the read portion of the store, the partial store does not complete its update, leaving the original data unchanged, marked with NotData. No trap is presented, and the L2 Notdata Error register is not updated.

16.9.14 L2 Cache Data NotData Error for DMA Access (NDDM)

16.9.14.1 DMA Read (L2ND)

When a NotData error is detected on a DMA read, the error information (nddm, rw, vcid = L2_CSR_REG.errorsteer, PA{39:4}) is captured in the L2 Notdata Error register. Hardware returns the data with a UE error indicator back to the DMA requestor. In addition, if the L2 Cache Error Enable nceen bit is set, hardware generates an L2ND error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2nd bit is set, and if the DESR.f bit is clear, hardware records the error in the DESR by setting l2nd. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.9.14.2 DMA Write Partial (L2ND)

When a NotData error is detected the error information (nddm, rw, vcid = L2_CSR_REG.errorsteer, PA{39:4}) is captured in the L2 NotData Error register. If the L2 Cache Error Enable nceen bit is set, hardware generates an L2ND error to the virtual processor specified in L2_CSR_REG.errorsteer. If the SPARC CERER.l2nd bit is set, and if the DESR.f bit is clear, hardware records the error in the DESR by setting l2nd. The contents of the DESR.erroraddr field are undefined. If SETER.de is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, a disrupting *sw_recoverable_error* trap is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If SETER.de is not set, or PSTATE.ie is not set and HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

The DMA write partial will not complete its update, leaving the original data unchanged marked with NotData.

16.9.15 L2 Cache Data NotData Error for Writeback

When a NotData error is detected on an eviction, no error information is logged since the NotData is simply being moved to memory. Hardware indicates the error on the data being written to memory, where the DRAM controller will write back the data with signaling ECC.

16.9.16 L2 Software Error Scrubbing Support

Errors which are software recoverable leave the L2 cache with a correctable error, which then needs to be scrubbed to prevent repetitive traps for the same soft error. Flushing the data from the cache to memory will cause the error to be corrected. UltraSPARC T2 provides a mechanism for L2 flushing through a variant of prefetch called invalidate cache entry (PrefetchICE), described in *L2 Cache Flushing* on page 437.

16.10 L2 Error Registers

16.10.1 L2 Error Enable Register

Each L2 bank has an Error Enable register that controls the reporting of L2 errors for that bank back to the initiator of an operation (or to the virtual processor specified in L2_CSR_REG.errorsteer if no initiator exists or can be readily identified). The L2 Error Enable register, the format of which is shown in TABLE 16-20, is available at address AA 0000 0000_{16} or BA 0000 0000_{16}. Address bits 8:6 select the cache bank, and address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

TABLE 16-20 Error Enable Register – L2_ERROR_EN_REG (AA 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2	debug_trig_en	0	RW	Trigger enable on L2 cache errors for debug.
1	nceen	0	RW	If set to 1, report uncorrectable errors.
0	ceen	0	RW	If set to 1, report correctable errors.

Note | Errors are always logged in the L2_ERROR_STATUS_REG and L2_ERROR_ADDRESS_REG regardless of the setting of the nceen and ceen bits in L2_ERROR_EN_REG. L2_ERROR_EN_REG only controls whether or not the error is reported back to the appropriate virtual processor (the requestor or the virtual processor specified in L2_CSR_REG.errorsteer).

16.10.2 L2 Error Status Register

Each L2 bank has an Error Status register that contains status on L2 errors for that bank. The status bits in this register are cleared by writing a 1 to the bit. The error register is not cleared on reset, so software can examine its contents after an error-induced reset. The L2 Error Status register, the format of which is shown in TABLE 16-21, is available at address AB 0000 0000_{16} or BB 0000 0000_{16}. Address bits 8:6 select the cache bank, and address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

Note | L2_ERROR_STATUS is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

TABLE 16-21 L2 Error Status Register – L2_ERROR_STATUS_REG (AB 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	POR Value	R/W	Description
63	meu	0	R/W1C	Multiple uncorrected or fatal errors (one or more uncorrected or fatal errors were not logged).
62	mec	0	R/W1C	Set by *any* error detected after vec = 1 (one or more corrected errors were not logged).
61	rw	0	RW	Specifies whether the error access was a read or write. Set to 1 for a write, 0 for a read.
60	—	0	RW	Set to 0.
59:54	vcid	0	RW	ID of virtual processor that encountered error.
53	ldac	0	R/W1C	Set to 1 if the error was an L2 cache data array access correctable error.
52	ldau	0	R/W1C	Set to 1 if the error was an L2 cache data array access uncorrectable error.
51	ldwc	0	R/W1C	Set to 1 if the error was an L2 cache data array writeback correctable error.
50	ldwu	0	R/W1C	Set to 1 if the error was an L2 cache data array writeback uncorrectable error.
49	ldrc	0	R/W1C	Set to 1 if the error was an L2 cache data array dma access correctable error.
48	ldru	0	R/W1C	Set to 1 if the error was an L2 cache data array dma access uncorrectable error.
47	ldsc	0	R/W1C	Set to 1 if the error was an L2 cache data array scrub correctable error.
46	ldsu	0	R/W1C	Set to 1 if the error was an L2 cache data array scrub uncorrectable error.

TABLE 16-21 L2 Error Status Register – L2_ERROR_STATUS_REG (AB 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	POR Value	R/W	Description
45	ltc	0	R/W1C	Set to 1 if the error was an L2 cache tag array correctable error.
44	lrf	0	R/W1C	Set to 1 if the error was an L2 cache directory uncorrectable error.
43	lvf	0	R/W1C	Set to 1 if the error was an L2 cache VUAD uncorrectable error.
42	dac	0	R/W1C	Set to 1 if the error was a DRAM access correctable error.
41	dau	0	R/W1C	Set to 1 if the error was a DRAM access uncorrectable error.
40	drc	0	R/W1C	Set to 1 if the error was a DRAM dma access correctable error.
39	dru	0	R/W1C	Set to 1 if the error was a DRAM dma access uncorrectable error.
38	dsc	0	R/W1C	Set to 1 if the error was a DRAM scrub correctable error or a recoverable DRAM link error.
37	dsu	0	R/W1C	Set to 1 if the error was a DRAM scrub uncorrectable error or an unrecoverable DRAM link error.
36	vec	0	R/W1C	Set to 1 if the register contains a valid correctable error.
35	veu	0	R/W1C	Set to 1 if the register contains a valid uncorrectable or fatal error.
34	lvc	0	R/W1C	Set to 1 if the error was a L2 cache VUAD correctable error.
33:28	—1	0	RO	*Reserved*
27:0	synd	X	RW	Parity or ECC syndrome.

The vec bit is set on all cycles where a correctable error is encountered. The veu bit is set on all cycles where an uncorrectable or fatal error is encountered.

The synd field is only valid for LVF, LVC, LDAC, LDAU, LDRU and LDRC errors. The rw bit will be set to 1 for a partial store, DMA write, and atomic load/store operation if they detect an error while performing the read part of the L2 read-modify-write operation. For DAC and DAU errors, if an L2 fill happens before the data is returned to the requestor, the DAC or DAU error will be reported to the virtual processor specified in L2_CSR_REG.errorsteer, and the rw bit will be 0. The vcid field is valid only for LDAC, LDAU, DAC, and DAU errors where the error is detected synchronous with the load or store operation. If the error is reported at the time of the L2 cache fill operation, vcid will contain 0. For all other error types, this field is not valid.

TABLE 16-22 summarizes the rw, vcid, moda, and synd fields.

TABLE 16-22 rw, vcid, moda, and synd Fields Summary

Error	rw	vcid	moda	synd
During L2 fill for DAC, DAU, DRC, DRU	X	L2_CSR_REG. errorsteer	X	X
LDAC, LDAU	1 for atomics and partial stores, else 0	VCID	0	synd_127_96{6:0}, synd_95_64{6:0}, synd_63_32{6:0}, synd_31_0{6:0}
LDWC, LDWU	X	X	X	X
LDRC, LDRU	1 for write, else 0	X	X	synd_127_96{6:0}, synd_95_64{6:0}, synd_63_32{6:0}, synd_31_0{6:0} for write, X for read
LDSC,LDSU	X	X	X	X
LTC	X	X	X	X
LRF	X	X	X	X
LVC, LVF	X	L2_CSR_REG. errorsteer	X	14'b0,synd_vd{6:0},synd_ua{6:0}
DAC, DAU	1 for atomics and stores, else 0	vcid	0	X
DRC, DRU	1 for write, else 0	X	X	X

The dsc and dsu bits are logged in L2 Error Status register to notify software to check the DRAM error registers, and are set regardless of the status of the other bits. Setting dsc and/or dsu does not cause any logging of the error address or syndrome in the L2 Error Address register. It also does not update the vec/veu/mec/meu bits.

Note | A dsc or dsu error will not be reported to the core responsible fo r handling the error until an L2 miss occurs. The L2 cannot send unsolicited packets to the cores, and the L2 miss response packet has space to report these types of errors.

Note | The syndrome is *not* logged on a L2 tag correctable error as well as for NotData errors.

With regard to the remaining ue, ce bits, if multiple errors occur in the same cycle, the meu or mec bit is set and only the highest-priority error is logged based on the following priority shown in TABLE 16-23 (errors with the same priority are mutually exclusive).

TABLE 16-23 Priority for Simultaneous FE, UE, and CE Errors

Error	Priority	Bit Set if Higher Priority Error
LVF (FE)	1	Not Applicable
LRF (FE)	2	meu
LDAU (UE)	3	meu
LDSU (UE)	3	meu
LDWU (UE)	4	meu
LDRU (UE)	5	meu
DAU (UE)	6	meu
DRU (UE)	6	meu
LVC (CE)	7	mec
LTC (CE)	8	mec
LDAC (CE)	9	mec
LDSC (CE)	9	mec
LDWC (CE)	10	mec
LDRC (CE)	11	mec
DAC (CE)	12	mec
DRC (CE)	12	mec

The syndrome, rw, moda, vcid, and address are captured for the highest-priority error in that cycle.

If FE, UE, or CE errors occur in a cycle when an error status bit is already set (indicating a previous error exists that hasn't been cleared from the error status register), the information in TABLE 16-24 applies.

TABLE 16-24 UE,CE Errors Occurring in a Cycle Where Error Status Bit Already Set

Existing Error	Error	Priority	Bit Set if Highest-Priority Error	Bit Set if Higher-Priority Error in Same Cycle
LVF/LRF	LVF	1	meu	Not Applicable
LVF/LRF	LRF	2	meu	meu
LDAU/LDSU/LDWU/LDRU/DRU/DAU	LVF	1	lvf, meu	Not Applicable
LDAU/LDSU/LDWU/LDRU/DRU/DAU	LRF	2	lrf, meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDAU	3	meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDSU	3	meu	meu

TABLE 16-24 UE,CE Errors Occurring in a Cycle Where Error Status Bit Already Set

Existing Error	Error	Priority	Bit Set if Highest-Priority Error	Bit Set if Higher-Priority Error in Same Cycle
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDWU	4	meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDRU	5	meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	DAU	6	meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	DRU	6	meu	meu
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LVC	7	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LTC	8	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDAC	9	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDSC	9	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDWC	10	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	LDRC	11	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	DRC	12	mec	mec
LVF/LRF/LDAU/LDSU/LDWU/LDRU/DRU/DAU	DAC	12	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LVF	1	lvf, mec	Not Applicable
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LRF	2	lrf, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDAU	3	ldau, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDSU	3	ldsu, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDWU	4	ldwu, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDRU	5	ldru, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	DAU	6	dau, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	DRU	6	dru, mec	meu
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LVC	7	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LTC	8	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDAC	9	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDSC	9	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDWC	10	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	LDRC	11	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	DRC	12	mec	mec
LVC/LTC/LDAC/LDSC/LDWC/LDRC/DRC/DAC	DAC	12	mec	mec

For the cases above where the “Bit Set If Highest-Priority Error” column contains a value besides mec and meu, the syndrome, rw, moda, vcid, and address for the highest-priority error in that cycle will overwrite the existing syndrome, rw, moda, vcid, and address.

Once set, error status bits are only cleared by software. Hardware will never clear a set status bit. If a software write of the error register happens on the same cycle as an error, the setting of bits by the error will be based on the register state before the write, following the rules of TABLE 16-24. The setting of fields by the error will take

precedence over the same field being updated by the write; however, fields that are not changed by the error will be updated by the write (for example, if the register has the vec and ltc bits set and software does a write to clear those bit on the same cycle as a L2 cache data uncorrectable error, the error register would end up with the veu and ldau bits set, the vec and ltc bits cleared, and the rw, moda, vcid, and synd fields would contain the values for the LDAU error). The rw, moda, vcid, and synd fields are always considered to be set by an error, even if the value they are being set to is undefined (that is, set to X in TABLE 16-22).

Note When writing the error status register on the same cycle that another error occurs, the error status register state before the register write is applied determines which bits will be updated in the register. If the error occurring on the same cycle as the write is same or lower priority than the error currently logged in the register that is being cleared by the write, this will result in the error status register having only either the vec and mec pair of bits set (for a correctable error) or veu and meu pair of bits set (for an uncorrectable error) after both the new error and the write which clears the old error bits are applied.

Note In case VEC and MEC bits are set and there is a write to the error status register to clear the set bits, if a UE or CE occurs in the same cycle of the write that would set MEC bit according to TABLE 16-24 on page 280, the MEC bit will not get cleared but VEC bit will get cleared. A subsequent read of the error status register would see MEC bit set without VEC bit set . Similarly , in case VEU and MEU bits are set and there is a write to the error status register to clear the set bits, if a UE occurs in the same cycle of the write that would set MEU bit according to TABLE 16-24 on page 280, the MEU bit will not get cleared but VEU will get cleared. A subsequent read of the error status register would see MEU bit set without VEU bit set.

Programming Note

To minimize the possibility of missing notification of an error, software should clear any multiple error indication as soon as possible, since UltraSPARC T2 provides no indication of the number of multiple errors represented by the multiple error bit.

An example of clearing behavior for the case where an DAC error is followed closely by an DAU error would be to first log that an DAC error was seen (which is indicated by the dac bit still being set in the Error Status register), and then to do a write to the Error Status register with bits 62, 42 and 36 set to clear the mec, vec, and dac bits. The vcid, moda, rw, and synd fields of error status would then be captured to memory or a register, the physical address for the DAU would be read from the Error Address register to memory or a register, and software could do a write to the Error Status register with bits 41 and 35 set to clear the dau and veu bits and put the error status register back in a state where it can capture full error information.

Software would then invoke the code to shoot down all the TLB entries for the page with the bad cache line, force the bad cache line from L2 to memory, and kill all processes that had access to the bad cache line. If another correctable error happened after the DAU but before the write that cleared the mec, vec, and dac bits, that error would not be logged. If another correctable error happened simultaneously to or after the write that cleared the mec, vec, and dac bits, but before or simultaneously to the write that cleared the dau and veu bits, that error would be captured by the mec and vec bits being set after the dau/veu-clearing write. If another uncorrectable error happened after the DAU, but before or simultaneously to the write that cleared the dau and veu bits, that error would be captured by the meu and veu bits being set after the dau/veu-clearing write.

16.10.3 L2 Error Address Register

Each L2 bank has an Error Address register that contains the address for the L2 error within that bank. The error register is not cleared on reset, so software can examine its contents after an error-induced reset. The L2 Error Address register is available at address AC 0000 0000_{16} or BC 0000 0000_{16}. Address bits 8:6 select the cache bank, and address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

Implementation Note

Within the L2 cache itself, PA {17:9} indicates the set, and PA {8:6} indicates the bank, if all eight banks are enabled. See *L2 Bank Enable* on page 433 for details on which bits are used to select the set and the bank if fewer than all eight L2 banks are enabled.

Programming Note | If L2_IDX_HASH_EN_STATUS.enb_hp = 1, the address stored in the L2_ERROR_ADDRESS_REG will contain the hashed PA{17:11} ← {(PA{32:28} **xor** PA{17:13}) **::** (PA{19:18} **xor** PA{12:11})}.

TABLE 16-25 shows the format of the L2 Error Address register.

TABLE 16-25 L2 Error Address Register – L2_ERROR_ADDRESS_REG (AC 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	POR Value	R/W	Description
39:4	address	X	RW	Error address.
3:0	—	0	RO	*Reserved*

Note | L2_ERROR_ADDRESS is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

TABLE 16-26 lists the bits captured for each of the FE/UE/CE error types.

TABLE 16-26 Bits Captured for Each Error Type

Error	Address Bits	Address Contents
LDAC	39:4	Physical address of quadword accessed.
LDAU	39:4	Physical address of quadword accessed.
LDWC	39:6	Physical address of cache line.
LDWU	39:6	Physical address of cache line.
LDRC (rw = 0)	39:6	Physical address of cache line.
LDRC (rw = 1)	39:4	Physical address of quadword accessed.
LDRU (rw = 0)	39:6	Physical address of cache line.
LDRU (rw = 1)	39:4	Physical address of quadword accessed.
LDSC	21:9	Cache index (21:18 way, 17:9 set).
LDSU	21:9	Cache index (21:18 way, 17:9 set).
LTC	39:4	Physical address of quadword accessed.
LRF	15:6	Directory index (15:12 panel, 11:9 CoreID, 8:7 L1way{1:0}, 6 IcacheDir). If IcacheDir = 1, panel = address{10,9,5,L1way{2}}, else panel = address{10,9,5,4}.
LVF	39:4	Physical address of quadword accessed.
LVC	39:4	Physical address of quadword accessed.

TABLE 16-26 Bits Captured for Each Error Type *(Continued)*

Error	Address Bits	Address Contents
DAC	39:6	Physical address of cache line.
DAU	39:6	Physical address of cache line.
DRC	39:6	Physical address of cache line.
DRU	39:6	Physical address of cache line.

For a given error, the unused bits in the address field are not guaranteed to be zero, and should be masked by software. This register is writable by software for register diagnostics and isn't expected to be written during normal operation. However, in the event it is written on the same cycle that an error is reported, the update from the error will take precedence over the write.

16.10.4 L2 NotData Error Register

Each L2 bank has a NotData Error register that contains the status and address for the L2 NotData error within that bank. The error register is not cleared on reset, so software can examine its contents after an error-induced reset. The L2 NotData Error register is available at address AE 0000 0000_{16} and BE 0000 0000_{16}. Address bits 8:6 select the cache bank, address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

Programming Note | If L2_IDX_HASH_EN_STATUS.enb_hp = 1, the address stored in the L2_NOTDATA_ERROR_REG will contain the hashed PA{17:11} ← {(PA{32:28} **xor** PA{17:13}) **::** (PA{19:18} **xor** PA{12:11})}.

TABLE 16-27 shows the format of the L2 NotData Error register.

TABLE 16-27 L2 NotData Error Register – L2_NOTDATA_ERROR_REG (AE 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
63:52	—	0	RO	*Reserved*
51	mend	0	R/W1C	Multiple NotData errors, one or more NotData errors were not logged.
50	rw	X	RW	Specifies whether the NotData error access was a read or write. Set to 1 for a write or atomic, 0 for a read.
49	ndsp	0	R/W1C	Set to 1 if the error was a L2 cache data array access NotData error (from SPARC core).
48	nddm	0	R/W1C	Set to 1 if the error was a L2 cache data array Dma access NotData error.
47:46	—	0	RO	*Reserved*

TABLE 16-27 L2 NotData Error Register – L2_NOTDATA_ERROR_REG (AE 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
45:40	vcid	X	RW	ID of virtual processor that encountered error for ndsp, or L2_CSR_REG.errorsteer for nddm.
39:4	address	X	RW	Error address. Physical address of quadword accessed.
3:0	—	0	RO	*Reserved*

Note | L2_NOTDATA_ERROR is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

This register is writable by software for register diagnostic reasons and isn't expected to be written during normal operation. However, in the event it is written on the same cycle that an error is reported, the update from the error will take precedence over the write.

If multiple errors occur in the same cycle, the mend bit is set and only the highest-priority error is logged based on the following priority shown in TABLE 16-28.

TABLE 16-28 Priority for Simultaneous NotData Errors

Error	Priority	Bit Set if Higher Priority Error
NDSP	1	Not Applicable
NDDM	2	mend

The rw, vcid, and address are captured for the highest-priority NotData error in that cycle.

If NotData errors occur in a cycle when an error status bit is already set (indicating a previous error exists that hasn't been cleared from the error status register), the mend bit is set.

16.10.5 L2 Error Injection Register

Each cache bank has an error injection register for use in injecting errors to test error functionality or error handling code. L2 tag, VUAD, and data array errors can be injected via the diagnostic access, so the L2 error injection register only provides for the injection of directory parity errors. Errors can be injected either single/double-shot or continuously (due to restrictions in implementation, a true single-shot mode was not possible, because in a few cases the hardware must introduce two errors back-to-back). Once the enb_hp bit set, either the first (and possibly second) subsequent operation (for sdshot = 1), or all subsequent operations (for sdshot = 0)

that cause a directory update will result in the parity of the directory entry to be inverted. When in single- or double-shot mode, after the injected error(s) are generated, the enb bit is automatically reset by the hardware to 0. The L2 Error Injection register, the format of which is shown in TABLE 16-29, is available at address AD 0000 0000_{16} or BD 0000 0000_{16}. Address bits 8:6 select the cache bank, address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

TABLE 16-29 L2 Error Injection Register – L2_ERROR_INJECT_REG (AD 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	POR Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	sdshot	0	RW	Controls type of error injection. 1 = single or double shot; 0 = continuous.
0	enb_hp	0	RW	Enables directory error injection.

Note | L2_ERROR_INJECT_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.11 DRAM Error Descriptions

Each 128-bit data block in memory is protected by QEC/OED (Quad Error Correct, Octal Error Detect) ECC. This ECC code supports Extended ECC for x4 DRAM chips, where the complete failure of any aligned 4-bit block can be corrected, and any error where exactly two 4-bit blocks are in error is recognized as an uncorrectable error.

L2 cache handling of DRAM detected errors for loads is processed differently, depending on the following conditions:

- If the error is detected on the bytes of the line requested by the core (critical bytes), or in the other (non-critical) bytes of the line, and
- If the error is indicated in the L2 cache during the line fill into the L2 cache .

If the error was in the block requested for load/ifetch/prefetch (the 16-byte primary line for loads , 32-byte primary line for instruction fetches) and the error got detected before the line fill , the error is delivered to the requesting virtual processor with the data in the L2_Load_Return packet with the err field indicating correctable or uncorrectable error. The error type (DSC/DSU/DAC/DAU), rw, vcid, and moda fields are captured in the L2 Cache ESR, and the PA{39:6} is captured in the L2 Cache Error Address register. The error information (DSC/DSU/DAC/DAU/DBU/FBR/FBU) is captured in the DRAM ESR.

If the cache is filled before the data is returned to the requestor on a load/ifetch and error got delivered to the L2 cache from DRAM on that line fill, or if the error is delivered to the L2 cache from DRAM during the line fill of a TTE fetch, the error is reported to the virtual processor indicated by L2_CSR_REG.errorsteer, with the data in the Error_Indication_(L2) packet with the err field indicating correctable or uncorrectable error. The error type (DSC/DSU/DAC/DAU) field is captured in the L2 cache ESR, and the PA 39:6 is captured in the L2 Cache Error Address register. The error information (DSC/DSU/DAC/DAU/DBU/FBR/FBU) is captured in the DRAM ESR.

If the error is delivered to the L2 cache during the line fill, upto two errors could be presented: one to the strand indicated by the L2_CSR_REG.errorsteer on the line fill, and conditionally an L2 cache detected error reported on the err field of the data return packet to the requestor, when the L2 Cache is accessed for the load/ifetch/ TTE data after the line fill.

In case the line fill detected uncorrectable error from DRAM ,the subsequent L2 cache detected error can be only Notdata Error to the critical quarter-line since the Uncorrectable Error would already been reported on the line fill. In case the line fill detected correctable error from DRAM, there will not be a subsequent L2 cache detected error since the line fill would fill corrected data from DRAM.

The following errors are detected in the MCU and sent to the L2 cache, where they are returned to the requesting core/strand, using the L2_Load_Return packet/ L2_Ifetch_Return packet:

- DAC, DAU, DBU, FBR, and FBU errors related to a load/ifetch/prefetch access when the error (as indicated by the err field of the packet) is observed on critical data prior to the line fill.

The following errors are detected in the MCU and sent to the L2 cache, where they are returned to the core/strand indicated by the L2_CSR_REG.errorsteer, using the Error_Indication_(L2) packet:

- DSC and DSU errors
- DAC, DAU, DBU, FBR, and FBU errors related to a store,TTE fetch,atomic access
- DAC, DAU, DBU, FBR, and FBU errors related to a load/ifetch/prefetch access when the error (as indicated by the err field of the packet) is observed on the line fill .

The flow of DRAM errors through the L2cache is shown in FIGURE 16-2.

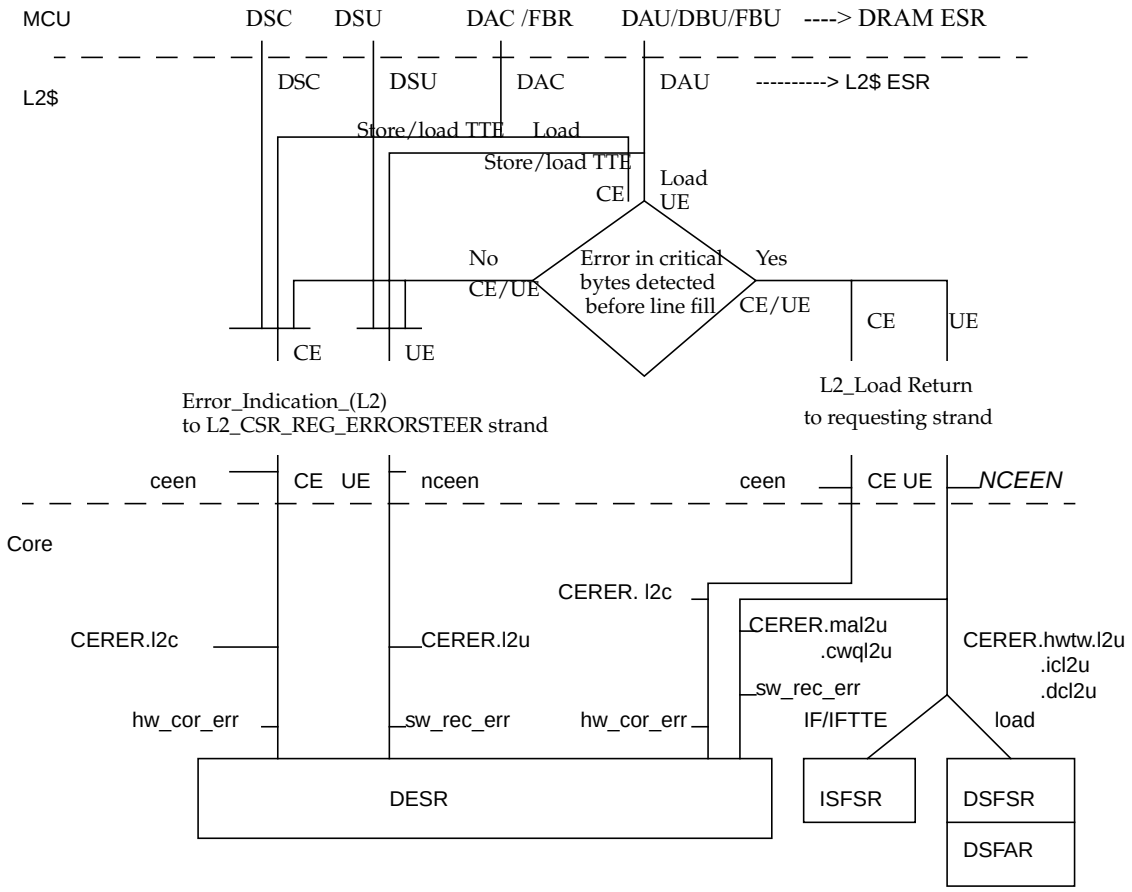


FIGURE 16-2 DRAM Error Flow From MCU to L2 Cache

Note | DRAM correctable errors (DAC, DSC, FBR) can result in two traps if error reporting is enabled both through the L2 and through the MCU Correctable/Recoverable Error Count register described in *MCU Correctable/Recoverable Count Errors* on page 334.

16.11.1 DRAM Correctable Error for Access (DAC)

The DRAM detects correctable data ECC errors and reports them to the L2 cache as DAC errors. These errors errors are indicated by setting the dac,vec bits and the synd field in the DRAM ESR

Note | If L2_CONTROL_REG.dis = 1 (the L2 cache is disabled), no logging or trap generation is performed for DRAM correctable ECC errors encountered on 32-bit and 64-bit stores.

For disrupting *sw_recoverable_error* trap conditions, if SETER.de is not set (or PSTATE.ie is clear and HPSTATE.hpriv is set), the error is still recorded in the DESR, but the trap remains pending until software sets SETER.de (and PSTATE.ie is set or HPSTATE.priv is clear).

For disrupting *hw_corrected_error* trap conditions, if SETER.dhcce is not set (or PSTATE.ie is clear and HPSTATE.hpriv is set), the error is still recorded in the DESR, but the trap remains pending until software sets SETER.dhcce (and PSTATE.ie is set or HPSTATE.priv is clear).

If either the L2 Cache Error Enable ceen or CERER.l2c_socc bit is reset, the error is ignored, but the error information is still captured in the DRAM Error Status register, the L2 Cache Error Status register, and the L2 Cache Error Address register.

These errors are logged and cause traps without regard to the ECC and FBR Error Count registers described in sections 25.12.4 and 25.12.9.

16.11.1.1 Load Miss / Instruction Fetch Miss / Prefetch Miss

The handling of the error in the L2 cache and the virtual processor depend on whether the error is reported on critical load data and whether the load occurs before the line fill.

Critical load data delivered before L2 cache line fill. When the load miss/ ifetch miss/prefetch miss request replay occurs before the line fill in the L2 cache and detects the correctable DRAM error on its critical bytes, the error information (dac,vec, rw, vcid, and moda) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

Hardware corrects the error before the data is placed into the L2 cache and returned to the virtual processor.

If the L2 Cache Error Enable ceen bit is set, an L2C error is signaled to the requesting virtual processor. The error information is delivered in the err field of an L2_Load_Return packet with the data. The requesting virtual processor indicates a *sw_recoverable_error* disrupting error condition. These errors are handled as follows:

- If the access was an Ifetch request, and CERER.icl2c is set and DESR.f is clear, hardware encodes ICL2C in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware takes a disrupting *sw_recoverable_error* trap.
- If the access was a load request, and CERER.dcl2c is set and DESR.f is clear, hardware encodes DCL2C in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware takes a disrupting *sw_recoverable_error* trap.
- In all cases, if SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Critical load data delivered after line fill. When the load miss/ifetch miss/ prefetch miss request replay occurs after the line fill in the L2 cache and a correctable DRAM error is presented to the L2 cache during the line fill, the error information (dac,vec and vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable ceen bit is set, an L2C error is signaled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of an Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.dhcce bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap to the requesting virtual processor.
- If SETER.dhcce is not set, or PSTATE.ie is not set, or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.11.1.2 Atomic Miss / TTE Miss

On an atomic request or a TTE miss request, the line fill into the L2 cache with corrected data occurs first. Then the critical data is accessed from the L2 cache.

When the correctable DRAM error is presented to the L2 cache on an atomic/TTE miss request, the error information (dac,vec and vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable ceen bit is set, an L2C error is signaled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.dhcce bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap to the requesting virtual processor.
- If SETER.dhcce is not set, or PSTATE.ie is not set, or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.11.1.3 Partial Store Miss

When the correctable DRAM error is presented to the L2 cache, the error information (dac,vec and vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable ceen bit is set, an L2C error is signalled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of an Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.DHCCE bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap to the requesting virtual processor.
- If SETER.dhcce is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.11.1.4 Store Miss

When the correctable DRAM error is presented to the L2 cache, the error information (dac,vec and vcid = L2_CSR_REG.errorsteer) is also captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register. Hardware corrects the error before the data is placed into the L2 cache (the new store data will also be correct).

If the L2 Cache Error Enable ceen bit is set, an L2C error is signaled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of an Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.dhcce bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap to the requesting virtual processor.
- If SETER.dhcce is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

If the L2 Cache is disabled (L2Control register dis bit is set), no logging occurs in the L2 Cache Error Status register and L2 Cache Error Address register, and no trap is presented.

16.11.1.5 DMA Read (DRC/DAC)

When a correctable error is presented to the L2 cache, the error information (drc,vec and rw) bits are captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register. Hardware corrects the error and returns it to the DMA requestor. In addition, if the L2 Cache Error Enable ceen bit is set, hardware generates an L2C error to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *sw_recoverable_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.
- If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

L2 Error status reports DRC (DMA correctable) for this error, while DRAM reports DAC.

16.11.1.6 DMA Write Partial (DRC/DAC)

When a correctable error is presented to the L2 cache, the error information (drc,vec and rw) bits are captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register. Hardware corrects the error in the L2 cache line. In

addition, if the L2 Cache Error Enable ceen bit is set, hardware generates an L2C error to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the SPARC CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.dhcce bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap to the virtual processor specified in L2_CSR_REG.errorsteer.
- If SETER.dhcce is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

L2 Error status reports DRC (DMA correctable) for this error, while memory reports DAC.

16.11.2 DRAM Correctable ECC Error for Scrub (DSC/FBR)

The DRAM controller detects two types of correctable errors that are reported to the L2 as DSC:

- Correctable ECC error found during a scrub, the error information (dsc, synd) is captured in the DRAM Error Status register, and PA{39:4} in the DRAM Error Address register.
- Recoverable FBD link error associated with DRAM accesses. These errors are indicated by the fbr bit set in the DRAM ESR. The error information is captured in the DRAM FDB Error Syndrome register.

When a correctable scrub error is presented to the L2 cache, DSC bit is set in the L2 Cache Error Status register. Hardware corrects the error in memory by writing back the corrected data and parity (this rewrite will be unable to correct a permanently failed bit). In addition, if the L2 Cache Error Enable ceen bit is set, hardware generates an L2C error to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *hw_corrected_error* disrupting error condition. These errors are handled as follows:

- If the CERER.l2c_socc bit is set and DESR.f is clear, hardware encodes l2c in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.dhcce bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *hw_corrected_error* trap.

- If SETER.dhcce is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.dhcce and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note that the dsc bit is set in the L2 Cache ESR regardless of the status of any other bits in the register. The dsc bit is logged to notify software to check the DRAM error register.

16.11.2.1 FBD Channel Recoverable Error (FBR)

There are four FBDIMM channel recoverable error conditions that the DRAM controller detects: CRC error on data, Alert Frames from the AMBs, Alert Asserted in Status Frames from the AMBs, or Status Frame Parity Error. When one of these conditions is detected, the DRAM controller will attempt to recover from the condition. If it successfully recovers, an FBR will be logged in the MCU ESR, the error condition will be logged in the MCU Error Syndrome Register, and a DSC will be reported to the L2.

16.11.3 DRAM Uncorrectable Error for Access (DAU/DBU/FBU)

The DRAM detects three types of uncorrectable errors and reports them to the L2 cache as DAU errors:

- Uncorrectable data ECC errors, indicated by setting the dau,veu bits and the synd field in the DRAM ESR.
- Out-of-bounds errors to nonexistent DRAM addresses (for details on out-of-bounds addresses, see *Access to Nonexistent Memory* on page 369). An out-of-bounds error is signaled as a cache line with all data marked with an uncorrectable error. For each 32-bit chunk, the data is loaded into the L2 cache with signaled ECC. These errors are indicated in the DRAM ESR with the dbu bit set. Note that writeback and block stores to out-of-bounds addresses are dropped with no error indication provided. Refer to TABLE 17-3 on page 370 for the complete list of OOB address ranges for different memory configurations.
- Unrecoverable FBD link error associated with DRAM accesses. These errors are indicated by the fbu bit set in the DRAM ESR. The error information is captured in the DRAM FDB Error Syndrome register.

16.11.3.1 Load Miss/Instruction Fetch Miss

The handling of the error in the L2 cache and the virtual processor depend on whether the error is reported on critical load data and whether the load occurs before the line fill. The line fill data is loaded into the L2 cache with signaled ECC.

Critical load data delivered before L2 cache line fill. When the load/ifetch miss request replay occurs before the line fill in the L2 cache and detects the uncorrectable DRAM error on its critical bytes(the 16-byte primary line for loads/ prefetches, 32-byte primary line for instruction fetches), the error information (dau,veu, rw, vcid, and moda) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

For each 32-bit chunk with an error, the data is loaded into the L2 cache with corrupted ECC. The L1 cache data is loaded with bad parity.

If the L2 Cache Error Enable nceen bit is set, an L2U error is signaled to the requesting virtual processor. The error information is delivered in the err field of a L2_Load_Return packet with the data. The requesting virtual processor indicates a precise error condition. These errors are handled as follows:

- If the access was an instruction fetch and the SPARC CERER.icl2u bit is set and SETER.pscce is set, hardware records icl2u in the ISFSR and takes a precise *instruction_access_error* trap. The VA of the instruction access is in TPC[TL]. If SETER.pscce is clear, hardware does not take a trap nor does it record any error in the ISFSR. Instead, it continues executing using the (corrupt) data from memory.
- If the access was a data fetch and the CERER.dcl2u bit is set and SETER.pscce is set, hardware records dcl2u in DSFSR, and takes a precise *data_access_error* trap. The VA of the data access is *not* logged in DSFSR. Software must examine the L2 ESRs to determine the failing physical address. If SETER.pscce is clear, hardware does not take a trap nor does it record any error in DSFSR or DSFAR. Instead, it continues executing using the corrupt data.

Critical load data delivered after L2 cache line fill. When the load/ifetch request replay occurs after the line fill in the L2 cache and an uncorrectable DRAM error is presented to the L2 cache during the line fill, the error information (dau,veu and vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable nceen bit is set, an L2U error is signaled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *sw_recoverable_error* disrupting error condition. These errors are handled as follows:

- If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes L2U in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.

- If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.11.3.2 Atomic Miss/TTE Miss

Since the atomic access or TTE miss service always occurs after the L2 cache line fill, the L2 cache presents the error after the line fill. The L2 cache also writes the data with signaled ECC. The error information (dau,veu and vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable nceen bit is set, an L2U error is signaled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *sw_recoverable_error* disrupting error condition. These errors are handled as follows:

- If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes l2u in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.
- If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

When the atomic accesses the L2 cache after the line fill , it will detect a NotData condition on the load and store accesses. See *Atomic Hit (DCL2ND)* on page 274.

For an atomic, if the error was in the word to be updated, the atomic operation will not complete its update, leaving the original data and the bad ECC unchanged.

When the TTE miss accesses the L2 cache after the line fill , it will detect a NotData condition . See *TTE Request for ITLB (ITL2ND)* on page 273 and *TTE Request for DTLB (DTL2ND)* on page 273

16.11.3.3 Prefetch Miss

When an uncorrectable error is presented to the L2 cache, the L2 cache presents the error after the line fill. The L2 cache also writes the data with signaled ECC. The error information (dau,veu, vcid = L2_CSR_REG.errorsteer) is captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

If the L2 Cache Error Enable nceen bit is set, an L2U error is signalled to the virtual processor specified in L2_CSR_REG.errorsteer. The error information is delivered in the err field of a Error_Indication_(L2) packet. That virtual processor indicates a *sw_recoverable_error* disrupting error condition. These errors are handled as follows:

- If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes l2u in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.
- If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Prefetch does not load data into the requesting core's data cache.

16.11.3.4 Store Miss

When an uncorrectable error is presented to the L2 cache, the error information (dau,veu, rw, vcid) is captured *i*n the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register. For each 32-bit chunk with an error, the data is loaded into the L2 cache with signaled ECC. For a partial store, the partial store will not complete its update, leaving the original data and the bad ECC unchanged. In addition, if the L2 Cache Error Enable nceen bit is set, a disrupting L2U error is generated to the virtual processor specified in L2_CSR_REG.errorsteer.

If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes L2U in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.

If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATDE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

16.11.3.5 DMA Read (DRU/DAU)

When the uncorrectable error is presented to the L2 cache, the error information (dru,veu, vcid = L2_CSR_REG.errorsteer) are captured *i*n the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register. Hardware returns the data back to the DMA requestor, with a UE error indicator for each 128-bit chunk with an error. In addition, if the L2 Cache Error Enable nceen bit is set, an L2U error is reported to the virtual processor specified in L2_CSR_REG.errorsteer.

If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes l2u in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.

If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

L2 Error status will report DRU (DMA uncorrectable) for this error, while the MCU DRAM will report DAU.

16.11.3.6 DMA Write Partial (DRU/DAU)

When the uncorrectable error is presented to the L2 cache, the error information (dru,veu and vcid = L2_CSR_REG.errorsteer) are captured in the L2 Cache Error Status register and PA{39:6} in the L2 Cache Error Address register.

These errors occur for DMA 8-byte writes where the line is read from DRAM into the L2 cache where the DMA write data is merged. For each 32-bit chunk with an error, the data is loaded into the L2 cache with signaled ECC. In addition, if the L2 Cache Error Enable nceen bit is set, an L2U error is reported to the virtual processor specified in L2_CSR_REG.errorsteer.

If the SPARC CERER.l2u_socu bit is set, and the DESR.f bit is clear, hardware encodes l2u in the DESR. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.

If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

L2 Error status will report DRU (DMA uncorrectable) for this error, while the MCU DRAM will report DAU.

16.11.3.7 FBD Channel Unrecoverable CRC Error (FBU)

When the MCU detects a CRC error on a read packet, it will retry the read transaction. If the error condition persists, the MCU will force a Fast Reset of the FBD channel and retry the read transaction. If the error persists after the Fast Reset, the error is logged as an FBU in the MCU ESR, a CRC error is logged in the MCU Error Syndrome Register, and the error is reported to the L2 as a DAU along with the corrupted data from the channel.

16.11.4 DRAM Uncorrectable ECC Error for Scrub (DSU/ FBU)

When an uncorrectable ECC error is found during a scrub, the information (dsu, synd) is captured in the DRAM Error Status register, with PA{39:4} captured in the DRAM Error Address register.

When the uncorrectable ECC error for scrub is presented to the L2 cache, the error information (dsu) is captured in the L2 Cache Error Status register. In addition, if the L2 Cache Error Enable nceen bit is set, an L2U error is reported to the virtual processor specified in L2_CSR_REG.errorsteer.

If the SPARC CERER.l2u_socu bit is set and the DESR.f bit is clear, hardware encodes l2u in the desr. The contents of the DESR.erroraddr field are undefined. In addition, if the SETER.de bit is set, and PSTATE.ie is set or HPSTATE.hpriv is clear, hardware generates a disrupting *sw_recoverable_error* trap.

If SETER.de is not set or PSTATE.ie is not set or HPSTATE.hpriv is set, hardware keeps the trap request pending until software either a) resets DESR.f, clearing the trap request, or b) sets SETER.de and either sets PSTATE.ie or clears HPSTATE.hpriv, causing hardware to take the trap.

Note that the dsu bit is set in the L2 Cache ESR regardless of the status of any other bits in the register. The dsc bit is logged to notify software to check the DRAM error register.

16.11.4.1 FBD Channel Unrecoverable Status Frame Parity and Alert Frame Errors

When the MCU detects a Status Frame Parity r an Alert Frame Error, it first issues a Soft Channel Reset on the FBD channel to try to clear the condition. If the error condition persists, it forces a Fast Reset of the FBD channel. If the error conditon persists after the Fast Reset, an FBU is logged in the MCU ESR, the error condition is logged in the MCU Error Syndrome Register, and the error is reported to the L2 as a DSU.

16.11.5 DRAM Software Error Scrubbing Support

Some errors will leave the DRAM with a correctable error, which then needs to be scrubbed to prevent repetitive traps for effectively the same soft error. DRAM scrubbing can be done through the L2 cache. To scrub, load the line into the L2, dirty it without modifying it (via a CAS that matches on the compare and has the same store data as the data returned on the load), then use a PrefetchICE to flush the line back to DRAM.

16.12 DRAM Error Registers

This section describes the error control and log registers for the DRAM. Each DRAM channel has its own set of error registers.

16.12.1 DRAM Error Status Register

This register contains status on DRAM errors. The status bits in this register are cleared by writing a 1 to the bit position.

Note | DRAM_ERROR_STATUS is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

TABLE 16-30 shows the format of the DRAM Error Status register.

TABLE 16-30 DRAM Error Status Register – DRAM_ERROR_STATUS_REG ($84\ 0000\ 0280_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63	meu	0	R/W1C	Multiple uncorrected errors, one or more uncorrected errors were not logged.
62	mec	0	R/W1C	Multiple corrected errors, one or more corrected errors were not logged.
61	dac	0	R/W1C	Set to 1 if the error was a DRAM access correctable error.
60	dau	0	R/W1C	Set to 1 if the error was a DRAM access uncorrectable error.
59	dsc	0	R/W1C	Set to 1 if the error was a DRAM scrub correctable error.
58	dsu	0	R/W1C	Set to 1 if the error was a DRAM scrub uncorrectable error.
57	dbu	0	R/W1C	Set to 1 if the error was an access to a nonexistent DRAM address (address out of bounds).
56	meb	0	R/W1C	Set to 1 if there were multiple out-of-bound errors
55	fbu	0	RW1C	Set to 1 if the error was a FBDIMM channel unrecoverable error
54	fbr	0	R/W1C	Set to 1 if the error was a FBDIMM channel recoverable error
53:16	—	0	RO	*Reserved*
15:0	synd	X	RW	ECC syndrome.

If both correctable and uncorrectable errors occur in the same cycle, the mec bit is set and only the appropriate bit for the uncorrectable error is set. With the exception of DBU, the syndrome is captured for the highest-priority error in that cycle (DSU, DAU, and FBU are priority 1, DSC, DAC, and FBR are priority 2). The syndrome remains unchanged if the error is a DBU. The address is loaded if the highest-priority error in the cycle is either a DSU or DSC.

If there are multiple errors, in different 16-byte chunks, in a single access, they are treated as multiple errors, since there is only logging state to describe a single 16-byte chunk error.

Note | MCU FBDIMM Unrecoverable error (FBU) can be injected through NCU; but error is not reported back to NCU, unlike FBR.

If errors occur in a cycle where an error status bit is already set, TABLE 16-31 applies.

TABLE 16-31 Errors When Error Status Bit Is Already Set

Existing Error	Error	Priority	Bit Set
DSU/DAU/FBU	DSU, DAU, or FBU	1	meu
DSU/DAU/FBU	DBU	1	dbu
DSU/DAU/FBU	DSC, DAC or FBR	2	mec
DBU	DSU	1	dsu
DBU	DAU	1	dau
DBU	FBU	1	fbu
DBU	DBU	1	meb
DBU	DSC	2	dsc
DBU	DAC	2	dac
DBU	FBR	2	fbr
DSC/DAC/FBR	DSU	1	dsu
DSC/DAC/FBR	DAU	1	dau
DSC/DAC/FBR	DBU	1	dbu
DSC/DAC/FBR	FBU	1	fbu
DSC/DAC/FBR	DSC, DAC, or FBR	2	mec

For the cases above where the “Bit Set” column contains a value besides meb, mec, and meu, the syndrome (for DSU/DAU/FBU/DSC/DAC/FBR) and address (for DSU/DSC) for the highest-priority error will overwrite the existing syndrome and address.

Once set, error status bits are only cleared by software. Hardware will never clear a set status bit. If a software write of the error register happens on the same cycle as an error, the setting of bits by the error will be based on the register state before the write. The setting of fields by the error will take precedence over the same field being updated by the write; however, fields that are not changed by the error will be

updated by the write (for example, if the register has the dac bit set and software does a write to clear that bit on the same cycle as a DRAM scrub uncorrectable error, the error register would end up with the dac bit cleared and the mec bit set, and the rw and synd fields would contain the values for the error).

16.12.2 DRAM Error Address Register

The DRAM Error Address register contains the physical address for the DRAM scrub error. DRAM load access address for errors are expected to be logged by L2 controller.

Note | DRAM_ERROR_ADDRESS is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

TABLE 16-32 shows the format of the DRAM Error Address register.

TABLE 16-32 DRAM Error Address Register – DRAM_ERROR_ADDRESS_REG ($84\ 0000\ 0288_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:40	—	0	RO	*Reserved*
39:4	address	X	RW	Error address.
3:0	—	0	RO	*Reserved*

This register is writable by software for register diagnostic reasons and isn't expected to be written during normal operation. However, in the event it is written on the same cycle that an error is reported, the update from the error will take precedence over the write.

16.12.3 DRAM Error Injection Register

Each DRAM channel has an Error Injection register for use in injecting DRAM errors to test error functionality or error handling code. The DRAM Error Injection register only provides for the injection of bad ECC on data written to memory. To inject an ECC error on data read from memory, bad ECC must be set up via a memory write, and then the memory locations with the bad ECC accessed with a read. Errors can be injected either single-shot or continuously. Once the enb_hp bit is set, either the first subsequent operation (for sshot = 1), or all subsequent operations (for sshot = 0) that cause a DRAM write will **xor** eccmask with the normally generated ECC when writing memory. When in single-shot mode, after the first injected error is generated, the sshot and enb_hp are automatically reset by the hardware to 0.

TABLE 16-33 shows the format of the DRAM Error Injection register.

TABLE 16-33 DRAM Error Injection Register – DRAM_ERROR_INJECT_REG ($84\ 0000\ 0290_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	enb_hp	0	RW	Enables error injection.
30	sshot	0	RW	Controls type of error injection. 1 = single shot; 0 = continuous.
29:16	—	0	RO	*Reserved*
15:0	eccmask	0	RW	ECC mask for error injection. The mask is **xor**ed with the computed ECC.

Note | DRAM_ERROR_INJECT_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.12.4 DRAM Error Counter Register

Each DRAM channel has an Error Counter register for use in counting DRAM correctable ECC errors and generating a trap when the counter decrements to 0. Each 16-byte chunk with an error will cause the count field to decrement by one.

When count transitions from 1 to 0, the corresponding mcu{0-3} ECC bit is set in the SOC_ERROR_STATUS_REG described in *SOC Error Status Register* on page 336, which if enabled will generate a *sw_recoverable_error* trap to the strand in the SOC_ERROR_STEER_REG.

TABLE 16-34 shows the format of the DRAM Error Counter register.

TABLE 16-34 DRAM Error Counter Register – DRAM_ERROR_COUNTER_REG ($84\ 0000\ 0298_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:16	—	X	RO	*Reserved*
15:0	COUNT	0	RW	Counter that decrements with each ECC correctable error unless already 0.

Note | DRAM_ERROR_COUNTER_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.12.5 DRAM Error Location Register

Each DRAM channel has an Error Location register for software to sample to locate a bad memory part. The register is set on a correctable DRAM error if a correctable error is not already logged in the MCU ESR. The bit set corresponds to the nibble that was corrected. Bits 35:32 correspond to the ECC nibbles and 31:0 to data nibbles. When in dual-channel mode, 35:18 refer to nibbles in channel 0 and 17:0 refer to nibbles in channel 1.

TABLE 16-35 shows the format of the DRAM Error Location register.

TABLE 16-35 DRAM Error Location Register – DRAM_ERROR_LOCATION_REG (84 0000 $02A0_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:36	—	0	RO	*Reserved*
35:0	location	X	RW	DRAM ECC error location.

Note DRAM_ERROR_LOCATION_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.12.6 DRAM Error Retry Register

Each DRAM channel has an Error Retry register for software to sample to locate a bad memory part. The register is set on each correctable DRAM error.

TABLE 16-36 shows the format of the DRAM Error Retry register.

TABLE 16-36 DRAM Error Retry Register – DRAM_ERROR_RETRY_REG (84 0000 $02A8_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:37	—	0	RW	*Reserved*
36	valid	0	RW	Error Retry register is valid.
35:20	synd2	0	RO	Syndrome from second retry read.
19:18	type2	0	RW	Result of second retry read.
17:2	synd1	0	RO	Syndrome from first retry read
1:0	type1	0	RW	Result of first retry read. 00 – No read. 01 – No error. 10 – Correctable error. 11 – Uncorrectable error.

Note | DRAM_ERROR_RETRY_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.12.7 DRAM FBD Error Syndrome Register

Each DRAM channel has an FBD Error Syndrome for FBD link errors. When an FBD link error is detected, the syndrome is captured in this register. For a recoverable link error, if the count field of the DRAM_FBR_COUNT_REG register is zero, the corresponding mcu{0-3} FBR{} bit is set in the SOC_ERROR_STATUS_REG register as described in *SOC Error Status Register* on page 336. If enabled, setting of this bit generates a *hw_corrected_error* trap to the strand in the SOC_ERROR_STEER_REG register. Once the valid bit is set, no further FBD link errors are logged, so software will need to clear the valid bit to enable further FBD link error detection.

TABLE 16-37 shows the format of the DRAM FBD Error Syndrome register.

TABLE 16-37 DRAM FBD Error Syndrome Register – DRAM_FBD_ERROR_SYND_REG (84 0000 $0C00_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63	valid	0	RW	Valid.
62:30	—	0	RO	*Reserved*
29:18	alert1	0	RW	AMB alert bits set in Channel 1.
17:6	alert0	0	RW	AMB alert bits set in Channel 0.
5	softreset	0	RW	MCU issued a soft channel reset command to the channel.
4	fastreset	0	RW	MCU performed a fast reset of a channel due to a soft channel reset for the error not being effective.
3	sfpe	0	RW	Status frame parity error.
2	aa	0	RW	Alert asserted.
1	afe	0	RW	Alert frame error.
0	c	0	RW	CRC error.

Note | DRAM_FBD_ERROR_SYND_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.12.8 DRAM FBD Injected Error Source Register

When the NCU signals the MCU to inject an error, this register determines into which error detection logic the error will be injected.

TABLE 16-38 shows the format of the DRAM FBD Injected Error Source register.

TABLE 16-38 DRAM FBD Error Syndrome Register – DRAM_FBD_INJ_ERROR_SRC_REG ($84\ 0000\ 0c08_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
62:2	—	0	RO	*Reserved*
1:0	errorsource	0	RW	Source of the error. 0 – CRC error. 1 – Alert frame error. 2 – Alert asserted. 3 – Status frame parity error.

Note | DRAM_FBD_INJ_ERROR_SRC_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

Note | Continuous injection of Alert Frame errors may cause a system to hang. The memory controller will be continually processing the Alert Frame errors and will not have time to service DRAM read requests.

Note | When injecting Alert Asserted errors, the memory controller will only report one AA error to the L2 until the MCU Error Syndrome register is cleared.

16.12.9 DRAM FBR Count Register

This register controls the sending of FBD recoverable error interrupts to the NCU. Each DRAM channel has an FBR Count register for use in counting FBD recoverable channel errors and generating a trap when the counter decrements to 0. Each FBR will cause the count field to decrement by one.

When count transitions from 1 to 0, the corresponding mcu{0-3} FBR bit is set in the SOC_ERROR_STATUS_REG described in *SOC Error Status Register* on page 336, which if enabled will generate a *hw_corrected_error* trap to the strand in the SOC_ERROR_STEER_REG.

TABLE 16-39 shows the format of the DRAM FBR Count register.

TABLE 16-39 DRAM FBR Count Register – DRAM_FBR_COUNT_REG (84 0000 $0C10_{16}$) (Count 4 Step 4096)

Bit	Field	POR Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	countone	0	RW	If set, the MCU will generate an interrupt to the NCU on every FBR error.
15:0	count	0	RW	Count of FBR errors, decremented on every FBR error if countone is not set. When this value decrements from 1 to 0, an FBR interrupt will be sent to the NCU.

Note | DRAM_FBR_COUNT_REG is not cleared on warm reset to allow software to examine the error information that may have required the warm reset.

16.13 Block Loads and Stores

UltraSPARC T2 supports 64-byte block load and store access to ASI_BLK_P, ASI_BLK_S, ASI_BLK_COMMIT_P, ASI_BLK_COMMIT_S, ASI_BLK_PL, ASI_BLK_SL, ASI_BLK_AIUP, ASI_BLK_AIUS, ASI_BLK_AIUPL, ASI_BLK_AIUSL.

Loads for these operations consist of four 16-byte "helper" loads, while stores are composed of eight 8-byte "helper" stores. Store buffer errors encountered on any of the helper stores will be logged the same as if it was a non-helper store and will not affect the issuing of subsequent helper stores. For block stores, if there is an error on any of the helper stores, the entire block store will be terminated and no memory updates performed.

For the helper loads, errors (including L2C) will be accumulated on all four helpers and reported (and trapped if enabled) after the final helper completes. The error reported will be the highest-priority error (ND $\rightarrow$ UE $\rightarrow$ CE) if one or more errors are detected. If multiple errors of the highest priority are detected, the error reported will be the one associated with the earliest helper.

The block load is considered a single operation, and even if multiple uncorrectable or correctable errors (or a combination of the two) are encountered, they are treated as a single error with respect to updating the D-SFSR. The floating-point register file will be updated with load results up to the point of the earliest helper that encountered an uncorrectable error. Loading of results into the floating-point register file will be suppressed for the helper that encountered the uncorrectable error and any subsequent helpers.

Programming Note

When a block load encounters an error, the destination registers of the helper operations issued before the helper encountering the error will be updated. This implies that the exception taken for the error will not be strictly precise. However, the TPC and TNPC will still be updated as if the exception were fully precise and TPC will point to the block load instruction that encountered the error.

16.14 CMP Error Summary

TABLE 16-40 summarizes CMP error handling. Column heads and error-type abbreviations are described in TABLE 16-41 on page 314.

TABLE 16-40 SPARC, L2 and DRAM Error Handling Summary *(1 of 5)*

Error Type	Error	L2 ESR ISFSR/DSFSR/ DESR/DFESR	(NotData ESR = ND)	DRAM Status	PA	Syn	Trap	Trap Type	Section
ITLB tag parity	CE	ITTP				is	p	IM	16.7.1.2
IT tag multiple hit	CE	ITTM				is	p	IM	16.7.1.1
ITLB Data Parity	CE	ITDP				is	p	IM	16.7.1.3
ITLB MRA uncorrectable	UE	ITMU				is	p	IM	16.7.1.4
ITLB L2 correctable	CE	ITL2C	LDAC		L	I	di	E	16.9.1.1
ITLB L2 uncorrectable	UE	ITL2U	LDAU		L	I	p	IM	16.9.6.1
ITLB L2 NotData	ND	ITL2ND	ND.NDSP		N		p	IM	16.9.13.1
Icache valid bit	CE	ICVP				de	di	C	16.7.3.1
Icache tag parity	CE	ICTP				de	di	C	16.7.3.2
Icache tag multiple hit	CE	ICTM				de	di	C	16.7.3.3
Icache data parity	CE	ICDP				de	di	C	16.7.3.4
Icache L2 correctable	CE	ICL2C	LDAC		L	I	di	E	16.9.1.3
Icache L2 uncorrectable	UE	ICL2U	LDAU		L	I	p	I	16.9.6.3
Icache L2 NotData	ND	ICL2ND	ND.NDSP		N		p	I	16.9.14.2
IRF correctable ECC error	CE	IRFC				ds	p	P	16.7.5
IRF uncorrectable ECC error	UE	IRFU				ds	p	P	16.7.5
FRF correctable ECC error	CE	FRFC				ds	p	P	16.7.6
FRF uncorrectable ECC error	UE	FRFU				ds	p	P	16.7.6
DTLB tag parity	CE	DTTP				ds	p	DM	16.7.2.2
DT tag multiple hit	CE	DTTM				ds	p	DM	16.7.2.1
DTLB data parity	CE	DTDP				ds	p	DM	16.7.2.3
DTLB MRA uncorrectable	UE	DTMU				ds	p	DM	16.7.2.4
DTLB L2 correctable	CE	DTL2C	LDAC		L	I	di	E	16.9.1.2
DTLB L2 uncorrectable	UE	DTL2U	LDAU		L	I	p		16.9.6.2
DTLB L2 NotData	ND	DTL2ND	ND.NDSP		N		p	DM	16.9.13.2
Dcache valid bit	CE	DCVP				de	di	C	16.7.4.1
Dcache tag parity	CE	DCTP				de	di	C	16.7.4.2
Dcache tag multiple hit	CE	DCTM				de	di	C	16.7.4.3
Dcache Data Parity	CE	DCDP				de	di	C	16.7.4.4

TABLE 16-40 SPARC, L2 and DRAM Error Handling Summary *(2 of 5)*

Error Type	Error	ISFSR/DSFSR/ DESR/DFESR	L2 ESR (NotData ND)	DRAM Status	PA	Syn	Trap	Trap Type	Section
Dcache L2 Correctable	CE	DCL2C	LDAC		L	l	di	E	16.9.1.4
Dcache L2 Uncorrectable	UE	DCL2U	LDAU		L	l	p	D	16.9.6.4 and 16.9.6.5
Dcache L2 NotData	ND	DCL2ND	ND.NDSP		N		p	D	16.9.13.4 and 16.9.13.5
Store Buffer Data Load hit Correctable ECC	CE	SBDLC				ds	p	P	16.7.7.1
Store Buffer Data Load hit Uncorrectable ECC	UE	SBDLU				ds	p	P	16.7.7.2
Store Buffer Data PCX read or ASI store Correctable	CE	SBDPC				de	di	C	16.7.7.4
Store Buffer Data PCX read or ASI store Uncorrectable	UE	SBDPU				de	di	E	16.7.7.5
Store Buffer Address PCX read or ASI ring store read Uncorrectable	DE	SBDIOU				df	df	S	16.7.7.6
Store Buffer Address PCX read or ASI ring store read Address Bit Parity	DE	SBAPP				df	df	S	16.7.7.7
TSA Correctable	CE	TSAC				ds	p	P	16.7.10
TSA Uncorrectable	UE	TSAU				ds	p	P	16.7.10
MRA Uncorrectable	UE	MRAU				ds	p	P	16.7.11
SCA Correctable	CE	SCAC				ds	p	P	16.7.8
SCA Uncorrectable	UE	SCAU				ds	p	P	16.7.8
Tick_Compare Correctable Precise	CE	TCCP				ds	p	P	16.7.9
Tick_Compare Correctable Disrupting	CE	TCCD				de	di	E	16.7.9
Tick_Compare Uncorrectable Precise	UE	TCUP				ds	p	P	16.7.9
Tick_Compare Uncorrectable Disrupting	UE	TCUD				de	di	E	16.7.9
IO error from Load from noncacheable address	UE	SOCU			S	s	p	D	
IO error from Ifetch from noncacheable address	UE	ICL2U			S	s	p	I	

TABLE 16-40 SPARC, L2 and DRAM Error Handling Summary *(3 of 5)*

Error Type	Error	ISFSR/DSFSR/DESR/DFESR	L2 ESR (NotData ESR = NDND)	DRAM Status	PA	Syn	Trap	Trap Type	Section
L2$ data ecc: LD_h, If_h, ATOM_h, PF_h	CE		LDAC		L	I	di	E	
L2$ data ecc: LD_h, ATOM_h	UE	LDAU	LDAU		LS	I	p	D	
L2$ data ecc: PF_h	UE		LDAU		L	I	di	E	
L2$ data ecc: If_h	UE	LDAU	LDAU		LS	I	p	I	
L2$ data NotData: LD_h, ATOM_h	ND	DCL2ND	ND.NDSP		N		p	D	
L2$ data NotData: PF_h	ND		ND.NDSP		N		di	E	
L2$ data NotData: If_h	ND	ICL2ND	ND.NDSP		N		p	I	
L2$ data ecc: PST_h	CE		LDAC		L	I	di	C	
L2$ data ecc: PST_h	UE		LDAU		L	I	di	E	
L2$ data ecc: PST_h	ND	L2ND	ND.NDSP		N	de	di	E	
L2$ data ecc: wb	CE		LDWC		L	I	di (ES)	C	
L2$ data ecc: wb	UE		LDWU		L	I	di (ES)	E	
L2$ data ecc:dma_read	CE		LDRC		L	I	di (ES)	E	
L2$ data ecc:dma_read	UE		LDRU		L	I	di (ES)	E	
L2$ data ecc:dma_read	ND	L2ND	ND.NDDM		N	de	di (ES)	E	
L2$ data ecc:dma_write_partial	CE		LDRC		L	I	di (ES)	C	
L2$ data ecc:dma_write_partial	UE		LDRU		L	I	di (ES)	E	
L2$ data ecc:dma_write_partial	ND	L2ND	ND.NDDM		N	de	di (ES)	E	
L2$ data ecc: scrub	CE		LDSC		L	I	di (ES)	C	
L2$ data ecc: scrub	UE		LDSU		L	I	di (ES)	E	
L2$ tag ecc: all refs	CE		LTC		L	—	di (ES)	C	
L2 $ dir parity: scrub	FE		LRF		L	I	F	R	
L2 $ vuad ecc: all refs	FE		LVF		L	I	F	R	
L2 $ vuad ecc: all refs	CE		LVC		L	I	di (ES)	C	
Dram ECC error: LD_m_cc, If_m_c32B, PF_m_cc	CE		DAC	DAC	L	D	di	E	
FBDIMM recoverable error: LD_m_cc, If_m_c32B, PF_m_cc	CE		DAC	FBR	L	F	di	E	
Dram ECC error: LD_m_cc, ATOM_m_cc	UE	LDAU	DAU	DAU	LS	D	p	D	
Dram ECC error: PF_m_cc	UE		DAU	DAU	L	D	di	E	

TABLE 16-40 SPARC, L2 and DRAM Error Handling Summary *(4 of 5)*

Error Type	Error	ISFSR/DSFSR/ DESR/DFESR	L2 ESR (NotData ESR = ND)	DRAM Status	PA	Syn	Trap	Trap Type	Section
Dram ECC error: If_m_c32B	UE	LDAU	DAU	DAU	LS	D	p	I	
FBDIMM unrecoverable error: LD_m_cc, ATOM_m_cc	UE	LDAU	DAU	FBU	LS	F	p	D	
FBDIMM unrecoverable error: PF_m_cc	UE		DAU	FBU	L	F	di	E	
FBDIMM unrecoverable error: If_m_c32B	UE	LDAU	DAU	FBU	LS	F	p	I	
Dram address out of bounds: LD_m, ATOM_m	UE	LDAU	DAU	DBU	LS	D	p	D	
Dram address out of bounds: PF_m	UE		DAU	DBU	L	D	di	E	
Dram address out of bounds: If_m	UE	LDAU	DAU	DBU	LS	D	p	I	
Dram ECC error: ST_m, PST_m, LD_m_ncc, If_m_nc32B, ATOM_m, PF_m_ncc	CE		DAC	DAC	L	D	di (ES)	C	
Dram ECC error: dma_read_req, dma_write_partial	CE		DRC	DAC	L	D	di (ES)	C	
FBDIMM recoverable error: Status Frame Parity Error, Alert Frames, AMB Alert Asserted, CRC	CE		DSC	FBR	L	F	di (ES)	C	
Dram ECC error: ST_m,PST_m, LD_m_ncc, If_m_nc32B, ATOM_m, PF_m_ncc	UE		DAU	DAU	L	D	di (ES)	E	
Dram ECC error: dma_read_req, dma_write_partial	UE		DRU	DAU	L	D	di (ES)	E	
FBDIMM unrecoverable error: CRC	UE		DAU	FBU	L	F	di (ES)	E	
FBDIMM unrecoverable error: Status Frame Parity Error, Alert Frame	UE		DSU	FBU	L	F	di (ES)	E	
Dram address out of bounds: ST_m, PST_m	UE		DAU	DBU	L	D	di (ES)	E	

TABLE 16-40 SPARC, L2 and DRAM Error Handling Summary *(5 of 5)*

Error Type	Error	ISFSR/DSFSR/ DESR/DFESR	L2 ESR (NotData ND)	DRAM ESR = Status	PA	Syn	Trap	Trap Type	Section
Dram address out of bounds: dma_read_req, dma_write_partial	UE		DRU (DMA read) DAU (DMA write)	DBU	L	D	di (ES)	E	
Dram Scrub error	CE		DSC	DSC	D	D	di (ES)	C	
Dram Scrub error	UE		DSU	DSU	D	D	di (ES)	E	

TABLE 16-41 describes the column heads and error-type abbreviations (in alphabetical order) of TABLE 16-40.

TABLE 16-41 Description of Column Heads and Error-Type Abbreviations in TABLE 16-40

Column or Error-Type Abbr	Meaning
Error	FE – fatal error; UE – uncorrected error; CE – corrected error, ND - NotData error, DF – deferred error
PA logging	S – D-SFAR; N – L2 NotData error; L – L2 error address; D – DRAM error address
SYN logging	is – I-SFSR; ds – D-SFSR; de – DESR; df – DFESR; l – L2 error status; D – DRAM error status; F – DRAM FBD error syndrome
Trap	p – precise to requestor; di – disrupting to requestor; di (ES) – disrupting to the virtual processor specified in L2_CSR_REG.errorsteer; f - fatal warm reset to all virtual processors
Trap Type	I – *instruction_access_error*; IM – *instruction_access_MMU_error*; D – *data_access_error,* DM – *data_access_MMU_error*; P - *internal_processor_error,*; E - *sw_recoverable_error,* C – *hw_corrected_error*; S – *store_error*; R – *warm_reset*
ATOM_h	Atomic operation hit
ATOM_m_cc	Atomic operation miss critical 16-byte chunk
ATOM_m_ncc	Atomic operation miss noncritical chunk
dma_read	DMA read any size
dma_write_partial	subline DMA write
If_h	I-fetch hit
If_m_c32B	Ifetch miss critical 32-byte chunk
If_m_nc32B	Ifetch miss noncricital 32-byte chunk
LD_h	Load hit
LD_m_cc	Load miss critical 16-byte chunk
LD_m_ncc	Load miss noncritical chunk
PF_h	Prefetch hit
PF_m_cc	Prefetch miss critical 16-byte chunk
PF_m_ncc	Prefetch miss noncritical chunk

TABLE 16-41 Description of Column Heads and Error-Type Abbreviations in TABLE 16-40 *(Continued)*

Column or Error-Type Abbr	Meaning
PST_h	Partial Store hit
PST_m	Partial Store miss
wb	writeback to memory

16.15 Boot ROM Interface (SSI)

TABLE 16-42 describes the SSI's handling of errors. The error indication on read returns is delivered regardless of the SSI_TIMEOUT.erren bit, where it is up to the virtual processor to ignore the error or receive it based on whether or not CERER.icl2u and CERER.dcl2u bit are set. Logging the error and sending an error trap request are controlled by the erren bit. Note that returning zeros on an I-fetch timeout will tend to cause an *illegal_instruction* trap.

TABLE 16-42 SSI Error Handling

Error	TType	Severity	Core Error Info	Trap Taken	Error Returned on Transaction	NCU Error Info (if SSI_TIMEOUT.erren =1)
SSI parity error	Read	Uncorrectable	I-SFSR = icl2u or D-SFSR = dcl2u	*instruction_access_error* or *data_access_error*	Yes (with data)	SSI_LOG. parity
SSI parity error	Write	Uncorrectable	Not Recorded	None	No	SSI_LOG. parity
SSI timeout	Read	Uncorrectable	I-SFSR = icl2u or D-SFSR = dcl2u	*instruction_access_error* or *data_access_error*	Yes (with data = 0)	SSI_LOG. tout
SSI timeout	Write	Uncorrectable	Not Recorded	None	No	SSI_LOG. tout

Note | I-fetch to any other I/O space other than SSI boot ROM space would result in error packet being returned to the requesting SPC by NCU.

16.15.1 SSI Parity Error

SSI has serial parity on all requests and responses. Odd parity on any response will be treated as a parity error.

On reads, the SSI block will return the data, but marked with an error indication, which will tend to cause an NCU error at the requesting SPARC. For both reads and writes, the SSI block will issue an error interrupt via int_man{1} (if SSI_TIMEOUT.erren is set).

16.15.2 SSI Timeout

SSI only supports a single transaction outstanding at any time, and write transactions receive a positive acknowledgement to inform UltraSPARC T2 of their completion. Whenever UltraSPARC T2 issues a read or write transaction, it starts a timer to the value specified in SSI_TIMEOUT[TIMEVAL], which then decrements by 1 every SSI cycle. If the time underflows before the transaction completes, it is treated as a timeout.

On reads, the SSI block will return zeros, but marked with an error indication, which will tend to cause an NCU error at the requesting SPARC. For both reads and writes, the SSI block will issue an error interrupt via int_man{1} (if SSI_TIMEOUT.*ERREN* is set).

16.15.3 SSI Error Registers

The serial bus interface to the Boot ROM is called SSI, hence the registers dealing with errors on this interface are SSI registers.

TABLE 16-43 and TABLE 16-44 define the format of the SSI Timeout and Log registers, respectively.

TABLE 16-43 SSI Timeout Register– SSI_TIMEOUT (FF 0001 0088_{16})

Bit	Field	Initial Value	R/W	Description
63:25	—	X	RO	*Reserved*
24	erren	0	RW	Enables error logging and error interrupt generation in the SSI.
23:0	timeval	200000_{16}	RW	Number of SSI cycles before an unacknowledged request causes a timeout error.

The default value for timeout is about 40 msec.

TABLE 16-44 217 SSI Log Register – SSI_LOG (FF 0000 0018_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	parity	0	RW1C	Parity error detected on response.
0	tout	0	RW1C	No response before TIMEVAL.

16.16 Error Injection Summary

Most of the large arrays on UltraSPARC T2 have some error protection, and also the capability of getting an error injected. TABLE 16-45 specifies the programmatic interface for injection, plus some notes of the type of injection.

TABLE 16-45 Error Injection Summary

Error	Control	Notes
ITLB data parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
ITLB CAM parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
DTLB data parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
DTLB CAM parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
Icache data parity	ASI_ICACHE_INSTR	Optionally flip parity on ASI write.
Icache tag parity	ASI_ICACHE_TAG	Optionally flip parity on ASI write.
Dcache data parity	ASI_DCACHE_DATA	Flip parity bits under mask on ASI write.
Dcache tag parity	ASI_DCACHE_TAG	Optionally flip parity on ASI write.
Int RegFile ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
FP RegFile ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
Scratchpad array ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
Tick_compare Array ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
TSA ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
MRA parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
Store buffer CAM parity	ASI_ERROR_INJECT_REG	Continuous parity flip on update.
Store buffer data ECC	ASI_ERROR_INJECT_REG	Continuous ECC XOR on update.
L2 data ECC	L2_DIAG_DATA	Write 4B data and computed ECC.
L2 tag ECC	L2_DIAG_TAG	Write tag and computed ECC.
L2 directory parity	L2_ERROR_INJECT_REG	Single/double or continuous parity flip on update.
L2 UA ECC	L2_DIAG_UA	Write UA bits and computed ECC.
L2 VD ECC	L2_DIAG_VD	Write VD bits and computed ECC.
DRAM ECC	DRAM_ERROR_INJECT_REG	Continuous syndrome XOR on write.
SOC errors	SOC_ERROR_INJECT_REG	Continuous parity/ECC corrupt.

16.17 SOC Error Descriptions

An error in the system-on-the-chip (SOC) can occur on the major types of SOC operations. The SOC detects correctable and uncorrectable errors; it does not detect NotData errors. Any error in the SOC can be made fatal under the control of the SOC Fatal Error Enable register, see *SOC Fatal Error Enable Register* on page 346 for details.

The SOC operation types are as follows:

- Error on load to I/O space (PIO load)
- Error on store to I/O space (PIO store)
- Errors on an interrupt request
- Errors on DMA reads and writes
- Error interrupts from the MCU due to an error count registers

These errors are discussed in the following sections.

16.18 PIO Load Errors

Three classes of PIO load errors are detected in the SOC: fatal errors (FE), which cause a warm reset; uncorrectable errors (UE), which generate a *data_access_error* trap to the requesting strand or *sw_recoverable_error* to the strand specified in the SOC Error Steering Register ; or correctable errors (CE), which generate a *hw_corrected_error* trap to the requesting strand .

See TABLE 16-46 for details.

TABLE 16-46 SOC Errors for a PIO Load Request

Error	SOCESR bit	Recommended Error Type	Error Info (Core)	Error Info (NCU)	Trap
NCU uncorrectable on command or thread_id from PCX	ncupcxue{19}	FE	DESR = SOCU if SOC EIE bit (SOC Error Interrupt Enable bit) set	NCUSYN (etag, rqtyp, cpu_id, thread_id, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII command parity error or Ctag UE	siidmuue{1}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU Ctag UE	ncuctague{22}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (ctag)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU data parity	ncudataparity{14}	FE	D-SFSR = SOCU, DESR = SOCU if SOC EIE bit set	NCUSYN (ctag)	Warm reset if configured as fatal; otherwise, *data_access_error*. and *sw_recoverable_error* if SOC EIE bit set

TABLE 16-46 SOC Errors for a PIO Load Request

Error	SOCESR bit	Recommended Error Type	Error Info (Core)	Error Info (NCU)	Trap
NCU Mondo Table Parity	ncumondotable{15}	UE	D-SFSR = SOCU, DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *data_access_error* , and *sw_recoverable_error* if SOC EIE bit set
NCU DMU data parity	ncudmuue{21}	FE	DESR = SOCU, DESR = SOCU IF SOC EIE BIT SET	NCUSYN (etag, reqtype, cpuid, strandid, pa)	Warm reset if configured as fatal; otherwise *data_access_error* , and, *sw_recoverable_error* if SOC EIE bit set
NCU FIFO output error	ncucpxue{20}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise *sw_recoverable_error* if SOC EIE bit set
NCU timeout, unmapped error or uncorrected error	—	UE	D-SFSR = SOCU	Not recorded	Warm reset if configured as fatal; otherwise, *data_access_error*.
NCU PCX Data	ncupcxdata{18}	UE	DESR = SOCU if SOC EIE bit set	NcuSyn (Etag, Rqtyp, Cpu_id, Thread_id, and PA)	Warm reset if configured as fatal; otherwise sw_recoverable_error if SOC EIE bit set
SII Ctag CE	siidmuce{3}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise *hw_corrected_error* if SOC EIE bit set.
NCU Ctag CE	ncuctagce{23}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set.

16.18.1 Uncorrectable PIO Load Errors Recommended as Fatal

The following uncorrectable PIO Load errors are recommended to be set as fatal in the SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. Fatal errors associated with a PIO load cause the PIO load to be dropped and a warm reset to be initiated.NCUPCXDATANCUPCXDATA

16.18.1.1 Uncorrectable NCU FIFO Errors (NcuPcxUE)

The NCU provides DEDSEC protection on the NCU FIFO. If an uncorrectable error is detected on the command or thread ID fields, bit 19, ncupcxue, in the SOC ESR is set. The error information (etag, rqtyp, cpu_id, thread_id, and PA{39:0}) is recorded in the NCUSYN register.

No response is returned, so the requesting strand will become hung. Therefore, this error needs to be fatal.

16.18.1.2 Uncorrectable SII Ctag/Command Parity Errors (SiiDmuCtagUE)

For PIO load return requests (through the DMU) to the SII, the Ctag ECC is checked and also the command parity. If an uncorrectable Ctag ECC error or command parity error is detected, bit 1, siidmuctague, in the SOC ESR is set. The error information (etag, ctag, and PA{39:0}) is recorded in the NCUSIISYN register.

No response is returned, so the requesting strand will become hung. Therefore, this error needs to be fatal.

16.18.1.3 Uncorrectable NCU Ctag Error (NcuCtagUe)

The NCU also checks Ctag ECC from the SII. If an uncorrectable Ctag ECC error is detected, bit 22, ncuctague, in the SOC ESR is set. The error information (Ctag) is recorded in the NCUSYN register. A return packet is never generated to the requesting strand and that strand will become hung. Therefore, this error needs to be fatal.

16.18.1.4 NCU Data Parity Error (NcuDataParity)

The NCU checks data parity for returning PIO load requests from both the NIU and DMU (via the SII). If a data parity error is detected, bit 14, ncudataparity, in the SOC ESR is set. The error information (Ctag) is recorded in the NCUSYN register. For a PIO load, this error does not need to be fatal, because an error packet is returned to the requesting strand. However, since the ncudataparity bit in the SOCESR is shared with interrupts from the SII, which do need to be fatal, the fatal error will also occur for interrupts when the ncuctague bit is set.

16.18.1.5 NCU FIFO Output Error (NcuCpxUe)

The NCU checks the output FIFO to the CPX. If the NCU detects an error in the output FIFO, bit 20, ncucpxue, in the SOC ESR is set.No other error information is captured.

No response is returned to the requesting strand, which will become hung. Therefore, this error needs to be fatal.

16.18.1.6 NCU Parity Error from NCU DMU PIO Req FIFO (NcuDmuUe)

If NCU detects a parity error from the NCU DMU PIO Req FIFO on a PIO load, bit 21, ncudmuue, in the SOC ESR is set. The error information (etag, rqtyp, cpu_id, thread_id and PA{39:0}) is recorded in the NCUSYN register. NCU returns data on CPX packet marked with UE. NCU squashes the PIO load request to DMU.

However since the parity error can be in cpu_id[2:0] bits , the load return might happen to the wrong CPU. Hence this error needs to be fatal.

16.18.2 Uncorrectable PIO Load Errors

The following uncorrectable PIO load errors are recommended to *not* be set as fatal in the SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For uncorrectable SOC detected errors associated with the PIO load, the NCU Load Return packet is sent to the requesting virtual processor with the err field indicating uncorrectable error.

16.18.2.1 NCU Mondo Table Error (NcuMondoTable)

A PIO load can access the Mondo Table in the NCU. If the NCU detects a data parity error, bit 15, ncumondotable, in the SOC ESR is set. No other error is information captured.

16.18.2.2 NCU PCX FIFO Data Parity Error (NcuPCXData)

A PIO load request to NCU that has parity error in the data (payload) logs ncupcxdata bit 18 in the SOC ESR, eventhough the data is ignored by NCU. The error information (etag, rqtyp, cpu_id, thread_id and PA{39:0}) is recorded in the NCUSYN register. NCU returns data on CPX packet without any error bit set.

16.18.2.3 Other Uncorrectable NCU Errors

Errors may have been detected for the PIO load request. These errors are reported as bits in the packet from the SII. The NCU checks the error bits 29:31 in the SII to NCU packet. If bit 31, timeout; bit 30, Unmapped address error (dmuae); or bit 29, uncorrected error from DMU (dmuue), is set, a UE is reported back to the thread that issue the PIO load. Bits 29:31 are logically **or**ed. No ESR is set in the SOC.

Note For Timeout error, the specific timeout error could be logged at bit 21 (primary error) or bit 53 (secondary error) of the PEU Other Event Status Clear register. If it is a primary error, the TLP header is logged at the PEU Transmit Other Event Header1/Header2 registers.The specific timeout error could also be logged in bit 14 (primary error) or bit 46 (secondary error) of the PEU Uncorrectable Error Status Clear register. If it is primary error, the TLP header is logged in the PEU Transmit Uncorrectable Error Header1/Header2 registers.

For Unmapped Address error, the specific unmapped address error is logged in bit 20 (primary error) or bit 52 (secondary error) of the PEU Uncorrectable Error Status Clear register. If it is a primary error, the original sending TLP header is logged in the PEU Transmit Uncorrectable Error Header1/Header2 registers. The completion header is logged at the PEU Receive Uncorrectable Error Header1/Header2 registers.

16.18.3 Correctable PIO Load Errors

The following correctable PIO load errors are recommended to *not* be set as fatal in the SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For correctable SOC detected errors associated with the PIO load, the NCU Load Return packet is sent to the requesting virtual processor with the err field indicating correctable error.

16.18.3.1 Correctable SII Ctag Error (SiiDmuCtagCE)

If a correctable Ctag error is detected, the Ctag is corrected and bit 3, siidmuctagce, in the SOC ESR is set. No other error information is provided. Data parity is not checked at this point. It will be checked in the NCU.

16.18.3.2 Correctable NCU Etag Error

If a correctable Ctag error is detected by the NCU, the Ctag is corrected and bit 2, ncuctagce, in the SOC ESR is set. No other error information is provided.

16.19 PIO Store Errors

There are two classes of PIO store errors detected in the SOC, fatal errors (FE), which cause a warm reset, and uncorrectable errors (UE), which generate a *sw_recoverable_error*.

The PIO store errors are presented in TABLE 16-47.

TABLE 16-47 SOC Errors for a PIO Store Request

Error	SOCESR bit	Recommended Error Type	Error Info (core)	Error Info (NCU)	Trap
NCU uncorrectable on command or thread_id from PCX	ncupcxue{19}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (etag, rqtyp, cpu_id, thread_id, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
NCU store data parity	ncupcxdata{18}	UE	DESR = SOCU if SOC EIE bit set	NCUSYN (etag, rqtyp, cpu_id, thread_id, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
NCU DMU Credit parity	ncudmucredit{42}	UE	DESR = SOCU if SOC EIE bit set	Not Recorded	Warm reset if configured as fatal; otherwise, sw_recoverable_error if SOC EIE bit set
NCU FIFO output error	ncucpxue{20}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
NCU DMU data parity	ncudmuue{21}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (etag, reqtype, cpuid, strandid, pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.

16.19.1 Uncorrectable PIO Store Errors Recommended as Fatal

The following uncorrectable PIO Store errors are recommended to be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. Fatal errors associated with a PIO store cause the PIO store to be dropped and a warm reset to be initiated.

16.19.1.1 Uncorrectable NCU FIFO Errors (NcuPcxUE)

The NCU provides DEDSEC protection on the NCU FIFO. If an uncorrectable error is detected on the command or thread ID fields, bit 19, ncupcxue, in the SOC ESR is set. The error information (etag, rqtyp, cpu_id, thread_id, and PA{39:0}) is recorded in the NCUSYN register.

No store ACK is returned, so the requesting strand will become hung. Therefore this error needs to be fatal.

16.19.1.2 NCU FIFO Output Error (NcuCpxUe)

The NCU checks the output FIFO to the CPX. If the NCU detects an error in the output FIFO, bit 20, ncucpxue, in the SOC ESR is set. No other error information is captured.

No store ACK is returned to the requesting strand, which will become hung. Therefore this error needs to be fatal.

16.19.1.3 NCU Parity Error from NCU DMU PIO Req FIFO (NcuDmuUe)

If the NCU detects a parity error from the NCU DMU PIO Req FIFO on a PIO Store, bit 21, ncudmuue, in the SOC ESR is set. The error information (etag, rqtyp, cpu_id, thread_id and PA{39:0}) is recorded in the NCUSYN register. NCU returns store ACK on CPX packet without any error. NCU squashes the PIO store request to DMU.

However since the parity error can be in cpu_id[2:0] bits , the store ACK might happen to the wrong CPU. Hence this error needs to be fatal.

16.19.2 Uncorrectable PIO Store Errors

The following uncorrectable PIO Store errors are *not* recommended to be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. The SOC sends an Error Indication (SOC) packet to the CPX with the err field indicating uncorrectable error.

16.19.2.1 NCU Store Data Parity Error (NcuPcxData)

The NCU checks data parity on the store data. If a data parity error is detected, bit 18, ncupcxdata, in the SOC ESR is set. No other error information is provided.

16.19.2.2 NCU DMU Credit Parity (NcuDmuCredit)

If NCU detects a parity error on the token returned from DMU to NCU after completion of a PIO Store Request , bit 42, ncudmucredit, in the SOC ESR is set . No other error information is provided.

16.20 Interrupt Errors

Interrupts can be sourced from either the NIU or the DMU (for PCI mondo interrupts). There are three classes of interrupt errors detected in the SOC: fatal errors (FE), which cause a warm reset; uncorrectable errors (UE), which generate a *sw_recoverable_error* trap to the strand specified in the SOC Error Steering register; or correctable errors (CE), which generate a *hw_corrected_error* trap to the strand specified in the SOC Error Steering register. The SOC Error Steering register is described in *SOC Error Steering Register* on page 345. Interrupt errors are presented in TABLE 16-48.

TABLE 16-48 SOC Errors for Interrupts

Error	SOCESR bit	Recommended Error Type	Error Info (core)	Error Info (NCU)	Trap
SII command parity error or Ctag UE	siidmuue{1}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU Ctag UE	ncuctague{22}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (ctag)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
DMU mondo ACK credit parity	dmuncucredit {9}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU data parity	ncudataparity14}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (ctag)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU mondo FIFO parity	ncumondofifo {16}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set

TABLE 16-48 SOC Errors for Interrupts

Error	SOCESR bit	Recommended Error Type	Error Info (core)	Error Info (NCU)	Trap
NCU interrupt table parity	ncuinttable{17}	FE	DESR = SOCU if SOC EIE bit set	NCUSYN (etag, reqtype, cpuid, strandid)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU FIFO output error	ncucpxue{20}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set
NCU mondo table parity	NCU Mondo Table{15}	UE	D-SFSR= SOCU, and DESR= SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *data_access_error.* and *sw_recoverable_error* if SOC EIE bit set
SII Ctag CE	siidmuce{3}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set
NCU Ctag CE	ncuctagce{23}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set

16.20.1 Uncorrectable Interrupt Errors Recommended as Fatal

The following uncorrectable Interrupt errors are recommended to be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. Fatal errors associated with an interrupt cause the interrupt to be dropped, and a warm reset to be initiated.

16.20.1.1 Uncorrectable SII Ctag/Command Parity Errors (SiiDmuCtagUE)

For interrupt requests (through the DMU) to the SII, the Ctag ECC is checked and also the command parity. If an uncorrectable Ctag ECC error or command parity error is detected, bit 1, siidmuctague, in the SOC ESR is set. The error information (etag, ctag and pa{39:0}) is recorded in the NCUSIISYN register.

For interrupts, this error condition does not have to be a fatal error, since the thread_id is not contained in the Ctag for interrupts. However, since the siidmuctague bit in the SOC ESR is shared with PIO loads, which do need to be fatal, the fatal error will also occur for interrupts when the siidmuctague bit is set.

16.20.1.2 Uncorrectable NCU Ctag Error (NcuCtagUe)

The NCU also checks Ctag ECC from the SII. If an uncorrectable Ctag ECC error is detected, bit 22, ncuctague, in the SOC ESR is set. The error information (etag, ctag) is recorded in the NCUSYN register.

For interrupts, this error condition does not have to be a fatal error, since the thread_id is not contained in the Ctag for interrupts. However, since the ncuctague bit in the SOC ESR is shared with PIO loads, which do need to be fatal, the fatal error will also occur for interrupts when the ncuctague bit is set.

16.20.1.3 DMU Mondo Ack Credit Parity(DmuNcuCredit)

DMU checks parity of Mondo Ack Credit from NCU. If a parity error is detected, bit 9, dmuncucredit , in the SOC ESR is set. There is no other information captured.

DMU Interrupt control unit keeps waiting for Ack forever, no future interrupt requests can be issued. So this needs to be a fatal error.

16.20.1.4 NCU Data Parity Error (NcuDataParity)

The NCU checks data parity for the interrupt requests from both the NIU and DMU (via the SII). If a data parity error is detected, bit 14, ncudataparity, in the SOC ESR is set. The error information (Etag, Ctag) is recorded in the NCUSYN register.

The interrupt is lost, so this needs to be a fatal error.

16.20.1.5 NCU FIFO Output Error (NcuCpxUe)

The NCU checks the output FIFO to the CPX. If the NCU detects an error in the output FIFO, bit 20, ncucpxue, in the SOC ESR is set. There is no other error information captured.

The interrupt is lost, so this needs to be a fatal error.

16.20.1.6 NCU Mondo FIFO Parity Error (NcuMondoFifo)

The NCU checks the data parity of the Mondo FIFO during interrupt processing. If a data parity error is detected during the read of the NCU Mondo FIFO, bit 16, ncumondofifo, in the SOC ESR is set. No other error information is captured.

NCU does not send data return CPX packet for load, so thread hangs. So this needs to be a fatal error.

16.20.1.7 NCU Interrupt Table Parity Error (NcuIntTable)

The NCU checks the data parity of the NCU Interrupt table during the processing of a non-Mondo interrupt. If a data parity error is detected during the read of the NCU Interrupt table, bit 17, ncuinttable, in the SOC ESR is set. The error information (etag, rqtyp, cpu_id, thread_id) is recorded in the NCUSYN register.

The interrupt is lost, so this needs to be a fatal error.

16.20.2 Uncorrectable Interrupt Errors

All uncorrectable Interrupt errors other than ncumondotable are recommended to be set as fatal. For uncorrectable SOC detected errors associated with an interrupt, the SOC sends the error Indication (SOC) packet to the CPX with the err field indicating uncorrectable error. The packet is sent to the target thread of the interrupt.

16.20.2.1 NCU Mondo Table Parity Error (NcuMondoTable)

NCU checks for data parity error while reading the Mondo table for a PIO read request.If the NCU detects an error, bit 15, ncumondotable, in the SOC ESR is set. There is no other error information captured. PIO load returns UE on the CPX packet to the requesting core.

16.20.3 Correctable Interrupt Errors

The following correctable Interrupt errors are recommended *not* to be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For correctable SOC detected errors associated with an interrupt, the SOC sends the Error Indication (SOC) packet to the CPX with the err field indicating correctable error. The packet is sent to the target thread of the interrupt

16.20.3.1 Correctable SII Ctag Error (SiiDmuCtagCE)

If a correctable Ctag error is detected, the Ctag is corrected and bit 3, siidmuctagce, in the SOCESR is set. No other error information is provided. Data parity is not checked at this point. It will be checked in the NCU.

16.20.3.2 Correctable NCU Ctag Error (NCUCtagCE)

If a correctable Ctag error is detected by the NCU, the Ctag is corrected and bit 23, ncuctagce, in the SOC ESR is set. No other error information is provided.

16.21 DMA Reads and Writes

I/O DMA requests to the L2 cache can be made from either the DMU or the NIU. There are three classes of DMA errors detected in the SOC: fatal errors (FE), which cause a warm reset; uncorrectable errors (UE), which generate a *sw_recoverable_error* trap to the strand specified in the SOC Error Steering register; or correctable errors (CE), which generate a *hw_corrected_error* trap to the strand specified in the SOC Error Steering register. The SOC Error Steering register is described in *SOC Error Steering Register* on page 345.

For DMA writes, errors can be detected in the SII section of the SOC and the L2 cache. See the L2 cache section for details on DMA errors.

For DMA reads, errors can be detected in the SII, SIO, NIU, and DMU sections of the SOC, and the L2 cache.

For DMA errors, the transaction continues to completion for correctable and uncorrectable errors. The e and ue bits in the I/O block packet indicate uncorrectable errors. For fatal errors, the transaction continues until the warm reset is effective.

TABLE 16-49 SOC Errors for DMA Read and Write Request

Error	SOC ESR bit	Recommended Error Type	Error Info (Core)	Error Info (NCU)	Trap
SII command parity error or Ctag UE from DMU	siidmuue{1}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII command parity error or Ctag UE from NIU	siiniuue{0}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SIO Ctag UE	sioctague{25}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
DMU Ctag UE	dmuctague{11}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.

TABLE 16-49 SOC Errors for DMA Read and Write Request *(Continued)*

Error	SOC ESR bit	Recommended Error Type	Error Info (Core)	Error Info (NCU)	Trap
NIU Ctag UE	niuctague{28}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
DMU Credit parity from SII	dmusiicredit{12}	FE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII address parity from DMU	siidmuaparity{7}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII address parity from NIU	siiniuaparity{4}	FE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII data parity from DMU	siidmudparity{5}	UE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SII data parity from NIU	siiniudparity{6}	UE	DESR = SOCU if SOC EIE bit set	NCUSIISYN (etag, ctag, and pa)	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
DMU Data parity from SIO	dmudataparity {13}	UE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
NIU Data parity from SIO	niudataparity{29}	UE	DESR = SOCU if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *sw_recoverable_error* if SOC EIE bit set.
SIO Ctag CE	sioctagce{26}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, hw_corrected_error if SOC EIE bit set.
SII Ctag CE from DMU	siidmuce{3}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set.
SII Ctag CE from NIU	siiniuce{2}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set.
DMU Ctag CE	dmuctagce{10}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set.
NIU Ctag CE	niuctagce{27}	CE	DESR = SOCC if SOC EIE bit set	Not recorded	Warm reset if configured as fatal; otherwise, *hw_corrected_error* if SOC EIE bit set.

16.21.1 Uncorrectable DMA Errors Recommended as Fatal

The following uncorrectable DMA errors are recommended to be set as fatal in the SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For fatal errors, the DMA transaction attempts to continue (with the write being quashed) until the warm reset occurs.

16.21.1.1 Uncorrectable SII Ctag ECC Error or Command Parity Error (SiiDmuCtagUe, SiiNiuCtagUe)

If an uncorrectable Ctag ECC error or command parity error is detected by the SII on a DMA access from either the DMU or the NIU, a fatal error is presented. The DMA access attempts to complete by setting the e bit in the packet to the L2 cache; however, the DMA write is quashed. If the DMA request is sourced from the DMU, the bit 1, siidmuctague, in the SOC ESR is set; if the request is sourced from the NIU, bit 0, siiniuctagce, in the SOC ESR is set. The error information (etag, ctag and pa[39:0]) is recorded in the NCUSIISYN register.

This error does not have to be fatal. However, the siidmuctague bit is shared with the PIO load access, which must be fatal. The siiniuctagce bit is made fatal for consistency.

16.21.1.2 Uncorrectable SIO Ctag ECC Error (SioCtagUe)

The L2 cache provides a response to the SIO for DMA reads, DMA 8-byte writes and write invalidates. If an uncorrectable Ctag ECC error is detected by the SIO on a DMA access, bit 25, sioctague, in the SOC ESR is set. No other error information is provided in the SOC. This error is recommended as fatal because it would cause indeterministic machine behavior.

16.21.1.3 Uncorrectable DMU Ctag ECC Error (DmuCtagUe)

If an uncorrectable Ctag ECC error is detected by the DMU on a DMA read access from the SIO, bit 11, dmuctague, in the SOC ESR is set. No other error information is provided in the SOC. This error is recommended as fatal because it would cause indeterministic machine behavior.

16.21.1.4 Uncorrectable NIU Ctag ECC Error (NiuCtagUe)

If an uncorrectable Ctag ECC error is detected by the NIU on a DMA read access from the SIO, bit 28, niuctague, in the SOC ESR is set. No other error information is provided in the SOC. This error is recommended as fatal because it would cause indeterministic machine behavior.

16.21.1.5 DMU Credit Parity error (DmuSiiCredit)

If the DMU detects a parity error in the credit field during a DMA write acknowledge from the SII, bit 12, dmusiicredit, in the SOC ESR is set. No other error information is provided in the SOC.This error is recommended as fatal because it would cause indeterministic machine behavior.

16.21.1.6 SII Address Parity Error (SiiDmuAParity, SiiNiuAParity)

If an address parity error is detected by the SII on a DMA write from either the DMU or the NIU, the write is squashed. For a DMA read, the e bit is set in the packet sent to the L2 cache. If the DMA request is sourced from the DMU, bit 7, siidmuaparity, in the SOC ESR is set; if the request is sourced from the NIU, bit 4, siiniuaparity, in the SOC ESR is set. The error information (etag, ctag, and PA{39:0}) is recorded in the NCUSIISYN register. This error is recommended as fatal because it would cause indeterministic machine behavior.

16.21.2 Uncorrectable DMA Errors

The following uncorrectable DMA errors are recommended to *not* be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For uncorrectable SOC detected errors associated with a DMA read or write, the transaction continues with the DMA write being squashed in the L2 cache; the DMA read continues with the e bit set in the I/O Block header for address parity errors, and the ue bit set for data parity errors.

The SOC sends the Error Indication (SOC) packet to the CPX with the err field indicating uncorrectable error. The packet is sent to the target thread of the trap request.

16.21.2.1 SII Data Parity Error (SiiDmuDParity, SiiNiuDParity)

If a data parity error is detected by the SII on a 64 byte DMA write from either the DMU or the NIU, data is corrupted (poisoned) by flipping the two least significant bits of the Data ECC field. If the data is sourced from the DMU, bit 5, siidmudparity,

in the SOC ESR is set; if the data is sourced from the NIU, bit 6, siiniudparity, in the SOC ESR is set. The error information (etag, ctag, and PA{39:0}) is recorded in the NCUSIISYN register.

However if there is a data parity error on a partial DMA write (less than or equal to 8 bytes) from either the DMU or the NIU, SII cannot detect the error and poison the data going to L2. As a result, silent data corruption happens in L2 and eventually memory.

16.21.2.2 DMU Data Parity Error (DmuDataParity)

On a DMA read access, if a data parity error is detected by the DMU on the data from the SIO, bit 13, dmudataparity, in the SOC ESR is set. No other error information is provided in the SOC.

16.21.2.3 NIU Data Parity Error (NiuDataParity)

On a DMA read access, if a data parity error is detected by the NIU on the data from the SIO, bit 29, niudataparity, in the SOC ESR is set. No other error information is provided in the SOC.

16.21.3 Correctable DMA Errors

The following correctable DMA errors are recommended to *not* be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. For correctable SOC detected errors associated with a DMA access, the SOC sends the Error Indication (SOC) packet to the CPX with the err field indicating correctable error. The packet is sent to the thread indicated in the Error Steering register field of the NCU Strand Enable Status register.

For correctable errors, the DMA read or write operation completes.

16.21.3.1 Correctable SII Ctag ECC Error (SiiDmuCtagCe, SiiNiuCtagCe)

If a correctable Ctag ECC error is detected by the SII on a DMA access from either the DMU or the NIU, the access completes. If the DMA request is sourced from the DMU, bit 3, siidmuctagce, in the SOC ESR is set; if the request is sourced from the NIU, bit 2, siiniuctagce, in the SOC ESR is set. No other error information is provided in the SOC.

16.21.3.2 Correctable SIO Ctag ECC Error (SioCtagCe)

If a correctable Ctag ECC error is detected by the SIO on a DMA access from either the DMU or the NIU, the access completes, and bit 26, sioctagce, in the SOC ESR is set. No other error information is provided in the SOC.

16.21.3.3 Correctable DMU Ctag ECC Error (DmuCtagCe)

If a correctable Ctag ECC error is detected by the DMU on a DMA read access, the access completes, and bit 10, dmuctagce, in the SOC ESR is set. No other error information is provided in the SOC.

16.21.3.4 Correctable NIU Ctag ECC Error (NiuCtagCe)

If a correctable Ctag ECC error is detected by the NIU on a DMA read access, the access completes, and bit 27, niuctagce, in the SOC ESR is set. No other error information is provided in the SOC.

16.22 MCU Correctable/Recoverable Count Errors

The MCU correctable/recoverable count errors are recommended to *not* be set as fatal in SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. There are four MCU FBdimms (MCU0 –MCU3. Each MCU FBdimm has an associated MCU Syndrome register, MCU ECC Correctable Error Count register, MCU Recoverable Error Count register, and associated error bits in the SOC ESR.

When the count registers transition from 1 to 0, an error signal is sent to the SOC. The SOC sets the corresponding bit in the SOC ESR, and sends the Error Indication (SOC) packet to the CPX with the err field indicating correctable error. The packet is sent to the thread indicated by the SOC Error Steering register.

TABLE 16-50 SOC Errors for MCU Error Count Interrupts

Error	SOCESR Bit	Recommended Error Type	Error Info (Core)	Error Info (MCU)	Trap
MCU0 correctable ECC error count reg = 0	mcu0ecc{32}	CE	DESR = SOCC	DRAM ESR0 (dsc, dac)	Warm reset if configured as fatal; otherwise, *hw_corrected_error.*
MCU1 correctable ECC error count reg = 0	mcu1ecc{35}	CE	DESR = SOCC	DRAM ESR1 (dsc, dac)	Warm reset if configured as fatal; otherwise, *hw_corrected_error.*
MCU2 correctable ECC error count reg = 0	mcu2ecc{38}	CE	DESR = SOCC	DRAM ESR2 (dsc, dac)	Warm reset if configured as fatal; otherwise, *hw_corrected_error.*

TABLE 16-50 SOC Errors for MCU Error Count Interrupts *(Continued)*

Error	SOCESR Bit	Recommended Error Type	Error Info (Core)	Error Info (MCU)	Trap
MCU3 correctable ECC error count reg = 0	mcu3ecc{41}	CE	DESR = SOCC	DRAM ESR3 (dsc, dac)	Warm reset if configured as fatal; otherwise, *hw_corrected_error*.
MCU0 recoverable error count reg = 0	mcu0fbr{31}	CE	DESR = SOCC	DRAM FDB SYN REG0 (alert, soft reset, fast reset, error source)	Warm reset if configured as fatal; otherwise, *hw_corrected_error*.
MCU1 recoverable error count reg = 0	mcu1fbr{34}	CE	DESR = SOCC	DRAM FDB SYN REG1 (alert, soft reset, fast reset, error source)	Warm reset PF if configured as fatal; otherwise, *hw_corrected_error*.
MCU2 recoverable error count reg = 0	mcu2fbr{37}	CE	DESR = SOCC	DRAM FDB SYN REG2 (alert, soft reset, fast reset, error source)	Warm reset if configured as fatal; otherwise, *hw_corrected_error*.
MCU3 recoverable error count reg = 0	mcu3fbr{40}	CE	DESR = SOCC	DRAM FDB SYN REG3 (alert, soft reset, fast reset, error source)	Warm reset if configured as fatal; otherwise, *hw_corrected_error*.

16.22.1 MCU ECC Correctable Errors (Mcu0ECC, Mcu1ECC, Mcu2ECC, Mcu3ECC)

Each MCU maintains a DRAM_ERROR_COUNT_REGISTER for counting correctable ECC errors and generating an error signal when the count decrements to zero.

When the count transitions from 1 to 0, the corresponding error bit in the SOC ESR is set: bit 32, mcu0ecc, for MCU0; bit 35, mcu1ecc, for MCU1; bit 38, mcu2ecc, for MCU2; and bit 41, mcu3ecc, for MCU3. Additional error information (DSC, DAC) is provided in the DRAM Error Status register in the corresponding MCU.

16.22.2 MCU Recoverable Errors (Mcu0Fbr, Mcu1Fbr, Mcu2Fbr, Mcu3Fbr)

Each MCU maintains a DRAM_FBR_COUNT_REG for counting recoverable link errors and generating an error signal when the count decrements to zero.

When the count transitions from 1 to 0 or if the count_one field in MCU Count register or MCU fbrcnt are set to 1, the corresponding error bit in the SOC ESR is set: bit 31, mcu0fbr, for MCU0; bit 34, mcu1fbr, for MCU1; bit 37, mcu2fb, for MCU2; and bit 40, mcu3fbr, for MCU3.

Additional error information (alert, soft reset, fast reset, error source) is captured in the corresponding DRAM FDB Syndrome register.

16.23 SOC Error Registers

This section describes the error control and log registers for the SOC. The SOC error registers are located in the NCU.

Error injection for the SOC error sources are provided by the SOC Error Injection register. The SOC provides a double-buffered error register for *sw_recoverable_error* or *hw_corrected_error* handling. All errors that are enabled by the SOC Error Log Enable register are captured in the SOC Error Status register. In addition, when a *sw_recoverable_error* or *hw_corrected_error* trap request is generated, the contents of the SOC Error Status register are copied to the SOC Pending Error Status register, and the SOC Error Status register is cleared. Each error source can generate no trap request, a *sw_recoverable_error* trap request, a *hw_corrected_error* trap request, or a warm reset of the UltraSPARC T2 chip under control of the SOC Error Interrupt Enable and SOC Fatal Error Enable registers. The strand receiving a trap request takes a disrupting trap, subject to masking and other restrictions of the specific trap.

16.23.1 SOC Error Status Register

The Error Status register contains status on SOC errors. The status bits in this register are cleared by writing a 0 to the bit position. This register is not changed on a warm reset to allow inspection of the error status by software following the warm reset. This register is cleared on generation of a trap request, and its contents are copied to the SOC Pending Error Status register.

Note | All SOC errors may be set to fatal using the SOC Fatal Error Enable register described in *SOC Fatal Error Enable Register* on page 346. The error type column lists the error type assuming the error has not been set to be fatal.

TABLE 16-51 shows the format of the SOC Error Status register.

TABLE 16-51 SOC Error Status Register – SOC_ERROR_STATUS_REG ($80\ 0000\ 3000_{16}$) *(1 of 4)*

Bit	Field	Error Type	Initial Value	R/W	Error Information (core)	Error Information (SOC)	Description
63	v	—	0	RW	—	—	Multiple uncorrected errors, one or more uncorrected errors were not logged.
62:43	SPARE4	—	0_{16}	RW	—	—	*Reserved* for errors to be assigned in future versions of chip if required
42	ncudmucredit	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if an uncorrectable parity error is detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	CE	0	RW	DESR = SOCC	DRAM ESR3 (dsc, dac)	Set to 1 if MCU 3 detected a correctable DRAM ECC error with its DRAM Error Count register reaching zero.
40	mcu3fbr	CE	0	RW	DESR = SOCC	DRAM FBD SYN REG3 (alert, soft reset, fast reset, error source)	Set to 1 if MCU 3 detected a FBDIMM recoverable error with its DRAM Recoverable Link Error Count register reaching zero.
39	SPARE3	—	0	RW	—	—	*Reserved* for errors to be assigned in future versions of chip if required
38	mcu2ecc	CE	0	RW	DESR = SOCC	DRAM ESR2 (dsc, dac)	Set to 1 if MCU 3 detected a correctable DRAM ECC error with its DRAM Error Count register reaching zero.
37	mcu2fbr	CE	0	RW	DESR = SOCC	DRAM FDB SYN REG2 (alert, soft reset, fast reset, error source)	Set to 1 if MCU 3 detected a FBDIMM recoverable error with its DRAM Recoverable Link Error Count register reaching zero.
36	SPARE2	—	0	RW	—	—	*Reserved* for errors to be assigned in future versions of chip if required
35	mcu1ecc	CE	0	RW	DESR = SOCC	DRAM ESR1 (dsc, dac)	Set to 1 if MCU 3 detected a correctable DRAM ECC error with its DRAM Error Count register reaching zero.

TABLE 16-51 SOC Error Status Register – SOC_ERROR_STATUS_REG ($80\ 0000\ 3000_{16}$) *(2 of 4)*

Bit	Field	Error Type	Initial Value	R/W	Error Information (core)	Error Information (SOC)	Description
34	mcu1fbr	CE	0	RW	DESR = SOCC	DRAM FDB SYN REG1 (alert, soft reset, fast reset, error source)	Set to 1 if MCU 3 detected a FBDIMM recoverable error with its DRAM Recoverable Link Error Count register reaching zero.
33	SPARE1	—	0	RW	—	—	*Reserved* for errors to be assigned in future versions of chip if required
32	mcu0ecc	CE	0	RW	DESR = SOCC	DRAM ESR0 (dsc, dac)	Set to 1 if MCU 0 detected a correctable DRAM ECC error with the DRAM Error Count register reaching zero.
31	mcu0fbr	CE	0	RW	DESR = SOCC	DRAM FDB SYN REG0 (alert, soft reset, fast reset, error source)	Set to 1 if MCU 0 detected a FBDIMM recoverable error with the DRAM Recoverable Link Error Count register reaching zero.
30	SPARE0	—	0	RW	—	—	*Reserved* for errors to be assigned in future versions of chip if required
29	niudataparity	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the SIO detected a CTAG uncorrected error from the old FIFO.
24	spare		0	RW	—	—	Hardware does not log any error for this or can inject any error for this, but asserts interrupt on software write.
23	ncuctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the NCU detected a CTAG corrected error on an interrupt write or a PIO read return from the SII.

TABLE 16-51 SOC Error Status Register – SOC_ERROR_STATUS_REG ($80\ 0000\ 3000_{16}$) *(3 of 4)*

Bit	Field	Error Type	Initial Value	R/W	Error Information (core)	Error Information (SOC)	Description
22	ncuctague	UE	0	RW	DESR = SOCU	NCUSYN (ctag)	Set to 1 if the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return from the SII.
21	ncudmuue	UE	0	RW	D-SFSR = SOCU, DESR = SOCU	NCUSYN (etag, rqtyp, cpu_id, thread_id, and PA)	Set to 1 if the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the NCU detected an error in the output FIFO to the crossbar.
19	ncupcxue	UE	0	RW	DESR = SOCU	NCUSYN (etag, rqtyp, cpu_id, thread_id, and PA)	Set to 1 if the NCU detected a command or Thread_id parity error in PIO/CSR commands from the processors (PCX FIFO).
18	ncupcxdata	UE	0	RW	DESR = SOCU	NCUSYN (etag, rqtyp, cpu_id, thread_id, and PA)	Set to 1 if the NCU detected a data parity error in PIO/CSR data from the processors (PCX FIFO).
17	ncuinttable	UE	0	RW	DESR = SOCU	NCUSYN (etag, rqtyp, cpu_id, thread_id, and pa)	Set to 1 if the NCU detected an error while reading the interrupt table for a non-mondo interrupt request.
16	ncumondofifo	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the NCU detected an error while reading the mondo FIFO for an interrupt request.
15	ncumondotable	UE	0	RW	D-SFSR = SOCU, DESR = SOCU	Not recorded	Set to 1 if the NCU detected an error while reading the mondo table for a PIO read request.
14	ncudataparity	UE	0	RW	D-SFSR = SOCU, DESR = SOCU	NCUSYN (ctag)	Set to 1 if the NCU detected a data parity error for interrupt write or PIO read return data through the SII from the DMU.
13	dmudataparity	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the DMU detected a data parity error in a DMA read return from the SIO.
12	dmusiicredit	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the DMU detected a parity error in the DMA write acknowledge credit from the SII.

TABLE 16-51 SOC Error Status Register – SOC_ERROR_STATUS_REG ($80\ 0000\ 3000_{16}$) *(4 of 4)*

Bit	Field	Error Type	Initial Value	R/W	Error Information (core)	Error Information (SOC)	Description
11	dmuctague	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the DMU detected an uncorrected CTAG error in the DMA read return from the SIO.
10	dmuctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the DMU detected a corrected CTAG error in the DMA read return from the SIO.
9	dmuncucredit	UE	0	RW	DESR = SOCU	Not recorded	Set to 1 if the DMU detected a parity error in the mondo acknowledge credit from NCU.
8	dmuinternal	UE	0	RW	Not recorded	Not recorded	Set to 1 if the DMU detected an internal error.
7	siidmuaparity	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and pa)	Set to 1 if the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and pa)	Set to 1 if the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and pa)	Set to 1 if the SII detected a parity error on data for DMA write transactions from DMU FIFO.
4	siiniuaparity	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and PA)	Set to 1 if the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the SII detected a corrected ECC CTAG error on a transaction from the DMU FIFO.
2	siiniuctagce	CE	0	RW	DESR = SOCC	Not recorded	Set to 1 if the SII detected a corrected ECC CTAG error on a transaction from the NIU FIFO.
1	siidmuctague	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and pa)	Set to 1 if the SII detected an uncorrectable ECC CTAG error or Command Parity error on a transaction from the DMU FIFO.
0	siiniuctague	UE	0	RW	DESR = SOCU	NCUSIISYN (etag, ctag, and pa)	Set to 1 if the SII detected an uncorrectable ECC CTAG error or Command Parity error on a transaction from the NIU FIFO.

If a software write of the error register happens on the same cycle as an error, the setting of bits by the error will be based on the register state before the write. The setting of fields by the error will take precedence over the same field being update by the write; however, fields that are not changed by the error will be updated by the write (for example, if the register has the ncucpxue bit set and software does a

write to clear that bit and the v bit on the same cycle as a NCUCTAGCE error, the error register would end up with the ncucpxue bit cleared and the ncuctagce and v bits set).

16.23.2 SOC Error Log Enable Register

This register enables the logging of SOC errors.

TABLE 16-52 shows the format of the SOC Error Log Enable register.

TABLE 16-52 SOC Error Log Enable Register – SOC_ERROR_LOG_ENABLE_REG ($80\ 0000\ 3008_{16}$) *(1 of 3)*

Bit	Field	Initial Value	R/W	Description
63:43	SPARE4	0_{16}	RW	*Reserved* (for errors to be assigned in future versions of chip if required).
42	ncudmucredit	1_{16}	RW	Set to 1 to log an uncorrectable parity error is detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	1_{16}	RW	Set to 1 to log MCU 3 exceeded data CE threshold.
40	mcu3fbr	1_{16}	RW	Set to 1 to log MCU 3 generated a FBDIMM recoverable error.
39	SPARE3	1_{16}	RW	*Reserved* (for errors to be assigned in future versions of chip if required).
38	mcu2ecc	1_{16}	RW	Set to 1 to log MCU 2 exceeded data CE threshold.
37	mcu2fbr	1_{16}	RW	Set to 1 to log MCU 2 generated a FBDIMM recoverable error.
36	SPARE2	1_{16}	RW	*Reserved* (for errors to be assigned in future versions of chip if required).
35	mcu1ecc	1_{16}	RW	Set to 1 to log MCU 1 exceeded data CE threshold.
34	mcu1fbr	1_{16}	RW	Set to 1 to log MCU 1 generated a FBDIMM recoverable error.
33	SPARE1	1_{16}	RW	*Reserved* (for errors to be assigned in future versions of chip if required).
32	mcu0ecc	1_{16}	RW	Set to 1 to log MCU 0 exceeded data CE threshold.
31	mcu0fbr	1_{16}	RW	Set to 1 to log MCU 0 generated a FBDIMM recoverable error.
30	SPARE0	1_{16}	RW	Spare 0 (for errors to be assigned in future versions of chip if required).
29	niudataparity	1_{16}	RW	Set to 1 to log the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	1_{16}	RW	Set to 1 to log the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	1_{16}	RW	Set to 1 to log the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	1_{16}	RW	Set to 1 to log the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	1_{16}	RW	Set to 1 to log the SIO detected a CTAG uncorrected error from the old FIFO.
24	SPARE	1_{16}	RW	Hardware does not log any error for this or can inject any error for this, but asserts interrupt on software write.
23	ncuctagce	1_{16}	RW	Set to 1 to log the NCU detected a CTAG corrected error on an interrupt write or a PIO read return.

TABLE 16-52 SOC Error Log Enable Register – SOC_ERROR_LOG_ENABLE_REG ($80\ 0000\ 3008_{16}$) *(2 of 3)*

Bit	Field	Initial Value	R/W	Description
22	ncuctague	1_{16}	RW	Set to 1 to log the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return.
21	ncudmuue	1_{16}	RW	Set to 1 to log the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	1_{16}	RW	Set to 1 to log the NCU detected an error in the Output FIFO to the crossbar.
19	ncupcxue	1_{16}	RW	Set to 1 to log the NCU detected an error in PIO/CSR commands from the processors.
18	ncupcxdata	1_{16}	RW	Set to 1 to log the NCU detected an error in PIO/CSR data from the processors.
17	ncuinttable	1_{16}	RW	Set to 1 to log the NCU detected an error while reading the interrupt table.
16	ncumondofifo	1_{16}	RW	Set to 1 to log the NCU detected an error while reading the mondo FIFO.
15	ncumondotable	1_{16}	RW	Set to 1 to log the NCU detected an error while reading the mondo table.
14	ncudataparity	1_{16}	RW	Set to 1 to log the NCU detected an parity error for interrupt write or PIO read return data.
13	dmudataparity	1_{16}	RW	Set to 1 to log the DMU detected a parity error in a DMA read return from the SIO.
12	dmusiicredit	1_{16}	RW	Set to 1 to log the DMU detected a parity error in the DMA write acknowledge credit from the SII.
11	dmuctague	1_{16}	RW	Set to 1 to log the DMU detected an uncorrected error in the DMA read return from the SIO.
10	dmuctagce	1_{16}	RW	Set to 1 to log the DMU detected a corrected error in the DMA read return from the SIO.
9	dmuncucredit	1_{16}	RW	Set to 1 to log the DMU detected a parity error in the Mondo acknowledge credit from NCU.
8	dmuinternal	1_{16}	RW	Set to 1 to log the DMU detected an internal error.
7	siidmuaparity	1_{16}	RW	Set to 1 to log the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	1_{16}	RW	Set to 1 to log the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	1_{16}	RW	Set to 1 to log the SII detected a parity error on data for DMA transactions from DMU FIFO.
4	siiniuaparity	1_{16}	RW	Set to 1 to log the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	1_{16}	RW	Set to 1 to log the SII detected a corrected error on a transaction from the DMU FIFO.

TABLE 16-52 SOC Error Log Enable Register – SOC_ERROR_LOG_ENABLE_REG ($80\ 0000\ 3008_{16}$) *(3 of 3)*

Bit	Field	Initial Value	R/W	Description
2	siiniuctagce	1_{16}	RW	Set to 1 to log the SII detected a corrected error on a transaction from the NIU FIFO.
1	siidmuctague	1_{16}	RW	Set to 1 to log the SII detected an uncorrected error on a transaction from the DMU FIFO.
0	siiniuctague	1_{16}	RW	Set to 1 to log the SII detected an uncorrected error on a transaction from the NIU FIFO.

16.23.3 SOC Error Interrupt Enable Register

This register controls which errors will generate a Error Indication (SOC) error packet to the CPX. If the eie bit is set, Error Indication (SOC) error packet will always be sent to the CPX irrespective of whether the error caused the transaction to be terminated or not. The Correctable SOC errors will set an error code of 01_2 in the err field of the packet while uncorrectable SOC errors set an error code of 10_2 in the err field.

TABLE 16-53 shows the format of the SOC Error Interrupt Enable register.

TABLE 16-53 SOC Error Interrupt Enable Register – SOC_ERROR_INTERRUPT_ENABLE_REG ($80\ 0000\ 3010_{16}$)

Bit	Field	Initial Value	R/W	Description
62:43	SPARE4	0	RW	Set to 1 to request trap on errors for whatever error bits would get assigned to bits 62:43 in future versions of chip.
42	ncudmucredit	0	RW	Set to 1 to request trap on an uncorrectable parity error is detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	0	RW	Set to 1 to request trap on MCU 3 exceeded data CE threshold.
40	mcu3fbr	0	RW	Set to 1 to request trap on MCU 3 generated a FBDIMM recoverable error.
39	spare3	0	RW	Set to 1 to request trap error for whatever error bit would get assigned to bits 39 in future versions of chip.
38	mcu2ecc	0	RW	Set to 1 to request trap on MCU 2 exceeded data CE threshold.
37	mcu2fbr	0	RW	Set to 1 to request trap on MCU 2 generated a FBDIMM recoverable error.
36	SPARE2	0	RW	Set to 1 to request trap error for whatever error bit would get assigned to bits 36 in future versions of chip.
35	mcu1ecc	0	RW	Set to 1 to request trap on MCU 1 exceeded data CE threshold.
34	mcu1fbr	0	RW	Set to 1 to request trap on MCU 1 generated a FBDIMM recoverable error.
33	SPARE1	0	RW	Set to 1 to request trap error for whatever error bit would get assigned to bits 33 in future versions of chip

TABLE 16-53 SOC Error Interrupt Enable Register – SOC_ERROR_INTERRUPT_ENABLE_REG ($80\ 0000\ 3010_{16}$) *(Continued)*

Bit	Field	Initial Value	R/W	Description
32	mcu0ecc	0	RW	Set to 1 to request trap on MCU 0 exceeded data CE threshold.
31	mcu0fbr	0	RW	Set to 1 to request trap on MCU 0 generated a FBDIMM recoverable error.
30	SPARE0	0	RW	Set to 1 to request trap error for whatever error bit would get assigned to bits 30 in future versions of chip
29	niudataparity	0	RW	Set to 1 to request trap on the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	0	RW	Set to 1 to request trap on the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	0	RW	Set to 1 to request trap on the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	0	RW	Set to 1 to request trap on the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	0	RW	Set to 1 to request trap on the SIO detected a CTAG uncorrected error from the old FIFO.
24	spare	0	RW	HW does not log any error for this or can inject any error for this, but requests trap on SW write.
23	ncuctagce	0	RW	Set to 1 to request trap on the NCU detected a CTAG corrected error on an interrupt write or a PIO read return.
22	ncuctague	0	RW	Set to 1 to request trap on the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return.
21	ncudmuue	0	RW	Set to 1 to request trap on the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	0	RW	Set to 1 to request trap on the NCU detected an error in the output FIFO to the crossbar.
19	ncupcxue	0	RW	Set to 1 to request trap on the NCU detected an error in PIO/CSR commands from the processors.
18	ncupcxdata	0	RW	Set to 1 to request trap on the NCU detected an error in PIO/CSR data from the processors.
17	ncuinttable	0	RW	Set to 1 to request trap on the NCU detected an error while reading the interrupt table.
16	ncumondofifo	0	RW	Set to 1 to request trap on the NCU detected an error while reading the mondo FIFO.
15	ncumondotable	0	RW	Set to 1 to request trap on the NCU detected an error while reading the mondo table.
14	ncudataparity	0	RW	Set to 1 to request trap on the NCU detected an parity error for interrupt write or PIO read return data.
13	dmudataparity	0	RW	Set to 1 to request trap on the DMU detected a parity error in a DMA read return from the SIO.
12	dmusiicredit	0	RW	Set to 1 to request trap on the DMU detected a parity error in the DMA write acknowledge credit from the SII.

TABLE 16-53 SOC Error Interrupt Enable Register – SOC_ERROR_INTERRUPT_ENABLE_REG ($80\ 0000\ 3010_{16}$) *(Continued)*

Bit	Field	Initial Value	R/W	Description
11	dmuctague	0	RW	Set to 1 to request trap on the DMU detected an uncorrected error in the DMA read return from the SIO.
10	dmuctagce	0	RW	Set to 1 to request trap on the DMU detected a corrected error in the DMA read return from the SIO.
9	dmuncucredit	0	RW	Set to 1 to request trap on the DMU detected a parity error in the Mondo acknowledge credit from NCU.
8	dmuinternal	0	RW	Set to 1 to request trap on the DMU detected an internal error.
7	siidmuaparity	0	RW	Set to 1 to request trap on the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	0	RW	Set to 1 to request trap on the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	0	RW	Set to 1 to request trap on the SII detected a parity error on data for DMA transactions from DMU FIFO.
4	siiniuaparity	0	RW	Set to 1 to request trap on the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	0	RW	Set to 1 to request trap on the SII detected a corrected error on a transaction from the DMU FIFO.
2	siiniuctagce	0	RW	Set to 1 to request trap on the SII detected a corrected error on a transaction from the NIU FIFO.
1	siidmuctague	0	RW	Set to 1 to request trap on the SII detected an uncorrected error on a transaction from the DMU FIFO.
0	siiniuctague	0	RW	Set to 1 to request trap on the SII detected an uncorrected error on a transaction from the NIU FIFO.

16.23.4 SOC Error Steering Register

This register controls which virtual processor will be sent SOC error disrupting trap requests.

TABLE 16-53 shows the format of the SOC Error Steering register.

TABLE 16-54 SOC Error Steering Register – SOC_ERROR_STEER_REG ($90\ 0104\ 1000_{16}$)

Bit	Field	Initial Value	R/W	Description
62:6	—	0	RO	*Reserved*
5:0	vcid	0	RW	ID of virtual processor that will be target of SOC error trap requests.

Notes Software should program the Error Steering register the same in NCU and L2 so that multiple errors for the same error reported by both L2 and NCU go to same thread.

For errors reported multiple times (through PIO load return precise trap on crossbar and then SOC error packet and disruptive trap) the vcid can be different for the traps.

If the Error Steering register points to a virtual processor that is not available or not enabled, any SOC error disurpting trap request will be dropped.

If the Error Steering register points to a virutal processor that is parked, any SOC error disrupting trap request will be held pending by target virtual processor until it is running and the trap is unmasked.

If the Error Steering register points to a virutal processor that is halted, any SOC error disrupting trap request will cause the target virtual processor to return to running state; the trap request will be held pending until the trap is unmasked.

16.23.5 SOC Fatal Error Enable Register

This register controls which errors will generate a fatal error, resetting UltraSPARC T2.

TABLE 16-55 shows the format of the SOC Fatal Error Enable register.

TABLE 16-55 SOC Fatal Error Enable Register – SOC_FATAL_ERROR_ENABLE_REG ($80\ 0000\ 3020_{16}$) *(1 of 3)*

Bit	Field	Initial Value	R/W	Description
62:43	SPARE4	0	RW	Set to 1 to make fatal for whatever error bit would get assigned to bits 62:43 in future versions of chip
42	ncudmucredit	0	RW	Set to 1 to make fatal an uncorrectable parity error detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	0	RW	Set to 1 to make fatal MCU 3 exceeded data CE threshold.
40	mcu3fbr	0	RW	Set to 1 to make fatal MCU 3 generate an FBDIMM recoverable error.
39	SPARE3	0	RW	Set to 1 to make fatal whatever error bit would get assigned to bits 39 in future versions of chip
38	mcu2ecc	0	RW	Set to 1 to make fatal MCU 2 exceeded data CE threshold.
37	mcu2fbr	0	RW	Set to 1 to make fatal MCU 2 generated a FBDIMM recoverable error.
36	SPARE2	0	RW	Set to 1 to make fatal for whatever error bit would get assigned to bits 36 in future versions of chip

TABLE 16-55 SOC Fatal Error Enable Register – SOC_FATAL_ERROR_ENABLE_REG ($80\ 0000\ 3020_{16}$) *(2 of 3)*

Bit	Field	Initial Value	R/W	Description
35	mcu1ecc	0	RW	Set to 1 to make fatal MCU 1 exceeded data CE threshold.
34	mcu1fbr	0	RW	Set to 1 to make fatal MCU 1 generated a FBDIMM recoverable error.
33	SPARE1	0	RW	Set to 1 to make fatal for whatever error bit would get assigned to bits 33 in future versions of chip
32	mcu0ecc	0	RW	Set to 1 to make fatal MCU 0 exceeded data CE threshold.
31	mcu0fbr	0	RW	Set to 1 to make fatal MCU 0 generated a FBDIMM recoverable error.
30	SPARE0	0	RW	Set to 1 to make fatal for whatever error bit would get assigned to bits 30 in future versions of chip.
29	niudataparity	0	RW	Set to 1 to make fatal the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	0	RW	Set to 1 to make fatal the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	0	RW	Set to 1 to make fatal the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	0	RW	Set to 1 to make fatal the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	0	RW	Set to 1 to make fatal the SIO detected a CTAG uncorrected error from the old FIFO.
24	spare	0	RW	HW does not log any error for this or can inject any error for this, but reuests trap on SW write.
23	ncuctagce	0	RW	Set to 1 to make fatal the NCU detected a CTAG corrected error on an interrupt write or a PIO read return.
22	ncuctague	0	RW	Set to 1 to make fatal the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return.
21	ncudmuue	0	RW	Set to 1 to make fatal the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	0	RW	Set to 1 to make fatal the NCU detected an error in the output FIFO to the crossbar.
19	ncupcxue	0	RW	Set to 1 to make fatal the NCU detected an error in PIO/CSR commands from the processors.
18	ncupcxdata	0	RW	Set to 1 to make fatal the NCU detected an error in PIO/CSR data from the processors.
17	ncuinttable	0	RW	Set to 1 to make fatal the NCU detected an error while reading the interrupt table.
16	ncumondofifo	0	RW	Set to 1 to make fatal the NCU detected an error while reading the mondo FIFO.
15	ncumondotable	0	RW	Set to 1 to make fatal the NCU detected an error while reading the mondo table.
14	ncudataparity	0	RW	Set to 1 to make fatal the NCU detected an parity error for interrupt write or PIO read return data.

TABLE 16-55 SOC Fatal Error Enable Register – SOC_FATAL_ERROR_ENABLE_REG ($80\ 0000\ 3020_{16}$) *(3 of 3)*

Bit	Field	Initial Value	R/W	Description
13	dmudataparity	0	RW	Set to 1 to make fatal the DMU detected a parity error in a DMA read return from the SIO.
12	dmusiicredit	0	RW	Set to 1 to make fatal the DMU detected a parity error in the DMA write acknowledge credit from the SII.
11	dmuctague	0	RW	Set to 1 to make fatal the DMU detected an uncorrected error in the DMA read return from the SIO.
10	dmuctagce	0	RW	Set to 1 to make fatal the DMU detected a corrected error in the DMA read return from the SIO.
9	dmuncucredit	0	RW	Set to 1 to make fatal the DMU detected a parity error in the Mondo acknowledge credit from NCU.
8	dmuinternal	0	RW	Set to 1 to make fatal the DMU detected an internal error.
7	siidmuaparity	0	RW	Set to 1 to make fatal the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	0	RW	Set to 1 to make fatal the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	0	RW	Set to 1 to make fatal the SII detected a parity error on data for DMA transactions from DMU FIFO.
4	siiniuaparity	0	RW	Set to 1 to make fatal the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	0	RW	Set to 1 to make fatal the SII detected a corrected error on a transaction from the DMU FIFO.
2	siiniuctagce	0	RW	Set to 1 to make fatal the SII detected a corrected error on a transaction from the NIU FIFO.
1	siidmuctague	0	RW	Set to 1 to make fatal the SII detected an uncorrected error on a transaction from the DMU FIFO.
0	siiniuctague	0	RW	Set to 1 to make fatal the SII detected an uncorrected error on a transaction from the NIU FIFO.

16.23.6 SOC Pending Error Status Register

The Pending Error Status register contains the state of the SOC_ERROR_STATUS_REG when the disrupting trap request was generated as a result of an SOC error logged that had its corresponding bit set in SOC_ERROR_INTERRUPT_ENABLE_REG. The valid bit of this register prevents further disrupting trap requests from being generated by the SOC. This register is not changed on a warm reset to allow inspection of the error status by software following the warm reset.

TABLE 16-56 shows the format of the SOC Pending Error Status register.

TABLE 16-56 SOC Pending Error Status Register – SOC_PENDING_ERROR_STATUS_REG ($80\ 0000\ 3028_{16}$) *(1 of 2)*

Bit	Field	Initial Value	R/W	Description
63	v	0	RW	Valid bit, prevents generation of further SOC *sw_recoverable_error* traps.
62:43	SPARE4	0	RW	*Reserved* for future versions of the chip.
42	ncudmucredit	0	RW	Set to 1 to if an uncorrectable parity error is detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	0	RW	Set to 1 if MCU 3 exceeded data CE threshold.
40	mcu3fbr	0	RW	Set to 1 if MCU 3 generated a FBDIMM recoverable error.
39	SPARE3	0	RW	*Reserved* for future versions of the chip.
38	mcu2ecc	0	RW	Set to 1 if MCU 2 exceeded data CE threshold.
37	mcu2fbr	0	RW	Set to 1 if MCU 2 generated a FBDIMM recoverable error.
36	SPARE2	0	RW	*Reserved* for future versions of the chip.
35	mcu1ecc	0	RW	Set to 1 if MCU 1 exceeded data CE threshold.
34	mcu1fbr	0	RW	Set to 1 if MCU 1 generated a FBDIMM recoverable error.
33	spare1	0	RW	*Reserved* for future versions of the chip.
32	mcu0ecc	0	RW	Set to 1 if MCU 0 exceeded data CE threshold.
31	mcu0fbr	0	RW	Set to 1 if MCU 0 generated a FBDIMM recoverable error.
30	SPARE0	0	RW	*Reserved* for future versions of the chip.
29	niudataparity	0	RW	Set to 1 if the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	0	RW	Set to 1 if the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	0	RW	Set to 1 if the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	0	RW	Set to 1 if the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	0	RW	Set to 1 if the SIO detected a CTAG uncorrected error from the old FIFO.
24	spare	0	RW	Hardware does not log any error for this or can inject any error for this, but requests trap on SW write.
23	ncuctagce	0	RW	Set to 1 if the NCU detected a CTAG corrected error on an interrupt write or a PIO read return.
22	ncuctague	0	RW	Set to 1 if the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return.
21	ncudmuue	0	RW	Set to 1 if the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	0	RW	Set to 1 if the NCU detected an error in the output FIFO to the crossbar.
19	ncupcxue	0	RW	Set to 1 if the NCU detected an error in PIO/CSR commands from the processors.
18	ncupcxdata	0	RW	Set to 1 if the NCU detected an error in PIO/CSR data from the processors.

TABLE 16-56 SOC Pending Error Status Register – SOC_PENDING_ERROR_STATUS_REG ($80\ 0000\ 3028_{16}$) *(2 of 2)*

Bit	Field	Initial Value	R/W	Description
17	ncuinttable	0	RW	Set to 1 if the NCU detected an error while reading the interrupt table.
16	ncumondofifo	0	RW	Set to 1 if the NCU detected an error while reading the mondo FIFO.
15	ncumondotable	0	RW	Set to 1 if the NCU detected an error while reading the mondo table.
14	ncudataparity	0	RW	Set to 1 if the NCU detected an parity error for interrupt write or PIO read return data.
13	dmudataparity	0	RW	Set to 1 if the DMU detected a parity error in a DMA read return from the SIO.
12	dmusiicredit	0	RW	Set to 1 if the DMU detected a parity error in the DMA write acknowledge credit from the SII.
11	dmuctague	0	RW	Set to 1 if the DMU detected an uncorrected error in the DMA read return from the SIO.
10	dmuctagce	0	RW	Set to 1 if the DMU detected a corrected error in the DMA read return from the SIO.
9	dmuncucredit	0	RW	Set to 1 if the DMU detected a parity error in the Mondo acknowledge credit from NCU.
8	dmuinternal	0	RW	Set to 1 if the DMU detected an internal error.
7	siidmuaparity	0	RW	Set to 1 if the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	0	RW	Set to 1 if the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	0	RW	Set to 1 if the SII detected a parity error on data for DMA transactions from DMU FIFO.
4	siiniuaparity	0	RW	Set to 1 if the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	0	RW	Set to 1 if the SII detected a corrected error on a transaction from the DMU FIFO.
2	siiniuctagce	0	RW	Set to 1 if the SII detected a corrected error on a transaction from the NIU FIFO.
1	siidmuctague	0	RW	Set to 1 if the SII detected an uncorrected error on a transaction from the DMU FIFO.
0	siiniuctague	0	RW	Set to 1 if the SII detected an uncorrected error on a transaction from the NIU FIFO.

16.23.7 SOC Error Injection Register

This register controls the injection of errors. Continuous errors are generated for any error types with their bit set in the register.

TABLE 16-57 shows the format of the SOC Error Injection register.

TABLE 16-57 SOC Error Injection Register – SOC_ERROR_INJECTION_REG ($80\ 0000\ 3018_{16}$) *(1 of 3)*

Bit	Field	Initial Value	R/W	Description
63:43	SPARE4	0	RW	*Reserved* for future versions of the chip.
42	ncudmucredit	0	RW	Set to 1 to enable error injection for an uncorrectable parity error is detected on the credit token bus to NCU for DMU PIO write credits.
41	mcu3ecc	0	RW	Set to 1 to enable error injection on MCU 3 exceeded data CE threshold.
40	mcu3fbr	0	RW	Set to 1 to enable error injection on MCU 3 generated a FBDIMM recoverable error.
39	mcu3fbu	0	RW	Set to 1 to enable error injection on MCU 3 generated a FBDIMM unrecoverable error.
38	mcu2ecc	0	RW	Set to 1 to enable error injection on MCU 2 exceeded data CE threshold.
37	mcu2fbr	0	RW	Set to 1 to enable error injection on MCU 2 generated a FBDIMM recoverable error.
36	mcu2fbu	0	RW	Set to 1 to enable error injection on MCU 2 generated a FBDIMM unrecoverable error.
35	mcu1ecc	0	RW	Set to 1 to enable error injection on MCU 1 exceeded data CE threshold.
34	mcu1fbr	0	RW	Set to 1 to enable error injection on MCU 1 generated a FBDIMM recoverable error.
33	mcu1fbu	0	RW	Set to 1 to enable error injection on MCU 1 generated a FBDIMM unrecoverable error.
32	mcu0ecc	0	RW	Set to 1 to enable error injection on MCU 0 exceeded data CE threshold.
31	mcu0fbr	0	RW	Set to 1 to enable error injection on MCU 0 generated a FBDIMM recoverable error.
30	mcu0fbu	0	RW	Set to 1 to enable error injection on MCU 0 generated a FBDIMM unrecoverable error.
29	niudataparity	0	RW	Set to 1 to enable error injection on the NIU detected a data parity error in the DMA read return from the SIO.
28	niuctague	0	RW	Set to 1 to enable error injection on the NIU detected a CTAG uncorrected error in the DMA read return from the SIO.
27	niuctagce	0	RW	Set to 1 to enable error injection on the NIU detected a CTAG corrected error in the DMA read return from the SIO.
26	sioctagce	0	RW	Set to 1 to enable error injection on the SIO detected a CTAG corrected error from the old FIFO.
25	sioctague	0	RW	Set to 1 to enable error injection on the SIO detected a CTAG uncorrected error from the old FIFO.
24	SPARE	0	RW	Hardware does not log any error for this or can inject any error for this, but asserts requests trap on software write.
23	ncuctagce	0	RW	Set to 1 to enable error injection on the NCU detected a CTAG corrected error on an interrupt write or a PIO read return.
22	ncuctague	0	RW	Set to 1 to enable error injection on the NCU detected a CTAG uncorrected error on an interrupt write or a PIO read return.

TABLE 16-57 SOC Error Injection Register – SOC_ERROR_INJECTION_REG ($80\ 0000\ 3018_{16}$) *(2 of 3)*

Bit	Field	Initial Value	R/W	Description
21	ncudmuue	0	RW	Set to 1 to enable error injection on the NCU detected a parity error in the NCU DMU PIO Req FIFO.
20	ncucpxue	0	RW	Set to 1 to enable error injection on the NCU detected an error in the output FIFO to the crossbar.
19	ncupcxue	0	RW	Set to 1 to enable error injection on the NCU detected an error in PIO/ CSR commands from the processors.
18	ncupcxdata	0	RW	Set to 1 to enable error injection on the NCU detected an error in PIO/ CSR data from the processors.
17	ncuinttable	0	RW	Set to 1 to enable error injection on the NCU detected an error while reading the interrupt table.
16	ncumondofifo	0	RW	Set to 1 to enable error injection on the NCU detected an error while reading the mondo FIFO.
15	ncumondotable	0	RW	Set to 1 to enable error injection on the NCU detected an error while reading the mondo table.
14	ncudataparity	0	RW	Set to 1 to enable error injection on the NCU detected an parity error for interrupt write or PIO read return data.
13	dmudataparity	0	RW	Set to 1 to enable error injection on the DMU detected a parity error in a DMA read return from the SIO.
12	dmusiicredit	0	RW	Set to 1 to enable error injection on the DMU detected a parity error in the DMA write acknowledge credit from the SII.
11	dmuctague	0	RW	Set to 1 to enable error injection on the DMU detected an uncorrected error in the DMA read return from the SIO.
10	dmuctagce	0	RW	Set to 1 to enable error injection on the DMU detected a corrected error in the DMA read return from the SIO.
9	dmuncucredit	0	RW	Set to 1 to enable error injection on the DMU detected a parity error in the Mondo acknowledge credit from NCU.
8	dmuinternal	0	RW	Set to 1 to enable error injection on the DMU detected an internal error.
7	siidmuaparity	0	RW	Set to 1 to enable error injection on the SII detected a parity error on address field for DMA transactions from DMU FIFO.
6	siiniudparity	0	RW	Set to 1 to enable error injection on the SII detected a parity error on data for DMA transactions from NIU FIFO.
5	siidmudparity	0	RW	Set to 1 to enable error injection on the SII detected a parity error on data for DMA transactions from DMU FIFO.
4	siiniuaparity	0	RW	Set to 1 to enable error injection on the SII detected a parity error on address field for DMA transactions from NIU FIFO.
3	siidmuctagce	0	RW	Set to 1 to enable error injection on the SII detected a corrected error on a transaction from the DMU FIFO.

TABLE 16-57 SOC Error Injection Register – SOC_ERROR_INJECTION_REG ($80\ 0000\ 3018_{16}$) *(3 of 3)*

Bit	Field	Initial Value	R/W	Description
2	siiniuctagce	0	RW	Set to 1 to enable error injection on the SII detected a corrected error on a transaction from the NIU FIFO.
1	siidmuctague	0	RW	Set to 1 to enable error injection on the SII detected an uncorrected error on a transaction from the DMU FIFO.
0	siiniuctague	0	RW	Set to 1 to enable error injection on the SII detected an uncorrected error on a transaction from the NIU FIFO.

16.23.8 SOC SII Error Syndrome Register

This register logs the SII Error Syndrome for the errors listed in the etag field. This register is not changed on a warm reset to allow inspection of the syndrome by software following the warm reset.

TABLE 16-58 shows the format of the SOC SII Error Syndrome register.

TABLE 16-58 SOC SII Error Syndrome Register – SOC_SII_ERROR_SYNDROME_REG ($80\ 0000\ 3030_{16}$)

Bit	Field	Initial Value	R/W	Description
63	v	0	RW	Valid bit, set to 1 when the syndrome has been logged.
62:59	—	0	RO	*Reserved*
58:56	etag	0	RW	Error tag: 7 – SIIDMUAPARITY; 6 – SIINIUDPARITY; 5 – SIIDMUDPARITY; 4 – SIINIUAPARITY; 1– SIIDMUCTAGUE; 0 – SIINIUCTAGUE
55:40	ctag	0	RW	ctag field from the header.
39:0	pa	0	RW	Physical address.

16.23.9 SOC NCU Error Syndrome Register

This register logs the NCU Error Syndrome for the NCUCTAGUE, NCUDMUUE, NCUPCXUE, NCUPCXDATA, NCUINTTABLE, and NCUDATAPARITY errors. This register is not changed on a warm reset to allow inspection of the syndrome by software following the warm reset.

TABLE 16-59 shows the format of the SOC NCU Error Syndrome register.

TABLE 16-59 SOC NCU Error Syndrome Register – SOC_NCU_ERROR_SYNDROME_REG ($80\text{-}0000\text{-}3038_{16}$)

Bit	Field	Initial Value	R/W	Description
63	v	0	RW	Valid bit, set to 1 when the syndrome has been logged.
62	g	0	RW	Valid bit for CTAG field. If g = 0, none of the other field valid bits (r, c, s, or p) will be set and bits 15:0 indicate the Ctag. If v = 1, the other field valid bits can be nonzero and bits 39:0 indicate the PA.
61	r	0	RW	Valid bit for reqtype field.
60	c	0	RW	Valid bit for coreid field.
59	s	0	RW	Valid bit for strandid field.
58	p	0	RW	Valid bit for pa field.
57:56	—	0	RO	*Reserved*
55:51	etag	0	RW	Error tag: 42 – NCUDMUCredit; 41 – MCU3ECC; 40 – MCU3FBR; 38 – MCU2ECC; 37 – MCU2FBR; 35 – MCU1ECC; 34 – MCU1FBR; 32 – MCU0ECC; 31 – MCU0FBR; 29 – NIUDATAPARITY; 28 – NIUCTAGUE; 27 – NIUCTAGCE; 26 – SIOCTAGCE; 25 – SIOCTAGUE; 24 – TESTMODE; 23 – NCUCTAGCE; 22 – NCUCTAGUE; 21 – NCUDMUUE; 20 – NCUCPXUE; 19 – NCUPCXUE; 18 – NCUPCXDATA; 17 – NCUINTTABLE 16 – NCUMONDOFIFO; 15 – NCUMONDOTABLE 14 – NCUDATAPARITY; 13 – DMUDATAPARITY; 12 – DMUSIICREDIT 11 – DMUCTAGUE; 10 – DMUCTAGCE; 9 – DMUNCUCREDIT 8 – DMUINTERNAL; 3 – SIIDMUCTAGCE; 2 – SIINIUCTAGCE
50:46	reqtype	0	RW	Request type
45:43	coreid	0	RW	Physical strand ID.
42:40	strandid	0	RW	Strand ID on physical core.
39:0	pa_ctag	0	RW	Physical address{39:0} if p is set. If g bit is set, contains ctag in 15:0.

CHAPTER 17

Memory Controller

17.1 Overview

UltraSPARC T2 interfaces to external registered DDR2 fully buffered DIMMs (FBDs) through unidirectional high-speed links. UltraSPARC T2 II interfaces directly to external registered DDR2 DIMMs. There are four memory branches on UltraSPARC T2. Each memory branch services 64-byte read and write requests from two L2 cache banks of the on-chip L2 Cache unit.

The features of the UltraSPARC T2 memory controller are as follows:

- Uses 10-bit southbound and 14-bit northbound FBD channel protocols running at 12 times the SDRAM cycle rate.
- Supports 256-Mbit DRAM components for x4 data width; supports 512-Mbit, 1-Gbit, and 2-Gbit DRAM components for x4 and x8 data widths.
- Maximum memory of 128 Gbytes per branch using sixteen 8-Gbyte DDR2 FBDs
- Supports up to 16 ranks of DDR2 DIMMs per branch (8 pairs of double-sided FBDs)
- Supports registered DDR2 DIMMs of clock frequency up to 400 MHz
- Supports 128 bits of write data and 16 bits ECC per SDRAM cycle and 256 bits of read data and 32 bits ECC per SDRAM cycle.
- Supports DDR2 SDRAM burst length of 4 when using both FBD channels in a branch, burst length of 8 when using a single channel per branch.
- ECC generation, check, correction, and Extended ECC.
- Programmable DDR2 SDRAM power throttle control
- System peak memory bandwidth (4 branches): 50 Gbytes/s for reads, 25 Gbytes/s for writes.
-

Note | UltraSPARC T2 does not support the FBD Hot Plug feature.

17.2 Memory Terminology

A few of the more common memory and DRAM terms are described here.

bank Most DDR SDRAM chips are broken up into four or eight logical banks internally to enable full pipelining of memory operations.

channel Port connecting processor chip to DIMM.

DIMM Dual Inline Memory Module. Industry-standard SDRAM module package. A stick of memory.

DRAM chip Single chip inside the DIMM. We differentiate the type by how many bits it outputs and its capacity. (x4 means 4-bit output, x8 means 8-bit output, x16, x32 etc., and 256-Mbit or 512-Mbit capacity). Most common ones are the x4, x8 outputs.

rank A data group that can be accessed from a DIMM. Each DIMM has two chip selects. When a DIMM has two ranks, each chip select accesses DRAMs on one side of the DIMM independently. When a DIMM has one rank, both chip selects must be asserted at the same time to access all DRAMs on the DIMM. For x4 SDRAMs, single-rank DIMMs have 18 devices and double-rank DIMMs have 36 devices.

RAS/CAS RAS stands for "row address strobe." When this signal is asserted, a particular bank is enabled. It is also often referred to as "active" command. CAS stands for "column address strobe." When this signal is asserted, the column address and Read/Write signals are transmitted.

refresh DRAM requires what is often referred to as "refresh" cycle. Every row in the DRAM requires a refresh access every 15.6µS/7.8µS.

single-channel mode A low-power configuration with one DIMM per memory channel. Only one FBD channel is used, and the memory burst length is 8.

17.3 Fully Buffered DIMM (FBD) Terminology

A few of the more common FBD terms are described here.

advanced memory buffer (AMB) Buffers memory traffic between the host and the SDRAMs. Requests are sent by the host to the AMB across a high-speed link, and the AMB drives the requests to the SDRAMs using the DDR2 protocol.

bit lane A differential pair of signals in one direction.

cyclic redundancy code (CRC) An error detection code sent with data across the FBD link to protect the data from errors. When a CRC error is detected, the faulty frame must be retransmitted.

DDR branch A minimum aggregation of DDR channels that operate in lock-step to support error correction. A rank spans a branch. In UltraSPARC T2, a branch consists of one or two DDR channels.

DDR channel A channel that consists of a data channel with 72 bits of data and an addr/cntrl channel.

DDR data channel A data channel that consists of 72 bits of data divided into 18 data groups.

FBD Fully buffered DIMM.

frame Groups of bits containing commands or data sent across the link over 12 cycles.

Linear Feedback Shift register (LFSR) A shift register where the data input to the last register is a function of the outputs of other registers.

link High-speed parallel differential point-to-point interface.

northbound (NB) The direction of signals running from the farthest DIMM toward the host.

slot Socket for a DIMM.

southbound (SB) The direction of signals running from the host controller toward the DIMMs.

training sequence (TS) A sequence of bits sent per bit lane from the host to the FBDs to initialize the channel operation.

unit interval (UI) Average time interval between voltage transitions of a signal. Approximately 200 ps for DIMMs running at 800 MHz.

17.4 DRAM Branch Configuration

Each DRAM branch can have a different memory size and a different kind of DIMM (for example, a different number of ranks or different CAS latency). Software should not use address space larger than four times the lowest memory capacity in a channel because the cache lines are interleaved across channels.

The MCU employs the following design requirements:

- x4 and x8 DRAM parts are supported. Extended ECC is not supported for x8 DRAM parts.
- DIMM capacity, configuration, and timing parameters cannot be different within a memory branch.
- DRAM banks are always closed after read or write command by issuing an autoprecharge command.
- Burst length is 4 (bl = 4) when using a two channels per DDR branch. Burst length is 8 (bl = 8) when using a single channel per branch.
- There is a fixed 1 dead cycle for switching commands from one rank on a DIMM to the other rank on the same DIMM.
- Reads, writes, and refreshes across DDR branches have no relationship to each other. They are all independent.

There are four independent DDR branches per CPU chip, each controlled by a separate MCU. Each branch can be configured with one or two channels and supports up to 16 ranks of DIMMs as shown in FIGURE 17-1. Each channel can be populated with up to eight single- or dual-rank FBDs. When a branch is configured with two channels, the two FBDs that share the same AMB ID are accessed in lock-step. Data is returned 144 bits per frame for 8 frames in single channel mode and 288 bits per frame for 4 frames in dual channel mode. In either mode, the total data transfer size is 512 bits, or 64 bytes, the cache line size for the L2 cache.

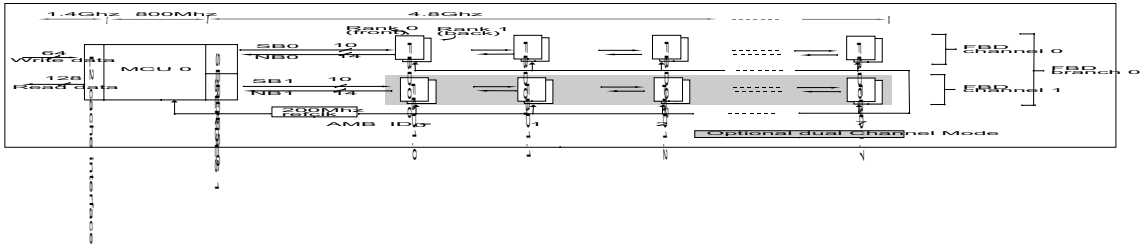


FIGURE 17-1 DDR Branch Configuration

Each FBD contains four or eight internal banks that can be controlled independently. These internal banks are controlled inside the SDRAM chips themselves. Accesses can overlap between different internal banks. In a normal configuration, every read and write operation to SDRAM will generate a burst length of 4 with 16 bytes of data transferred every half memory clock cycle. In single-channel mode, reads and writes will have a burst length of 8 with 8 bytes of data transferred every half memory cycle.

TABLE 17-1 shows the memory organizations supported by UltraSPARC T2. Table 26-2 shows how the MCU should be programmed for the supported DIMM configurations.

TABLE 17-1 UltraSPARC T2 Memory Configurations

DIMM	Base Device	Part	Ranks	# of Devices	Min. Memory per Branch	Max. Memory per Branch
512 MB	256 Mb	x4	1	18	512 MB	8 GB
1 GB	256 Mb	x4	2	36	1 GB	16 GB
1 GB	512 Mb	x4	1	18	1 GB	16 GB
2 GB	512 Mb	x4	2	36	2 GB	32 GB
2 GB	1 Gb	x4	1	18	2 GB	32 GB
4 GB	1 Gb	x4	2	36	4 GB	64 GB
4 GB	2 Gb	x4	1	18	4 GB	64 GB
8 GB	2 Gb	x4	2	36	8 GB	128 GB
512 MB	512 Mb	x8	1	9	512 MB	8 GB
1 GB	512 Mb	x8	2	18	1 GB	16 GB
1 GB	1 Gb	x8	1	9	1 GB	16 GB
2 GB	1 Gb	x8	2	18	2 GB	32 GB
2 GB	2 Gb	x8	1	9	2 GB	32 GB
4 GB	2 Gb	x8	2	18	4 GB	64 GB

TABLE 17-2 MCU programming for supported DIMMs

DIMM	Base Device	Ranks	8 bank mode	RAS Address Width	CAS Addr Width	Stacked
512 MB	256 Mb (64 Mb x4)	1	0	D_{16}	B_{16}	0
1 GB	256 Mb (64 Mb x4)	2	0	D_{16}	B_{16}	1
1 GB	512 Mb (128 Mb x4)	1	0	E_{16}	B_{16}	0
2 GB	512 Mb (128 Mb x4)	2	0	E_{16}	B_{16}	1
2 GB	1 Gb (256 Mb x4)	1	1	E_{16}	B_{16}	0
4 GB	1 Gb (256 Mb x4)	2	1	E_{16}	B_{16}	1
4 GB	2 Gb (512 Mb x4)	1	1	F_{16}	B_{16}	0
8 GB	2 Gb (512 Mb x4)	2	1	F_{16}	B_{16}	1
512 MB	512 Mb (64 Mb x8)	1	0	D_{16}	A_{16}	0

DIMM	Base Device	Ranks	8 bank mode	RAS Address Width	CAS Addr Width	Stacked
1 GB	512 Mb (64 Mb x8)	2	0	D_{16}	A_{16}	1
1 GB	1 Gb (128 Mb x8)	1	0	E_{16}	A_{16}	0
2 GB	1 Gb (128 Mb x8)	2	0	E_{16}	A_{16}	1
2 GB	2 Gb (256 Mb x8)	1	1	E_{16}	A_{16}	0
4 GB	2 Gb (256 Mb x8)	2	1	E_{16}	A_{16}	1

17.5 FBD Channel Configuration

The FBD specification supports two southbound channel configurations and five northbound channel configurations. UltraSPARC T2 will support both southbound configurations—the 10-bit and 10-bit failover modes—and two of the northbound configurations, the 14-bit and 14-bit failover modes. These modes support data packets of 64-bit data and 8-bit ECC. The 10-bit southbound mode provides 22 bits of CRC while the 10-bit failover mode has 10 bits of CRC. The 14-bit northbound mode provides 24 bits of CRC on read data (12 bits per 72-bit data packet), and the 14-bit failover mode provides 12 bits of CRC (6 bits per 72-bit data packet).

During channel initialization, software will determine if a channel can be fully utilized (10-bit southbound or 14-bit northbound mode) or if a failover mode must be used in which one of the bit lanes is muxed out.

17.5.1 FBD Channel Initialization

There are two ways to initialize the FBD channels. The first way uses a hardware state machine for initialization. This initialization triggered by writing to the Channel Reset register. In addition to a hardware state machine, the FBD channels can also be initialized through a software interface. This allows more flexibility in the initialization over the dedicated hardware state machine. Software must perform the following sequence of events to initialize an FBD channel:

1. Because the SerDes PLLs are enabled before the TXBCLKIN is stable, the transmit FIFO which tracks the phase drift between TXBCLKIN and the SerDes internal clock may be too far off center when the MCU begins transmitting data. In order to re-center this FIFO, software must toggle the TX_ENFTP bit in the SERDES_CONFIG_BUS_REG from 0 to 1 and back to 0.

2. Drive Electrical Idle on the SB channel's TX outputs by setting the Channel State register to disable. Channels must remain in `Disable` state for at least tDisable (51 frames) before transitioning to `Calibrate` state.

3. To transition to `Calibrate` state, set Channel State register to calibrate for longer than twice tClkTrain time (42 frames). Once the AMBs are in the `Calibrate` state, they must remain in this state for at least tCalibrate time (480K frames).

4. Drive Electrical Idle on SB channel to transition AMBs to `Disable` state. Remain in `Disable` state for at least tDisable time (51 frames).

5. Set the Channel State register to training to begin driving TS0 patterns on the SB channel to transition the AMBs to the `Training` state. The TS0 patterns are sent to the last AMB until TS0 patterns are received on the northbound channel with the AMB ID from the last AMB. Software will use the Training State Loopback registers to determine how many correct TS0 patterns have been received on the northbound channel. This training requires approximately 275 frames with eight DIMMs per channel. After several correct TS0 patterns have been received on 13 of 14 of the bit lanes, initialization can proceed to step 5.

6. Set the Channel State register to testing to begin driving TS1 patterns on the SB channel to transition the AMBs to the `Testing` state. The IBIST engine within the MCU will take over after the TS1 header has been sent, and it will signal the MCU upon its completion so the MCU can send the trailer and begin the next training sequence. After several TS1 patterns with the AMB ID of the last AMB have been received correctly, and software/IBIST has determined that at least 9 southbound and 13 northbound bit lanes are working, initialization can proceed to step 6.

7. Set the Channel State register to polling to begin driving TS2 patterns on the SB channel to transition the AMBs to the `Polling` state. Continue sending TS2 patterns to the last AMB until correct TS2 patterns are received on the NB channel. This determines the read round-trip delay for the channel. TS2 patterns can be sent to intermediate AMBs to determine which channel protocols they support and to check that they can properly merge their data into the NB data stream. AMBs that are not able to merge their data into the NB data stream correctly will assert their data_merge_error status bit. Once initialization reaches the L0 state, software can check these bits by using AMB Configuration register read commands to determine how to adjust the COMMAND_TO_DATA_INCR registers in the AMBs to increase the channel latency.

8. Set the Channel State register to config to begin driving TS3 patterns on the SB channel to transition the AMBs to the `Config` state. The TS3 patterns program the configuration of the SB and NB channels (always 10 SB and 14 SB for UltraSPARC T2) and which channel bits are muxed out if using a failover mode. TS3 patterns are issued until the patterns are correctly received on the NB channel.

9. Set the Channel State register to l0 to transition AMBs to `L0` state. After four consecutive NOPs have been sent on the SB channel, the channel is ready to accept channel and DRAM commands.

17.5.2 Interconnect BIST (IBIST)

Interconnect BIST provides a mechanism for system level testing of the FBD channel connections. The memory controller has IBIST transmit and receive engines, each with a pattern generater. The transmit engine sends patterns to an AMB on the southbound channel. The AMB loops the patterns back to the receive engine which checks the patterns for errors.

IBIST is optionally run during the TS1 stage of the FBD channel initialization sequence. The following sequence is used to run IBIST.

1. Set AMBID in FBD_CHANNEL_STATE_REG to the target AMB.
2. Set channel to be test in TS1_SB_NB_MAPPING_REG.
3. Set SBTS0CNT field in SBFIBINIT_REG and TRAINING_STATE_MIN_TIME_REG such that SBTS0CNT * 12 is greater than TRAINING_STATE_MIN_TIME_REG + 240.
4. Start RX engine by setting start bit in NBFIBPORTCTL_REG.
5. Start TX engine by setting start bit in SBFIBPORTCTL_REG.

After starting the TX engine, the memory controller will sequence the initialization state machine to the TS0 state and then to the TS1 state, buring which IBIST will be run. After IBIST completes, the memory controller will continue sequencing to the TS2 and TS3 states and complete in L0 state. Error status for the IBIST run will be logged in the NBFIBPORTCTL_REG.

17.6 AMB Initialization

1. 1.5V, 1.8V, and 3.3V power supplies come up:
   - RESET# asserted low while power supplies are coming up.
   - CKEs are low upon 1.8V power-up.
2. BIOS queries SPD on all the FBDs on the channel to determine operating conditions:
   - Channel frequency, compatible DIMMs, DRAM, and AMB parameters
3. Clocks up and stable at required frequency.
   - Reference clocks should be stable for at least 1 ms before RESET# deasserted.
   - DRAM clocks (CLK/CLK) may be toggling at this time.
4. RESET# deasserted high.

- CKEs to DRAMs remain low.

5. No in-band or SMBus transactions for at least 2 ms after RESET# deasserted.
6. AMB parameters critical for robust link initialization are programmed via SMBus.
   - Architected link registers:
     i. LINKPARNXT: link frequency. **Note:** some AMBs may use this write to trigger PLL init. After writing to LINKPARNXT, 200 µs is required prior to any in-band activity. Note: It is generally satisfied by additional SMBus activity.
     ii. FBDSBCFGNXT: SB transmitter drive strength, de-emphasis setting, and pass-through mode
     iii. FBDNBCFGNXT: NB transmitter drive strength, de-emphasis setting, and pass-through mode
   - Personality bytes from SPD needed for link inititalization
     i. PERSBYTE{5:0}NXT: Remaining Personality bytes are not required for link init and may be loaded over the high speed FBD configuration register accesses
   - These next register values must be transferred to the matching current registers before the FBD link leaves the `Disable` state
     i. Updates may be done right after the NXT register is updated when link is in electrical idle. Updates must be complete before the beginning of training.
7. FBD link is initialized including `Calibration` state.
   - Refer to *FBD Channel Configuration*" for initialization sequence.
8. Remaining AMB configuration is loaded over high speed FBD channel
   - CMD2DATA, remaining Personality bytes, other SPD parameters, DRAM parameters (MTR, DRT, DRC, etc.), Errors enabled, etc.
   - These AMB registers are loaded through the MCU Configuration register Access Address and Data registers
9. FBD Link goes through fast reset (no calibration) to establish the desired configuration.
   - DRAM clocks should be stable at this time (that is, after link train).
10. DRAM interface can now be established
    a. MRS/EMRS set up and DRAM initialization sequence using DCALCSR and DCALADDR.
       i. Refer to *Memory Initialization* for DRAM initialization sequence

ii. The DCALCSR and DCALADDR registers are accessed through the MCU Configuration register Access Address and Data registers. The address associated with the DRAM command is placed in the DCALADDR register. The command is programmed in the DCALCSR with bit 31 set to initiate the command. Software then polls this register until bit 31 is reset, indicating that the command has completed.

b. DRAM interface calibrated using DCALCSR.

c. Optionally, Membist functionality can be used to test the DRAMs.

d. DRAMs can be initialized with MemBist.

e. AMB autorefresh engine is enabled at this time.

11. Refresh must now be transferred to the host.
    - Option 1:
        i. Use fast reset on the link with DRAMs in self-refresh.
        ii. Clear DSREFTC.dissrexit to enable fast self-refresh exit when link is re-established
        iii. Put the link in disable state which automatically puts the DRAMs in self-refresh.
        iv. Start the refresh engine on the host.
        v. Bring up the link again.
        vi. Host starts sending refresh commands as soon as L0 state is reached.
    - Option 2:
        i. Write control register to disable autorefresh engine followed by
        ii. Clear DAREFTC.arefen to turn off autorefresh.
        iii. Host then immediately takes over sending refresh commands.
12. Host now has complete control of the FBD Channel.

17.7 Memory Initialization

The power-up sequence for the SDRAMs is same as the JEDEC specification (JC 42.3). Software controls the sequence through the DDR Calibration Control and Status register and the DDR Calibration Address register within the AMB. These are accessed with AMB Configuration register reads and writes using the MCU's Configuration Register Access Address and Data registers.

17.7.1 Power On

Apply power, and maintain CKE below 0.2 * VDDQ and ODT at a low state. All other inputs may be undefined. Once RESETn is asserted to the AMB, it will drive CKE low.

17.7.2 Clocks Stable

Clocks to the DIMMs start as soon as power is enabled to the AMB. The clocks should be stable for at least 1 ms before RESETn is deasserted.

17.7.3 Assert CKE

Software has to write to the DRAM Controller Mode register within the AMB to enable CKEs to the DRAMs.

17.7.4 Software Configuration

Software needs to set all the DRAM configuration registers within the AMB to the desired value.

17.7.5 Pause for 200 µs

DDR SDRAM initialization requires a 200 µs wait after clock has been stable. The memory controller counts a fixed number of cycles to pause for this time.

17.7.6 Pause 400 ns

Controller waits for 400 ns before issuing precharge all command.

17.7.7 `precharge_all` Command

Controller issue a `precharge_all` command for all banks of the device and for all DIMMs present. This is done with a single command to all the DIMMs.

17.7.8 Issue EMRS(2) Write Command

Even though these registers contain no useful information, it is required that a write to this register be performed. The default data of 0 is written by controller.

17.7.9 Issue EMRS(3) Write Command

Even though these registers contain no useful information, it is required that a write to this register be performed. The default data of 0 is written by controller.

17.7.10 Issue EMRS(1) write command to enable DLL

Memory controller issues the extended mode register set command for DLL enable.

17.7.11 Reset DLL

Controller issues mode register set command for DLL reset.

17.7.12 `precharge_all` command

`precharge_all` command is issued for all banks of the devices and for all DIMMs present by the controller.

17.7.13 Two Auto Refresh Cycles

JEDEC specifies to issue two or more autorefresh commands, and the memory controller issues two autorefresh commands.

17.7.14 Set Mode Register to Configure the Device

Mode register command to initialize device operation is issued by the memory controller.

17.7.15 200 Cycles After DLL Reset, Set OCD Default Command

Controller counts the number of cycles from the DLL reset to this command and if it meets 200 cycle count, OCD default command is issued to the DIMMs.

17.7.16 Perform OCD Calibration

Software must read the ocd_sense_pull[up | down] CSRs within the AMBs and adjust each DRAM device accordingly. The encoded commands for the adjustments are written into the AMBs' write data FIFOs. When the OCD adjust command is issued, the data from the FIFOs is driven to the DRAMs.

The following pseudocode can be used to adjust the impedance of the SDRAM lines.

```
foreach (DRAM rank) {
  foreach (pullup, pulldown) {
    reset impedance strength;
    sense OCD;
    until (all DRAM devices set) {
       adjust each device accordingly;
       sense OCD;
    }
  }
}
```

17.7.17 Set OCD Exit Command

An OCD exit command is required to get out of the OCD calibration mode and is issued by the controller.

Note | To guarantee ODT off, vref must be valid and a low level must be applied to the ODT pin.

17.7.18 Initialization Complete

After the above step is performed, the DDR2 SDRAM initialization is finished, normal memory accesses are now allowed.

17.8 RAS Feature Overview

17.8.1 ECC and Extended ECC

The data sent to the DRAMs is protected by SEC-DED error correction. Galois field multiplication techniques are used to generate 16 bits of ECC for each 128 bits of data.

Extended ECC (failover) is a feature of the ECC bits, where they contain enough information to correct any nibble within a 128-bit word. If a single x4 SDRAM component fails, the SDRAM, which would normally hold parity information for double error detection, will be reused for normal data or single error correction information. Thus, when in failover mode, single-bit errors can still be detected and corrected, but multiple errors can no longer be detected. The DRAM Fail-Over Status register and the DRAM Fail-Over Mask register control the operation of the failover mode. Extended ECC is not supported in single channel mode.

Appendix A of the UltraSPARC T2 Programmer's Reference Manual discusses the ECC and Extended ECC algorithms for the UltraSPARC T2 MCU.

17.8.2 Memory Scrubbing

Memory scrubbing refers to the regeneration of ECC for data in memory and the correction of single-bit errors and detection of double-bit errors. When scrubbing is enabled through the DRAM Scrub Enable register described in *DRAM Scrub Enable Register* on page 374, at the end of the time interval defined by the DRAM Scrub Frequency register described in *DRAM Scrub Frequency Register* on page 372, a memory scrub request is issued to the DIMMs. The scrubbing requests have priority over L2 cache requests. First, a scrubbing read request is issued to the DIMMs, and L2 requests to the same bank as the scrub request are blocked. When the scrubbing read data returns, the error detection and correction logic is used on the data. ECC is regenerated and compared with the ECC data read from memory. If an error is detected, subsequent L2 cache transactions are halted and a single-bit or double-bit error is flagged in the DRAM Error Status register as well as being signaled to the L2 cache; then the MCU generates additional requests to the SDRAMs to collect more information on the error, as detailed in the following section. After the scrubbing transaction completes, the L2 cache requests are able to proceed.

Once a scrubbing request is sent, the time interval counter is reset and begins counting down again, and the scrub address is incremented to the next memory location.

Implementation Note There can only be one outstanding scrub command to the DIMMs. So, having a very low number in scrub frequency register does not issue a lot of scrubs.

17.8.3 ECC Error Handling

When an error occurs on a scrub read or an L2 cache read request, the MCU will flag the error to the L2 and log its error in the MCU ESR. Then it will try to determine if the error is a hard error or a transient error. When the error occurs, subsequent L2 requests will be blocked. The MCU will perform the first retry read and log its ECC status in the Retry Status register. If the first retry read does not have an uncorrectable error, the corrected read data is written back to the SDRAM, and a second retry read will be issued and its status will also be logged. Only the status from the original read will be sent to the L2 cache and logged in the ESR. One of the virtual processors must perform a register read to check the status of the subsequent reads. Full details on error handling can be found in *L2 Cache Error Descriptions* on page 256 and *L2 Error Registers* on page 276.

17.8.4 Data Poisoning

Data poisoning involves marking known corrupt data in memory with bad ECC so that any later access will get an ECC error. MCU memory poisoning is performed by flipping ECC check bits 15, 9, 5, and 0. This will generate a failing syndrome of 8221_{16} which, when encountered on a read, will most likely indicate poisoned data.

17.9 Access to Nonexistent Memory

Load accesses from nonexistent memory will take a *data_access_error* trap. Instruction fetches from nonexistent memory will take an *instruction_access_error* trap. Store accesses to nonexistent memory will be silently discarded by the system.

Please refer to TABLE 17-3 for out-of-bound address ranges for different memory configurations.

TABLE 17-3 Out-of-Bound Address Ranges for Different Memory Configurations

DIMMs per Channel	DIMM Capacity	Ranks	Dual Channel Memory	Dual Channel Out-of-Bound	Single Channel Memory	Single Channel Out-of-Bound
1	256 Mb × 4	1	1 GB	PA{39:32}	512 MB	PA{39:31}
2	256 Mb × 4	1	2 GB	PA{39:33}	1 GB	PA{39:32}
4	256 Mb × 4	1	4 GB	PA{39:34}	2 GB	PA{39:33}
8	256 Mb × 4	1	8 GB	PA{39:35}	4 GB	PA{39:34}
1	256 Mb × 4	2	2 GB	PA{39:33}	1 GB	PA{39:32}
2	256 Mb × 4	2	4 GB	PA{39:34}	2 GB	PA{39:33}
4	256 Mb × 4	2	8 GB	PA{39:35}	4 GB	PA{39:34}
8	256 Mb × 4	2	16 GB	PA{39:36}	8 GB	PA{39:35}
1	512 Mb × 4	1	2 GB	PA{39:33}	1 GB	PA{39:32}
2	512 Mb × 4	1	4 GB	PA{39:34}	2 GB	PA{39:33}
4	512 Mb × 4	1	8 GB	PA{39:35}	4 GB	PA{39:34}
8	512 Mb × 4	1	16 GB	PA{39:36}	8 GB	PA{39:35}
1	512 Mb × 4	2	4 GB	PA{39:34}	2 GB	PA{39:33}
2	512 Mb × 4	2	8 GB	PA{39:35}	4 GB	PA{39:34}
4	512 Mb × 4	2	16 GB	PA{39:36}	8 GB	PA{39:35}
8	512 Mb × 4	2	32 GB	PA{39:37}	16 GB	PA{39:36}
1	1 Gb × 4	1	4 GB	PA{39:34}	2 GB	PA{39:33}
2	1 Gb × 4	1	8 GB	PA{39:35}	4 GB	PA{39:34}
4	1 Gb × 4	1	16 GB	PA{39:36}	8 GB	PA{39:35}
8	1 Gb × 4	1	32 GB	PA{39:37}	16 GB	PA{39:36}
1	1 Gb × 4	2	8 GB	PA{39:35}	4 GB	PA{39:34}
2	1 Gb × 4	2	16 GB	PA{39:36}	8 GB	PA{39:35}
4	1 Gb × 4	2	32 GB	PA{39:37}	16 GB	PA{39:36}
8	1 Gb × 4	2	64 GB	PA{39:38}	32 GB	PA{39:37}
1	2 Gb × 4	1	8 GB	PA{39:35}	4 GB	PA{39:34}
2	2 Gb × 4	1	16 GB	PA{39:36}	8 GB	PA{39:35}
4	2 Gb × 4	1	32 GB	PA{39:37}	16 GB	PA{39:36}
8	2 Gb × 4	1	64 GB	PA{39:38}	32 GB	PA{39:37}
1	2 Gb × 4	2	16 GB	PA{39:36}	8 GB	PA{39:35}
2	2 Gb × 4	2	32 GB	PA{39:37}	16 GB	PA{39:36}
4	2 Gb × 4	2	64 GB	PA{39:38}	32 GB	PA{39:37}
8	2 Gb × 4	2	128 GB	PA{39:39}	64 GB	PA{39:38}

17.10 Power Management

The power used by the SDRAMs will be a significant portion of the system power usage. Some high-performance systems may be able to handle the maximum power consumption rates, but low-cost systems may need to limit their power usage due to cooling issues, etc. The power throttling scheme of UltraSPARC T2 limits the number of SDRAM memory access transactions during a specified time period. This is done by counting the number of banks that are opened (that is, activate cycles) during this time. Since all of the write and read transactions of UltraSPARC T2 use auto-precharge, the number of banks opened is equivalent to the number of write and read transactions. If the number of transactions during this time period exceeds a preprogrammed limit, no more memory transactions are dispatched until the time period expires. More details on memory power management can be found in *Memory Access Throttle Control* on page 405.

17.11 DRAM Control and Status Registers

This section describes the control registers and diagnostic access for the DRAM. Each DRAM branch has its own set of control, status, and error registers. Note that each DRAM branch requires that all DIMMs on that branch be of exactly the same kind. There is no requirement on the kind of DIMMs used by different DRAM branches. It is better if the size of physical memory on each branch is equal, but it would still work if the software programs the registers to the lowest value of the branches.

Note: When operating in Partial-Bank mode, the CSRs in the disabled memory controllers will not be accessible because the clocks to these controllers will be turned off to reduce power.

17.11.1 DRAM CAS Address Width Register

TABLE 17-4 shows the format of the DRAM CAS Address Width register.

TABLE 17-4 DRAM CAS Address Width Register – DRAM_CAS_ADDR_WIDTH_REG ($84\ 0000\ 0000_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	width	B_{16}	RW	Defines the number of bits of CAS address width. Any value other than A_{16} or B_{16} will be ignored. Reset only on POR.

17.11.2 DRAM RAS Address Width Register

This register indicates the number of RAS Address bits for a x4 DRAM part of a given density. When x8 parts are being used, this register should be programmed for the x4 part of the same density. The hardware will make the address width correction based on the DRAM CAS Address Width register.

TABLE 17-5 shows the format of the DRAM RAS Address Width register.

TABLE 17-5 DRAM RAS Address Width Register – DRAM_RAS_ADDR_WIDTH_REG ($84\ 0000\ 0008_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	width	F_{16}	RW	Defines the number of bits of RAS address width. Legal values are D_{16} to F_{16}. Values outside this range will be treated as F_{16}. Reset only on POR.

17.11.3 DRAM CAS Latency Register

TABLE 17-6 shows the format of the DRAM RAS Address Width register.

TABLE 17-6 DRAM CAS Address Latency Register – DRAM_CAS_LAT_REG ($84\ 0000\ 0010_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2:0	lat	3_{16}	RW	Defines the CAS latency. 2_{16} = CL of 2; 3_{16} = CL of 3; and so on. Reset only on POR.

17.11.4 DRAM Scrub Frequency Register

TABLE 17-7 shows the format of the DRAM Scrub Frequency register.

TABLE 17-7 DRAM Scrub Frequency Register – DRAM_SCRUB_FREQ_REG ($84\ 0000\ 0018_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	freq	FFF_{16}	RW	Defines how often scrubbing should be performed in terms of DRAM clocks. Reset only on POR. A smaller number causes more frequent scrubbing.

17.11.5 DRAM Refresh Frequency Register

TABLE 17-8 shows the format of the DRAM Refresh Frequency register.

TABLE 17-8 DRAM Refresh Frequency Register – DRAM_REFRESH_FREQ_REG ($84\ 0000\ 0020_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:13	—	0	RO	*Reserved*
12:0	freq	820_{16} (for 266 MHz DIMMs)	RW	Defines how often refresh should be performed in terms of DRAM clocks. It is 2080 (820_{16}) cycles at 266 MHz clock, 2600 ($A28_{16}$) for 333 MHz, and 3120 ($C30_{16}$) for 400 MHz. These values are approximately 7.8us. When this register value is written, the refresh counter value is reset. A smaller value generates more frequent refresh requests. This register is only reset on POR.

17.11.6 DRAM Refresh Counter Register

This register is the free-running counter that triggers refresh when it equals DRAM_REFRESH_FREQ_REG. Refreshes are not issued to the DRAMs if the DRAM_DIMM_INIT_REG bit 0 is set to 1.

TABLE 17-9 shows the format of the DRAM Refresh Counter register.

TABLE 17-9 DRAM Refresh Counter Register – DRAM_REFRESH_COUNTER_REG ($84\ 0000\ 0038_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:13	—	0	RO	*Reserved*
12:0	count	0	RW	Free running counter to issue refresh command when it equals Refresh frequency register. This register is reset to 0 when the DRAM Refresh Frequency register is written or when the count equals or exceeds the DRAM Refresh Frequency value. Reset on POR, WMR, or DBR.

17.11.7 DRAM Scrub Enable Register

This register controls whether scrub should happen in background to DRAM.

TABLE 17-10 shows the format of the DRAM Scrub Enable register.

TABLE 17-10 DRAM Scrub Enable Register – DRAM_SCRUB_ENABLE_REG (84 0000 0040_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	enab	0	RW	If 1, scrub is enabled. Reset only on POR.

17.11.8 DRAM RAS to RAS Different Bank Delay Register

TABLE 17-11 shows the format of the DRAM RAS to RAS Different Bank Delay register.

TABLE 17-11 DRAM RAS to RAS Different Bank Delay Register – DRAM_TRRD_REG (84 0000 0080_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	2_{16}	RW	Trrd delay. Reset only on POR.

17.11.9 DRAM RAS to RAS Same Bank Delay Register

TABLE 17-12 shows the format of the DRAM RAS to RAS Same Bank Delay register.

TABLE 17-12 DRAM RAS to RAS Sam Bank Delay Register – DRAM_TRC_REG (84 0000 $0088)_{16}$ (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:5	—	0	RO	*Reserved*
4:0	delay	C_{16}	RW	Trc delay. This is by default equals to tRAS + tRP. Reset only on POR.

17.11.10 DRAM RAS to CAS Delay Register

TABLE 17-13 shows the format of the DRAM RAS to CAS Delay register.

TABLE 17-13 DRAM RAS to CAS Delay Register – DRAM_TRCD_REG (84 0000 0090_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	3_{16}	RW	Trcd delay. Reset only on POR.

17.11.11 DRAM Write to Read CAS Delay Register

TABLE 17-14 shows the format of the DRAM Write to Read CAS Delay register.

TABLE 17-14 DRAM Write to Read CAS Delay Register – DRAM_TWTR_REG (84 0000 0098_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	0_{16}	RW	Twtr delay. Actual delay is this register value + CL – 1 + BL/2 + iWTR. Reset only on POR. DRAM controller supports a total of these values not to exceed F_{16}.

17.11.12 DRAM Read to Write CAS Delay Register

TABLE 17-15 shows the format of the DRAM Read to Write CAS Delay register.

TABLE 17-15 DRAM Read to Write CAS Delay Register – DRAM_TRTW_REG (84 0000 $00A0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	0_{16}	RW	Trtw delay. Actual delay is this register value + BL/2 + 2 tCK. Reset only on POR. DRAM controller supports a total of these values not to exceed $F._{16}$

17.11.13 DRAM Internal Read to Precharge Delay Register

TABLE 17-16 shows the format of the DRAM internal Read to Precharge Delay register.

TABLE 17-16 DRAM internal Read to Precharge Delay Register – DRAM_TRTP_REG (84 0000 00A8$_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2:0	delay	2_{16}	RW	Trtp delay. The minimum value for this register is 2_{16}. If software tries to program a value less than 2_{16}, the register will be loaded with the value 2_{16}. Reset only on POR.

17.11.14 DRAM Active to Precharge Delay Register

TABLE 17-17 shows the format of the DRAM Active to Precharge Delay register.

TABLE 17-17 DRAM Active to Precharge Delay Register – DRAM_TRAS_REG (84 0000 00B0$_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	9_{16}	RW	Tras delay. Reset only on POR.

17.11.15 DRAM Precharge Command Period Register

TABLE 17-18 shows the format of the DRAM Precharge Command Period register.

TABLE 17-18 DRAM Precharge Command Period Register – DRAM_TRP_REG (84 0000 00B8$_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	3_{16}	RW	Trp delay. Reset only on POR.

17.11.16 DRAM Write Recovery Period Register

TABLE 17-19 shows the format of the DRAM Write Recovery Period register.

TABLE 17-19 DRAM Write Recovery Period Register – DRAM_TWR_REG (84 0000 $00C0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	delay	3_{16}	RW	Twr delay. Reset only on POR.

17.11.17 DRAM Autorefresh to Active Period Register

TABLE 17-20 shows the format of the DRAM Autorefresh to Active Period register.

TABLE 17-20 DRAM Autorefresh to Active Period Register – DRAM_TRFC_REG (84 0000 $00C8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:7	—	0	RO	*Reserved*
6:0	delay	27_{16}	RW	Trfc delay (195 ns @ 200MHz). Reset only on POR.

17.11.18 DRAM Mode Register Set Command Period Register

TABLE 17-21 shows the format of the DRAM Mode Register Set Command Period register.

TABLE 17-21 DRAM Mode Register Set Command Period Register – DRAM_TMRD_REG (84 0000 $00D0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	delay	2_{16}	RW	Tmrd delay. Reset only on POR.

17.11.19 DRAM Four-Activate Window Register

TABLE 17-22 shows the format of the DRAM Four-Activate Window Register.

TABLE 17-22 DRAM Four-Activate Window Register – DRAM_FAWIN_REG (84 0000 $00D8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:5	—	0	RO	*Reserved*
4:0	mode	A_{16}	RW	Tfaw. Number of cycles in which four activate commands may be sent to a DIMM. Reset only on POR.

17.11.20 DRAM Internal Write to Read Command Delay Register

TABLE 17-23 shows the format of the DRAM Internal Write to Read Command Delay register.

TABLE 17-23 DRAM Internal Write to Read Command Delay Register – DRAM_TIWTR_REG (84 0000 $00E0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	delay	2_{16}	RW	Tiwtr delay. Reset only on POR.

17.11.21 DRAM Precharge Wait Register During Power Up

This register is no longer used by hardware.

This register controls the wait time before a precharge can be issued during the power up sequence.

TABLE 17-24 shows the format of the DRAM Precharge Wait Register During Power Up.

TABLE 17-24 DRAM Precharge Wait Register – DRAM_PRECHARGE_WAIT_REG (84 0000 $00E8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	bunch	55_{16}	RW	Equivalent to 80 dram cycles, which is 400 ns @ 200 MHz.

17.11.22 DRAM DIMM Stacked Register

TABLE 17-25 shows the format of the DRAM DIMM Stacked register.

TABLE 17-25 DRAM DIMM Stacked Register – DRAM_DIMM_STACK_REG ($84\ 0000\ 0108_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	stack	0_{16}	RW	Set to 1 if DIMMs are stacked, 0 otherwise. Reset only on POR.

17.11.23 DRAM Extended Mode (2) Register

TABLE 17-26 shows the format of the DRAM Extended Mode (2) register.

TABLE 17-26 DRAM Extended Mode (2) Register – DRAM_EXT_WR_MODE2_REG ($84\ 0000\ 0110_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:15	—	0	RO	*Reserved*
14:0	mreg	0_{16}	RW	Default value of Extended mode register (2) defined by JEDEC. Reset only on POR.

17.11.24 DRAM Extended Mode (1) Register

TABLE 17-27 shows the format of the DRAM Extended Mode (1) register.

TABLE 17-27 DRAM Extended Mode (1) Register – DRAM_EXT_WR_MODE1_REG ($84\ 0000\ 0118_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:15	—	0	RO	*Reserved*
14:0	mreg	18_{16}	RW	Extended mode register (1) defined by JEDEC. Reset only on POR.

17.11.25 DRAM Extended Mode (3) Register

TABLE 17-28 shows the format of the DRAM Extended Mode (3) register.

TABLE 17-28 DRAM Extended Mode (3) Register – DRAM_EXT_WR_MODE3_REG ($84\ 0000\ 0120_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:15	—	0	RO	*Reserved*
14:0	mreg	0_{16}	RW	Extended mode register (3) defined by JEDEC. Reset only on POR.

17.11.26 DRAM 8 Bank Mode Register

This register indicates whether a x4 DRAM part of a given density has 4 or 8 internal banks. When x8 parts are being used, this register should be programmed for the x4 part of the same density. The hardware will make the 8 bank mode correction based on the DRAM CAS Address Width register.

Note: When the CAS Address Width register is programmed to A_{16}, the 8 bank mode register will always be read as a 1; however, the hardware will use the value that was written to this register.

TABLE 17-29 shows the format of the DRAM 8 Bank Mode register.

TABLE 17-29 DRAM 8 Bank Mode Register – DRAM_8_BANK_MODE_REG ($84\ 0000\ 0128_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	disable	1_{16}	RW	If 1, the dram devices have 8 banks instead of 4 banks. Reset only on POR.

17.11.27 DRAM Branch Disabled Register

TABLE 17-30 shows the format of the DRAM Branch Disabled register.

TABLE 17-30 DRAM Branch Disabled Register – DRAM_BRANCH_DISABLED_REG ($84\ 0000\ 0138_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	disable	0_{16}	RW	Set to 1 if branch is not present or is disabled. Reset only on POR.

17.11.28 DRAM Select Low Order Address Bits Register

TABLE 17-31 shows the format of the DRAM Select Low Order Address Bits register.

TABLE 17-31 DRAM Select Low Order Address Bits Register – DRAM_SEL_LO_ADDR_BITS_REG ($84\ 0000\ 0140_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	bunch	0	RW	If 1, then bits after the bank are used for stack and rank instead of the top most bits of PA. Reset only on POR.

17.11.29 DRAM Single Channel Mode Register

TABLE 17-32 shows the format of the DRAM Single Channel Mode register.

TABLE 17-32 DRAM Single Channel Mode register – DRAM_SNGL_CHNL_MODE_REG ($84\ 0000\ 0148_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	mode	0	RW	If 1, enables single channel mode for FBD. Burst length becomes 8. Reset only on POR.

17.11.30 DRAM DIMM Initialization Register

TABLE 17-33 shows the format of the DRAM DIMM Initialization register.

TABLE 17-33 DRAM DIMM Initialization Register – DRAM_DIMM_INIT_REG ($84\ 0000\ 01A0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	cke	0	RW	CKE enable to the controller. Set to 1 to assert CKE high to the DIMMS.
0	init	1_{16}	RW	Software must clear this register upon completing the initialization of the MCU and AMB registers and the DRAM initialization sequence.

17.11.31 DRAM Mode Reg Write Status Register

This register is no longer used by hardware.

TABLE 17-34 shows the format of the DRAM Mode Reg Write Status register.

TABLE 17-34 DRAM Mode Reg Write Status Register – DRAM_MODE_WRITE_STATUS_REG ($84\ 0000\ 0208_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	done	0_{16}	RO	1 if mode reg write done, 0 otherwise. Also its reset to 0 if register $1A0_{16}$ is written to. Reset only on POR.

17.11.32 DRAM Initialization Status Register

TABLE 17-35 shows the format of the DRAM Initialization Status register.

TABLE 17-35 DRAM HW Initialization Status Register – DRAM_INIT_STATUS_REG ($84\ 0000\ 0210_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	done	0_{16}	RO	1 if DRAM initialization sequence is done, 0 otherwise. Also its reset to 0 if register $1A0_{16}$ is written to. Reset only on POR.

17.11.33 DRAM DIMMs Present Register

TABLE 17-36 shows the format of the DRAM DIMMs Present register.

TABLE 17-36 DRAM DIMMs Present Register – DRAM_DIMMS_PRESENT_REG ($84\ 0000\ 0218_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	X	RO	*Reserved*
3:0	num	1_{16}	RW	Number of DIMMs per channel. Reset only on POR.

17.11.34 DRAM Failover Status Register

Each DRAM branch has a failover status register which can be set by software to move a nibble of data from a bad chip to replace one of the ECC chips.

TABLE 17-37 shows the format of the DRAM Fail-Over Status register.

TABLE 17-37 DRAM Fail-Over Status Register – DRAM_FAILOVER_STATUS_REG (84 0000 0220_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	en	0_{16}	RW	Set if in fail-over mode. Reset only on POR.

17.11.35 DRAM Failover Mask Register

Each DRAM branch has a failover status mask register which can be set by software to specify which nibble to ignore the data from.

TABLE 17-38 shows the format of the DRAM Fail-Over Mask register.

TABLE 17-38 DRAM Fail-Over Mask Register – DRAM_FAILOVER_MASK_REG (84 0000 0228_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:35	—	0	RO	*Reserved*
34:0	mask	0_{16}	RW	Mask register to shift data. Reset only on POR.

17.11.36 Power Down Mode Register

TABLE 17-39 shows the format of the DRAM Power Down Mode register.

TABLE 17-39 Power Down Mode Register (84 0000 0238_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	enb_hp	0_{16}	RW	Enables the use of DRAM power down mode for power saving.

Note The Power Down Mode register should not be changed during normal system operation. Doing so may cause transactions to be dropped or memory contents to be updated incorrectly.

17.11.37 FBD Channel State Register

TABLE 17-40 shows the format of the FBD Channel State register.

TABLE 17-40 FBD Channel State Register – FBD_CHANNEL_STATE_REG ($84\ 0000\ 0800_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	0_{16}	RO	*Reserved*
7	mdisable	0_{16}	RW	Disable AMB data merging for TS2 patterns.
6:3	ambid	0_{16}	RW	Target AMB for training sequences. Reset only on POR.
2:0	state	0_{16}	RW	State in initialization sequence: 0_{16} = Disable; 1_{16} = Calibrate; 2_{16} = Training; 3_{16} = Testing; 4_{16} = Polling; 5_{16} = Config, 6_{16} = L0.

17.11.38 FBD Fast Reset Flag Register

TABLE 17-41 shows the format of the FBD Fast Reset Flag register.

TABLE 17-41 FBD Fast Reset Flag Register – FBD_FAST_RESET_FLAG_REG ($84\ 0000\ 0808_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0_{16}	RO	*Reserved*
3	sync_ier	0_{16}	RW	Enable use of ier bit in Sync command. ier will be issued in last sync frame before a channel reset.
2:1	sync_r	0_{16}	RW	Indicates which status register will be received from the AMBs.
0	fastreset	0_{16}	RW	Causes MCU to use the Fast Reset sequence for FBDIMM channel initialization. Reset only on POR.

17.11.39 FBD Channel Reset Register

Writing to bit 0 of this register causes the MCU to sequence through the AMB initialization states. If the FBD Fast Reset Flag is set, the Calibration stage will not be included in the initialization.

TABLE 17-42 shows the format of the FBD Channel Reset register.

TABLE 17-42 FBD Channel Reset Register – FBD_CHNL_RESET_REG ($84\ 0000\ 0810_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1	err	0_{16}	RW	Set to 1 if an error occurred during FBD channel initialization.
0	fbdinit	0_{16}	RW	Causes MCU to initialize FBD channels. Reset to 0 when initialization is complete.

17.11.40 TS1 Southbound to Northbound Mapping Register

TABLE 17-43 shows the format of the TS1 Southbound to Northbound Mapping register.

TABLE 17-43 TS1 Southbound to Northbound Mapping Register – TS1_SB_NB_MAPPING_REG ($84\ 0000\ 0818_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:4	—	0_{16}	RO	*Reserved*
3	ibrx_chnl	0_{16}	RW	Indicates which channel for IBIST to check. Reset only on POR.
2:0	mapping	0_{16}	RW	Determines how targeted AMB maps data from SB bit lanes to NB bit lanes. Reset only on POR

17.11.41 TS1 Test Parameter Register

TABLE 17-44 shows the format of the TS1 Test Parameter register.

TABLE 17-44 TS1 Test Parameter Register – TS1_TEST_PARAMETER_REG ($84\ 0000\ 0820_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:24	—	0_{16}	RO	*Reserved*
23:0	param	0_{16}	RW	AMB test parameters for TS1 sequence. Reset only on POR.

17.11.42 TS3 Failover Configuration Register

TABLE 17-45 shows the format of the TS3 Failover Configuration register.

TABLE 17-45 TS3 Failover Configuration Register – TS3_FAILOVER_CONFIG_REG ($84\ 0000\ 0828_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15:12	sbconfig1	F_{16}	RW	Indicates which southbound lanes for channel 1 will be used. Reset only on POR.
11:8	nbconfig1	F_{16}	RW	Indicates which northbound lanes for channel 1 will be used. Reset only on POR.
7:4	sbconfig0	F_{16}	RW	Indicates which southbound lanes for channel 0 will be used. Reset only on POR.
3:0	nbconfig0	F_{16}	RW	Indicates which northbound lanes for channel 0 will be used. Reset only on POR.

17.11.43 Electrical Idle Detected Register

TABLE 17-46 shows the format of the Electrical Idle Detected register.

TABLE 17-46 Electrical Idle Detected Register – ELECTRICAL_IDLE_DETECTED_REG ($84\ 0000\ 0830_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:28	—	0_{16}	RO	*Reserved*
27:14	electidle1	$3FF_{16}$	RO	Electrical Idle detected from bit lanes in channel 1
13:0	electidle0	$3FF_{16}$	RO	Electrical Idle detected from bit lanes in channel 0

17.11.44 Disable State Period Register

TABLE 17-47 shows the format of the Disable State Period register.

TABLE 17-47 Disable State Period Register – DISABLE_STATE_PERIOD_REG ($84\ 0000\ 0838_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:10	—	0_{16}	RO	*Reserved*
9:0	count	FF_{16}	RW	Counter value for the Disable state. Once Disable state is entered, a counter will count to this value. Once the count is reached, the Disable Done bit will be set. Reset only on POR.

17.11.45 Disable State Period Done Register

TABLE 17-48 shows the format of the Disable State Period Done register.

TABLE 17-48 Disable State Period Done Register – DISABLE_STATE_PERIOD_DONE_REG ($84\ 0000\ 0840_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	0_{16}	RO	*Reserved*
0	done	0_{16}	RW	If set, indicates that the counter for the Disable state period has completed counting.

17.11.46 Calibrate State Period Register

TABLE 17-49 shows the format of the Calibrate State Period register.

TABLE 17-49 Calibrate State Period Register – CALIBRATE_STATE_PERIOD_REG $(84\ 0000\ 0848)_{16}$ (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:20	—	0_{16}	RO	*Reserved*
19:0	count	0_{16}	RW	Counter value for the Calibrate state. Once Calibrate state is entered, a counter will count to this value. Once the count is reached, the calibrate_done bit will be set. Reset only on POR.

17.11.47 Calibrate State Period Done Register

TABLE 17-50 shows the format of the Calibrate State Period Done register.

TABLE 17-50 Calibrate State Period Done Register – CALIBRATE_STATE_PERIOD_DONE_REG $(84\ 0000\ 0850_{16})$ (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:1	—	0_{16}	RO	*Reserved*
0	done	0_{16}	RW	If set, indicates that the counter for the Calibrate state period has completed counting.

17.11.48 Training State Minimum Time Register

TABLE 17-51 shows the format of the Training State Minimum Time register.

TABLE 17-51 Training State Minimum Time Register – TRAINING_STATE_MIN_TIME_REG $(84\ 0000\ 0858_{16})$ (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15:0	count	FF_{16}	RW	Minimum number of frames for Training state before starting to check for Done. Reset only on POR.

17.11.49 Training State Done Register

TABLE 17-52 shows the format of the Training State Done register.

TABLE 17-52 Training State Done Register – TRAINING_STATE_DONE_REG ($84\ 0000\ 0860_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1	timeout	0_{16}	RW	Set if MCU does not recognize 16 `Training` state sequences from FBDs before `Training` state timeout period has elapsed.
0	done	0_{16}	RW	Set when MCU has recognized 16 training state sequences from FBDs after minimum `Training` state period has elapsed.

17.11.50 Training State Timeout Register

TABLE 17-53 shows the format of the Recalibration Duration register.

TABLE 17-53 Training State Timeout Register – TRAINING_STATE_TIMEOUT_REG ($84\ 0000\ 0868_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15:0	count	$FFFF_{16}$	RW	Timeout period for `Training` state. Reset only on POR.

17.11.51 Testing State Done Register

TABLE 17-54 shows the format of the Testing State Done register.

TABLE 17-54 Testing State Done Register – TESTING_STATE_DONE_REG ($84\ 0000\ 0870_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1	timeout	0_{16}	RW	Set if MCU does not recognize 16 `Testing` state sequences from FBDs before testing state timeout period has elapsed.
0	done	0_{16}	RW	Set when MCU has recognized 16 `Testing` state sequences from FBDs.

17.11.52 Testing State Timeout Register

TABLE 17-55 shows the format of the Testing State Timeout register.

TABLE 17-55 Testing State Timeout Register – TESTING_STATE_TIMEOUT_REG (84 0000 0878_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	0_{16}	RO	*Reserved*
7:0	count	FF_{16}	RW	Timeout period for `Testing` state. Reset only on POR.

17.11.53 Polling State Done Register

TABLE 17-56 shows the format of the Polling State Done register.

TABLE 17-56 Polling State Done Register – POLLING_STATE_DONE_REG (84 0000 0880_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1	timeout	0_{16}	RW	Set if MCU does not recognize 16 `Polling` state sequences from FBDs before Polling State Timeout period has elapsed.
0	done	0_{16}	RW	Set when MCU has recognized 16 `Polling` state sequences from FBDs.

17.11.54 Polling State Timeout Register

TABLE 17-57 shows the format of the Polling State Timeout register.

TABLE 17-57 Polling State Timeout Register – POLLING_STATE_TIMEOUT_REG (84 0000 0888_{16}) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	0_{16}	RO	*Reserved*
7:0	count	FF_{16}	RW	Timeout period for `Polling` state. Reset only on POR.

17.11.55 Config State Done Register

TABLE 17-58 shows the format of the Config State Done register.

TABLE 17-58 Config State Done Register – CONFIG_STATE_DONE_REG ($84\ 0000\ 0890_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1	timeout	0_{16}	RW	Set if MCU does not recognize 16 `Config` state sequences from FBDs before Config State Timeout period has elapsed.
0	done	0_{16}	RW	Set when MCU has recognized 16 `Config` state sequences from FBDs.

17.11.56 Config State Timeout Register

TABLE 17-59 shows the format of the Config State Timeout register.

TABLE 17-59 Config State Timeout Register – CONFIG_STATE_TIMEOUT_REG ($84\ 0000\ 0898_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	count	FF_{16}	RW	Timeout period for `Config` state. Reset only on POR.

17.11.57 DRAM Per-Rank CKE Register

Writing this register or the `cke` bit of the DIMM Initialization register, $84\ 0000\ 01A0_{16}$, will cause the MCU to send a CKE command to the FBDIMMs. Each bit corresponds to a rank in a fully populated FBDIMM branch. The enable bits for the CKE command are qualified by the number of DIMMS in the branch, whether the DIMMs are stacked and whether the DIMM Initialization register `cke` bit is set.

Memory traffic must be disabled by setting bit 0 of the DIMM Initialization register when using the Per-Rank CKE register because timing checks are not performed between CKE commands and DRAM transactions.

TABLE 17-60 shows the format of the DRAM Per-Rank CKE register.

TABLE 17-60 PER_RANK_CKE_REG (84 0000 08A0$_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15	d7r1	1_{16}	RW	CKE enable for DIMM 7, Rank 1. Reset only on POR.
14	d7r0	1_{16}	RW	CKE enable for DIMM 7, Rank 0. Reset only on POR.
13	d6r1	1_{16}	RW	CKE enable for DIMM 6, Rank 1. Reset only on POR.
12	d6r0	1_{16}	RW	CKE enable for DIMM 6, Rank 0. Reset only on POR.
11	d5r1	1_{16}	RW	CKE enable for DIMM 5, Rank 1. Reset only on POR.
10	d5r0	1_{16}	RW	CKE enable for DIMM 5, Rank 0. Reset only on POR.
9	d4r1	1_{16}	RW	CKE enable for DIMM 4, Rank 1. Reset only on POR.
8	d4r0	1_{16}	RW	CKE enable for DIMM 4, Rank 0. Reset only on POR.
7	d3r1	1_{16}	RW	CKE enable for DIMM 3, Rank 1. Reset only on POR.
6	d3r0	1_{16}	RW	CKE enable for DIMM 3, Rank 0. Reset only on POR.
5	d2r1	1_{16}	RW	CKE enable for DIMM 2, Rank 1. Reset only on POR.
4	d2r0	1_{16}	RW	CKE enable for DIMM 2, Rank 0. Reset only on POR.
3	d1r1	1_{16}	RW	CKE enable for DIMM 1, Rank 1. Reset only on POR.
2	d1r0	1_{16}	RW	CKE enable for DIMM 1, Rank 0. Reset only on POR.
1	d0r1	1_{16}	RW	CKE enable for DIMM 0, Rank 1. Reset only on POR.
0	d0r0	1_{16}	RW	CKE enable for DIMM 0, Rank 0. Reset only on POR.

17.11.58 L0s Duration Register

Note The *enb_hp* bit of this register should always be set to 0. UltraSPARC T2 does not fully support the AMB L0s state, and enabling this feature will cause unrecoverable channel errors.

TABLE 17-61 shows the format of the L0s Duration register.

TABLE 17-61 L0s Duration Register – L0S_DURATION_REG (84 0000 08A8$_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:7	—	0_{16}	RO	*Reserved*
6	enb_hp	0_{16}	RW	Enables use of L0s state.
5:0	count	$2A_{16}$	RW	Determines the number of frames that the branch will be in L0s state. Legal values are 20_{16} to $2A_{16}$. Values below 20_{16} will be treated as 20_{16}, and values above $2A_{16}$ will be treated as $2A_{16}$. Reset only on POR.

17.11.59 Channel Sync Frame Frequency Register

TABLE 17-62 shows the format of the Channel Sync Frame Frequency register.

TABLE 17-62 Channel Sync Frame Frequency Register – CHANNEL_SYNC_FRAME_FREQ_REG ($84\ 0000\ 08B0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:6	—	0_{16}	RO	*Reserved*
5:0	freq	$2A_{16}$	RW	Frequency at which sync frames are sent on channels. Reset only on POR.

17.11.60 Channel Read Latency Register

TABLE 17-63 shows the format of the Channel Read Latency register.

TABLE 17-63 Channel Read Latency Register – CHANNEL_READ_LAT_REG ($84\ 0000\ 08B8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15:8	latency1	FF_{16}	RW	Read latency for channel 1. Determined during `Polling` state.
7:0	latency0	FF_{16}	RW	Read latency for channel 0. Determined during `Polling` state.

17.11.61 Channel Capability Register

TABLE 17-64 shows the format of the Channel Capability register.

TABLE 17-64 Channel Capability Register – CHANNEL_CAPABILITY_REG ($84\ 0000\ 08C0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:10	—	0_{16}	RO	*Reserved*
9:5	capabil1	0_{16}	RO	Channel capabilities for selected AMB in channel 1. Determined during `Polling` state.
4:0	capabil0	0_{16}	RO	Channel capabilities for selected AMB in channel 0. Determined during `Polling` state.

17.11.62 Loopback Mode Control Register

TABLE 17-65 shows the format of the Loopback Mode Control register.

TABLE 17-65 Loopback Mode Control Register – LOOPBACK_MODE_CNTL_REG (84 0000 $08C8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:2	—	0_{16}	RO	*Reserved*
1:0	mode	0_{16}	RW	Branch Loopback Mode. 0 = Loopback Mode disabled; 10 = Place low-order NB data on SB bus; 11 = Place high-order NB data on SB bus

17.11.63 SerDes Configuration Bus Register

TABLE 17-66 shows the format of the SerDes Configuration Bus register.

TABLE 17-66 SerDes Configuration Bus Register – SERDES_CONFIG_BUS_REG (84 0000 $08D0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:30	—	0_{16}	RO	*Reserved*
29:28	rxtx_rate	0_{16}	RW	Receiver/transmitter operating rate. Reset only on POR.
27	tx_cm	0_{16}	RW	Transmitter common mode. Reset only on POR.
26:24	tx_swing	1_{16}	RW	Transmitter output swing. Reset only on POR.
23:20	tx_de	0_{16}	RW	Transmitter de-emphasis. Reset only on POR.
19	tx_enftp	0_{16}	RW	Transmitter enable fixed TXBCLKIN phase. Reset only on POR.
18:16	rx_term	0_{16}	RW	Receiver termination. Reset only on POR.
15	—	0_{16}	RO	*Reserved*
14:12	rx_cdr	0_{16}	RW	Receiver Clock/Data Recovery algorithm. Reset only on POR.
11:8	rx_eq	0_{16}	RW	Receiver adaptive equalizer configuration. Reset only on POR.
7:6	—	0_{16}	RO	*Reserved*
5:2	pll_mpy	6_{16}	RW	PLL multiplier. reset only on POR.
1:0	pll_lb	0_{16}	RW	PLL loopback bandwidth. Reset only on POR.

17.11.64 SerDes Transmitter and Receiver Differential Pair Inversion Register

TABLE 17-67 shows the format of the SerDes Transmitter and Receiver Differential Pair Inversion register.

TABLE 17-67 SerDes Transmitter and Receiver Differential Pair Inversion Register – SERDES_INVPAIR_REG ($84\ 0000\ 08D8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:48	—	0_{16}	RO	*Reserved*
47:42	tx1_invpair	0_{16}	RW	Invert Channel 1 TXPi/TXNi if bit is 1. Reset only on POR.
41:28	tx0_invpair	0_{16}	RW	Invert Channel 0 TXPi/TXNi if bit is 1. Reset only on POR.
27:14	rx1_invpair	0_{16}	RW	Invert Channel 1 RXPi/RXNi if bit is 1. Reset only on POR.
13:0	rx0_invpair	0_{16}	RW	Invert Channel 0 RXPi/RXNi if bit is 1. Reset only on POR.

17.11.65 SerDes Test Configuration Bus Register

TABLE 17-68 shows the format of the SerDes Test Configuration Bus register.

TABLE 17-68 SerDes Test Configuration Bus Register – SERDES_TEST_CONFIG_BUS_REG ($84\ 0000\ 08E0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:32	—	0_{16}	RO	*Reserved*
31	fsr1_tx_entest	0_{16}	RW	Enable testing of FSR1 transmit ports. Reset only on POR.
30	fsr0_tx_entest	0_{16}	RW	Enable testing of FSR0 transmit ports. Reset only on POR.
29	fsr1_rx_entest	0_{16}	RW	Enable testing of FSR1 receive ports. Reset only on POR.
28	fsr0_rx_entest	0_{16}	RW	Enable testing of FSR0 receive ports. Reset only on POR.
27	fsr1_invpatt	0_{16}	RW	FSR1 invert polarity. Reset only on POR.
26:25	fsr1_rate	0_{16}	RW	FSR1 operating rate. Reset only on POR.
24	fsr1_enbspls	0_{16}	RW	FSR1 receiver pulse boundary scan. Reset only on POR.
23	fsr1_enbsrx	0_{16}	RW	FSR1 receiver boundary scan. Reset only on POR.
22	fsr1_enbstx	0_{16}	RW	FSR1 transmitter boundary scan. Reset only on POR.
21:20	fsr1_loopback	0_{16}	RW	FSR1 loopback. Reset only on POR.
19:18	fsr1_clkbyp	0_{16}	RW	FSR1 clock bypass. Reset only on POR.
17	fsr1_enrxpatt	0_{16}	RW	FSR1 enable RX patterns. Reset only on POR.
16	fsr1_entxpatt	0_{16}	RW	FSR1 enable TX patterns. Reset only on POR.
15:14	fsr1_testpatt	3_{16}	RW	FSR1 test patterns. Reset only on POR.
13	fsr0_invpatt	0_{16}	RW	FSR0 invert polarity. Reset only on POR.
12:11	fsr0_rate	0_{16}	RW	FSR0 operating rate. Reset only on POR.
10	fsr0_enbspls	0_{16}	RW	FSR0 receiver pulse boundary scan. Reset only on POR.
9	fsr0_enbsrx	0_{16}	RW	FSR0 receiver boundary scan. Reset only on POR.
8	fsr0_enbstx	0_{16}	RW	FSR0 transmitter boundary scan. Reset only on POR.
7:6	fsr0_loopback	0_{16}	RW	FSR0 loopback. Reset only on POR.

TABLE 17-68 SerDes Test Configuration Bus Register – SERDES_TEST_CONFIG_BUS_REG ($84\ 0000\ 08E0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
5:4	fsr0_clkbyp	0_{16}	RW	FSR0 clock bypass. Reset only on POR.
3	fsr0_enrxpatt	0_{16}	RW	FSR0 enable RX patterns. Reset only on POR.
2	fsr0_entxpatt	0_{16}	RW	FSR0 enable TX patterns. Reset only on POR.
1:0	fsr0_testpatt	3_{16}	RW	FSR0 test patterns. Reset only on POR.

17.11.66 SerDes PLL Status Register

TABLE 17-69 shows the format of the SerDes PLL Status register.

TABLE 17-69 SerDes PLL Status Register – SERDES_PLL_STATUS_REG ($84\ 0000\ 08e8_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:6	—	0_{16}	RO	*Reserved*
5:3	fsr1_stspll	0_{16}	RO	PLL lock status for FSR1 macros
2:0	fsr0_stspll	0_{16}	RO	PLL lock status for FSR0 macros

17.11.67 SerDes Test Status Register

TABLE 17-70 shows the format of the SerDes Test Status register.

TABLE 17-70 SerDes Test Status Register – SERDES_TEST_STATUS_REG ($84\ 0000\ 08F0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:48	—	0_{16}	RO	*Reserved*
47:38	fsr1_tx_testfail	0_{16}	RO	Test status for FSR1 transmit ports
37:28	fsr0_tx_testfail	0_{16}	RO	Test status for FSR0 transmit ports
27:14	fsr1_rx_testfail	0_{16}	RO	Test status for FSR1 receive ports
13:0	fsr0_rx_testfail	0_{16}	RO	Test status for FSR1 receive ports

17.11.68 Configuration Register Access Address Register

TABLE 17-71 shows the format of the Configuration Register Access Address register.

TABLE 17-71 Configuration Register Access Address Register – CONFIG_REG_ACCESS_ADDR_REG ($84\ 0000\ 0900_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0_{16}	RO	*Reserved*
15	channel	0_{16}	RW	Channel of Configuration register access.
14:11	amb	0_{16}	RW	AMB ID of Configuration register access.
10:2	data	0_{16}	RW	Address for Configuration register read or write. Bits 10:8 are function and bits 7:2 are offset within the function
1:0	—	0_{16}	RO	*Reserved*

17.11.69 Configuration Register Access Data Register

TABLE 17-72 shows the format of the Configuration Register Access Data register.

TABLE 17-72 Configuration Register Access Data Register – CONFIG_REG_ACCESS_DATA_REG ($84\ 0000\ 0908_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:32	—	0_{16}	RO	*Reserved*
31:0	data	0_{16}	RW	Data for Configuration register read or write. Writing to this register generates a Configuration register write on the FBD Channel. Reading from this register generates a Configuration register read on the FBD Channel.

17.11.70 AMB IBIST SBFIBPORTCTL and SBFIBPGCTL Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-73 shows the format of the AMB IBIST SBFIBPORTCTL and SBFIBPGCTL registers.

TABLE 17-73 AMB IBIST SBFIBPORTCTL and SBFIBPGCTL Register – SBFIBPORTCTL_SBFIBPGCTL_REG ($84\ 0000\ 0E80_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:56	—	0_{16}	RO	*Reserved*
55	sbfibportctl_ sbnbmap	0_{16}	RW	Loopback mapping bit. Reset only on POR.
54:38	—	0_{16}	RO	*Reserved*

TABLE 17-73 AMB IBIST SBFIBPORTCTL and SBFIBPGCTL Register – SBFIBPORTCTL_SBFIBPGCTL_REG ($84\ 0000\ 0E80_{16}$) (Count 4 Step 4096) *(Continued)*

Bit	Field	Initial Value	R/W	Description
37	sbfibportctl_ autoinvswpen	0_{16}	RW	Auto Inversion sweep enable. Reset only on POR.
36	—	0_{16}	RO	*Reserved*
35	sbfibportctl_ loopcon	0_{16}	RW	Loop forever. Reset only on POR.
34	sbfibportctl_ complete	0_{16}	RW1C	IBIST done flag. Reset only on POR.
33	sbfibportctl_ mstrmd	1_{16}	RO	Master mode enable. Reset only on POR.
32	sbfibportctl_ ibistr	0_{16}	RW	IBIST start. Setting this bit initiates a channel reset in which IBIST is run during Testing stage. Reset only on POR.
31:26	sbfibpgctl_ ovrloopcnt	F_{16}	RW	Overall Loop Count. Reset only on POR.
25:21	sbfibpgctl_ cnstgencnt	0_{16}	RW	Constant generator loop counter. Reset only on POR.
20	sbfibpgctl_ cnstgenset	0_{16}	RW	Constant generator setting. Reset only on POR.
19:13	sbfibpgctl_ modloopcnt	F_{16}	RW	Modulo-N loop counter. Reset only on POR.
12:10	sbfibpgctl_ modperiod	1_{16}	RW	Period of the modulo-*n* counter. Reset only on POR.
9:3	sbfibpgctl_ pattloopcnt	F_{16}	RW	Pattern buffer loop counter. Reset only on POR.
2:0	sbfibpgctl_ ptgenord	0_{16}	RW	Pattern generator order. Reset only on POR.

17.11.71 AMB IBIST SBFIBPATTBUF1 and SBFIBTXMSK Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-75 shows the format of the AMB IBIST SBFIBPATTBUF1 and SBFIBTXMSK registers.

TABLE 17-74 AMB IBIST SBFIBPATTBUF1 and SBFIBTXMSK Register – SBFIBPATTBUF1_SBFIBTXMSK_REG ($84\ 0000\ 0E88_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:56	—	0_{16}	RO	*Reserved*
55:32	sbfibpattbuf1	$2CCFD_{16}$	RW	IBIST pattern buffer. Reset only on POR.
31:10	—	0_{16}	RO	*Reserved*
9:0	sbfibtxmsk	$3FF_{16}$	RW	Selects which channels to enable for testing. Reset only on POR.

17.11.72 AMB IBIST SBFIBTXSHFT Register

This 32-bit register is documented in the FBDIMM AMB Specification.

TABLE 17-75 shows the format of the AMB IBIST SBFIBTXSHFT register.

TABLE 17-75 AMB IBIST SBFIBTXSHFT Register – SBFIBTXSHFT_REG ($84\ 0000\ 0E90_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:10	—	0_{16}	RO	*Reserved*
9:0	sbfibtxshft	$3FF_{16}$	RW	Transmitter Inversion Shift register. Reset only on POR.

17.11.73 AMB IBIST SBFIBPATTBUF2 and SBFIBPATT2EN Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-76 shows the format of the AMB IBIST SBFIBPATTBUF2 and SBFIBPATT2EN registers.

TABLE 17-76 AMB IBIST SBFIBPATTBUF2 and SBFIBPATT2EN Register – SBFIBPATTBUF2_SBFIBPATT2EN_REG ($84\ 0000\ 0EA0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:56	—	0_{16}	RO	*Reserved*
55:32	sbfibpattbuf2	$FD3302_{16}$	RW	IBIST Pattern Buffer 2. Reset only on POR.
31:10	—	0_{16}	RO	*Reserved*
9:0	sbfibpatt2en	0_{16}	RW	IBIST Pattern Buffer 2 enable. Reset only on POR.

17.11.74 AMB IBIST SBFIBINIT and SBIBISTMISC Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-77 shows the format of the AMB IBIST SBFIBINIT and SBIBISTMISC registers.

TABLE 17-77 AMB IBIST SBFIBINIT and SBIBISTMISC Register – SBFIBINIT_SBIBISTMISC_REG ($84\ 0000\ 0EB0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63	—	0_{16}	RO	*Reserved*
62:53	sbts0cnt	$C8_{16}$	RW	Southbound TS0 count. Reset only on POR.
52:45	sbts1cnt	1_{16}	RW	Southbound TS1 count. Reset only on POR.
44:35	sbdiscnt	100_{16}	RW	Southbound disable state count. Reset only on POR.
34	sbcaliben	1_{16}	RW	Southbound calibration enable. Reset only on POR.
33	—	0_{16}	RO	*Reserved*
32	sbfibiniten	0_{16}	RW	IBIST initialization enable. This bit performs no function in the MCU. The IBIST Start bit of the SBFIBPORTCTL register must be written to initiate IBIST. Bit Reset only on POR.
31:24	—	0_{16}	RO	*Reserved*
23:20	ambid	0_{16}	RW	Value of AMB ID field during TS0. Reset only on POR.
19:0	sbibistcalperiod	$61A80_{16}$	RW	Number of cycles to drive 1 during Calibration. Reset only on POR.

17.11.75 AMB IBIST NBFIBPORTCTL and NBFIBPGCTL Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification. For NBFIBPORTCTL, since the MCU will always be a slave on the northbound port, Bits 0, 1, 22, and 23 will not be used by the MCU and will be read only. Bit 1 will be 0_2 and bits 23:22 will be 0_{16}.

TABLE 17-78 shows the format of the AMB IBIST NBFIBPORTCTL and NBFIBPGCTL registers.

TABLE 17-78 AMB IBIST NBFIBPORTCTL and NBFIBPGCTL Register – NBFIBPORTCTL_NBFIBPGCTL_REG (84 0000 $0EC0_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:54	—	0_{16}	RO	*Reserved*
53:44	nbfibportctl_ errcnt	0_{16}	RW1C	Error counter. Reset only on POR.
43:40	nbfibportctl_ errlnnum	0_{16}	RO	Error lane number. Reset only on POR.
39:38	nbfibportctl_ errstat	0_{16}	RW1C	Port error status. Reset only on POR.
37	nbfibportctl_ autoinvswpen	0_{16}	RW	Auto inversion sweep enable. Reset only on POR.
36	nbfibportctl_ stoponerr	0_{16}	RW	Stop IBIST on error. Reset only on POR.
35	nbfibportctl_ loopcon	0_{16}	RW	Loop forever. Reset only on POR.
34	nbfibportctl_ complete	0_{16}	RW1C	IBIST done flag. Reset only on POR.
33	nbfibportctl_ mstrmd	0_{16}	RO	Master mode enable. Reset only on POR.
32	nbfibportctl_ ibistr	0_{16}	RW	IBIST start. Reset only on POR.
31:26	nbfibpgctl_ ovrloopcnt	F_{16}	RW	Overall loop count. Reset only on POR.
25:21	nbfibpgctl_ cnstgencnt	0_{16}	RW	Constant generator loop counter. Reset only on POR.
20	nbfibpgctl_ cnstgenset	0_{16}	RW	Constant generator setting. Reset only on POR.
19:13	nbfibpgct_ modloopcnt	F_{16}	RW	Modulo-*n* loop counter. Reset only on POR.
12:10	nbfibpgctl_ modperiod	1_{16}	RW	Period of modulo-*n* counter. Reset only on POR.
9:3	nbfibpgctl_ pattloopcnt	F_{16}	RW	Pattern buffer loop counter. Reset only on POR.
2:0	nbfibpgctl_ ptgenord	0_{16}	RW	Pattern generation order. Reset only on POR.

17.11.76 AMB IBIST NBFIBPATTBUF1 Register

This 32-bit register is documented in the FBDIMM AMB Specification.

TABLE 17-79 shows the format of the AMB IBIST NBFIBPATTBUF1 register.

TABLE 17-79 AMB IBIST NBFIBPATTBUF1 Register – NBFIBPATTBUF1_REG (84 0000 $0EC8_{16}$) (Count 4 Step 4096)

Field	Bit	Initial Value	R/W	Description
—	63:56	0_{16}	RO	*Reserved*
nbfibpattbuf1	55:32	$2CCFD_{16}$	RW	IBIST pattern buffer. Reset only on POR.
—	31:0	0_{16}	RO	*Reserved*

17.11.77 AMB IBIST NBFIBRXMSK Register

This 32-bit register is documented in the FBDIMM AMB Specification.

TABLE 17-80 shows the format of the AMB IBIST NBFIBRXMSK register.

TABLE 17-80 AMB IBIST NBFIBRXMSK Register – NBFIBRXMSK_REG (84 0000 $0ED0_{16}$) (Count 4 Step 4096)

Field	Bit	Initial Value	R/W	Description
—	64:46	0_{16}	RO	*Reserved*
nbfibrxmsk	45:32	$3FFF_{16}$	RW	NB IBIST receiver mask. Reset only on POR.
—	31:0	0_{16}	RO	*Reserved*

17.11.78 AMB IBIST NBFIBRXSHFT and NBFIBRXLNERR Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-81 shows the format of the AMB IBIST NBFIBRXSHFT and NBFIBRXLNERR register.

TABLE 17-81 AMB IBIST NBFIBRXSHFT and NBFIBRXLNERR Register – NBFIBRXSHFT_NBFIBRXLNERR_REG (84 0000 $0ED8_{16}$) (Count 4 Step 4096)

Field	Bit	Initial Value	R/W	Description
—	64:46	0_{16}	RO	*Reserved*
nbfibrxshft	45:32	$3FFF_{16}$	RW	Receiver Inversion Shift register. Reset only on POR.
—	31:14	0_{16}	RO	*Reserved*
nbfibrxlnerr	13:0	0_{16}	RO	Receiver error status. Reset only on POR.

17.11.79 AMB IBIST NBFIBPATTBUF2 and NBFIBPATT2EN Register

This pair of 32-bit registers are documented in the FBDIMM AMB Specification.

TABLE 17-82 shows the format of the AMB IBIST NBFIBPATTBUF2 and NBFIBPATT2EN register.

TABLE 17-82 AMB IBIST NBFIBPATTBUF2 and NBFIBPATT2EN Register – NBFIBPATTBUF2_NBFIBPATT2EN_REG (84 0000 $0EE0_{16}$) (Count 4 Step 4096)

Field	Bit	Initial Value	R/W	Description
—	63:56	0_{16}	RO	*Reserved*
nbfibpattbuf2	55:32	$FD3302_{16}$	RW	IBIST Pattern Buffer 2. Reset only on POR.
—	31:14	0_{16}	RO	*Reserved*
nbfibpatt2en	13:0	0_{16}	RW	IBIST Pattern Buffer 2 enable. Reset only on POR.

17.11.80 Other DRAM Registers

Other DRAM registers are defined in other chapters, particularly in the Performance Instrumentation chapter, Power Management chapter, and HW Debug chapter.

CHAPTER 18

Power Management

18.1 SPARC Power Management

The UltraSPARC T2 virtual processor contains extensive power management features. Besides clock gating at the core, there is clock gating within functional units down to the L1 clock headers.

As a safety device, UltraSPARC T2 includes an ASI register to enable lower-level clock gating.

The strands on each physical core share the `ASI_SPARC_PWR_MGMT` register. The `ASI_SPARC_PWR_MGMT` register, described in TABLE 18-1, is a hyperprivileged, read-write register located at ASI $4E_{16}$, VA 0_{16}.

TABLE 18-1 `ASI_SPARC_PWR_MGMT` Register Format

Bit	Field	Remarks
63:16	—	*Reserved*
15	dc	If 1, enables power management in data cache.
14	ic	If 1, enables power management in instruction cache.
13	ifu_cmu	If 1, enables power management in instruction cache miss unit.
12	ifu_ftu	If 1, enables power management in instruction fetch logic excluding instruction cache.
11	ifu_ibu	If 1, enables power management in instruction buffers.
10	dec	If 1, enables power management in decoder.
9	pku	If 1, enables power management in pick unit.
8	lsu	If 1, enables power management in load-store unit excluding data cache.
7	exu	If 1, enables power management in integer execution unit.
6	fgu	If 1, enables power management in floating-point and graphics unit.
5	tlu	If 1, enables power management in trap unit.
4	gkt	If 1, enables power management in crossbar interface.

TABLE 18-1 ASI_SPARC_PWR_MGMT Register Format

Bit	Field	Remarks
2	pmu	If 1, enables power management in performance monitor unit.
1	mmu	If 1, enables power management in memory management unit.
0	misc	If 1′ enables power management in remainder of SPARC core.

Reserved bits read as zeroes and are ignored on writes. The POR value of this register is 0. Writes to this register take effect immediately; this register is synchronizing so that all subsequent activity in the core is performed with the new setting of the power management enables when the next instruction is fetched from the thread that issued the write.

18.2 CPU Throttle Control

The CPU_THROTTLE_CTL register, shown in TABLE 18-2, displays the current value of the three CPU_THROTTLE_CTL pins on UltraSPARC T2. The CPU_THROTTLE_CTL pins are controlled by an external agent that does thermal monitoring of the UltraSPARC T2 chip.

CPU throttling is done by limiting instruction issue for each physical core over an eight-cycle window. At the lowest throttle setting of 0, a physical core issues an instruction on all eight cycles of a window. With each increase in the throttle setting, the physical core issues one fewer instruction per eight cycles, until at the maximum throttle setting of 7, the physical core issues only one instruction each eight-cycle window.

TABLE 18-2 Chip CPU Throttle Control – CPU_THROTTLE_CTL ($98\ 0000\ 0828_{16}$)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2:0	throttle	0	RO	Controls the issuing of instructions to the pipeline across an 8-cycle window as follows (I – issue, N – no issue): 0 – I I I I I I I I 1 – I I I I I I I N 2 – I I I N I I I N 3 – I I N I I N I N 4 – I N I N I N I N 5 – N N I N N I N I 6 – N N N I N N N I 7 – N N N N N N N I

18.3 Memory Access Throttle Control

18.3.1 DRAM Open Bank Max Register

The DRAM Open Bank Max register controls the maximum number of banks that can be open across the entire memory. A single register was desired; however, for implementation reasons, a copy of this register is present for each memory controller. All four registers must be set to the same value for the power throttle feature to function properly. Setting the registers to different values will result in undefined behavior.

TABLE 18-3 shows the format of the DRAM Open Bank Max register.

TABLE 18-3 DRAM Open Bank Max Register – DRAM_OPEN_BANK_MAX_REG ($84\ 0000\ 0028_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16:0	freq	$1FFFF_{16}$	RW	Maximum banks open across the entire memory at any given time. Reset only on POR.

18.3.2 DRAM Programmable Time Counter Register

This register controls a programmable time value in which max banks can be open.

TABLE 18-4 shows the format of the DRAM Programmable Time Counter register.

TABLE 18-4 DRAM Programmable Time Counter Register – DRAM_PROG_TIME_CNTR_REG ($84\ 0000\ 0048_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	bunch	$FFFF_{16}$	RW	The time value in DRAM clocks. Reset only on POR.

18.4 DRAM Refresh Asynchronicity

Memory refresh operations draw more current than normal memory operations. To avoid current spikes, and to ease the burden on the system power distribution, the UltraSPARC T2 memory controller allows memory refresh asynchronicity, guaranteeing that all refreshes are at least 100 nsec from any other refresh.

Each DRAM controller has a background refresh timer, programmable to trigger a refresh every *n* cycles. Each time it triggers, the controller will schedule a refresh for all banks within a single rank on its channel. Subsequent refreshes will cycle through the ranks attached to that channel. When a refresh is scheduled, it is a high-priority request, so it is guaranteed to be issued in a relatively small number of cycles.

To get asynchronicity, the refresh counters on the different channel controllers must be started up $n/4$ DRAM cycles apart, after which they will remain neatly out of sync.

CHAPTER **19**

Configuration and Diagnostics Support

19.1 ASI_LSU_CONTROL_REG

Each virtual processor has a hyperprivileged `ASI_LSU_CONTROL_REG` register at ASI 45_{16}, VA{63:0} = 0, which contains fields that control several memory-related hardware functions. Nonprivileged or supervisor access to this register causes a *privileged_action* trap.

The format of the register is shown in TABLE 19-1.

TABLE 19-1 LSU Control Register – ASI_LSU_CONTROL_REG (ASI 45_{16}, VA 0_{16})

Bit	Field	Initial Value	R/W	Description
63:35	—	0	RO	*Reserved*
34:33	mode	0	RW	Watchpoint mode. Controls whether VA, PA, or no addresses are checked for data watchpoints. TABLE 19-2 lists the settings of the mode field.
32:25	bm	0	RW	Data watchpoint byte mask. For virtual or physical address data watchpointing, the ASI_DMMU_VA_WATCHPOINT register contains the virtual or physical address of a 64-bit word to be watched. The 8-bit bm controls which byte(s) within the 64-bit word should be watched. If all 8 bits are cleared, the watchpoint is disabled. If the watchpoint is enabled and a data reference overlaps any of the watched bytes in the watchpoint mask, a *VA_watchpoint* or *PA_watchpoint* trap is generated.
24	re	0	RW	Data watchpoint Read and Write Enable. If re or we is set, a read or write that matches the virtual or physical address in ASI_DMMU_VA_WATCHPOINT and the vm byte masks causes a *VA_watchpoint* or *PA_watchpoint* trap. Both re and we can be set to place a watchpoint for either a read or write access. Atomic operations are considered both a read and a write, and watchpoints for atomics are enabled if re, we, or both are set.
23	we	0	RW	
22:5	—	0	RO	*Reserved*
4	se	0	RW	Speculation enable. If set, loads are predicted to hit in the primary cache, branches are predicted not taken, and the floating-point unit predicts exceptions and runs in a pipelined mode.
3	dm	0	RW	DMMU enable. If cleared, when HPSTATE.hpriv = 0 the DMMU treats all data addresses as real addresses and performs a real-to-physical translation. See *Translation* on page 128 for more details.

TABLE 19-2 Mode Field Settings

Mode Field	Watchpoint Function
00_2	Watchpoint disabled
01_2	
10_2	Match on data physical address
11_2	Match on data virtual address

TABLE 19-3 VA and PA Data Watchpoint Byte Mask Examples

Watchpoint Mask	Address of Bytes Watched 7654 3210
00_{16}	Watchpoint disabled
01_{16}	0000 0001
32_{16}	0011 0010
FF_{16}	1111 1111

TABLE 19-1 LSU Control Register – ASI_LSU_CONTROL_REG (ASI 45_{16}, VA 0_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
2	im	0	RW	IMMU enable. If cleared, when HPSTATE.hpriv = 0 and HPSTATE.red = 0 the IMMU treats all instruction addresses as real addresses and performs a real-to-physical translation. See *Translation* on page 128.
1	dc	0	RW	Dcache enable. If cleared, the primary data cache does not allocate a line on a miss.
0	ic	0	RW	Icache enable. If cleared, the primary instruction cache does not allocate a line on a miss. **Note:** The Dcache and Icache are still kept coherent by UltraSPARC T2 when the dc and ic bits are set to 0. This includes updating the Dcache when stores from a strand hit in the Dcache.

19.2 Watchpoint Support

UltraSPARC T2 supports both instruction VA and data VA/PA watchpoints.

19.2.1 ASI_DMMU_WATCHPOINT

Note | Also see the ASI_LSU_CONTROL_REG definition in *ASI_LSU_CONTROL_REG* on page 407, which contains control bits for data watchpoint support.

Each virtual processor has a hyperprivileged ASI_DMMU_WATCHPOINT register at ASI 58_{16}, VA{63:0} = 38_{16} that is used for controlling the address (or address range if bytes are masked) for data PA and VA watchpoints. The format of the register is shown in TABLE 19-4.

TABLE 19-4 DMMU Watchpoint – ASI_DMMU_WATCHPOINT (ASI 58_{16}, VA $38)_{16}$

Bit	Field	Initial Value	R/W	Description
63:48	va_high	X	RO	Ignores writes, returns value of VA{47} on reads.
47:40	va	X	RW	VA watchpoint address{47:40}.
39:3	vapa	X	RW	VA/PA watchpoint address{39:3}.
2:0	—	0	RO	Bits 2:0 of the VA/PA watchpoint address are all zeros, watchpointing is done for 8-byte memory locations.

Nonprivileged and supervisor access to this register causes a *privileged_action* trap.

Notes The VA_watchpoint trap is never generated while executing in hyperprivileged mode. This implies that the AS_IF_USER ASIs will not generate a *VA_watchpoint* trap when accessed in hyperprivileged mode, even though a *VA_watchpoint* trap might be generated when accessed in user mode. The PA_watchpoint trap can be taken in hyperprivileged mode.

For quadword accesses, VA and PA watchpoint checking is only done on the lower 8 bytes of the access. This implies that a *VA_watchpoint* or *PA_watchpoint* trap will only be generated for a quadword load if vapa{3} is set to zero.

For quadword accesses, VA and PA watchpoint checking is only done on the lower 8 bytes of the access. This implies that a *VA_watchpoint* or *PA_watchpoint* trap will only be generated for a quadword load if vapa{3} is set to zero.

Implementation Note The data VA watchpoint comparison is done only on VA{47:3}, and does not mask ASI_DMMU_WATCHPOINT{47:32} when PSTATE.am = 1. This implies that out-of-range accesses to the VA hole will generate a *VA_watchpoint* trap if bits 47:3 and the byte mask match for data. In addition, when PSTATE.am = 1, bits 47:32 in the VA will be zero and ASI_DMMU_WATCHPOINT needs to have bits 47:32 loaded with zeros to produce a data VA watchpoint match.

19.2.2 ASI_IMMU_VA_WATCHPOINT

Each physical core has a pair of hyperprivileged ASI_IMMU_VA_WATCHPOINT register at ASI 50_{16}, VA{63:0} = 38_{16} that are used for controlling the address and enabling of instruction VA watchpoints. Strands 0-3 on the physical core share one of the registers, and strands 4-7 on the physical core share the other register. The format of the register is shown in TABLE 19-5.

TABLE 19-5 IMMU VA Watchpoint – ASI_IMMU_VA_WATCHPOINT (ASI 50_{16}, VA $38)_{16}$

Bit	Field	Initial Value	R/W	Description
63:48	va_high	X	RO	Ignores writes, returns value of VA{47} on reads.
47:2	va	X	RW	VA watchpoint address {47:2}.
1	—	0	RO	*Reserved*
0	enable	0	RW	If 1, instruction VA watchpoint is enabled. If 0, instruction VA watchpoint is disabled.

Nonprivileged and supervisor access to this register causes a *privileged_action* trap.

Notes

The *instruction_VA_watchpoint* trap is never generated while executing in hyperprivileged mode.

No *instruction_VA_watchpoint* trap is generated by UltraSPARC T2 for real-to-physical translations (RA → PA in TABLE 12-15 on page 129).

Implementation Note

The instruction VA watchpoint comparison is done only on VA{47:2}, and does not mask `ASI_IMMU_VA_WATCHPOINT`{47:32} when PSTATE.am = 1. This implies that out-of-range accesses to the VA hole will generate an *instruction_VA_watchpoint* trap if bits 47:2 match. In addition, when PSTATE.am = 1, bits 47:32 in the VA will be zero and `ASI_IMMU_VA_WATCHPOINT` needs to have bits 47:32 loaded with zeros to produce an instruction VA watchpoint match.

19.3 Breakpoint Support

19.3.1 ASI_INST_MASK_REG

Each physical core has a pair of hyperprivileged `ASI_INST_MASK_REG` registers at ASI 42_{16}, VA{63:0} = 8_{16}. Strands 0–3 on the physical core share one of the registers, and strands 4–7 on the physical core share the other register. This register is used to disable execution of a particular instruction or class of instructions for diagnostic or debug purposes, under the control of HPSTATE.ibe. If HPSTATE.ibe = 1 and any of the enable fields are set to 1, execution of any instruction that matches all the enabled bits in the inst field of this register results in an *instruction_breakpoint* trap. Nonprivileged or supervisor access to this register causes a *privileged_action* trap.

TABLE 19-6 defines the format of the `ASI_INST_MASK_REG` register.

TABLE 19-6 SPARC Instruction Mask Register – `ASI_INST_MASK_REG` (ASI 42_{16}, VA 8_{16})

Bit	Field	Initial Value	R/W	Description
63:39	—	0	RO	*Reserved*
38	enb31_30	0	RW	Enable matching on INST 31:30.
37	enb29_25	0	RW	Enable matching on INST 29:25.
36	enb24_19	0	RW	Enable matching on INST 24:19.
35	enb18_14	0	RW	Enable matching on INST 18:14.

TABLE 19-6 SPARC Instruction Mask Register – ASI_INST_MASK_REG (ASI 42_{16}, VA 8_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
34	enb13	0	RW	Enable matching on INST 13.
33	enb12_5	0	RW	Enable matching on INST 12:5.
32	enb4_0	0	RW	Enable matching on INST 4:0.
31:0	inst	0	RW	Instruction pattern to be trapped.

Programming Note While there is a pair of instruction mask registers per physical processor core, each strand is under the control of its own HPSTATE.ibe bit. All HPSTATE.ibe bits for the strands in a physical processor core must be set to have the physical processor core fully trap on the instruction pattern.

19.3.2 Trap on Control Transfer

If PSTATE.tct is set, a taken control transfer will result in a *control_transfer_instruction* trap. Control transfers include conditional branches, jumps, done, and retry. The trap is precise and taken before the instruction executes. Once the trap is taken, PSTATE.tct is cleared. The virtual address of the control transfer instruction will be contained in TPC[TL].

19.4 Instruction and Data Cache Control

19.4.1 ASI_LSU_DIAG_REG

Each physical core has a hyperprivileged ASI_LSU_DIAG_REG register at ASI 42_{16}, VA{63:0} = 10_{16}. This register disables associativity in the primary instruction and/or data caches for diagnostic or debug purposes. Nonprivileged or supervisor access to this register causes a *privileged_action* trap.

TABLE 19-7 defines the format of the ASI_LSU_DIAG_REG register.

TABLE 19-7 LSU Diagnostic Register – ASI_LSU_DIAG_REG (ASI 42_{16}, VA 10_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1	dassocdis	0	RW	If 1, the Dcache replacement algorithm is not LRU but instead uses bits 12:11 of the virtual address to determine the replacement way when a miss occurs
0	iassocdis	0	RW	If 1, the Icache replacement algorithm is not LRU but instead uses bits 13:11 of the virtual address to determine the replacement way when a miss occurs.

19.5 L1 I-Cache Diagnostic Access

19.5.1 ASI_ICACHE_INSTR

Each virtual processor shares a read/write, hyperprivileged ASI_ICACHE_INSTR register at ASI 66_{16}, VA 0_{16}– $7FF8_{16}$ which provides diagnostic access to the Icache data array.

TABLE 19-8 describes the VA format when accessing ASI_ICACHE_INSTR.

TABLE 19-8 ASI_ICACHE_INSTR Address Format

VA Bits	Field	Remarks
63:15	—	*Reserved*
14:12	way{2:0}	Selects cache way to be read or written.
11:6	index{5:0]	Selects cache index to be read or written.
5:3	word	Selects aligned 4 bytes out of 32 bytes in cache line to be read or written. (See below for data dependency on bit 4.)
2:0	—	*Reserved*

Note | The diagnostic VA is shifted left one bit from what would normally be used to index the cache to fetch an instruction (for example, bits 11:3 are used instead of bits 10:2).

The data field is interpreted as shown in TABLE 19-9.

TABLE 19-9 ASI_ICACHE_INSTR Register

Operation Type	Data Bits	Field	Remarks
Store if VA[4]==0	63:33	—	*Reserved*
	32	perrinj	**xor** this bit with generated parity bit
	31:0	instr{31:0}	32 bits of instruction data
Store if VA[4]==1	63:33	—	*Reserved*
	32	instr{0}	bit 0 of instruction data
	31	instr{1}	bit 1 of instruction data
	...	...	...
	1	instr{31}	bit 31 of instruction data
	0	perrinj	**xor** this bit with generated parity bit
Data read if VA[4]==0	63:33	—	*Reserved*
	32	parity	Parity bit read from array
	31:0	instr{31:0}	32 bits of instruction data read from array
Data read if VA[4]==1	63:33	—	*Reserved*
	32	instr{0}	bit 0 of instruction data read from array
	31	instr{1}	bit 1 of instruction data read from array
	...	...	...
	1	instr{31}	bit 31 of instruction data read from array
	0	parity	Parity bit read from array

Note Since hardware does not keep diagnostic stores to the data array consistent with the L2 reverse directory, the L2 cache, and other instruction and data caches, software must flush the L2 cache or otherwise restore system caches to a coherent state after performing a diagnostic store to the data array.

The bit ordering of the data to or from the instruction cache depends on the word address. Specifically, on read or write via this ASI, if bit 4 of the virtual address is set, then the parity and instruction bits (32:0) must be bit-reversed (i.e., bit 0 becomes bit 32, bit 1 becomes bit 31, etc.) to be consistent with functional read and write paths.

19.5.2 ASI_ICACHE_TAG

Each virtual processor has a shared, hyperprivileged, read-write ASI_ICACHE_TAG register located at ASI 67_{16}, VA 0_{16}–$FFE0_{16}$, which provides diagnostic access to the Icache tags and valid bit array. Diagnostic reads of the tag and valid bit arrays do not cause parity errors. The VA format is shown in TABLE 19-10.

TABLE 19-10 ASI_ICACHE_TAG Address Format

VA Bits	Field	Remarks
63:17	—	*Reserved*
16	perren	**xor**s this value with the computed tag parity bit (even parity is computed and stored; ignored on a read)
15	vb_err_en	Forces the slave copy of the valid bit to be the inversion of the master copy (ignored on a read)
14:12	way{2:0}	Identifies the way whose tag and valid bit will be accessed
11:6	index{5:0}	Identifies the index whose tag and valid bit will be accessed
5:0	—	*Reserved*

Note The diagnostic VA is shifted left one bit from what would normally be used to index the cache to fetch an instruction (for example, bits 11:3 are used instead of bits 10:2).

The data format is shown in TABLE 19-11.

TABLE 19-11 ASI_ICACHE_TAG Register

Operation Type	Data bits	Field	Remarks
Store	63:31	—	*Reserved*
	30:2	tag{39:11}	Contents of the tag array.
	1	valid{1}	valid bit master copy.
	0	—	*Reserved*. valid bit slave copy (ignored on write).
Load	63:32	—	*Reserved*
	31	Tag Parity	parity bit of the tag entry.
	30:2	tag{39:11}	Contents of the tag array.
	1	valid{1}	valid bit master copy.
	0	valid{0}	valid bit slave copy.

Notes Since hardware does not keep diagnostic stores to the tag and valid bit arrays consistent with the L2 reverse directory, the L2 cache, and other instruction and data caches, software must flush the L2 cache or otherwise restore system caches to a coherent state after performing a diagnostic store to the L1 tags and valid bit array.

The Icache Tag includes physical address bit 39, even though in normal system operation bit 39 will always be 0 for any cacheable access.

19.6 L1 D-Cache Diagnostic Access

19.6.1 ASI_DCACHE_DATA

Each virtual processor has a shared, hyperprivileged, read-write ASI_DCACHE_DATA register located at ASI 46_{16}, VA 0_{16}–$1FFFF8_{16}$ that is used for diagnostic accesses to the L1 data cache data array. Stores update 64 bits of the data array at one time and can optionally invert each computed byte parity bit independently. Hardware computes and stores even parity. Loads specify whether to load parity or data bits. Loads that specify parity bits load all eight byte parity bits at a time. Loads that specify data bits load 64 data bits at a time.

TABLE 19-12 describes the VA format for a store to ASI_DCACHE_DATA.

TABLE 19-12 ASI_DCACHE_DATA Store Address Format

VA Bits	Field	Remarks
63:21	—	*Reserved*
20:13	perrmask{7:0}	**xor**s this value with computed parity bits; bit 7 corresponds to parity for bits 63:56, bit 6 to bits 55:48, etc.
12:11	way{1:0}	Selects cache way to be written.
10:4	Index{6:0}	Selects cache index to be written.
3	doubleword	Selects aligned 8 bytes out of 16 bytes in cache line to be written.
2:0	—	*Reserved*

For a load to ASI_DCACHE_DATA, the VA is interpreted as follows.

TABLE 19-13 ASI_DCACHE_DATA Load Address Format

VA Bits	Field	Remarks
63:14	—	*Reserved*
13	data_notparity	If 1, selects data field to be returned by load; if 0, selects parity bits to be returned by load.
12:11	way{1:0}	Selects cache way to be written.
10:4	index{6:0}	Selects cache index to be read.
3	doubleword	Selects aligned 8 bytes out of 16 bytes in cache line to be read.
2:0	—	*Reserved*

The data field is interpreted as follows.

TABLE 19-14 ASI_DCACHE_DATA Format

Operation Type	Data Bits	Field	Remarks
Store	63:0	data{63:0}	Contains the 64-bit data field to be written to the data array.
Data read	63:0	data{63:0}	Contains the 64-bit data field read from the data array.
Parity read	63:8	—	*Reserved.* Read as all zeroes.
	7:0	parity{7:0}	Contains the 8 parity bits read from the data array.

Note Since hardware does not keep diagnostic stores to the data array consistent with the L2 reverse directory, the L2 cache, and instruction and other data caches, software must flush the L2 cache or otherwise restore system caches to a coherent state after performing a diagnostic store to the data array.

19.6.2 ASI_DCACHE_TAG

Each virtual processor has a shared, read/write, hyperprivileged ASI_DCACHE_TAG register located at ASI 47_{16}, VA 0_{16}–$7FF0_{16}$. This register reads and writes the contents of the Dcache valid bit and tag arrays for diagnostic purposes.

The VA field is interpreted as follows.

TABLE 19-15 ASI_DCACHE_TAG Address Format

VA Bits	Field	Remarks
14	vb_err_en	Forces the slave copy of the valid bit to be the inversion of the master copy (ignored on a read).
13	perren	**xor**s this value with the computed tag parity bit (even parity is computed and stored; ignored on a read).
12:11	way{1:0}	Identifies the way whose tag and valid bit will be accessed.
10:4	index{6:0}	Identifies the index whose tag and valid bit will be accessed.
3:0	—	*Reserved*

The data field is interpreted as follows.

TABLE 19-16 ASI_DCACHE_TAG Format

Operation Type	Data Bits	Field	Remarks
Store	63:31	—	*Reserved*
	30:2	tag{39:11}	Contents of the tag array.
	1	valid{1}	valid bit master copy.
	0	—	*Reserved.* valid bit slave copy (ignored on write).
Load	63:32	—	*Reserved*
	31	Tag Parity	Parity bit of the tag entry.
	30:2	tag{39:11}	Contents of the tag array.
	1	valid{1}	valid bit master copy.
	0	valid{0}	valid bit slave copy.

Loads from this ASI do not cause parity errors if the stored parity is bad or if the valid bit copies differ from each other.

Note | Since hardware does not keep diagnostic stores to the tag and valid bit arrays consistent with the L2 reverse directory, the L2 cache, and instruction and other data caches, software must flush the L2 cache or otherwise restore system caches to a coherent state after performing a diagnostic store to the L1 tags and valid bit array.

19.7 Integer Register File

To read or write the IRF data portions, normal SPARC integer loads and stores or arithmetic instructions can be used. Instructions that read the IRF data may result in ECC errors. These ECC errors can be suppressed by setting CERER.irf or SETER.pscce to 0.

To read the ECC check bits for the IRF, software can load from the ASI_IRF_ECC_REG, described below.

19.7.1 ASI_IRF_ECC_REG

Each virtual processor has a hyperprivileged, read-only ASI_IRF_ECC_REG located at ASI 48_{16}, VA 0_{16}–$F8_{16}$. A load to this register with VA bits 7:3 set to the 5-bit register number to be read returns the ECC bits of the IRF register entry accessed by VA bits 4:0, in the current register window. Nonprivileged or supervisor access to this register causes a *privileged_action* trap. Writes to this register cause a *DAE_invalid_ASI*.

The format of the ASI_IRF_ECC_REG address is as follows.

TABLE 19-17 ASI_IRF_ECC_REG address

Bit	Field	Remarks
63:8	—	*Reserved*. Ignored
7:3	irf_index	Integer register number to be read. Note this should be the same as the index contained in the DSFARwhen an error occurs. See Table 16-11 on page 248 for details.
2:0	—	*Reserved*

When a read to this register occurs, the format of the returned data is as follows.

TABLE 19-18 ASI_IRF_ECC_REG format

Bit	Field	Remarks
63:8	—	*Reserved.* Reads return zeroes.
7:0	ecc	ECC bits read from IRF.

A read of this register never results in an ECC error.

19.8 Floating-Point Register File

To read or write the FRF data portions, normal SPARC FGU loads and stores or arithmetic instructions can be used. Instructions that read the FRF data may result in ECC errors. These ECC errors can be suppressed by setting CERER.frf or CETER.pscce to 0.

To read the ECC check bits for the FRF, software can load from the ASI_FRF_ECC_REG, described below.

19.8.1 ASI_FRF_ECC_REG

Each virtual processor has a hyperprivileged, read-only ASI_FRF_ECC_REG located at ASI 49_{16}, VA 0_{16}–$F8_{16}$. A load to this register with VA bits 7:3 set to the double-precision 6-bit register number returns the even and odd ecc bits for the even and odd halves of the FRF register entry accessed by VA bits 7:3. Nonprivileged or supervisor access to this register causes a *privileged_action* trap. Writes to this register cause a *DAE_invalid_ASI*. The format of the ASI_FRF_ECC_REG is as follows.

TABLE 19-19 ASI_FRF_ECC_REG Address

Bit	Field	Remarks
63:8	—	*Reserved*. Ignored
7:3	frf_index{4:0}	Index in the FRF register file to be read. The contents of this field specify the precise index of the FRF to be read, not the encoded SPARC V7 FRF register number; e.g., an FRF_index{4:0} value of 10110_2 will read FRF index 22.
2:0	—	*Reserved*. Ignored.

When a read to this register occurs, the format of the returned data is as follows.

TABLE 19-20 ASI_FRF_ECC_REG Format

Bit	Field	Remarks
63:14	—	*Reserved*. Reads return zeroes.
13:7	ecc_even	ECC bits read from FRF for even half of register.
6:0	ecc_odd	ECC bits read from FRF for odd half of register.

A read of this register never results in an ECC error.

19.9 Store Buffer — ASI_STB_ACCESS

Diagnostic access to the store buffer is complicated by the width of each entry and the fact that all stores (even diagnostic ASI stores) are written to the store buffer.

Each strand has a hyperprivileged, read-only ASI_STB_ACCESS register located at ASI $4A_{16}$, VA 0_{16}–$3F_{16}$. Diagnostic software can read any entry (belonging to its own strand) by using the ASI_STB_ACCESS register.

The store buffer is divided into three parts. The cam field contains bits 39:3 of the physical address, 8 byte marks, and a parity bit. The data array contains two subfields. The first field consists of a 64-bit data field and two 7-bit SEC/DED ecc fields (ECC is stored on a 32-bit word basis). The second field of the data array consists of encoded privilege-level information (encoded with hamming [?] distance of 2 between valid codes for a total of 3 bits). Finally, as will be explained shortly, there is a read that returns the current store buffer pointer for the thread.

To read a store buffer entry, diagnostic software issues a load to the ASI_STB_ACCESS register, specifying the entry and subfield of the store buffer to read in the virtual address of the load, as detailed below.

Software cannot write directly into a store buffer entry. All stores are placed into the store buffer. This includes stores to ASI registers as well as stores to memory and I/O space.

Software can inject an error into the store buffer by setting the stau or stdu bit in the ASI_ERROR_INJECT_REG. Hardware **xor**s the values of the computed ecc bits with ASI_ERROR_INJECT_REG.eccmask{6:0} if stdu is set, or **xor**s the computed parity for the address with ASI_ERROR_INJECT_REG.eccmask{6:0} if stau is set. If multiple bits are set, hardware **xor**s each field independently. Once software sets the enb_hp bit in the ASI_ERROR_INJECT_REG, all future stores will be written to the store buffer with errors injected according to which of the stau or stdu bits is set. Errors cease to be injected once software resets the enb_hp bit. Note that hardware will not inject an error for a store to the ASI_ERROR_INJECT_REG.

Software has control over the contents of any of the store buffer subfields. For example, the contents of the store data field and the type of store determine the settings of the data field data bits and byte mark bits. The privilege level of the store also determine the contents of the privilege level field.

The format of the va field for accesses to ASI_STB_ACCESS is detailed below.

TABLE 19-21 ASI_STB_ACCESS Address Format

Operation Type	Bit	Field	Remarks
Store buffer read	63:9	—	*Reserved*
	8:6	field	Determines subset of store buffer entry to be read as follows: 000 – Store buffer data 001 – Store buffer data ECC (14 bits) 010 – Store buffer control and address parity 011 – Store buffer address and byte marks 100 – Current store buffer pointer 101 – *Reserved* 110 – *Reserved* 111 – *Reserved*
	5:3	entry	Contains store buffer entry ID (of this thread's store buffer) to read.
	2:0	—	*Reserved*

The data field of a load or store to the ASI_STB_ACCESS field is detailed below.

TABLE 19-22 ASI_STB_ACCESS Format

Operation Type	Bit	Field	Remarks
STB data read	63:0	data	Contains data in the store buffer entry addressed by ASI_STB_ACCESS.
STB data ECC read	63:14	—	*Reserved.* Read as all zeroes.
	13:7	ecc_odd	ECC for bits 63:32 of the data field.
	6:0	ecc_even	ECC for bits 31:0 of the data field.
STB control read	63:4	—	*Reserved*. Read as all zeroes.
	3	c_p	Parity over bm_asi and PA{39:3}.
	2:0	control	Privilege level encoding: 000 – user; 011 – priv; 101 – hpriv; all others – parity error).
STB CAM read	63:48	—	Read as all zeroes.
	47:40	bm_asi	Byte marks{7:0} or ASI value.
	39:3	pa{39:3}	Contents of PA cam field.
	2:0	—	*Reserved.* Read as all zeroes.

19.10 Scratchpad Registers

Each strand has a hyperprivileged, read-only ASI_SCRATCHPAD_ACCESS register located at ASI 59_{16}, VA 0_{16}–78_{16}.

Diagnostic software can read the data bits or ecc bits using this register. The format and usage of this register are described below.

TABLE 19-23 ASI_SCRATCHPAD_ACCESS Address Format

Operation Type	Bit	Field	Remarks
Load	63:7	—	*Reserved*
	6	data_np	If 1, return the data entry contents; if 0, return the eight 8 ecc bits in bit positions 7:0 of the returned data field.
	5:3	index	Index in Scratchpad array to be read.
	2:0	—	*Reserved*

The data returned by an access to the ASI_SCRATCHPAD_ACCESS register is defined as follows.

TABLE 19-24 ASI_SCRATCHPAD_ACCESS Format

Operation Type	Data Bits	Field	Remarks
Data read	63:0	data{63:0}	Contains the 64-bit data field read from the data array.
ECC read	63:8	—	*Reserved*. Read as all zeroes
	7:0	ecc{7:0}	Contains the eight ecc bits read from the data array.

Diagnostic reads to the scratchpad array using this ASI never result in ECC errors.

19.11 Tick Compare

Each strand has a hyperprivileged, read-only ASI_TICK_ACCESS register located at ASI $5A_{16}$, VA 0_{16}–38_{16}.

Diagnostic software can read the data or ecc bits using this register.

Implementation Note The intdis bit of the various TICK_CMPR registers is inverted before being used in the ECC calculation.

The format and usage of this register are described below.

TABLE 19-25 ASI_TICK_ACCESS Address Format

Operation Type	Bit	Field	Remarks
Load	63:6	—	*Reserved*
	5	data_np	If 1, returns the data entry contents; if 0, returns the 8 ecc bits in bit positions 7:0 of the returned data field.
	4:3	index{1:0}	Index in TICK_CMPR array to be read: 00 – TICK_CMPR; 01 – STICK_CMPR; 10 – HSTICK_COMPARE; 11 – *Reserved*.
	2:0	—	*Reserved*

The data returned by an access to the ASI_TICK_ACCESS register is defined as follows.

TABLE 19-26 ASI_TICK_ACCESS Format

Operation Type	Data Bits	Field	Remarks
Data read	63:0	data{63:0}	Contains the 64-bit data field read from the data array
ECC read	63:8	—	*Reserved*. Read as all zeroes
	7:0	ecc{7:0}	Contains the 8 ecc bits read from the data array

Diagnostic reads to the TICK_CMPR array using this ASI never result in ECC errors.

In order to cause an ECC error, software can first set up the ASI_ERROR_INJECT register as described in *ASI_ERROR_INJECT_REG* on page 255. Then, software can write to the TICK_CMPR registers using normal instructions.

19.12 Trap Stack Array (TSA)

Each strand has a hyperprivileged, read-only ASI_TSA_ACCESS register located at ASI $5B_{16}$, VA 0_{16}–38_{16}.

Diagnostic software can read the ecc bits by using this register. The format and usage of this register are described below.

TABLE 19-27 ASI_TSA_ACCESS Address Format

Operation Type	Bit	Field	Remarks
Load	63:6	—	*Reserved*
	5:3	index{2:0}	Index in TSA to be read
	2:0	—	*Reserved*

The data returned by an access to the ASI_TSA_ACCESS register is defined as follows:.

TABLE 19-28 ASI_TSA_ACCESS Format

Operation Type	Data Bits	Field	Remarks
ECC read	63:16	—	*Reserved*. Read as all zeroes
	15:0	ecc{15:0}	Contains the 16 ecc bits read from the data array

Diagnostic reads to the TSA using this ASI never result in ECC errors.

To cause an ECC error, software can first set up the ASI_ERROR_INJECT register as described in *ASI_ERROR_INJECT_REG* on page 255. Then, software can write to the TSA using normal instructions.

The TSA stores several logical registers. TABLE 19-29 documents the internal organization of the array and the mapping of the contents of the logical registers to their physical locations within the array. Note that some bits of some entries of the array are reserved for hardware use. If a correctable ECC error occurs in these bits or an error occurs in the ecc bits themselves, software can correct the error by doing the following:

1. Disabling correctable ECC checking on the TSA by setting CERER.tsac to 0.
2. Reading any logical register in the entry with an error.
3. Writing this logical register back with the same data as was read.

This causes the ECC to be recalculated.

TABLE 19-29 TSA Entry Contents

Entry	Data Bits	Logical Description	Remarks
0	151		Not software-accessible; reserved for hardware VA out-of range detection in the event that a trap handler enables translation or changes PSTATE.am. If an error causes this bit to flip from 1 to 0 and the trap handler enables translation or changes PSTATE.am, it is possible that the hardware will fail to take an *instruction_address_range* or *instruction_real_range* trap after the done or retry. If an error causes this bit to flip from 0 to 1 and the trap handler enables translation or changes PSTATE.am, it is possible that the hardware will incorrectly take an *instruction_address_range* or *instruction_real_range* trap after the DONE or RETRY instruction.
	150		Not software-accessible; reserved for hardware VA out-of-range detection in the event that a trap handler enables translation or changes PSTATE.am. If an error causes this bit to flip from 1 to 0 and the trap handler enables translation or changes PSTATE.am, it is possible that the hardware will fail to take an *instruction_address_range* or *instruction_real_range* trap after the done or retry. If an error causes this bit to flip from 0 to 1 and the trap handler enables translation or changes PSTATE.am, it is possible that the hardware will incorrectly take an *instruction_address_range* or *instruction_real_range* trap after the DONE or RETRY instruction.

TABLE 19-29 TSA Entry Contents *(Continued)*

Entry	Data Bits	Logical Description	Remarks
	149:142	ECC for bits 151,150,133:68	
	141:134	ECC for bits 67:0	
	133	—	Not software-accessible; *reserved* for hardware usage
	132:131	TSTATE{41:40}	TSTATE1.GL{1:0}
	130:123	TSTATE1{39:32}	TSTATE1.CCR
	122:115	TSTATE1{31:24}	TSTATE1.ASI
	114	HTSTATE{10}	HTSTATE.ibe
	113	TSTATE1{17}	TSTATE1.PSTATE.cle
	112	TSTATE1{16}	TSTATE1.PSTATE.tle
	111	TSTATE1{20}	TSTATE1.PSTATE.tct
	110	HTSTATE1{2}	HTSTATE.priv
	109	HTSTATE1{5}	HTSTATE1.red
	108	TSTATE1{12}	TSTATE1.PSTATE.pef
	107	TSTATE1{11}	TSTATE1.PSTATE.am
	106	TSTATE1{10}	TSTATE1.PSTATE.priv
	105	TSTATE1{9}	TSTATE1.PSTATE.ie
	104	HTSTATE1{0}	HTSTATE1.tlz
	103:101	TSTATE1{2:0}	TSTATE1.CWP{2:0}
	100:92	TT1{8:0}	
	91:46	TPC1{47:2}	
	45:0	TNPC1{47:2}	
1–5		All	Same as entry 0; Trap Stack entries 2-6

TABLE 19-29 TSA Entry Contents *(Continued)*

Entry	Data Bits	Logical Description	Remarks
6	149:142	ECC for bits 133:67	
	141:134	ECC for bits 66:0	
	133:92	—	*Reserved*. Unused
	91:80	mondo_queue_tail{17:6}	
	79:68	dev_queue_tail{17:6}	
	67:46	—	*Reserved*. Unused
	45:34	mondo_queue_head{17:6}	
	33:22	dev_queue_head{17:6}	
	21:0	—	*Reserved*. Unused
7	149:142	ECC for bits 133:67	
	141:134	ECC for bits 66:0	
	133:92	—	*Reserved*. Unused
	91:80	res_err_queue_tail{17:6}	
	79:68	nonres_err_queue_tail{17:6}	
	67:46	—	*Reserved*. Unused
	45:34	res_err_queue_head{17:6}	
	33:22	nonres_err_queue_head{17:6}	
	21:0	—	*Reserved*. Unused

19.13 MMU Register Array (MRA)

Each strand has a hyperprivileged, read-only `ASI_MRA_ACCESS` register located at ASI 51_{16}, VA 0_{16}–38_{16}.

Diagnostic software can read the parity bits using this register. The format and usage of this register are described below.

TABLE 19-30 ASI_MRA_ACCESS Address Format

Operation Type	Bit	Field	Remarks
Load	63:6	—	*Reserved*
	5:3	index{2:0}	Index in MRA array to be read.
	2:0	—	*Reserved*

The data returned by an access to the ASI_MRA_ACCESS register is defined as follows.

TABLE 19-31 ASI_MRA_ACCESS Data Format

Operation Type	Data Bits	Field	Remarks
Parity read	63:2	—	*Reserved*. Read as all zeros.
	1:0	pty{1:0}	Contains the parity bits for each 41-bit subfield.

Diagnostic reads to the MRA using this ASI never result in parity errors.

To cause a parity error, software can first set up the ASI_ERROR_INJECT register as described in *ASI_ERROR_INJECT_REG* on page 255. Then, software can write to the MRA using ASI store instructions.

The MRA stores several MMU configuration registers. TABLE 19-32 documents the internal organization of the array and the mapping of the contents of the logical registers to their physical locations within the array. Note that bits 81:78 of the array are not used for entries 3:0. Software can correct any MRA parity error by reloading the entry from a “clean” copy in memory. Hardware generates the correct parity for the entry as it is written back to the array.

TABLE 19-32 MRA Entry Contents

Entry	Bit	Logical Description	Remarks
0	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81:77	—	*Reserved*
	76:75	z_tsb_config_0{62:61}	
	74:48	z_tsb_config_0{39:13}	
	47:39	z_tsb_config_0{8:0}	
	38	—	*Reserved*
	37:36	z_tsb_config_1{62:61}	
	35:9	z_tsb_config_1{39:13}	
	8:0	z_tsb_config_1{8:0}	
1	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81:77	—	*Reserved*
	76:75	z_tsb_config_2{62:61}	
	74:48	z_tsb_config_2{39:13}	
	47:39	z_tsb_config_2{8:0}	
	38	—	*Reserved*
	37:36	z_tsb_config_3{62:61}	
	35:9	z_tsb_config_3{39:13}	
	8:0	z_tsb_config_3{8:0}	
2	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81:77	—	*Reserved*
	76:75	nz_tsb_config_0{62:61}	
	74:48	nz_tsb_config_0{39:13}	
	47:39	nz_tsb_config_0{8:0}	
	38	—	*Reserved*
	37:36	nz_tsb_config_1{62:61}	
	35:9	nz_tsb_config_1{39:13}	
	8:0	nz_tsb_config_1{8:0}	

TABLE 19-32 MRA Entry Contents *(Continued)*

Entry	Bit	Logical Description	Remarks
3	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81:77	—	*Reserved*
	76:75	nz_tsb_config_2{62:61]	
	74:48	nz_tsb_config_2{39:13]	
	47:39	nz_tsb_config_2{8:0]	
	38	—	*Reserved*
	37:36	nz_tsb_config_3{62:61]	
	35:9	nz_tsb_config_3{39:13]	
	8:0	nz_tsb_config_3{8:0]	
4	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81	real_range_0{63]	Unused
	80:27	real_range_0{53:0]	
	26:0	physical_offset_0{39:13]	
5	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81	real_range_1{63]	Unused
	80:27	real_range_1{53:0]	
	26:0	physical_offset_1{39:13]	
6	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81	real_range_2{63]	Unused
	80:27	real_range_2{53:0]	
	26:0	physical_offset_2{39:13]	
7	83	Parity for bits 81:41	
	82	Parity for bits 40:0	
	81	real_range_3{63]	Unused
	80:27	real_range_3{53:0]	
	26:0	physical_offset_3{39:13]	

19.14 L2 Cache Registers

This section discusses L2 control and status registers.

19.14.1 L2 Control Register

Each cache bank has a control register. To enable the L2 cache, the dis bit must be set to 0 for all enabled banks. The L2 cache can be disabled by setting the dis bit for all enabled banks to 1. Operation when the dis bit of the enabled L2 banks are not all the same is undefined. Note that while the L2 cache can be disabled during normal operation, it must be completely flushed of dirty cache lines before being disabled because once disabled it will no longer participate in the coherence protocol (that is, when disabled, the L2 cache is treated as if all its contents are invalid). Likewise, reenabling the L2 cache requires that the L2 cache be completely invalidated before clearing the dis bit since the data in the L2 cache can be stale. In addition, before disabling the L2 cache, the L1 instruction and data caches must be disabled in all strands' ASI_LSU_CONTROL_REG registers, as the L1 caches cannot operate properly when the L2 cache is disabled. The L2 Control register is available at address A9 0000 0000_{16} or B9 0000 0000_{16}. Address bits 8:6 select the cache bank, address bits 31:9 and 5:3 are ignored (that is, the register aliases across the address range).

The L2 Control register format is shown in TABLE 19-33.

TABLE 19-33 L2 Control Register – L2_CONTROL_REG (A9 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
63:23	—	X	RO	*Reserved*
22	dirclear	0	WO	Writing 1 clears I-Cache and D-Cache directories.
21	dbgen	0	RW	Unused
20:15	errorsteer	0	RW	Specifies the physical core (bits 20:18) and strand (bits 17:15) that receives all the L2 errors whose cause can't be linked to a specific virtual processor.
14:3	scrubinterval	0	RW	Interval between scrubbing of adjacent sets in L2 in processor core clocks. In units of 1M cycles.
2	scrubenable	0	RW	If set to 1, enable hardware scrub.
1	dmmode	0	RW	If set to 1, address bits 21 to 18 indicate the replacement way (direct-mapped mode). If set to 0, all L2 ways are enabled under LRU control.
0	dis	1	RW	If set to 1, disable the L2 cache. If set to 0, enable the L2 cache.

19.14.2 L2 Bank Available

The L2 Bank Available register format is shown below.

TABLE 19-34 L2 Bank Available – BANK_AVAIL ($80\ 0000\ 1018_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	rsvd	0	RO	*Reserved*
7:0	avail	FF_{16}	RO	L2 bank available programmed by eFuse.

19.14.3 L2 Bank Enable

The L2 Bank Enable register, described in TABLE 19-35, allows software to enable a legal combination of L2 banks. Initially this register is set identical to BANK_AVAIL. Software must program BANK_ENABLE to a legal combination of banks, as listed in TABLE 19-36.

Note | After programming BANK_ENABLE to a legal combination of banks, SW needs to initiate a warm reset to have the effect take place, i.e have all appropriate banks be enabled or disabled.

Note | The BANK_AVAIL.avail field is **and**ed with value written to BANK_ENABLE, preventing software from being able to enable a bank that is not available.

TABLE 19-35 L2 Bank Enable – BANK_ENABLE ($80\ 0000\ 1020_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	rsvd	0	RO	*Reserved*
7:0	enable	BANK_AVAIL.avail	RW	L2 banks enabled.

TABLE 19-36 Legal Values for BANK_ENABLE.enable.

Value	Description
03_{16}	2-bank mode
$0C_{16}$	2-bank mode
$0F_{16}$	4-bank mode
30_{16}	2-bank mode
33_{16}	4-bank mode

TABLE 19-36 Legal Values for BANK_ENABLE.enable.

$3C_{16}$	4-bank mode
$C0_{16}$	2-bank mode
$C3_{16}$	4-bank mode
CC_{16}	4-bank mode
$F0_{16}$	4-bank mode
FF_{16}	8-bank mode

TABLE 19-37 3 bank pair to 2 bank pair remap of BANK_ENABLE.enable.

BANK_ENABLE.enable	Remapped To
$3F_{16}$	$0F_{16}$
CF_{16}	$0F_{16}$
$F3_{16}$	$F0_{16}$
FC_{16}	$F0_{16}$

TABLE 19-36 specifies the following rules:

1. Two banks enabled: Any one of BA01, BA23, BA45, or BA67.
2. Four banks enabled: Any one of {BA01, BA23},{BA23, BA45}, {BA45, BA67}, {BA67, BA01}, {BA01,BA45}, or {BA23,BA67}.
3. Eight banks enabled: {BA01, BA23, BA45, BA67}.

Note | In case Software programs 3 bank pairs to be enabled in the L2 Bank Enable register in violation of the legal values indicated in TABLE 19-36, UltraSPARC T2 hardware remaps it to 2 bank pairs as per TABLE 19-37 on page 434 .

WARNING! | Because the L2 reverse directory can accommodate no more than 1/8 of the L1 cache lines per virtual processor, in the partial bank enabled configurations, the number of SPC cores enabled through CMP registers must be less than or equal to *n*, where *n* is the number of available banks.

Notes | Software needs to ensure PA{39} = 0 in 8-bank, PA{39:38} = 0_2 in 4-bank mode and PA{39:37} = 000_2 in 2-bank mode.

PA{39:38} = 0_2 in 4-bank mode and PA{39:37} = 000_2 in 2-bank mode.

Note Because a pair of L2 banks are directly connected to a single memory port, in a 2-bank configuration, only a single memory port will be usable. In a 4-bank configuration, only a pair of memory ports will be usable.

Note In 2 or 4 bank enabled modes, other than accesses to Bank Enable,Bank Enable Status,Index Hash Enable, Index Hash Enable Status registers which reside in NCU block, UltraSPARC T2 would have CSR/Diagnostic accesses to all of disabled L2 banks aliased and remapped to enabled banks by SPC.

Implementation Notes In the 2-bank configuration, the L2 bank is selected by PA{6}, with PA{6} = 0 selecting the even bank and PA{6} = 1 selecting the odd bank. In the 4-bank configuration, the L2 index is selected by PA{7:6}. PA{7} selects between the pair of banks enabled, with PA{7} = 0 selecting the lower numerical pair of banks (except for the case where banks BA67 and BA01 are enabled, where PA{7} = 0 selects BA67), and PA{7} = 1 selecting the higher numerical pair of banks (again except for the case where banks BA67 and BA01 are enabled, where PA{7} = 1 selects BA01). PA{6} selects between the even and odd banks of the bank pair selected by PA{7}, with PA{6} = 0 selecting the even bank and PA{6} = 1 selecting the odd bank.

In the 2-bank configuration, the L2 index is formed from PA{15:7} and the tag from PA{37:16} (and PA{39:38} is zeroed out). The usual addr{10,9,5,4} accesses reverse directory CAM panel (1 of 16, each is 32 way), but PA{8:7} is used to access 8 out of the 32 ways in the directory CAM. In the 4-bank configuration, the L2 index is formed from PA{16:8} and the tag from PA{38:17} (and PA{39} is zeroed out). The usual addr{10,9,5,4} accesses reverse directory CAM panel (1 of 16, each is 32 way), but PA{8} accesses 16 out of the 32 ways in the directory CAM.

19.14.4 L2 Bank Enable Status

The Bank Enable Status register shows the status of the hardware signals generated as a result of the BANK_ENABLE register. Software can use this register to determine when the effects of a write to BANK_ENABLE are complete. The preview versions of the status change immediately after BANK_ENABLE is written, while the non-

preview version change only after the next warm reset when all appropriate banks have been enabled or disabled.SW needs to initiate a warm reset to have all appropriate banks have been enabled or disabled.

TABLE 19-38 L2 Bank Enable Status – BANK_ENABLE_STATUS (80 0000 1028_{16})

Bit	Field	Initial Value	R/W	Description
63:13	—	0	RO	*Reserved*
12	pm_preview	0	RO	L2 partial mode enable. Set if BANK_ENABLE is not FF_{16}.
11	ba67_preview	1	RO	Banks 6 and 7 enable. Set if bits 7:6 of BANK_ENABLE are 3_{16}.
10	ba45_preview	1	RO	Banks 4 and 5 enable. Set if bits 5:4 of BANK_ENABLE are 3_{16}.
9	ba23_preview	1	RO	Banks 3 and 2 enable. Set if bits 3:2 of BANK_ENABLE are 3_{16}.
8	ba01_preview	1	RO	Banks 1 and 0 enable. Set if bits 1:0 of BANK_ENABLE are 3_{16}.
7:5	—	0	RO	*Reserved*
4	pm	0	RO	L2 partial mode enabled. Set if 4 or 2 L2 banks are enabled.
3	ba67	1	RO	Banks 6 and 7 enabled. Set if L2 banks 7:6 are enabled.
2	ba45	1	RO	Banks 4 and 5 enabled. Set if L2 banks 5:4 are enabled.
1	ba23	1	RO	Banks 3 and 2 enabled. Set if L2 banks 3:2 are enabled.
0	ba01	1	RO	Banks 1 and 0 enabled. Set if L2 banks 1:0 are enabled.

19.14.5 L2 Index Hash Enable

The L2 has a hashing function available to relieve hot-spotting in certain L2 sets. Hashing is controlled through the L2_IDX_HASH_EN and L2_IDX_HASH_EN_STATUS registers. The hashing enable can be changed only following a warm reset. After a warm reset, the value written into L2_IDX_HASH_EN is sent the L2 and reflected in L2_IDX_HASH_EN_STATUS once the L2 completes hashing setup.

TABLE 19-39 L2 Index Hash Enable – L2_IDX_HASH_EN (80 0000 1030_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	enb_hp	0	RW	If 1, enable L2 index hashing following next warm reset. If 0, disable L2 index hashing following next warm reset.

Implementation Note The hashing function used by UltraSPARC T2 sets the effective PA{17:11} ← {(PA{32:28} **xor** PA{17:13}) **::** (PA{19:18} **xor** PA{12:11})}. This hashing function is unaffected by the BANK_ENABLE_STATUS register settings.

19.14.6 L2 Index Hash Enable Status

TABLE 19-40 L2 Index Hash Enable Status – L2_IDX_HASH_EN_STATUS (80 0000 1038_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	enb_hp	0	RO	Set if L2 index hashing is enabled.

19.14.7 Other L2 Registers

Other L2 Cache registers are defined in other chapters, particularly the error control and status registers are in the Error Handling chapter.

19.15 L2 Cache Flushing

UltraSPARC T2 provides Prefetch [address], 18_{16} (PrefetchICE) for invalidating and coherently committing an L2 cache line to memory. This PrefetchICE is only available in hyperprivileged mode. Execution of a PrefetchICE instruction while in user or privileged mode is a NOP. TABLE 19-41 lists the address format for the PrefetchICE.

TABLE 19-41 PrefetchICE Address Format

Bit	Field	Description
63:40	—	*Reserved*
39:37	key	Must be 011. Use of any other value places UltraSPARC T2 in an undefined state.
36:22[1]	—	*Reserved*, can be any value.
21:18[2]	way	Selects way in cache set.
17:9[3]	set	Selects cache set in bank.
8:6[4]	bank	Selects cache bank.
5:0	—	Reserved, can be any value.

1. When index hashing is enabled, bits 32:28 will affect the set selection.
2. When index hashing is enabled, bits 19:18 will affect the set selection.
3. When four L2 banks are enabled, set is specified in bits 16:8 and bit 17 may be any value, while when two banks are enabled, set is specified in bits 15:7 and bits 17:16 may be any value.
4. When four L2 banks are enabled, bank is specified in bits 7:6 only, while when two banks are enabled, bank is specified in bit 6 only.

Note The PrefetchICE address specifies which particular L2 cache way, set, and bank are committed to memory. To flush a specific address, software must either issue a PrefetchICE to all 16 possible ways or, if software can guarantee that the address will not be refetched during the flushing code, software can use diagnostic reads to find which way contains the cache address and then issue a PrefetchICE to that specific way. A series of PrefetchICE's to a particular L2 bank can be guaranteed complete if followed by any load to that bank (even diagnostic load). The completion of the load will indicate that the PrefetchICE instructions are complete.

Programming Notes When L2 index hashing is enabled (as specified in L2_IDX_HASH_EN_STATUS), software will need to generate the nonhashed index for use by the PrefetchICE. While bits 32:28 can be set to zero to remove their effect on hashing, bits 19:18 are part of the hash calculation, and the bits 12:11 of the desired set must be **xor**ed with bits 19:18 to generate the bits 12:11 used in the PrefetchICE. For example, if software desired to flush all 16 ways of bank 0, set 0, it would need to generate the following addresses: 60 0000 0000_{16}, 60 0004 0800_{16}, 60 0008 1000_{16}, 60 000C 1800_{16}, 60 0010 0000_{16}, 60 0014 0800_{16}, 60 0018 1000_{16}, 60 001C 1800_{16}, 60 0020 0000_{16}, 60 0024 0800_{16}, 60 0028 1000_{16}, 60 002C 1800_{16}, 60 0030 0000_{16}, 60 0034 0800_{16}, 60 0038 1000_{16}, 60 003C 1800_{16}.

When four or two L2 banks are enabled (as specified in BANK_ENABLE_STATUS), PrefetchICE will shift the set and bank bits (for 4 banks set is 16:8 and bank is 7:6, for 2-banks set is 15:7 and bank is 6). The way will always be specified in bits 21:18 and will not be shifted based on BANK_ENABLE_STATUS.

Note If a PrefetchICE instruction detects a tag parity error it issues a scrub of the tag array index and gets replayed after the scrub. It completes normally if there is no further tag parity error. If the PrefetchICE instruction detects a VUAD correctable error, the PrefetchICE instruction just completes as normal and the VUAD error gets silently corrected (without being logged). Since the VUAD error gets corrected early in the L2 pipeline, the PrefetchICE instruction always gets the correct value of the Dirty bits in its eviction pass and hence completes normally.

Note | A PrefetchICE instruction when used to invalidate a line from UltraSPARC T2 L2 Cache might cause a tag parity error or data correctable/uncorrectable error to be detected and reported . In case the reporting has been turned off by software prior to the issue of the Prefetch ICE, the error would still be in pending state until the next miss is serviced by the L2 cache. In order not to have this error reported in the future, software should clear this internal error reporting pending state in L2 cache by forcing a miss to that particular L2 bank after the Prefetch ICE . One way to force the miss is to issue a load to the same physical address as the line which got invalidated by Prefetch ICE.

19.16 L2 Cache Diagnostic Access

This section describes the control registers and diagnostic access for the L2 cache.

Diagnostic accesses will functionally work in the midst of other operations, but the results of L2 diagnostic accesses are inherently somewhat indeterminate because the state of the L2 cache is a moving target. Of course, diagnostic writes have the capability of putting UltraSPARC T2 into an inconsistent or illegal state.

TABLE 19-42 shows the breakdown of the L2 address range.

TABLE 19-42 Breakdown of the L2 Diagnostic Address Range

Address Range (8 MSBs of the 40-bit Address)	Assigned to:	Comment
$A0_{16}$–$A3_{16}$/$B0_{16}$–$B3_{16}$	L2 Data	Diagnostic access to the L2 data array.
$A4_{16}$–$A5_{16}$/$B4_{16}$–$B5_{16}$	L2 Tag	Diagnostic access to the L2 tag array.
$A6_{16}$–$A7_{16}$/$B6_{16}$–$B7_{16}$	L2 Tag VUAD	Diagnostic access to the L2 VUAD array.
$A8_{16}$–AF_{16}/$B8_{16}$–BF_{16}	L2 Registers	Error, control, and status registers.

19.16.1 L2 Data Diagnostic Access

Diagnostic access to the L2 data array is done through 64-bit read/writes that access a 32-bit data subblock along with the corresponding 7-bit ecc. The format for addressing the entire L2 cache is shown in TABLE 19-43.

Note | Diagnostic loads of the L2 data do not check the ecc, and thus cannot generate an ECC error.

Programming Note L2 Data diagnostic addresses are not affected by L2 index hashing (as specified in L2_IDX_HASH_EN_STATUS). If index hashing is enabled and software wants to access a line corresponding to a specific PA, software must adjust the L2 Data diagnostic address accordingly.

Programming Note In case L2 is configured in partial bank mode, for a L2 data diagnostic access, the L2 index gets picked up from bits 16:8 in 4 bank mode and 15:7 in 2 bank mode from the L2 data diagnostic address packet which is similar to the way L2 index is constructed for any cacheable access to L2 in partial bank modes. Bits 17:6 should be the same as the original PA in partial bank mode.

TABLE 19-43 Format 6 L2 Data Diagnostic Addressing

Bit	Field	Description
63:40	—	*Reserved*
39:32	select	Must be $A0_{16}$, $A1_{16}$, $A2_{16}$, $A3_{16}$, $B0_{16}$, $B1_{16}$, $B2_{16}$, or $B3_{16}$ to select L2 data diagnostic access.
31:23	—	*Reserved*, can be any value (that is, the data diagnostic access is aliased throughout the A0 – A3/B0–B3 address range).
22	oddeven	Selects 32 bit word from 64 bit word selected by word field.
21:18	way	Selects way in cache set.
17:9	set	Selects cache set in bank.
8:6	bank	Selects cache bank.
5:3	word	Selects 64 bit word in 64 byte cache line.
2:0	—	All zero for 64-bit access.

The data format is shown in TABLE 19-44.

TABLE 19-44 23 L2 Diagnostic Data – L2_DIAG_DATA (A0 0000 0000_{16}) (Count 2, Step 4194304) (Count 393216, Step 8)

Bit	Field	Initial Value	R/W	Description
63:39	—	X	RO	*Reserved*
38:7	data	X	RW	Data.
6:0	ecc	X	RW	ECC value for data.

19.16.2 L2 Tag Diagnostic Access

Diagnostic access to the L2 tag array is done through 64 bit read/writes that access the tag along with the corresponding 6-bit ecc. The format for addressing the entire L2 cache is shown in TABLE 19-45.

Note | Diagnostic loads of the L2 tag do not check the ecc, and thus cannot generate an ECC error.

Programming Note | L2 Tag diagnostic addresses are not affected by either partial L2 banks (as specified in BANK_ENABLE_STATUS) or L2 index hashing (as specified in L2_IDX_HASH_EN_STATUS). If either index hashing is enabled or not all banks are enabled and software wants to access a line corresponding to a specific PA, software must adjust the L2 Tag diagnostic address accordingly. In case L2 is configured in partial bank mode, for a L2 tag diagnostic access, the L2 index gets picked up from 17:9 of the L2 Tag Diagnostic address packet . Hence software should precompute the physical index from the specific PA based on the partial bank configuration (16:8 for 4 bank and 15:7 for 2 bank) and present in on bits 17:9 of the L2 Tag diagnostic packet. However bits 8:6 should be the same as the original PA in partial bank mode.

TABLE 19-45 Format 7 L2 Tag Diagnostic Addressing

Bit	Field	Description
63:40	—	*Reserved*
39:32	select	Must be $A4_{16}$, $A5_{16}$, $B4_{16}$, or $B5_{16}$ to select L2 tag diagnostic access.
31:22	—	*Reserved*, can be any value (i.e., the tag diagnostic access is aliased throughout the A4 – A5/B4 – B5 address range).
21:18	way	Selects way in cache set.
17:9	set	Selects cache set in bank.
8:6	bank	Selects cache bank.
5:3	—	*Reserved*, can be any value.
2:0	—	All zero for 64-bit access.

The data format is shown in TABLE 19-46.

TABLE 19-46 24 L2 Diagnostic Tag – L2_DIAG_TAG (A4-0000-0000$_{16}$) (Count 49152, Step 64)

Bit	Field	Initial Value	R/W	Description
63:30	—	X	RO	*Reserved*
27:6	tag	X	RW	Tag, corresponds to addr{39:18}[1].
5:0	ecc	X	RW	ECC value for tag.

1. With the L2 in 4 bank mode, the tag contains addr{38:17} if index hashing is disabled, and {addr{38:18}, addr{32}^addr{17}} if index hashing is enabled. With the L2 in 2-bank mode, the tag contains addr{37:16} if index hashing is disabled, and {addr{36:18},addr{32:31} **xor** addr{17:16}} if index hashing is enabled.

Note | Restraint must be exercised when using this register to inject errors into the L2 tags. Since a tag correction takes much longer than an injection and since correction causes a retry for an access that discovers an error, repeatedly injecting errors at a high rate can prevent forward progress of another virtual processor accessing a line at that cache index (set). Injecting each error just once (potentially for many individual errors) or ensuring that the injections to the same index are at least 1000 cycles apart are both sufficient to avoid this problem.

19.16.3 L2 VUAD Diagnostic Access

Diagnostic access to the L2 VUAD array is done through a pair of address access ranges. The first accesses the valid and dirty bits for an entire set plus the ECC for those bits across the set via 64-bit read/writes. The format for addressing the entire L2 cache is shown in TABLE 19-47.

Note | Diagnostic loads of the VUAD do not check ecc, and thus cannot generate ECC errors.

Programming Note | L2 VUAD diagnostic addresses are not affected by L2 index hashing (as specified in L2_IDX_HASH_EN_STATUS). If index hashing is enabled software wants to access a line corresponding to a specific PA, software must adjust the L2 VUAD diagnostic address accordingly.

Programming Note In case L2 is configured in partial bank mode, for a L2 VUAD diagnostic access, the L2 physical index gets picked up from bits 16:8 in 4 bank mode and 15:7 in 2 bank mode from the L2 VUAD diagnostic address packet which is similar to the way L2 index is constructed for any cacheable access to L2 in partial bank modes.Bits 17:6 should be the same as the original PA in partial bank mode.

TABLE 19-47 Format 8 L2 VD Diagnostic Addressing

Bit	Field	Description
63:40	—	*Reserved*
39:32	select	Must be $A6_{16}$, $A7_{16}$, $B6_{16}$, or $B7_{16}$ to select L2 VD diagnostic access.
31:23	—	*Reserved*, can be any value (i.e., the VD diagnostic access is aliased throughout the A6–A7/B6–B7 address range where bit 22 is 1).
22	vdsel	Must be set to 1
21:18	—	*Reserved*, can be any value (i.e., the VD diagnostic access is aliased throughout the A6– A7/B6–B7 address range where bit 22 is 1).
17:9	set	Selects cache set in bank.
8:6	bank	Selects cache bank.
5:3	—	*Reserved*, can be any value.
2:0	rsvd4	All zero for 64 bit access

The data format is shown in TABLE 19-48.

TABLE 19-48 25 L2 Diagnostic VD – L2_DIAG_VD (A6 0040 0000_{16}) (count 4096 step 64)

Bit	Field	Initial Value	R/W	Description
63:39	—	X	RO	*Reserved*
38:32	vdecc	X	RW	ECC for all dirty and valid bits.
31:16	valid	X	RW	Valid bits for way 15 down to way 0.
15:0	dirty	X	RW	Dirty bits for way 15 down to way 0.

The second range accesses the AD bits for an entire set via 64 bit read/writes. The format for addressing the entire L2 cache is shown in TABLE 19-49.

TABLE 19-49 Format 9 L2 UA Diagnostic Addressing

Bit	Field	Description
63:40	—	*Reserved*
39:32	select	Must be $A6_{16}$, $A7_{16}$, $B6_{16}$, or $B7_{16}$ to select L2 UA diagnostic access.
31:23	—	*Reserved*, can be any value (i.e., the UA diagnostic access is aliased throughout the A6–A7/B6–B7 address range where bit 22 is 0).

TABLE 19-49 Format 9 L2 UA Diagnostic Addressing *(Continued)*

Bit	Field	Description
22	vdsel	Must be set to 0.
21:18	—	*Reserved*, can be any value (i.e., the UA diagnostic access is aliased throughout the A6–A7/B6–B7 address range where bit 22 is 0).
17:9	set	Selects cache set in bank.
8:6	bank	Selects cache bank.
5:3	—	*Reserved*, can be any value.
2:0	rsvd4	All zero for 64 bit access.

The data format is shown in TABLE 19-50.

TABLE 19-50 26 L2 Diagnostic UA – L2_DIAG_UA (A6 000 00000_{16}) (Count 4096 Step 64)

Bit	Field	Initial Value	R/W	Description
63:39	—	X	RO	*Reserved*
38:32	uaecc	X	RW	ECC for all used and alloc bits.
31:16	used	X	RW	Used bits for way 15 down to way 0.
15:0	alloc	X	RW	Allocated bits for way 15 down to way 0.

19.17 Built-In Self-Test (BIST)

19.17.1 L2 BIST Control

On UltraSPARC T2, each L2 bank has its BIST control coming from the Test Control Unit (TCU). The software-visible BIST state of each L2 bank is also available in TCU. Hence the L2 Control register at address A8 0000 0000_{16}/B8 0000 0000_{16} is reserved and returns a 64-bit zero value when read.

CHAPTER 20

Hardware Debug Support

20.1 SPARC Core Debug Features

Each physical UltraSPARC T2 SPARC core contains several hardware debug features. Some (breakpoints and watchpoints, for example) have already been mentioned in Chapter 19, *Configuration and Diagnostics Support*.

20.1.1 Shadow Scan

Each physical UltraSPARC T2 SPARC core supports the ability to capture a subset of each strand's state for inspection via a shadow scan facility. The shadow scan is invoked by JTAG commands (see Appendix H, *JTAG (IEEE 1149.1) Scan Interface*). Please refer to section *Shadow Scan Chains* on page 982 for Shadow Scan in UltraSPARC T2.

The TCU continually specifies a strand ID to each physical UltraSPARC T2 SPARC core. In response, the physical core atomically captures the state as described in the following table in a scan string. The TCU then accesses the scan string and captures it in a JTAG-visible register for presentation over the JTAG interface.

TABLE 20-1 lists the state that can be gathered from each running virtual processor.

TABLE 20-1 SPARC Shadow Scan State

Data Bits	Field	Remarks
117:72	VA{47:2}	Virtual address of last instruction executed by that strand
71	ibe	HPSTATE.ibe
70	cle	PSTATE.cle
69	tle	PSTATE.tle
68	tct	PSTATE.tct

TABLE 20-1 SPARC Shadow Scan State

Data Bits	Field	Remarks
67	hpriv	HPSTATE.hpriv
66	red	HPSTATE.red
65	pef	PSTATE.pef
64	am	PSTATE.am
63	priv	PSTATE.priv
62	ie	PSTATE.ie
61	tlz	HPSTATE.tlz
60:58	TL{2:0}	TL
57:12	TPC{47:2}	TPC for the last trap
11:3	TT{8:0}	TT for the last trap
2:0	TL_FOR_TT{2:0}	TL for the last trap

VA is updated when an instruction is executed. When the TCU changes the strand ID that controls which strand's state is captured, the VA field can capture spurious values. Between the time that the TCU updates the strand ID and the time that the new strand executes an instruction, the VA field reflects either the VA of the last instruction that executed on the previous strand ID or a spurious value captured during the strand ID transition. After the first instruction executes on the new strand, the VA field correctly reflects the VA for the new strand. There is no valid bit or strand ID captured in the shadow scan to distinguish when the VA has updated for the new strand.

HPSTATE, PSTATE, and TL are updated dynamically. Thus, they correctly reflect updates after a trap, a DONE, a RETRY, or a software write to `%tl`. These fields also update immediately after the TCU changes the strand ID.

TPC, TT, and TL_FOR_TT, however, 0

are updated only when the strand takes a trap. They are not updated for DONE, RETRY, or software writes to `%tl`. For example, if the processor traps from TL = 0 to TL = 1 to TL = 2 and then uses DONE and/or RETRY to get back to TL = 0, shadow scan will still reflect TT{2}, TPC{2}, and TL_FOR_TT will still be 2. Similarly, if the processor traps out to TL = 2 and then software writes TL to 1 or 0, shadow scan will still show TT{2}, TPC{2}, and TL_FOR_TT will still be 2. Simiarly, after TCU changes the strand ID, these fields reflect state for the previous strand ID until the new strand takes a trap. There is no valid bit or strand ID captured in the shadow scan to distinguish when these fields have updated for the new strand.

On the JTAG bus, bit 117 appears first, followed by bit 116, sequentially down to and including bit 0.

If multiple traps occur after the state was atomically captured into the shadow scan string, but while the shadow scan string is being scanned, only the state belonging to the last trap remains to be (potentially) captured on the next capture point. Due to the length of time to perform a shadow scan relative to the time between traps, sampling via shadow scan can miss several traps.

All eight core shadow scans are scanned serially as one chain, with strand 0 closest to TDI and strand 7 closest to TDO. Any strand marked unavailable in the CMP STRAND_AVAILABLE register will not be included when scanned via TDI to TDO. The shadow scan chain for a given strand is placed in that strand's second scan chain during ATPG test mode; they are accessible otherwise only via JTAG shadow scan instructions (that is, not during JTAG serial scan).

20.1.2 SPARC Debug Event Control Register

There is one debug control register per physical UltraSPARC T2 SPARC core (ASI_DECR). Each event type field in the ASI_DECR contains two bits that encode the type of response for that event as shown in TABLE 20-2.

TABLE 20-2 ASI_DECR Debug Event Response Type Encoding

DECR Event Response Type Encoding	Response If Debug Event Occurs
00	Debug event disabled.
01	Soft stop.
10	Hard stop.
11	Pulse TRIGOUT/ Watchpoint.

All strands of a physical UltraSPARC T2 SPARC core share a hyperprivileged, read/write, Debug Event Control register located at ASI 45_{16}, VA 8_{16}. The DECR controls the debug event response (hard stop or soft stop or watchpoint) for an associated core debug event if that event occurs.

The format of the ASI_DECR register is described in TABLE 20-3.

TABLE 20-3 ASI_DECR Register (ASI 45_{16}, VA 8_{16})

Bit	Field	Initial Value	R/W	Description
63:62	iwa_de	0 (preserved across warm and debug reset)	RW	Instruction breakpoint match debug event enable.
61:60	iva_de	0 (preserved across warm and debug reset)	RW	Instruction virtual address match debug event enable.
59:58	dva_de	0 (preserved across warm and debug reset)	RW	Data virtual address match debug event enable.
57:56	dpa_de	0 (preserved across warm and debug reset)	RW	Data physical address match debug event enable.
55:54	tct_de	0 (preserved across warm and debug reset)	RW	Trap on Control Transfer debug event enable.
53:52	pe_de	0 (preserved across warm and debug reset)	RW	Precise error event (an event which would be recorded in the I-SFSR or D-SFSR) debug event enable.
51:50	de_de	0 (preserved across warm and debug reset)	RW	Disrupting error event (an event which would be recorded in the DESR) debug event enable.
49:48	df_de	0 (preserved across warm and debug reset)	RW	Deferred error event (an event which would be recorded in the DFESR) debug event enable.
47:46	pm_de	0 (preserved across warm and debug reset)	RW	Performance monitor event which causes a performance counter to wrap debug event enable.
45:0	—	0	RO	*Reserved*

20.1.3 Disable Overlap and Single-Stepping Modes

Each core can disable overlap of instruction execution within all threads. In this mode, each thread issues one instruction, waits for the instruction to commit and any associated memory operations to be globally observed, then fetches and executes the next instruction. This mode essentially disables pipelining for all of a physical core's strands.

Additionally, each core has the capability of executing one instruction from each enabled strand. This is referred to as single-step.

These two features are controlled through JTAG only by the TAP_DOSS_ENABLE, TAP_DOSS_MODE, TAP_SS_REQUEST, TAP_DOSS_STATUS, TAP_CS_MODE, TAP_CS_STATUS JTAG commands described in TABLE H-1 on page 971.

20.1.4 ASI_RST_VEC_MASK

All physical cores share a hyperprivileged, read/write ASI_RST_VEC_MASK register located as ASI 45_{16}, VA 18_{16}. Reserved bits read as zero and are ignored on write. The contents of this register are preserved across warm reset. In normal operation,

resets trap to the *RSTVADDR* **or** (TT ← 5) (FF FFFF FFF0 $0000_{16}x0$), which maps to ROM. To enhance repeatability, UltraSPARC T2 can direct resets to RAM instead. To redirect resets to RAM, hyperprivileged software can set the vec_mask bit of the ASI_RST_VEC_MASK register.

TABLE 20-4 ASI_RST_VEC_MASK Register (ASI 45_{16}, VA 18_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	vec_mask	0 (preserved across warm and debug reset)	RW	If 1, set *RSTV* to 0000 0000 0000 0000_{16}. If 0 set *RSTV* to FFFF FFFF F000 0000_{16}.

20.2 SOC Debug Features

20.2.1 SOC Debug Event Control Register

There is one debug control register for SOC (SOC_DECR). SOC_DECR controls the debug response type (hard stop or watchpoint) for an associated SOC debug event if that event occur. Each event type field in SOC_DECR contains two bits that encode the type of response for that event as shown in the following table.

TABLE 20-5 SOC_DECR Debug Event Response Type Encoding

DECR Event Response Type Encoding	Response If Debug Event Occurs
00	Debug event disabled
01	*Reserved*
10	Hard-stop
11	Pulse TRIGOUT/Watchpoint

The format of the SOC_DECR is described in the following table.

TABLE 20-6 SOC_DECR Register – SOC_DECR ($86\ 0000\ 0010_{16}$)

Bit	Field	Initial Value	R/W	Description
63:22	—	0	RO	*Reserved*
21:20	se_de	0 (preserved across warm and debug reset)	RW	SOC (SII, SIO, NCU, DMU, PEU) error debug event enable.
19:18	me_de	0 (preserved across warm and debug reset)	RW	MCU error debug event enable.
17:16	l2e_de	0 (preserved across warm and debug reset)	RW	L2 error debug event enable.
15:14	l2b7_de	0 (preserved across warm and debug reset)	RW	L2 PA match bank 7 debug event enable.
13:12	l2b6_de	0 (preserved across warm and debug reset)	RW	L2 PA match bank 6 debug event enable.
11:10	l2b5_de	0 (preserved across warm and debug reset)	RW	L2 PA match bank 5 debug event enable.
9:8	l2b4_de	0 (preserved across warm and debug reset)	RW	L2 PA match bank 4 debug event enable.
7:6	l2b3_de	0 (preserved across warm and debug reset)	RW	L2 PA match bank 3 debug event enable.
5:4	L2B2_DE	0 (preserved across warm and debug reset)	RW	L2 PA match bank 2 debug event enable.
3:2	L2B1_DE	0 (preserved across warm and debug reset)	RW	L2 PA match bank 1 debug event enable.
1:0	L2B0_DE	0 (preserved across warm and debug reset)	RW	L2 PA match bank 0 debug event enable.

20.2.2 L2 Cache Debug Features

20.2.2.1 L2 Address Mask and Compare Registers

Each L2 Cache bank has a pair of registers per bank to control debug based on PA matching. The L2_MASK_REG controls which bits are compared against the values set in the L2_COMP_REG. A debug event is asserted if data and L2_MASK_REG = L2_COMP_REG and data is valid.

Since there are eight L2 cache banks in UltraSPARC T2, there are eight debug events based on PA matching. Debug response for these events are controlled by SOC_DECR{15:0}.

TABLE 20-7 L2 Address Mask Register – L2_MASK_REG (AF 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
63:52	—	0	RO	*Reserved*
51:48	ttype	0 (preserved across warm and debug reset)	RW	ttype{3:0} mask.
47:46	—1	0	RO	*Reserved*
45:40	vcid	0 (preserved across warm and debug reset)	RW	Virtual processor ID mask.
39:34	—	0	RO	*Reserved* (Physical Address{39:34} ignored.)
33:2	addr	0 (preserved across warm and debug reset)	RW	Physical Address{33:2} mask.
1:0	—3	0	RO	*Reserved* (Physical Address{1:0} ignored.)

TABLE 20-8 L2 Address Compare Register – L2_COMP_REG (BF 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
63:52	—	0	RO	*Reserved*
51:48	ttype	0 (preserved across warm and debug reset)	RW	ttype{3:0} compare value.
47:46	—1	0	RO	*Reserved*
45:40	vcid	0 (preserved across warm and debug reset)	RW	Virtual processor ID compare value.

TABLE 20-8 L2 Address Compare Register – L2_COMP_REG (BF 0000 0000_{16}) (Count 8 Step 64)

Bit	Field	Initial Value	R/W	Description
39:34	—	0	RO	*Reserved* (Physical Address{39:34} ignored.)
33:2	addr	0 (preserved across warm and debug reset)	RW	Physical Address{33:2} compare value.
1:0	—3	0	RO	*Reserved* (Physical Address{1:0} ignored.)

20.2.2.2 L2 Shadow Scan

Shadow scan for L2 Error registers is controlled via JTAG. The contents to be captured in the shadow scan are as shown in TABLE 20-9. Refer also to *Shadow Scan Chains* on page 982 for L2 shadow scan.

TABLE 20-9 L2 Shadow Scan State

Data Bits	Remarks
141:84	L2 Error Status register.
83:36	Notdata error register.
35:0	FE/UE/CE Error Address register.

All eight L2 Tag shadow scan contents are captured at the same time, and are available at TDO with L2T7 first and L2T0 last . JTAG instructions to support L2 Tag shadow scan are shown in TABLE D-1.

On the JTAG bus, bits appear in the order:
15:0, 31:16, 47:32 ,63:48, 79:64, 95:80, 111:96, 127:112, 141:128.

20.2.3 Debug Event Trigger Enables

The registers described in this section contain the masks for generating debug events in SOC_DECR.

20.2.3.1 PCI-Express Debug Event Trigger Enable Register

PCI-Express Debug Event Trigger Enable register is described in *DMU Core and Block Interrupt Enable Register (0063180016 / 016)* on page 554.

20.2.3.2 DRAM Debug Event Trigger Enable Register

Each DRAM controller has a register that contains the debug event trigger enable for DRAM controller related debug events.

TABLE 20-10 shows the format of the DRAM Debug Trigger Enable register.

TABLE 20-10 DRAM Debug Event Trigger Enable Register – DRAM_DBG_TRG_EN_REG ($84\ 0000\ 0230_{16}$) (Count 4 Step 4096)

Bit	Field	Initial Value	R/W	Description
63:5	—0	0_{16}	RO	*Reserved*
4	dtm_mask1	0_{16}	RW	If set to 1, mask off CRC data for FBD Channel 1 going to Debug bus. Reset only on POR.
3	dtm_mask0	0_{16}	RW	If set to 1, mask off CRC data for FBD Channel 0 going to Debug bus. Reset only on POR.
2	dbg_trig_en	0_{16}	RW	Trigger enable for DRAM errors. dbg_en is cleared when triggered. Reset only on POR.
1	mask_err	0_{16} (preserved across WMR and DBR)	RW	If set to 1, all LFSR mismatches on ALERT frame patterns coming in from the AMB are masked.
0	kp_lnk_up	0_{16} (preserved across WMR and DBR)	RW	If set to 1, MCU keeps sending out sync pulses on the southbound links during the entire duration of the debug/warm reset, thereby keeping the links enabled during the duration of the debug/warm reset. Clearing this bit to 0 after the debug/warm reset takes the MCU's FBDIMM interface state machine to L0 state, where it gets ready to dispatch new read/write requests from SPC's/SOC to the DIMMs.

20.2.3.3 NCU Debug Event Trigger Enable Register

The NCU block has a register that contains the debug event trigger enable for NCU-related debug events.

The format of the NCU Debug Event Trigger Enable register is shown in TABLE 20-11.

TABLE 20-11 NCU Debug Event Trigger Enable Register – NCU_DBG_TRG_EN_REG ($80\ 0000\ 4000_{16}$)

Bit	Field	Initial Value	RW	Description
63:1	—	0	RO	*Reserved*
0	dbg_trig_en	0_{16}	RW	Trigger enable for SOC Error Status register errors.

20.2.3.4 L2 Debug Event Trigger Enable Register

The L2 Debug Event Trigger Enable is described in *L2 Error Enable Register* on page 276.

20.3 TCU Debug Support

The TCU handles debug events requests from the SPARC virtual processors directly or from the SOC. The response to these requests is to stop the clocks (hard or soft) or pass the watchpoint signal to the I/O pins. A set of registers is provided in the TCU to assist in control of these responses to debug event requests.

TCU is not designed to handle burst CSR read requests from SPARC virtual processors, that is, a CSR read request cannot be followed immediately by another CSR read request, otherwise the second one may be dropped and no read data will be returned and the thread issuing the second request may hang. Users should program the second CSR read request after the data for the first one has returned. In the case of multiple SPARC threads accessing TCU CSRs, some mechanism (such as a semaphore lock) should be used to guarantee only one thread accesses any TCU CSR register at a given time.

20.3.1 Action in Response to a Soft-Stop Event

If the DECR bits for a particular event are configured for a soft-stop (set to 01_2), and that event occurs, the following sequence of operations results. The UltraSPARC T2 core or SOC sends a soft stop request to the TCU. The TCU lowers the strand_running inputs for all strands. The strand stops issuing instructions and waits for all core activity to quiesce. "Quiesce" means that all in-flight instructions have completed (or taken exceptions) all memory references issued by the core have been globally observed. Each strand lowers STRAND_RUNNING_STATUS as it quiesces. Once all strands of a physical core quiesce, the TCU subsequently stops the UltraSPARC T2 core's clocks.

The cycle when the stop occurs is a function of the value of the TCU cycle counter (refer to *TCU Cycle Counter Register* on page 456) as well as the transmission delay from the core to the TCU and from the TCU to the clock network in the core. If the TCU cycle counter is non-zero when the core generates a soft-stop request, the TCU will decrement the cycle counter until it reaches 0. When it reaches 0, the TCU will stop the core's clocks (note that it may take several cycles before the processor clocks stop after the counter reaches 0 due to the propagation delay from the TCU to the core clock network).

20.3.2 Action in Response to a Hard-Stop Event

If the DECR bits for an event are set to 10_2 and that event occurs, the UltraSPARC T2 core requests the TCU to stop the clocks as soon as the TCU cycle counter reaches 0. TCU does not wait for SPC to quiesce before stopping the clocks to the SPCs on a hard stop request.

Note | A hard stop request from any SPC or SOC would have TCU stopping the clocks to the entire chip.

20.3.3 Action in Response to an External Hard Stop

TRIGIN pin when asserted from the system will have TCU initiate a hard stop of UltraSPARC T2 based on the values programmed in the TCU Clock CLK Stop Delay register and TCU Clock Domain Stop register. Please refer to TABLE H-15 on page 981 for description of TCU Clock CLK Stop Delay register.

20.3.4 TCU Debug Event Counter Register

The 32-bit TCU Debug Even Counter register must be zero before the cycle counter is enabled. If it is non-zero, then each SPC debug event request received at the TCU will decrement it by 1, and each SOC debug event request received at the TCU will decrement it by 4; when zero is reached, the cycle counter will begin decrementing with the next debug event request. No differentiation is made regarding debug event requests, so it is up to the user to ensure that only one type of debug event is enabled when using the debug event counter. Debug event requests consist of soft stop, hard stop, and watchpoint requests from SPC, and hard stop, watchpoint requests from SOC blocks. These debug events have been described in ASI_DECR and SOC_DECR registers.

The Debug Event Counter is only recognized when TCU_DEBUG_CONTROL_REG.enable = 0. When TCU_DEBUG_CONTROL_REG.enable = 1, the debug event counter is disabled. The debug event counter is accessed with JTAG instruction TAP_DE_COUNT; default value upon reset is zero. Refer to *TCU Debug Control Register* on page 456 for the description of the TCU Debug Control register.

The format of the TCU Debug Event Counter register is shown in TABLE 20-12.

TABLE 20-12 TCU Debug Event Counter Register – TCU_DEBUG_EVENT_COUNTER_REG ($85\ 0000\ 0011 8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	counter	0	RW	Debug event counter.

20.3.5 TCU Cycle Counter Register

The 64-bit TCU Cycle Counter register counts off of UltraSPARC T2 core clock and can delay the response to a debug event. For example, if the TCU receives a hard-stop request, the cycle counter will begin counting down with each CMP clock cycle, and when it reaches zero, the hard stop will be performed. All debug event requests from the SPARC cores or a hard-stop request from SOC logic will be delayed by the cycle counter.

These actions are only valid when TCU_DEBUG_CONTROL_REG.enable = 0. For behavior when TCU_DEBUG_CONTROL_REG.enable = 1, see the next section. The cycle counter is loaded with JTAG instruction TAP_CYCLE_COUNT; default value on reset is zero.

The format of the TCU Cycle Counter register is shown in TABLE 20-13.

TABLE 20-13 TCU Cycle Counter Register – TCU_CYCLE_COUNTER_REG ($85\ 0000\ 0100_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	counter	0	RW	Cycle counter.

20.3.6 TCU Debug Control Register

The TCU has a 4-bit register to control responses to debug events, the TCU Debug Control Register (DCR). When bit 2 of TCU DCR is 0, the cycle counter and debug event counters perform as described above.

When bit 2 of the TCU DCR is set to 1, the lower 32 bits of the cycle counter are treated as a reset counter. In this mode, the reset counter is decremented with each core clock cycle after the power-on reset (POR) sequence ends. Once zero is reached, a watchpoint, a hard clock stop, or a clock stretch can be performed, or the upper 32 bits of the cycle counter can then be used. In this mode (bit 2 of TCU DCR = 1) the debug event counter will be ignored.

Note | A hard stop request from any SPC or SOC would have TCU stopping the clocks to the entire chip.

The behavior of the debug event and cycle counters is determined by the values in the TCU DCR as specified in TABLE 20-14. The TCU DCR is loaded with JTAG instruction TAP_TCU_DCR; default value on reset is 0.

The format of the TCU Debug Control Register is shown in TABLE 20-14.

TABLE 20-14 TCU Debug Control Register – TCU_DEBUG_CONTROL_REG ($85\ 0000\ 0108_{16}$)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3	soft_stop	0_{16}	RW	Enables all strands to soft-stop upon a debug event.
2	enable	0_{16}	RW	Trigger enable for the action field.
1:0	action	0_{16}	RW	Refer to Description in TABLE 20-15, below.

TABLE 20-15 TCU Debug Actions

soft_stop{3}	enable{2}	action{1:0}	Description
0/1	0	xx	Debug event and cycle counter recognize SPC debug events.
x	1	00	Watchpoint pulsed.
x	1	01	Hard stop and watchpoint pulsed.
x	1	10	Clock stretch and watchpoint pulsed.
x	1	11	Clock stretch and watchpoint, followed by hard stop and watchpoint.

The soft_stop bit 3 when set will cause TCU to soft-stop all SPCs when any SPC requests a soft stop. It is only active when bit 2 is 0. The following actions are valid when bit 2 of the TCU DCR is set to 1:

- **Watchpoint.** If the TCU DCR bits 2:0 are set to 100, then a single pulse of an external chip pin (TRIGOUT) will occur when the reset counter reaches zero. The upper 32 bits of the cycle counter are ignored.
- **Hard stop.** A hard clock stop will be performed if the TCU DCR bits 2:0 are set to 101, as specified in *Action in Response to a Hard-Stop Event* on page 455, and a watchpoint pulse generated, when the reset counter reaches zero. The upper 32-bits of the cycle counter are ignored.
- **Clock Stretch.** If the TCU DCR bits 2:0 are set to 110, then a clock-stretch signal will be pulsed out of the TCU when the reset counter reaches zero, and a watchpoint pulse will also be generated. The upper 32 bits of the cycle counter are ignored.

- **Clock Stretch then Hard Stop.** If the TCU DCR bits 2:0 are set to 111, then when the reset counter reaches zero, a clock stretch will be triggered and a watchpoint pulse will also be generated. Then the upper 32 bits of the cycle counter will be allowed to count down to trigger a clock hard stop and a second watchpoint will also be generated.

20.3.7 SW/JTAG Trigger Output Register

The format of the SW/JTAG Trigger Output register is shown in TABLE 20-16.

TABLE 20-16 SW/JTAG Trigger Output Register – TCU_TRIGOUT_REG ($85\ 0000\ 0110_{16}$)

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	trigout	0_{16}	RW	Pulses TRIGOUT when written to 1; reset automatically after being written.

20.4 Debug Port Support

UltraSPARC T2 has a 166-pin-wide debug port that is used as an observability vehicle to promote repeatability, tester characterization, chip debug, and general SPC and SOC debug. The debug port can be enabled through software and JTAG access to the Debug Port Configuration register, described in *Debug Port Configuration Register* on page 464. The debug port can be configured into any one of six observability modes. The following are the different observability modes of the debug port based on bits 3:1 of the Debug Port Configuration register:

- **000_2: SOC observability mode.** UltraSPARC T2 reset state (Reset State Machine Output), MCU,SII → L2,L2 → SIO signals to help chip debug (sent out on 159 pins).
- **001_2: Tester charac/SPC debug mode.** {SPC_id,thread_id} on per L2 bank basis and SPC instruction commit status on per SPC basis, sent out on 160 pins.
- **010_2: Repeatability mode.** SII and NCU inputs from DMU and NIU on debug port double-pumped on 166 pins.
- **011_2: Core and SOC debug.** SII and NCU inputs from DMU and SPC instruction commit status on per SPC basis.
- **100_2: NIU debug mode.** NIU debug signals on 32 output pins of debug port with 2 clk pins, 5 select signals on 5 input pins; rest of debug port not driven or driven but not sampled by LA. To debug NIU logic problems in the lab.

- **101_2: PCI_EX debug mode.** PCI_EX critical signals to debug PCI_EX logic problems. Sixteen from DMU @ iol2clk and 16 from PEU @ pcl2clk to be driven out on debug port as feedthroughs on 32 pins. The DMU signals will be retimed in MIO, but the PEU signals will go out as feedthroughs. Also the two clock signals (one for DMU, one for PEU) will be driven on the debug port for the LA to align the data against.
- **110_2–111_2:** Reserved for future use.

The following sections describe these modes in detail.

20.4.1 SOC Observability Mode

This mode will be used to capture a variety of critical SOC signals which will be helpful to debug UltraSPARC T2. The following is the breakup of the signals in this mode.

- 5-bit encoded state for reset state machine (has 20 states) from RST block to MIO block to monitor reset state on the tester and LA. Sent out at sys_clk frequency from reset block in UltraSPARC T2 (feedthrough in MIO) on 5 pins.
- Each MCU will send the following *new* signals to DBG block, which will be useful to debug MCU hangs/scheduler issues or MCU error handling issues on both FBDIMM channel errors and ECC errors. These signals will all be synchronized by MCU to the iol2clk domain and sent to DBG block. This leads to a total of 21 wires per MCU. Since there are 4 MCUs, this will lead to a total of 84 wires to DBG block from all MCUs together. DBG block will drive this information out on 84/2 = 42 pins of the debug port @ io2xclk.
  - mcu_dbg_rd_req_in_0{3:0} – Read request from L2 bank 0 to MCU (id + valid).
  - mcu_dbg_rd_req_in_1{3:0} – Read request from L2 bank 1 to MCU (id + valid).
  - mcu_dbg_rd_request_out{4:0} – Read ACK from MCU to L2 bank 0 or 1 (id + valid + dest_l2_bank).
  - mcu_dbg_wr_req_in_0 – Write request valid from L2 bank 0.
  - mcu_dbg_wr_req_in_1 – Write request valid from L2 bank 1.
  - mcu_dbg_wr_req_out{1:0} – 0, 1, 2, 3 writes completed at DRAM indication. (MCU dispatches up to a maximum of three writes on any cycle on two FBDIMM channels; then samples information coming FBDIMM channels to see if there were any errors. If no errors were reported, MCU interprets this as all writes completed.)
  - mcu_dbg_mecc_err – MCU has detected an MECC error on a L2 read or scrub.
  - mcu_dbg_secc_err – MCU has detected a SECC error on a L2 read or scrub.
  - mcu_dbg_fbd_err – MCU has detected a FBDIMM channel error.
  - mcu_dbg_err_mode – FBDIMM interface logic has gone into error handling mode. This bit stays on until error handling complete.

- SII and SIO will send the following signals to DBG block, which will be useful to debug L2 hang cases (SII sent DMA request to L2, L2 never sends an ACK or data return back):
  - sii_dbg_l2t[0:7]_req{1:0} – Req type encoded on 2 bits from sii to each l2t bank (00 – no request; 01 – RDD; 10 – WRI; 11 – WR8).
  - l2t[0:7]_dbg_sii_iq_dequeue – L2 dequeue from IQ.
  - l2t[0:7]_dbg_sii_wib_dequeue – L2 dequeue from IOWB.
  - l2b[0:7]_dbg_sio_ctag_vld – response valid from L2 to SIO.
  - l2b[0:7]_dbg_sio_ack_type – Read or write ACK from L2 to SIO
  - l2b[0:7]_dbg_sio_ack_test – ACK to DMU or NIU

Which leads to a total of $(7 \times 8) = 56$ wires for all L2 banks together @ 1.4 GHz to DBG block. DBG block will drive this information out on $56 \times 2 = 112$ pins of the debug port @ io2xclk.

Thus, the total number of debug pins that will be used up in the SOC observability mode will be 42 + 112 = 154.

20.4.2 Tester Characterization / SPC Debug Mode

The signals that will be observed on the debug port in this mode will be used for general SPC debug and tester characterization of multithreaded diagnostics and also for SPC speed binning on the tester. Each SPC will have four signals driven to the DBG block, and each L2 bank will have six signals driven to the DBG block. All these signals will be at CMP clk frequency, that is, 1.4 GHz nominal. Eight virtual processors and eight L2 banks will lead to a total of (4 + 6) x 8 = 80 signals at 1.4 GHz driven to the DBG block. Since the debug port will drive the signals out @ io2xclk, the DBG block will sample two consecutive cycles of these 80-bit wires and drive out 160 signals @ io2xclk to the debug pins for LA sampling.

For each SPC, these four wires are chosen as follows. Since each core has two thread groups, we have the following encoding per thread group, using 2 bits/thread group:

- 00_2 – Instruction noncommitted
- 01_2 – Control Transfer instruction committed in pipe
- 10_2 – Integer or FPU instruction committed in pipe
- 11_2 – Load/Store instruction committed in pipe

That is, every instruction committed per cycle in each thread group will be visible. for each L2 bank, the six wires are vcid{5:0} {spc_id{2:0}, thread_id{2:0}} of each crossbar packet to that bank on every cycle.

The combination of these two groups of signals will be adequate to keep track of execution of instructions in both single and multithreaded diagnostics on the tester and also could be useful for SPC speed binning on the tester.

20.4.3 Repeatability Mode

In this mode, a total of 353 signals (in iol2clk domain) will be routed to DBG block (from NIU and DMU). From , 166 wires will get driven at io2xclk to the debug pins. These signals capture both inbound DMA and PIO returns from NIU and PCI_EX blocks in UltraSPARC T2 to SII and NCU and will be used as bus traces for checkpoint/replay scheme in UltraSPARC T2. These 353 signals and rate conversion to debug port frequency are listed below.

1. dmu_ncu_wrack_vld;
2. dmu_ncu_wrack_tag{3:0};
3. dmu_ncu_stall; // total 6 bits @ iol2clk = 3 pins @ io2xclk (DDR)
4. dmu_ncu_vld;
5. dmu_ncu_data{31:0}; // 33 bits get driven over four clocks. Eight clocks minimum before next set of four clocks, so total of 132 bits to be emptied over 12 iol2clks, that is, 66 bits DDR over 12 clocks, that is, 6 pins @ io2xclk (DDR).
6. niu_ncu_stall;
7. niu_ncu_vld;
8. niu_ncu_data{31:0}; // 34 bits @ iol2clk = 17 pins @ io2xclk (DDR)
9. dmu_sii_hdr_vld;
10. dmu_sii_reqbypass;
11. dmu_sii_datareq;
12. dmu_sii_datareq16;
13. dmu_sii_data{127:0};
14. dmu_sii_be{15:0}; // 148 bits @ iol2clk = 74 pins @ io2xclk (DDR)
15. niu_sii_hdr_vld;
16. niu_sii_reqbypass;
17. niu_sii_datareq;
18. niu_sii_data{127:0};
19. niu_sio_dq; // 132 bits @ iol2clk = 66 pins @ io2xclk (DDR)

Total = 66 + 74 + 17 + 3 + 6 = 166 pins @ io2xclk (DDR)

20.4.4 Core and SOC Debug mode.

SII and NCU inputs from DMU and SPC instruction commit status on per SPC basis:

1. dmu_ncu_wrack_vld;
2. dmu_ncu_wrack_tag{3:0};
3. dmu_ncu_stall; // total 6 bits @ iol2clk = 3 pins @ io2xclk(DDR)
4. dmu_ncu_vld;
5. dmu_ncu_data{31:0}; // 33 bits get driven over four clocks. Eight clocks minimum before next set of four clks, so a total of 132 bits will be emptied over 12 iol2clks, that is, 66 bits DDR over 12 clocks, that is, 6 pins @ io2xclk (DDR)
6. dmu_sii_hdr_vld;
7. dmu_sii_reqbypass;

8. dmu_sii_datareq;
9. dmu_sii_datareq16;
10. dmu_sii_data{127:0};
11. dmu_sii_be{15:0}; // 148 bits @ iol2clk = 74 pins @ io2xclk (DDR)

Each SPC will have four signals driven to the DBG block at core clk frequency. Since there are eight cores, this will lead to a total of (4) x 8 = 32 signals @ l2 clk driven to DBG block. Since the debug port will drive the signals out @ io2xclk, the DBG block will sample two consecutive cycles of these 32-bit wires and drive out 64 signals @ io2xclk to the debug pins for LA sampling.

For each SPC, these four wires are chosen as follows.

Since each core has 2 thread groups, we have the following encoding per thread group using 2 bits/thread group:

- 00_2 – Instruction non committed
- 01_2 – Control Transfer instruction committed in pipe
- 10_2 – Integer or FPU instruction committed in pipe
- 11_2 – Load/Store instruction committed in pipe

Thus total number of pins used to drive out debug info in this mode = (3 + 6 + 74 + 64) = 147.

20.4.5 NIU Debug Mode

In this mode, debug port will not run at io2xclk but will just act as a feedthrough for NIU debug signals (NIU's own debug port). The NIU will have its own 32-wire debug port accompanied by two clock signals, where 32 wires will come from up to two different clock domains in NIU. The maximum frequency is, however, iol2clk frequency.

The NIU debug port will output state machine states and flags (for example, FIFO Full and empty) from internal blocks that will be needed to debug internal blocks. There will also be some statistics for tracking packets during debug.

NIU_debug_interface = {niu_mio_debug_clock{1:0}, niu_mio_debug_data{31:0}}

niu_mio_debug_data{31:0} contains up to two buses worth of control signals /state machine states, each of which could be at a different frequency.

niu_mio_debug_clock{1:0} carries up to two clocks that debug_data{31:0} reference.

Also NIU will need 5 input signals to select the state machines to display on these 32 pins . In this exclusive NIU debug mode configuration, these 5 NIU input debug mode select signals will be driven from DBG1 block from "niu_dbg_sel" CSR bits of Debug Port Configuration register in TABLE 20-18 on page 464.

TABLE 20-17 below shows the NIU debug mode select encodings.

TABLE 20-17 NIU Debug Mode Select Encodings

Debug Mode	test_sel	debug_clk
Functional mode	00000_2	{0,0}
Reserved	00001_2 - 10000	N/A for UltraSPARC T2
Debug_TXC	10001_2	{iol2clk,iol2clk}
Debug_TDMC	10010_2	{iol2clk,iol2clk}
Debug_RDMC	10011_2	{iol2clk,iol2clk}
Debug_ZCP	10100_2	{iol2clk,iol2clk}
Debug_IPP	10101_2	{iol2clk,iol2clk}
Debug_FFLP	10110_2	{iol2clk,iol2clk}
Debug_PIO	10111_2	{iol2clk,iol2clk}
Debug_MAC	11000_2	{iol2clk,iol2clk}
Reserved	11001_2	N/A for UltraSPARC T2
Debug_Meta	11010_2	{iol2clk,iol2clk}
Debug_SMX	11011_2	{iol2clk,iol2clk}
Reserved	11100_2 - 11111_2	N/A for UltraSPARC T2

20.4.6 PCI_EX Debug Mode

In this mode, debug port will not run at io2xclk. The PCI_EX will have its own 32 wire debug port, where 16 wires will come from DMU @ iol2clk and 16 from PEU at pc_clk. The 16 wires from DMU will be retimed in MIO before being sent out, but the PEU wires will go out as feedthrough.

These wires from DMU and PEU will be used to carry internal state machine information out to an LA for stand-alone debug of PCI_EX logic problems. UltraSPARC T2's PCI-Express subsystem (DMU and PEU) will provide 32 (16 for each unit) debug signals to DBG block. Each of DMU and PEU will have two debug buses (A and B): eight bits per bus. Each debug bus will be controlled by its own debug select register which is software programmable and will be programmed to output a variety of internal state machine and critical debug information for DMU and PEU sub-blocks. These buses are listed below.

- peu_mio_debug_bus_a{7:0}
- peu_mio_debug_bus_b{7:0}

- dmu_mio_debug_bus_a{7:0}
- dmu_mio_debug_bus_b{7:0}

Some of the DMU/PEU subblocks that are currently POR to be on these debug buses are CLU, CMU, CRN, DSN, ILU,IMU, MMU, PMU, PSB, RMU, TMU, SB, ETL, ITL, PMC, RSB, CTB, CSPL.

Beyond these wires, two clocks will also be sent to MIO block from DMU and PEU to help the LA to align the DMU and PEU signals with respect to the DMU and PEU clock. These two clock signals will be called dmu_mio_debug_clk and peu_mio_debug_clk.

20.4.6.1 Debug Bus A

The PCI-Express controls for debug bus A are described in *PEU Debug Select A Register (0068300016 / 016)* on page 593.

20.4.6.2 Debug Bus B

The PCI-Express controls for debug bus B are described in *PEU Debug Select B Register (0068300816 / 016)* on page 593.

20.4.7 Debug Port Configuration Register

The Debug Port Configuration register enables the debug port in one of six modes.

TABLE 20-18 shows the format of the Debug Port Configuration register.

TABLE 20-18 Debug Port Configuration – DEBUG_PORT_CONFIG ($86\ 0000\ 0000_{16}$)

Bit	Field	Initial Value	R/W	Description
63:62	imp_ctrl	0 (preserved on WMR/ DBR)	RW	MIO driver impedance control: 11 – Strong driver 10 – Nominal driver 01 – Weak driver 00 – Low power driver
61:10	—	0	RO	*Reserved*
9:5	niu_dbg_del	0 (preserved on WMR/ DBR)	RW	NIU Debug select bits

TABLE 20-18 Debug Port Configuration – DEBUG_PORT_CONFIG ($86\ 0000\ 0000_{16}$) *(Continued)*

Bit	Field	Initial Value	R/W	Description
4	debug_train	0 (preserved on WMR/ DBR)	RW	When set to 1, enables training for Debug port in modes 000, 001, 010, and 011.
3:1	debug_conf	0 (preserved on WMR/ DBR)	RW	Debug Port Configuration: 000 – SOC observability 001 – Tester characterization/SPC debug 010 – Repeatability 011 – CORE_SOC debug 100 – NIU debug 101 – PCI Express debug 110–111 – *Reserved*
0	debug_en	0 (preserved on WMR/ DBR)	RW	When set to 1, enables Debug port output drivers.

20.4.8 Debug Port Training Sequence

The data coming out from UltraSPARC T2 on the debug port needs to be aligned properly in the Logic Analyzer connected to the debug port after cancelling out signal to clock and signal to signal skews. To support this, UltraSPARC T2 would provide training sequence in all the debug port modes. For all modes other than the NIU debug mode, this sequence will be a repetitive pattern of three 1's, followed by a single 0. This asymmetrical pattern will ease the alignment and deskewing of the data bits in the LA in case the skew for some bits is as large as one cycle. For the NIU debug mode signals, the training pattern is programmable by software programming any training pattern it wants in the training_vector{31:0} register in the specific NIU block.

20.4.9 IO Quiesce Control Register

The IO Quiesce Control register format is shown in TABLE 20-19.

TABLE 20-19 IO Quiesce Control Register – IO_QUIESCE_CONTROL ($86\ 0000\ 0008_{16}$)

Bit	Field	Initial Value	R/W	Description
63:4	—	X	RO	*Reserved*
3	niu_stall_done	X	RO	Set to 1 by hardware when NIU stall completes in hardware. Cleared by hardware when NIU_STALL cleared from 1 to 0 by software.

TABLE 20-19 IO Quiesce Control Register – IO_QUIESCE_CONTROL ($86\ 0000\ 0008_{16}$)

Bit	Field	Initial Value	R/W	Description
2	dmu_stall_done	X	RO	Set to 1 by hardware when DMU stall completes in hardware. Cleared by hardware when DMU_STALL cleared from 1 to 0 by SW.
1	niu_stall	0 (preserved on WMR/ DBR)	RW	When set to 1, causes NIU traffic to stall. When cleared to 0 from 1, causes NIU traffic to resume.
0	dmu_stall	0 (preserved on WMR/ DBR)	RW	When set to 1, causes DMU traffic to stall. When cleared to 0 from 1, causes DMU traffic to resume.

20.5 Repeatability Support

To effectively run processor tests in the post-silicon phase with or without the presence of I/O (on UltraSPARC T2's XAUI and PCI_EX interfaces) and to debug them, we need to have a high level of repeatability within UltraSPARC T2's synchronous clock domains. These include the core clock domain (covers SPARC cores, crossbar, L2's, portions of SII, SIO, NCU), the DRAM domain (covers MCU logic before SerDes), and the I/O clock domain (rest of SII, SIO, NCU).

This will allow us to run a group of tests many times, with slightly different starting parameters (for example, SPARC threads starting at slightly different times or with different cache initialization) that shouldn't affect the outcome, looking for failing corner cases. When a failing case is found, the test and the particular seed parameters will be used to simulate the test in the pre-silicon environment, to see what caused the failure.

The overall approach involves very close interaction between some debug software (part of Hypervisor software) and UltraSPARC T2 chip hardware. This is commonly known as the checkpoint/replay mechanism, where the debug software will periodically put the synchronous portion of the chip (as described before) into an `idle` state (idle all threads other than one, and also stall I/O into the synchronous domain) at what are called checkpoints. Once the synchronous portion of the chip is put into this `idle` state, the debug software will dump all software-visible state of the machine to memory, and then initiate a debug reset of UltraSPARC T2.

The debug software will initiate by writing a 1 to the dbr_gen bit in reset_gen register.

20.5.1 Debug Reset

The debug reset is a flavor of l in UltraSPARC T2 which is identical to the functionality of warm reset other than it will not reset the NIU and PCI_EX blocks and the MCU FBDIMM links. (The NIU and PCI_EX block can keep running, the debug software has already quiesced their traffic to the SII and NCU blocks, so for NIU some packets will be dropped, and for PCI_EX back pressure will be exerted on the external device).

Notes

Program counter continues to advance for some time after the debug reset request from the RST block till the reset takes effect on the SPC blocks.

PCI_EX device has a hardcoded timeout of 50 msec, so the duration of the back pressure exerted by UltraSPARC T2 during checkpoint and debug reset should be well below 50 msec so as not to starve any access from the device by more than that amount (to avoid PCI_EX errors and system panics). This means that checkpoint snap, debug reset, FBDIMM initialization and resumption of I/O traffic to synchronous domain of UltraSPARC T2 should finish well before 50 msec.

UltraSPARC T2, like previous Sun processors, keeps a fair amount of architected state unchanged for warm reset. Also contents of arrays (TLBs, L1/L2 caches, etc.) are unchanged. Please refer to Table 11-13 in section 11.8 for a list of UltraSPARC T2 software-visible state that will be lost over debug reset and will need to be retrieved after debug reset. So before invoking debug reset, software should copy the software-visible state that loses value over debug reset to memory, and retrieve it back from memory after the reset.

The duration of the debug reset is small enough (in the range of 40 μsecs), so to address the data integrity in DRAM during the debug reset, UltraSPARC T2 will either (1) address it through self-refresh during the debug reset or (2) autorefresh in small intervals before going to debug reset and then doing some in small intervals after coming out of debug reset. Besides, UltraSPARC T2 would keep the FDIMM links up during debug reset and keep issuing sync pulses to the AMBs.

After debug reset, the reset vector will be fetched from memory from a different location ($00000002 0_{16}$) from a regular reset. This is because the boot code for a debug reset will be different from a regular reset. The boot code will do several things at the beginning including program the Memory refresh registers, reinstate the software-visible state to the state before reset for those states that lose value over debug reset), and remove the stall of inbound I/O to the synchronous domain of the chip from NIU and PCI_EX, before enabling all threads to start executing.

In normal operation POR and warm reset both trap to the *RSTVADDR* | 20_{16} (FFFF FFFF F000 0020_{16}), which maps to ROM. To enhance repeatability, UltraSPARC T2 will have the capability of directing POR, WMR, or DBR to RAM. To POR or WMR or DBR from RAM at location ($0\ 0000\ 0020_{16}$), hyperprivileged software can set the `ASI_WMR_VEC_MASK` register.

The idea is that by capturing the software-visible state of UltraSPARC T2 (in the synchronous domain of the chip) on the last checkpoint before the failure and by initializing the synchronous portion of UltraSPARC T2 to known state, we can create a common starting point between silicon and the synchronous portion of the chip in the pre-silicon environment. Then by running the same code sequence on the SPARC core from the last checkpoint to the failure point and capturing the I/O traffic to the SII, NCU inputs (synchronous I/O interface of UltraSPARC T2: debug port mode 000_2) from DMU, NIU on the debug port lossless and feeding it back to the same nodes in the pre-silicon environment, we can create the event sequence in pre-silicon environment leading to the failure.

Note | For checkpoint/replay, we do not need to observe the FBDIMM interface on the debug port. This is because once the links are trained, data will always come back to the MCU data return FIFO in a fixed latency from the time of issue of the request. After link training, MCU logic will record this latency (in terms of MCU clocks) in the MCU Channel Read Latency register. So the debug software can probe this value and feed that same latency to the equivalent point in the pre-silicon environment and thereby achieve cycle accuracy with respect to silicon without having to probe the FBDIMM interface.

Thus, checkpoint/replay approach intrudes on the state of the machine in the context of the tests or applications running on the chip, in that it periodically halts all threads and I/O and takes the machine to `reset` state. This might change the timing of events to cause the bug to manifest itself later than usual, but eventually it will, with millions of cycles of instructions executed in between checkpoints. And when it does, it can be recreated in pre-silicon environment.

20.5.2 Keeping FBDIMM Links Up During Debug Reset

If debug reset did reset the whole MCU, the FBDIMM links will have to be retrained after reset deassertion, and this will change the FBDIMM data round-trip latency for subsequent requests until the next debug reset. Debug software can live with this by reading the MCU Channel Read Latency register after every debug reset, or MCU needs to keep sending sync pulses during the debug reset.

To support the latter, MCU will keep a small amount of logic running during warm/debug reset while the rest of it gets reset through the flush mechanism. This logic will consist of (1) logic to keep the links enabled and generate sync pulses in a fixed

repetitive manner under software control and (2) logic to keep incrementing the read pointers of the northbound MCU FIFOs and two synchronizers per FIFO (this way, during debug/warm reset, the read and the write pointers constantly increment and are always offset by 2: delay through the two synchronizers).

Also each MCU would support two new CSR bits, located in the DRAM Debug Event Trigger Enable register for software to control this feature:

- **kp_lnk_up.**

  When written to $_2$,

  - Keeps the southbound links enabled during the duration of the debug reset to send out the sync pulses.
  - Selects the output of the sync pulse gen logic in the new MCU control module to generate sync pulses.

  When written to 0,

  - Selects the output of the regular sync pulse gen logic in MCU.
  - Clears the counter for the regular sync pulse gen logic in MCU.
  - Takes the MCU FBDIMM interface state machine to L0 state, where it is ready to dispatch new read/write requests to the DIMMs.

- **mask_err.**

  When written to 1,

  - Makes MCU mask all the errors it normally detects on LFSR mismatches on ALERT frame patterns coming in from AMB.

  Cleared by MCU hardware 4K cycles after reset when the LFSRs are realigned by the MCU.

Note | Both kp_lnk_up and mask_err bits are protected on warm reset/ debug reset.

Thus, the interaction between software and hardware to achieve determinism on FMDIMM interface after debug reset is as follows:

1. After making sure no transactions are pending in MCU, software sets kp_lnk_up and mask_err just before initiating debug reset.
2. Debug reset happens. The entire MCU gets reset other than the control logic module, which has its clock running keeping the sync pulses going and the FIFO read pointer incrementing every cycle.
3. Debug reset finishes. The MCU FBDIMM interface state machine comes up in DISABLED state. Sync ACKs keep coming, but since the mask_err bit is set, no errors are flagged. MCU logic counts 4K cycles after reset and realigns the LFSRs and clears the mask_err bit.

4. After a certain time T1 (but always fixed from the deassertion of debug reset), software writes a 0 to the kp_lnk_up bit. This clears the sync pulse gen counter, takes the FBDIMM interface state machine to L0 state, and selects the sync pulse gen counter output to generate the sync pulses.
5. After a time T2 from the point where software wrote kp_lnk_up with 0, the first fetch is issued on the southbound link. T2 should be the same all the time.

Note | The FBDIMM links would be retrained after every warm reset.

20.5.3 I/O Quiescing in UltraSPARC T2 During Checkpoint

An inherent requirement for checkpoint/replay in UltraSPARC T2 is to stall I/O to the synchronous domain of the chip (SII and NCU inputs) from NIU and PCI_EX blocks. This is part of the effort to get the chip to a quiescent state on every checkpoint before dumping software-visible state and asserting debug reset to get the synchronous portion of the chip to a known state.

This I/O quiescing will get implemented in UltraSPARC T2 under software control by having the DBG module contain a couple of CSR bits (niu_stall and dmu_stall) in UltraSPARC T2 I/O Quiesce Control register (refer to *IO Quiesce Control Register* on page 465) which software can set to 1's by writing a 1 to them. Once these bits are set, the debug module in UltraSPARC T2 will assert a couple of signals, called dbg_niu_stall and dbg_dmu_stall to NIU and DMU, respectively.

On seeing the assertion of these two signals, NIU and DMU should suspend all transactions to SII and NCU at any convenient point (for NIU, this can be at a packet boundary—whatever is easy to implement and creates the fewest corner cases) and send back niu_dbg_stall_ack and dmu_dbg_stall_ack to the debug module after they have received all pending ACKs and data returns from SIU and NCU. At the point at which these two ACKs are sent to the DBG block, the NIU → SII, NCU and DMU → SII, NCU interfaces will be considered as having quiesced. This applies to interrupts also. Neither DMU nor NIU should send any interrupt requests to NCU or SII after having sent the ACKs. On sampling niu_dbg_stall_ack and dmu_dbg_stall_ack signals, DBG block will set niu_stall_done and dmu_stall_done bits in the two I/O Quiesce Control registers. The debug software that will have been polling these status bits will then see that both bits are set and will proceed to dump software-visible state of machine to memory and then initiate a debug reset.

Note that even during the time this interface is quiesced, the Xaui and PCI_EX interface SerDes links are active and running.

NIU will be dropping packets but is guaranteed never to corrupt any data on Rx and Tx DMA. Also, in case NIU drops packets and does not send ACKs back to the device, the driver will retry. The whole NIU/networking architecture is built on the

assumption that there will never be any back pressure exerted to the device. So packets can be dropped and the driver will retry. Also, all errors recorded in NIU are sticky (persist until cleared by software) and all interrupts issued by NIU log the event causing the interrupts in NIU CSRs that are software-visible. So even if interrupts cannot go through and are dropped during the quiescent time, software can come later and examine the error bits and ping NCU to dispatch interrupts to SPARC core. Also, with respect to multiple interrupts issued from the same source, NIU logs the event for the first interrupt only. So it is okay to react to the first interrupt and ignore the following ones.

While PCI_EX interface with SIU and NCU is quiesced, the PCI_EX block will keep the link running but will exert back pressure to the device in the following way. In the PCI-Ex link, a flow control mechanism prevents the FIFO overflow. When the Ingress Queue is full, the PEU will stop updating the flow control credit to the remote device. When the remove device runs out of credit and does not receive any new credit from PEU, it stops sending any new packets. Also, all PCI_EX errors are sticky and, as long as the CSR bits are not cleared, will generate interrupts eventually. In case there are interrupts pending to be dispatched to SII from DMU after DMU is asked to stall, the interrupt packet will be stuck at some queue in DMU. When the interface is later unquiesced, the interrupt will be able to move upstream again and reach SII eventually.

After debug reset, the reset code will clear the niu_stall and dmu_stall CSR bits in DBG block, which will cause DBG block to assert a couple of signals to NIU and DMU called dbg_niu_resume and dbg_dmu_resume. On receiving these “resume” signals, NIU and DMU will unquiesce their respective interfaces with SII and NCU and continue issuing transactions to SII and NCU.

20.6 Clock/PLL Observability

UltraSPARC T2 has two pins called PLL_CHAR_OUT{1:0} dedicated for PLL/clock observability output.

CHAPTER 21

PCI Express Interface Unit (PIU)

21.1 Overview

The PCI Express (PCIE) interface integrates the functionality of a host to PCI Express bridge onto the UltraSPARC T2 system-on-a-chip. In PCI Express terminology [1] this is called a PCI Express Root Complex. This allows an I/O subsystem based on PCI Express to be connected to an UltraSPARC T2 processor in flexible configurations using standard PCI Express components such as

- PCI Express to PCI Express bridges and switches;
- PCI Express to PCI and/or PCI Express to PCI-X bridges;
- Native PCI Express devices.

The use of7 PCI Express allows the use of inexpensive commodity components while supporting excellent bandwidth to high throughput I/O devices. Interoperability with PCI and PCI-X devices is achieved through bridges, and compatibility with existing software is achieved by using standard PCI abstractions. The signaling technology of PCI Express has high bandwidth per pin, giving a low pin count and reduced system cost.

21.1.1 Supported Features

UltraSPARC T2 supports one x8 PCI Express port operating at 2.5 GHz. PCI Express uses 8-bit/10-bit encoding, full-duplex point-to-point links, and each lane carries 1 bit of data per cycle. The maximum theoretical bandwidth per lane is therefore 2.0 Gbit/s (in each direction of the point-to-point link). Since UltraSPARC T2 has eight lanes, this is a total of 16.0 Gbit/s or 2.0 Gbyte/s (per direction).

The PCI Express Interface has the following features:

- Integrated into the UltraSPARC T2 system-on-chip.
- One x8 2.5GHz PCI Express port with internal SerDes.

- Maps PCI Express configuration, PCI Express I/O and PCI Express 32-bit and 64-bit memory address spaces into UltraSPARC T2 physical address space (downbound operation).
- Maps UltraSPARC T2 physical address space into PCI Express memory address space (upbound operation). Both bypass and translated accesses through an MMU are supported. Support the hypervisor requirement for device specific translation.
- Translates PCI Express MSI/MSI-X interrupts into UltraSPARC T2 mondo interrupts.
- PCI Express legacy INTX support.
- PCI Express Power Management support.

The design of the UltraSPARC T2 PCI Express interface includes three cores:

- **DMU (Data Management Unit).** This is derived from the DMC (Data Management Core) block of the Fire ASIC. In general, the DMU is compatible with the software-visible feature set of the Fire ASIC DMC. This includes architecture, programming model, memory mappings, registers, IOMMU in sun4u mode, error handling, and interrupt architecture. DMU also includes a new glue logic block called DSN(DMC ↔ SIU/NCU). DSN performs the protocol modification for the interface between DMC ↔ SIU and DMC ↔ NCU.
- **PEU (PCI Express Unit).** This is imported as a third-party IP core that implements the lower levels of the PCI Express protocol. The physical layer, data link layer and some basic function of transaction layer will be provided from this IP core. The major transaction layer functions and its associated registers will be primary from the Fire ASIC TLU 1.0. A glue logic unit—TLG—is developed and placed between the third-party IP core and Fire ASIC TLU 1.0.
- **PSR (PCI-Express SerDes).** TI provides the PCI-Express SerDes macro for UltraSPARC T2. Please refer to the WIZ6C2xxN5x0 Data sheet for SerDes's detailed information. A glue logic unit, PCS, is developed between PSR and the third-party IP core.

The combination of the DMU, PEU, and PSR is called the PCI Express Interface Unit, abbreviated to PIU.

Note that the UltraSPARC T2 DMU will be clocked at 350 MHz and the PEU at 250 MHz, with the crossing between these two clock domains inside PEU.

21.1.2 Features Not Supported

The following limitations are inherited from the Fire ASIC:

- No isochronous transfers (only one virtual channel is supported).
- No lock operations.
- No wake system from PME (that is, no Vaux support).

- No root complex configuration space. Specifically, this means that there is no Root Complex Register block (optional in PCI Express) and no PCI Express Capability Structure associated with a Root Port.

21.1.3 PIU Specification Relative to Fire ASIC

This section documents the changes in the UltraSPARC T2 PIU relative to the Fire ASIC. The reference is version 2.0 of the Fire Programmer's Reference Manual [2].

The following features are not present in the PIU:

- EBUS interface, GPIO pins, and I2C ports are removed.
- The 40 external interrupt pins are removed.
- JBUS interface, JBC (JBUS Core), and JBUS coherency mechanisms are removed. All JBC CSRs are removed from the UltraSPARC T2 PCI Express Interface, though some are moved to other UltraSPARC T2 units (see the chapter's introductory paragraphs). In particular, the Fire ASIC JBC registers that control the partitioning of the 64-Gbyte noncacheable PIO space are moved out to the UltraSPARC T2 NCU. Also, the registers and mechanisms associated with JBUS reset are removed. Reset functionality on UltraSPARC T2 is the responsibility of the Reset Unit.
- LPU (Link Physical Unit) and the CSRs associated with the LPU are removed. Similar features are provided by the CXPL core from external IP vendor (see Section 16.3, "PEU Specification").
- All physical addresses are reduced from 43 bits to 40 bits.
- Only PCI Express Port A is supported on UltraSPARC T2, and it is an x8 link.
- PCI Express Port B is removed. All CSRs and datapaths associated with PCI Express Port B are removed.
- The NCU supports only 1-, 2-, 4-, and 8-byte-aligned PIO accesses to DMU space. The Fire ASIC supports 16-byte and 64-byte JBUS accesses to the PCI MEM32/ MEM64 subregions of the 64-Gbyte noncacheable region. There is no equivalent mechanism on UltraSPARC T2.

The following features are added relative to the Fire ASIC:

- The JBUS interface is replaced with a direct connection to the physical address space of UltraSPARC T2. The registers and address spaces of the DMU appear directly in the UltraSPARC T2 noncacheable memory space (via the UltraSPARC T2 NCU). The DMU has direct access to the UltraSPARC T2 memory system for DMA and DVMA (via the UltraSPARC T2 SIU). The register and address space map is relocated within the UltraSPARC T2 physical address space (compared with an UltraSPARC T1 chip and external Fire ASIC), but otherwise the relative properties of the register address map are retained. In most cases, register bit-field assignments are unchanged for registers that are common to both the Fire ASIC and UltraSPARC T2.

- The IOMMU is extended with a new sun4v mode to support hypervisor requirements. This allows partitioning of the I/O devices and protection of DMA memory. The Fire IOMMU behavior is retained in the backward-compatible sun4u mode. sun4v mode provides the following additional features:
  - Added support in IOMMU for X-PAR
  - Memory translation for 64-bit addresses

For more information on the IOMMU changes see Section 21.3.6, *sun4v Mode IOMMU*, on page 501.

The following features are provided in a similar way to the Fire ASIC but with some necessary changes:

- Interrupt mondo sources corresponding to external interrupt pins, I2C and JBC are removed. The interrupt numbers (INO) of the remaining interrupts are retained without change. It is desirable to separate interrupts from DMU versus those from PEU since PEU is imported IP from a third-party. INO 62 is used for DMU and 63 (previously JBC) is used for PEU.
- The Fire ASIC can steer mondo interrupts to any of 32 processors via JBUS. To support the 64 threads on UltraSPARC T2, the field t_jpid in the interrupt mapping registers must be expanded from 5 bits to 6 bits. On the Fire ASIC, t_jpid occupies bits 30:26, while on UltraSPARC T2 it will be 30:25 (bit 31 is already used) and it is renamed to threadid.
- Physical addresses on JBUS are 43 bits, while on UltraSPARC T2 they are 40 bits. A number of CSRs hold physical addresses and these are resized to 40 bits. The affected DMU registers include:
  - MMU TSB Control register
  - 64 x MMU TTE Cache Physical Tag registers
  - 64 x MMU TTE Cache Data registers
  - MMU TTE Cache Flush Address register
- The decoding of PCI-Express 64-bit address space to implement bypass mode will be
  - Bits 63:50 – all set to 1.
  - Bits 49:39 – all set to 0.
  - Bits 38:0 – 39-bit address in UltraSPARC T2 main memory.
- The names JBC, PEC and LPU should be eliminated from the PRM. Other sub-units like IMU, MMU, ILU and PEU will probably be retained in register names.

21.1.4 Address Map

The Fire ASIC provides three addressable regions in JBUS address space. A JBUS physical address is 43 bits decoded as follows:

1. An 8-Mbyte noncacheable region used for PIO access to Fire ASIC CSRs:

2. A 64-Gbyte noncacheable region that maps to subregions for PCI Express configuration, PCI Express I/O, and PCI Express 32-bit and 64-bit memory address spaces:

3. A 64-Gbyte cacheable region that is not used.

The 8-Mbyte and 64-Gbyte noncacheable regions are supported by UltraSPARC T2, though they have different base addresses than the Fire ASIC. Within these regions the address decoding on UltraSPARC T2 is identical to the Fire ASIC. Thus, CSRs inside the 8-Mbyte noncacheable region, and subregions inside the 64-Gbyte noncacheable regions, have the same offsets (relative to the appropriate base) on UltraSPARC T2 as on the Fire ASIC.

21.1.5 PIU CSR Summary

TABLE 16-2 summarizes the registers contained in the PCI Express interface compared with the Fire ASIC PRM [2]. The first column contains the offset of the register, the second column describes the register and the third column indicates whether the register has been added (A), modified (M), removed (R), or is unchanged (U) relative to the Fire ASIC PRM. If the status of the register has not yet been determined or is in flux, then it is marked with "?".

All accesses to these registers must be 8 byte wide, 8 byte aligned. The base address of these registers in the UltraSPARC T2 physical address map is 88 0000 0000_{16}.

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(1 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
PCIE CSR Space		
600000_{16}–$600FF8_{16}$	*Reserved*	—
601000_{16}–601098_{16}	Interrupt Mapping registers (Mondo 0 - 19)	R
$6010A0_{16}$–$6011D8_{16}$	Interrupt Mapping registers (Mondo 20 - 59)	M (6-bit t_jpid)
$6011E0_{16}$–$6011E8_{16}$	Interrupt Mapping registers (Mondo 60 - 61)	R
$6011F0_{16}$–$6011F8_{16}$	Interrupt Mapping registers (Mondo 62 - 63)	M (6-bit T_JPID)
601200_{16}–$6013F8_{16}$	*Reserved*	—
601400_{16}–601598_{16}	Interrupt Clear registers (Mondo 0–19)	R
$6014A0_{16}$–$6015D8_{16}$	Interrupt Clear registers (Mondo 20–59)	U
$6015E0_{16}$–$6015E8_{16}$	Interrupt Clear registers (Mondo 60–61)	R
$6015F0_{16}$–$6015F8_{16}$	Interrupt Clear registers (Mondo 62–63)	U
601600_{16}–$6019F8_{16}$	*Reserved*	—
$601A00_{16}$	Interrupt Retry Timer register	U

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(2 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
$601A08_{16}$	*Reserved*	—
$601A10_{16}$	Interrupt State Status register 1	U
$601A18_{16}$	Interrupt State Status register 2	U
$601A20_{16}$–$60AFF8_{16}$	*Reserved*	—
$60B000_{16}$	INTX Status register	U
$60B008_{16}$	INT A Clear register	U
$60B010_{16}$	INT B Clear register	U
$60B018_{16}$	INT C Clear register	U
$60B020_{16}$	INT D Clear register	U
$60B028_{16}$–$60FFF8_{16}$	*Reserved*	—
610000_{16}	Event Queue Base Address register	U
610008_{16}–$610FF8_{16}$	*Reserved*	—
611000_{16}–611118_{16}	Event Queue Control Set register	U
611120_{16}–$6111F8_{16}$	*Reserved*	—
611200_{16}–611318_{16}	Event Queue Control Clear register	U
611320_{16}–$6113F8_{16}$	*Reserved*	—
611400_{16}–611518_{16}	Event Queue State register	U
611520_{16}–$6115F8_{16}$	*Reserved*	—
611600_{16}–611718_{16}	Event Queue Tail register	U
611720_{16}–$6117F8_{16}$	*Reserved*	—
611800_{16}–611918_{16}	Event Queue Head register	U
611920_{16}–$61FFF8_{16}$	*Reserved*	—
620000_{16}–$6207F8_{16}$	MSI Mapping register	U
620800_{16}–$627FF8_{16}$	*Reserved*	—
628000_{16}–$6287F8_{16}$	MSI Clear registers	U
628800_{16}–$62BFF8_{16}$	*Reserved*	—
$62C000_{16}$	Interrupt Mondo Data 0 register	M
$62C008_{16}$	Interrupt Mondo Data 1 register	U
$62C010_{16}$–$62FFF8_{16}$	*Reserved*	—
630000_{16}	ERR COR Mapping register	U
630008_{16}	ERR NONFATAL Mapping register	U
630010_{16}	ERR FATAL Mapping register	U
630018_{16}	pm PME Mapping register	U
630020_{16}	PME To ACK Mapping register	U
630028_{16}–$630FF8_{16}$	*Reserved*	—

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(3 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
631000_{16}	IMU Error Log Enable register	U
631008_{16}	IMU Interrupt Enable register	U
631010_{16}	IMU Interrupt Status register	U
631018_{16}	IMU Error Status Clear register	U
631020_{16}	IMU Error Status Set register	U
631028_{16}	IMU RDS Error Log register	U
631030_{16}	IMU SCS Error Log register	U
631038_{16}	IMU EQS Error Log register	U
631040_{16}–$6317F8_{16}$	*Reserved*	—
631800_{16}	DMC Core and Block Interrupt Enable register	M
631808_{16}	DMC Core and Block Error Status register	U
631810_{16}	Multi Core Error Status register	R
631818_{16}–$631FF8_{16}$	*Reserved*	—
632000_{16}	IMU Performance Counter Select register	U
632008_{16}	IMU Performance Counter Zero register	U
632010_{16}	IMU Performance Counter One register	U
632018_{16}–$633FF8_{16}$	*Reserved*	—
634000_{16}	MSI 32-bit Address register	U
634008_{16}	MSI 64-bit Address register	U
634010_{16}	*Reserved*	—
634018_{16}	Mem 64 PCIE Offset register	U
634020_{16}–$63FFF8_{16}$	*Reserved*	—
640000_{16}	MMU Control and Status register	M (sun4v)
640008_{16}	MMU TSB Control register	M (40-bit addr)
640010_{16}–$6400F8_{16}$	*Reserved*	—
640100_{16}	MMU TTE Cache Flush Address register	M (40-bit addr). Relocate the address to NCU space
640108_{16}	MMU TTE Cache Invalidate register	U
640110_{16}–$640FF8_{16}$	*Reserved*	—
641000_{16}	MMU Error Log Enable register	M (sun4v)
641008_{16}	MMU Interrupt Enable register	M (sun4v)
641010_{16}	MMU Interrupt Status register	M (sun4v)
641018_{16}	MMU Error Status Clear register	M (sun4v)
641020_{16}	MMU Error Status Set register	M (sun4v)
641028_{16}	MMU Translation Fault Address register	U
641030_{16}	MMU Translation Fault Status register	U

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(4 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
641038_{16}–$641FF8_{16}$	*Reserved*	—
642000_{16}	MMU Performance Counter Select register	U
642008_{16}	MMU Performance Counter Zero register	U
642010_{16}	MMU Performance Counter One register	U
642018_{16}–$645FF8_{16}$	*Reserved*	—
646000_{16}–$6461F8_{16}$	MMU TTE Cache Virtual Tag registers	M (sun4v)
646200_{16}–$646FF8_{16}$	*Reserved*	—
647000_{16}–$6471F8_{16}$	MMU TTE Cache Physical Tag registers	M (40-bit addr and sun4v)
647200_{16}–$647FF8_{16}$	*Reserved*	—
648000_{16}–$648FF8_{16}$	MMU TTE Cache Data registers	M (40-bit addr and sun4v)
649000_{16}–649078_{16}	MMU DEV2IOTSB registers	A (sun4v)
649080_{16}–$6490F8_{16}$	*Reserved*	—
649100_{16}–$6491F8_{16}$	MMU IOTSBDESC registers	A (sun4v)
649200_{16}–$650FF8_{16}$	*Reserved*	—
651000_{16}	ILU Error Log Enable register	U
651008_{16}	ILU Interrupt Enable register	U
651010_{16}	ILU Interrupt Status register	U
651018_{16}	ILU Error Status Clear register	U
651020_{16}	ILU Error Status Set register	U
651028_{16}–$6517F8_{16}$	*Reserved*	—
651800_{16}	PEU Core and Block Interrupt Enable register	U
651808_{16}	PEU Core and Block Interrupt Status register	U
651810_{16}–$651FF8_{16}$	*Reserved*	—
652000_{16}	ILU Diagnostic register	M
652008_{16}–$652FF8_{16}$	*Reserved*	—
653000_{16}	DMU Debug Select register for DMU Debug Bus A	M
653008_{16}	DMU Debug Select register for DMU Debug Bus B	M
653010_{16}–$6530F8_{16}$	*Reserved*	—
653100_{16}	DMU PCI Express Configuration register	U
653108_{16}–$65FFF8_{16}$	*Reserved*	—
660000_{16}–$6600F8_{16}$	Packet Scoreboard DMA register set	U
660100_{16}–$663FF8_{16}$	*Reserved*	—
664000_{16}–664078_{16}	Packet Scoreboard PIO register set	U
664080_{16}–$66FFF8_{16}$	*Reserved*	—
670000_{16}–$6700F8_{16}$	Transaction Scoreboard register set	U

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(5 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
670100_{16}	Transaction Scoreboard Status register	U
670108_{16}–$67FFF8_{16}$	*Reserved*	—
680000_{16}	PEU Control register	M
680008_{16}	PEU Status register	M
680010_{16}	PEU PME Turn Off Generate register	U
680018_{16}	PEU Ingress Credits Initial register	U
680020_{16}–$6800F8_{16}$	*Reserved*	—
680100_{16}	PEU Diagnostic register	M
680108_{16}–$6801F8_{16}$	*Reserved*	—
680200_{16}	PEU Egress Credits Consumed register	U
680208_{16}	PEU Egress Credit Limit register	U
680210_{16}	PEU Egress Retry Buffer register	U
680218_{16}	PEU Ingress Credits Allocated register	U
680220_{16}	PEU Ingress Credits Received register	U
680228_{16}–$680FF8_{16}$	*Reserved*	—
681000_{16}	PEU Other Event Log Enable register	U
681008_{16}	PEU Other Event Interrupt Enable register	U
681010_{16}	PEU Other Event Interrupt Status register	U
681018_{16}	PEU Other Event Status Clear register	M
681020_{16}	PEU Other Event Status Set register	M
681028_{16}	PEU Receive Other Event Header1 Log register	U
681030_{16}	PEU Receive Other Event Header2 Log register	U
681038_{16}	PEU Transmit Other Event Header1 Log register	U
681040_{16}	PEU Transmit Other Event Header2 Log register	U
681048_{16}–$681FF8_{16}$	*Reserved*	—
682000_{16}	PEU Performance Counter Select register	M
682008_{16}	PEU Performance Counter Zero register	U
682010_{16}	PEU Performance Counter One register	U
682018_{16}	PEU Performance Counter Two register	U
682020_{16}–$682FF8_{16}$	*Reserved*	—
683000_{16}	PEU Debug Select A register	M
683008_{16}	PEU Debug Select B register	M
683010_{16}–$68FFF8_{16}$	*Reserved*	—
690000_{16}	PEU Device Capabilities register	U
690008_{16}	PEU Device Control register	U

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(6 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
690010_{16}	PEU Device Status register	U
690018_{16}	PEU Link Capabilities register	U
690020_{16}	PEU Link Control register	U
690028_{16}	PEU Link Status register	M
690030_{16}	PEU Slot Capabilities register	U
690038_{16}–$690FF8_{16}$	*Reserved*	—
691000_{16}	PEU Uncorrectable Error Log Enable register	U
691008_{16}	PEU Uncorrectable Error Interrupt Enable register	U
691010_{16}	PEU Uncorrectable Error Interrupt Status register	U
691018_{16}	PEU Uncorrectable Error Status Clear register	U
691020_{16}	PEU Uncorrectable Error Status Set register	U
691028_{16}	PEU Receive Uncorrectable Error Header1 Log register	U
691030_{16}	PEU Receive Uncorrectable Error Header2 Log register	U
691038_{16}	PEU Transmit Uncorrectable Error Header1 Log register	U
691040_{16}	PEU Transmit Uncorrectable Error Header2 Log register	U
691048_{16}–$6A0FF8_{16}$	*Reserved*	—
$6A1000_{16}$	PEU Correctable Error Log Enable register	U
$6A1008_{16}$	PEU Correctable Error Interrupt Enable register	U
$6A1010_{16}$	PEU Correctable Error Interrupt Status register	U
$6A1018_{16}$	PEU Correctable Error Status Clear register	U
$6A1020_{16}$	PEU Correctable Error Status Set register	U
$6A1028_{16}$–$6E1FF8_{16}$	*Reserved*	—
$6E2000_{16}$	PEU CXPL/SerDes Revision register	A
$6E2008_{16}$	PEU CXPL AckNak Latency Threshold register	A
$6E2010_{16}$	PEU CXPL AckNak Latency Timer register	A
$6E2018_{16}$	PEU CXPL Replay Timer Threshold register	A
$6E2020_{16}$	PEU CXPL Replay Timer register	A
$6E2028_{16}$–$6E2038_{16}$	*Reserved*	—
$6E2040_{16}$	PEU CXPL DLL Vendor DLLP Message register	A
$6E2048_{16}$	*Reserved*	—
$6E2050_{16}$	PEU CXPL LTSSM Control register	A
$6E2058_{16}$	PEU CXPL DLL (Data Link Layer) Control register	—
$6E2060_{16}$	PEU CXPL MACL/PCS (Media Access Layer / Physical Coding Sublayer) Control register	A
$6E2068_{16}$	PEU CXPL MACL Lane Skew/Receiver Detection/Bit Lock Timer Control register	A

TABLE 21-1 CSR Map: Changes From the Fire 2.0 ASIC to UltraSPARC T2 PIU *(7 of 7)*

Offset	Register	A = Added M = Modified R = Removed U = Unchanged
$6E2070_{16}$	PEU CXPL MACL Symbol Number Control register	A
$6E2078_{16}$	PEU CXPL MACL Symbol Timer register	A
$6E2080_{16}$–$6E20F8_{16}$	*Reserved*	—
$6E2100_{16}$	PEU CXPL Core Status register	A
$6E2108_{16}$	PEU CXPL Event/Error Log Enable register	A
$6E2110_{16}$	PEU CXPL Event/Error Interrupt Enable register	A
$6E2118_{16}$	PEU CXPL Event/Error Interrupt Status register	A
$6E2120_{16}$	PEU CXPL Event/Error Status Clear register	A
$6E2128_{16}$	PEU CXPL Event/Error Status Set register	A
$6E2130_{16}$	PEU Link Bit Error Counter I register	A
$6E2138_{16}$	PEU Link Bit Error Counter II register	A
$6E2140_{16}$–$6E21F8_{16}$	*Reserved*	A
$6E2200_{16}$	PEU SerDes PLL Control/Status register	A
$6E2208_{16}$–$6E22F8_{16}$	*Reserved*	A
$6E2300_{16}$–$6E2338_{16}$	PEU SerDes Receiver Lane Control register	A
$6E2340_{16}$–$6E2378_{16}$	*Reserved*	—
$6E2380_{16}$–$6E23B8_{16}$	PEU SerDes Receiver Lane Status register	A
$6E23C0_{16}$–$6E23F8_{16}$	*Reserved*	—
$6E2400_{16}$–$6E2438_{16}$	PEU SerDes Transmitter Control register	A
$6E2480_{16}$–$6E24B8_{16}$	PEU SerDes Transmitter Status register	A
$6E2500_{16}$–$6E2508_{16}$	PEU SerDes Test Configuration register	A
$6E24C0_{16}$–$6FFF8_{16}$	*Reserved*	—

21.1.6 References

[1] PCI Express Base Specification, Revision 1.0a, April 15, 2003
(http://www.pcisig.com/specifications/pciexpress)
[2] Fire Programmers' Reference Manual, available from either
http://wikis.sun.com/display/FOSSdocs/Home
or
http://www.sun.com/processors/documentation.html

Sun-Internal Note *Fire 1.0 silicon release (associated with MAS 2.2 specification and LSI PCI-E Link Core 04/21/04, rel 1.05 specification), Part No.: 8xx-xxxx-xx, Current Revision: 1.0, Release Date: April 28, 2004. (http://asics.east/twiki/bin/vfs/vobs/fire/ops/arch/docs/spec/prog_ref.pdf?web=Asic_Products&topic=WebHome)*

21.2 PIU Programmer's Reference

21.2.1 Organization

The PIU programmer's reference consists of the following sections:

- **Introduction.** Gives an overview of this document and a high-level overview of PIU from a software perspective.
- **Operational Overview.** Provides a detailed description of the functionality within PIU.
- **CSR Fields and Bits.** Defines and documents the registers within PIU.
- **Error Event Summary.** Defines all the error events and documents the error handling for each individual error event.
- **Operational Sequences**. provides software sequences/algorithms for
  - PIU Link Training Sequence
  - PIU Hot Reset Sequence
  - PIU Link Disable Sequence
  - PIU Drain State and Recovery Sequence
  - PIU SerDes Clock and Electrical Configuration Sequence.
  - PIU Deterministic Mode Behavior and Sequence

21.2.2 PIU Overview

The PCI Express Interface integrates the functionality of a host to PCI Express bridge onto the UltraSPARC T2 system-on-chip. In PCI Express terminology this is called a PCI Express Root Complex.

PCI Express ports are point-to-point dual simplex interfaces. A x1 (pronounced "by 1") PCI Express port provides a low voltage differential signal pair (one in each direction) running at 2.5 Gbit/s ($8_2 10_2$). The maximum theoretical bandwidth for a x1 is 2 Gbit/s in each direction or 512 Mbytes/s aggregate. A x4 PCI Express port provides four differential signal pairs that produces a maximum theoretical aggregate bandwidth of 2 Gbytes/s (1 Gbyte/s in each direction).

PIU provides one x8 PCI Express port with a maximum theoretical aggregate bandwidth of 4 Gbytes/s (2 Gbytes/s in each direction).

PIU provides the following high-level functionality:

- Mapping PCI Express configuration, PCI Express I/O, and PCI Express 32-bit and 64-bit memory address spaces into UltraSPARC T2 address space (downbound operation).

- Mapping UltraSPARC T2 address space into PCI Express memory address space (upbound operation). Both bypass and translated accesses through an MMU are supported.
- Translating PCI Express MSI interrupts into Mondo interrupts.
- PCI Express legacy INTx support.
- PCI Express Power Management support (excluding wake on PME).

PIU does not support the following PCI Express features:

- Peer-to-peer communication (that is, between two I/O devices)
- Isochronous transfers (only one virtual channel is supported)
- Lock operations
- Wake system from PME (that is, no Vaux support)

21.3 Operational Overview

21.3.1 PIU Overview

PIU has an interface to the NCU, SIU, and one independent PCI Express leaf.

The PCI Express leaf contains an MMU, a TSB cache, 36 event queues, MSI mapping state, and Interrupt mondo controller.

On PCI Express, the requester ID is the combination of a requester's bus number, device number, and function number that uniquely identifies the requester. PIU's PCI Express requester ID is programmable by software through DMC PCI Express Configuration register.

21.3.2 NCU to PIU (Downbound) Transactions

21.3.2.1 UltraSPARC T2 Physical Address Partitioning

The 40-bit physical base addresses for UltraSPARC T2 are decoded as follows:

- An 8-Mbyte noncacheable region used for PIO access to on-chip PIU Internal CSRs:
  - {39:32 = 88_{16} to select 8-Mbyte noncacheable region in PIU
  - {31:23} = 0
  - {22:00} = offset within the 8-Mbyte region

- A 64-Gbyte noncacheable region that maps to subregions for off-chip PCI-Express configuration, PCI Express I/O and PCI Express 32-bit and 64-bit memory address spaces:
  - {39:36} = C_{16} to select 64-Gbyte noncacheable region in off-chip PCI-Express space.
  - {35:00} = offset within the 64-Gbyte region

In terms of the NCU mapping from 40-bit physical addresses to units, the UltraSPARC T2 PCI Express Unit is mapped as shown in TABLE 21-2.

TABLE 21-2 NCU Address-to-Unit Mapping

MSB Address Range {39:32}	Assignment
88_{16}	On-chip PIU CSRs (contains 8-Mbyte noncacheable region)
$C0_{16}$ –CF_{16}	Off-chip PCI-Express Spaces (64-Gbyte noncacheable region)

The registers to partition these noncacheable regions into subregions are provided by the NCU and described in Chapter 15, *Noncacheable Unit (NCU) and Boot ROM Interfaces*.

The 8-Mbyte noncacheable configuration region contains all of PIU's registers.

- A partition of the 8-Mbyte noncacheable configuration region can be found in TABLE 21-6 on page 488.
  - Writes to an address in a reserved range in PIU's 8-Mbyte noncacheable configuration region are dropped.
  - Reads to an address in a reserved range in PIU's 8-Mbyte noncacheable configuration region will return an Unmapped Read Error Packet.
  - A map of the registers within the configuration region can be found in TABLE 21-1 on page 477. **[is this xref OK?]**
  - Writes to an address marked as *Reserved* in TABLE 21-23 are silently dropped. No error is logged. **[is this xref OK?]**
  - Reads to an address marked as *Reserved* in TABLE 21-23 will return an Unmapped Read Error Packet. No error is logged. **[is this xref OK?]**
- All accesses to the 8-Mbyte noncacheable configuration region are big endian.

Accesses to PIU's 64-Gbyte noncacheable region *outside* the subregions mapped by the Address Mask/Match registers (described below) are treated as follows:

- Reads will return an Unmapped Read Error Packet.
- Writes are dropped.

The 64-Gbyte noncacheable space is further broken up into multiple subregions. PIU decodes and responds to three separate subregions for the PCIE port. Each subregion's offset and size are programmed during chip initialization via offset base and offset mask registers provided by NCU.

The three subregions in a PCI Express leaf are the PCI Express configuration and I/O subregion, the 32-bit addressable PCI Express memory subregion, and the 64-bit addressable PCI Express memory subregion.

The subregions within the 64-Gbyte noncacheable region are summarized in TABLE 21-3.

TABLE 21-3 64-Gbyte Noncacheable Region Partitioning

Subregion	Size	UltraSPARC T2-to-PIU Transactions
PCIE CFG/IO	512 Mbit	noncacheable read (4 byte max) noncacheable write (4 byte max)
PCIE Mem32	16 Mbyte – 4 Gbyte	noncacheable read noncacheable write
PCIE Mem64	16 Mbyte – 64 Gbyte	noncacheable read) noncacheable write

Only certain size accesses are allowed to the different 64-Gbyte and 8-Mbyte regions. TABLE 21-4 lists the allowable accesses for noncacheable write transactions and TABLE 21-5 lists the allowable accesses for noncacheable read transactions

TABLE 21-4 Allowable PIO Write Access Sizes

	64-Gbyte space			8-Mbyte space
Size in Bytes	Conf/IO	Mem32	Mem 64	CSR
1	Y	Y	Y	N
2	Y	Y	Y	N
4	Y	Y	Y	N
8	N	Y	Y	Y
Partial Store Instruction causes 8-byte masked stores where the byte mask indicates the particular byte to be stored. The byte mask can be any 8-bit value.	N	Y	Y	N

Note | For VIS Partial Store instruction with all byte masks are off (0 byte write), address [2] will be set to zero.

TABLE 21-5 Allowable PIO Read Access Sizes

	64 G space			8 M space
Size in Bytes	**Conf/IO**	**Mem32**	**Mem 64**	**CSR**
1	Y	Y	Y	N
2	Y	Y	Y	N
4	Y	Y	Y	N
8	N	Y	Y	Y

21.3.2.2 8-Mbyte Noncacheable Configuration Region

The Configuration Region is further divided as summarized in TABLE 21-6.

TABLE 21-6 8-Mbyte Noncacheable Configuration Region

Offset Range	Name	Size	Description
$00\ 0000_{16}$ – $3F.FFF8_{16}$	Reserved Range	4 Mbytes	Not used
$40\ 0000_{16}$ – $4F\ FFF8_{16}$	Reserved Range	1 Mbyte	Not used
$50\ 0000_{16}$ – $51\ FFF8_{16}$	Reserved Range	0.2 Mbyte	Not used
$52\ 0000_{16}$ – $53\ FFF8_{16}$	Reserved Range	0.2 Mbyte	Not used
$54\ .0000_{16}$ – $5F\ FFF8_{16}$	Reserved Range	0.6 Mbyte	Not used
$60\ 0000_{16}$ – $6F\ FFF8_{16}$	PCI Express-A Leaf CSRs	1 Mbyte	PCI Express-PIU internal registers, MMU, PCI Express Interrupts, etc.
$70\ 0000_{16}$ – $7F\ FFF8_{16}$	Reserved Range	1 Mbyte	Not used

For more information on the registers contained within these subdivisions, see *CSR Fields and Bits* on page 528.

21.3.2.3 PCI Express Configuration and I/O Subregion

A PCI Express configuration and I/O address subregion is a fixed size of 512 Mbytes. The offset within PIU's 64-Gbyte noncacheable region and subregion size are set using the PCIE-Cfg/IO Offset Base register and the PCIE Cfg/IO Offset Mask register provided by NCU. The offset base and offset mask registers for this subregion must be initialized for a 512-Mbyte subregion. Failure to do so will result in undefined behavior. The first 256 Mbytes of the PCI Express configuration and I/O address subregion maps to PCI Express configuration space. The last 256 Mbytes of this subregion maps to PCI Express IO space. All configuration and I/O packets are issued with the virtual channel ID equal to zero (the default virtual channel).

The mapping from the software-relative (that is, byte addressing, no byte masks) physical address {35:00} to the PCI Express Configuration and I/O subregion is as follows:

- 35:29 – Configuration and I/O Address Match. These bits determine if the access to PIU maps into one of the Configuration & I/O subregions.
- 28 – Configuration or I/O operation. 0x0 (1'b0) signifies a configuration operation. 0x1 (1'b1) signifies a I/O operation.
- 27:00 – Configuration- or I/O space-specific addressing information. See *Accessing PCI Express Configuration Space* and *Accessing PCI Express I/O Space* for more information on how these address bits are used.

Accessing PCI Express Configuration Space. The mapping from the 28-bit physical address offset within the PCI Express Configuration space to the Configuration Addressing information is as follows:

- 27:20 (8) – PCI Express Bus Number
- 19:15 (5) – Device Number
- 14:12 (3) – Function Number
- 11:00 (12) – PCI Express Configuration Register Address. Address bits {00:01} are used to determine the byte enables on PCI Express.

> **Note** The Configuration Address Space per Device function has been extended in PCI Express from 256 bytes to 4096 bytes. The Configuration Register Address has been extended to 12 bits (from 8) and the bus number, device number, and function number have been shifted left 4 bits. Therefore, software that accesses configuration registers for PCI Express devices and PCI devices sitting behind a PCI Express to PCI bridge will be different on PIU-based systems.

If the Bus Number for a configuration access matches the Bus Number in the DMC PCI Express Configuration register, a Type 0 Configuration Packet is generated. The Bus Number in the DMU PCI Express Configuration register is programmable by software. If the Bus Number is not equal to the Bus Number in DMC PCI Express Configuration register, a Type 1 Configuration Packet is generated.

A PCI Special Cycle can be generated on a PCI bus attached (directly or indirectly) to a PCI Express to PCI bridge by sending a Type 1 Configuration Packet with the appropriate bus number, all 1's for the Device Number ($1F_{16}$) and Function Number (7_{16}), and all zeros for the Configuration Address; for example, partialaddress{19:00} = $FF000_{16}$.

PIU does not implement any registers within PCI Express configuration space.

See the PCI and PCI Express specifications for more detail on accessing Configuration Space.

Accessing PCI Express I/O Space. The mapping from the 28-bit physical address offset within the PCI Express I/O space to the PCI Express I/O address is as follows:

- 27:0 – PCI Express I/O Address. Address bits 28:31 are always set to zero.

21.3.2.4 PCI Express 32-bit Addressing Memory Subregion

A PCI Express 32-bit Addressing Memory subregion is a programmable size ranging from 16 Mbytes to 4 Gbytes. The offset within PIU's 64-Gbyte noncacheable region and subregion size are set using the PCI Express Bus Mem32 Offset Base register and the PCI Express Bus Mem32 Offset Mask register provided by NCU. The offset base and offset mask registers for this subregion must be initialized for a subregion between 16 Mbytes and 4 Gbytes. Failure to do so will result in undefined behavior.

The mapping from the software-relative relative (that is, byte addressing, no byte masks) physical address {35:00} to the PCI Express 32-bit Addressing Memory subregion is as follows:

- 35:(X+1) – Address Match. These bits determine if the access to PIU maps into the 32-bit Addressing Memory subregions. X is determined by the size of the Mem32 subregion, which is set by the PCI Express Bus Mem32 Offset Mask register.
- X:00 – PCI Express 32-bit memory address. The PCI Express Mem32 address bits 31:(X+1) are set to zero. X is determined by the size of the Mem32 subregion, which is set by the PCI Express Bus Mem32 Offset Mask register.

Note | The lower 4-Gbyte space can only be set up by PCI-Express Mem32 Offset Base register and PCI-Express Mem32 Offset Mask register.

21.3.2.5 PCI Express 64-bit Addressing Memory Subregion

A PCI Express 64-bit Addressing Memory subregion is a programmable size ranging from 16 Mbytes to 64 Gbytes. The offset within PIU's 64-Gbyte noncacheable region and subregion size are set using the PCI Express Bus Mem64 Offset Base register and the PCI Express Bus Mem64 Offset Mask register provided by NCU. The offset base and offset mask registers for this subregion must be initialized for a subregion between 16 Mbytes and 64 Gbytes. Failure to do so will result in undefined behavior. All downbound packets to this subregion are issued with the virtual channel ID equal to zero (the default virtual channel).

The mapping from the software relative (that is, byte addressing, no byte masks) physical address {35:00} to the PCI Express 64-bit Addressing Memory subregion is as follows:

- 35:(X+1) – Address Match. These bits determine if the access to PIU maps into one of the 64-bit Addressing Memory subregions. X is determined by the size of the Mem64 subregion, which is set by the PCI Express Bus Mem64 Offset Mask register.

- X:00 – Offset within PCI Express 64-bit memory address.

The PCI Express Mem64 PIO address is formed from

{offset{63:36}, (offset{35:24} | pio addr{35:24}), pio addr{23:2}}

where pio addr{35:0} = {(pa{35:24} & ~mask{35:24}), pa{23:0}} and offset{63:24} is from the Mem64 PCIE Offset register.

Note | The lower 4-Gbyte space cannot be setup by PCI-Express Mem64 Offset Base register and PCI-Express Mem64 Offset Mask register.

21.3.3 PIU to Memory (Upbound) Transactions

21.3.3.1 PCI Express → PIU

PCI Express defines the following four address spaces:

- PCI Express configuration space
- PCI Express I/O space
- PCI Express memory space
- PCI Express message space

PIU does *not* respond to PCI Express Configuration space or I/O space transactions.

PIU does respond to PCI Express memory space transactions. This is the space in which DVMA and DMA (MMU bypass) activity takes place. PIU does not support peer-to-peer transactions; however, PCI Express switches or PCI Express-to-PCI bridges may support peer-to-peer transactions which would require no interaction with PIU. See *PCI Express Memory Transaction Processing* on page 491 for more information on how PIU handles PCI Express memory space transactions.

PCI Express defines two sets of message groups that live within the PCI Express message space: Standard Messages and Advanced Switching messages. PIU will not generate or respond to Advanced Switching messages. PIU will generate and/or respond to a subset of the Standard Messages. See *PCI Express Message Transaction Processing* on page 492 for more information on how PIU handles Standard Messages.

PCI Express Memory Transaction Processing. The final destination and address translation of a PCI Express Memory transaction is based on the following:

- IOMMU mode of operation: sun4u or sun4v
- PCI Express addressing mode used: 64-bit vs. 32-bit in sun4u mode.
- Value of Translation Enable (te) in the MMU Control register
- Value of Bypass Enable (be) in the MMU Control register

- Value of PCI Express address bits {63:39} when a 64-bit address is used

The IOMMU is a functional unit that is part of the UltraSPARC T2 PCI-Express subsystem. Its purpose is to map PCI-Express virtual addresses to UltraSPARC T2 system physical addresses during DMA transfers. An IOMMU is present on all I/O (PCI/PCI-X/PCI-Express) host bridges for Solaris/SPARC systems, but it is not required by the PCI/PCI-X/PCI-Express architecture.

PCI Express Message Transaction Processing. PCI Express defines the following message groups within the Standard Messaging group:

- Interrupt signaling
- Locked transaction support
- Error signaling
- Power management
- Slot power limit support
- Payload defined
- Vendor specific
- Hotplug signaling

The Interrupt Signaling message group consists of the Assert_INTx message and the Deassert_INTx message. MSIs are implemented via a memory write transaction. They are not messages and therefore, not included with the Interrupt signaling messages. PIU supports both Interrupt signaling messages and MSIs. See *Message Signaled Interrupts (MSIs and MSI-Xs)* on page 520 for more information about MSIs.

Locked transaction messages are *not* supported by PIU.

The Error Signaling message group consists of Correctable Error detected (ERR_COR), Uncorrectable Error detected (ERR_UNC), and Fatal Error Occurred (ERR_FATAL). PIU supports all error messages. See *PCI Express Messages* on page 520 for more information on Error Signaling message handling.

The Power Management message group consists of the following messages:

- PM_Active_State_NAK
- PM_PME
- PME_Turn_Off
- PME_TO_Ack

All Power Management messages are supported by PIU. See *Power Management* on page 523 for more information on Power Management message handling. PIU does not support powering on a system from a PME message.

PIU will *not* automatically generate the set_slot_power_limit message when it detects that a link has transitioned from down to up. PIU only generates the set_slot_power_limit message when it detects software updating the PEU slot capability register, even with the same value. PEU uses the Slot Power Limit Scale and Slot Power Limit Value fields in the PEU Slot Capabilities register when generating the message. These fields are programmable by software.

PIU does not support Payload Defined messages (messages with data), vendor-specific messages, and advanced switching messages. PIU will silently drop these messages as required by the PCI Express specification.

The Hotplug message group consists of the following messages:

- Attention_Indicator_On
- Attention_Indicator_Blink
- Attention_Indicator_Off
- Power_Indicator_On
- Power_Indicator_Blink
- Power_Indicator_Off
- Attention_Button_Pressed

PIU does not support the Hotplug message group. If a system requires these messages, it can use a switch that supports them. See Section 2.8.1.8. Hot Plug Signaling Messages on page 96 of the PCI Express Specification v1.0(RC1) for more information on switch support for PCI Express Hotplug messages.

21.3.4 UltraSPARC T2 IOMMU (sun4u vs. sun4v)

The UltraSPARC T2, IOMMU provides two modes of operation: sun4u (X-PARs disable) mode and sun4v (X-PARs enable) mode. It provides the capability to allow the software to enable/disable X-PARs functionality via an on/off switch. This switch is effected using the sun4v bit of the MMU Control/Status register as follows:

- sun4u (X-PARs disable) mode: This is the default mode. The IOMMU function is similar to the IOMMU function in Fire IO bridge 1.0 silicon when in this mode.
  - 32-bit access is enabled.
  - 64-bit access is disabled.
  - Bypass is normally enabled (it can be disabled if there is a security concern).
- sun4v (X-PARs enable) mode: In this mode, the function allows IOMMU to guarantee isolation.
  - 32-bit access is enabled and is treated as 64-bit access with upper 32-bits all zero.
  - 64-bit access is enabled.
  - Bypass is normally disabled (it can be enabled if there is a need for this application and security has been taken care of somewhere else).

21.3.5 sun4u Mode IOMMU and Bypass Operation

In sun4u mode, the IOMMU provides the similar programming model as Fire ASIC IOMMU. The MMU provides a virtual to physical address translation on a range of upbound PCI Express transactions. See TABLE 21-7 for more information on which transactions are translated by the MMU. The MMU allows a collection of

discontiguous physical memory pages to be represented as a contiguous virtual I/O address range. It also provides a level of protection by detecting a subset of invalid memory reads and writes to addresses that are marked as invalid I/O addresses.

21.3.5.1 Translation

TABLE 21-7 shows the various ways that PIU, as a PCI Express target device, deals with PCI Express Memory Space addresses.

TABLE 21-7 PCI Express Memory Transaction Behavior (sun4u mode)

Mode	te	be	Addr{63:39}	Addr{31:0}	Result
32-bit	0	X	N/A	N/A	MMU TRN_ERR
32-bit	1	X	N/A	FIGURE 21-1	FIGURE 21-1
64-bit	X	0	X	N/A	MMU BYP_ERR
64-bit	X	1	0000 0000 0000 0000 0000 0000 0_2 – 1111 1111 1111 1011 1111 1111 1_2	N/A	MMU BYP_OOR
64-bit	X	1	1111 1111 1111 1100 0000 0000 0_2	N/A	Bypass (DMA)
64-bit	X	1	1111 1111 1111 1100 0000 0000_1_2 – 1111 1111 1111 1111 1111 1111 1_2		MMU BYP_OOR

The size of the TSB table determines the range on the virtual addresses that are valid. Addresses not in the valid range cause a translation out-of-range error (TRN_OOR). FIGURE 21-1 illustrates virtual address configurations. TSB sizes and page sizes not diagrammed in the figure are not valid and produce undefined results.

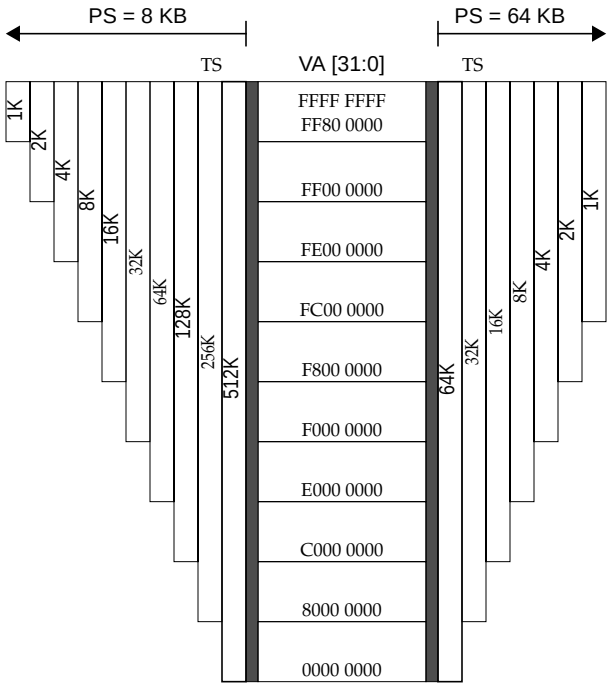


FIGURE 21-1 Translation Size Virtual Address Configurations (sun4u mode)

In MMU translation mode, the physical address is obtained by performing a virtual-to-physical translation through the MMU. See *sun4u Mode IOMMU and Bypass Operation* on page 493 for more information on MMU operation.

PIU does not support peer-to-peer operations between I/O devices. PCI Express switches and PCI Express-to-PCI bridges may support peer-to-peer operations.

In bypass mode, UltraSPARC T2 physical memory address{38:0} = PCI Express{38:0}. (Note that {38:0} is a logical representation of the address bits.)

21.3.5.2 Translation Storage Buffer Overview

The MMU fetches translation information from a Translation Storage Buffer (TSB) in system memory. The TSB consists of an array of 8-byte Translation Table Entries (TTEs) that provide mapping information for the virtual DMA pages.

The base size of a page that a single TTE maps to is set while initializing the MMU. This can be set to either an 8-Kbyte or 64-Kbyte page. The base page size is used by the MMU to determine the index into the TSB (that is, bits 15:00 of the address are ignored when calculating the index into the TSB for a base page size of 64 Kbytes, bits 12:00 of the address are ignored for an 8-Kbyte page).

The size of the TSB can be initialized from 1-Kbyte to 512-Kbyte entries in power-of-2 increments. Therefore, the TSB must reside in a physically contiguous buffer ranging from 8 Kbytes to 4 Mbytes since each entry consumes 8 bytes. The DVMA address space available for an 8-Kbyte base page ranges from 8 Mbytes to 4 Gbytes depending on the size of the TSB. The DVMA address space available for a 64-Kbyte base page ranges from 64 Mbytes to 4 Gbytes depending on the size of the TSB (a DVMA address space larger than 4 Gbytes is not supported by PIU).

PIU does *not* support mixed page size TTEs (that is, both 8-Kbyte and 64-Kbyte TTEs in the TSB at the same time). This feature is no longer required since the TSB cache is cacheline based now (previous bridges cached individual TTEs). Existing code will still function correctly with no performance degradation (that is, this change is backward compatible). See *TSB Cache Operation, Initialization, and Debug* on page 499 for more information on the TSB cache.

The base address of the TSB table has to be aligned on an 8-Kbyte boundary. The lower order 13 bits are assumed to be 0_{16}. Tables larger than 8 Kbytes are only constrained to be on 8K boundaries rather than having to be size aligned.

TSB Indexing. An index into the TSB is calculated from the DVMA address (VirtAddr) as follows:

```
if (base page size == 8KB)
   Page Bits = 13;
else
   Page Bits = 16;
Index = (VirtAddr >> Page Bits) & (Table Size - 1);
```

TABLE 21-8 TSB Indexing

	8-Kbyte Base Page		64 Kbyte-Base Page	
TSB Size	**VA Space Size**	**TSB Index {3}**	**VA Space Size**	**TSB Index {3}**
1K entries	8 MB	VA{22:13}	64 MB	VA{25:16}
2K entries	16 MB	VA{23:13}	128 MB	VA{26:16}
4K entries	32 MB	VA{24:13}	256 MB	VA{27:16}
8K entries	64 MB	VA{25:13}	512 MB	VA{28:16}
16K entries	128 MB	VA{26:13}	1 GB	VA{29:16}
32K entries	256 MB	VA{27:13}	2 GB	VA{30:16}
64K entries	512 MB	VA{28:13}	4 GB	VA{31:16}

TABLE 21-8 TSB Indexing *(Continued)*

	8-Kbyte Base Page		64 Kbyte-Base Page	
TSB Size	**VA Space Size**	**TSB Index {3}**	**VA Space Size**	**TSB Index {3}**
128K entries	1 GB	VA{29:13}	Not allowed[1]	—
256K entries	2 GB	VA{30:13}	Not allowed[1]	—
512K entries	4 GB	VA{31:13}	Not allowed[1]	—

1. Hardware does not prevent illegal combinations from being programmed. Programming an illegal combination into the MMU will result in undefined behavior.

21.3.5.3 Translation Table Entry (TTE)

A TTE entry has valid information only when the Valid bit (data_v) is set. TTEs have the following format.

TABLE 21-9 TTE Data Format (sun4u mode)

Bit	Field	Description
63:48	dev_key	Device Key -- 16-bit PCI-Express Requestor ID (bus #, device #, function #)
47:39	—	*Reserved* for software use, is ignored by hardware.
38:13	data_pa	Contains bits 38:13 of physical address. Bits 15:13 are not used for 64K page.
12:7	data_soft	*Reserved* for software use, is ignored by hardware.
6:5	—	*Reserved* for software use, is ignored by hardware.
5:3	fnm	Function number mask
2	key_valid	Device key valid bit
1	data_w	Set if this page is writable
0	data_v	Valid bit (1 = TTE entry has valid mapping)

Note The data_size, localbus, context, and cacheable fields have been removed from the TTE for PIU. They are marked as *Reserved*. They should be ignored by hardware for software backward compatibility in order to maintain compatibility with any software that may use these fields. data_size is no longer required since the hardware will cache a cacheline of TTEs (that is, there is no benefit to having 64K pages in 8K TTE entries). localbus is not meaningful for PCI Express. context was used to manage a software coherent TSB cache, which is not present in PIU. PIU does not support generating noncoherent memory transactions; therefore, the cacheable field is not required.

Note | The dev_key, key_valid, and fnm fields are added into the TTE for PIU. This allows the PIO to perform TTE key protection in sun4u mode. The full 16-bit PCI express requester ID of the device that initiated the DMA transaction is compared with the 16-bit dev_key value. This comparison only happens if the key_valid bit is set. If it is not, any device will have access to the physical addresses mapped by the IOTTE. To handle the PCI Express phantom function numbers, the fnm field is used. This function number mask will be **and**ed with the function numbers in the source requester ID and also **and**ed with the function number in the dev_key prior to the comparison. This implementation is to avoid software programming inconsistency between the fnm field and dev_key value field in the TTE.

21.3.5.4 Bypass and MMU Initialization

There are two paths that an upbound transaction can take based on the memory transaction addressing mode. PIU will attempt to translate (using the MMU) a 32-bit mode memory transaction and "bypass" a 64-bit mode memory transaction. Each of these paths can be individually enabled/disabled without side effects on the other. See TABLE 21-7 on page 494 for more information how upbound memory transactions are handled.

Bypass is disabled/enabled via the Bypass Enable (be) bit in the MMU Control register. When Bypass is enabled, upbound 64-bit mode memory transactions with the upper 14 address bits 63:50 set to all 1's; address bits 49:39 set to zero are converted to the appropriate memory operation. The lower 39 bits (38:0) of the address are used as the cacheable physical address.

If bypass is disabled or if address bits 63:39 are not set to the correct values stated earlier, an error is flagged, which optionally generates an interrupt. (A 64-bit address that is not addressed to PIU should never reach PIU since PCI Express links are point-to-point, for example, a switch should never forward to PIU a packet that is meant for another device.)

The MMU is disabled/enabled via the Translation Enable (te) bit in the MMU Control register. When the MMU is enabled, upbound 32-bit mode memory transactions will attempt to be translated to the 39 bit (38:0) cacheable physical memory address. To translate, an index into the TSB is calculated (see *TSB Indexing* on page 496 for more information) from the I/O address, and the TTE is fetched. If the TTE is valid, PIU will check to see if the transaction is a write and the page is marked as writable in the TTE. If the transaction is a read or is a valid write transaction, PIU will convert the transaction to the appropriate memory operation using the physical address specified in the TTE.

If the MMU is disabled or if one or more of the assertions stated above are not valid, an error is flagged, which optionally generates an interrupt. If the TTE is invalid or the TSB does not cover the address in question (that is, the TSB size is less than maximum size and the address is outside the valid address range specified by tsb_size), an error is generated.

Note | The MMU Enable behavior has changed from previous generation Sun bridges. If the MMU is disabled, it will no longer pass through DMA traffic. Instead it will ignore upbound reads and writes, log an error, and optionally generate an interrupt.

During normal operation, bypass and the MMU should be initialized as follows:

- Translations should be enabled (1).
- Bypass should be enabled (1).
- Process Disable should be enabled (0).

TSB Cache Operation, Initialization, and Debug. The TSB resides in system memory and contains TTEs that provide mapping information for the virtual DMA pages. See *Translation Storage Buffer Overview* on page 495 for more information on the TSB.

PIU implements a software coherent cache to locally store recently accessed TTEs in order to reduce DVMA transfer latencies. PIU never modifies the contents of the TTEs therefore and will never write back the contents of the cache to system memory. Cache eviction is based on an LRU algorithm. PIU supports a software flush mechanism via the PCIE TSB Cache Flush register.

PIU caches reads to the TSB. A TSB cache entry is based on the L2 cacheline size, which is 64 bytes. Each cacheline can hold up to 8 valid TTE entries. Up to 64 cache entries can be stored in the TSB cache at any given time, giving a maximum total of 512 TTEs that can be cached.

Software may need to use a more intelligent TTE allocation algorithm to realize the full performance potential of the hardware-coherent TSB cache. Since each entry in the TSB cache consists of eight TTEs, initializing one TTE entry could invalidate seven other TTE entries currently in the cache (that is, they would be removed from the cache but still would be "valid" TTEs). Therefore, a simple sequential allocation scheme may not be the best approach. It is important to note that the TTE allocation algorithm will have to balance efficient cache utilization with fragmentation issues that arise when you sparsely populate the cachelines.

Three bits control the behavior of PIU's TSB cache: TSB Cache Mode, cm, (2 bits), and TSB Snoop Enable, se, (1 bit). When translation is disabled, TSB Cache Mode has no meaning and is ignored. TSB Cache Snoop Enable is independent of translation enable and TSB cache mode behavior.

During normal operation, the TSB cache control bits should be initialized to the following:

- TSB Cache Mode should be enabled (3).
- TSB Cache Snoop should be enabled (1).

The TSB cache can only be read-from/written-to via PIOs when the TSB cache is disabled. The Translation enable setting has no effect on accesses to the TSB cache. Attempts to read from or write to the TSB cache when it is enabled are flagged as an error and optionally generate an interrupt.

Note | The TTEs can be initialized to valid settings for debug purposes. See the TSB Cache Mode table, below, for more information on this behavior.

MMU TSB Cache Mode

cm	Description
00	TSB cache is disabled. A single cacheline, ha coherent TLB is used to cache the last-used T
01	TSB cache is enabled and locked. Misses cau The TSB in main memory is not used.
10	TSB cache is enabled and locked. Misses use
11	TSB cache is enabled (default)

If Translation is enabled and the TSB cache is disabled (cm = 0) PIU uses a single cacheline, software-coherent TLB to cache the last used TTE. If TSB Snoop Enable is disabled, software must flush the TLB entry to ensure that the TLB is consistent with the TSB during unmap operations. There are no restrictions on when the TLB entry can be flushed. It is expected that this functionality will only be used for bringup/debug if issues with the main TSB cache are discovered.

21.3.5.5 IOMMU Translation Flow

- Performs range check based on page size and IOTSB size.
  - If out-of-range is detected, set trn_oor, nullify the transaction.
- Fully associative lookup in virtual tag buffer, comparing the virtual page number against the fields of all tag array entries.
- Cache hit if the virtual page number match the field of tag array entry.
  - Retrieved the physical page number from the appropriate I/O TTE in the matching entry.

    The following errors are detected in parallel. There is no priority between these errors.

    – If TTE cache data parity error is detected, set ttc_dpe, nullify the transaction.

    – If TTE is invalid, set tte_inv, nullify the transaction.

– If the wrt bit is not set and the current transaction is a write transaction, set tte_prt, nullify the write transaction.

– Compare the full 16-bit of the PCI-Express requester ID and the device key value if key_valid field is enable. If device key value doesn't match the PCI-Express requester ID, set key_err, nullify the transaction.

- Construct UltraSPARC T2 physical address.

- Cache miss if no entries have match on virtual page number.

  If tablewalk is disabled, set tbw_dme, nullify the transaction.

  - Form the index of I/O TSB table, fetch the IOTTEs from memory.

    The following errors are detected in parallel.

    – If the SIU return the tablewalk data with UE or DSN detects the tablewalk data with parity error, set tbw_err, nullify the transaction.

    – If the IOMMU detects parity error on the tablewalk data locally, set tbw_dpe, nullify the transaction.

    – If the IOMMU detects that the allocated replacement entry mismatch with the replacement entry in the tablewalk data, set tbw_une, nullify the transaction.

  - Install IOTTE in cache, replacing entry in cache if full and updating tag and data array portions of cache.

    The following errors are detected in parallel, there is no priority between theses errors.

    – If TTE is invalid, set tte_inv, nullify the transaction.

    – If wrt bit is not set and the current transaction is the write transaction, set tte_prt, nullify the write transaction.

    – Compare the full 16-bit of the PCI-Express requester ID and the device key value if key_valid field is enable. If device key value doesn't match the PCI-Express requester ID, set key_err, nullify the transaction.

- Construct UltraSPARC T2 physical address.

21.3.6 sun4v Mode IOMMU

In sun4v mode, IOMMU provides a higher level of memory protection by applying device-specific translation to each PCI-Express DVMA transaction. PCI-E devices will be restricted to translations that are specific to them. To implement this scheme, two tables will be implemented in hardware. One will be an I/O Translation Storage Buffer (IOTSB) descriptor table. This table will hold the information that specifies the IOTSB for each device. The second table will hold indices into the descriptor table. The IOTSB identifier will be 7 bits: 6 bits from the PCI bus ID of the requesting device and the MSB from the virtual address of the request. However, the IOTSB descriptor table will only have 32 entries. This table then maps the 7-bit IOTSB

identifiers into the 5-bit index for the IOTSB descriptor table. This approach allows 32 independent IOMMU partitions to be created. The PCI bus ID is 8 bits wide; a configuration bit in IOMMU control/status register selects whether it is the least-significant 6 bits of the bus ID or the least-significant 6 bits of the bus ID with one bit right-shifted out. Since bit 63 of the virtual address are included in the lookup, a set of devices with a particular bus ID can use two different IOMMU partitions. This allows, for example, up to two different IOTSBs with two different page sizes to be used by the translations for those devices. FIGURE 21-2 illustrates the use of these two tables.

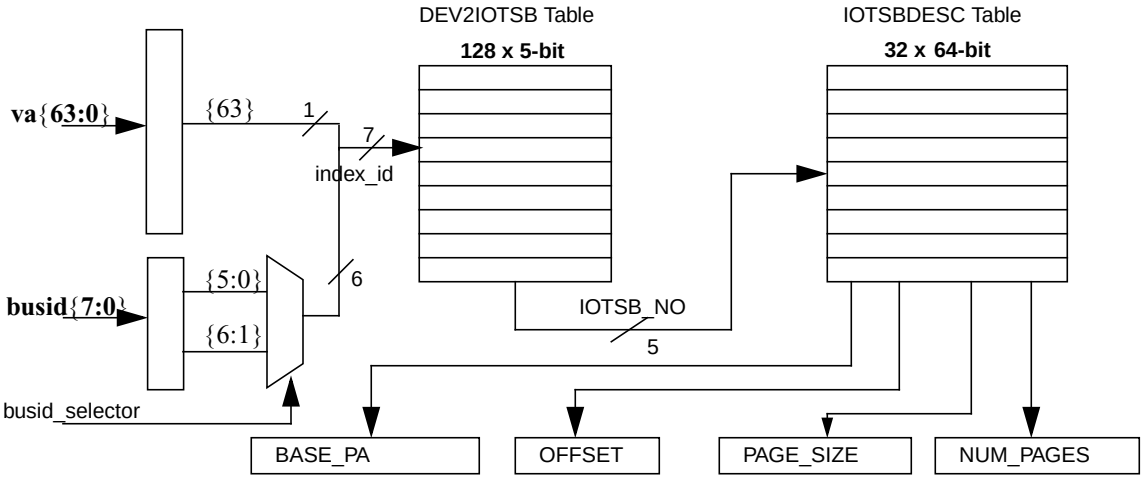


FIGURE 21-2 Reading IOTSB descriptor for a request

21.3.6.1 IOTSB Descriptor Table

Since the IOMMU will support multiple IOTSBs, there exists a table to keep track of the defining information for each. The information stored in each entry of this table includes the base PA, the page size, the number of pages, and an address offset to be subtracted from the PCI-Express virtual address when determining which IOTTE to use for the translation. The IOTSB descriptor table will hold this information for 32 IOTSBs. The IOMMU supports 40 bits of PCI Express virtual address and 39 bits of physical address. Accesses in the 64-bit PCI Express virtual address range that are outside of the 40-bit virtual address range will be treated as out-of-range errors and logged. The 32-bit PCI Express virtual address range is zero-extended to the supported 40-bit space and is always in range.

TABLE 21-10 Field Description for IOTSB Descriptor Table Entries

Bit	Field	Description
63	valid	Valid bit (0 = not valid, 1 = valid)
62:61	*Reserved* for SW PAR for HW	These two bits should be ignored by software. Hardware uses these two bit for odd parity protection. Bit 62: {63, {59:32}} Bit 61:{31:0}
60	—	*Reserved*
59:34	base_pa{38:13}	Starting PA of IOTSB.
33:7	offset{26:0}	Address offset.
6:4	page_size{2:0}	Page size for this IOTSB.
3:0	num_pages{3:0}	Number of pages in the IOTSB.

The page_size field indicates how large each page in the IOTSB is. UltraSPARC T2 IOMMU support four different page sizes, which are calculated by $2^{(13 + 3 * \text{page\_size})}$. TABLE 21-11 shows which sizes are supported (these are the same as the UltraSPARC T2 core MMU).

TABLE 21-11 Page Size Programming

page_size{2:0}	Page Size	Index Range
000_2	8K	12–0
001_2	64K	15–0
010_2	*Reserved*	*Reserved*
011_2	4M	21–0
100_2	*Reserved*	*Reserved*
101_2	256M	27–0
110_2	*Reserved*	*Reserved*
111_2	*Reserved*	*Reserved*

The offset field allows for an arbitrary linear relocation (at page-size granularity) between the PCI Express virtual address and the adjusted virtual address used in the IOMMU lookup. For example, it allows the PCI Express address range for a particular IOMMU partition to start at an arbitrary base address and be relocated down to 0 so that the indices into its IOTSB translation tables can start at 0 (ensuring that space is not wasted at the beginning of the translation table). The offset is subtracted to the transaction address with the index bits shifted out. The result of the subtraction is the page number for this address. If the underflow happens after

the subtraction, the out-of-range error will be logged. The number of pages for each IOTSB is calculated as $2^{(10 + num\_pages)}$. If the page number of the current address is larger than the number of pages in the IOTSB, the out-of-range error will be logged.

The base_pa is the physical address of the first IOTTE in the page table. When the base_pa is added to the transaction page number left-shifted 3 bits, we have the PA of the IOTTE for this DMA transaction. To make programming and hardware design of this table simple, it is proposed that each entry contain 64 bits. Thus, each entry can be programmed with a single PIO write. To do this, we have limited the base_pa to start on an 8-Kbyte boundary (26 bits need to be stored) and limited the offset to 27 bits. These are sufficient offset bits to allow arbitrary relocation of the virtual address space at page-size granularity.

With an 8-Kbyte page size, the 40-bit virtual address size has 27 bits of page number requiring a 27-bit offset. With larger page sizes the page number is smaller requiring fewer bits of offset. The offset subtraction to yield the page number is performed with a 27-bit subtraction and underflow detection. The resulting page number is also subjected to an out-of-range check versus the num_pages limit. Note that the IOMMU provides complete isolation and protection between the virtual address spaces of each I/O partition: Translation tables are separate for each I/O partition and translations in the IOMMU cache are tagged with the IOTSB number (5 bits) as well as the adjusted virtual page number (27 bits). TABLE 21-12 lists the legal page size and IOTSB size combinations. The calculation of the IOTTE address is summarized in FIGURE 21-3.

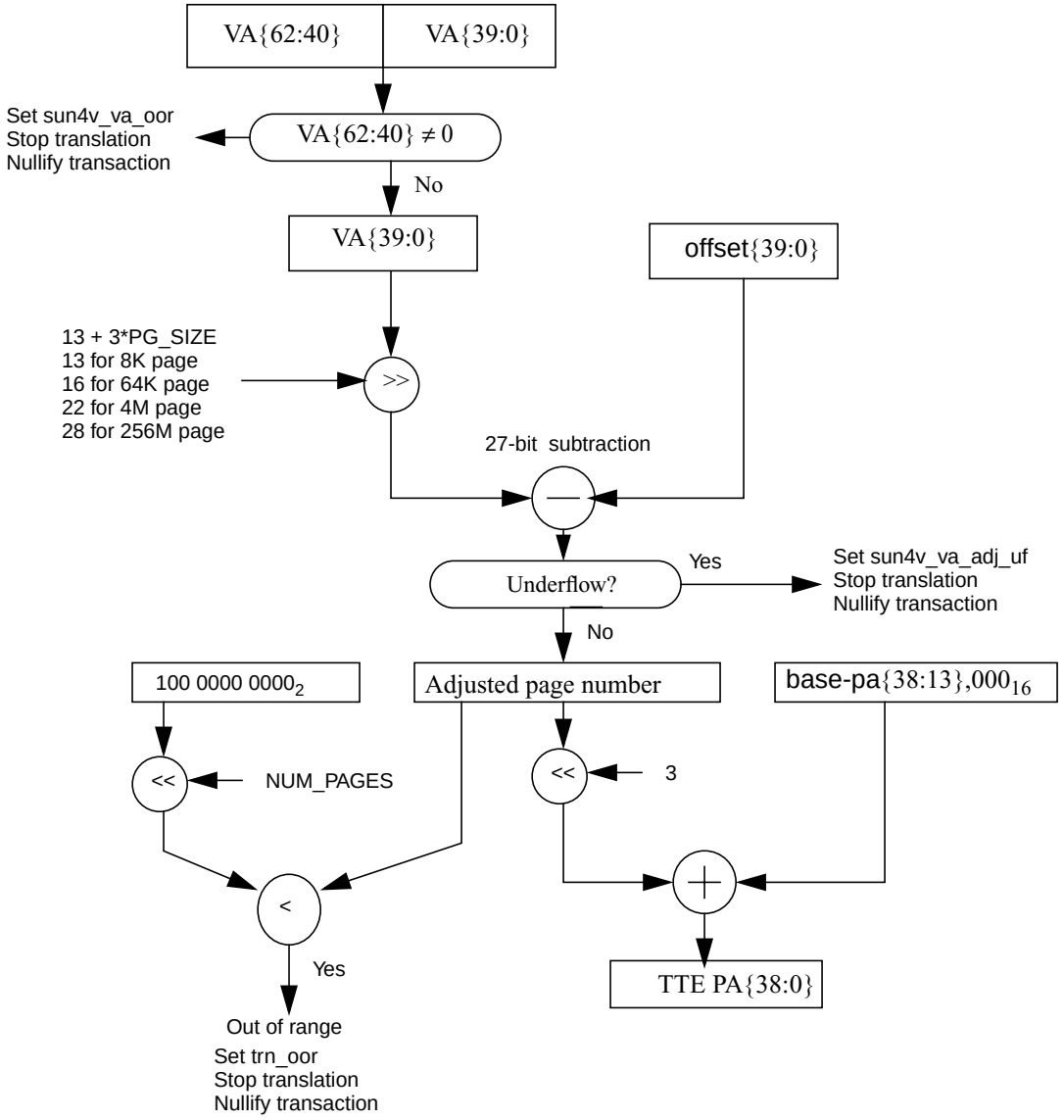
Note | It is recommended that the starting VA from a partition, the offset, and the IOTSB size are chosen so that the lowest adjusted virtual page number always selects the first IOTSB entry and the partition can access the full IOTSB table size with VA{62:40} = 0.

TABLE 21-12 sun4v IOTSB Indexing With Legal Combination of Page Size and IOTSB Size

	8-Kbyte Base Page		64-Kbyte Base Page		4-Mbyte Base Page		256-Mbyte Base Page	
IOTSB Size	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**
1K	8 MB	ADJ_VA {22:13}	64 MB	ADJ_VA {25:16}	4 GB	ADJ_VA {31:22}	256 GB	ADJ_VA {37:28}
2K	16 MB	ADJ_VA {23:13}	128 MB	ADJ_VA {26:16}	8 GB	ADJ_VA {32:22}	512 GB	ADJ_VA {38:28}
4K	32 MB	ADJ_VA {24:13}	256 MB	ADJ_VA {27:16}	16 GB	ADJ_VA {33:22}	1 TB	ADJ_VA {39:28}
8K	64 MB	ADJ_VA {25:13}	512 MB	ADJ_VA {28:16}	32 GB	ADJ_VA {34:22}		
16K	128 MB	ADJ_VA {26:13}	1 GB	ADJ_VA {29:16}	64 GB	ADJ_VA {35:22}		
32K	256 MB	ADJ_VA {27:13}	2 GB	ADJ_VA {30:16}	128 GB	ADJ_VA {36:22}		

TABLE 21-12 sun4v IOTSB Indexing With Legal Combination of Page Size and IOTSB Size *(Continued)*

	8-Kbyte Base Page		64-Kbyte Base Page		4-Mbyte Base Page		256-Mbyte Base Page	
IOTSB Size	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**	**VA Space**	**IOTSB Index**
64K	512 MB	ADJ_VA {28:13}	4 GB	ADJ_VA {31:16}	256 GB	ADJ_VA {37:22}		
128K	1 GB	ADJ_VA {29:13}	8 GB	ADJ_VA {32:16}	512 GB	ADJ_VA {38:22}		
256K	2 GB	ADJ_VA {30:13}	16 GB	ADJ_VA {33:16}	1 TB	ADJ_VA {39:22}		
512K	4 GB	ADJ_VA {31:13}	32 GB	ADJ_VA {34:16}				
1M	8 GB	ADJ_VA {32:13}	64 GB	ADJ_VA {35:16}				
2M	16 GB	ADJ_VA {33:13}	128 GB	ADJ_VA {36:16}				
4M	32 GB	ADJ_VA {34:13}	256 GB	ADJ_VA {37:16}				
8M	64 GB	ADJ_VA {35:13}	512 GB	ADJ_VA {38:16}				
16M	128 GB	ADJ_VA {36:13}	1 TB	ADJ_VA {39:16}				
32M	256 GB	ADJ_VA {37:13}						



Priority Error List: sun4v_va_oor, sun4v_va_adj_uf, trn_oor.

FIGURE 21-3 IOTTE Address Calculation (After Reading IOTSB Descriptor From IOTSBDESC)

21.3.6.2 DEV2IOTSB Table

The DEV2IOTSB table essentially holds 5-bit x 128 indices into the IOTSB descriptor table. Each IOTSB identifier is generated by concatenating bits 63 of the VA onto the selected 6 bits of the PCI bus ID for the device that initiated the transaction. The index table maps the 7-bit IDs into the matching entry in the descriptor table. To simplify programming and reduce hardware area, the table will be arranged as 16 registers each containing 8 x 5-bit values. For convenience, each 5-bit value is packed with 3 reserved bits to byte boundaries within the register.

If the least-significant 6 bits of bus ID are selected, an index into the table is va{63}, bus_id{5:3}, and the values within a register are selected by bus_id{2:0}. If the least-significant 6 bits of bus ID with one bit right-shifted out are selected, an index into the table is va{63}, bus_id{6:4} and the values within a register are selected by bus_id{3:1}.

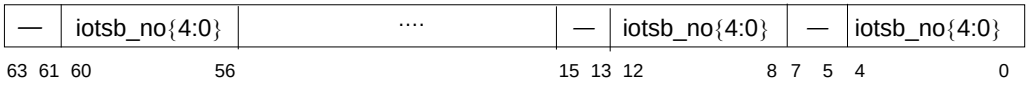


FIGURE 21-4 Index Table Entry Format

21.3.6.3 IOMMU Translation Lookup Flow

Cache hit case flow. ■DMA request from PCI-E arrives, sent to IOMMU.

- If VA{62:40} ≠ 0, set sun4v_va_oor, nullify the transaction.
- Form index_id{6:0} from the bus_id and address bits 63 of the PCI-E address. Use index_id{6:0} as index to retrieve the iotsb_no{4:0} value in DEV2IOTSB table.
- Use iotsb_no{4:0} as index to retrieve the offset, page size, base address and number of pages from IOTSB Descriptor table.

  The following is the priority of error reporting during this stage.

  - If IOTSBDESC parity error is detected, set iotsbdesc_dpe, nullify the transaction.
  - If no IOTSBDESC parity error but valid bit ≠ 1, set iotsbdesc_inv, nullify the transaction.
  - If no IOTSBDESC parity error, valid bit = 1_2 but invalid page size is detected, set sun4v_inv_pg_size, nullify the transaction.
- Form the adjusted PCI-E virtual page number by shifting the address and subtracting the offset.

  The following is the priority of error reporting during this stage.

  - If underflow occurs, set sun4v_va_adj_uf, nullify the transaction.
  - If no underflow occurs but adjusted VPN exceeds the IOTSB size, set trn_oor, nullify the transaction.

- Masking the lower 3 bits of the adjusted PCI-E virtual page number address and associative lookup in VA tag array, comparing iotsb_no and adjusted PCI-E virtual page number with lower 3 bits masked against those fields of all tag array entries.
- Cache hit if iotsb_no and adjusted PCI-E virtual page number match the entry in VA tag array.
- Retrieve the physical page frame number and the device key value from the appropriate TTE in the matching entry of the data array portion of the TTE cache.

  The following errors are detected in parallel. There is no priority among these errors.

  - If TTE cache data parity error is detected, set ttc_dpe, nullify the transaction.
  - If TTE is invalid, set tte_inv, nullify the transaction.
  - If the wrt bit is not set and the current transaction is a write transaction, set tte_prt, nullify the write transaction.
  - Compares the full 16-bit of the PCI-Express requester ID and the device key value if key_valid field is enable. If device key value doesn't match the PCI-Express requester ID, set key_err, nullify the transaction.
- Construct UltraSPARC T2 physical address.

Cache miss case flow. ■DMA request from PCI-E arrives, sent to IOMMU.

- If VA{62:40} $\neq 0$, set sun4v_va_oor, nullify the transaction. (same as cache hit flow).
- Form index_id{6:0} from the bus_id and address bits 63 of the PCI-E address. Use index_id{6:0} as index to retrieve the iotsb_no{4:0} value in IOTSBDESC table.

  The following is the priority of error reporting during this stage (same as cache hit flow).

  - If IOTSBDESC parity error is detected, set iotsbdesc_dpe, nullify the transaction.
  - If no IOTSBDESC parity error but valid bit $\neq 1$, set iotsbdesc_inv, nullify the transaction.
  - If no IOTSBDESC parity error, valid bit = 1_2, but invalid page size is detected, set sun4v_inv_pg_size, nullify the transaction.
- Use iotsb_no{4:0} as index to retrieve the offset and page size from IOTSB Descriptor table.
- Form the adjusted PCI-E virtual page number by shifting the address and subtracting the offset.

  The following is the priority of error reporting during this stage. (same as cache hit flow).

  - If underflow occurs, set sun4v_va_adj_uf, nullify the transaction.
  - If no underflow occurs but adjusted VPN exceeds the IOTSB size, set trn_oor, nullify the transaction.

- Masking the lower 3 bits of the adjusted PCI-E virtual page number and associative lookup in VA tag array, comparing iotsb_no and adjusted PCI-E address with lower 3-bit masked against those fields of all tag array entries.
- Cache miss if no entry match the iotsb_no and adjusted PCI-E.
  - If tablewalk is disabled, set tbw_dme, nullify the transaction.
- Form index for IOTTE and check its size. Fetch the IOTTE from IOTSB in main memory.

  The following errors are detected in parallel.
  - If the SIU returns the tablewalk data with UE or DSN detects the tablewalk data with parity error, set tbw_err, nullify the transaction.
  - If the IOMMU detects parity error on the tablewalk data locally, set tbw_dpe, nullify the transaction.
  - If the IOMMU detects that the allocated replacement entry number mismatch with the replacement entry number in the tablewalk data, set tbw_une, nullify the transaction.
- Install IOTTE in cache, replacing entry if it is full and updating tag and data array portions of cache.

  The following errors are detected in parallel. There is no priority among these errors.
  - If TTE is invalid, set tte_inv, nullify the transaction.
  - If the wrt bit is not set and the current transaction is the write transaction, set tte_prt, nullify the write transaction.
  - Compares the full 16-bit of the PCI-Express requester ID and the device key value if key_valid field is enable. If device key value doesn't match the PCI-Express requester ID, set key_err, nullify the transaction.
- Construct UltraSPARC T2 physical address.

21.3.7 TTE Cache Invalidation Flow (for Both sun4u and sun4v)

- Processor as part of unbind operation or other operation requiring teardown of mapping does store to location in IOTSB in host memory containing the IOTTE to invalidate. Note: All eight IOTTEs in the entry are invalidated.
- Processor does a PIO write to UltraSPARC T2 IOMMU TTE Cache Flush CSR. PIO write data carries the UltraSPARC T2 physical address associated with the given IOTTE. Note: All eight IOTTEs in the entry are invalidated.
- UltraSPARC T2 IOMMU does associative lookup on PA tag portion of IOTTE cache.

- If there is no match, UltraSPARC T2 IOMMU does nothing; otherwise, UltraSPARC T2 IOMMU clears the valid bit for all IOTTEs occupying the entry that matched on the PA tag.

21.3.8 IOMMU Function Description

21.3.8.1 IOTTE Format

The hardware will expect the IOTTEs to include the 26-bit PPN, a write-allowed bit, and a valid bit. Each entry of an IOTTE is 8-bytes and is formatted as shown in TABLE 7.1. This is the same as the Fire TTE format except that data_pa has been resized for 39 bits of physical address space and data_v is moved to bit 0. At sun4v mode, hardware will also expect 16-bit dev_key field, 3-bit Function Number mask and 1-bit key_valid value.

TABLE 21-13 IOTTE Format

Bit	Field	Description
63:48	dev_key	Device Key -- 16-bit PCI-Express requestor ID (bus #, device #, function #)
47:39	—	*Reserved*
38:13	data_pa	Physical Address{38:13}
12:6	data_soft	*Reserved* for software use
5:3	fnm	sun4v mode, function number mask
3	—	*Reserved*
2	key_valid	Device key Valid bit.
1	data_w	Write bit (0 = nonwriteable, 1 = writeable)
0	data_v	Valid bit (0 = not valid, 1 = valid)

The data_pa field is simply the physical page number for the requested data. data_v and data_w control whether the IOTTE is valid and if that page is writeable. Note that depending on the page size for the IOTSB where this IOTTE resides, lower bits of the data_pa may be ignored by hardware.

Note | The sun4v IOTTE format is the same as the sun4u's format.

21.3.8.2 New Fields Introduced by UltraSPARC T2 PIU

The full 16-bit PCI express device ID of the device that initiated the DMA transaction is compared with the 16-bit dev_key value. This comparison only happens if the key_valid bit is set. If it is not, any device will have access to the physical addresses mapped by the IOTTE.

To handle the PCI Express phantom function numbers, the fnm field is used. This function number mask will be **and**ed with the function number in the source device ID and also **and**ed with the function number in the dev_key prior to the comparison. This implementation is to avoid programming inconsistency between the fnm field and dev_key value field in the IOTTE. The following table gives the valid programming of the fnm field.

TABLE 21-14 IOTTE fnm Field

fnm Value	Description	PCIE Dev Capabilities Reg {4:3}
000	All function numbers used for phantom functions	11
001	2 MSBs used for phantom functions	10
011	MSB used for phantom functions	01
111	No phantom function numbers	00

Nonfatal errors will be reported for the following:

- Translation request for DMA write to a page with data_w cleared.
- Translation request to page with data_v cleared.
- Translation request when key_valid and dev_key not equal to source device ID **an**ded with the fnm (lower 3 bits only).

TABLE 21-15 Physical Tag Format

Bit	Field	Type	Description
63:39	—	RO	*Reserved*
38:6	tag	RW	Translation table address tag {38:6}
5:1	—	RO	*Reserved*
0	valid	RW	Valid (0 = not valid, 1 = valid)

TABLE 21-16 Virtual Tag Format

Bit	Field	Type	Description
63:57	—	RO	*Reserved*
56:40	cnt	RW	LRU counter
39:16	vpn or Adjusted vpn	RW	Virtual page number in sun4u mode Adjusted virtual page number in sun4v mode
15:11	$0\ 0000_2$or iotsb_no	RW	00000_2 in sun4u mode iotsb_no for IOTSBDESC in sun4v mode
10:1	—	RO	*Reserved*
0	valid	RW	Valid (0 = not valid, 1 = valid)

TABLE 21-17 TTE Cache Data Format

Bit	Field	Type	TTE Field	Description
63:48	key	RW	dev_key	Device Key -- 16-bit PCIE requester ID
47:44	*Reserved* for software PAR for hardware	RO	NA	Odd Parity ({63:52};{51:48,38:31};{30:19};{18:13,5:0})
43:39	—	RO	NA	*Reserved*
38:13	ppn	RW	data_pa	Physical page number {38:13}
12:6	—	RO	NA	*Reserved*
5:3	fnm	RW	fnm	Function number mask
2	key_vld	RW	key_valid	Device key valid bit.
1	wrt	RW	data_w	Writable (0 = not writeable, 1 = writeable)
0	vld	RW	data_v	Valid (0 = not valid, 1 = valid)

21.3.9 Interrupt Model

PCI Express supports two different interrupt mechanisms: INTx emulation and Message signaled interrupts (MSI). INTx refers to four PCI Interrupt pins, named INTA-D, abbreviated as INTx. On PCI, each device can assert one interrupt per function, with a total of four discreet interrupt pins or wires. The wiring of these interrupt signals on PCI is platform- and system-specific. MSIs are defined in PCI 2.2 as an optional feature but have not been widely adopted by PCI devices, since a PCI device must also support INTx in order to work in all existing systems. MSI support is required in PCI-X devices.

INTx emulation provides software compatibility for firmware and bridge drivers (assuming the PCI Express bridge is register compatible with an existing PCI bridge). However, INTx emulation is expected to be deprecated in future versions of the PCI Express specification. MSIs are mandatory in PCI Express and are software compatible with existing PCI device drivers.

MSIs are inband posted (no acknowledge generated) writes to a preprogrammed address (either 32 bit or 64 bit). MSIs can be generated from PCI Express devices or PCI-X devices behind a PCI Express to PCI-X bridge. Interrupts on the various Sun processor buses are also inband “writes.” However, a processor can NACK an interrupt, causing the interrupter to have to resend the interrupt at a later time.

When PIU generates an interrupt, the source of the interrupt could be from one of PIU’s internal blocks; from a PCI Express INTx Assertion Packet; or from one of PIU’s Event Queues (due to one or more MSIs, Power Management messages, error messages, etc., being received). All interrupt sources are maskable.

PIU uses a subset of interrupt state machine in Schizo and Tomatillo **[Replace code names?]**. An interrupt is always a “level”-sensitive interrupt. An interrupt can be in one of three states: `idle`, `received`, or `pending`. `idle` represents the state where

no interrupts are reported. `received` indicates that an interrupt has been detected and should be delivered to the processor if/when the valid bit is set in its mapping register. `pending` is the state when the interrupt has been queued to be or has been sent to the processor. Any subsequent detection of the same interrupt is ignored until software resets the state machine back to `idle`. The state register for each level-sensitive interrupt can be set to any desired state by software via the Interrupt Clear registers.

Note | Software should ensure that the source of a level-sensitive interrupt is cleared before clearing the interrupt state register via the Interrupt Clear register; otherwise, PIU will continue to reissue the interrupt each time the state register is set to `idle`.

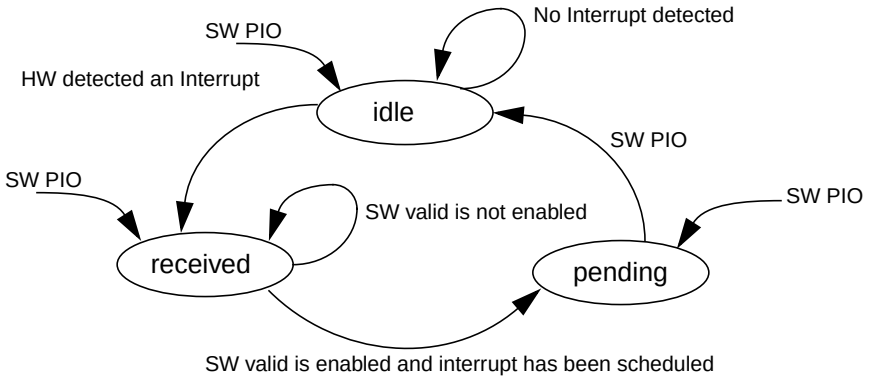


FIGURE 21-5 Interrupt Mondo State Machine

21.3.9.1 Internal Interrupts

The PIU can generate two internal interrupts: a DMU internal interrupt (INO 62) and a PEU block internal interrupt (INO 63).

21.3.9.2 PCI Express INTx Emulation

PIU generates four level-sensitive signals based on state bits that are asserted when PIU sees the appropriate Assert_INTx message and deasserted when it sees the appropriate Deassert_INTx message. Each of the four INTx interrupts has a dedicated interrupt mondo.

From the PCI Express Base Specification, Rev. 1.0a [04/15/03], 2.2.8.1. INTx Interrupt Signaling — Rules, page 65.

"The Assert_INTx/Deassert_INTx Message pairs constitute four 'virtual wires' for each of the legacy PCI interrupts designated A, B, C, and D. The following rules describe the operation of these virtual wires:

- The components at both ends of each Link must track the logical state of the four virtual wires using the Assert/Deassert Messages to represent the active and inactive transitions (respectively) of each corresponding virtual wire.
- An Assert_INTx represents the active going transition of the INTx (x = A, B, C, or D) virtual wire.
- A Deassert_INTx represents the inactive going transition of the INTx (x = A, B, C, or D) virtual wire."

Notes | Duplicate Assert_INTx/Deassert_INTx Messages have no effect but are not errors.

If a PCI-E link goes down, software must clear the INTx state (related to that port) to ensure consistent INTx state.

INTx legacy interrupts are inband, so they can be stalled due to data traffic. An interrupt deassert could be late because it was stalled, it may not be coming because the device did not deassert the interrupt, or it may not be coming because the interrupt is shared by another device. For the second two cases, software will be interrupted again.

INTx legacy interrupts should be treated just like the external interrupts with the additional caveat that you may get spurious INTx legacy interrupts due to a late-arriving INTx deassert message.

21.3.9.3 Event Queue Interrupts

PIU uses event queues to queue up MSIs and valid PCI Express messages received which require software notification. An event queue is tied to a specified processor and generates only one outstanding interrupt mondo for one or more writes to the event queue. Multiple event queues exist primarily to provide hardware-based interrupt distribution among the processors. However, it is possible for software to use an additional event queue for just error messages, for example.

Internal to PIU, each event queue exports a level signal (called NOT_EMPTY for this discussion) which is asserted whenever head ≠ tail. NOT_EMPTY is deasserted if head = tail.

A processor interrupt mondo will be sent to the software programmable target processor if the event queue is enabled, NON_EMPTY is asserted, and the event queue's interrupt state is `idle`. Each event queue has a dedicated interrupt mondo.

Event queues are aligned on a 64-byte boundary. An entry in the event queue is called an event queue record. Event queue records are 64-byte aligned and are 64 bytes large. The base address of an event queue is an MMU address. It can be a

virtual I/O address or a physical address (using MMU bypass). If it is a bypass address, the queue must reside in physically contiguous memory. The head pointer and tail pointer are indexes into the event queue.

Items in the queue are written to the end of the queue using the tail index. Items are removed from the front of the queue using the head index (An item enters the queue at the end of the queue and is removed from the queue when it reaches the front of a queue). “With queues, we speak of the front and the rear of the queue; things enter at the rear and are removed when they ultimately reach the front position.” (Knuth)

When the head index and the tail index are equal, the queue is considered empty. When (tail + 1) = head, the queue is considered full. This means the event queue is full at 127 entries. If PIU attempts to write to the event queue when it is full, it will drop the write, set the overflow bit, and generate an internal interrupt signaling the error.

PIU writes event queue records to the queue using the event queue address indexed by the value in tail. In PIU, the tail is incremented *before* the write has successfully completed. Therefore, the event queue’s tail register cannot be used by software to determine the number of event queue records present in the event queue. PIU never updates the value in head. The event queue head is only writable by software.

System software reads an event record by reading the event queue record at the address of the queue indexed by the value in head. The event queue head may only be incremented after the system software successfully reads one or more event records. Once initialized by system software, the value in tail is never updated by system software. The event queue head is updated by writing a new value to the event queue’s head register. Since the actual head is written to the event queue’s head register, more than one event queue record can be freed up with a single write to the register.

Since software cannot use the tail register to determine the number of event queue records present in the event queue, it is suggested that software should use the first 64-bit value in the event queue record to determine when it has processed the last event queue record. To accomplish this, software should zero the event queue before use and then look for a nonzero value in the first 64-bit value in the event queue record to indicate whether an event queue record is present at the current head. If the head contents are zero, the event queue is empty. Therefore, when software processes an event queue record, it must zero out the first 64-bit value in the event queue *before* writing to the event queue head register to indicate it has processed the event queue record.

For PIU, all event queues live in a contiguous memory region, where the first event queue must be aligned on a programmable 512-Kbyte boundary. Each event queue has a fixed size of 128 entries (one 8-Kbyte page) but is considered full at 127 entries.

Each event queue has the following state included; enable/disable event queue, head index, tail index, overflow error occurred, and the event queue state (`idle`, `active`, `error`). Software can induce any of state transitions with the exception of `idle` to `error`.

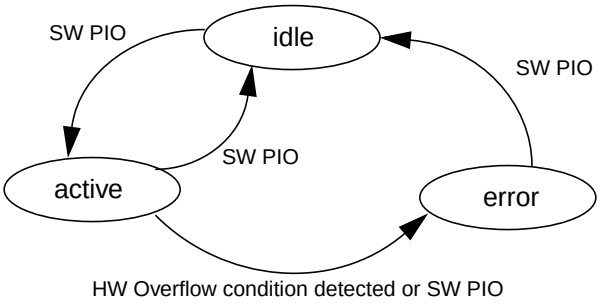


FIGURE 21-6 Event Q State Machine

21.3.9.4 Interrupt Mondos

An interrupt mondo consists of an address and two 64-bit data words. The address identifies the packet as an interrupt packet and also identifies the target thread which will receive the interrupt.

PCI Express Interface Unit can generate 42 unique interrupt mondos. The 42 interrupt mondos break down as follows: 4 interrupts for PCI Express INTx Emulation (INO 20-23), 36 event queue interrupts (INO 24-59), a PCIE leaf internal interrupt (INO 62), and a PEC block interrupt (INO 63). For more details see Chapter 15, *Noncacheable Unit (NCU) and Boot ROM Interfaces*. Each interrupt uses the interrupt state machine specified in *Interrupt Model* on page 512.

PIU can send two differently formatted interrupt mondo's based on the setting of mdo_mode in the Interrupt Mapping registers. The interrupt mondo format is specified below (The last six 64-bit data words are always set to 0).

```
if (mdo_mode = 0)
 mdata0{63:0} = {0_16, threadid{5:0}, ino{5:0}}
 mdata1{63:0} = {0_16}
} else {
    mdata0{63:0}} = {data0{63:12}, threadid{5:0}, ino{5:0}};
     mdata1{63:0} = {data1{63:0}}
}
```

*1 - data0 → Interrupt Mondo Data 0 register ($0062C000_{16}$ / 0_{16})

*2 - data1 → Interrupt Mondo Data 1 register ($0062C008_{16}$ / 0_{16})

Note | To properly use event queues, the interrupt trap handler should support, directly or via a device-specific extension, reading event queue records and scheduling interrupt handlers (based on the event queue records) to run at the appropriate interrupt level. An event queue can have multiple event queue records present at any given time. Each event queue record corresponds to a unique interrupt with its own interrupt level (that is, an event queue can have event queue records enqueued that will run at different interrupt levels).

PIU also supports four interrupt groups that allow multiple interrupts to be issued at a time. Schizo and Tomatillo **[replace code names]** can issue only one interrupt at a time. Interrupt mondos are assigned to an interrupt group. A given interrupt group can have only one interrupt outstanding at a time (that is, until a processor acknowledges the interrupt). It is expected that software will use one interrupt group per processor present in the system.

Interrupt Mondo State includes valid/invalid entry, current interrupt state, Target thread ID, bit 63 for data0 (1-bits), and data1 (64-bits), interrupt group.

21.3.9.5 Interrupt Mondo INO Mapping Table

TABLE 21-18 Interrupt Mondo INO Mapping Table

INO	Function
INO 0 - 19	*Reserved*
INO 20 - 23	Four interrupts for PCI Express INTx Emulation • 20 INTA • 21 INTB • 22 INTC • 23 INTD
INO 24 - 59	36 Event Queue interrupts
INO 60 - 61	*Reserved*
INO 62	DMU internal interrupt
INO 63	PEU internal interrupt

21.3.9.6 Interrupt Relocation

Not all the event queues have to be used. In normal system operation, it is expected that there will be no more than one event queue active per processor in the system (at least for handling MSIs). Software typically retargets interrupts when more threads come online or when some threads go offline. Both conventional discrete interrupts and MSIs may be retargeted to different threads upon software request.

To retarget an interrupt, software must clear the valid bit of the interrupt mapping register, and then busy-wait until the interrupt state is no longer `pending`. If an interrupt is not in progress when the valid bit is cleared, this is immediately satisfied. If an interrupt is currently being handled, software may wait quite a while. Once the interrupt state is `idle`, software can then change the thread ID field of the mapping register to a new thread ID, and then renable the interrupt by setting the valid bit in the interrupt mapping register.

An MSI cannot be dynamically relocated to a new event queue unless software can ensure that the MSI is not in progress and that the device will not send that particular MSI while the MSI is the process of being retargeted.

21.3.9.7 Data Flushing

No special sync or sync flush operations within PIU are required to synchronize data with interrupts. PIU flushes all DMA writes in PCI Express's virtual channel #0 before returning the response to a PIO read. PCI to PCI Express bridges will also flush writes before returning the response to a PIO read.

Note | PCIE devices can generate zero-byte PCIE reads and/or writes to PIU for data synchronization. PIU will handle these operations without error *as long as* they are to a valid MMU mapped address. *The PCIE specification does not specify how the address for the zero-byte operation is determined. This could be an issue if PCIE devices are hardwired to work with Intel bridges. We must look at this when selecting PCIE devices.*

The latest PCIE Specification (1.0, July 22) says the following things about zero byte reads and writes.

- A write request with a length of 1 DW with no bytes enabled is permitted and has no effect at the Completer.
- If a read request of 1 DW specifies that no bytes are enabled to be read (1st DW BE{3:0} field = 0000_2), the corresponding Completion must specify a length of 1 DW and include a data payload of 1 DW. The contents of the data payload within the Completion packet is unspecified and may be any value.

- A memory read request of 1 DW with no bytes enabled, or "zero-length read," may be used by devices as a type of flush request. For a requester, the flush semantic allows a device to ensure that previously issued posted writes have been completed at their PCI Express destination.
- The flush semantic has wide application, and all Completers must implement the functionality associated with this semantic. Because a requester may use the flush semantic without comprehending the characteristics of the Completer, Completers must ensure that zero-length reads do not have side effects. This is really just a specific case of the rule that in a non-prefetchable space, nonenabled bytes must not be read at the Completer. Note that the flush applies only to traffic in the same traffic class as the zero-length read.

21.3.9.8 Event Queue Records

Event queues (EQ) are used to queue interrupts with additional state information. An entry in the event queue is called an event queue record. An event queue record can be written by DMU as a result of a valid PCI Express message received which requires software notification or a MSI/MSI-X being received. More information on what messages generate event queue records and how they are directed to a particular event queue can be found in the following sections.

An MSI/MSI-X as seen on the PCIE link looks like any other DMA write to PIU. The only way PIU knows that it is a MSI/MSI-X is if it matches a range that is set up by two registers depending upon the DMA is a 64-bit-addressed DMA or a 32-bit-addressed DMA. These two registers (634000_{16}, 634008_{16}) are set up by software to whatever PCIE address the software wants. *Any* DMA write seen matching in this range will be treated as an MSI/MSI-X and will be sent to the IMU and removed from normal DMA pipeline. So if external devices have normal DMA going to the same address range as is set up to be the range for MSIs that would be *bad*.

Note This address matching for MSI/MSI-X *cannot* be disabled. The default address for both the 64-bit and the 32-bit MSI address matching register is an address of all zeros. Any DMA going to this address even if the necessary steps to set up MSIs are not taken *will* be removed from the DMA pipeline and passed to the IMU. It is mandatory that no DMA address be sent to this address range if the MSI functionality is not being utilized.

After the MSIs/MSI-As makes it to the IMU, that PCIE address is no longer used, and it is discarded. No matter what it is, it will not make it to the MMU. A new address is created from the EQ base address register + EQ number offset plus Tail pointer. Depending on what software set the EQ base address register to determines what the MMU does with the Event Q write. If software set it up as a bypass address, it will be treated as *bypass* by the MMU; if software set it up to be translated, it will be translated.

An event queue record contents include PCI Express message/PIU Message, Fmt & Type, Packet Length, PCI Express Address{15:2}, Requester ID, MSI Data{15:0}, msgcode{7:0}. Not all of the fields are valid for a given message. Software can process an event queue record with a single 64-bit read from the event queue record.

All the fields in the EQ record are from the PCIE header that the MSI/MSI-X/Msg came with. Please see the PCIE specification for all the details about that. These fields are put there unmodified. The length is in dw and for any "good" MSI/MSI-X, it should always be 1 or it will be detected as a malformed MSI.

TABLE 21-19 Event Queue Record Format

Data Words	Bit	Field	Description
0	63	—	*Reserved*. Will be set to zero.
0	62:56	fmt/type	111 1000 for 64-bit addressed MSI. 101 1000 for 32-bit addressed MSI. 011 0xxx for Messages where xxx complies with routing rules in PCIE spec.
0	55:46	length	Set to the length of the data.
0	45:32	addr{15:2}	Copy of the address bits 15:2.
0	31:16	rid	Requester ID.

0	15:00	data0	if (fmt/type = 111 1000 or 101 1000) {
			data0 ← MSI data{15:0}
			15:8 = byte 1 of MSI/MSI-X Data
			7:0 = byte 0 of MSI/MSI-X Data
			}
			else data0 ← (0_2, MSGCODE{7:0})

1	63:16	addr{63:16}	Copy of the address bits 63:16.

1	15:00	data1	if (FMT/TYPE = 1111000 or 1011000) {
			data1 ← MSI-X data{31:16} (Field *not* valid for MSI)
			15:8 = byte 3 of MSI-X Data (Field *not* valid for MSI)
			7:0 = byte 2 of MSI-X Data (Field *not* valid for MSI)
			}
			else data1 ← (copy of address bits{13:0}, 0_2)

2 - 7	63:00	—	*Reserved*. Will be set to zero.

PCI Express Messages. PIU will generate event queue records for the following PCI Express messages: Correctable error, Uncorrectable error, Fatal error, PME, and PME_TO_Ack. Each of these messages can be individually targeted to an event queue using the PCI Express Message Mapping registers. For example, Uncorrectable error and Fatal error may be targeted to event queue #15; and Correctable error, PME, and PME_TO_Ack, and may be targeted to event queue #14. As an example for this case, PME messages will then always generate event queue records in event queue #14.

Message Signaled Interrupts (MSIs and MSI-Xs). MSIs and MSI-Xs are inband posted writes to a preprogrammed address that can be either a 32-bit or a 64-bit address. MSIs and MSI-Xs allow up to 32 unique interrupts to be generated from

each function within a device. PIU will check either the 32-bit or the 64-bit address on a PCI Express write to determine if it is a MSI/MSI-X. If the incoming write is 64-bit-addressed the MSI 64-bit address register ($0063\ 4008_{16}$) will be used for the comparison. If the incoming write is 32 bit addressed, the MSI 32-bit Address register ($0063\ 4000_{16}$) will be used for the comparison. In sun4v mode, and if bit 63 of EQ Base address register is set to 0, bit 63 of the MSI 64-bit address register will not be used for comparison. The address used to determine if the write is an MSI or MSI-X is the same. Two separate address *cannot* be specified to distinguish MSIs from MSI-Xs. PIU relies on SW to distinguish MSIs from MSI-Xs by identifying the requester ID in the EQ record.

A 16-bit Message Data field is transmitted as part of the MSI. A 32-bit Message Data field is transmitted as part of the MSI-X. The contents of this field are programmable and are located in the function's PCI/PCI Express configuration space (capability list).

If a device function reports that it is Multiple Message Capable, it will also report the maximum number of unique MSIs or MSI-Xs that it can generate (from 1 to 32 in powers-of-2). System software can enable all or a subset of the MSIs or MSI-Xs (from 1 to MAX in powers-of-2). If a device function is configured to generate multiple unique MSIs or MSI-Xs, the device will use the lower data bits in the MSI 16-bit data or the lower data bits in the MSI-X 32-bit field to signify the interrupt number (from 0 to MAX - 1). The number of bits used is n where 2^n = MAX. The lower n bits of the Message Data field located in the function's PCI configuration space are ignored for this case.

UltraSPARC T2 implements an MSI/MSI-X mapping table, located within IMU, to provide interrupt distribution and interrupt coalescing (for MSIs/MSI-Xs that do not require interrupts to be acknowledged by software). Since PCI Express does not allow us to set an individual 16-bit data field for each unique MSI or the individual 32-bit data field for each unique MSI-X that the device function supports, the MSI/MSI-X mapping table will map an MSI/MSI-X to an event queue. It also will prevent an interrupt being written multiple times into an event queue. This prevents the device from overflowing the event queue with the same interrupt. The number of MSIs/MSI-Xs supported by PIU is limited by the size of the MSI/MSI-X mapping table. PIU mapping table supports 256 MSIs/MSI-Xs; therefore, PIU supports up to 256 MSIs/MSI-Xs. The MSI/MSI-X mapping table is indexed into using the lower 8 bits of the 16-bit MSI data field or the lower 8 bits of the32-bit MSI-X data field.

A device function's MSIs/MSI-Xs must be mapped both aligned and contiguous within the MSI mapping table (that is, if the function supports four interrupts, those four interrupt must take up four contiguous entries and be aligned on a four-entry boundary in the MSI/MSI-X mapping table).

The MSI/MSI-X mapping state includes valid/invalid entry, OK to write to EQ, and the EQ number to write Event Record to.

Note | Before processing an MSI, software must set the "OK to write to EQ" bit in the MSI state and then read back the MSI mapping register to ensure that the write completed.

21.3.10 EQ Address Translation

UltraSPARC T2 supports two types of EQ address. One is physical address which bypasses the IOMMU translation, the other one is virtual address which will be translated by IOMMU. Since IOMMU supports two different modes: SUN4U mode/ SUN4V mode. The EQ virtual address would be formed differently in different mode, and IOMMU will be operated differently while translating the EQ virtual address.

21.3.10.1 SUN4U mode EQ address translation:

SUN4U mode EQ address translation is enables by clearing SUN4V_EN bit in IOMMU control and status register and also clearing bit[63] of Event Queue Base address register. During this mode, EQ_Base[62:32} are don't care but are recommended to be programmed as zero.

The 32-bit EQ virtual address seen by IOMMU will be as follows:

{EQ_Base[31:19], EQ_Number[5:0], EQ_Tail_Pointer[6:0], 6'b0}

SW needs to make sure the EQ_Base{31:19} is programmed appropriate so that EQ virtual address will not exceed the virtual address space defined in IOMMU.

21.3.10.2 SUN4V mode MSI/MSI-X EQ address translation:

SUN4V mode MSI/MSI-X EQ address translation is enabled by setting SUN4V_EN bit in IOMMU control/status register and clearing bit[63] of Event Queue Base Address Register.

During this mode, the following is the programming guideline to ensure that IOMMU can be operated correctly.

- UltraSPARC T2 supports up to 256 MSI/MSI-X from PCI-E devices and it also supports 36 EVENT Queues and 32 IOTSBs. SW should partition the PCIE devices into the 36 Event Queues based on the IOMMU partition. The devices which are in the same IOMMU partition shares the same EQs.
- During this mode, the VA bit[63] and requester ID{BDF#) from the MSI/MSI-X PCI-E post write TLP header will be used for IOMMU partition. Bit[63] of MSI/ MSI-X mapping register will not be used to compare against the VA bit[63] of PCI-E post write TLP header.

- During this mode, EQ_BASE[62:40] should be programmed to zero to avoid SUN4V_VA_OOR reported in IOMMU. The 64-bit virtual address seen by IOMMU should be

  {VA[63], 22'b0, EQ_BASE{39:19], EQ_Number[5:0], EQ_Tail_Pointer[6:0], 6'b0}.

SW need to make sure that EQ_BASE{39:19] is programmed appropriate so that the adjust_vpn will not exceed any VA address space for all partitions.

- During this mode, the EQ address for PCI-Express messages will continue to be a physical address and bypass the IOMMU translation. The bypass address for PCI-Express message EQ seen by IOMMU will be

  {14'b1, 11'b0, EQ_BASE{38:19], EQ_Number[5:0], EQ_Tail_Pointer[6:0], 6'b0}

21.3.11 Power Management

PIU does not support the PCI Express link power management state L2, since it does not support being powered from auxiliary power. Therefore, PIU cannot be used to provide features such as Wake On Lan (WOL).

PIU supports PCI Express link active state power management state L0s and L1. Both of these behaviors can be enabled/disabled by software and default to disabled.

PIU supports entering the PCI Express link power management state L1 on an upstream port due to a request from the downstream device attached to that port. This behavior can be enabled/disabled by software and defaults to disabled. PIU will never generate a request downstream to enter L1 state.

There are four power management messages in PCI Express:

- PM_Active_State_NACK
- PME_Turn_Off
- PM_PME
- PME_TO_Ack

PIU will silently handle PM_Active_State_NACK messages. There is no software visibility for these messages.

The PME_Turn_Off message will be generated by PIU when software writes the gen_pme_off bit in the PME Turn Off Generate register.

The PM_PME and PME_TO_Ack messages will be written to programmable event queue(s) by PIU. PIU will do no further processing on behalf of these messages. It is up to software to handle these messages (that is, time-out a PME_TO_Ack message in response to a PME_Turn_Off message).

The following describes the sequence required to transition from PCI Express link power management state L0 to PCI Express link power management state L2/L3 Ready.

Software directs all functions of a downstream component to D3(hot). The downstream component then initiates the transition of the link to L1 as required by the PCI Express specification. Software then waits for the link to transition into L1 (reading the PEU CXPL Core Status register to determine the link LTSSM state). If the link transitions to L1 (13_{16}, l1_idle), software then causes PIU to broadcast the PM_Turn_Off message by writing to the pto bit in the PME Turn Off Generate register. This message causes the link to transition back to L0 in order to send the PM_Turn_Off message and to enable the downstream component to respond with PM_TO_Ack. The PM_TO_Ack message is placed in an Event Queue by PIU for software to process. It is up to software to time-out if a PME_TO_Ack is not received. After the PM_TO_Ack is sent, the downstream component then initiates the L2/L3 Ready transition protocol. Software then waits for the link to transition into L2/L3 Ready (reading the PEU CXPL Core Status register to determine the link power state). If the link transitions to L2/L3 Ready (14_{16}, l2_idle), software can then power-off the link by disable the SerDes (writing all 1_2's to bit 33:16 of DMU ILU Diagnostic register).

Handling of the failure to transition to L1 or L2/L3 Ready is software implementation specific.

21.3.12 Endianness

Since PCI Express is a serial bus, PIU does not need to worry about the "byte twisting" done on previous generations of Sun's PCI bridges. PIU ensures that the byte stream order is maintained.

PCI Express packet headers are big endian. PCI Express data is endian neutral (that is, a byte stream). For places where a portion of the PCI Express header/data is placed into memory (for example, event queue records), they will maintain big-endian structure (that is, PIU will never swap bytes around).

21.3.13 Error Register Overview

For *each* error bit, five required registers are implemented in the PIU. An optional sixth register may be implemented for some error types to log any helpful additional information. This register may be shared between multiple but similar errors. The registers are as follows:

- **Interrupt Status register.** Indicates whether the particular error is set and enabled for interrupt. Software will read this register to determine the error interrupt source.

- **Error Status Clear register.** Used by software to clear the error. Can also be read by software to determine if the error is set, regardless of whether it will cause an interrupt (for the case when interrupts are not enabled).
- **Interrupt Enable register.** Used by software to enable a particular error for interrupt.
- **Error Log Enable register.** Used by software to enable a particular error for logging. Errors must be enabled for logging in order to be enabled for and cause an interrupt.
- **Error Status Set register.** Used by software to force a particular error. If interrupts are enabled for the particular error, then setting this register will cause an interrupt. Diagnostic use only.
- **Error Log register.** This is an optional register that holds any necessary additional information that is associated with a particular error. Such information might be transaction address, ID, command type, etc.

For each error bit, PIU maintains Primary and Secondary Error register bits, used as follows: On the first occurrence of an error the Primary bit is set. If enabled, an interrupt will occur and software will handle and eventually clear the error status from the registers. If any error from the same group were to occur again before software cleared the Primary event, then the Secondary bit will be set and if enabled, an interrupt will occur. If any error from the same group were to occur after hardware handled and cleared the Primary condition, then, of course, the Primary bit will again be set rather than the Secondary.

Programming Notes

The secondary error interrupt should not be enabled without the primary error interrupt enabled. Hardware will still function properly only reporting the secondary error (should one occur), but software has the potential to miss primary errors by programming in this manner.

Error Log registers are freely loaded, only holding data should an error occur (for example, the log register stops loading and is "locked"). Once an error has occurred, the log register can be read for relevant data. Writing to the corresponding Error Status Clear register will clear the error and reenable the log register to start clocking data once again.

Errors may be grouped when errors share a common log register because they require the same additional information. In this case, only information associated with the primary error is captured. Should any secondary errors or other primary errors in the same group occur, their information will not be captured since the log register is already locked by the first primary error.

21.3.14 Performance Register Overview

PIU provides nine performance counter registers throughout the chip to aid in analysis, debugging, performance-tuning applications, and link-stability analysis. The registers track events in the physical layer, link layer, transaction layer, MMU and interrupt unit.

PIU Performance Counter registers are listed in TABLE 21-20.

TABLE 21-20 PIU Performance Counter Registers

Location (block)	Register	Size (bits)	Clock Domain
DMU	IMU Performance Counter Zero	64	IO Clk
DMU	IMU Performance Counter One	64	IO Clk
DMU	MMU Performance Counter Zero	64	IO Clk
DMU	MMU Performance Counter One	64	IO Clk
PEU	PEU Performance Counter Zero	64	250 MHz
PEU	PEU Performance Counter One	64	250 MHz
PEU	PEU Performance Counter Two	64	250 MHz
PEU	PEU Bit Error Counter I	64	250 MHz
PEU	PEU Bit Error Counter II	64	250 MHz

Each register group contains a corresponding Select register used to select the specific input to the counter, except the Bit Error counters.

Programming Note Counting does not occur when the select is set to zero. Counters are cleared by software when writing them to zero (or by reset) and not by transitioning the select. If the select is transitioned between various events without resetting the counters, an aggregate value will be contained in the counters (that is, the counters keep on counting without returning to zero).

See the specific register descriptions for additional details on the counters.

21.3.15 Link Layer Thresholds

There are three types of thresholds in link layer:

- Ack/Nak Latency Timer Threshold (PEU CXPL DLL AckNak Latency Threshold register). Round-trip latency time limit; ACK/NAK timer will expire when it reaches this limit.
- Replay Timer Threshold (PEU CXPL DLL Replay Timer Threshold register). Replay time limit; replay timer will expire when it reaches this limit. A replay is reinitiated due either to reception of a NACK or to replay timer expiration.
- Bit Lock Timer Threshold (PEU CXPL MACL Lane Skew/Receiver/Bit_Lock Control register). Bit lock timer limit; PEU starts to enable SerDes's comma detection logic when bit lock timer (per lane basis) reaches the limit. Bit lock timer is reset whenever SerDes detect L0s signal and resumes the counting whenever SerDes detect valid differential signals in the receiver channels.

21.3.15.1 Ack/Nak Latency Timer Threshold

The value to program into PEU CXPL DLL AckNak Latency Threshold register is a function of the link operating widths and max payload size. TABLE 21-21 is the recommended value for software to program this register.

TABLE 21-21 ACK/NACK Latency Timer Threshold Table

Max Payload Size	Link Operation Width - X1	Link Operation Width - X2	Link Operation Width - x4	Link Operation Width - x8
128 byte	ED_{16} (237 dec)	80_{16} (128 dec)	49_{16} (73 dec)	43_{16} (67 dec)
256 byte	$1A0_{16}$ (416 dec)	$D9_{16}$ (217 dec)	76_{16} (118 dec)	$6B_{16}$ (107 dec)
512 byte	$22F_{16}$ (559 dec)	121_{16} (289 dec)	$9A_{16}$ (154 dec)	56_{16} (86 dec)

If L0s is enabled, the Tx_L0s adjustment factor will be added to the ACK/NACK Threshold register by hardware. Software doesn't have to do it.

If Extended Sync = 0, then:

Tx_L0s adjustment factor = (received FTS number) x 4.

If Extended Sync = 1, then:

Tx_L0s adjustment factor = (4096) x 4.

The received FTS number is derived from rcv_fts_num{35:28} in the PEU CXPL Core Status register.

21.3.15.2 Replay Timer Threshold

The value to program into PEU CXPL DLL Replay Timer Threshold register is a function of the link operating widths and max payload size. TABLE 21-22 is the recommended value for software to program this register These values are taken from PCI-E spec plus a margin of 25%.

TABLE 21-22 Replay Timer Threshold Table

Max Payload Size	Link Operation Width - X1	Link Operation Width - X2	Link Operation Width - x4	Link Operation Width - x8
128 byte	379_{16} (889 dec)	$1E0_{16}$ (480 dec)	112_{16} (274 dec)	FC_{16} (252 dec)
256 byte	618_{16} (1560 dec)	$32D_{16}$ (813 dec)	$1BA_{16}$ (442 dec)	192_{16} (402 dec)
512 byte	831_{16} (2097 dec)	$43B_{16}$ (1083 dec)	242_{16} (578 dec)	143_{16} (323 dec)

The Rx_L0s adjustment factor will be added to the Replay Timer Threshold register by hardware. Software doesn't have to do it.

Rx_L0s adjustment factor = (Transmit FTS number) x 4.

The Transmit FTS number is derived from n_fts{23:16} in the PEU CXPL MACL/PCS control register.

21.3.16 Register Reset Behavior and Software Access to PIU

Two reset signals are internal to PIU and referenced in the Reset Value field of all register descriptions contained in Section 21.4.

TABLE 21-23 PIU Reset Value Definitions

Reset	Comment
rst_l	All registers reset except error log/status registers, (see actual register descriptions)
por_l	All registers reset (including those reset by rst_l)

In general terms, rst_l is considered a warm reset since most registers are reset by it, but some are not and instead maintain their state across the reset activity. In contrast, por_l is considered a power-on reset since all registers in PIU are affected by it.

21.4 CSR Fields and Bits

21.4.1 General Information

PIU's Configuration and Status registers (CSRs) are described in this section. The following abbreviations are used throughout this chapter:

- RO — Read-only: a register field that is not writeable.
- RW - Read/write: A standard register field.
- RW1C — Read/write 1 to clear: Writing a 0 to bits in this field has no effect, but writing a 1 to a bit in this field will cause that bit to be set to 0.
- RW1S —Read/write 1 to set: Writing a 0 to bits in this field has no effect, but writing a 1 to a bit in this field will cause that bit to be set to 1.
- WO — Write-only: reads of this field return 0.
- DMA — Direct Memory Access: A transaction initiated by a leaf block resulting in a memory transaction.
- DVMA — Direct Virtual Memory Access: A subclass of DMA transactions in which the address of the leaf transaction is treated as a virtual address and translated by an MMU.

Note that *Reserved* bits read as 0.

21.4.1.1 Access Size

Most registers in PIU have a natural access size of 8 bytes. Register accesses must always be done with the appropriate access size, or undefined behavior will result.

21.4.1.2 Unimplemented Addresses

Any address within PIU's register address space that is not documented here as belonging to a specific register is reserved and should not be read or written by software. Although PIU will "own" and respond to its entire memory address range, a read to an unimplemented address can return a read error or undefined data and a write will be ignored. Since PIU will cover its entire allocated address space, system hangs due to accesses to unimplemented addresses will be reduced, although this behavior should still be avoided by software.

21.4.1.3 Physical Addresses

All register addresses are documented as offsets within PIU's 8-Mbyte noncacheable configuration region. For more details about PIU's address regions, see Chapter 15, *Noncacheable Unit (NCU) and Boot ROM Interfaces*.

21.4.2 Register Map

TABLE 21-24 summarizes all the on-chip registers within PIU. All accesses to these registers must be 8 byte, 8 byte aligned. All offsets that are not listed are *Reserved*.

TABLE 21-24 PIO Register Map *(1 of 5)*

Offset	Register
PCIE CSR Space	
$6010A0_{16}$–$6011D8_{16}$, $6011F0_{16}$, $6011F8_{16}$	Interrupt Mapping
$6014A0_{16}$–$6015D8_{16}$, $6015F0_{16}$, $6015F8_{16}$	Interrupt Clear
$601A00_{16}$	Interrupt Retry Timer
$601A10_{16}$	Interrupt State Status register 1
$601A18_{16}$	Interrupt State Status register 2
$60B000_{16}$	INTX Status
$60B008_{16}$	INT A Clear
$60B010_{16}$	INT B Clear
$60B018_{16}$	INT C Clear
$60B020_{16}$	INT D Clear
610000_{16}	Event Queue Base Address
611000_{16}–611118_{16}	Event Queue Control Set
611200_{16}–611318_{16}	Event Queue Control Clear
611400_{16}–611518_{16}	Event Queue State
611600_{16}–611718_{16}	Event Queue Tail
611800_{16}–611918_{16}	Event Queue Head
620000_{16}–$6207F8_{16}$	MSI Mapping
628000_{16}–$6287F8_{16}$	MSI Clear
$62C000_{16}$	Interrupt Mondo Data 0
$62C008_{16}$	Interrupt Mondo Data 1
630000_{16}	ERR COR Mapping
630008_{16}	ERR NONFATAL Mapping
630010_{16}	ERR FATAL Mapping
630018_{16}	pm PME Mapping
630020_{16}	PME To ACK Mapping
631000_{16}	IMU Error Log Enable
631008_{16}	IMU Interrupt Enable
631010_{16}	IMU Interrupt Status

TABLE 21-24 PIO Register Map *(2 of 5)*

Offset	Register
631018_{16}	IMU Error Status Clear
631020_{16}	IMU Error Status Set
631028_{16}	IMU RDS Error Log
631030_{16}	IMU SCS Error Log
631038_{16}	IMU EQS Error Log
631800_{16}	DMC Core and Block Interrupt Enable
631808_{16}	DMC Core and Block Error Status
632000_{16}	IMU Performance Counter Select
632008_{16}	IMU Performance Counter Zero
632010_{16}	IMU Performance Counter One
634000_{16}	MSI 32-bit Address
634008_{16}	MSI 64-bit Address
634018_{16}	Mem 64 PCIE Offset
640000_{16}	MMU Control and Status
640008_{16}	MMU TSB Control
NCU 8000002030_{16}	MMU TTE Cache Flush Address
640108_{16}	MMU TTE Cache Invalidate
641000_{16}	MMU Error Log Enable
641008_{16}	MMU Interrupt Enable
641010_{16}	MMU Interrupt Status
641018_{16}	MMU Error Status Clear
641020_{16}	MMU Error Status Set
641028_{16}	MMU Translation Fault Address
641030_{16}	MMU Translation Fault Status
642000_{16}	MMU Performance Counter Select
642008_{16}	MMU Performance Counter Zero
642010_{16}	MMU Performance Counter One
646000_{16}–$6461F8_{16}$	MMU TTE Cache Virtual Tags
647000_{16}–6471F8	MMU TTE Cache Physical Tags
648000_{16}–$648FF8_{16}$	MMU TTE Cache Data

TABLE 21-24 PIO Register Map *(3 of 5)*

Offset	Register
649000_{16}–649078_{16}	MMU DEV2IOTSB
649100_{16}–$6491F8_{16}$	MMU IOTSBDESC
651000_{16}	ILU Error Log Enable
651008_{16}	ILU Interrupt Enable
651010_{16}	ILU Interrupt Status
651018_{16}	ILU Error Status Clear
651020_{16}	ILU Error Status Set
651800_{16}	PEU Core and Block Interrupt Enable
651808_{16}	PEU Core and Block Interrupt Status
652000_{16}	ILU Diagnostic
653000_{16}	DMU Debug Select register for DMU Debug Bus A
653008_{16}	DMU Debug Select register for DMU Debug Bus B
653100_{16}	DMU PCI Express Configuration
660000_{16}–$6600F8_{16}$	Packet Scoreboard DMA register set
664000_{16}–664078_{16}	Packet Scoreboard PIO register set
670000_{16}–$6700F8_{16}$	Transaction Scoreboard register set
670100_{16}	Transaction Scoreboard Status
680000_{16}	PEU Control
680008_{16}	PEU Status
680010_{16}	PEU PME Turn Off Generate
680018_{16}	PEU Ingress Credits Initial
680100_{16}	PEU Diagnostic
680200_{16}	PEU Egress Credits Consumed
680208_{16}	PEU Egress Credit Limit
680210_{16}	PEU Egress Retry Buffer
680218_{16}	PEU Ingress Credits Allocated
680220_{16}	PEU Ingress Credits Received
681000_{16}	PEU Other Event Log Enable
681008_{16}	PEU Other Event Interrupt Enable
681010_{16}	PEU Other Event Interrupt Status

TABLE 21-24 PIO Register Map *(4 of 5)*

Offset	Register
681018_{16}	PEU Other Event Status Clear
681020_{16}	PEU Other Event Status Set
681028_{16}	PEU Receive Other Event Header1 Log
681030_{16}	PEU Receive Other Event Header2 Log
681038_{16}	PEU Transmit Other Event Header1 Log
681040_{16}	PEU Transmit Other Event Header2 Log
682000_{16}	PEU Performance Counter Select
682008_{16}	PEU Performance Counter Zero
682010_{16}	PEU Performance Counter One
682018_{16}	PEU Performance Counter Two
683000_{16}	PEU Debug Select A
683008_{16}	PEU Debug Select B
690000_{16}	PEU Device Capabilities
690008_{16}	PEU Device Control
690010_{16}	PEU Device Status
690018_{16}	PEU Link Capabilities
690020_{16}	PEU Link Control
690028_{16}	PEU Link Status
690030_{16}	PEU Slot Capabilities
691000_{16}	PEU Uncorrectable Error Log Enable
691008_{16}	PEU Uncorrectable Error Interrupt Enable
691010_{16}	PEU Uncorrectable Error Interrupt Status
691018_{16}	PEU Uncorrectable Error Status Clear
691020_{16}	PEU Uncorrectable Error Status Set
691028_{16}	PEU Receive Uncorrectable Error Header1 Log
691030_{16}	PEU Receive Uncorrectable Error Header2 Log
691038_{16}	PEU Transmit Uncorrectable Error Header1 Log
691040_{16}	PEU Transmit Uncorrectable Error Header2 Log
$6A1000_{16}$	PEU Correctable Error Log Enable
$6A1008_{16}$	PEU Correctable Error Interrupt Enable

TABLE 21-24 PIO Register Map *(5 of 5)*

Offset	Register
$6A1010_{16}$	PEU Correctable Error Interrupt Status
$6A1018_{16}$	PEU Correctable Error Status Clear
$6A1020_{16}$	PEU Correctable Error Status Set
$6E2000_{16}$	PEU CXPL/SerDes Revision
$6E2008_{16}$	PEU CXPL AckNak Latency Threshold
$6E2010_{16}$	PEU CXPL AckNak Latency Timer
$6E2018_{16}$	PEU CXPL Replay Timer Threshold
$6E2020_{16}$	PEU CXPL Replay Timer
$6E2040_{16}$	PEU CXPL Vendor DLLP Message
$6E2050_{16}$	PEU CXPL LTSSM Control
$6E2058_{16}$	PEU CXPL DLL (Data Link Layer) Control
$6E2060_{16}$	PEU CXPL MACL/PCS (Media Access Layer / Physical Coding Sublayer) Control
$6E2068_{16}$	PEU CXPL MACL Lane Skew/Receiver Detection/Bit Lock Timer Control
$6E2070_{16}$	PEU CXPL MACL Symbol Number
$6E2078_{16}$	PEU CXPL MACL Symbol Timer
$6E2100_{16}$	PEU CXPL Core Status
$6E2108_{16}$	PEU CXPL Event/Error Log Enable
$6E2110_{16}$	PEU CXPL Event/Error Interrupt Enable
$6E2118_{16}$	PEU CXPL Event/Error Interrupt Status
$6E2120_{16}$	PEU CXPL Event/Error Status Clear
$6E2128_{16}$	PEU CXPL Event/Error Set
$6E2130_{16}$	PEU Link Bit Error Counter I
$6E2138_{16}$	PEU Link Bit Error Counter II
$6E2200_{16}$	PEU SerDes PLL Control
$6E2300_{16}$–$6E2338_{16}$	PEU SerDes Receiver Lane Control
$6E2380_{16}$–$6E23B8_{16}$	PEU SerDes Receiver Lane Status
$6E2400_{16}$–$6E2438_{16}$	PEU SerDes Transmitter Control
$6E2480_{16}$–$6E24B8_{16}$	PEU SerDes Transmitter Status
$6E2500_{16}$–$6E2508_{16}$	PEU SerDes Test Configuration

21.4.3 PCI-E Registers

21.4.3.1 Interrupt Mapping Registers ($006010A0_{16}$-$006011D8_{16}$, $006011F0_{16}$, $006011F8_{16}$ / 0_{16})

The Interrupt Mapping registers are 42 registers, 1 for each mondo, starting from Mondo 20–Mondo 59, Mondo 62, and Mondo 63. These are read/write registers that software uses to set up each of the supported 42 Mondos. Through these registers, software on a per-mondo basis can set up the mondo mode, thread ID of the target CPU the mondo is destined for, enable and disable the mondo's valid bit, and select which Interrupt Controller (group controller) the mondo will use.

TABLE 21-25 Interrupt Mapping Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63	mdo_mode	rst_l	0_{16}	RW	This bit is used to select which of the two mondo formats the mondo will use. A value of 1 = data-bearing mondo; a value of 0 = ,non-data-bearing mondo (normal mondo) The value of this bit will be used as bit 63 of the first data word in the Mondo vector. In general, EQ mondos should have this bit set to 1 and non EQ mondos should set this bit to 0.
62:32	—				*Reserved*
31	v	rst_l	0_{16}	RW	Valid bit: When set to 0, interrupt will not be dispatched to CPU from this PCIE leaf. Has no other impact on interrupt state.
30:25	threadid	rst_l	0_{16}	RW	Thread ID of processor that this interrupt will be sent to.
24:10	—				*Reserved*
9:6	int_cntrl_num	rst_l	0_{16}	RW	Interrupt controller number. This is used to select which Interrupt controller will issue the interrupt. This is a 1-hot value; only 1 bit may be selected at a time. Valid values are as follows: 0000 – No controller selected. 0001 – Interrupt controller 0. 0010 – Interrupt controller 1. 0100 – Interrupt controller 2. 1000 – Interrupt controller 3. If other values are programmed, that is a programming error and the results are undefined
5:0	—				*Reserved*

21.4.3.2 Interrupt Clear Registers ($006014A0_{16}$–$06015D8_{16}$, $006015F0_{16}$, $006015F8_{16}$ / 0_{16})

The Interrupt Clear registers are 42 registers, 1 for each mondo, starting from Mondo 20–Mondo59, Mondo 62, and Mondo 63. These are read/write registers that software uses to read and write the state of the supported 42 mondos. Through these registers, software on a per-mondo basis can at any time obtain the current value of the mondo state or direct the mondo in to one of the legal states

.

TABLE 21-26 Interrupt Clear Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63:2	—				*Reserved*
1:0	int_state	rst_l	0_{16}	RW	Writing of the register, the value of the lower two bits are used to control the state bits for the interrupt state machine associated with this interrupt. The following values my be written: 00 – Set the state machine to `idle` state. 01 – Set the state machine to `received` state. 10 – *Reserved*. If this value is used it is a programming error. The results are undefined. 11 – Set the state machine to `pending` state. When reading from this register, the actual state of the associated interrupt state machine are read. The legal values are the same as listed above.

21.4.3.3 Interrupt Retry Timer Register ($00601A00_{16}$ / 0_{16})

The Interrupt Retry Timer register is a read/write register that software uses to set up the value of the interrupt retry timer used by all four of the interrupt controllers. When the IMU receives a NACK for a given mondo, the value from this register is loaded into the interrupt controller counter, which then begins to decrement the value by 1 on every clock cycle. When the value of the counter reaches 0, the mondo that was NACKED is then retried. A value of all 0's means that the mondo will be retried on the very next cycle.

TABLE 21-27 Interrupt Retry Timer Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:25	—				*Reserved*
24:0	limit	rst_l	0_{16}	RW	Limit the retry interval in clock cycles (DMU FREQ)

21.4.3.4 Interrupt State Status Register 1 ($00601A10_{16}$ / 0_{16})

The Interrupt State Status Register 1 is a read-only register that software uses to read the state of the first 32 mondos (0–31) supported by PIU all at once. Each mondo is represented by two bits where the same state encodings from the Interrupt Clear registers are used.

TABLE 21-28 Interrupt State Status Register 1

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	state	rst_l	0_{16}	RO	State Values for mondos 0 through 31 Each state is 2 bits in the register with the MSB being the 2nd bit of mondo 31 and the LSB being the 1st bit of mondo 0

21.4.3.5 Interrupt State Status Register 2 ($00601A18_{16}$ / 0_{16})

The Interrupt State Status Register 2 is a read-only register that software uses to read the state of the second 32 Mondos (32–63) supported by PIU all at once. Each mondo is represented by 2 bits where the same state encodings from the Interrupt Clear Registers are used.

TABLE 21-29 Interrupt State Status Register 2

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	state	rst_l	0_{16}	RO	State Values for mondos 32 through 63. Each state is 2 bits in the register, with the MSB being the second bit of mondo 63 and the LSB being the first bit of mondo 32.

21.4.3.6 INTX Status Register ($0060B000_{16}$ / 0_{16})

The INTX Status register is a read-only register that software uses to obtain the status of the four emulated INTX interrupts in PIU. Each of the four bits will be updated to reflect the current hardware status, which is affected by the reception of either Assert and De-Assert PCIE messages or a write by software to any of the four INT X Clear registers.

TABLE 21-30 INTX Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:4	—				*Reserved*
3	int_a	rst_l	0_{16}	RO	INT A Status. 0 = no INTX; 1 = INTX This register will be set when an assert INT A message is received and will be cleared when a deassert INT A message is received or when cleared via the INT A Clear register by software.

TABLE 21-30 INTX Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
2	int_b	rst_l	0_{16}	RO	INT B Status. 0 = no INTX; 1 = INTX. This register will be set when an assert INT B message is received and will be cleared when a deassert INT B message is received or when cleared via the INT B Clear register by software.
1	int_c	rst_l	0_{16}	RO	INT C Status. 0 = no INTX; 1 = INTX. This register will be set when an assert INT C message is received and will be cleared when a deassert INT C message is received or when cleared via the INT C Clear register by software.
0	int_d	rst_l	0_{16}	RO	INT D Status. 0 = no INTX 1 = INTX. This register will be set when an assert INT D message is received and will be cleared when a deassert INT D message is received or when cleared via the INT D Clear register by software.

21.4.3.7 INT A Clear Register ($0060B008_{16}$ / 0_{16})

The INT A Clear register is a read/write 1 to the Clear register, which software uses to complete two tasks. First, software can obtain the current hardware status of the INT A just as in bit 3 of the INTX Status register. Second, software can clear the INT A interrupt by writing a value of 1_2 to bit 0 of this register. This functionality is required in case the PCIE link goes down; software must clear the int_a bit in order to maintain consistent INTX state on the fabric for that port. These registers should only be used for that purpose and *never* during normal operation. If they are used, results are undefined and interrupts *may be lost*.

TABLE 21-31 INT A Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:1	—				*Reserved*
0	clr	rst_l	0_{16}	RW1C	Write 0 has no effect. Write 1 will clear the int_a bit of the INTX Status register. When reading, the value of the int_a bit from the INTX Status register will be returned

21.4.3.8 INT B Clear Register ($0060B010_{16}$ / 0_{16})

The INT B Clear register is a read/write 1 to the Clear register, which software uses to complete two tasks. First, software can obtain the current hardware status of the INT B just as in bit 2 of the INTX Status register. Second, software can clear the INT A interrupt by writing a value of 1_2 to bit 0 of this register. This functionality is required in case the PCIE link goes down; software must clear the int_b bit in order to maintain consistent INTX state on the fabric for that port. These registers should only be used for that purpose and *never* during normal operation. If they are used, results are undefined and interrupts *may be lost*.

TABLE 21-32 INT B Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:1	—				*Reserved*
0	clr	rst_l	0_{16}	RW1C	Write 0 has no effect. Write 1 will clear the int_b bit of the INTX Status register. When reading, the value of the int_b bit from the INTX Status register will be returned

21.4.3.9 INT C Clear Register ($0060B018_{16}$ / 0_{16})

The INT C Clear register is a read/write 1 to the Clear register, which software uses to complete two tasks. First, software can obtain the current hardware status of the INT C just as in bit 1 of the INTX Status register. Second, software can clear the INT C interrupt by writing a value of 1_2 to bit 0 of this register. This functionality is required in case the PCIE link goes down; software must clear the int_c bit in order to maintain consistent INTX state on the fabric for that port. These registers should only be used for that purpose and *never* during normal operation. If they are used, results are undefined and interrupts *may be lost*.

TABLE 21-33 INT C Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:1	—				*Reserved*
0	clr	rst_l	0_{16}	RW1C	Write 0 has no effect. Write 1 will clear the int_c bit of the INTX Status register. When reading, the value of the int_c bit from the INTX Status register will be returned

21.4.3.10 INT D Clear Register ($0060B020_{16}$ / 0_{16})

The INT D Clear register is a Read/Write 1 to Clear register that software uses to complete two tasks. First, software can obtain the current hardware status of the INT D just as in bit 0 of the INTX Status register. Second, software can clear the INT D interrupt by writing a value of 1_2 to bit 0 of this register. This functionality is required in case the PCIE link goes down; software must clear the int_d bit in order to maintain consistent INTX state on the fabric for that port. These registers should only be used for that purpose and *never* during normal operation. If they are used results are undefined and interrupts *may be lost*.

TABLE 21-34 INT D Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:1	—				*Reserved*
0	clr	rst_l	0_{16}	RW1C	Write 0 has no effect. Write 1 will clear the int_d bit of the INTX Status register. When reading, the value of the int_d bit from the INTX Status register will be returned

21.4.3.11 Event Queue Base Address Register (00610000_{16} / 0_{16})

The Event Queue Base Address register is used as the base address for all event queue writes issued by PIU. It is the start of 512-Kbyte aligned contiguous memory region containing thirty-six 8-Kbyte pages, 1 page per event queue. This register can be set up to have EQ writes use a physical address or a virtual address that needs to be translated by the MMU. In either case, the address must be set up correctly. Refer to the addressing sections of the PRM for details.

TABLE 21-35 Event Queue Base Address Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:19	address	rst_l	0_{16}	RW	EQ base address, 512-Kbyte aligned. This address has to be a properly formatted physical or sun4v/sun4u virtual address. To generate physical bypass address, bits 63:50 set to all 1's, bits 49:39 set to all zeros. The lower bits of address{38:19} + EQ Number + Tail Pointer are used as the physical address to access the system memory. To generate sun4u mode virtual address[sun4v disable], bit 63 needs to be zero, bits 62:32 are don't care but are recommended to be set to zero. address{31:19} + EQ Number + Tail Pointer are used for IOMMU translation. To generate SUN4V mode virtual address[sun4v enable], bits 63:40 need to be zero. VA{63} and requester ID from the MSI/MSI-X are used for IO partition, address{39:19} + EQ Number + Tail Pointer are used for IOMMU translation. However, during this mode (SUN4V mode/EQ translation mode), the EQ address for PCI-Express message will not be translated. It will be automatically treated as bypass address.
18:0	—				*Reserved*

21.4.3.12 Event Queue Control Set Register (00611000_{16}–00611118_{16} / 0_{16})

The Event Queue Control Set registers are a set of 36 consecutive registers, 1 for each EQ. These are write-only registers which software uses to complete two tasks. First, software uses these registers to make an EQ active. The EQ will move from the `idle` to `active` state when a 1 is written to the `en` bit. Second, software uses this register to inject an EQ overflow error. When a 1 is written to the enoverr bit, the overflow error status bit for the EQ will be set and it will transition to `error_state`.

TABLE 21-36 Event Queue Control Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:58	—				*Reserved*

TABLE 21-36 Event Queue Control Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
57	enoverr	rst_l	0_{16}	WO	0 = no action;1 = set overr bit. A read of this register is not allowed. This bit should only be set if the EQ is currently in the `active` state. Setting this bit when the EQ is `idle` will cause undetermined results.
56:45	—				*Reserved*
44	en	rst_l	0_{16}	WO	0 = no action;1 = enable EQ. EQ will be running when state = `active`, A read of this register is not allowed. This bit should only be written when the EQ is currently `idle`. If the EQ is not in the `idle` state this operation will have no effect on the state of the EQ.
43:0	—				*Reserved*

21.4.3.13 Event Queue Control Clear Register (00611200_{16}–00611318_{16} / 0_{16})

The Event Queue Control Clear registers are a set of 36 consecutive registers, 1 for each EQ. These are write-only registers that software uses to complete three tasks. First, software uses these registers to disable an EQ. The EQ will move from the `active` to `idle` state when a 1 is written to the dis bit. Second, software uses this register to clear an EQ overflow error. When a 1 is written to the coverr bit, the overflow error status bit for the EQ will be cleared. Third, software uses this register to take an EQ from the `error` to `idle` state when a 1 is written to e2i bit.

TABLE 21-37 Event Queue Control Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:58	—				*Reserved*
57	coverr	rst_l	0_{16}	WO	0 = no action; 1 = clear overr bit, A read of this register is not allowed
56:48	—				*Reserved*
47	e2i	rst_l	0_{16}	WO	0 = no action;1 = go from `error` to `idle`. A read of this register is not allowed. This bit should only be written when the EQ is currently in the `error` state. If the EQ is not in the `error` state, this operation will have no effect on the state of the EQ.
46:45	—				*Reserved*
44	dis	rst_l	0_{16}	WO	0 = no action; 1 = disable EQ. A read of this register is not allowed. This bit should only be set if the EQ is currently in the `active` state. If the EQ is not in the `active` state this operation will have no effect on the state of the EQ.
43:0	—				*Reserved*

21.4.3.14 Event Queue State Register (00611400_{16}–00611518_{16} / 1_{16})

The Event Queue State registers are a set of 36 consecutive registers, 1 for each EQ. These are read-only registers which software uses to determine the current state of a given Event Queue. Each EQ is represented by a 3-bit, 1-hot encoding to represent each of the three possible EQ state values:
001 – `idle`, 010 – `active`, 100 – `error`.

TABLE 21-38 Event Queue State Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:3	—				*Reserved*
2:0	state	rst_l	1_{16}	RO	Event Queue state: 001 – `idle`, 010 – `active`, 100 – `error`

21.4.3.15 Event Queue Tail Register (00611600_{16}–00611718_{16} / 0_{16})

The Event Queue Tail registers are a set of 36 consecutive registers, 1 for each EQ. These registers should be considered read-only registers for software, which uses them to complete two tasks. First, software uses these registers to obtain the value of the current hardware tail pointer. Second, software uses this register to obtain the value EQ overflow error bit.

Note that even though the tail pointer field is writeable by software it should not be written except for diagnostic purposes. Writing this field in normal operation will cause unpredictable results and loss of interrupts. It is also possible for software to initialize the tail pointer to a value other than 0 but it is not necessary and should only be done with good reason

TABLE 21-39 Event Queue Tail Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:58	—				*Reserved*
57	overr	rst_l	0_{16}	RO	1 = EQ overflow occurred.
56:7	—				*Reserved*
6:0	tail	rst_l	0_{16}	RW	Value of the current hardware tail pointer. In normal operation it is read by software and written by hardware.

21.4.3.16 Event Queue Head Register (00611800_{16}–00611918_{16} / 0_{16})

The Event Queue Head registers are a set of 36 consecutive registers, 1 for each EQ. These registers are read/write registers for software, which uses them to read and write the hardware value of the EQ head pointer.

TABLE 21-40 Event Queue Head Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:7	—				*Reserved*
6:0	head	rst_l	0_{16}	RW	EQ head pointer. Initialize by software, written by software during operation.

21.4.3.17 MSI Mapping Register (00620000_{16}–$006207F8_{16}$ / 0_{16})

The MSI Mapping registers are a series of 256 consecutive registers, 1 for each MSI. These are read/write registers that software uses to set up each of the supported 256 MSI. Through these registers, software on a per-MSI basis can enable and disable the MSI's valid bit and select which event queue the MSI will use. In addition, software can also obtain status on the value of the eqwr_n bit, which is used by hardware to determine if it is OK to send an MSI to the EQ or if it is a duplicate.

TABLE 21-41 MSI Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid, A received MSI of this number will be treated as an error. 1 = valid, A received MSI of this number will be routed to the EQ specified in the 'eqnum' field
62	eqwr_n	rst_l	0_{16}	RO	0 = OK to write to; a received MSI of the number will be sent to the EQ specified. 1 = MSI already in EQ, received MSI of the number will be treated as a duplicate; software must clear this bit *before* calling the client's interrupt handler.
61:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event queue number.

21.4.3.18 MSI Clear Registers (00628000_{16}-$006287F8_{16}$ / 0_{16})

The MSI Clear registers are a series of 256 consecutive registers, 1 for each MSI. These are read/write 1 to clear registers that software uses to clear the eqwr_n bit for each of the supported 256 MSIs.

TABLE 21-42 MSI Clear Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63	—				*Reserved*
62	eqwr_n	rst_l	0_{16}	RW1C	Write 0 has no effect. Write 1 will clear the eqwr_n bit of the MSI Mapping register. When reading, the value of the eqwr_n bit from the MSI Mapping Register will be returned
61:0	—				*Reserved*

21.4.3.19 Interrupt Mondo Data 0 Register ($0062C000_{16}$ / 0_{16})

The Interrupt Mondo Data 0 register is a read/write register that software uses to set up the value it wants for the upper bits of Data Word 0 for a data-bearing mondo. This value will be inserted into the upper bits of Data Word 0 when the mondo's mode is selected to be 1_2 in the Interrupt Mapping register.

TABLE 21-43 Interrupt Mondo Data 0 Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:12	data	rst_l	0_{16}	RW	Data 0 word, bits 63:12 of mondo used for a data bearing mondos with the mode bit set to 1.
11:6	spare	rst_l	0_{16}	RW	Spare bits
5:0	—				*Reserved*

21.4.3.20 Interrupt Mondo Data 1 Register ($0062C008_{16}$ / 0_{16})

The Interrupt Mondo Data 1 register is a read/write register that software uses to set up the value it wants for Data Word 1 for a data-bearing mondo. This value will be inserted into Data word 1 when the mondo's mode is selected to be 1_2 in the Interrupt Mapping register.

TABLE 21-44 Interrupt Mondo Data 1 Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	data	rst_l	0_{16}	RW	Data 1 word of mondo used for data bearing mondos with the mode bit set to 1.

21.4.3.21 ERR COR Mapping Register (00630000_{16} / 0_{16})

The ERR COR Mapping register. is a read/write register that software uses to set up how hardware should handle the reception of an ERR COR message. Through this register, software enables and disables the ERR COR message's Valid bit and selects which of the 36 EQs the message will use. Note that setting the eqnum field to a value larger than 35 will yield unpredictable results.

TABLE 21-45 ERR COR Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid; a received message of this type will be treated as an error. 1 = valid; a received message of this type will be routed to the EQ specified in the eqnum field.
62:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event queue number.

21.4.3.22 ERR NONFATAL Mapping Register (00630008_{16} / 0_{16})

The ERR NONFATAL Mapping register is a read/write register that software uses to set up how hardware should handle the reception of an ERR NONFATAL message. Through this register, software enables and disables the ERR NONFATAL message's Valid bit and selects which of the 36 EQs the message will use. Note that setting the eqnum field to a value larger than 35 will yield unpredictable results.

TABLE 21-46 ERR NONFATAL Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid; a received message of this type will be treated as an error. 1= Valid; a received message of this type will be routed to the EQ specified in the eqnum field
62:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event queue number.

21.4.3.23 ERR FATAL Mapping Register (00630010_{16} / 0_{16})

The ERR FATAL Mapping register is a read/write register that software uses to set up how hardware should handle the reception of a ERR FATAL message. Through this register, software enables and disables the ERR FATAL message's valid bit and select which of the 36 EQs the message will use. Note setting the eqnum field to a value larger than 35 will yield unpredictable results.

TABLE 21-47 ERR FATAL Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid; a received message of this type will be treated as an error. 1 = valid; a received message of this type will be routed to the EQ specified in the eqnum field
62:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event Queue Number

21.4.3.24 PM PME Mapping Register (00630018_{16} / 0_{16})

The PM PME Mapping register is a read/write register that software uses to set up how hardware should handle the reception of a PM PME message. Through this register, software enables and disables the PM PME message's valid bit and selects which of the 36 EQs the message will use. Note setting the eqnum field to a value larger than 35 will yield unpredictable results.

TABLE 21-48 PM PME Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid; a received message of this type will be treated as an error. 1 = valid; a received message of this type will be routed to the EQ specified in the eqnum field
62:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event Queue Number

21.4.3.25 PME To ACK Mapping Register (00630020_{16} / 0_{16})

The PME To ACK Mapping register is a read/write register that software uses to set up how hardware should handle the reception of a PME To ACK message. Through this register, software enables and disables the PME To ACK message's Valid bit and selects which of the 36 EQs the message will use. Note setting the eqnum field to a value larger than 35 will yield unpredictable results.

TABLE 21-49 PME To ACK Mapping Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	v	rst_l	0_{16}	RW	0 = not valid; a received message of this type will be treated as an error. 1 = valid; a received message of this type will be routed to the EQ specified in the eqnum field
62:6	—				*Reserved*
5:0	eqnum	rst_l	0_{16}	RW	Event Queue Number

21.4.3.26 IMU Error Log Enable Register (00631000_{16} / $7FFF_{16}$)

The IMU Error Log Enable register enables the logging for all of the possible IMU errors detected by PIU. By default, all the errors are enabled to be logged.

This and following IMU error registers have been implemented in accordance with *Fire 2.0 MAS, Appendix* E and have chosen to use the Error Grouping feature. This allows more than one error to be grouped to the same error logging register. For full details, see *Fire 2.0 MAS Appendix E*.

The IMU has the following error groups: SCS Group and the RDS Group.

The errors associated with each group are as follows:

- RDS Group:
  msi_mal_err; msi_par_err; pmeack_mes_not_en; pmpme_mes_not_en; fatal_mes_not_en; nonfatal_mes_not_en; cor_mes_not_en; msi_not_en; spare bit 0; spare bit 1
- SCS Group: eq_not_en

- EQS Group: eq_over

Note that since the RDS sub-block detects multiple possible errors for an MSI transaction, there is a priority order for the MSI errors in the RDS group. Of the three detected MSI errors, the priority order will be msi_par_err, msi_mal_err, msi_not_en. The eqwr_n bit will not be affected by any of these three errors and will remain unchanged and hold the value from the last nonerrored MSI.

TABLE 21-50 IMU Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:15	—				*Reserved*
14:10	spare_log_en	por_l	$1F_{16}$	RW	Spare error, error log enable bits
9	eq_over_log_en	por_l	1_{16}	RW	EQ overflow error, error log enable bit
8	eq_not_en_log_en	por_l	1_{16}	RW	EQ not enabled, error log enable bit
7	msi_mal_err_log_en	por_l	1_{16}	RW	Malformed MSI, error log enable bit
6	msi_par_err_log_en	por_l	1_{16}	RW	MSI data parity error, error log enable bit
5	pmeack_mes_not_en_log_en	por_l	1_{16}	RW	PME to ACK message not enabled, error log enable bit
4	pmpme_mes_not_en_log_en	por_l	1_{16}	RW	PM PME message not enabled, error log enable bit
3	fatal_mes_not_en_log_en	por_l	1_{16}	RW	Fatal message not enabled, error log enable bit
2	nonfatal_mes_not_en_log_en	por_l	1_{16}	RW	Nonfatal message not enabled, error log enable bit
1	cor_mes_not_en_log_en	por_l	1_{16}	RW	Correctable message not enabled, error log enable bit
0	msi_not_en_log_en	por_l	1_{16}	RW	MSI not enabled, error log enable bit

21.4.3.27 IMU Interrupt Enable Register (00631008_{16} / 0_{16})

The IMU Error Log Enable register enables the interrupt for all of the possible IMU errors detected by PIU. By default, all the errors are not enabled to generate an interrupt.

TABLE 21-51 IMU Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:47	—				*Reserved*
46:42	spare_s_int_en	rst_l	0_{16}	RW	Spare error, secondary interrupt enable bits
41	eq_over_s_int_en	rst_l	0_{16}	RW	EQ Overflow error, secondary interrupt enable bit
40	eq_not_en_s_int_en	rst_l	0_{16}	RW	EQ not enabled, secondary interrupt enable bit
39	msi_mal_err_s_int_en	rst_l	0_{16}	RW	Malformed MSI, secondary interrupt enable bit
38	msi_par_err_s_int_en	rst_l	0_{16}	RW	MSI data parity error, secondary interrupt enable bit
37	pmeack_mes_not_en_s_int_en	rst_l	0_{16}	RW	PME to ACK message not enabled, secondary interrupt enable bit

TABLE 21-51 IMU Interrupt Enable Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
36	pmpme_mes_not_en_s_int_en	rst_l	0_{16}	RW	PM PME message not enabled, secondary interrupt enable bit
35	fatal_mes_not_en_s_int_en	rst_l	0_{16}	RW	Fatal message not enabled, secondary interrupt enable bit
34	nonfatal_mes_not_en_s_int_en	rst_l	0_{16}	RW	Nonfatal message not enabled, secondary interrupt enable bit
33	cor_mes_not_en_s_int_en	rst_l	0_{16}	RW	Correctable Message not enabled, secondary interrupt enable bit
32	msi_not_en_s_int_en	rst_l	0_{16}	RW	MSI not enabled, secondary interrupt enable bit
31:15	—				*Reserved*
14:10	spare_p_int_en	rst_l	0_{16}	RW	Spare error, primary interrupt enable bits
9	eq_over_p_int_en	rst_l	0_{16}	RW	EQ overflow error, primary interrupt enable bit
8	eq_not_en_p_int_en	rst_l	0_{16}	RW	eq not enabled, primary interrupt enable bit
7	msi_mal_err_p_int_en	rst_l	0_{16}	RW	Malformed MSI, primary interrupt enable bit
6	msi_par_err_p_int_en	rst_l	0_{16}	RW	MSI data parity error, primary interrupt enable bit
5	pmeack_mes_not_en_p_int_en	rst_l	0_{16}	RW	PME to ACK message not enabled, primary interrupt enable bit
4	pmpme_mes_not_en_p_int_en	rst_l	0_{16}	RW	PM PME message not enabled, primary interrupt enable bit
3	fatal_mes_not_en_p_int_en	rst_l	0_{16}	RW	Fatal message not enabled, primary interrupt enable bit
2	nonfatal_mes_not_en_p_int_en	rst_l	0_{16}	RW	Nonfatal message not enabled, primary interrupt enable bit
1	cor_mes_not_en_p_int_en	rst_l	0_{16}	RW	Correctable message not enabled, primary interrupt enable bit
0	msi_not_en_p_int_en	rst_l	0_{16}	RW	MSI not enabled, primary interrupt enable bit

21.4.3.28 IMU Interrupt Status Register (00631010_{16} / 0_{16})

The IMU Interrupt Status Register is a read-only register that software reads to obtain the source of interrupt for the IMU Block. (mondo number 62). This register will contain a 1 in all bits for which an interrupt was generated.

TABLE 21-52 IMU Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:47	—				*Reserved*
46:42	spare_s	rst_l	0_{16}	RO	Spare error, secondary error status. Bit 1 = error received
41	eq_over_s	rst_l	0_{16}	RO	EQ overflow secondary error status. Bit 1 = error received

TABLE 21-52 IMU Interrupt Status Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
40	eq_not_en_s	rst_l	0_{16}	RO	EQ not enabled secondary error status. Bit 1 = error received
39	msi_mal_err_s	rst_l	0_{16}	RO	Malformed MSI secondary error status. Bit 1 = error received
38	msi_par_err_s	rst_l	0_{16}	RO	MSI Data Parity secondary error status. Bit 1 = error received
37	pmeack_mes_not_en_s	rst_l	0_{16}	RO	PME to ACK message not enabled secondary error status. Bit 1 = error received
36	pmpme_mes_not_en_s	rst_l	0_{16}	RO	pm PME message not enabled secondary error status. Bit 1 = error received
35	fatal_mes_not_en_s	rst_l	0_{16}	RO	Fatal message not enabled secondary error status. Bit 1 = error received
34	nonfatal_mes_not_en_s	rst_l	0_{16}	RO	Nonfatal message not enabled secondary error status. Bit 1 = error received
33	cor_mes_not_en_s	rst_l	0_{16}	RO	Correctable message not enabled secondary error status. Bit 1 = error received
32	msi_not_en_s	rst_l	0_{16}	RO	MSI not enabled secondary error status. Bit 1 = error received
31:15	—				*Reserved*
14:10	spare_p	rst_l	0_{16}	RO	Spare error, primary error status. Bit 1 = error received
9	eq_over_p	rst_l	0_{16}	RO	EQ overflow primary error status. Bit 1 = error received
8	eq_not_en_p	rst_l	0_{16}	RO	EQ not enabled primary error status. Bit 1 = error received
7	msi_mal_err_p	rst_l	0_{16}	RO	Malformed MSI primary error status. Bit 1 = error received
6	msi_par_err_p	rst_l	0_{16}	RO	MSI Data Parity Primary Error Status. Bit 1 = error received
5	pmeack_mes_not_en_p	rst_l	0_{16}	RO	PME to ACK message not enabled primary error status. Bit 1 = error received
4	pmpme_mes_not_en_p	rst_l	0_{16}	RO	pm PME message not enabled primary error status. Bit 1 = error received
3	fatal_mes_not_en_p	rst_l	0_{16}	RO	Fatal message not enabled primary error status. Bit 1 = error received
2	nonfatal_mes_not_en_p	rst_l	0_{16}	RO	Nonfatal message not enabled primary error status. Bit 1 = error received
1	cor_mes_not_en_p	rst_l	0_{16}	RO	Correctable message not enabled primary error status. Bit 1 = error received
0	msi_not_en_p	rst_l	0_{16}	RO	MSI not enabled primary error status. Bit 1 = error received

21.4.3.29 IMU Error Status Clear Register (00631018_{16} / 0_{16})

The IMU Error Status Clear register is a read, write 1 to clear register that serves two purposes for software. First, software can read this register to determine which IMU errors have been logged by PIU but have not necessarily caused an interrupt (mondo 62). This register will contain a 1 in all bits where errors were logged. Second, software uses this register to clear the IMU errors detected by PIU. It writes a 1 to the bit(s) that it wants to clear.

TABLE 21-53 IMU Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:47	—				*Reserved*
46:42	spare_s	por_l	0_{16}	RW1C	Spare error, secondary error status. Bit 1 = error received
41	eq_over_s	por_l	0_{16}	RW1C	EQ overflow secondary error status. Bit 1 = error received
40	eq_not_en_s	por_l	0_{16}	RW1C	EQ not enabled secondary error status. Bit 1 = error received
39	msi_mal_err_s	por_l	0_{16}	RW1C	Malformed MSI secondary error status. Bit 1 = error received
38	msi_par_err_s	por_l	0_{16}	RW1C	MSI data parity secondary error status. Bit 1 = error received
37	pmeack_mes_not_en_s	por_l	0_{16}	RW1C	PME to ACK message not enabled primary error status. Bit 1 = error received
36	pmpme_mes_not_en_s	por_l	0_{16}	RW1C	PM PME message not enabled primary error status. Bit 1 = error received
35	fatal_mes_not_en_s	por_l	0_{16}	RW1C	Fatal message not enabled primary error status. Bit 1 = error received
34	nonfatal_mes_not_en_s	por_l	0_{16}	RW1C	Nonfatal message not enabled primary error status. Bit 1 = error received
33	cor_mes_not_en_s	por_l	0_{16}	RW1C	Correctable message not enabled primary error status. Bit 1 = error received
32	msi_not_en_s	por_l	0_{16}	RW1C	MSI not enabled secondary error status. Bit 1 = error received
31:15	—				*Reserved*
14:10	spare_p	por_l	0_{16}	RW1C	Spare error, primary error status. Bit 1 = error received
9	eq_over_p	por_l	0_{16}	RW1C	EQ overflow primary error status. Bit 1 = error received
8	eq_not_en_p	por_l	0_{16}	RW1C	EQ not enabled primary error status. Bit 1 = error received
7	msi_mal_err_p	por_l	0_{16}	RW1C	Malformed MSI primary error status. Bit 1 = error received
6	msi_par_err_p	por_l	0_{16}	RW1C	MSI data parity primary error status. Bit 1 = error received

TABLE 21-53 IMU Error Status Clear Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
5	pmeack_mes_not_en_p	por_l	0_{16}	RW1C	PME to ACK message not enabled primary error status. Bit 1 = error received
4	pmpme_mes_not_en_p	por_l	0_{16}	RW1C	pm PME message not enabled primary error status. Bit 1 = error received
3	fatal_mes_not_en_p	por_l	0_{16}	RW1C	Fatal message not enabled primary error status. Bit 1 = error received
2	nonfatal_mes_not_en_p	por_l	0_{16}	RW1C	Nonfatal message not enabled primary error status. Bit 1 = error received
1	cor_mes_not_en_p	por_l	0_{16}	RW1C	Correctable message not enabled primary error status. Bit 1 = error received
0	msi_not_en_p	por_l	0_{16}	RW1C	MSI not enabled primary error status. Bit 1 = error received

21.4.3.30 IMU Error Status Set Register (00631020_{16} / 0_{16})

The IMU Error Status Set register is a read, write 1 to set register that serves two purposes for software. First, software can read this register to determine which IMU errors have been logged by PIU but have not necessarily caused an interrupt (mondo 62). This register will contain a 1 in all bits where errors were logged. Second, software uses this register to set the IMU errors detected by PIU. It writes a 1 to the bit(s) it wants to set. *This should only be used for diag purposes and error testing.*

TABLE 21-54 IMU Error Status Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:47	—				*Reserved*
46:42	spare_s	por_l	0_{16}	RW1S	Spare error, secondary error status. Bit 1 = error received
41	eq_over_s	por_l	0_{16}	RW1S	EQ overflow secondary error status. Bit 1 = error received
40	eq_not_en_s	por_l	0_{16}	RW1S	EQ not enabled secondary error status. Bit 1 = error received
39	msi_mal_err_s	por_l	0_{16}	RW1S	Malformed MSI secondary error status. Bit 1 = error received
38	msi_par_err_s	por_l	0_{16}	RW1S	MSI data parity secondary error status. Bit 1 = error received
37	pmeack_mes_not_en_s	por_l	0_{16}	RW1S	PME to ACK message not enabled primary error status. Bit 1 = error received
36	pmpme_mes_not_en_s	por_l	0_{16}	RW1S	PM PME message not enabled primary error status. Bit 1 = error received

TABLE 21-54 IMU Error Status Set Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
35	fatal_mes_not_en_s	por_l	0_{16}	RW1S	Fatal message not enabled primary error status. Bit 1 = error received
34	nonfatal_mes_not_en_s	por_l	0_{16}	RW1S	Nonfatal message not enabled primary error status Bit 1 = error received
33	cor_mes_not_en_s	por_l	0_{16}	RW1S	Correctable message not enabled primary error status. Bit 1 = error received
32	msi_not_en_s	por_l	0_{16}	RW1S	MSI not enabled secondary error status. Bit 1 = error received
31:15	—				*Reserved*
14:10	spare_p	por_l	0_{16}	RW1S	Spare error, primary error status. Bit 1 = error received
9	eq_over_p	por_l	0_{16}	RW1S	EQ overflow primary error status. Bit 1 = error received
8	eq_not_en_p	por_l	0_{16}	RW1S	EQ not enabled primary error status. Bit 1 = error received
7	msi_mal_err_p	por_l	0_{16}	RW1S	Malformed MSI primary error status. Bit 1 = error received
6	msi_par_err_p	por_l	0_{16}	RW1S	MSI data parity primary error status. Bit 1 = error received
5	pmeack_mes_not_en_p	por_l	0_{16}	RW1S	PME to ACK message not enabled primary error status. Bit 1 = error received
4	pmpme_mes_not_en_p	por_l	0_{16}	RW1S	pm PME message not enabled primary error status. Bit 1 = error received
3	fatal_mes_not_en_p	por_l	0_{16}	RW1S	Fatal message not enabled primary error status. Bit 1 = error received
2	nonfatal_mes_not_en_p	por_l	0_{16}	RW1S	Nonfatal message not enabled primary error status. Bit 1 = error received
1	cor_mes_not_en_p	por_l	0_{16}	RW1S	Correctable message not enabled primary error status. Bit 1 = error received
0	msi_not_en_p	por_l	0_{16}	RW1S	MSI not enabled primary error status. Bit 1 = error received

21.4.3.31 IMU RDS Error Log Register (00631028_{16} / 0_{16})

The IMU RDS Error Log register captures information when the first primary error is detected by IMU in the RDS group. Note that until an error in the IMU RDS Group is logged, this register does not contain valid data and will update on every cycle.

The RDS error logging register will *not* change for the MSI-X and *will not* have all four bytes of the MSI-X data stored in it. It will only contain the first two bytes of MSI data.

TABLE 21-55 IMU RDS Error Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:58	type	por_l	0_2	RW	The lowest 6 bits of the type of the errored transaction as seen by the IMU in the RDS pipe stage: 111000 – 64-bit addressed MSI 011000 – 32-bit addressed MSI 110xxx - Message where xxx complies with the routing code in PCIE spec
57:48	length	por_l	0_2	RW	The length of the errored transaction.
47:32	req_id	por_l	0_2	RW	The requester ID of the errored transaction.
31:24	tlp_tag	por_l	0_2	RW	The TLP tag of the errored transaction.
23:16	be_mess_code	por_l	0_2	RW	The message code of the error is associated with a message. The first and last byte enabled if the error is associated with an MSI.
15:0	msi_data	por_l	0_2	RW	The MSI data if the error is associated with an MSI.

21.4.3.32 IMU SCS Error Log Register (00631030_{16} / 0_{16})

The IMU SCS Error Log register captures information when the first primary error is detected by IMU in the SCS group. Note that until an error in the IMU SCS Group is logged, this register does not contain valid data and will update on every cycle.

TABLE 21-56 IMU SCS Error Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:58	type	por_l	0_2	RW	The lowest 6 bits of the type of the errored transaction as seen by the IMU in the SCS pipe stage: 011000 – MSI 010000 – Message
57:48	length	por_l	0_2	RW	The length of the errored transaction. At this stage of the pipeline in the IMU, the length has already been changed to this. Always will be 10_{16}.
47:32	req_id	por_l	0_2	RW	The requester ID of the errored transaction.
31:24	tlp_tag	por_l	0_2	RW	The TLP tag of the errored transaction.
23:16	be_mess_code	por_l	0_2	RW	The message code of the error is associated with a message. The first and last byte enabled if the error is associated with an MSI.
15:6	—				*Reserved*
5:0	eq_num	por_l	0_2	RW	EQ number that the transaction tried to go to but was not enabled.

21.4.3.33 IMU EQS Error Log Register (00631038_{16} / 0_{16})

The IMU EQS Error Log captures information when the first primary error is detected by IMU in the EQS group. Note that until an error in the IMU EQS Group is logged, this register does not contain valid data and will update on every cycle.

TABLE 21-57 IMU EQS Error Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:6	—				*Reserved*
5:0	eq_num	por_l	0_2	RW	EQ number that the transaction tried to go to but was not enabled.

21.4.3.34 DMU Core and Block Interrupt Enable Register (00631800_{16} / 0_{16})

The DMU Core and Block Interrupt Enable register enables interrupts at a block and core level. This register allows software to mask interrupts it does not want to see with a core- and block-level granularity. Each of the supported DMU errors falls under one of the block enables, either the MMU or IMU. If a particular block is not enabled, no errors from that block will generate an interrupt no matter what the individual error interrupt enable is set to.

Note that if a condition occurs that should have caused an interrupt via mondo 62 and a particular enable is disabled at that time (not causing the mondo to be sent), if that enable is then enabled before clearing the stored offending condition, an interrupt via mondo 62 will then be sent.

TABLE 21-58 DMU Core and Block Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	dmu	rst_l	0_{16}	RW	The enable bit to enable all operations from the DMC that would cause an interrupt via mondo 62. 1 = Core-level interrupt enabled; 0 = Core-level interrupt disabled
62	debug_trig_en	rst_l	0_{16}	RW	Debugging trigger enable for PCI_EX errors. This is the **or**ed signal of Mondo 63 and Mondo 62.
61:2	—				*Reserved*
1	mmu	rst_l	0_{16}	RW	The enable bit to enable all operations from the MMU that would cause an interrupt via mondo 62. 1 = Block-level interrupt enabled; 0 = Block-level interrupt disabled.
0	imu	rst_l	0_{16}	RW	The enable bit to enable all operations from the IMU that would cause an interrupt via mondo 62. 1 = Block-level interrupt enabled; 0 = Block-level interrupt is disabled.

21.4.3.35 DMU Core and Block Error Status Register (00631808_{16} / 0_{16})

The DMU Core and Block Error Status register is a read-only register that software uses to determine which block in the DMC core cause an interrupt when it received a mondo 62 from the DMC core. A value of 1 means that particular block caused an interrupt to be generated.

TABLE 21-59 DMC Core and Block Error Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:2	—				*Reserved*
1	mmu	rst_l	0_{16}	RO	The MMU has an interrupt that needs to be processed via mondo 62.
0	imu	rst_l	0_{16}	RO	The IMU has an interrupt that needs to be processed via mondo 62.

21.4.3.36 IMU Performance Counter Select Register (00632000_{16} / 0_{16})

The IMU Performance Counter Select register provides the select inputs for both of the performance counters in the IMU block. These selects are used to select which of the seven possible events each counter should count. Both counters default to disabled. When selecting an event to count, the counting will begin as soon as this register is updated. It should be noted that by changing the value of the select, the counter register *does not* reset to 0. It will continue to count at its current value If the counter needs to start at 0, it should be cleared before the new value of the select is written.

TABLE 21-60 IMU Performance Counter Select Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:8	sel1	rst_l	0_{16}	RW	The values for Select 1 are the same as for Select 0.
7:0	sel0	rst_l	0_{16}	RW	Select for counter 0: 00 – None 01 – Clock cycles 02 – Total number of mondos issued 03 – Total number of MSIs issued 04 – Total number of mondo NACKs 05 – Total Number of EQ Writes 06 – Total number of EQ mondos

21.4.3.37 IMU Performance Counter Zero Register (00632008_{16} / 0_{16})

The IMU Performance Counter Zero register reads and writes the value of counter zero for the IMU block. This counter will increment once for every clock cycle that the selected event chosen by the IMU Performance Counter Select register occurs. This value can be set and cleared by software.

TABLE 21-61 IMU Performance Counter Zero Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	cnt	rst_l	0_{16}	RW	Counter.

21.4.3.38 IMU Performance Counter One Register (00632010_{16} / 0_{16})

The IMU Performance Counter One register reads and writes the value of counter one for the IMU block. This counter will increment once for every clock cycle that the selected event chosen by the IMU Performance Counter Select register occurs. This value can be set and cleared by software.

TABLE 21-62 IMU Performance Counter One Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	cnt	rst_l	0_{16}	RW	Counter.

21.4.3.39 MSI/MSI-X 32-bit Address Register (00634000_{16} / 0_{16})

The MSI/MSI-X 32-bit Address register is used by hardware to compare against the address of incoming PCIE 32-bit addressed memory write commands. If the address matches bits 31:16, the PCIE command is considered to be an MSI/MSI-X and is thus passed to the IMU to be processed. This value of this register is programmed by software.

TABLE 21-63 MSI/MSI-X 32-bit Address Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:16	addr	rst_l	0_{16}	RW	Upper 16 bits of the 32-bit address, which translates into an MSI interrupt.
15:0	—				*Reserved*

21.4.3.40 MSI/MSI-X 64-bit Address Register (00634008_{16} / 0_{16})

The MSI/MSI-X 64-bit Address register is used by hardware to compare against the address of incoming PCIE 64-bit addressed memory write commands. If the address matches bits 63:16, the PCIE command is considered to be an MSI/MSI-X and is thus passed to the IMU to be processed. If IOMMU sun4v mode and bit 63 of Event Q base address register is set to 0, bit 62:16 is used for comparison. This value of this register is programmed by software.

TABLE 21-64 MSI/MSI-X 64-bit Address Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	addr	rst_l	0_{16}	RW	Upper 48 bits of the 64-bit address, which translates into an MSI interrupt.
15:0	—				*Reserved*

21.4.3.41 Mem 64 PCIE Offset Register (00634018_{16} / 0_{16})

The Mem 64 PCIE Offset register serves two purposes. First, it is written by software to select the value of the upper bits of the address for a 64-bit addressed PIO initiated by PIU, which hardware will place in the PCIE command header before the packet is sent. Second, the register contains 24 bits of spare control and status bits that could be used in a metal spin. This register currently has no functionality.

TABLE 21-65 Mem 64 PCIE Offset Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:24	addr	rst_l	0_{16}	RW	The upper 40 bits of the 64 bit PIO address. The 64-bit PIO address is formed from {offset{63:36}, (offset{35:24} | (PA{35:24} & ~mask{35:24})}.
23	spare_control_load_7	rst_l	0_{16}	RW	Spare Loadable Control register 7.
22	spare_control_load_6	rst_l	0_{16}	RW	Spare Loadable Control register 6.
21	spare_control_load_5	rst_l	0_{16}	RW	Spare Loadable Control register 5.
20	spare_control_load_4	rst_l	0_{16}	RW	Spare Loadable Control register 4.
19	spare_control_load_3	rst_l	0_{16}	RW	Spare Loadable Control register 3.
18	spare_control_load_2	rst_l	0_{16}	RW	Spare Loadable Control register 2.
17	spare_control_load_1	rst_l	0_{16}	RW	Spare Loadable Control register 1.
16	spare_control_load_0	rst_l	0_{16}	RW	Spare Loadable Control register 0.
15:8	spare_status	rst_l	0_{16}	RW	Spare Status registers: RW from SW; writable from HW.
7:2	spare_control	rst_l	0_{16}	RW	Spare Control Register: RW from SW; RO from HW.
1	pio_control_1	rst_l	0_{16}	RW	Enables the ILU logic which detects the EHB write/read pointer and stalls the CRM PIO dispatch logic to guarantee that the PIO portion of the EHB is never full.

TABLE 21-65 Mem 64 PCIE Offset Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
0	pio_control_0	rst_l	0_{16}	RW	Enables the CRM state machine that controls the PIO read counter and PIO post-write counter to guarantee that the PIO portion of the EHB is never full if there is at least one outstanding PIO read.

21.4.3.42 MMU Control and Status Register (00640000_{16} / 0_{16})

The MMU Control and Status register enables the functions of the MMU and indicates MMU status.

TABLE 21-66 MMU Control and Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:52	—				*Reserved*
51:48	spares	rst_l	0_{16}	RO	Spare status bits.
47:46	—				*Reserved*
45	paq	rst_l	0_{16}	RO	Physical Address Queue not empty.
44	vaq	rst_l	0_{16}	RO	Virtual Address Queue not empty.
43	tpl	rst_l	0_{16}	RO	Translation pipeline not empty.
42	tip	rst_l	0_{16}	RO	Tablewalk in progress.
41:40	tcm	rst_l	0_{16}	RO	Tablewalk Cache Mode. Cache mode when tablewalk was initiated.
39:20	—				*Reserved*
19:16	sparec	rst_l	0_{16}	RW	Spare control bits.
15:13	—				*Reserved*
12	pd	rst_l	0_{16}	RW	Process Disable. When set, addresses are not processed. Otherwise, addresses are processed. During normal operation, this bit will be clear.
11	—				*Reserved*
10	se	rst_l	0_{16}	RW	TSB Cache Snoop Enable. When set, snoop addresses which hit will invalidate cache entries and or the TLB. During the normal operation, this bit will be set.
9:8	cm	rst_l	0_{16}	RW	Cache Mode. 00 – Cache disabled, TLB used, tablewalks enabled. 01 = Cache enabled but locked, tablewalks disabled. 10 = Cache enabled but locked, TLB used, tablewalks enabled. 11 = Cache enabled, TLB used for update, tablewalks enabled. During normal operation, its value is 11.
7:4	—				*Reserved*
3	busid_sel	rst_l	0_{16}	RW	Busid Select. 0 = use busid{6:1} for DEV2IOTSB index; 1 = use busid{5:0} for DEV2IOTSB index.

TABLE 21-66 MMU Control and Status Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
2	sun4v_en	rst_l	0_{16}	RW	sun4v IOMMU enable bit. 0 = sun4v disabled (default is SUN4U mode); 1 = sun4v enabled
1	be	rst_l	0_{16}	RW	Bypass Enable. When set, bypass mode is enabled. Otherwise, an attempt to use bypass mode causes an error. During normal operation, this bit will be set.
0	te	rst_l	0_{16}	RW	Translation Enable. When set, translation mode is enabled. Otherwise, an attempt to use translation mode causes an error. During normal operation, this bit will be set.

21.4.3.43 MMU TSB Control Register (00640008_{16} / 0_{16})

The TSB Control register contains the pointer to the first entry of the TSB table. Together with part of the virtual address it uniquely identifies the address where hardware should fetch the TTE from the TSB table. The TSB table has to be aligned on an 8-Kbyte boundary. The lower-order 13 address bits are assumed to be 0_{16} during TSB table lookup. Tables larger than 8-Kbytes are only constrained to be on 8-Kbyte boundaries rather than having to be size aligned. The TSB size and TSB page size are also configured in this register.

.

TABLE 21-67 MMU TSB Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:39	—				*Reserved*
38:13	tb	rst_l	0_{16}	RW	Base address. Most-significant 26 bits of the TSB physical address.
12:9	—				*Reserved*
8	ps	rst_l	0_{16}	RW	Page size. 0 = 8 KB; 1 = 64 KB.
7:4	—				*Reserved*
3:0	ts	rst_l	0_{16}	RW	TSB table size measured in the number of 8-byte entries. 0 – 1K 1 – 2K 2 – 4K 3 – 8K 4 – 16K 5 – 32K 6 – 64K 7 – 128K 8 – 256K 9 – 512K A-F – *Reserved*

21.4.3.44 MMU TTE Cache Flush Address Register (8000002030_{16} / 0_{16})

The MMU TTE Cache Flush Address register is implemented in NCU, not inside the IOMMU. The entire physical address of this register is specified above. The register is documented here for IOMMU documentation completeness.

The MMU TTE Cache Flush Address register flushes entries from the TTE cache if the address matches the physical addresses of the TSB entry corresponding to the cacheline to be flushed. All the TTEs are flushed if this address (flsh_addr{38:6}) matches their physical tag. Note that the entire 64-bit flsh_addr{63:0} is passed into DMU through the existent NCU-DMU interface, but only flsh_addr{38:6} is used inside the IOMMU for TTE Cache flush.

TABLE 21-68 MMU TTE Cache Flush Address Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	flsh_addr	rst_l	0_{16}	RW	Flush address. Only flsh_addr{38:6} is used in IOMMU.

21.4.3.45 MMU TTE Cache Invalidate Register (00640108_{16} / 0_{16})

The MMU TTE Cache Invalidate register flushes the indicated entry from the TTE cache. Bit 0 flushes entry 0, etc. Writing all 1's will flush the entire cache. Any write will also flush the TLB. Writes to this register are not dependent upon the state of the se or cm bits in the MMU Control and Status register.

TABLE 21-69 MMU TTE Cache Invalidate Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	flsh_tte	rst_l	0_{16}	W	Flush TTE cache entry.

21.4.3.46 MMU Error Log Enable Register (00641000_{16} / $1FFFFF_{16}$)

The MMU Error Log Enable register enables the logging of errors.

TABLE 21-70 MMU Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:21	—				*Reserved*
20:0	en	por_l	$1FFFFF_{16}$	RW	Error log enables.

21.4.3.47 MMU Interrupt Enable Register (00641008_{16} / 0_{16})

The MMU Interrupt Enable register enables errors to generate interrupts.

TABLE 21-71 MMU Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52:32	en_s	rst_l	0_{16}	RW	Secondary interrupt enables.
31:21	—				*Reserved*
20:0	en_p	rst_l	0_{16}	RW	Primary interrupt enables.

21.4.3.48 MMU Interrupt Status Register (00641010_{16} / 0_{16})

The MMU Interrupt Status register shows the errors that are currently logged and are enabled to generate interrupts.

TABLE 21-72 MMU Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52:32	err_s	rst_l	0_{16}	RO	Secondary interrupt status.
31:21	—				*Reserved*
20:0	err_p	rst_l	0_{16}	RO	Primary interrupt status.

21.4.3.49 MMU Error Status Clear Register (00641018_{16} / 0_{16})

The MMU Error Status Clear register shows the errors that are currently logged. Primary errors include data logged with those errors. Secondary errors do not have any data logged with them. All errors belong to the same error group, which means that any primary error logged will cause any additional error to be logged as secondary.

TABLE 21-73 MMU Error Status Clear Register *(1 of 3)*

Bits	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52	sun4v_key_err_s	por_l	0_{16}	RW1C	Device key mismatch secondary error (sun4u/sun4v).
51	sun4v_va_adj_uf_s	por_l	0_{16}	RW1C	sun4v virtual address adjustment secondary error.
50	sun4v_va_oor_s	por_l	0_{16}	RW1C	sun4v virtual address out of range secondary error, VA{62:40} ≠ 0.
49	iotsbdesc_dpe_s	por_l	0_{16}	RW1C	IOTSBDESC data parity secondary error.
48	iotsbdesc_inv_s	por_l	0_{16}	RW1C	IOTSBDESC invalid secondary error.
47	tbw_dpe_s	por_l	0_{16}	RW1C	Tablewalk data parity secondary error.
46	tbw_err_s	por_l	0_{16}	RW1C	Tablewalk secondary error.

TABLE 21-73 MMU Error Status Clear Register *(2 of 3)*

Bits	Field	Reset Name	Reset Value	R/W	Description
45	tbw_ude_s	por_l	0_{16}	RW1C	Tablewalk unexpected data return secondary error.
44	tbw_dme_s	por_l	0_{16}	RW1C	Tablewalk disabled miss secondary error.
43	spare3_s	por_l	0_{16}	RW1C	Spare secondary error.
42	spare2_s	por_l	0_{16}	RW1C	Spare secondary error.
41	ttc_cae_s	por_l	0_{16}	RW1C	TTE cache CSR access secondary error.
40	ttc_dpe_s	por_l	0_{16}	RW1C	TTE cache data parity secondary error.
39	tte_prt_s	por_l	0_{16}	RW1C	TTE write bit not set on write secondary error.
38	tte_inv_s	por_l	0_{16}	RW1C	TTE valid bit not set secondary error.
37	trn_oor_s	por_l	0_{16}	RW1C	Translation access out-of-range secondary error.
36	trn_err_s	por_l	0_{16}	RW1C	Translation access when te = 0 secondary error.
35	spare1_s	por_l	0_{16}	RW1C	Spare secondary error.
34	sun4v_inv_pg_sz_s	por_l	0_{16}	RW1C	sun4v Invalid Page Size secondary error. Illegal value is placed in the IOTSBDESC table.
33	byp_oor_s	por_l	0_{16}	RW1C	Bypass access out-of-range secondary error.
32	byp_err_s	por_l	0_{16}	RW1C	Bypass access when be = 0 secondary error.
31:21	—				*Reserved*
20	sun4v_key_err_p	por_l	0_{16}	RW1C	device key mismatch primary error (sun4u/sun4v).
19	sun4v_va_adj_uf_p	por_l	0_{16}	RW1C	sun4v virtual address adjustment underflow primary error.
18	sun4v_va_oor_p	por_l	0_{16}	RW1C	sun4v virtual address out of range primary error, VA{62:40} ≠ 0.
17	iotsbdesc_dpe_p	por_l	0_{16}	RW1C	IOTSBDESC data parity primary error.
16	iotsbdesc_inv_p	por_l	0_{16}	RW1C	IOTSBDESC invalid primary error.
15	tbw_dpe_p	por_l	0_{16}	RW1C	Tablewalk data parity primary error.
14	tbw_err_p	por_l	0_{16}	RW1C	Tablewalk primary error.
13	tbw_ude_p	por_l	0_{16}	RW1C	Tablewalk unexpected data return primary error.
12	tbw_dme_p	por_l	0_{16}	RW1C	Tablewalk disabled miss primary error.
11	spare3_p	por_l	0_{16}	RW1C	Spare primary error.
10	spare2_p	por_l	0_{16}	RW1C	Spare primary error.
9	ttc_cae_p	por_l	0_{16}	RW1C	TTE cache CSR access primary error.
8	ttc_dpe_p	por_l	0_{16}	RW1C	TTE cache data parity primary error.
7	tte_prt_p	por_l	0_{16}	RW1C	TTE write bit not set on write primary error.
6	tte_inv_p	por_l	0_{16}	RW1C	TTE valid bit not set primary error.
5	trn_oor_p	por_l	0_{16}	RW1C	Translation access out-of-range primary error.
4	trn_err_p	por_l	0_{16}	RW1C	Translation access when te = 0 primary error.
3	spare1_p	por_l	0_{16}	RW1C	Spare primary error.
2	sun4v_inv_pg_sz_p	por_l	0_{16}	RW1C	sun4v Invalid Page Size primary error. Illegal value is placed in the IOTSBDESC table.

TABLE 21-73 MMU Error Status Clear Register *(3 of 3)*

Bits	Field	Reset Name	Reset Value	R/W	Description
1	byp_oor_p	por_l	0_{16}	RW1C	Bypass access out-of-range primary error.
0	byp_err_p	por_l	0_{16}	RW1C	Bypass access when be = 0 primary error.

21.4.3.50 MMU Error Status Set Register (00641020_{16}/ 0_{16})

The MMU Error Status Set register allows for error bits to be set to simulate actual error occurrence.

TABLE 21-74 MMU Error Status Set Register

Bits	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52	sun4v_key_err_s	por_l	0_{16}	RW1S	Device key mismatch secondary error (sun4u/sun4v).
51	sun4v_va_adj_uf_s	por_l	0_{16}	RW1S	sun4v virtual address adjustment underflow secondary error.
50	sun4v_va_oor_s	por_l	0_{16}	RW1S	sun4v virtual address out of range secondary error, VA{62:40} ≠ 0.
49	iotsbdesc_dpe_s	por_l	0_{16}	RW1S	IOTSBDESC data parity secondary error.
48	iotsbdesc_inv_s	por_l	0_{16}	RW1S	IOTSBDESC invalid secondary error.
47	tbw_dpe_s	por_l	0_{16}	RW1S	Tablewalk data parity secondary error.
46	tbw_err_s	por_l	0_{16}	RW1S	Tablewalk secondary error.
45	tbw_ude_s	por_l	0_{16}	RW1S	Tablewalk unexpected data return secondary error.
44	tbw_dme_s	por_l	0_{16}	RW1S	Tablewalk disabled miss secondary error.
43	spare3_s	por_l	0_{16}	RW1S	Spare secondary error.
42	spare2_s	por_l	0_{16}	RW1S	Spare secondary error.
41	ttc_cae_s	por_l	0_{16}	RW1S	TTE cache CSR access secondary error.
40	ttc_dpe_s	por_l	0_{16}	RW1S	TTE cache data parity secondary error.
39	tte_prt_s	por_l	0_{16}	RW1S	TTE write bit not set on write secondary error.
38	tte_inv_s	por_l	0_{16}	RW1S	TTE valid bit not set secondary error.
37	trn_oor_s	por_l	0_{16}	RW1S	Translation access out-of-range secondary error.
36	trn_err_s	por_l	0_{16}	RW1S	Translation access when te = 0 secondary error.
35	spare1_s	por_l	0_{16}	RW1S	Spare secondary error.
34	sun4v_inv_pg_sz_s	por_l	0_{16}	RW1S	SUN4V Invalid Page Size secondary error. Illegal value is placed in the IOTSBDESC table.
33	byp_oor_s	por_l	0_{16}	RW1S	Bypass access out-of-range secondary error.
32	byp_err_s	por_l	0_{16}	RW1S	Bypass access when be = 0 secondary error.
31:21	—				*Reserved*
20	sun4v_key_err_p	por_l	0_{16}	RW1S	Device key mismatch primary error (both sun4u/sun4v).

TABLE 21-74 MMU Error Status Set Register *(Continued)*

Bits	Field	Reset Name	Reset Value	R/W	Description
19	sun4v_va_adj_uf_p	por_l	0_{16}	RW1S	sun4v virtual address adjustment underflow primary error.
18	sun4v_va_oor_p	por_l	0_{16}	RW1S	sun4v virtual address out of range primary error, VA{62:40} ≠ 0.
17	iotsbdesc_dpe_p	por_l	0_{16}	RW1S	IOTSBDESC data parity primary error.
16	iotsbdesc_inv_p	por_l	0_{16}	RW1S	IOTSBDESC invalid primary error.
15	tbw_dpe_p	por_l	0_{16}	RW1S	Tablewalk data parity primary error.
14	tbw_err_p	por_l	0_{16}	RW1S	Tablewalk primary error.
13	tbw_ude_p	por_l	0_{16}	RW1S	Tablewalk unexpected data return primary error.
12	tbw_dme_p	por_l	0_{16}	RW1S	Tablewalk disabled miss primary error.
11	spare3_p	por_l	0_{16}	RW1S	Spare primary error.
10	spare2_p	por_l	0_{16}	RW1S	Spare primary error.
9	ttc_cae_p	por_l	0_{16}	RW1S	TTE cache CSR access primary error.
8	ttc_dpe_p	por_l	0_{16}	RW1S	TTE cache data parity primary error.
7	tte_prt_p	por_l	0_{16}	RW1S	TTE write bit not set on write primary error.
6	tte_inv_p	por_l	0_{16}	RW1S	TTE valid bit not set primary error.
5	trn_oor_p	por_l	0_{16}	RW1S	Translation access out-of-range primary error.
4	trn_err_p	por_l	0_{16}	RW1S	Translation access when te = 0 primary error.
3	spare1_p	por_l	0_{16}	RW1S	Spare primary error.
2	sun4v_inv_pg_sz_p	por_l	0_{16}	RW1S	sun4v Invalid Page Size primary error. Illegal value is placed in the IOTSBDESC table.
1	byp_oor_p	por_l	0_{16}	RW1S	Bypass access out-of-range primary error.
0	byp_err_p	por_l	0_{16}	RW1S	Bypass access when be = 0 primary error.

21.4.3.51 MMU Translation Fault Address Register (00641028_{16} / 0_{16})

The MMU Translation Fault Address register aids debugging of software errors. It logs the virtual address / adjusted virtual address on any primary error. This register is undefined unless a primary error is set. Diagnostic software can write this register after a primary error is set.

TABLE 21-75 MMU Translation Fault Address Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:2	va	por_l	0_2	RW	Virtual address.
1:0	—				*Reserved*

21.4.3.52 MMU Translation Fault Status Register (00641030_{16} / 0_{16})

The MMU Translation Fault Status register provides the TTE cache entry address, transaction type, and requester ID on any primary error. This register is undefined unless a primary error is set. Diagnostic software can write this register after a primary error is set.

TABLE 21-76 MMU Translation Fault Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:41	—				*Reserved*
40:32	entry	por_l	0_2	RW	TTE cache entry address.
31:23	—				*Reserved*
22:16	type	por_l	0_2	RW	Transaction type.
15:0	id	por_l	0_2	RW	Requester ID.

21.4.3.53 MMU Performance Counter Select Register (00642000_{16} / 0_{16})

The MMU Performance Counter Select register provides the selects for the performance counters.

TABLE 21-77 MMU Performance Counter Select Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:8	sel1	rst_l	0_{16}	RW	Select for counter 1: 00 – None 01 – Clock cycles 02 – Total translations 03 – Total stall cycles 04 – Total translation misses 05 – Tablewalk stall cycles 06 – Bypass mode translations 07 – Translation mode translations 08 – Flow control stall cycles 09 – Total cache entries flushed
7:0	sel0	rst_l	0_{16}	RW	Select for counter 0: Values are as above.

21.4.3.54 MMU Performance Counter Zero Register (00642008_{16} / 0_{16})

The MMU Performance Counter Zero register counts the occurrences of the events selected by the sel0 field in the MMU Performance Counter Select register.

TABLE 21-78 MMU Performance Counter Zero Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	cnt	rst_l	0_{16}	RW	Counter.

21.4.3.55 MMU Performance Counter One Register (00642010_{16} / 0_{16})

The MMU Performance Counter One register counts the occurrences of the events selected by the sel1 field in the MMU Performance Counter Select register.

TABLE 21-79 MMU Performance Counter One Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	cnt	rst_l	0_{16}	RW	Counter.

21.4.3.56 MMU TTE Cache Virtual Tag Registers (00646000-$006461F8_{16}$ / 0_{16})

The MMU TTE Cache Virtual Tag registers give software access to the virtual tags of the TTE cache. When these registers are loaded for test purposes, the cnt fields should all be unique with the highest number being the valid entry that will be replaced first. Access can only occur when cm = 0 in the MMU Control and Status register.

TABLE 21-80 MMU TTE Cache Virtual Tag Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63:57	—				*Reserved*
56:40	cnt	rst_l	x_2	RW	LRU counter.
39:16	vpn	rsl_l	x_2	RW	Virtual page number in sun4u mode; adjusted virtual page number in sun4v mode.
15:11	iotsb_no	rsl_l	x_2	RW	0_2 in sun4u mode; iotsb_no for IOTSBDESC in sun4v mode.
10:1	—				*Reserved*
0	vld	rst_l	0_2	RW	Valid.

21.4.3.57 MMU TTE Cache Physical Tag Registers (00647000_{16}–$006471F8_{16}$ / 0_{16})

The MMU TTE Cache Physical Tag registers give software access to the physical tags of the TTE cache. Access can only occur when cm = 0 in the MMU Control and Status register.

TABLE 21-81 MMU TTE Cache Physical Tag Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63:39	—				*Reserved*
38:6	tag	rst_l	x_2	RW	Physical tag.
5:1	—				*Reserved*
0	vld	rst_l	0_2	RW	Valid.

21.4.3.58 MMU TTE Cache Data Registers (00648000_{16}–$00648FF8_{16}$ / $0)_{16}$

The MMU TTE Cache Data registers give software access to the data of the TTE cache. Access can only occur when cm = 0 in the MMU Control and Status register. The data includes odd parity which must be calculated and written when these registers are written.

TABLE 21-82 MMU TTE Cache Data Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63:48	key	por_l	x_2	RW	device key -- 16-bit PCIE requester ID.
47:44	par	por_l	x_2	RO	Odd parity: Bit 47 = parity for bits 63:52 Bit 46 = parity for bits 51:48, 38:31 Bit 45 = parity for bits 30:19 Bit 44 = parity for bits 18:13, 5:0
43:39	—				*Reserved*
38:13	ppn	por_l	x_2	RW	Physical Page Number {38:13}.
12:6	—				*Reserved*
5:3	fnm	por_l	x_2	RW	sun4v mode, function number mask; sun4u mode, *Reserved*
2	key_valid	por_l	x_2	RW	sun4v mode, device key Valid bit; sun4u mode, *Reserved*
1	wrt	por_l	x_2	RW	Write.
0	vld	por_l	x_2	RW	Valid.

21.4.3.59 MMU DEV2IOTSB Registers (00649000_{16}–00649078_{16} / 0_{16})

The MMU DEV2IOTSB registers map from a PCI Express ID to an index into the IOTSB descriptor table. The PCI Express ID is a 7-bit value formed from a virtual address bit and bits from the PCI Express busid, as described in *DEV2IOTSB Table* on page 507. Bits 6:3 of this ID select one of the 16 MMU DEV2IOTSB registers. Bits 2:0 of this ID select one of the eight iotsb_no fields within that register.

TABLE 21-83 MMU DEV2IOTSB Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63:61	—				*Reserved*
60:56	iotsb_no_7	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
55:53	—				*Reserved*
52:48	iotsb_no_6	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
47:45	—				*Reserved*
44:40	iotsb_no_5	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
39:37	—				*Reserved*
36:32	iotsb_no_4	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
31:29	—				*Reserved*
28:24	iotsb_no_3	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
23:21	—				*Reserved*
20:16	iotsb_no_2	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
15:13	—				*Reserved*
12:8	iotsb_no_1	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.
7:5	—				*Reserved*
4:0	iotsb_no_0	por_l	0_2	RW	5-bit index into the IOTSB descriptor table.

21.4.3.60 MMU IOTSBDESC Registers (00649100_{16}–$006491F8_{16}$ / 0_{16})

The MMU IOTSBDESC registers define 32 I/0O TSB entries. This array of registers is indexed by the 5-bit iotsb_no index as part of the IOMMU lookup process in sun4v mode, as described in *IOTSB Descriptor Table* on page 502.

TABLE 21-84 MMU IOTSBDESC Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
63	valid	por_l	0_2	RW	Valid bit. 0 = not valid; 1 = valid.
62:61	par	por_l	0_2	RO	These two bits should be ignored by software. Hardware uses these two bit for odd parity protection. Bit 62: {63, {59:32}} Bit 61:{31:0}
60	—				*Reserved*

TABLE 21-84 MMU IOTSBDESC Registers

Bit	Field	Reset Name	Reset Value	R/W	Description
59:34	base_pa	por_l	0_2	RW	Starting physical address of IOTSB, PA{38:13}.
33:7	offset	por_l	0_2	RW	Address offset, offset{26:0}.
6:4	page_size	por_l	0_2	RW	Page size for this iotsb.
3:0	num_pages	por_l	0_2	RW	Number of pages in the iotsb.

21.4.3.61 ILU Error Log Enable Register (00651000_{16} / $F0_{16}$)

This is ILU Error Log Enable register with three spare bits. It enables the logging for all possible ILU errors detected by PIU. By default all errors are enabled to be logged.

TABLE 21-85 ILU Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:8	—				*Reserved*
7	spare3	por_l	1_{16}	RW	Spare error.
6	spare2	por_l	1_{16}	RW	Spare error.
5	spare1	por_l	1_{16}	RW	Spare error.
4	ihb_pe	por_l	1_{16}	RW	Ingress header buffer parity error.
3:0	—				*Reserved*

21.4.3.62 ILU Interrupt Enable Register (00651008_{16} / 0_{16})

The ILU Interrupt Enable register enables the interrupt for all possible ILU errors detected by PIU. By default all errors are not enabled to generate an interrupt

TABLE 21-86 ILU Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:40	—				*Reserved*
39	spare3_s	rst_l	0_{16}	RW	Spare secondary error.
38	spare2_s	rst_l	0_{16}	RW	Spare secondary error.
37	spare1_s	rst_l	0_{16}	RW	Spare secondary error.
36	ihb_pe_s	rst_l	0_{16}	RW	Ingress header buffer parity secondary error.
35:8	—				*Reserved*
7	spare3_p	rst_l	0_{16}	RW	Spare primary error.
6	spare2_p	rst_l	0_{16}	RW	Spare primary error.
5	spare1_p	rst_l	0_{16}	RW	Spare primary error.
4	ihb_pe_p	rst_l	0_{16}	RW	Ingress header buffer parity primary error.

TABLE 21-86 ILU Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
3:0	—				*Reserved*

21.4.3.63 ILU Interrupt Status Register (00651010_{16} / 0_{16})

The ILU Interrupt Status register is a read-only register that software reads to obtain the source of interrupt for the ILU block. This register will contain a 1 in all bits for which an interrupt was generated.

TABLE 21-87 ILU Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:40	—				*Reserved*
39	spare3_s	rst_l	0_{16}	RO	Spare secondary error.
38	spare2_s	rst_l	0_{16}	RO	Spare secondary error.
37	spare1_s	rst_l	0_{16}	RO	Spare secondary error.
36	ihb_pe_s	rst_l	0_{16}	RO	Ingress header buffer parity secondary error.
35:8	—				*Reserved*
7	spare3_p	rst_l	0_{16}	RO	Spare primary error.
6	spare2_p	rst_l	0_{16}	RO	Spare primary error.
5	spare1_p	rst_l	0_{16}	RO	Spare primary error.
4	ihb_pe_p	rst_l	0_{16}	RO	Ingress header buffer parity primary error.
3:0	—				*Reserved*

21.4.3.64 ILU Error Status Clear Register (00651018_{16} / 0_{16})

The ILU Error Status Clear register is a read, write 1 to clear register that serves two purposes for software. First, software can read this register to determine which ILU errors have been logged by PIU but have not necessarily caused an interrupt (mondo 63). This register will contain a 1 in all bits where errors were logged. Second, software uses this register to clear the ILU errors detected by PIU. It writes a 1 to the bit(s) that it wants to clear.

.

TABLE 21-88 ILU Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:40	—				*Reserved*
39	spare3_s	por_l	0_{16}	RW1C	Spare secondary error.
38	spare2_s	por_l	0_{16}	RW1C	Spare secondary error.
37	spare1_s	por_l	0_{16}	RW1C	Spare secondary error.

TABLE 21-88 ILU Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
36	ihb_pe_s	por_l	0_{16}	RW1C	Ingress header buffer parity secondary error.
35:8	—				*Reserved*
7	spare3_p	por_l	0_{16}	RW1C	Spare primary error.
6	spare2_p	por_l	0_{16}	RW1C	Spare primary error.
5	spare1_p	por_l	0_{16}	RW1C	Spare primary error.
4	ihb_pe_p	por_l	0_{16}	RW1C	Ingress header buffer parity primary error.
3:0	—				*Reserved*

21.4.3.65 ILU Error Status Set Register (00651020_{16} / 0_{16})

The ILU Error Status Set register is a read, write 1 to set register that serves two purposes for software. First, software can read this register to determine which ILU errors have been logged by PIU but have not necessarily caused an interrupt (Mondo 63). This register will contain a 1 in all bits where errors were logged. Second, software uses this register to set the ILU errors detected by PIU. It writes a 1 to the bit(s) that it wants to clear. *This should only be used for diagnostic purposes and error testing.*

TABLE 21-89 ILU Error Status Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:40	—				*Reserved*
39	spare3_s	por_l	0_{16}	RW1S	Spare secondary error.
38	spare2_s	por_l	0_{16}	RW1S	Spare secondary error.
37	spare1_s	por_l	0_{16}	RW1S	Spare secondary error.
36	ihb_pe_s	por_l	0_{16}	RW1S	Ingress header buffer parity secondary error.
35:8	—				*Reserved*
7	spare3_p	por_l	0_{16}	RW1S	Spare primary error.
6	spare2_p	por_l	0_{16}	RW1S	Spare primary error.
5	spare1_p	por_l	0_{16}	RW1S	Spare primary error.
4	ihb_pe_p	por_l	0_{16}	RW1S	Ingress header buffer parity primary error.
3:0	—				*Reserved*

21.4.3.66 PEU Core and Block Interrupt Enable Register (00651800_{16} / 0_{16})

The PEU Core and Block Interrupt Enable register is described in TABLE 21-90.

TABLE 21-90 PEU Core and Block Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	pec	rst_l	0_{16}	RW	The mask bit to mask the PEU interrupt.
62:4	—				*Reserved*
3	pec_ilu	rst_l	0_{16}	RW	ILU block interrupt enable bit.
2	pec_ue	rst_l	0_{16}	RW	Uncorrectable error interrupt enable bit.
1	pec_ce	rst_l	0_{16}	RW	Correctable error interrupt enable bit.
0	pec_oe	rst_l	0_{16}	RW	Other event interrupt enable bit.

21.4.3.67 PEU Core and Block Interrupt Status Register (00651808_{16} / 0_{16})

The PEU Core and Block Interrupt Status register is described in TABLE 21-91.

TABLE 21-91 PEU Core and Block Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:4	—				*Reserved*
3	ilu	rst_l	0_{16}	RW	ILU header parity error.
2	ue	rst_l	0_{16}	RW	Uncorrectable error from PEU.
1	ce	rst_l	0_{16}	RW	Correctable error from PEU.
0	oe	rst_l	0_{16}	RW	Other event from PEU.

21.4.3.68 DMU ILU Diagnostic Register (00652000_{16} / 0_{16})

The DMU ILU Diagnostic register allows for EHB/EDB error injection and allows software to control PLL-enable, TX-lane-enable and RX-lane-enable bits. When PEU enters L23 ready state, software can optionally power down the SerDes by disable all the PLL/TX-lane/RX-lane enable bits.

TABLE 21-92 DMU ILU Diagnostic Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:34	—				*Reserved*
33	enpll1	rst_l	1_{16}	RW	Enable PEU SerDes PLL for Lane 4 – Lane 7: 0 = disable; 1 = enable.
32	enpll0	rst_l	1_{16}	RW	Enable PEU SerDes PLL for Lane0 – Lane3: 0 = disable; 1 = enable.
31	entx7	rst_l	1_{16}	RW	Enable PEU SerDes Lane 7's Transmitter: 0 = disable; 1 = enable.
30	entx6	rst_l	1_{16}	RW	Enable PEU SerDes Lane 6's Transmitter: 0 = disable; 1 = enable.
29	entx5	rst_l	1_{16}	RW	Enable PEU SerDes Lane 5's Transmitter: 0 = disable; 1 = enable.
28	entx4	rst_l	1_{16}	RW	Enable PEU SerDes Lane 4's Transmitter: 0 = disable; 1 = enable.

TABLE 21-92 DMU ILU Diagnostic Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
27	entx3	rst_l	1_{16}	RW	Enable PEU SerDes Lane 3's Transmitter: 0 = disable; 1 = enable.
26	entx2	rst_l	1_{16}	RW	Enable PEU SerDes Lane 2's Transmitter: 0 = disable; 1 = enable.
25	entx1	rst_l	1_{16}	RW	Enable PEU SerDes Lane 1's Transmitter: 0 = disable; 1 = enable.
24	entx0	rst_l	1_{16}	RW	Enable PEU SerDes Lane 0's Transmitter: 0 = disable; 1 = enable.
23	enrx7	rst_l	1_{16}	RW	Enable PEU SerDes Lane 7's Receiver: 0 = disable; 1 = enable.
22	enrx6	rst_l	1_{16}	RW	Enable PEU SerDes Lane 6's Receiver: 0 = disable; 1 = enable.
21	enrx5	rst_l	1_{16}	RW	Enable PEU SerDes Lane 5's Receiver; 0 = disable; 1 = enable.
20	enrx4	rst_l	1_{16}	RW	Enable PEU SerDes Lane 4's Receiver; 0 = disable; 1 = enable.
19	enrx3	rst_l	1_{16}	RW	Enable PEU SerDes Lane 3's Receiver: 0 = disable; 1 = enable.
18	enrx2	rst_l	1_{16}	RW	Enable PEU SerDes Lane 2's Receiver; 0 = disable; 1 = enable.
17	enrx1	rst_l	1_{16}	RW	Enable PEU SerDes Lane 1's Receiver: 0 = disable; 1 = enable.
16	enrx0	rst_l	1_{16}	RW	Enable PEU SerDes Lane 0's Receiver: 0 = disable; 1 = enable.
15:12	edi_par	rst_l	0_{16}	RW	Egress data parity invert bits.
11:8	ehi_par	rst_l	0_{16}	RW	Egress header parity invert bits.
7:6	—				*Reserved*
5	edi_trg	rst_l	0_{16}	RW1S	Egress data parity invert trigger. When set, the next valid write into the egress data buffer will invert parity based upon the value of the egress data parity invert bits. This bit will be cleared when this occurs.
4	ehi_trg	rst_l	0_{16}	RW1S	Egress header parity invert trigger. When set, the next valid write into the egress header buffer will invert parity based upon the value of the egress header parity invert bits. This bit will be cleared when this occurs.
3:2	rate scale	por_l	2'b0	RW	This fields control all the operating rate for receiver/transmit channel: 00_2 – Full Rate. Two data samples taken per PLL output clock cycle. 01_2 – Half Rate. One data samples taken per PLL output clock cycle. 10_2 – Quarter Rate. One data samples taken every two PLL output clock cycles. 11_2 – *Reserved*
1:0	—				*Reserved*

21.4.3.69 DMU Debug Select Register for DMU Debug Bus A

(00653000_{16} / 0_{16})

The DMU Debug Select Register for DMU debug bus A is described in TABLE 21-93.

TABLE 21-93 DMU Debug Select Register for DMU Debug Bus A

Bit	Field	Reset Name	Reset Value	R/W	Description
63:10	—				*Reserved*
9:6	block_sel	rst_l	0_{16}	RW	DMU block debug selects for DMU debug bus A: 0000 – All zeros 0001 – CLU block selects (Cache Line Unit) 0010 – CMU block selects (Context Manager Unit) 0011 – CRU block selects (CSR Request Unit) 0100 – DSN block selects 0101 – Training sequences select 0110 – ILU block selects (Interface Layer Unit) 0111 – All zeros 1000 – All zeros 1001 – IMU block selects (Interrupt Message Unit) 1010 – MMU block selects 1011 – PMU block selects (Packet Manager Unit) 1100 – PSB block selects (Packet Scoreboard Unit) 1101 – RMU block selects (Record Manager Unit) 1110 – TMU block selects (Transaction Manager Unit) 1111 – TSB block selects (Transaction Scoreboard Unit)
5:3	sub_sel	rst_l	0_{16}	RW	Select the sub-block for DMU debug bus A.
2:0	signal_sel	rst_l	0_{16}	RW	Select the signals for DMU debug bus A.

21.4.3.70 DMU Debug Select Register for DMU Debug Bus B

(00653008_{16} / 0_{16})

The DMU Debug Select Register for DMU debug bus B is described in TABLE 21-94.

TABLE 21-94 DMU Debug Select Register for DMU Debug Bus B

Bit	Field	Reset Name	Reset Value	R/W	Description
63:10	—				*Reserved*
9:6	block_sel	rst_l	0_{16}	RW	DMU block debug selects for DMU debug bus B: 0000 – All zeros 0001 – CLU block selects (Cache Line Unit) 0010 – CMU block selects (Context Manager Unit) 0011 – CRU block selects (CSR Request Unit) 0100 – DSN block selects 0101 – Training sequence selects 0110 – ILU block selects (Interface Layer Unit) 0111 – All zeros 1000 – All zeros 1001 – IMU block selects (Interrupt Message Unit) 1010 – MMU block selects 1011 – PMU block selects (Packet Manager Unit) 1100 – PSB block selects (Packet Scoreboard Unit) 1101 – RMU block selects (Record Manager Unit) 1110 – TMU block selects (Transaction Manager Unit) 1111 – TSB block selects (Transaction Scoreboard Unit)
5:3	sub_sel	rst_l	0_{16}	RW	Select the sub-block for DMU debug bus B.
2:0	signal_sel	rst_l	0_{16}	RW	Select the signals for DMU debug bus B.

21.4.3.71 DMU PCI Express Configuration Register (00653100_{16} / 0_{16})

The DMU PCI Express Configuration register specifies the requester ID used for PIU and the bus number used for Type 0 configuration requests.

TABLE 21-95 DMC PCI Express Configuration Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	bus_num	rst_l	0_{16}	RW	Bus number used for Type 0 configuration requests.
23:16	—				*Reserved*
15:0	req_id	rst_l	0_{16}	RW	Requester ID used by PIU.

21.4.3.72 Packet Scoreboard DMA Register Set (00660000_{16}–$006600F8_{16}$ / 0_{16})

The Packet Scoreboard DMA register set comprises 32 register sets of 41 bits each for storing and retrieving DMA record information associated with a given packet.

TABLE 21-96 Packet Scoreboard DMA Register Set

Bit	Field	Reset Name	Reset Value	R/W	Description
63:41	—				*Reserved*
40:0	entry	rst_l	0_{16}	RO	entry{40:36} – Transaction tag entry{35:31} – Context number entry{30:26} – Packet sequence number entry{25:22} – Cacheline total entry{21:12} – Length entry{11:0} – Byte count

21.4.3.73 Packet Scoreboard PIO Register Set (00664000_{16}–00664078_{16} / 0_{16})

The Packet Scoreboard PIO register set comprises 16 register sets of 6 bits each for storing and retrieving PIO Record Information associated with a given packet.

TABLE 21-97 Packet Scoreboard PIO Register Set

Bit	Field	Reset Name	Reset Value	R/W	Description
63:6	—				*Reserved*
5:0	entry	rst_l	0_{16}	RO	Thread ID

21.4.3.74 Transaction Scoreboard Register Set (00670000_{16}–$006700F8_{16}$ / 0_{16})

The Transaction Scoreboard register set comprises 32 register sets of 48 bits each for storing and retrieving DMA read record information associated with a given transaction

.

TABLE 21-98 Transaction Scoreboard Register Set

Bit	Field	Reset Name	Reset Value	R/W	Description
63:48	—				*Reserved*
47:0	entry	rst_l	0_{16}	RO	entry{47:45} – Traffic class entry{44:43} – Attribute entry{42:31} – Byte count entry{30:15} – Requestor ID entry{14:7} – Transaction layer packet entry{6:0} – Address alignment bits

21.4.3.75 Transaction Scoreboard Status Register (00670100_{16} / 1_{16})

The Transaction Scoreboard Status register reports the status on pending DMAs.

TABLE 21-99 Transaction Scoreboard Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:8	—				*Reserved*
7	full	rst_l	0_{16}	RO	TSB fullness. 0 = TSB is not full; 1 = TSB is full.
6:1	num_pnd_dma	rst_l	0_{16}	RO	Number of pending DMA transactions
0	empty	rst_l	1_{16}	RO	TSB emptiness. 0 = TSB is not empty; 1 = TSB is empty (no pending DMAs).

21.4.3.76 PEU Control Register (00680000_{16} / 1_{16})

The PEU Control register controls the transaction layer. It also provides configuration for the link layer.

TABLE 21-100 PEU Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	l0s_tim	por_l	0_{16}	RW	L0s Entry Timer Value. Values are in multiples of 32 ns, as follows: 00_{16} – 0 ns 01_{16} – 32 ns 02_{16} – 64 ns DA_{16} – 7.0 µs (recommended) FF_{16} – 8.2 µs
23:21	—				*Reserved*
20	npwr_en	por_l	0_{16}	RW	Nonposted Write Enable. 0 = Issue non-posted writes at will; 1 = Nonposted write will stall issue of all other requests until it completes.
19	—				*Reserved*
18:16	cto_sel	por_l	0_{16}	RW	Completion Timeout Select, as follows: 000_2 – Infinite 001_2 – 2^{26} symbol times (268 ms) 010_2 – 2^{25} symbol times (134 ms) 011_2 – 2^{24} symbol times (67.1 ms) 100_2 – 2^{23} symbol times (33.6 ms) 101_2 – 2^{22} symbol times (16.8 ms) 110_2 – 2^{21} symbol times (8.4 ms) 111_2 – 2^{10} symbol times (4 µs)

TABLE 21-100 PEU Control Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
15:0	config	por_l	101_{16}	RW	Configuration bits, as follows. config{0} – Port. 0 = upstream port; 1 = downstream port. config{1} – common clock configuration. 0 = This component and the component at the opposite end of this link are operating with asynchronous reference clock; 1 = This component and the component at the opposite end of this link are operating with a distributed common reference clock. config{6:2} – *Reserved* config{7} – Blunt-end loopback enable. 0 = be loopback disable; 1 = be loopback enable config{8} – Remain in `Detect.Quiet`. 0 = Processed with link training sequence; 1 = Remain in state `Detect.Quiet`. config{9} – Lane Flipping Enable, as follows: 0 = Lane Flipping Disable: Logical Lane 0 = Physical Lane 0 Logical Lane 1 = Physical Lane 1 Logical Lane 2 = Physical Lane 2 Logical Lane 3 = Physical Lane 3 Logical Lane 4 = Physical Lane 4 Logical Lane 5 = Physical Lane 5 Logical Lane 6 = Physical Lane 6 Logical Lane 7 = Physical Lane 7 1 = Lane Flipping Enable: Logical Lane 0 = Physical Lane 7 Logical Lane 1 = Physical Lane 6 Logical Lane 2 = Physical Lane 5 Logical Lane 3 = Physical Lane 4 Logical Lane 4 = Physical Lane 3 Logical Lane 5 = Physical Lane 2 Logical Lane 6 = Physical Lane 1 Logical Lane 7 = Physical Lane 0 config{10} – SPARE config{11} – SPARE config{12} – SPARE config{13} – SPARE config{14} – SPARE config{15} –- SPARE

21.4.3.77 PEU Status Register (00680008_{16} / 01_{16})

The PEU Status register reports the status from the link layer.

TABLE 21-101 PEU Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:9	—				*Reserved*
8	drain	rst_l	0_{16}	RW1C	Drain state bit. Set by hardware when in the drain state. Cleared by software when preparing to bring the link backup. **Note**: This bit is not set for the scenario of IHB Parity Error.
7:0	status	rst_l	01_{16}	RO	Status bit. status{2:0} – Data Link State, as follows: 001_2 – Data Link Inactive 010_2 – Data Link Init 100_2 –Data Link Active status{3} – Recovery status{7:4} – Power Management Controller State, as follows: 0000_2 – IDLE 0001_2 – pm_nak 0010_2 – pm_nak_wait 0011_2 – pm_ack_aspm_l1 0100_2 – pm_ack_l1 0101_2 – pm_ack_l23 0110_2 – pm_trans_l0 0111_2 – pm_trans_l0_aspm_l1 1000_2 – pm_trans_l0_l1 1001_2 – pm_trans_l0_l23 1010_2 – pm_timeout

21.4.3.78 PEU PME Turn Off Generate Register (00680010_{16} / 0_{16})

The PEU PME Turn Off Generate register generates a PME_turn_off message,

TABLE 21-102 PEU PME Turn Off Generate Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:1	—				*Reserved*
0	pto	rst_l	0_{16}	RW1S	Generate PME_turn_off message. 0 = not pending; 1 = pending

21.4.3.79 PEU Ingress Credits Initial Register (00680018_{16} / $10000200C0_{16}$)

The PEU Ingress Credits Initial register specifies the initial credits that are advertised. The completion header credit, completion data credit, and nonposted data credit are all advertised as infinite and cannot be changed. The sum of the non-posted header credit and posted header credit must be less than or equal to 30_{16}. The posted data credit must be less than or equal to $0C0_{16}$. The values take effect immediately by hardware for initial credit advertisement after power-on reset and before the link is up. After the link is up the first time, the value in only take effect through warm reset or PIU sub-system reset.

.

TABLE 21-103 PEU Ingress Credits Initial Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:60	—				*Reserved*
59:52	chc	por_l	0_{16}	RO	Completion header credit.
51:40	cdc	por_l	0_{16}	RO	Completion data credit.
39:32	nhc	por_l	10_{16}	RW	Nonposted header credit.
31:20	ndc	por_l	0_{16}	RO	Nonposted data credit.
19:12	phc	por_l	20_{16}	RW	Posted header credit.
11:0	pdc	por_l	$C0_{16}$	RW	Posted data credit.

21.4.3.80 PEU Diagnostic Register (00680100_{16} / 0_{16})

The PEU Diagnostic register allows for error injection and other diagnostic functions.

TABLE 21-104 PEU Diagnostic Register

Bits	Field	Reset Name	Reset Value	R/W	Description
63:60	—				*Reserved*
59:56	erbi_par	rst_l	0_{16}	RW	Egress retry buffer parity invert.
55	erbi_trg	rst_l	0_{16}	RW1S	Egress retry buffer parity invert trigger. When set, the next valid write into the egress retry buffer will invert parity, based on the value of the egress retry buffer parity invert bits. The bit will be cleared when this occurs.
54:48	—				*Reserved*

TABLE 21-104 PEU Diagnostic Register *(Continued)*

Bits	Field	Reset Name	Reset Value	R/W	Description
47:32	chk_dis	rst_l	0_{16}	RW	Disable error check, as follows: chk_dis{32} – Flow control max credit chk_dis{33} – Flow control finite update chk_dis{34} – *Reserved* chk_dis{35} – *Reserved* chk_dis{36} – Byte enable rules chk_dis{37} – 4KB boundary crossing chk_dis{38} – Receiver overflow chk_dis{39} – *Reserved* chk_dis{40} – Nonconfiguration with configuration retry status chk_dis{41} – TC mismatch in CPL/CPLD chk_dis{42} – Attribute mismatch in CPL/CPLD chk_dis{43} – Byte count mismatch in CPL/CPLD. chk_dis{44} – Lower address mismatch in CPL/CPLD chk_dis{47:45} – *Reserved*
31:16	epi_par	rst_l	0_{16}	RW	Egress parity invert.
15:12	idi_par	rst_l	0_{16}	RW	Ingress data parity invert.
11:8	ihi_par	rst_l	0_{16}	RW	Ingress header parity invert.
7	epi_trg	rst_l	0_{16}	RW1S	Egress parity invert trigger. When set, the next valid cycle on the Egress Transaction Packet interface will invert parity, based on the value of the egress parity invert bits. This bit will be cleared when this occurs.
6	idi_trg	rst_l	0_{16}	RW1S	Ingress data parity invert trigger. When set, the next valid write into the ingress data buffer will invert parity, based on the value of the ingress data parity invert bits. This bit will be cleared when this occurs.
5	ihi_trg	rst_l	0_{16}	RW1S	Ingress header parity invert trigger. When set, the next valid write into the ingress header buffer will invert parity, based on the value of the ingress header parity invert bits. This bit will be cleared when this occurs.
4	mrc_trg	rst_l	0_{16}	RW1S	Memory read capture trigger. When this bit is set, the next memory read request will be removed from the pipeline and, if enabled, will be captured in the Receive Other Event Header Log registers. This bit will be cleared when this occurs.
3:2	—				*Reserved*
1	epp_dis	rst_l	0_{16}	RW	Egress packet processing disable. When this bit is set, no packets will be transmitted.
0	ifc_dis	rst_l	0_{16}	RW	Ingress flow control update disable. When this bit is set, the ingress flow control allocated credits will not get updated when they become available. When this bit is subsequently cleared, credits will be restored as normal.

21.4.3.81 PEU Egress Credits Consumed Register $(00680200_{16} / 0)_{16}$

The PEU Egress Credits Consumed register indicates the current credits consumed by the egress block. It also indicates whether any of the header credits were advertised as being infinite.

TABLE 21-105 PEU Egress Credits Consumed Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	—				*Reserved*
62	chi	rst_l	0_{16}	RO	Infinite completion header credits.
61	nhi	rst_l	0_{16}	RO	Infinite nonposted header credits.
60	phi	rst_l	0_{16}	RO	Infinite posted header credits.
59:52	chc	rst_l	0_{16}	RO	Completion header credit.
51:40	cdc	rst_l	0_{16}	RO	Completion data credit.
39:32	nhc	rst_l	0_{16}	RO	Nonposted header credit.
31:20	ndc	rst_l	0_{16}	RO	Nonposted data credit.
19:12	phc	rst_l	0_{16}	RO	Posted header credit.
11:0	pdc	rst_l	0_{16}	RO	Posted data credit.

21.4.3.82 PEU Egress Credit Limit Register $(00680208_{16} / 0_{16})$

The PEU Egress Credit Limit register indicates the current credit limit advertised to the egress block. It also indicates whether any of the data credits were advertised as being infinite

.

TABLE 21-106 PEU Egress Credit Limit Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	—				*Reserved*
62	cdi	rst_l	0_{16}	RO	Infinite completion data credits.
61	ndi	rst_l	0_{16}	RO	Infinite nonposted data credits.
60	pdi	rst_l	0_{16}	RO	Infinite posted data credits.
59:52	chc	rst_l	0_{16}	RO	Completion header credit.
51:40	cdc	rst_l	0_{16}	RO	Completion data credit.
39:32	nhc	rst_l	0_{16}	RO	Nonposted header credit.
31:20	ndc	rst_l	0_{16}	RO	Nonposted data credit.
19:12	phc	rst_l	0_{16}	RO	Posted header credit.
11:0	pdc	rst_l	0_{16}	RO	Posted data credit.

21.4.3.83 PEU Egress Retry Buffer Register (00680210_{16} / 0_{16})

The PEU Egress Retry Buffer register indicates the credits consumed and the credit limit for the retry buffer. The granularity is *byte* in this register.

TABLE 21-107 PEU Egress Retry Buffer Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:48	—				*Reserved*
47:32	cc	rst_l	0_{16}	RO	Credits consumed.
31:16	—				*Reserved*
15:0	cl	rst_l	1000_{16}	RO	Credit limit.

21.4.3.84 PEU Ingress Credits Allocated Register (00680218_{16} / $10000200C0_{16}$)

The PEU Ingress Credits Allocated register indicates the credits currently allocated by the ingress block.

TABLE 21-108 PEU Ingress Credits Allocated Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:60	—				*Reserved*
59:52	chc	rst_l	0_{16}	RO	Completion header credit.
51:40	cdc	rst_l	0_{16}	RO	Completion data credit.
39:32	nhc	rst_l	10_{16}	RO	Nonposted header credit.
31:20	ndc	rst_l	0_{16}	RO	Nonposted data credit.
19:12	phc	rst_l	20_{16}	RO	Posted header credit.
11:0	pdc	rst_l	$0C0_{16}$	RO	Posted data credit.

21.4.3.85 PEU Ingress Credits Received Register (00680220_{16} / 0_{16})

The PEU Ingress Credits Received register indicates the credits currently received by the ingress block.

TABLE 21-109 PEU Ingress Credits Received Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:60	—				*Reserved*
59:52	chc	rst_l	0_{16}	RO	Completion header credit.
51:40	cdc	rst_l	0_{16}	RO	Completion data credit.

TABLE 21-109 PEU Ingress Credits Received Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
39:32	nhc	rst_l	0_{16}	RO	Nonposted header credit.
31:20	ndc	rst_l	0_{16}	RO	Nonposted data credit
19:12	phc	rst_l	0_{16}	RO	Posted header credit
11:0	pdc	rst_l	0_{16}	RO	Posted data credit

21.4.3.86 PEU Other Event Log Enable Register (00681000_{16} / $FFFFFF_{16}$)

The PEU Other Event Log Enable register controls which events will be logged in the PEU Other Event Status Clear register.

TABLE 21-110 PEU Other Event Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:24	—				*Reserved*
23:0	en	por_l	$FFFFFF_{16}$	RW	Log enables.

21.4.3.87 PEU Other Event Interrupt Enable Register (00681008_{16} / 0_{16})

The PEU Other Event Interrupt Enable register controls which events logged in the PEU Other Event Status Clear register will generate an interrupt.

TABLE 21-111 PEU Other Event Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:56	—				*Reserved*
55:32	en_s	rst_l	0_{16}	RW	Secondary interrupt enables.
31:24	—				*Reserved*
23:0	en_p	rst_l	0_{16}	RW	Primary interrupt enables.

21.4.3.88 PEU Other Event Interrupt Status Register (00681010_{16} / 0_{16})

The PEU Other Event Interrupt Status register indicates which events currently logged in the PEU Other Event Status Clear register are enabled to generate interrupts by the PEU Other Event Interrupt Enable register.

TABLE 21-112 PEU Other Event Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:56	—				*Reserved*
55:32	err_s	rst_l	0_{16}	RO	Secondary interrupt status.
31:24	—				*Reserved*
23:0	err_p	rst_l	0_{16}	RO	Primary interrupt status.

21.4.3.89 PEU Other Event Status Clear Register (00681018_{16} / 0_{16})

The PEU Other Event Status Clear register indicates that an event occurred. Primary events include data logged with these events. Secondary events have no data logged with them. All events belong to the same event group, meaning that any primary event logged will cause any additional event to get logged as secondary.

TABLE 21-113 PEU Other Event Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:56	—				*Reserved*
55	spare_s	por_l	0_{16}	RW1C	Spare secondary events.
54	mfc_s	por_l	0_{16}	RW1C	Malformed completion secondary event.
53	cto_s	por_l	0_{16}	RW1C	Completion timeout secondary event.
52	nfp_s	por_l	0_{16}	RW1C	No forward progress secondary event. Set When no TLP is successfully received between two consecutive recovery attempts.
51	lwc_s	por_l	0_{16}	RW1C	Link width change secondary event.
50	mrc_s	por_l	0_{16}	RW1C	Memory read capture secondary event.
49	wuc_s	por_l	0_{16}	RW1C	Write unsuccessful completion status secondary event.
48	ruc_s	por_l	0_{16}	RW1C	Read unsuccessful completion status secondary event.
47	crs_s	por_l	0_{16}	RW1C	Configuration request retry completion status secondary event.
46	iip_s	por_l	0_{16}	RW1C	Ingress interface parity error secondary event.
45	edp_s	por_l	0_{16}	RW1C	Egress data parity error secondary event.
44	ehp_s	por_l	0_{16}	RW1C	Egress header parity error secondary event.
43	spare	por_l	0_{16}	RW1C	Spare.
42	lrs_s	por_l	0_{16}	RW1C	Link reset secondary event.
41	ldn_s	por_l	0_{16}	RW1C	Link down secondary event.
40	lup_s	por_l	0_{16}	RW1C	Link up secondary event.
39:38	lpu_s	por_l	0_{16}	RW1C	Spare secondary events.
37	eru_s	por_l	0_{16}	RW1C	Egress retry buffer underflow error secondary event.
36	ero_s	por_l	0_{16}	RW1C	Egress retry buffer overflow error secondary event.
35	emp_s	por_l	0_{16}	RW1C	egress minimum packet error secondary event.

TABLE 21-113 PEU Other Event Status Clear Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
34	epe_s	por_l	0_{16}	RW1C	Egress protocol error secondary event.
33	erp_s	por_l	0_{16}	RW1C	Egress retry parity error secondary event.
32	eip_s	por_l	0_{16}	RW1C	Egress interface parity error secondary event.
31:24	—				*Reserved*
23	spare_p	por_l	0_{16}	RW1C	Spare primary events.
22	mfc_p	por_l	0_{16}	RW1C	Malformed completion primary event.
21	cto_p	por_l	0_{16}	RW1C	Completion timeout primary event.
20	nfp_p	por_l	0_{16}	RW1C	No forward progress primary event. Set When no TLP is successfully received between two consecutive recovery attempts.
19	lwc_p	por_l	0_{16}	RW1C	Link width change primary event.
18	mrc_p	por_l	0_{16}	RW1C	Memory read capture primary event.
17	wuc_p	por_l	0_{16}	RW1C	Write unsuccessful completion status primary event.
16	ruc_p	por_l	0_{16}	RW1C	Read unsuccessful completion status primary event.
15	crs_p	por_l	0_{16}	RW1C	Configuration request retry completion status primary event.
14	iip_p	por_l	0_{16}	RW1C	Ingress interface parity error primary event.
13	edp_p	por_l	0_{16}	RW1C	Egress data parity error primary event.
12	ehp_p	por_l	0_{16}	RW1C	Egress header parity error primary event.
11	lin	por_l	0_{16}	RW1C	CXPL interrupt event. This interrupt doesn't belong to the primary error group.
10	lrs_p	por_l	0_{16}	RW1C	Link reset primary event.
9	ldn_p	por_l	0_{16}	RW1C	Link down primary event.
8	lup_p	por_l	0_{16}	RW1C	Link up primary event.
7:6	lpu_p	por_l	0_{16}	RW1C	Spare primary events.
5	eru_p	por_l	0_{16}	RW1C	Egress retry buffer underflow error primary event.
4	ero_p	por_l	0_{16}	RW1C	Egress retry buffer overflow error primary event.
3	emp_p	por_l	0_{16}	RW1C	Egress minimum packet error primary event.
2	epe_p	por_l	0_{16}	RW1C	Egress protocol error primary event.
1	erp_p	por_l	0_{16}	RW1C	Egress retry parity error primary event.
0	eip_p	por_l	0_{16}	RW1C	Egress interface parity error primary event.

21.4.3.90 PEU Other Event Status Set Register (00681020_{16} / 0_{16})

The PEU Other Event Status Set register controls the setting of events in the PEU Other Event Status Clear register for testing.

TABLE 21-114 PEU Other Event Status Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:56	—				*Reserved*
55	spare_s	por_l	0_{16}	RW1S	Spare secondary events.
54	mfc_s	por_l	0_{16}	RW1S	Malformed completion secondary event.
53	cto_s	por_l	0_{16}	RW1S	Completion timeout secondary event.
52	nfp_s	por_l	0_{16}	RW1S	No forward progress secondary event. Set When no TLP is successfully received between two consecutive recovery attempts.
51	lwc_s	por_l	0_{16}	RW1S	Link width change secondary event.
50	mrc_s	por_l	0_{16}	RW1S	Memory read capture secondary event.
49	wuc_s	por_l	0_{16}	RW1S	Write unsuccessful completion status secondary event.
48	ruc_s	por_l	0_{16}	RW1S	Read unsuccessful completion status secondary event.
47	crs_s	por_l	0_{16}	RW1S	Configuration request retry completion status secondary event.
46	iip_s	por_l	0_{16}	RW1S	Ingress interface parity error secondary event.
45	edp_s	por_l	0_{16}	RW1S	Egress data parity error secondary event.
44	ehp_s	por_l	0_{16}	RW1S	Egress header parity error secondary event.
43	spare	por_l	0_{16}	RW1S	Spare
42	lrs_s	por_l	0_{16}	RW1S	Link reset secondary event.
41	ldn_s	por_l	0_{16}	RW1S	Link down secondary event.
40	lup_s	por_l	0_{16}	RW1S	Link up secondary event.
39:38	lpu_s	por_l	0_{16}	RW1S	Spare secondary events.
37	eru_s	por_l	0_{16}	RW1S	Egress retry buffer underflow error secondary event.
36	ero_s	por_l	0_{16}	RW1S	Egress retry buffer overflow error secondary event.
35	emp_s	por_l	0_{16}	RW1S	Egress minimum packet error secondary event.
34	epe_s	por_l	0_{16}	RW1S	Egress protocol error secondary event.
33	erp_s	por_l	0_{16}	RW1S	Egress retry parity error secondary event.
32	eip_s	por_l	0_{16}	RW1S	Egress interface parity error secondary event.
31:24	—				*Reserved*
23	spare_p	por_l	0_{16}	RW1S	Spare primary events.
22	mfc_p	por_l	0_{16}	RW1S	Malformed completion primary event.
21	cto_p	por_l	0_{16}	RW1S	Completion timeout primary event.
20	nfp_p	por_l	0_{16}	RW1S	No forward progress primary event. Set When no TLP is successfully received between two consecutive recovery attempts.
19	lwc_p	por_l	0_{16}	RW1S	Link width change primary event.
18	mrc_p	por_l	0_{16}	RW1S	Memory read capture primary event.
17	wuc_p	por_l	0_{16}	RW1S	Write unsuccessful completion status primary event.
16	ruc_p	por_l	0_{16}	RW1S	Read unsuccessful completion status primary event.
15	crs_p	por_l	0_{16}	RW1S	Configuration request retry completion status primary event.
14	iip_p	por_l	0_{16}	RW1S	Ingress interface parity error primary event.

TABLE 21-114 PEU Other Event Status Set Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
13	edp_p	por_l	0_{16}	RW1S	Egress data parity error primary event.
12	ehp_p	por_l	0_{16}	RW1S	Egress header parity error primary event.
11	lin	por_l	0_{16}	RW1S	CXPL interrupt event. This interrupt doesn't belong to the primary error group.
10	lrs_p	por_l	0_{16}	RW1S	Link reset primary event.
9	ldn_p	por_l	0_{16}	RW1S	Link down primary event.
8	lup_p	por_l	0_{16}	RW1S	Link up primary event.
7:6	lpu_p	por_l	0_{16}	RW1S	Spare primary events.
5	eru_p	por_l	0_{16}	RW1S	Egress retry buffer underflow error primary event.
4	ero_p	por_l	0_{16}	RW1S	Egress retry buffer overflow error primary event.
3	emp_p	por_l	0_{16}	RW1S	Egress minimum packet error primary event.
2	epe_p	por_l	0_{16}	RW1S	Egress protocol error primary event.
1	erp_p	por_l	0_{16}	RW1S	Egress retry parity error primary event.
0	eip_p	por_l	0_{16}	RW1S	Egress interface parity error primary event.

21.4.3.91 PEU Receive Other Event Header1 Log Register

(00681028_{16} / 0_{16})

The PEU Receive Other Event Header1 Log register holds the first eight bytes of the received header when one of the following primary other events in the PEU Logged Other Event Error Status Clear register is set:

- Ingress completion header error
- Memory read capture
- Write unsuccessful completion status
- Read unsuccessful completion status
- Configuration request retry completion status

Note | For the rest of OE errors or when there's no OE errors, the Receive Other Event Header1 Log register contains no valid header information, and software should ignore the content of this register.

Implementation Note | For warm reset protection, the value of PEU Receive Other Event Header1/Header2 Log register is warm reset protected only when a primary error bit is set in the PEU Other Event Status Clear register.

TABLE 21-115 PEU Receive Other Event Header1 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0_2	RW	hdr{63:56} – Header byte 0 hdr{55:48} – Header byte 1 hdr{47:40} – Header byte 2 hdr{39:32} – Header byte 3 hdr{31:24} – Header byte 4 hdr{23:16} – Header byte 5 hdr{15:8} – Header byte 6 hdr{7:0} – Header byte 7

21.4.3.92 PEU Receive Other Event Header2 Log Register

(00681030_{16} / 0_{16})

The PEU Receive Other Event Header2 Log register holds the second eight bytes of the received header when one of the following primary other events in the PEU Logged Other Event Error Status Clear register is set:

- Ingress completion header error
- Memory read capture
- Write unsuccessful completion status
- Read unsuccessful completion status
- Configuration request retry completion status

Note | For the rest of OE errors or where there is no OE errors, the Receive Other Event Header2 Log register contains no valid header information, and software should ignore the content of this register.

TABLE 21-116 PEU Receive Other Event Header2 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0_2	RW	hdr{63:56} – Header byte 8 hdr{55:48} – Header byte 9 hdr{47:40} – Header byte A hdr{39:32} – Header byte B hdr{31:24} – Header byte C or undefined hdr{23:16} – Header byte D or undefined hdr{15:8} – Header byte E or undefined hdr{7:0} – Header byte F or Undefined

21.4.3.93 PEU Transmit Other Event Header1 Log Register

(00681038_{16} / 0_{16})

The PEU Transmit Other Event Header1 Log register holds the first eight bytes of the original transmitted request header when one of the following primary other events in the PEU Logged Other Event Error Status Clear register is set on a completion:

- Completion Timeout Error
- Write Unsuccessful Completion Status
- Read Unsuccessful Completion Status
- Configuration Request Retry Completion Status.

Note | For the rest of OE errors or where there is no OE errors, the Transmit Other Event Header1 Log register contains no valid header information, and software should ignore the content of this register.

Implementation Note | For warm reset protection, the value of PEU Transmit Other Event Header1/Header2 Log register is warm reset protected only when a primary error bit is set in the PEU Other Event Status Clear register.

TABLE 21-117 PEU Transmit Other Event Header1 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0_2	RW	hdr{63:56} – Header byte 0 hdr{55:48} – Header byte 1 hdr{47:40} – Header byte 2 hdr{39:32} – Header byte 3 hdr{31:24} – Header byte 4 hdr{23:16} – Header byte 5 hdr{15:8} – Header byte 6 hdr{7:0} – Header byte 7

21.4.3.94 PEU Transmit Other Event Header2 Log Register

(00681040_{16} / 0_{16})

The PEU Transmit Other Event Header2 Log register holds the second eight bytes of the original transmitted request header when one of the following primary other events in the PEU Logged Other Event Error Status Clear register is set on a completion:

- Completion timeout error
- Write unsuccessful completion status
- Read unsuccessful completion status

- Configuration request retry completion status

For the rest of OE errors or where there is no OE errors, the Transmit Other Event Header2 Log register contains no valid header information, and software should ignore the content of this register.

TABLE 21-118 PEU Transmit Other Event Header2 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0_2	RW	hdr{63:56} – Header byte 8 hdr{55:48} – Header byte 9 hdr{47:40} – Header byte A hdr{39:32} – Header byte B hdr{31:24} – Header byte C or undefined hdr{23:16} – Header byte D or undefined hdr{15:8} – Header byte E or undefined hdr{7:0} – Header byte F or Undefined

21.4.3.95 PEU Performance Counter Select Register (00682000_{16} / 0_{16})

The PEU Performance Counter Select register provides the selects for the performance counters.

TABLE 21-119 PEU Performance Counter Select Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:18	—				*Reserved*
17:16	sel2	rst_l	0_{16}	RW	Select for counter 2, as follows: 00_2 = None 01_{16} = Total nonposted completion time in unit of 64-cycle, used together with 02_{16} (completion received) is sel1 to figure out the average PIO completion latency. The deviation of PIO latency ranges from 0 to 128 cycles. 02_{16} = Transmitted DMA read completion data words. 03_{16} = Received DMA write data words.

TABLE 21-119 PEU Performance Counter Select Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
15:8	sel1	rst_l	0_{16}	RW	Select for counter 1, as follows: 00_{16} = None 01_{16} = Clock cycles in 250 MHz frequency 02_{16} = Completions 10_{16} = Transmit posted credit not available cycles 11_{16} = Transmit nonposted credit not available cycles 12_{16} = Transmit completion credit not available cycles 13_{16} = Transmit credits (any) not available cycles 14_{16} = Retry buffer credit not available cycles 20_{16} = Received memory read packets 21_{16} = Received memory write packets 22_{16} = Receive credits below threshold cycles 23_{16} = Receive posted header credits exhausted cycles 24_{16} = Receive posted data credits below MPS threshold cycles 25_{16} = Receive nonposted header credits exhausted cycles 30_{16} = L0s cycles (receiver) 31_{16} = L0s transitions (receiver) 32_{16} = L0s cycles (transmitter) 33_{16} = L0s transitions (transmitter) 40_{16} = Receiver error 42_{16} = Bad TLP 43_{16} = Bad DLLP 44_{16} = REPLAY_NUM rollover 47_{16} = Replay timeout
7:0	sel0	rst_l	0_{16}	RW	Select for counter 0: Values are the same as counter 1.

21.4.3.96 PEU Performance Counter Zero Register (00682008_{16} / 0_{16})

The PEU Performance Counter Zero register contains the current count of events selected by the sel0 field in the PEU Performance Counter Control register.

TABLE 21-120 PEU Performance Counter Zero Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	cnt	rst_l	0_{16}	RW	Counter.

21.4.3.97 PEU Performance Counter One Register (00682010_{16} / 0_{16})

The PEU Performance Counter One register contains the current count of events selected by the sel1 field in the PEU Performance Counter Control register.

TABLE 21-121 PEU Performance Counter One Register

Field	Bits	Reset Name	Reset Value	R/W	Description
cnt	63:0	rst_l	0_{16}	RW	Counter

21.4.3.98 PEU Performance Counter Two Register (00682018_{16} / 0_{16})

The PEU Performance Counter Two register contains the current count of data selected by the sel2 field in the PEU Performance Counter Control register.

TABLE 21-122 PEU Performance Counter Two Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:0	cnt	rst_l	0_{16}	RW	Counter

21.4.3.99 PEU Debug Select A Register (00683000_{16} / 0_{16})

The Debug Select A register selects the output on debug port 0.

TABLE 21-123 PEU Debug Select a Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:9	—				*Reserved*
8:6	block	rst_l	0_{16}	RW	Block select in core, as follows: 000_2 – Constant zero 001_2 – Training sequence Selection 010_2 – ETL block 011_2 – ITL block 100_2 – PMC block 101_2 – RSB block 110_2 – CTB block 111_2 – CXPL core
5:3	module	rst_l	0_{16}	RW	Module select in block
2:0	signal	rst_l	0_{16}	RW	Signal select in sub-block

21.4.3.100PEU Debug Select B Register (00683008_{16} / 0_{16})

The Debug Select B register selects the output on debug port 1.

TABLE 21-124 PEU Debug Select B Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:9	—				*Reserved*
8:6	block	rst_l	0_{16}	RW	Block select in core, as follows: 000 – Constant zero 001 – Training sequence selection 010 – ETL block 011 – ITL block 100 – PMC block 101 – RSM block 110 – CTB block 111 – CXPL core
5:3	module	rst_l	0_{16}	RW	Module select in block.
2:0	signal	rst_l	0_{16}	RW	Signal select in module.

21.4.3.101PEU Device Capabilities Register (00690000_{16} / 2_{16})

The PEU Device Capabilities register identifies PCI Express device-specific capabilities.

.

TABLE 21-125 PEU Device Capabilities Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:12	—				*Reserved*
11:9	l1	rst_l	0_{16}	RO	Endpoint L1 acceptable latency. This field indicates the acceptable latency that can be withstood due to the transition from L1 state to L0 state. It is essentially an indirect measure of the internal buffering. This field must be 000_2 for root complex.
8:6	l0s	rst_l	0_{16}	RO	Endpoint L0s acceptable latency. This field indicates the acceptable latency that can be withstood due to the transition from L0s state to L0 state. It is essentially an indirect measure of the internal buffering. This field must be 000_2 for root complex.
5:3	—				*Reserved*
2:0	mps	rst_l	2_{16}	RO	Maximum payload size supported. This field indicates the maximum payload size that can be supported for TLPs. 010_2 = 512 bytes

21.4.3.102PEU Device Control Register (00690008_{16} / 0_{16})

The Device Control register controls PCI Express device-specific parameters.

TABLE 21-126 PEU Device Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:15	—				*Reserved*
14:12	mrrs	rst_l	0_{16}	RO	Maximum read request size. This field sets the maximum read request size for the device as a requester. Since the maximum read request size which can be issued is 16 bytes, the value is 000_2.
11:8	—				*Reserved*
7:5	mps	rst_l	0_{16}	RW	Maximum payload size. This field sets maximum TLP payload size for the device. As a receiver, the device must handle TLPs as large as the set value; as transmitter, the device must not generate TLPs exceeding the set value. 000_2 – 128 bytes 001_2 – 256 bytes 010_2 – 512 bytes 011_2 – 1024 bytes (not supported) 100_2 – 2048 bytes (not supported) 101_2 – 4096 bytes (not supported) 110_2 – *Reserved* 111_2 – *Reserved*
4:0	—				*Reserved*

21.4.3.103PEU Device Status Register (00690010_{16} / 0_{16})

The Device Control register provides information about PCI Express device-specific parameters.

TABLE 21-127 PEU Device Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:6	—				*Reserved*
5	tp	rst_l	0_{16}	RO	Transactions pending. This bit when set indicates that a port has issued nonposted requests that have not been completed.
4:0	—				*Reserved*

21.4.3.104PEU Link Capabilities Register (00690018_{16} / $15C81_{16}$)

The Link Capabilities register identifies PCI Express link-specific capabilities.

TABLE 21-128 PEU Link Capabilities Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	port	rst_l	0_{16}	RO	Port Number

TABLE 21-128 PEU Link Capabilities Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
23:18	—				*Reserved*
17:15	l1	rst_l	2_{16}	RO	L1 exit latency. 010_2 – 2 µs to less than 4 µs
14:12	l0s	rst_l	4_{16}	RO	L0s exit latency. 100_2 – 512 ns to less than 1 µs
11:10	aspm	rst_l	3_{16}	RO	Active state power management support, as follows: 00_2 – *Reserved* 01_2 – L0s entry supported 10_2 – *Reserved* 11_2 – L0s and L1 supported
9:4	width	rst_l	8_{16}	RO	Maximum link width. 001000_2 – x8
3:0	speed	rst_l	1_{16}	RO	Maximum link speed. 0001_2 – 2.5Gb/s

21.4.3.105PEU Link Control Register (00690020_{16} / 0_{16})

The Link Control register controls PCI Express link-specific parameters.

TABLE 21-129 PEU Link Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:8	—				*Reserved*
7	extsync	rst_l	0_{16}	RW	Extended sync. This bit when set forces the transmission of 4096 FTS ordered sets in the L0s state followed by a single SKP-ordered set prior to entering the L0 state, and the transmission of 1024 TS1 ordered sets in the L1 state prior to entering the recovery state.
6	clock	rst_l	0_{16}	RW	Common clock configuration. This bit when set indicates that this component and the component at the opposite end of this link are operating with a distributed common reference clock.
5	retrain	rst_l	0_{16}	RW	Retrain link. A write of 1 to this bit initiates link retraining by directing the physical layer LTSSM to the recovery state
4	disable	rst_l	0_{16}	RW	Link disable. This bit disables the link when set to 1.
3	rcb	rst_l	0_{16}	RO	Read completion boundary. 0 = 64 byte.
2	—				*Reserved*
1:0	aspm	rst_l	0_{16}	RW	Active state link PM Control. as follows: 00_2 Disabled 01_2 L0s entry enabled 10_2 L1 entry enabled 11_2 L0s and L1 entry supported

21.4.3.106PEU Link Status Register (00690028_{16} / 0_{16})

The Link Status register provides information about PCI Express link-specific parameters.

TABLE 21-130 PEU Link Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:13	—				*Reserved*
12	clock	rst_l	0_{16}	RO	Slot clock configuration. This bit indicates that the component uses the same physical reference clock that the platform provides.
11	train	rst_l	0_{16}	RO	Link Training. This indicates that link training is in progress (physical layer LTTSM in configuration or recovery state) or that 1 was written to the retrain link bit but link training has not yet begun.
10	spare	rst_l	0_{16}	RO	Spare
9:4	width	rst_l	0_{16}	RO	Negotiated link width, as follows: 000000_2 – link down 000001_2 – x1 000010_2 – x2 000100_2 – x4 001000_2 – x8 001100_2 – x12 (not supported) 010000_2 – x16 (not supported) 100000_2 – x32 (not supported)
3:0	speed	rst_l	0_{16}	RO	Link speed, as follows: 0000_2 – link down 0001_2 – 2.5 Gb/s

21.4.3.107PEU Slot Capabilities Register (00690030_{16} / 0_{16})

The Slot Capabilities register identifies PCI Express slot-specific capabilities. Any write to this register will cause the Set_Slot_Power_Limit message to be transmitted on the link.

TABLE 21-131 PEU Slot Capabilities Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:17	—				*Reserved*
16:15	spls	rst_l	0_{16}	RW	Set power limit scale. Specifies the scale used for the slot power limit value. 00_2 –1.0x 01_2 – 0.1x 10_2 – 0.01x 11_2 – 0.001x
14:7	splv	rst_l	0_{16}	RW	Set power limit value. In combination with the value of slot power limit scale, specifies the upper limit on power supplied by slot. Power limit (watts) calculated by multiplying the value in this field by the value in the slot power limit scale field.
6:0	—				*Reserved*

21.4.3.108PEU Uncorrectable Error Log Enable Register

(00691000_{16}/ $17F011_{16}$)

The PEU Uncorrectable Error Log Enable register controls which events will be logged in the PEU Uncorrectable Error Status Clear register.

TABLE 21-132 PEU Uncorrectable Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:21	—				*Reserved*
20:0	en	por_l	$17F011_{16}$	RW	Log enables.

21.4.3.109PEU Uncorrectable Error Interrupt Enable Register

(00691008_{16} / 0_{16})

The PEU Uncorrectable Error Interrupt Enable register controls which events logged in the PEU Uncorrectable Error Status Clear register will generate an interrupt.

TABLE 21-133 PEU Uncorrectable Error Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52:32	en_s	rst_l	0_{16}	RW	Secondary interrupt enables.
31:21	—				*Reserved*
20:0	en_p	rst_l	0_{16}	RW	Primary interrupt enables.

21.4.3.110PEU Uncorrectable Error Interrupt Status Register

(00691010_{16} / 0_{16})

The PEU Uncorrectable Error Interrupt Status register indicates which events currently logged in the PEU Uncorrectable Error Status Clear register are enabled to generate interrupts by the PEU Uncorrectable Error Interrupt Enable register.

TABLE 21-134 PEU Uncorrectable Error Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52:32	err_s	rst_l	0_{16}	R	Secondary interrupt status.
31:21	—				*Reserved*
20:0	err_p	rst_l	0_{16}	R	Primary interrupt status.

21.4.3.111PEU Uncorrectable Error Status Clear Register

(00691018_{16} / 0_{16})

The PEU Uncorrectable Error Status Clear register indicates that an uncorrectable error occurred. Primary errors include data logged with these errors. Secondary events have no data logged with them. All errors belong to the same error group, meaning that any primary error logged will cause any additional error to be logged as secondary.

TABLE 21-135 PEU Uncorrectable Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52	ur_s	por_l	0_{16}	RW1C	Unsupported request secondary error.
51	—				*Reserved*
50	mfp_s	por_l	0_{16}	RW1C	Malformed TLP secondary error.
49	rof_s	por_l	0_{16}	RW1C	Receiver overflow secondary error.
48	uc_s	por_l	0_{16}	RW1C	Unexpected completion secondary.
47	spare	por_l	0_{16}	RW1C	Spare bit.
46	cto_s	por_l	0_{16}	RW1C	Completion timeout secondary error.
45	fcp_s	por_l	0_{16}	RW1C	Flow control protocol secondary error.
44	pp_s	por_l	0_{16}	RW1C	Poisoned TLP secondary error.
43:37	—				*Reserved*
36	dlp_s	por_l	0_{16}	RW1C	Data link protocol secondary error.
35:33	—				*Reserved*
32	spare	por_l	0_{16}	RW1C	Spare bit.
31:21	—				*Reserved*
20	ur_p	por_l	0_{16}	RW1C	Unsupported request primary error.
19	—				*Reserved*
18	mfp_p	por_l	0_{16}	RW1C	Malformed TLP primary error.
17	rof_p	por_l	0_{16}	RW1C	Receiver overflow primary error.
16	uc_p	por_l	0_{16}	RW1C	Unexpected completion primary.
15	spare	por_l	0_{16}	RW1C	Spare bit.
14	cto_p	por_l	0_{16}	RW1C	Completion timeout primary error.
13	fcp_p	por_l	0_{16}	RW1C	Flow control protocol primary error.
12	pp_p	por_l	0_{16}	RW1C	Poisoned TLP primary error.
11:5	—				*Reserved*
4	dlp_p	por_l	0_{16}	RW1C	Data link protocol primary error.
3:1	—				*Reserved*
0	spare	por_l	0_{16}	RW1C	Spare bit.

21.4.3.112PEU Uncorrectable Error Status Set Register (00691020_{16} / 0_{16})

The Uncorrectable Error Status Set register controls setting errors in the PEU Uncorrectable Error Status Clear register for testing.

TABLE 21-136 PEU Uncorrectable Error Status Set Register

Bits	Field	Reset Name	Reset Value	R/W	Description
63:53	—				*Reserved*
52	ur_s	por_l	0_{16}	RW1S	Unsupported request secondary error.
51	—				*Reserved*
50	mfp_s	por_l	0_{16}	RW1S	Malformed TLP secondary error.
49	rof_s	por_l	0_{16}	RW1S	Receiver overflow secondary error.
48	uc_s	por_l	0_{16}	RW1S	Unexpected completion secondary.
47	spare	por_l	0_{16}	RW1S	Spare bit.
46	cto_s	por_l	0_{16}	RW1S	Completion timeout secondary error.
45	fcp_s	por_l	0_{16}	RW1S	Flow control protocol secondary error.
44	pp_s	por_l	0_{16}	RW1S	Poisoned TLP secondary error.
43:37	—				*Reserved*
36	dlp_s	por_l	0_{16}	RW1S	Data link protocol secondary error.
35:33	—				*Reserved*
32	spare	por_l	0_{16}	RW1S	Spare bit.
31:21	—				*Reserved*
20	ur_p	por_l	0_{16}	RW1S	Unsupported request primary error.
19	—				*Reserved*
18	mfp_p	por_l	0_{16}	RW1S	Malformed TLP primary error.
17	rof_p	por_l	0_{16}	RW1S	Receiver overflow primary error.
16	uc_p	por_l	0_{16}	RW1S	Unexpected completion primary.
15	spare	por_l	0_{16}	RW1S	Spare bit.
14	cto_p	por_l	0_{16}	RW1S	Completion timeout primary error.
13	fcp_p	por_l	0_{16}	RW1S	Flow control protocol primary error.
12	pp_p	por_l	0_{16}	RW1S	Poisoned TLP primary error.
11:5	—				*Reserved*
4	dlp_p	por_l	0_{16}	RW1S	Data link protocol primary error.
3:1	—				*Reserved*
0	spare	por_l	0_{16}	RW1S	Spare bit.

21.4.3.113PEU Receive Uncorrectable Error Header1 Log Register (00691028_{16} / 0_{16})

The PEU Receive Uncorrectable Error Header1 Log register holds the first eight bytes of the received header when a primary other event in the PEU Uncorrectable Error Status Clear register is set.

Implementation Note For warm reset protection, the value of PEU Receive Uncorrectable Error Header1/Header2 Log register is warm reset protected only when a primary error bit is set in the PEU Uncorrectable Error Status Clear register.

.

TABLE 21-137 PEU Receive Uncorrectable Error Header1 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0	RW	hdr{63:56} – Header byte 0 hdr{55:48} – Header byte 1 hdr{47:40} – Header byte 2 hdr{39:32} – Header byte 3 hdr{31:24} – Header byte 4 hdr{23:16} – Header byte 5 hdr{15:8 – Header byte 6 hdr{7:0} – Header byte 7

21.4.3.114PEU Receive Uncorrectable Error Header2 Log Register (00691030_{16} / 0_{16})

The PEU Receive Uncorrectable Error Header2 Log register holds the second eight bytes of the received header when a primary other event in the PEU Uncorrectable Error Status Clear register is set.

TABLE 21-138 PEU Receive Uncorrectable Error Header2 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0	RW	hdr{63:56} – Header byte 8 hdr{55:48} – Header byte 9 hdr{47:40} – Header byte A hdr{39:32} – Header byte B hdr{31:24} – Header byte C or undefined hdr{23:16} – Header byte D or undefined hdr{15:8} – Header byte E or undefined hdr{7:0} – Header byte F or undefined

21.4.3.115PEU Transmit Uncorrectable Error Header1 Log Register (00691038_{16} / 0_{16})

The PEU Transmit Uncorrectable Error Header1 Log register holds the first eight bytes of the original transmitted request header when a primary other event in the PEU Uncorrectable Error Status Clear Register is set on a completion.

Implementation Note For warm reset protection, the value of PEU Transmit Uncorrectable Error Header1/Header2 Log register is warm reset protected only when a primary error bit is set in the PEU Uncorrectable Error Status Clear register.

TABLE 21-139 PEU Transmit Uncorrectable Error Header1 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0	RW	hdr{63:56} – Header byte 0 hdr{55:48} – Header byte 1 hdr{47:40} – Header byte 2 hdr{39:32} – Header byte 3 hdr{31:24} – Header byte 4 hdr{23:16} – Header byte 5 hdr{15:8} – Header byte 6 hdr{7:0} – Header byte 7

21.4.3.116PEU Transmit Uncorrectable Error Header2 Log Register (00691040_{16} / 0_{16})

The PEU Transmit Uncorrectable Error Header2 Log register holds the second eight bytes of the original transmitted request header when a primary other event in the PEU Uncorrectable Error Status Clear register is set on a completion.

TABLE 21-140 PEU Transmit Uncorrectable Error Header2 Log Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:0	hdr	por_l	0	RW	hdr{63:56} – Header byte 8 hdr{55:48} – Header byte 9 hdr{47:40} – Header byte A hdr{39:32} – Header byte B hdr{31:24} – Header byte C or undefined hdr{23:16} – Header byte D or undefined hdr{15:8} – Header byte E or undefined hdr{7:0} – Header byte F or undefined

21.4.3.117PEU Correctable Error Log Enable Register ($006A1000_{16}$ / $11C1_{16}$)

The PEU Correctable Error Log Enable register controls which events will be logged in the PEU Correctable Error Status Clear register

TABLE 21-141 PEU Correctable Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:13	—				*Reserved*
12:0	en	por_l	$11C1_{16}$	RW	Log enables.

21.4.3.118PEU Correctable Error Interrupt Enable Register ($006A1008_{16}$ / 0_{16})

The PEU Correctable Error Interrupt Enable Register controls which events logged in the PEU Correctable Error Status Clear Register will generate an interrupt.

TABLE 21-142 PEU Correctable Error Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:45	—				*Reserved*
44:32	en_s	rst_l	0_{16}	RW	Secondary interrupt enables.
31:13	—				*Reserved*
12:0	en_p	rst_l	0_{16}	RW	Primary interrupt enables.

21.4.3.119PEU Correctable Error Interrupt Status Register ($006A1010_{16}$ / 0_{16})

The PEU Correctable Error Interrupt Status Register indicates which events currently logged in the PEU Correctable Error Status Clear Register are enabled to generate interrupts by the PEU Correctable Error Interrupt Enable Register

TABLE 21-143 PEU Correctable Error Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:45	—				*Reserved*
44:32	err_s	rst_l	0_{16}	RO	Secondary interrupt status
31:13	—				*Reserved*
12:0	err_p	rst_l	0_{16}	RO	Primary interrupt status

21.4.3.120PEU Correctable Error Status Clear Register ($006A1018_{16}$ / 0_{16})

The PEU Correctable Error Status Clear register indicates that a correctable error occurred. All errors belong to the same error group, meaning that any primary error logged will cause any additional error to get logged as secondary.

TABLE 21-144 PEU Correctable Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:45	—				*Reserved*
44	rto_s	por_l	0_{16}	RW1C	Replay timeout secondary error.
43:41	—				*Reserved*
40	rnr_s	por_l	0_{16}	RW1C	REPLAY_NUM rollover secondary error.
39	bdp_s	por_l	0_{16}	RW1C	Bad DLLP secondary error.
38	btp_s	por_l	0_{16}	RW1C	Bad TLP secondary error.
37:33	—				*Reserved*
32	re_s	por_l	0_{16}	RW1C	Receiver secondary error.
31:13	—				*Reserved*
12	rto_p	por_l	0_{16}	RW1C	Replay timeout primary error.
11:9	—				*Reserved*
8	rnr_p	por_l	0_{16}	RW1C	REPLAY_NUM rollover primary error.
7	bdp_p	por_l	0_{16}	RW1C	Bad DLLP primary error.
6	btp_p	por_l	0_{16}	RW1C	Bad TLP primary error.
5:1	—				*Reserved*
0	re_p	por_l	0_{16}	RW1C	Receiver primary error.

21.4.3.121PEU Correctable Error Status Set Register ($006A1020_{16}$ / 0_{16})

The PEU Correctable Error Status Set register controls setting errors in the PEU Correctable Error Status Clear register for testing.

TABLE 21-145 PEU Correctable Error Status Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:45	—				*Reserved*
44	rto_s	por_l	0_{16}	RW1S	Replay timeout secondary error.
43:41	—				*Reserved*
40	rnr_s	por_l	0_{16}	RW1S	REPLAY_NUM rollover secondary error.
39	bdp_s	por_l	0_{16}	RW1S	Bad DLLP secondary error.
38	btp_s	por_l	0_{16}	RW1S	Bad TLP secondary error.

TABLE 21-145 PEU Correctable Error Status Set Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
37:33	—				*Reserved*
32	re_s	por_l	0_{16}	RW1S	Receiver secondary error.
31:13	—				*Reserved*
12	rto_p	por_l	0_{16}	RW1S	Replay timeout primary error.
11:9	—				*Reserved*
8	rnr_p	por_l	0_{16}	RW1S	REPLAY_NUM rollover primary error.
7	bdp_p	por_l	0_{16}	RW1S	Bad DLLP primary error.
6	btp_p	por_l	0_{16}	RW1S	Bad TLP primary error.
5:1	—				*Reserved*
0	re_p	por_l	0_{16}	RW1S	Receiver primary error.

21.4.3.122PEU CXPL/SerDes Revision Register ($006E2000_{16}$ / $0)_{16}$

The PEU CXPL/SerDes Revision register contains information on the revision of the IP from an external vendor. The value will be hard-coded. It is put at 0_{16} temporarily.

TABLE 21-146 PEU CXPL/SerDes Revision Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:8	—				*Reserved*
7:4	cxpl_id	por_l	0	RO	CXPL core revision ID. Hard-coded.
3:0	serdes_id	por_l	0	RO	SerDes macro revision ID. Hard-coded.

21.4.3.123PEU CXPL DLL AckNak Latency Threshold Register ($006E2008_{16}$ / 43_{16})

The PEU CXPL Data Link Layer AckNak Latency Threshold register is from the alt field of "ACK Latency Timer and Replay Timer Register" in the Cascade CDM document.

TABLE 21-147 PEU CXPL DLL AckNak Latency Thresholds Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:0	ack_nak_thr	rst_l	43_{16}	RW	Ack Nak latency timer threshold, in terms of pcl2clk. This value is programmed by configured link width and max payload size.

21.4.3.124PEU CXPL DLL AckNak Latency Timer Register ($006E2010_{16}$ / 00_{16})

The PEU CXPL Data Link Layer AckNak Latency Timer register does not exist in the Cascade CDM

TABLE 21-148 PEU CXPL DLL AckNak Latency Timer

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:0	ack_nak_tmr	rst_l	0_{16}	RO	Ack Nak latency timer. This is a count-up register. Ack Nak latency timeout happens when the value in this register is equal to the value in AclNak Latency Threshold register.

21.4.3.125PEU CXPL DLL Replay Timer Threshold Register ($006E2018_{16}$ / FC_{16})

The PEU CXPL Data Link Layers register is from the rt field of "ACK Latency Timer and Replay Timer Register" in the Cascade CDM document.

TABLE 21-149 PEU CXPL DLL Replay Timer Threshold Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:0	rplay_tmr_thr	rst_l	FC_{16}	RW	Replay timer threshold. This value is programmed by configured link width and max payload size.

21.4.3.126PEU CXPL DLL Replay Timer Register ($006E2020_{16}$ / 00_{16})

The PEU CXPL Data Link Layer Replay Timer register does not exist in the Cascade CDM.

TABLE 21-150 PEU CXPL DLL Replay Timer Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:18	—				*Reserved*
17:16	replay_num	rst_l	0_{16}	RO	Replay number. The number of replays that have taken place
15:0	rplay_tmr	rst_l	0_{16}	RO	Replay timer. This is a count-up register. Replay timeout happens when the value in this register is equal to the value in Replay Timer Threshold register.

21.4.3.127PEU CXPL DLL Vendor DLLP Message Register

($006E2040_{16}$ / 00_{16})

The PEU CXPL Data Link Layer Vendor DLLP Message register is from the "Other Message Register" in the Cascade CDM document.

TABLE 21-151 PEU CXPL DLL Vendor DLLP Message Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:0	v_message	rst_l	0_{16}	RW	This register may be used to send a specific PCI-E DLLP message. The user writes the payload of the message it wants to send into this register, then sets other_message_request bit of PEU CXPL DLL Control register.

21.4.3.128PEU CXPL LTSSM Control Register ($006E2050_{16}$ / 00_{16})

The PEU CXPL Link Training Status State Machine register is from the lfs and fl fields of the "Port Force Link Register" in the Cascade CDM document.

TABLE 21-152 PEU CXPL LTSSM Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:9	—				*Reserved*
8	force_en	rst_l	0	RW	The force ltssm bit is enabled when it is written by 1. The force link pulse will force the ltssm into a specific state.
7:5	—				*Reserved*

TABLE 21-152 PEU CXPL LTSSM Control Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
4:0	forced_ltssm	rst_l	0_{16}	RW	CXPL LTSSM state will be forced to one of following state values when bit 15 is set. The following are the state encodings: `POLL_ACTIVE` = 02_{16} `POLL_COMPLIANCE` = 03_{16} `POLL_CONFIG` = 04_{16} `PRE_DETECT_QUIET` = 05_{16} `DETECT_WAIT` = 06_{16} `CFG_LINKWD_START` = 07_{16} `CFG_LINKWD_ACEPT` = 08_{16} `CFG_LANENUM_WAIT`= 09_{16} `CFG_LANENUM_ACEPT` = $0A_{16}$ `CFG_COMPLETE` = $0B_{16}$ `CFG_IDLE` = $0C_{16}$ `RCVRY_LOCK` = $0D_{16}$ `RCVRY_RCVRCFG` = $0E_{16}$ `RCVRY_IDLE` = $0F_{16}$ `L0` = 10_{16} `L0S` = 11_{16} `L123_SEND_IDLE` = 12_{16} `L1_IDLE` = 13_{16} `L2_IDLE` = 14_{16} `L2_WAKE` = 15_{16} `DISABLED_ENTRY` = 16_{16} `DISABLED_IDLE` = 17_{16} `DISABLED` = 18_{16} `LPBK_ENTRY` = 19_{16} `LPBK_ACTIVE` = $1A_{16}$ `LPBK_EXIT` = $1B_{16}$ `LPBK_EXIT_TIMEOUT` = $1C_{16}$ `HOT_RESET_ENTRY` = $1D_{16}$ `S_HOT_RESET` = $1F_{16}$

21.4.3.129PEU CXPL DLL Control Register ($006E2058_{16}$ / $00F900_{16}$)

The PEU CXPL Data Link Layer Control register controls CXPL data link behavior. The ack_freq field is from af field of the ACK frequency register. The other_message_request field is from the omr field of the Port link Control register. The ack_nack_disable is from the ad field of the Lane Skew register. The data_link_enable field is from the dle field of the Port Link Control register.

TABLE 21-153 PEU CXPL DLL Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—				*Reserved*
15:8	ack_freq (af)	rst_l	1_{16}	RW	ACK frequency, to allow up to 255 ACKs accumulated before send.
7:5	—				*Reserved*
4	flow_control_ disable	rst_l	0_{16}	RW	Flow control disable.
3	spare	rst_l	0_{16}	RW	SPARE bit
2	other_message_ request (omr)	rst_l	0_{16}	RW	Other message request when 1 is written.
1	ack_nack_disable (ad)	rst_l	0_{16}	RW	ACK/NACK disable.
0	data_link_enable (dle)	rst_l	1_{16}	RW	Data link layer enable. Software can hold the DLCMSM (data link control and management state machine) in the `delineative` state by setting this bit to 0.

21.4.3.130PEU CXPL MACL/PCS Control Register ($006E2060_{16}$ / 40000_{16})

The PEU CXPL Media Access Layer / Physical Coding Sublayer Control register is described in TABLE 21-154. .

TABLE 21-154 PEU CXPL MACL/PCS Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	link_num (ln)	rst_l	0_{16}	RW	Link number (0 to 255), advertised to the device at the remote link during link training.
23:16	n_fts (nfts)	rst_l	$1B_{16}$	RW	Number of NFTS. This is the N_FTS value which transmitted in the TS1/TS2 ordered sets by this port. This value is dependent on the bit/symbol lock time in SerDes.
15:14	—				*Reserved*
13:12	spare	rst_l	00_2	RW	Spare bits.
11:8	link_capable (lc)	rst_l	1000_2	RW	Link Mode Capable (1-hot): Bit 0 – x1 Bit 1 – x2 Bit 2 – x4 Bit 3 – x8
7:5	—				*Reserved*

TABLE 21-154 PEU CXPL MACL/PCS Control Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
4	fast_link_mode (flm)	rst_l	0_2	RW	Fast Link Mode. When asserted, all LTSSM timers and SerDes's receiver detection timers are set to fast mode for simulation purpose. LTSSM 2ms → 5 µs LTSSM 12ms → 30 µs LTSSM 24ms → 60 µs LTSSM 48 ms → 120 µs LTSSM 1024 TSs → 16 TSs. SerDes Tx charge timer 100 µs → 1us SerDes Tx discharge timer with no receiver 1 µs → 40 ns SerDes Tx discharge timer with receiver 30 µs → 400 ns
3	rx_high_imp_ dis	rst_l	0_2	RW	0 = SerDes's receiver channel is in high impedance mode (200K min). The 111_2 setting is taken effect for term field for all SerDes's receiver channel. 1 = The setting of term field in SerDes's Receiver Lane 0–7 Control Register is taken effect for SerDes's receiver channel.
2	elastical_buffer _disable	rst_l	0_2	RW	Elastic buffer disable.
1	scramble_ disable (sd)	rst_l	0_2	RW	Disable scrambling.
0	reset_assert (ra)	rst_l	0_2	RW1S	When this bit is asserted, it causes LTSSM to go to `Recovery` state, and eventually to `HOT_RESET` state. Software set this bit to issue the hot reset packet to the link. This field is reset by hardware when the LTSSM is in the `HOT_RESET` state.

21.4.3.131PEU CXPL MACL Lane Skew/Receiver Detection/ Bit Lock Timer Control Register ($006E2068_{16}$ / 00_{16})

PEU CXPL Medium Access Control Layer Lane Skew / Receiver Detection/ Bit Lock Timer Control register is described in TABLE 21-155.

TABLE 21-155 PEU CXPL MACL Lane Skew/Receiver Detection/Bit Lock Timer Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:26	—				*Reserved*
25	deskew_disable (dis)	rst_l	0_2	RW	De-skew disable, When asserted, disable lane skew logic is disabled on receive
24	spare	rst_l	0_2	RW	Spare bit.
23:16	bit_lock_timer_ threshold	rst_l	0_2	RW	The threshold value for the bit lock timer. This timer is used to delay the propagation of cma_aln_en bit of PEU SerDes's Receiver Lane 0–7 Control register to SerDes Macro. The cma_aln_en takes effect only when the bit lock timer reaches this threshold value.

TABLE 21-155 PEU CXPL MACL Lane Skew/Receiver Detection/Bit Lock Timer Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
15:9	spare	rst_l	0_2	RW	Spare.
8	force_rcv_ present_en	rst_l	0_2	RW	Force receiver present enable.
7	ln_7_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 7 0 = Force receiver not present at Physical Lane 7. When force_rcv_present_en = 0:, this bit is ignored.
6	ln_6_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1 1 = Force receiver present at Physical Lane 6 0 = Force receiver not present at Physical Lane 6. When force_rcv_present_en = 0, this bit is ignored.
5	ln_5_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 5 0 = Force receiver not present at Physical Lane 5. When force_rcv_present_en = 0, this bit is ignored.
4	ln_4_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 4 0 = Force receiver not present at Physical Lane 4. When force_rcv_present_en = 0, this bit is ignored.
3	ln_3_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 3 0 = Force receiver not present at Physical Lane 3. When force_rcv_present_en = 0, this bit is ignored.
2	ln_2_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 2 0 = Force receiver not present at Physical Lane 2. When force_rcv_present_en = 0, this bit is ignored.
1	ln_1_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 1 0 = Force receiver not present at Physical Lane 1. When force_rcv_present_en = 0, this bit is ignored.
0	ln_0_rcv_ present	rst_l	0_2	RW	When force_rcv_present_en = 1: 1 = Force receiver present at Physical Lane 0 0 = Force receiver not present at Physical Lane 0. When force_rcv_present_en = 0, this bit is ignored.

21.4.3.132PEU CXPL MACL Symbol Number Control Register

($006E2070_{16}$ / $33AA_{16}$)

All the fields of the PEU CXPL Medium Access Control Layer Symbol Number register are from Symbol Number register in the Cascade Document.

TABLE 21-156 PEU CXPL MACL Symbol Number Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:15	—				*Reserved*
14:12	spare	rst_l	011_2	RW	Spare bits.
11	—				*Reserved*
10:8	skip_symbols (ss)	rst_l	011_2	RW	Number of SKIP symbols in an SKIP ordered set.
7:4	spare	rst_l	A_{16}	RW	Spare bits
3:0	ts1_ts2_symbols (ts1/ts2)	rst_l	A_{16}	RW	Number of TS1/TS2 identifiers in a TS1 or TS2 ordered set.

21.4.3.133PEU CXPL MACL Symbol Timer Register ($006E2078_{16}$ / 3500_{16})

The PEU CXPL MACL Symbol Timer register is from Symbol Timer register in the Cascade Document.

TABLE 21-157 PEU CXPL MACL Symbol Timer Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:11	—				*Reserved*
10:0	skip_interval	rst_l	500_{16}	RW	SKIP interval value (between 1180 and 1538 symbol times).

21.4.3.134PEU CXPL Core Status Register ($006E2100_{16}$ / 0_{16})

The PEU CXPL Core Status register indicates all the status of the CXPL core.

TABLE 21-158 PEU CXPL Core Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:59	—				*Reserved*
58:56	tx_l0s_state	rst_l	0_{16}	RO	Transmit L0S state as follows: 000 – `S_L0S_XMT_ENTRY` 001 – `S_L0S_XMT_WAIT` 010 – `S_L0S_XMT_IDLE` 011 – `S_L0S_XMT_FTS` 100 – `S_L0S_XMT_SKIP` 101 – `S_L0S_XMT_EIDLE` 110 – `S_L0S_EXIT_WAIT`

TABLE 21-158 PEU CXPL Core Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
55:54	rv_l0s_state	rst_l	0_{16}	RO	Receiver L0S state: 00 – S_L0S_RCV_ENTRY; 01 – S_L0S_WAIT_IN_EIDLE 10 – S_L0S_RCV_IDLE 11 – S_L0S_RCV_FTS
53:52	int_fcsm_state	rst_l	0_{16}	RO	Init flow control state machine status: 00 – FC_IDLE 01 – FC_INIT1 11 – FC_INIT2 10 – FC_INIT_DONE
51:50	—				
49	rbuf_not_empty	rst_l	0_{16}	RO	Retry buffer not empty.
48:44	ltssm_state	rst_l	0_{16}	RO	LTSSM state.
43:36	rcv_polarity_rev	rst_l	0_{16}	RO	Receiver polarity reverse.
35:28	rcv_fts_num	rst_l	0_{16}	RO	Received FTS number.
27:20	rcv_link_num	rst_l	0_{16}	RO	Received link number.
19:12	pcs_lock_sts	rst_l	0_{16}	RO	Bit/byte synchronization status on Lane 7–0. Bit 19 is the status of Lane 7 and bit 12 is the status of Lane 0.
11:4	rcvr_detect_sts	rst_l	0_{16}	RO	Receiver detection status on Lane 7–0. Bit 11 is the status of Lane 7 and bit 4 is the status of Lane 0.
3	pcs_lane_rev	rst_l	0_{16}	RO	Lane reversal
2	pcs_align_sts	rst_l	0_{16}	RO	Align status, asserted when all enabled lanes are aligned.
1	sds_resdy_1	rst_l	0_{16}	RO	SerDes 1 ready (Lane 4–Lane 7).
0	sds_ready_0	rst_l	0_{16}	RO	SerDes 0 ready (Lane 0–Lane 3).

21.4.3.135PEU CXPL Event/Error Log Enable Register ($006E2108_{16}$)

The PEU CXPL Event/Error Log Enable register controls which events will be logged in the PEU CXPL Event/Error Status Clear register.

TABLE 21-159 PEU CXPL Event/Error Log Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	en_event	por_l	F_{16}	RW	Log enables for event.
23:18	—				*Reserved*
17:0	en_error	por_l	$3FFFF_{16}$	RW	Log enables for error.

21.4.3.136PEU CXPL Event/Error Interrupt Enable Register

$(006E2110_{16} / 0)_{16}$

The PEU CXPL Event/Error Interrupt Enable register controls which events logged in the PEU CXPL Event/Error Status Clear register will generate an interrupt.

TABLE 21-160 PEU CXPL Event/Error Interrupt Enable Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	en_event	rst_l	0_{16}	RW	Interrupt enables for event.
23:18	—				*Reserved*
17:0	en_error	rst_l	0_{16}	RW	Interrupt enables for error.

21.4.3.137PEU CXPL Event/Error Interrupt Status Register

$(006E2118_1 / 0_1)$

The PEU CXPL Event/Error Interrupt Status register indicates which events or errors currently logged in the PEU CXPL Event/Error Status Clear register are enabled to generate interrupts by the PEU CXPL Event/Error Interrupt Enable register.

TABLE 21-161 PEU CXPL Event/Error Interrupt Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31:24	en_event	rst_l	0_{16}	RO	Interrupt status for event.
23:18	—				*Reserved*
17:0	en_error	rst_l	0_{16}	R	Interrupt status for error.

21.4.3.138PEU CXPL Event/Error Status Clear Register

$(006E2120_{16} / 0_{16})$

The PEU CXPL Event/Error Status Clear register indicates that a CXPL event or error occurred.

TABLE 21-162 PEU CXPL Event/Error Status Clear Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*
31	evt_rcv_en_lb	por_l	0_{16}	RW1C	Received enable loopback.

TABLE 21-162 PEU CXPL Event/Error Status Clear Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
30	evt_rcv_dis_link	por_l	0_{16}	RW1C	Received disable link.
29	evt_rcv_hot_rst	por_l	0_{16}	RW1C	Received hot reset.
28	evt_rcv_eidle_exit	por_l	0_{16}	RW1C	Received electrical idle exit.
27	evt_rcv_eidle	por_l	0_{16}	RW1C	Received electrical idle.
26	evt_rcv_ts1	por_l	0_{16}	RW1C	Received TS1 order set.
25	evt_rcv_ts2	por_l	0_{16}	RW1C	Received TS2 order set.
24	evt_send_skp_b2b	por_l	0_{16}	RW1C	Two back-to-back skips have been sent.
23:18	—				*Reserved*
17	err_outstanding_skip	por_l	0_{16}	RW1C	Number of outstanding skips greater than 6.
16	err_elastic_fifo_undrflw	por_l	0_{16}	RW1C	Elastic FIFO underflow.
15	err_elstc_fifo_ovrflw	por_l	0_{16}	RW1C	Elastic FIFO overflow.
14	err_align	por_l	0_{16}	RW1C	Alignment error.
13	err_kchar_dllp_tlp	por_l	0_{16}	RW1C	KChar in DLLP/TLP error.
12	err_ill_end_pos	por_l	0_{16}	RW1C	Illegal END position error.
11	err_sync	por_l	0_{16}	RW1C	Lost bit/byte synchronization.
10	err_end_no_stp_sdp	por_l	0_{16}	RW1C	END without STP or SDP.
9	err_sdp_no_end	por_l	0_{16}	RW1C	SDP without END.
8	err_stp_no_end_edb	por_l	0_{16}	RW1C	STP without END or EDB.
7	err_ill_pad_pos	por_l	0_{16}	RW1C	Illegal PAD position.
6	err_multi_sdp	por_l	0_{16}	RW1C	Multiple SDPs.
5	err_multi_stp	por_l	0_{16}	RW1C	Multiple STPs.
4	err_ill_sdp_pos	por_l	0_{16}	RW1C	Illegal SDP position.
3	err_ill_stp_pos	por_l	0_{16}	RW1C	Illegal STP position.
2	err_unsup_dllp	por_l	0_{16}	RW1C	Unsupported DLLP.
1	err_src_tlp	por_l	0_{16}	RW1C	Source error TLP, receiving error TLP with inverted CRC and EDB.
0	err_sds_los	por_l	0_{16}	RW1C	SerDes loss signal — surprised removal.

21.4.3.139PEU CXPL Event/Error Status Set Register ($006E2128_{16}$ / 0_{16})

The PEU CXPL Event/Error Status Set register controls setting errors in the PEU CXPL Event/Error Status Clear register for testing.

TABLE 21-163 PEU CXPL Event/Error Status Set Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:32	—				*Reserved*

TABLE 21-163 PEU CXPL Event/Error Status Set Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
31	evt_rcv_en_lb	por_l	0_{16}	RW1S	Received enable loopback.
30	evt_rcv_dis_link	por_l	0_{16}	RW1S	Received disable link.
29	evt_rcv_hot_rst	por_l	0_{16}	RW1S	Received hot reset.
28	evt_rcv_eidle_exit	por_l	0_{16}	RW1S	Received electrical idle exit.
27	evt_rcv_eidle	por_l	0_{16}	RW1S	Received electrical idle.
26	evt_rcv_ts1	por_l	0_{16}	RW1S	Received TS1 order set.
25	evt_rcv_ts2	por_l	0_{16}	RW1S	Received TS2 order set.
24	evt_send_skp_b2b	por_l	0_{16}	RW1S	Two back to back skips have been sent.
23:18	—				*Reserved*
17	err_outstanding_skip	por_l	0_{16}	RW1S	Number of outstanding SKIPs greater than 6.
16	err_elastic_fifo_undrflw	por_l	0_{16}	RW1S	Elastic FIFO underflow.
15	err_elstc_fifo_ovrflw	por_l	0_{16}	RW1S	Elastic FIFO overflow.
14	err_align	por_l	0_{16}	RW1S	Alignment error.
13	err_kchar_dllp_tlp	por_l	0_{16}	RW1S	KChar in DLLP/TLP error.
12	err_ill_end_pos	por_l	0_{16}	RW1S	Illegal END position error.
11	err_sync	por_l	0_{16}	RW1S	Lost bit/byte synchronization.
10	err_end_edb_no_stp_sdp	por_l	0_{16}	RW1S	END or EDB without STP or SDP.
9	err_sdp_no_end	por_l	0_{16}	RW1S	SDP without END.
8	err_stp_no_end_edb	por_l	0_{16}	RW1S	STP without END or EDB.
7	err_ill_pad_pos	por_l	0_{16}	RW1S	Illegal PAD position.
6	err_multi_sdp	por_l	0_{16}	RW1S	Multiple SDPs.
5	err_multi_stp	por_l	0_{16}	RW1S	Multiple STPs.
4	err_ill_sdp_pos	por_l	0_{16}	RW1S	Illegal SDP position.
3	err_ill_stp_pos	por_l	0_{16}	RW1S	Illegal STP position.
2	err_unsup_dllp	por_l	0_{16}	RW1S	Unsupported DLLP.
1	err_src_tlp	por_l	0_{16}	RW1S	Source Error TLP, receiving error TLP with inverted CRC and EDB.
0	err_sds_los	por_l	0_{16}	RW1S	SerDes loss signal – surprised removal.

21.4.3.140PEU Link Bit Error Counter I Register ($006E2130_{16}$ /0_{16})

The PEU Link Bit Error Counter I register counts the total bit errors during the counter enable period. All the counters cannot roll over.

TABLE 21-164 PEU Link Bit Error Counter I Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63	ber_count_en	rst_l	0_2	RW	Enable *all* the bit error counters in both PEU link bit error counter I and PEU link bit error counter II registers.
62	ber_count_clr	rst_l	0_2	RW1S	Clear the bit error counters in both PEU link bit error counter I and PEU link bit error counter II registers. This bit will be cleared by hardware after all the bit error counters are reset to zero.
61:32	—				*Reserved*
31:24	cnt_bad_dllp	rst_l	0_2	RO	Number of DLLP LCRC errors (should not recount any error that has been reported in the Physical Layer).
23:16	cnt_bad_tlp	rst_l	0_2	RO	Number of TLP LCRC error s (should not recount any error that has been reported in the Physical Layer).
15:10	—				*Reserved*
9:0	cnt_pre	rst_l	0_2	RO	Number of physical receiver errors. (recount all cv violations/per lane).

21.4.3.141PEU Link Bit Error Counter II Register ($006E2138_{16}$ / 0_{16})

The PEU Link Bit Error Rate Counter II register counts the total 8-bit or 10-bit code encoding violations or 8-bit or 10-bit code disparity error per active lane during the bit error counter enable period All the counters cannot roll over.
.

TABLE 21-165 PEU Link Bit Error Counter II Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:62	—				*Reserved*
61:56	cnt_bad_ symbol_7	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violations or disparities error on Physical Lane 7 if this lane is active.
55:54	—				*Reserved*
53:48	cnt_bad_ symbol_6	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 6 if this lane is active.
47:46	—				*Reserved*
45:40	cnt_bad_ symbol_5	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 5 if this lane is active.
39:38	—				*Reserved*
37:32	cnt_bad_ symbol_4	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 4 if this lane is active.
31:30	—				*Reserved*
29:24	cnt_bad_ symbol_3	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 3 if this lane is active.
23:22	—				*Reserved*

TABLE 21-165 PEU Link Bit Error Counter II Register

Bit	Field	Reset Name	Reset Value	R/W	Description
21:16	cnt_bad_ symbol_2	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 2 if this lane is active.
15:14	—				*Reserved*
13:8	cnt_bad_ symbol_1	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 1 if this lane is active.
7:6	—				*Reserved*
5:0	cnt_bad_ symbol_0	rst_l	0_2	RO	Number of 8-bit/10-bit code encoding violation or disparity error on Physical Lane 0 if this lane is active.

21.4.3.142PEU SerDes PLL Control/Status Register ($006E2200_{16}$ / 1_{16})

Software can program these values any time, but the new value takes effect only after warm reset.

TABLE 21-166 PEU SerDes PLL Control/Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:8	—				*Reserved*
7:6	spare	por_l	0_2	RW	Spare bits.
5:4	lb	por_l	0_2	RW	Reference clock jitter and PLL loop bandwidth.
3:0	mpy	por_l	0001_2	RW	Reference clock multiplication. The default setting is based on 250 MHz reference clock, as follows: 0001_2 – for 250 MHz reference clock 0101_2 – 125 MHz reference clock 0111_2 – 100 MHz reference clock

21.4.3.143PEU SerDes Receiver Lane 0–7 Control Register

($006E2300_{16}$ – $006E2338_{16}$ / 444_{16})

The PEU SerDes Receiver Control register is described in TABLE 21-167.

TABLE 21-167 PEU SerDes Receiver Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:16	—		0_2	RO	*Reserved*
15	bsinrxn	por_l	0_2	RW	Boundary scan initialization for RXNi.
14	bsinrxp	por_l	0_2	RW	Boundary scan initialization for RXPi.
13:10	eq	por_l	0001_2	RW	Receiver equalizer configuration

TABLE 21-167 PEU SerDes Receiver Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
9:7	cdr	por_l	010_2	RW	Clock/data recovery algorithm selection. Default setting to second-order, high-precision threshold of 0 CDR Algorithm.
6:5	los	por_l	10_2	RW	Loss of signal detection. The threshold is in the range of 65 to 175 mVdifpp.
4:2	term	por_l	100_2	RW	Common mode termination is direct to VSSA. **Note:** The value of this field only takes effect after bit 3 of PEU CXPL MACL/PCS Control register is set. Before bit 3 of PEU CXPL MACL/PCS Control register is set, the value of 111_2 takes effect for all the SerDes's receiver channels.
1	cma_aln_en	por_l	1_2	RW	Comma symbol alignment enabled. Symbol alignment will be performed whenever a misaligned comma symbol is received. Deault is enabled.
0	entest	por_l	0_2	RW	Enable Test

21.4.3.144PEU SerDes Receiver Lane 0–7 Status Register

($006E2380_{16}$ – $006E23B8_{16}$ / 0_{16})

The PEU SerDes Receiver Status register is described in TABLE 21-168

TABLE 21-168 PEU SerDes Receiver Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:6	—		0_2	RO	*Reserved*
5	spare	por_l	0_2	RO	SPARE status bit.
4	spare	por_l	0_2	RO	SPARE status bit.
3	losdtct	por_l	0_2	RO	Loss of signal detect. Driven high asynchronously when a loss of signal (electrical idle) condition is detected for channel i. During the PCI-E L2 state, a low on this output indicates reception of a Beacon signal.
2	spare	por_l	0_2	RO	SPARE status bit.
1	sync	por_l	0_2	RO	Symbol alignment. When comma detection is enabled, this output is high when an aligned comma is received in the same cycle that the comma pattern is output on RDi. Alternatively, when an alignment jog is requested, it is high to indicate that the request has been completed. Synchronous to RXBCLKIN[i]. **Note:** The sync field could be altered by SerDes during warm reset sequence. This is due to the random data generator outputting the random data to sampler whenever the link is idle. The random data could form the comma character which causes comma detection logic detects a comma character.
0	testfail	por_l	0_2	RO	Test failure. Driven high when an error is encountered during a test sequence executed on channel i. Synchronous to RXBCLKIN[i].

21.4.3.145PEU SerDes Transmitter Lane 0 - 7 Control Register

($006E2400_{16}$ – $006E2438_{16}$ / $1F8_{16}$)

The PEU SerDes Transmit Control register is described in TABLE 21-169.

TABLE 21-169 PEU SerDes Transmit Control Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:11	—		0_2	RO	*Reserved*
10	spare	por_l	0_2	RW	Spare control bit – BSTX
9:6	de	por_l	0111_2	RW	De-emphasis. Selects one of 15 output de-emphasis settings from 4.76% to 71.42%.
5:3	swing	por_l	101_2	RW	Output swing. Selects one of 8 output amplitude settings between 125 and 1250 mVdfpp. The default setting is to select 1000 mVdfpp.
2	cm	por_l	1_2	RW	Common mode. Adjusts the common mode to suit the termination at the attached receiver.
1	invpair	por_l	0_2	RW	Invert polarity. Inverts polarity of TXPi and TXNi.
0	entest	por_l	0_2	RW	Enable test. Enables test modes specified via testcfg for this transmitter.

21.4.3.146PEU SerDes Transmitter Lane 0 - 7 Status Register

($006E2480_{16}$ – $006E24B8_{16}$ / 0_{16})

The PEU SerDes Transmitter Status register is described in TABLE 21-170.

TABLE 21-170 PEU SerDes Transmitter Status Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:2	—		0_2	RO	*Reserved*
1	rdtctip	por_l	0_{16}	RO	Receiver detect in progress.
0	testfail	por_l	0_{16}	RO	Test failure.

21.4.3.147PEU SerDes MACRO 0 - 1 Test Configuration Register ($006E2500_{16}$ – $006E2508_{16}$ / 03_{16})

The PEU SerDes Macro 0–1 Test Configuration register is described in TABLE 21-171.

TABLE 21-171 PEU SerDes MACRO 0 - 1 Test Configuration Register

Bit	Field	Reset Name	Reset Value	R/W	Description
63:20	—		0_2	RO	*Reserved*, not used
19:15	—		0_2	RO	*Reserved*, drive low.

TABLE 21-171 PEU SerDes MACRO 0 - 1 Test Configuration Register *(Continued)*

Bit	Field	Reset Name	Reset Value	R/W	Description
14	invpatt	por_l	0_2	RW	Invert polarity. Inverts polarity of Tx pattern generator in macros which have no transmitters.
13:12	rate	por_l	0_2	RW	Operating rate. Selects full, half or quarter rate operation for the Tx pattern generator in macros which have no transmitters.
11	—	por_l	0_2	RO	*Reserved*, drive low.
10	enbspt	por_l	0_2	RW	Receiver pulse boundary scan. Configures IEEE 1149.6 boundary scan comparators connected to RXPi and RXNi for EXTEST PULSE and EXTEST TRAIN when high and EXTEST when low.
9	enbsrx	por_l	0_2	RW	Receiver boundary scan. Enables IEEE 1149.6 boundary scan comparators connected to RXPi and RXNi, and the associated BSRXP and BSRXN bits of cfgrx[i]{27:0}.
8	enbstx	por_l	0_2	RW	Transmitter boundary scan. Enables IEEE 1149.6 boundary scan control of TXPi and TXNi from the BSTX bit of cfgtx[i]{23:0}.
7:6	loopback	por_l	0_2	RW	Loopback. Selects internal or bump/pad loopback.
5:4	clkbyp	por_l	0_2	RW	Clock bypass. Facilitates bypassing of the PLL with either REFCLKP/N or TESTCLKT, and bypassing of the recovered receiver clock with TESTCLKR.
3	enrxpatt	por_l	0_2	RW	Enable Rx patterns. Enables verification of test patterns in receivers.
2	entxpatt	por_l	0_2	RW	Enable Tx patterns. Enables generation of test patterns in transmitters.
1:0	testpatt	por_l	11_2	RW	Test Pattern. Selects one of two PRBS or two clock test patterns.

21.5 PIU Error Event Summary

PIU error events are summarized in TABLE 21-172 and TABLE 21-173.

TABLE 21-172 DMU Error Source and Handling Table *(1 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
DMU → SII DMA write data parity error	None in DMU	Generate bad parity on DMU → SII data. SII reports DMA write errors, logs address	DMA write is squashed with bad ECC on data.
DMU → SII PIO rd return/ Interrupt parity error	None in DMU	Generate bad parity on DMU → SII data, SII passes to NCU which logs, sends back to core	PIO loads get precise trap in core; interrupts are logged in NCU.
DMU → SII ECC error on CTAG from DMU to SII	None in DMU	SII checks DMA and logs, passes to NCU for PIO read and interrupts which logs.	Single bit ECC errors are corrected; double-bit errors cause writes to be squashed.
DMU → SII Parity on address in header from DMU to SII	None in DMU	SII checks DMA and logs, passes to NCU for PIO read cpl's and interrupts which logs.	DMA writes are squashed by clearing byte enables in SII.
DMU → SII Parity on cmd field of DSN → SII header	None in DMU	SII reports, destination is guessed and packet is passed on in error.	SII squashes any writes.
DMU → SII TO PIO rd cpl only	None in DMU	SII Sends to NCU PIO rd cpl has timed out	NCU handles.
DMU → SII UnMapped PIO rd cpl only	None in DMU	SII Sends to NCU PIO rd cpl with address errors are reported back to the NCU here	NCU interfaces with cores to handle.
SIO → DMU DMA read data return parity error, de bit set in header to local err.	Poison bit sent to ILU.	Single error bit to NCU, which logs, optional interrupt, poisoned data is detected at endpoint which reports back to initiating thread.	Parity is regenerated correctly, poison bit forwarded to ILU.
SII → DMU DMA write credit return parity error	None in DMU	DSN signals NCU with error bit. DSN drops this ACK back to DMU	One less credit ID to use in DMC. **Note:** All DMA writes, DMA reads, and interrupt mondos are affected because they all use the same tag pool. **Note:** Since DMA writes, DMA reads, and interrupt mondos all use the same tag pool and thus they are all affected when any one of them encounters a parity error or UE in the ACK or return.

TABLE 21-172 DMU Error Source and Handling Table *(2 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
SIO → DMU ECC error in header CTAG from SIO to DMU on DMA rd return	None in DMU	Two error bits sent to NCU for logging and optional interrupt Packet is never returned. Endpoint detects this and notifies the thread.	If CE, error is corrected. If UE, then packet is dropped from DSN to DMC and the credits are not released. This is considered a fatal error. The only way to get out of this situation is to issue warm reset.
DMU → NCU Parity on PIO write credit return to NCU	None in DMU	NCU checks and logs.	On error the credits are not released within the NCU.
NCU → DMU Parity error on MONDO ACK from NCU	None in DMU	Single error bit to NCU which logs, optional interrupt.	DSN drops credit return to IMU. This is considered a fatal error. The only way to get out of this situation is to issue warm reset.
MSI received but not enabled (DMU IMU – MSI Transaction)	MSI not-enabled bit (bit 0) is set in the IMU Error Status Clear register for primary detection. Bit 32 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, byte enables and MSI data is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an MSI to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
MSI Data Parity (DMU IMU – MSI Transaction)	MSI parity error bit (bit 6) is set in the IMU Error Status Clear register for primary detection. Bit 38 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, byte enables and MSI data is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an MSI to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
Malformed MSI (DMU IMU – MSI Transaction)	MSI malformed bit (bit 7) is set in the IMU Error Status Clear register for primary detection. Bit 39 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, byte enables and MSI data is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an MSI to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.

TABLE 21-172 DMU Error Source and Handling Table *(3 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Correctable message received but not enabled (DMU IMU – Message Transaction)	Correctable message not enabled bit (bit 1) is set in the IMU Error Status Clear register for primary detection. Bit 33 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and message code is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from a message to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
Nonfatal message received but not enabled (DMU IMU – Message Transaction)	Nonfatal message not enabled bit (bit 2) is set in the IMU Error Status Clear register for primary detection. Bit 34 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and message code is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from a message to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
Fatal message received but not enabled (DMU IMU – Message Transaction)	Fatal Message not enabled bit (bit 3) is set in the IMU Error Status Clear register for primary detection. Bit 35 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and message code is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an Message to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
PM PME message received but not enabled (DMU IMU – Message Transaction)	PM PME Message not enabled bit (bit 1) is set in the IMU Error Status Clear register for primary detection. Bit 33 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and message code is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an Message to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.

TABLE 21-172 DMU Error Source and Handling Table *(4 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
PME to ACK message received but not enabled (DMU IMU – Message Transaction)	PME to ACK Message not-enabled bit (bit 5) is set in the IMU Error Status Clear register for primary detection. Bit 37 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and message code is captured in the IMU RDS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an Message to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
EQ write to non-enabled EQ (DMU IMU – Message/MSI Transaction)	EQ not-enabled bit (bit 8) is set in the IMU Error Status Clear register for primary detection. Bit 40 is set for a secondary detection. PCIE transaction information including type, length, req_id, tlp_tag, and byte enables are captured in the IMU SCS Error Log register.	A maskable interrupt is generated.	The transaction is turned from an Message/MSI to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated.
EQ overflow error (DMU IMU – Message/MSI Transaction)	EQ Overflow bit (bit 9) is set in the IMU Error Status Clear register for primary detection. Bit 41 is set for a secondary detection. No other transaction information is captured.	A maskable interrupt is generated.	The transaction is turned from an Message/MSI to a NULL type in the IMU pipeline and passed up in order to retire the buffer space that was previously allocated. Also the EQ state will transition from `ACTIVE` to `ERROR`.
sun4u / sun4v IOMMU bypass mode error (Bypass access when be = 0)	MMU Error Status Clear register bit byp_err is set. Virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	A maskable interrupt is generated.Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u only MMU bypass mode out-of-range error	MMU Error Status Clear register bit byp_oor is set. Virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.

TABLE 21-172 DMU Error Source and Handling Table *(5 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
sun4v Only IOTSB descriptor invalid error	MMU Error Status Clear register bit iotsbdesc_inv is set. Adjusted virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated	Write transaction is terminated. Read transaction is completed with Completer Abort Status. The bit[39:13] of the MMU Translation Fault Address register doesn't contain any meaning info. SW should use bit[63] and the ID from the MMU Translation Fault Status register to identify the invalid entry in IOTSBDESC.
sun4v Only IOTSB descriptor data parity error	MMU Error Status Clear register bit iotsbdesc_dpe is set. Adjusted virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status. The bit[39:13] of the MMU Translation Fault Address register doesn't contain any meaning info. SW should use bit[63] and the ID from the MMU Translation Fault Status register to identify the parity error entry in IOTSBDESC.
sun4v Only sun4v VA out-of-range error (VA [62:40] ≠ 23'b0)	MMU Error Status Clear register bit sun4v_va_oor is set. Adjusted virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4v Only sun4v adjusted VA underflow error	MMU Error Status Clear register bit sun4v_adj_va_uf is set. Adjusted virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.

TABLE 21-172 DMU Error Source and Handling Table *(6 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
sun4u / sun4v IOTTE key protection error	MMU Error Status Clear register bit key_err is set. Virtual (SUN4U) / Adjusted virtual (SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU translation mode error	MMU Error Status Clear register bit trn_err is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU translation mode out-of-range error	MMU Error Status Clear register bit trn_oor is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU translation table entry invalid error	MMU Error Status Clear register bit tte_inv is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU translation table entry protection error	MMU Error Status Clear register bit tte_prt is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID, type, and RAM address are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.

TABLE 21-172 DMU Error Source and Handling Table *(7 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
sun4u / sun4v MMU translation table cache data parity error	MMU Error Status Clear register bit ttc_dpe is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID, type, and RAM address are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU Translation Table Cache Access Error (any CSR access to the cache while the cache is enabled)	MMU Error Status Clear register bit ttc_cae set.	A maskable interrupt is generated.	Write fails, read returns error status.
sun4u / sun4v MMU tablewalk disabled miss error (any miss in the cache of translation request)	MMU Error Status Clear register bit tbw_dme is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4u / sun4v MMU tablewalk unexpected data error (any miss in the cache of translation request)	MMU Error Status Clear register bit tbw_ude is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.

TABLE 21-172 DMU Error Source and Handling Table *(8 of 8)*

Event Detector	Information Captured	Reporting Mechanism	Impact
sun4u / sun4v MMU tablewalk error	MMU Error Status Clear register bit tbw_err set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
SUN4U / SUN4V MMU tablewalk data parity error	MMU Error Status Clear register bit tbw_dpe is set. Virtual(SUN4U) / Adjusted virtual(SUN4V) address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status.
sun4v Only Illegal page size detected during SUN4V IOMMU translation. The legal value for page sizes are as follows: 000_2 – 8K 001_2 – 64K 011_2 – 4M 101_2 – 256M The other values are illegal.	MMU Error Status Clear register bit sun4v_inv_pg_sz is set. Adjusted virtual address is logged in the MMU Translation Fault Address register. ID and Type are logged in the MMU Translation Fault Status register.	A maskable interrupt is generated.	Write transaction is terminated. Read transaction is completed with Completer Abort Status. The bit[39:13] of the MMU Translation Fault Address register doesn't contain any meaning info. SW should use bit[63] and the ID from the MMU Translation Fault Status register to identify the illegal page size entry in IOTSBDESC.

TABLE 21-173 PEU Error Source and Handling Table *(1 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Ingress header parity error.(From PEU IHB to DMU ILU)	ILU Error Status. Clear register bit (ihb_pe) is set.	A maskable interrupt is generated. (Mondo 62)	Header is dropped. Drain state (See Drain State Chapter for detail.) Fatal error, PIU Reset.
Ingress IO Rd – IORd (not supported)	PEU Uncorrectable Error. Status Clear register bit (ur) is set. Request header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers	A maskable interrupt is generated. (Mondo 63)	Completed with unsupported request status. Cpl NPH credits are recovered.
Ingress IO Wr – IOWr (not supported)	PEU Uncorrectable Error Status. Clear register bit (ur) is set. Request header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Completed with unsupported request status. Cpl NPD is dropped. NPH credits are recovered. No credit update for NPD (infinite).
Ingress Cfg Rd – CfgRd0 / CfgRd1 (not supported)	PEU Uncorrectable Error. Status Clear register bit (ur) is set. Request header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Completed with unsupported request status. Cpl NPH credits are recovered. No credits update for NPD (infinite).
Ingress Cfg Wr – CfgWr0 / CfgWr1 (not supported)	PEU Uncorrectable Error. Status Clear register bit (ur) is set. Request header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Completed with unsupported request status. Cpl NPD is dropped. NPH credits are recovered. No credits update for NPD.
Ingress Mem Rd Lock – MRdLk (not supported)	PEU Uncorrectable Error Status. Clear register bit (ur) is set. Request header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Completed with unsupported request status. CplL.k NPD is dropped. NPH credits are recovered. **Note:** MRdLk should not cause tablewalk or any trap in IOMMU.
Ingress unexpected completion (mismatches in tag or requester ID; ingress CplLk, CplDLk)	PEU Uncorrectable Error Status. Clear register bit (uc) is set. Completion header is logged in Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. No CHC// CDC credits update.

TABLE 21-173 PEU Error Source and Handling Table *(2 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Ingress Completion header error (a) mismatches in byte cnt, (chk can be disabled); addr, (chk can be disabled); length, TC (chk can be disabled) or Attr (chk can be disabled). (b) CplD with unsuccessful status; config-retry status for non-config regs (chk can be disabled); (c) CplD to PIO Config/ Wr reqs;) (d) Successful Cpl to PIO rd req.	PEU Other Event Status. Clear register bit (mfc) is set. Completion header is logged in Receive Other Event Header1 and Header2 Log registers. Original request header is logged in PEU Transmit Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. No CHC/ CDC credits update. An associated completion with timeout error status is generated by PEU.
Ingress MsgD (not supported) (exclude "Vendor Defined Type-1" messages, which are silently dropped)	PEU Uncorrectable Error Status. Clear register bit (ur) is set. Header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. PH/PD credits are recovered.
Completion timeout	PEU Uncorrectable Error Status. Clear register bit (cto) is set. Original request header is logged in PEU Transmit Uncorrectable Error Header1 and Header2 Log registers. PEU Other Event Status Clear register bit (cto) is set. Original request header is logged in PEU Transmit Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Completion is generated with timeout error status. Software determines impact.
Ingress MWr request (posted) with poisoned data	PEU Uncorrectable Error Status. Clear register bit (pp) is set. Request header is logged in Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. PH/PD credits are recovered.

TABLE 21-173 PEU Error Source and Handling Table *(3 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Ingress MsgD request (posted) with poisoned data	PEU Uncorrectable Error Status. Clear register bit (pp) is set. Also bit (ur) is set if it's not vendor-defined type 1 MsgD. Request header is logged in Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. PH/PD credits are recovered.
Ingress IOWr/CfgWr request with poisoned data	PEU Uncorrectable Error Status. Clear register bit (pp), bit (ur) is set. Request header is logged in Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. PH/PD Credits are recovered.
Ingress CplD with poisoned data	PEU Uncorrectable Error Status. Clear register bit (pp) is set. Completion header is logged in Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Software determines impact. No CH/CD credits update. (Infinite)
Ingress malformed packet: (a) max_payload_size violation for TLP including data (b) length size violation for TLP including data (c) TD field violation (d) Byte enable rules (RFE) (e) 4-KB boundary crossing (f) TC0 rule for msg. (g) TLP with undefined type field.	PEU Uncorrectable Error Status. Clear register bit (mfp) is set. Header is logged in PEU Receive Uncorrectable Error Header1 and Header2 Log registers.	A maskable interrupt is generated. Mondo 63)	Packet is dropped. Software determines impact. Credits are not recoverable. **Note:** PCI-Express 1.0 A states that all the credits are not recoverable for the malformed TLP packets. PCI-Express 1.1 states that credits are not recoverable for "ambiguous" malformed TLP packets. Credits are optionally not recoverable for nonambiguous TLP packets. However, spec doesn't distinguish between ambiguous and nonambiguous.
Link width change (can happen when LTSSM recovery leads to (reconfiguration.)	PEU Other Event Status. Clear register bit (lwc) is set.	A maskable interrupt is generated. (Mondo 63)	SW should reconfigure all the timer threshold registers in PEU.
No forward progress (when two consecutive retraining occur with no valid TLP received between them.)	PEU Other Event Status Clear register bit (nfp) is set.	A maskable interrupt is generated. (Mondo 63)	An indication that the remote device might be gummed up and cause a possible livelock in PCI-Express. Software determines impact.

TABLE 21-173 PEU Error Source and Handling Table *(4 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Ingress receiver overflow (error caused by flow-control buffer overflow condition)	PEU Uncorrectable Error Status. Clear register bit (rof) is set.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped. Credits are not recovered. Software determines impact.
Flow control protocol error (a) Flow control infinite update error (b) Flow control illegal size update error	PEU Uncorrectable Error Status Clear register bit (fcp) is set.	A maskable interrupt is generated. (Mondo 63)	This is caused by remote device's misbehavior. Software determines impact.
Data link protocol error. Sequence Number specified by the AckNak_Seq_Num doesn't correspond to an unacknowledged TLP or ACKD_SEQ	PEU Uncorrectable Error Status. Clear register bit (dlp) is set.	A maskable interrupt is generated. (Mondo 63)	This is caused by remote device's misbehavior. Software determines impact. DLLP is discarded. **Note:** It is not an error to receive an Ack DLLP when there are no outstanding unacknowledged TLPs, as long as the specified Sequence Number matches the value in ACKD_SEQ.
Replay timer timeout	PEU Correctable Error Status. Clear register bit (rtt) is set.	A maskable interrupt is generated. (Mondo 63)	Corrected by Link Layer Retry.
REPLAY_NUM rollover	PEU Correctable Error Status. Clear register bit (rnr) is set.	A maskable interrupt is generated. (Mondo 63)	Corrected by retraining the link.
Bad DLLP error (a) CRC not matched (b) EDB error	PEU Correctable Error Status. Clear register bit (bdp) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP Packet is dropped. **Note:** Unsupported DLLP is simply discard, and not considered as an error.
Bad TLP error –TLP error in Data Link Layer (a) CRC not matched & not a nullified TLP (b) CRC inverted & not a nullified TLP (c) CRC not inverted and EDB asserted (e) TLP sequence error	PEU Correctable Error Status. Clear register bit (btp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP Packet is dropped. Corrected by Data Link Layer Protocol. **Note:** Duplicate TLP is not an error. Ack DLLP is scheduled for transmission.

TABLE 21-173 PEU Error Source and Handling Table *(5 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Receiver error (a) 8-/10-bit decoder-related disparity error. (b) 8-/10-bit decoder-related code violation error. (c) Pipe elastic buffer overflow (d) Pipe elastic buffer underflow	PEU Correctable Error Status. Clear register bit (re) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP Receiver Error: DLLP is discarded. No error report from data link layer. TLP receiver error: TLP is discarded. NACK is scheduled. No error report from data link/transaction layer.
Memory read capture	PEU Other Event Status. Clear register bit (mrc) is set. Request header is logged in PEU Receive Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Diagnostic purpose. Read request is removed from the pipeline and requires software generated completion. No flow control credit is collected.
PIO write request completion with unsuccessful status	PEU Other Event Status. Clear register bit (wuc) is set. Completion header is logged in Receive Other Event Header1 and Header2 Log registers. Original request header is logged in PEU Transmit Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Software determines impact. NCU PIO credits is recovered. Unsuccessful completion packets are converted into internal "UR" record and returned to DMU for recovering NCU PIO credits. **Note:** When optional non-config CRS check is enabled, IO write completion with CRS should set OE mfc bit, but not OE wuc bit. When optional config CRS check is disabled, IO write completion with CRS should set OE wuc bit.

TABLE 21-173 PEU Error Source and Handling Table *(6 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
PIO read request completion with unsuccessful status	PEU Other Event Status. Clear register bit (ruc) is set. Completion header is logged in Receive Other Event Header1 and Header2 Log registers. Original request header is logged in PEU Transmit Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Software determines impact. NCU PIO credits are recovered. Unsuccessful completion packets are converted into internal UR record and returned to NCU for recovering PIO credits. **Note:** When optional non-config CRS check is enabled, IO read / memory read completion with CRS should set OE mfc bit, but not OE ruc bit. When optional config CRS check is disable, IO read /memory read completion with CRS should set OE ruc bit.
Completion with configuration retry status	PEU Other Event Status. Clear register bit (crs) is set. Completion header is logged in Receive Other Event Header1 and Header2 Log registers. Original request header is logged in PEU Transmit Other Event Header1 and Header2 Log registers.	A maskable interrupt is generated. (Mondo 63)	Software determines impact. NCU PIO credits is recovered. Unsuccessful completion packets are converted into internal "UR" record and returned to DMU/NCU for recovering NCU PIO credits.
Ingress interface parity error	PEU Other Event Status. Clear register bit (iip) is set.	A maskable interrupt is generated. (Mondo 63)	Packet is forwarded with bad parity. If the parity error is associated with header, it causes IHB parity error. If the parity error is associated with payload, it causes IHB parity error.
Egress data parity error (from EDB)	PEU Other Event Status. Clear register bit (edp) is set	A maskable interrupt is generated. (Mondo 63)	Drain state. Fatal error. If it happens during a transaction, the packet is terminated by inverting CRC and inserting EDB frame character.
Egress header parity error (from EHB)	PEU Other Event Status. Clear register bit (ehp) is set.	A maskable interrupt is generated. (Mondo 63)	
CXPL core interrupt	PEU Other Event Error Status. Clear register bit (lin) is set.	A maskable interrupt is generated. (Mondo 63)	Interrupt from CXPL core
Link reset	PEU Other Event Error Status. Clear register bit (lrs) is set.	A maskable interrupt is generated. (Mondo 63)	An indication that the remote device wants to reset the link. Software determines impact.
Link down transition	PEU Other Event Error Status. Clear register bit (ldn) is set.	A maskable interrupt is generated. (Mondo 63)	Drain state. Fatal error.

TABLE 21-173 PEU Error Source and Handling Table *(7 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Link up transition	PEU Other Event Error Status. Clear register bit (lup) is set.	A maskable interrupt is generated. (Mondo 63)	Status indication of link up.
Retry buffer underflow error	PEU Other Event Error Status. Clear register bit (eru) is set.	A maskable interrupt is generated. (Mondo 63)	Software determines impact.
Retry buffer overflow error	PEU Other Event Error Status. Clear register bit (ero) is set.	A maskable interrupt is generated. (Mondo 63)	Software determines impact.
Egress minimum packet error	PEU Other Event Error Status. Clear register bit (emp) is set.	A maskable interrupt is generated. (Mondo 63)	Packet is dropped.
Egress protocol error	PEU Other Event Error Status. Clear register bit (epe) is set.	A maskable interrupt is generated. (mondo 63)	Fatal.
Egress retry buffer parity error	PEU Other Event Error Status. Clear register bit (erp) is set.	A maskable interrupt is generated. (Mondo 63)	Fatal error. The current replay is aborted and the bad TLP is either dropped of truncated with an EDB symbol, depending on the location of parity error in the TLP packet. (a) If parity error happens at the first 16 byte of the TLP packet, the bad TLP packet is dropped. (b) If parity error happens after the first 16 TLP packet, it is truncated with an EDB symbol. If parity error occurs at every replay, the replay rollover error should occur. If replay rollover error happens twice, the no forward progress error will report. This is a fatal error and we recommend that software disable the link and retrain the link again.
Egress interface parity error	PEU Other Event Error Status. Clear register bit (eip) is set.	A maskable interrupt is generated. (Mondo 63)	Dummy check. No effect.
Skip greater than 6	PEU CXPL Error/Event Status. Clear register bit. (err_outstanding_skip) is set	A maskable interrupt is generated. (Mondo 63)	Software determines impact.
Elastic FIFO underflow	PEU CXPL Error/Event Status. Clear register bit. (err_elastic_fifo_undrflw) is set	A maskable interrupt is generated. (Mondo 63)	Clock compensation fails. TLP/DLLP packet will be corrupted. TLP/DLLP packet is dropped. NACK is scheduled for TLP packet. re is set.

TABLE 21-173 PEU Error Source and Handling Table *(8 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
Elastic FIFO overflow	PEU CXPL Error/Event Status. Clear register bit (err_elastic_fifo_ovrflw) is set.	A maskable interrupt is generated. (Mondo 63)	Clock compensation fails. TLP/DLLP packet will be corrupted. TLP/DLLP packet is dropped. NACK is scheduled for TLP packet. re is set.
De-skew alignment error	PEU CXPL Error/Event Status. Clear register bit (err_align) is set	A maskable interrupt is generated. (Mondo 63)	TLP/DLLP packet will be corrupted. TLP/DLLP packet is dropped. NACK is scheduled for TLP packet. re is set.
Kchar in DLLP/TLP error	PEU CXPL Error/Event Status. Clear register bit (err_unsup_dllp_tlp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP/DLLP Packet format rule violation.
Illegal END position error	PEU CXPL Error/Event Status. Clear register bit (err_ill_end_pos) is set.	A maskable interrupt is generated. (Mondo 63)	TLP/DLLP packet format rule violation. TLP/DLLP packet is dropped. NACK is scheduled for TLP packet.
Lost bit/byte synchronization	PEU CXPL Error/Event Status. Clear register bit (err_sync) is set.	A maskable interrupt is generated. (Mondo 63)	TLP/DLLP packet will be corrupted. re is set. Software should disable this error bits during link training and low power states to avoid false error reporting.
END/EDB symbol without STP/SDP in TLP or DLLP packet	PEU CXPL Error/Event Status. Clear register bit (err_end_edb_no_stp_sdp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP/DLLP packet format rule violation. TLP/DLLP packet is dropped. NACK may not be scheduled for the bad TLP packet.
DLLP SDP with no END	PEU CXPL Error/Event Status. Clear register bit (err_sdp_no_end) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP packet format rule violation. DLLP is dropped.
TLP STP with no END/EDB.	PEU CXPL Error/Event Status. Clear register bit (err_stp_no_end_edb) is set.	A maskable interrupt is generated. (Mondo 63)	TLP packet format rule violation. TLP is dropped. NACK is scheduled.
TLP with illegal PAD position	PEU CXPL Error/Event Status. Clear register bit (err_ill_pad_pos) is set.	A maskable interrupt is generated. (Mondo 63)	TLP packet format rule violation. TLP is dropped. NACK is scheduled.
DLLP with multiple SDP position	PEU CXPL Error/Event Status. Clear register bit (err_multi_sdp) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP packet format rule violation. DLLP is dropped.
TLP with multiple STP position	PEU CXPL Error/Event Status. Clear register bit (err_multi_stp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP packet format rule violation. TLP is dropped. NAK is scheduled.
DLLP with illegal SDP position	PEU CXPL Error/Event Status. Clear register bit (err_ill_sdp) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP packet format rule violation. DLLP is dropped.

TABLE 21-173 PEU Error Source and Handling Table *(9 of 9)*

Event Detector	Information Captured	Reporting Mechanism	Impact
TLP with illegal STP position	PEU CXPL Error/Event Status. Clear register bit (err_ill_stp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP packet format rule violation. TLP is dropped. NAK is scheduled.
Ingress unsupported DLLP	PEU CXPL Error/Event Status. Clear register bit (err_unsup_dllp) is set.	A maskable interrupt is generated. (Mondo 63)	DLLP packet is dropped. This is not an error case in PCI-Express spec.
TLP received with inverted CRC and EDB	PEU CXPL Error/Event Status. Clear register bit (err_src_tlp) is set.	A maskable interrupt is generated. (Mondo 63)	TLP packet is dropped. This is not an error case in PCI-Express spec.
Surprise removal or SerDes Los signal	PEU CXPL Error/Event Status. Clear register bit (err_sds_los) is set.	A maskable interrupt is generated. (Mondo 63)	Link is down. Drain state.

21.6 PIU Operation Sequence

21.6.1 PCI-Express Link Training Sequence

During fundamental reset (power on reset or warm reset), all the SerDes's receiver channels are in high impedance state. And after the fundamental reset, the SerDes's receiver channels continue to remain at high impedance state until software sets bit 3 (rx_high_imp_dis) of the PEU CXPL MACL/PCS Control register. Setting rx_high_imp_dis allows the termination selection (term field) from PEU SerDes's Receiver lane 0–7 Control register to take effect at the same time.

The remote device should repeat the `Detect.Quiet` and `Detect.Active` substates when PEU SerDes's receiver channel is in high impedance state.

The link does not proceed to train unless the rx_high_imp_dis bit 3 of the PEU CXPL MACL/PCS Control register is 1, where the bit defaults to 0 and "Remain in Detect.Quiet" bit 8 in PEU Control register is 0, where the bit defaults to 1. Therefore, after finishing all the CSR initialization, software must write to the PEU CXPL MACL/PCS Control register to set bit 3 to 1 and then write to PEU Control register to set bit 8 to 0. To prevent the link retrain automatically after it was down, software must set bit 8 in PEU Control register back to 1 after it receives a link up interrupt.

21.6.2 PCI-Express Hot Reset Sequence

Anytime that a software warm reset is initiated when a PCI Express link is up, the reset should be proceeded (if possible) by either a Hot Reset or a Link Disable. Hot Reset is initiated as follows:

1. Set link to stay in `Detect.Quiet` by setting bit 8 in the PEU Control register.
2. Set reset.assert bit of the PEU CXPL MAC/PCS Control register to 1.
3. Software can poll the ltssm_state{48:44} of the PEU Core Status register and wait until LTSSM enters `Detect.Quite` state. Software also makes sure that all the outstanding PIO read completions have returned to the threads.
4. Initiate software PIU local reset or warm reset.
5. After software PIU local reset or warm reset, set bit 3 of the PCU CXPL MACL/PCS Control register, and clear bit 8 in the PEU Control register to initiate link training.

21.6.3 PCI-Express Link Disable Sequence

Alternately, and preferred, a link can be disabled prior to the Warm Reset. Link disable uses the following steps:

1. Set link to stay in `Detect.Quiet` by setting bit 8 in the PEU Control register.
2. Set the disable bit (4) in the PEU Link Control register.
3. Software can poll the ltssm_state{48:44} of PEU Core Status register and wait until LTSSM enters `Detect.Quite` state. Software also makes sure that all the outstanding PIO read completions have returned to the threads.
4. Initiate software PIU local reset or warm reset.
5. After software PIU local reset or warm reset, set bit 3 of the PCU CXPL MACL/PCS Control register, and clear bit 8 in the PEU Control register to initiate link training.

21.6.4 `Drain` State

The purpose of `Drain` state is to install the PIU egress pipeline from NCU/SIU to PEU. This is required to enable internal PIOs issued by software to propagate to the CSR ring. In NCU, external and internal PIOs are stored in a single queue. If the pipeline stalls, internal PIOs could be blocked in the queue. Software would then be unable to read/clear the internal CSRs. `Drain` state alleviates this blockage and allows CSR access to software.

Four scenarios will get PIU into the `Drain` state:

- Ingress Header Buffer (IHB) parity error
- Egress Header Buffer (EHB) parity error
- Egress Data Buffer (EDB) parity error
- Link-down transition

When PIU is in the drain state, PIU does the following:

- Short-circuits PIO reads by manufacturing corresponding PIO completions with "Unsupported Request" status
- Silently drops PIO writes
- Stops receiving PCI-E traffic
- Silently drops outgoing DMA completions

The reason for completing PIO reads is to free the threads that issued the PIOs from waiting for their PIO completions and be able to issue new PIOs to internal CSRs. When any one of the above four scenarios occurs, the detector of the error will set its corresponding CSR error status bit if enabled and request for interrupt if enabled. Here are the steps software will take when it receives the interrupt:

1. PIO reads to the CSRs to identify the source of the problem. Note that once software identifies `Drain` state and root cause source of entering `Drain` state, software can ignore other error bits since during the flush operation, it could cause some miscellaneous errors to be reported.

2. A PIO write to the CSR to clear the error status bit.

3. If `Drain` state is caused by an IHB, and/or EHB, and/or EDB parity error, which are not recoverable since the PCI-E flow control credits are lost due to the header or data loss when a parity error occurs, follow the step 1, 2, and 3 of the procedure in the section of *PCI-Express Retrain the Link After Link Down*, below, and issue the SW PIU local reset.

4. Follow the entire procedures in the section of *PCI-Express Retrain the Link After Link Down* to get out of the `Drain` state if `Drain` state is caused by the link-down transition.

21.6.5 PCI-Express Retrain the Link After Link Down

The following procedure is followed if software wants to attempt to bring a PCI Express link up after it has gone down. Software must quiesce the system by executing the following three steps:

1. Software must shut down or suspend any process which may make PIO requests to the PCI Express link in question.

2. To ensure that no requests are outstanding, software issues PIO read requests to the link. All of these will complete with bus error status.

3. Software must ensure that no DMA read requests are outstanding by checking that no entries are on the Transaction Scoreboard.

Once the system is quiesced, software may bring up the link by executing the following three steps:

1. If another process has not already done so, the link-down bit in the PEU Other Event Status Clear register must be cleared by writing 1 to that bit.
2. The drain bit in the PEU Status register must be cleared by writing a 1 to that bit.
3. The Remain in `Detect.Quiet` bit in the PEU Control register must be cleared by writing a 0 to that bit. Once the link is up, the Remain in `Detect.Quiet` bit in the PEU Control register should be set.

21.6.6 PEU SerDes Clock and Electrical Configuration.

PEU provide control registers for SerDes's configuration. This allows OBP to overwrite the default SerDes's reference clock frequency and SerDes's electrical configuration for different platform. PEU supports three different SerDes reference clock: 100 MHz, 125 MHz, and 250 MHz. The default is 250 MHz.

The following is the expected OBP initialization sequence for switching reference clock frequency from default value (250 MHz) to either 125 MHz or 100 MHz.

- Step 1: Power-on reset.
- Step 2: After power-on reset, platform applies either 125 MHz or 100 MHz reference clock input to PEU SerDes's reference clock inputs.
- Step 3: For 125 MHz reference clock, OBP writes 0000 0000 0000 0005_{16} to the PEU SerDes PLL Control register (88 006E 2200_{16}). For 100 MHz Reference clock, OBP writes 0000 0000 0000 0007_{16} to the PEU SerDes PLL Control register.
- Step 4: OBP issues systemwide warm reset.
- Step 5: After warm reset, the link should be trained with the new reference clock.

21.6.7 PEU Deterministic Mode (DTM) Behavior and Sequence

The PEU deterministic mode provides a mechanism that allows the testing group to inject PCI-Express TLP packet through a normal tester to test the SOC functionality without involving the detailed PCI-Express link protocol, such as the complicated link training sequence and flow control initialization protocol.

The PEU deterministic mode also provides a mechanism that makes both the PEU and DMU operate at the same clock (100 MHz). This should eliminate the uncertainty of the asynchronous clock domain crossing between the PEU and DMU

during the normal operation and allows the cycle accuracy comparison between the simulation environment and the tester environment. Also, to ensure the deterministic behavior for incoming symbol crossing the elastic FIFO from the recovered clock domain to the I/O clock domain (100 MHz), it expects that the tester drives 100 MHz Refclk to both the UltraSPARC T2 refclk and PCI-Express SerDes refclk input and drives the serial data with 1.0 Gb/S data rate.

To make DTM mode work for PIU, the following is the expected initialization sequence.

1. After por_1 and por_2:
   a. JTAG writes 0000 0000 0000 0005_{16} to PEU SerDes's PLL Control register (88 006E 2200_{16}). This switches the MPY (multiplication factor) from default value (x5) to x10.
   b. JTAG writes 0000 0003 FFFF 0004_{16} or 0000 0003 FFFF 0008_{16} to DMU_ILU Diagnostic register ($0x8800652000_{16}$). This switches the rate scale from full rate to half rate or quarter rate.
   c. JTAG .....
   d. JTAG writes to the CCU CSR to enable the DTM mode.
2. JTAG issues warm reset.
3. During the warm reset, CCU sends a static signal, ccu_serdes_dtm, to PEU. After the warm reset, PEU uses this "static" signal to enter DTM mode and perform the following operations:
   a. Switch to dr_clk domain.
   b. Disable Data Link Layer Flow Control protocol. Mask bit 4 of PEU CXPL DLL Control register
   c. Disable Physical Layer Scrambler. Mask bit 1 of PEU CXPL MACL/PCS Control register
   d. Disable SerDes's Receiver channel's high impedance mode. Mask bit 3 of PEU CXPL MACL/PCS Control register.
   e. Switch the default Max Payload Size from 128 bytes to 512 bytes. Mask bit 6 of PEU Device Control register.
4. LTSSM remains at `Detect.Quiet` state.
5. Tester sends sufficient (TBD) FTS order sets for PEU to perform bit lock, symbol lock, and lane-to-lane de-skew. FTS order set is defined as K28.5 K28.1 K28.1 K28.1.

6. After Tester sends sufficient FTS order sets, Tester sends one SKP order set. When PEU detects this SKIP order set, a static signal is set to perform the following operations:

   a. Bypass LTSSM – Forces the physical layer LTSSM to move to `L0` state and always remain at `L0` state. It also forces the PEU active the entire eight lanes without lane reversals and polarity inversions

   b. Bypass DLCMSM – Forces the data link layer DLCMSM to move to `DL_Active` state and remain at `DL_Active` state. This is to bypass the flow control initialization sequence. Also, the infinite credits numbers are latched internally for DMA completion header, DMA completion data, PIO nonposted header, PIO nonposted data, PIO posted header, and PIO posted data.

   c. SKIP order set is defined as K28.5 K28.0 K28.0 K28.0.

7. JTAG can enable IOMMU bypass mode by writing $0000\ 0000\ 0000\ 0002_{16}$ into the IOMMU Control/Status register ($88\ 0064\ 0000_{16}$) during steps 4–6. Tester can start to send four DW TLP packets with 64-bit address memory access, where VA{63:39} = $1FFF800_{16}$. This address allows the TLP to bypass the IOMMU translation and allows IOMMU directly assign virtual address bits 38:2 to physical address bits 38:2.

During the DTM mode, the following signals from PEU PIPE (PHY interface for the PCI-Express, version 1.00) interface will be observed by tester through MIO pins.

- Tx_Data{7:0} – parallel pci-express egress bus for Lane 0
- Tx_Data{15:8} – parallel pci-express egress bus for Lane 1
- Tx_Data{23:16} – parallel pci-express egress bus for Lane 2
- Tx_Data{31:24} – parallel pci-express egress bus for Lane 3
- Tx_Data{39:32} – parallel pci-express egress bus for Lane 4
- Tx_Data{47:40} – parallel pci-express egress bus for Lane 5
- Tx_Data{55:48} – parallel pci-express egress bus for Lane 6
- Tx_Data{63:56} – parallel pci-express egress bus for Lane 7
- TxDataK{0} – 0 indicates a data byte; 1 indicates control byte for Lane 0
- TxDataK{1} – 0 indicates a data byte; 1 indicates control byte for Lane 1
- TxDataK{2} – 0 indicates a data byte; 1 indicates control byte for Lane 2
- TxDataK{3} – 0 indicates a data byte; 1 indicates control byte for Lane 3
- TxDataK{4} – 0 indicates a data byte; 1 indicates control byte for Lane 4
- TxDataK{5} – 0 indicates a data byte; 1 indicates control byte for Lane 5
- TxDataK{6} – 0 indicates a data byte; 1 indicates control byte for Lane 6
- TxDataK{7} – 0 indicates a data byte; 1 indicates control byte for Lane 7

CHAPTER 22

Network Interface Unit: Introduction

22.1 Features

- Line rate packet classification
  - ~33 MPkts/sec
  - L1 - L4 IPv4/v6 header parsing
  - Hierarchial classification : packet classes, VLAN (4K per port), ternary matches (256 entries) and hash function
  - 16 unique MAC addresses per port
- TCP/UDP/IP checksum offload
- Jumbo frame support (upto 9216B)
- Multiple DMA engines
  - Tx - 24 with DRR/gather support; Rx - 16 with WRED support
  - Flexible binding between DMAs and ports
  - Support CPU/thread affinity (avoid context switching)
- Virtualization support
  - DMA resource separation by logical groups (up to 8 partitions)
  - Descriptor address relocation
- Interrupts
  - Interrupt coalescing
  - Mailbox
- Multi-speed ports
  - 2 quad-speed (10M/100M/1G/10G)
  - 4x RGMII for 10M/100M/1G ports

22.2 Glossary

block A contiguous range (for example, 8 Kbyte, 32 Kbyte) of memory location. In UltraSPARC T2, this is a block of physically contiguous addresses. A block may be used to hold one or more packet buffers of the same size, or when used to store jumbo frames, multiple blocks are used to build a packet buffer.

logical address An address within a logical page.

logical condition (LC) A condition that, when true, may ultimately trigger an interrupt. It may be logically "level" in that the condition is constantly being evaluated, or it may be logically "edge triggered" in that a state is maintained when it first occurred. This state needs to be cleared to enable the hardware to detect the next occurrence.

logical device (LD) A term used generically to refer to a functional block that may ultimately cause an interrupt. This may be a transmit DMA channel, a receive DMA channel, a MAC, or other system-level blocks. Within a logical device, one can define one or more logical conditions. Each logical condition may be masked. A logical device may have up to two groups of logical conditions. Each group will have one "summary" flag (LDF). Depending on the logical conditions captured by the group, this flag may be level or may be edge-triggered. An unmasked logical condition, when true, *may* trigger an interrupt.

logical device flag (LDF) A logical **or** of some LCs within the LD. There are up to two flags per LD.

logical device group (LDG) A group of logical devices sharing an interrupt. A group may have only one LD. Currently, we support up to 64 logical device groups.

logical device group interrupt (LDGI) The interrupt associated with a LDG. This interrupt is controlled by a one-shot mechanism, that is, hardware will issue only one single interrupt, and software needs to arm the LDG again to enable it to issue another interrupt.

logical device interrupt mask (LDGIM) A per-LD mask that defines when the LC becomes true, whether a device may issue an interrupt.

logical device group state mask (LDGSM) A per-LDG mask that defines which part of the LDSV is visible to the LDG. This also defines which LD is part of a LDG.

logical device state vector (LDSV) A read-only state vector capturing the LDFs of *all* the LDs.

logical page A contiguous range of memory location. If an address posted by software is within a logical page, it will be translated to a physical address by replacing the base address of the logical page with the base address of the physical page. The size of the logical page is programmable.

receive block ring (RBR) A ring buffer of memory blocks posted by software.

receive completion ring (RCR) Stores the addresses of the buffers used to store incoming packets.

receive DMA channel (RDC) Consists of an RBR, an RCR, and a set of control and status registers. A receive DMA channel is selected after an incoming packet is classified. A packet buffer is derived from the pool and used to store the incoming packet. Each channel is capable of issuing interrupt to software based on the queue length of the receive completion ring or a timeout.

receive DMA channel table (RDC table) A table of 16 entries. Each entry contains one RDC. Each table defines the group of RDCs an incoming packet can be deposited into. The classification hardware will choose a table as an intermediate step before a final RDC is selected. The 0th entry of the table is the default RDC. This default RDC is used to queue error packets within the group. This default can be one of the RDC(s) in the group.

receive DMA channel group (RDC Group) A group of receive DMA channels stored in a corresponding RDC table.

receive packet classification Incoming packets will be classified based on layer 2/3/4 information. The result will select a Receive DMA Channel to store the packet.

system interrupt The Interrupt sent to the CPU. In UltraSPARC T2, this signal is sent to the NCU which will look up the (CPU ID, Interrupt Number) pair. Depending on the setting in the config space, the appropiate interrupt type will be issued.

transmit ring (TR) The data structure built-in system memory for software to post transmission requests.

transmit DMA channel (TDC) Consists of a transmit ring and a set of control and status registers. Software posts packets as a lists of gather pointers. Completion of packet transmission is indicated in status registers. Software needs to poll the status registers or enable hardware to issue interrupt after a specific packet is transmitted.

22.3 Functional Overview

A high-level block diagram is shown in FIGURE 22-1.

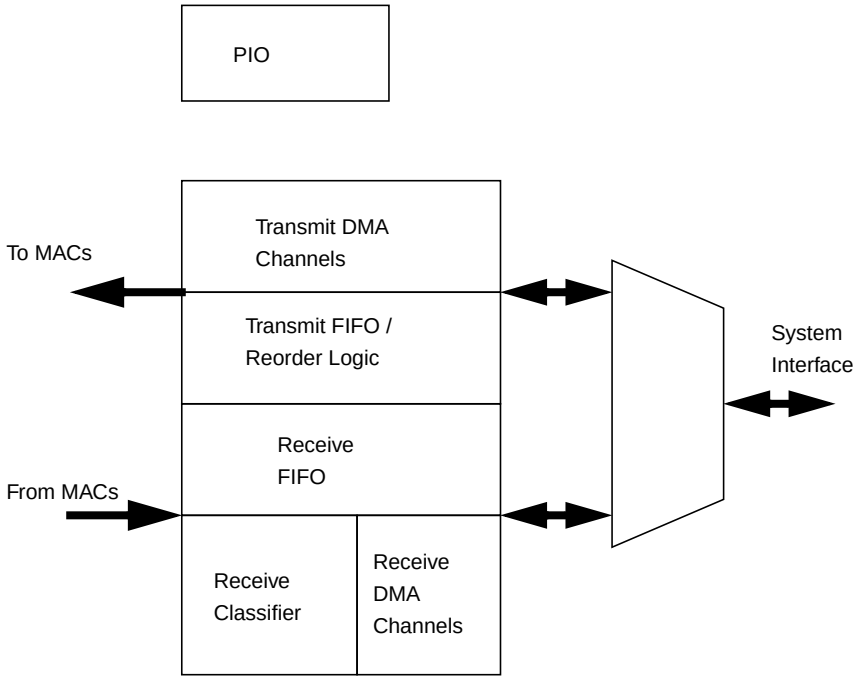


FIGURE 22-1 Ethernet Network I/O Subsystem Architecture

In UltraSPARC T2, System Interface is the datapath interconnecting the network interface sub-system to the system memory. The Program I/O (PIO) module is where memory-mapped I/O loads and stores to CSRs are dispatched to different functional units. The MACs are the Ethernet controllers, supporting the link protocol and statistics collection. Packets received are first classified based on the packet header information. The classification result determines the receive DMA channel. Transmit packets are posted into a transmit DMA channel. Each packet is made up of a gather list. Hardware supports checksum offload, on both receive and transmit data.

22.3.1 Resource Grouping And Virtualization Support

A general definition and usage of virtualization is beyond the scope of this document. For the purpose of this document, we concern ourself with accessing the device via PIOs and the address used in read/write requests. The latter relates to memory protection. Together these two features enable resource (both memory and DMAs) isolation, which is the basis of virtualization. In UltraSPARC T2, since the device is part of the processor, it is up to hypervisor how the hardware is presented to the OS. Not all hardware resources support virtualization directly. In some cases, it is to simplify the design, in others, it is the fundamental physical limit (for example, the MAC hardware state machine). Software driver access to these hardware blocks in an environment where virtualization is supported can be achieved via the virtualization management software layer.

In UltraSPARC T2, NIU is memory mapped to the system address. Hypervisor may organize the hardware as a two-function device. Within each function, two address ranges are defined: one for management one for virtualization. The entire device may be accessed through the management addresses. Virtualization addresses, on the other hand, only have accesses to a set of defined DMAs. The CSRs of multiple DMA channels can be grouped into an 8KB page within the virtualization address ranges. The grouping itself is defined by a table in the management address range.

To support memory protection, each transmit or receive DMA supports two logical pages. The addresses in the configuration registers, packet gather list pointers on the transmit side, and the allocated buffer pointer on the receive side will be relocated accordingly. The logical page registers are only accessable via the management address ranges.

In UltraSPARC T2, hypervisor software may expose an 8KB page, with a few DMAs defined, to the driver software, enabing the driver software to control the DMAs via PIOs. In addition, hypervisor also defines the logical page registers for these DMAs, which limits the addresses in the descriptors the driver posted. Together, this enforces DMA and memory resource the driver software may use.

Note that the datapath of a packet, as described in Section 22.3.4, is not affected by these features. The following figure summarizes the usage supported.

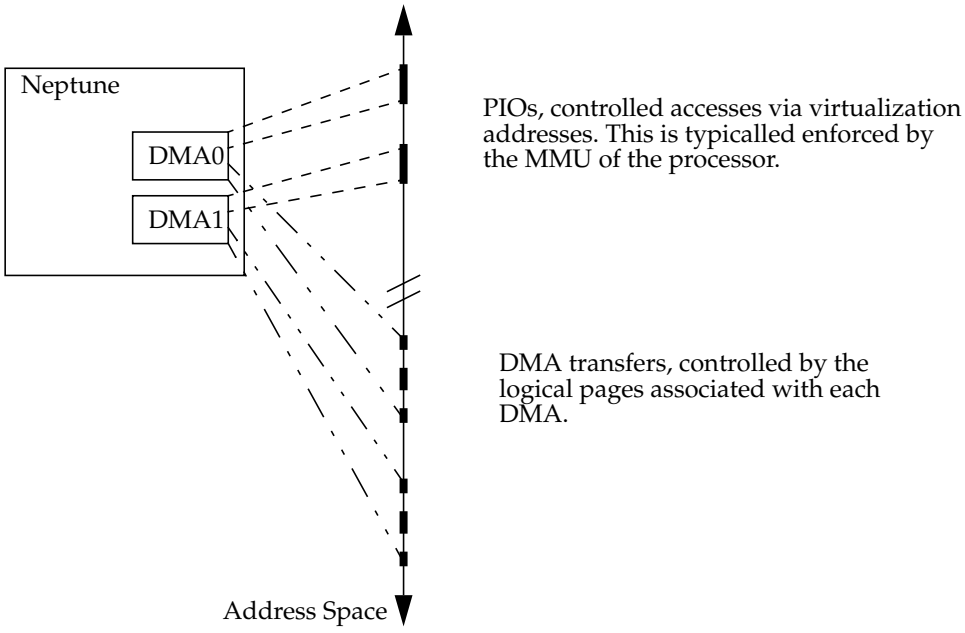


FIGURE 22-2 Virtualization support.

22.3.2 Interrupt Hierarchy

To support the sharing of available system interrupts, which may be less than the number of logical devices (LDs), LDs are grouped into logical device groups (LDGs). The state of the LDs that are part of an LDG may be read by software. Not all LDs belonging to a group can trigger an interrupt. This is controlled by the LD's interrupt mask. For example, a transmit DMA channel may be part of an LDG, and software may examine the flags associated with the transmit DMA by setting the LD's LDG number. However, the DMA will not trigger an interrupt if the corresponding bit in the interrupt mask is not asserted.

Associated with an LDG is a system interrupt control comprising an arm bit, a timer, and system interrupt data. System interrupt data is the data associated with the system interrupt and may be used to select the final interrupt sent to the processor. Driver software writes to the register to set the arm bit to 1 and to set the value of the timer. Hardware will start counting down the timer. An interrupt will only be issued if the timer is zero, the arm bit is set, and one or more LD's in LDG, have their flags set and not masked. This ensure that there is a period of "quiet" time between interrupt services.

Software needs to clear the state or adjust the conditions of individual LD after servicing. Note that hardware does not support any aggregate updates applied to the entire LDG.

For UltraSPARC T2, the system interrupt data is sent to the NCU to look up the hardware thread and interrupt number.

22.3.3 PIO and Datapath Interfaces

The UltraSPARC T2 system interface supports cache-line-size transfers. Logically, there are two classes of requests: ordered and bypass; they are queued separately in the SIU. An ordered request will not be issued to the memory system until older ordered and bypass requests are completed. Acknowledgments may return out of order. Bypass requests may be issued as long as the memory system can accept the request and may overtake older ordered requests that are enqueued or in transit to memory. In addition, write requests can be posted, and no acknowledgment will be returned. Packet data transfers, both receive and transmit, will be submitted as bypass requests. Control data (such as mailbox and completion ring updates) requests that affect the state of the DMA channels will be submitted as ordered requests.

For UltraSPARC T2, the NCU is the focal point where PIO requests will be dispatched to the NIU and where PIO read returns and interrupts are processed. It serializes the PIOs from different CPU threads to the NIU. It also has an internal table where, based on the system interrupt data, it looks up the CPU thread number and the interrupt number used. See Chapter 15, *Noncacheable Unit (NCU) and Boot ROM Interfaces*, for a more detailed discussion.

Internal to the PIO unit, a FIFO is used to queue up requests from CPUs. Requests will be read one by one and dispatched to the different functional units. Write requests may be dispatched to the functional unit so long as the functional unit can accept the request. Before a read request can be issued, *all* prior requests (read or write) have to be acknowledged.

22.3.4 Life of a Packet

A description of the receive path follows.

A more detailed block diagram of the receive classification logic and receive FIFO is shown in FIGURE 22-3.

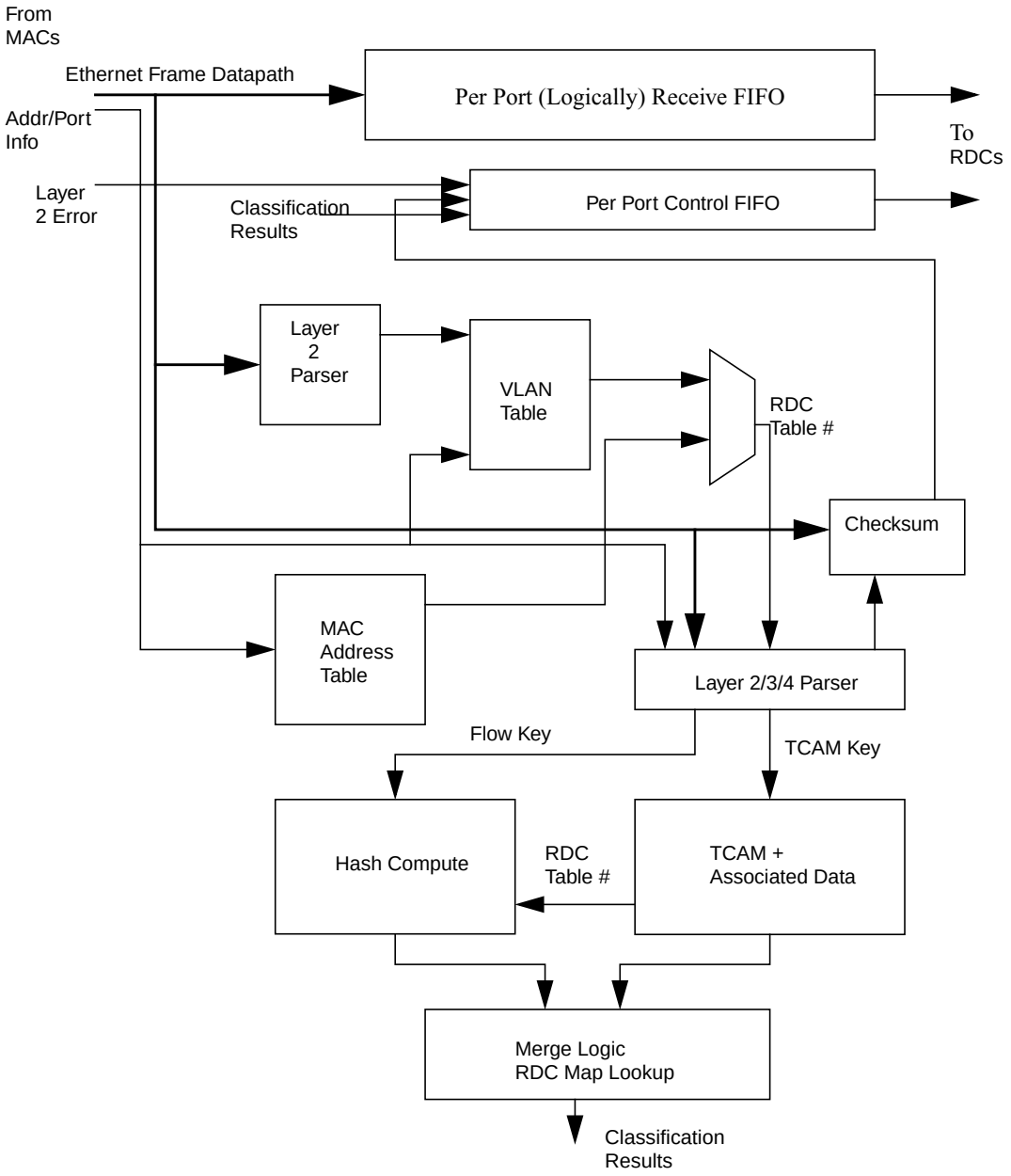


FIGURE 22-3 Receive Buffer and Classification Logic

MACs are the Ethernet media access controllers that support the Ethernet protocol. They contain the Layer 2 protocol logic, statistic counters, address matching, and filtering logic. The output from the MACs contains information on the destination address, whether it is one of the programmed individual addresses or an accepted group address, and the index associated with the address in that category.

Frames from different physical ports are stored temporarily in a per-port receive FIFO. While they are being stored into the FIFO, the frame will also be copied to the packet classification and checksum engines. The classification logic will determine which RDC group the packet belongs to and an offset into the RDC Table where the final RDC is determined. There are a total of eight RDC Tables.

The Layer 2 parser processes the Ethernet header to determine if the received frame contains a VLAN tag or LLC/SNAP header. For a VLAN-tagged packet, the VLAN ID is used to look up a VLAN table to determine the RDC table number for the packet. Hardware will also look up the MAC address table to determine an RDC table number based on the destination MAC address information. Software can program which of the two groups to use in subsequent classification. The output of the Layer 2 parser together with the resulting RDC table number will be passed to the Layer 2/3/4 parser.

The Layer 2/3/4 parser will examine the ethertype, tos/dscp field, and the protocol id/next header field to determine if the IP packet needs further classification. It is hardwired to recognize some fixed protocol such as TCP or UDP. It also supports a number of programmable Protocol IP numbers. If the packet needs further classification, it will generate a flow key and a TCAM key.

The TCAM key is sent to the TCAM unit for an associative search. If there is a match, the result may override the RDC table selection from L2 and/or contain an offset into the Layer 2 RDC table and ignore the result from the hash unit. The outputs from the hash unit and the TCAM unit will be merged to determine the RDC. This RDC is used to store the packet.

The outputs of the classification unit are stored into the Control FIFO.

Hardware supports checksum offload and will simply compare the calculated values with the values embedded in the frame. The result will be sent to software through the completion status. No discard decision is made. Note that checksum errors do not affect the L3/4 classification logic. Similarly, the error status will be sent to software through the completion status.

The Receive FIFO is logically organized per physical port. Layer 2/3/4 error information has to be logically synchronized with the classification result of the corresponding frame.

Logically, 16 receive DMA channels are available to incoming packets. The datapath engine is common across all DMA operations. It is also used to prefetch the receive blocks or update the completion ring of the RDCs. FIGURE 22-4 illustrates the functional blocks within the receive DMA channel logic block.

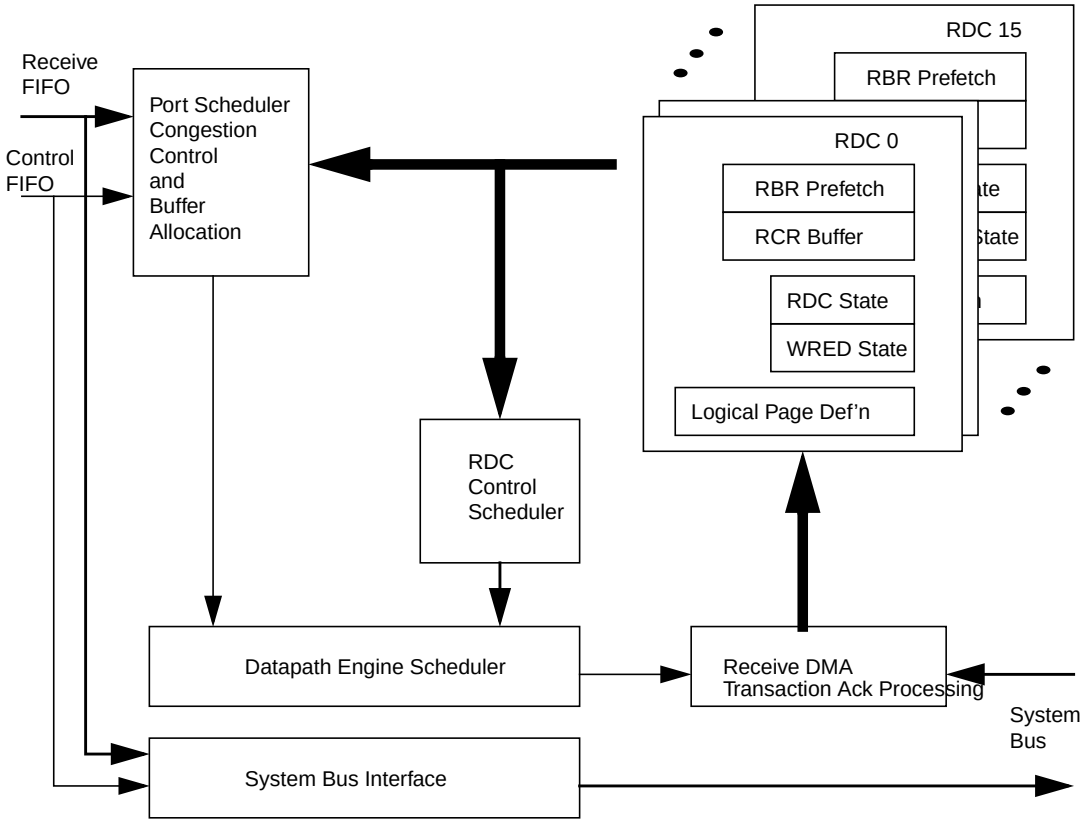


FIGURE 22-4 Receive DMA Channels

Each receive DMA channel (RDC) has a receive block ring (RBR), a receive completion ring (RCR), and the state associated with the RDC. Physically, they are allocated as ring buffers in DRAM. To support partitioning, each RDC supports two logical pages. All the addresses posted by software, such as the configuration of the ring buffers and buffer block addresses, are translated to physical addresses when used to reference system memory.

Software posts buffer blocks into the RBR. The size of each block is programmable, but fixed per channel. Software can specify up to three packet sizes for hardware to partition a block. Each block can only contain packet buffers of the same size.

To reduce the per-packet overhead, hardware maintains a prefetch buffer for the RBR and a tail buffer for the RCR. When the RBR prefetch is low, a request will be issued to the DRAM system to retrieve a cache of block addresses from the ring. Or

if the RCR tail buffer needs to be updated, a write request will be issued. The consistency of the RCR state is maintained by the hardware. The RDC control scheduler will maintain the fairness among the RDCs.

The port scheduler examines whether any frames are available from the Receive FIFO and the Control FIFO, and decides which port to service first. A deficit round robin scheduler is implemented. From the control header, the scheduler determines which RDC to check for congestion and retrieves a buffer to store the frame. Congestion is determined by a WRED algorithm applied on the receive completion ring. If the RDC is not congested, a buffer address is allocated according to the packet size. Packet data requests are issued as posted writes.

The datapath engine will fairly schedule the requests from the port scheduler and the RDC control scheduler and issue the requests to the DRAM.

The RCR buffer will be updated after issuing the write requests for the entire packet. The DMA status registers will be updated every time the RCR buffer is updated. Software may poll the DMA status registers to determine if any packet has been received. When the RCR queue length reaches a threshold or a timeout occurs, hardware may update the RCR buffer and at the same time, write the DMA status registers to a software defined mailbox. The software state will then be updated, and a logical device flag may be raised. The LDF may then lead to a system interrupt. Hardware will maintain the consistency of the DMA status registers and the RCR in that the status registers reflects the content of the RCR in system memory.

In UltraSPARC T2, there are 16 Transmit DMA Channels. FIGURE 22-5 shows the logical view of the transmit hardware. For UltraSPARC T2, there are two Ethernet ports. Each channel consists of a transmit ring and a set of control and status registers. Similar to the receive side, each channel supports two logical pages. Addresses in the transmit ring and configuration registers are subjected to a translation to convert to physical addresses.

The transmit ring is built from a ring buffer in system memory. Software posts packets into the transmit ring and signals the DMA hardware that packets have been queued. Each packet is built as a gather list. (Note that hardware will check to ensure that the total transfer size associated with a packet does not exceed the maximum hardware limit. This includes the internal headers.) When the transmit ring is not empty, hardware will prefetch the transmit ring into a per channel buffer.

Any DMA channel can be bound to one of the Ethernet ports by software. This is controlled by a mapping register at the per-port DRR scheduler. The DRR scheduler may switch to a different channel on packet boundary. This guarantees there will be no packet interleaving from different DMA channels. The scheduler will first acquire an available buffer for that port. If it is available, a memory request will be issued. A buffer tag identifying the buffer is needed to reorder potentially out-of-order read returns. This tag is linked to the request/ACK ID.

The Ethernet ports are serviced in round-robin order, and requests from different ports may be interleaved. See FIGURE 22-5.

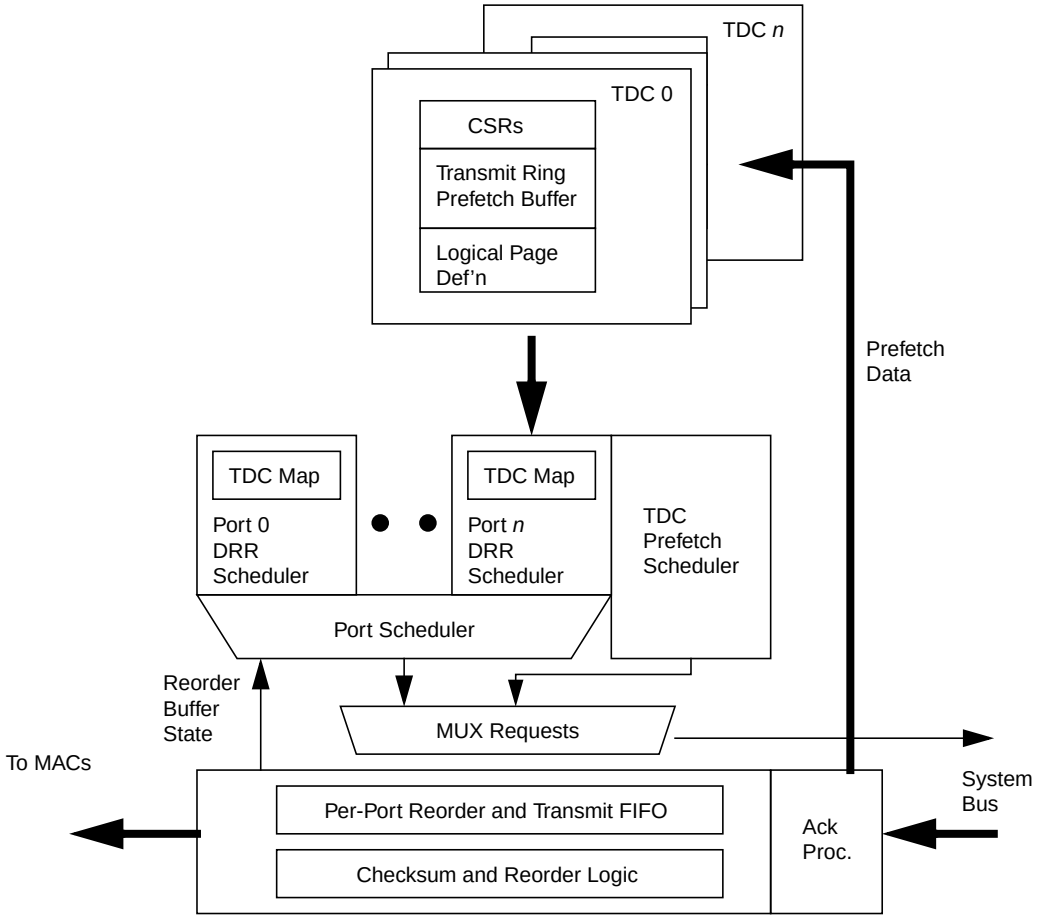


FIGURE 22-5 Transmit Functional Blocks

The transmit data requests and the prefetch requests share the same datapath to memory system. The returned acknowledgment is first processed to decide whether it is a prefetch or a transmit data. Transmit hardware also support checksum offload. This logic is embedded in the Reorder and Transmit FIFO logic.

When the entire packet has been received into the Transmit FIFO, the transmission of the packet is considered to be completed and the state of the DMA channel will be updated through the associated status register. A 12-bit wrap around counter, initialized to 0, keeps track of packets transmitted. Software needs to poll the status registers to determine the status. Alternatively, software may *mark* a packet so that an interrupt (if enabled) may be issued after the transmission of the packet. Similarly to the receive side, hardware may update the state of the DMA channel to a predefined mailbox after transmitting a marked packet.

Transmit and receive logic fairly share the same memory system interface.

22.3.5 Notes on Register Definition and DMA Addressing

The following convention is used in defining the addresses of one or more register blocks.

Register Name – Short_name (Base) [(count c step s)]

Register Name is the name of the register. Short_name is a short alias of the register. ***Base*** is the base address of block . ***count c*** step s indicates that there are *c* counts of the register in each block, the addresses of the registers are s apart. For example, if Base is 1000 0000_{16}, count is 3, and step is 8, then the addresses for the registers are 10000000, 10000008, and 10000010_{16}. Registers that hardware will not initialize are marked as Xs, and software needs to write proper values to these registers.

Internally NIU is a *little-endian* device. For UltraSPARC T2, only 8-byte naturally aligned register accesses are supported. All registers are defined to be 64 bits wide, though most of the registers have only 32 bits. Writes to the reserved fields will be ignored and have no side effect. Reads to the reserved fields will be accepted and zeros are returned.

For UltraSPARC T2, all DMA-related requests support 39-bit addressing or bits 38:0. Bit 39 is used internally for testing.

CHAPTER 23

Network Interface Unit: Interrupts and Virtualization

23.1 Address Assignment and Multifunction / Multidevice Support

Logically, NIU can be viewed as a platform with multiple independent hardware units. In UltraSPARC T2, NIU can be viewed as multiple single-function devices, or a single device with multiple functions, depending on how system software presents the hardware to the OS. Of the 39-bit address space, a fixed decode value for the higher-order 8 bits defines the NIU subregion. Refer to NCU section for the decode value. Internal to the NIU PIO block, the following address mapping is supported. Bits 26:24 identify a function and a memory region within a function. (Physically, the notion of a function in UltraSPARC T2 does not exist; it is here for description only.) There are two types of memory regions: management region and virtualization region. Within a memory region, all hardware registers are visible. In virtualization region, only selected hardware registers are accessible. The latter region supports hardware virtualization. The management regions of all functions are simply aliases of each other. Virtualization regions are different. However, whether certain registers are accessible within a function's management region, and what the access right (either read-only or read/write) is, depends on the device function number.

In UltraSPARC T2, each function has a management region and two virtualization regions.

Within the NIU address space, all defined registers are 64-bit entities. This does not mean that there is a physical device behind every bit field. In particular, reserved fields in a register are read-only, and will return zeros when read. (For example, there may be 32-bit registers in some logic block. These registers are extended to have 64-bit definition, with the higher-order 32 bits reserved. On reads, the higher-order 32 bits will return 0's.) Write accesses to address space not defined in the PRM will be silently discarded. Read requests to these addresses will cause a *data_error_exception* trap back to the CPU in UltraSPARC T2.

In a single-partition software model, the drivers associated with different functions may come up independently. The first driver that comes up may attempt to configure the device. The register DEV_FUNC_SR provides a simple test-and-set mechanism and a scratch pad register to facilitate locking and communication (described later in a this section). After configuration, software may decide to make the driver associated with function 0 as the master control. Thus, each internal hardware block has a function zero control (FZC) zone that normally may be accessed by any function. There is a control register bit, mpc bit in MULTI_PART_CTL register; if set, all the registers within these zones can only be written to via the function zero management space. Registers within this space *may* be read by any function.

The management region of NIU occupies a 24-bit address space.Bits 20–23 are used to select internal blocks. TABLE 23-1 shows the address allocated to different blocks. Bit 19 of an address is set to 1 for FZC zones. This enables a 20-bit address bus between PIO and other blocks. Note that in the register definition, the notation used is (block_name + $NNNNN_{16}$). The value $NNNNN_{16}$ should be interpreted as a 19-bit number. Bit 19 is implicit in the block name.

The following are used to generate the base addresses of different blocks. Note that the values are defined as variables. These are comment only.

RegisterBaseAddress PIO - 000000_{16}

RegisterBaseAddress FZC_PIO - 080000_{16}

RegisterBaseAddress FZC_MAC - 180000_{16}

RegisterBaseAddress FZC_IPP - 280000_{16}

RegisterBaseAddress FFLP - 300000_{16}

RegisterBaseAddress FZC_FFLP - 380000_{16}

RegisterBaseAddress ZCP - 500000_{16}

RegisterBaseAddress FZC_ZCP - 580000_{16}

RegisterBaseAddress DMC - 600000_{16}

RegisterBaseAddress FZC_DMC - 680000_{16}

RegisterBaseAddress TXC - 700000_{16}

RegisterBaseAddress FZC_TXC - 780000_{16}

RegisterBaseAddress PIO_LDSV - 800000_{16}

RegisterBaseAddress PIO_LDGIM - 900000_{16}

RegisterBaseAddress PIO_IMASK0 - $A00000_{16}$

RegisterBaseAddress PIO_IMASK1 - $B00000_{16}$

RegisterBaseAddress FZC_PROM - $C80000_{16}$

RegisterBaseAddress FZC_PIM - $D80000_{16}$

TABLE 23-1 Block Address Assignment

24-Bit Base Address ({23:19} Used in Block Selection)	Block Name	Physical Module
000000_{16}	PIO	PIO
080000_{16}	FZC_PIO	PIO
$1x00_{16}$	*Reserved*	
180000_{16}	FZC_MAC	MAC
200000_{16}	*Reserved*	
280000_{16}	FZC_IPP	IPP
300000_{16}	FFLP (Header Parser)	FFLP
380000_{16}	FZC_FFLP	FFLP
400000_{16}	*Reserved*	
480000_{16}	*Reserved*	
500000_{16}	ZCP	ZCP
580000_{16}	FZC_ZCP	ZCP
600000_{16}	DMC	600000_{16} – 630000_{16}: RDMC 640000_{16} – 670000_{16}: TDMC
680000_{16}	FZC_DMC	680000_{16} – $6B0000_{16}$: RDMC $6C0000_{16}$ – $6F0000_{16}$: TDMC
700000_{16}	TXC	TXC
780000_{16}	FZC_TXC	TXC
800000_{16}	PIO_LDSV (PIO address space for the logical device state vector)	PIO
880000_{16}	*Reserved*	
900000_{16}	*Reserved*	
980000_{16}	*Reserved*	
$A00000_{16}$	PIO_IMASK0 (PIO address space for logical device interrupt mask.)	PIO
$A80000_{16}$	*Reserved*	
$B00000_{16}$	PIO_IMASK1 (PIO address space for logical device interrupt mask.)	PIO
$B80000_{16}$ – $C00000_{16}$	*Reserved*	
$C80000_{1616}$	*Reserved*	PROM

TABLE 23-1 Block Address Assignment *(Continued)*

24-Bit Base Address ({23:19} Used in Block Selection)	Block Name	Physical Module
$D00000_{16}$	*Reserved*	
$D80000_{16}$	*Reserved*	PEU
$E00000_{16}$ – $F80000_{16}$	*Reserved*	

For UltraSPARC T2, the address regions are defined below. Note that two virtualization regions are defined for each function. Physically, the notion of a function in UltraSPARC T2 does not exist. It is used primarily for description only. System software may simply treat them as memory mapped I/O regions.

RegisterBaseAddress FUNC0_MNT – 0000000_{16}

RegisterBaseAddress FUNC0_VIR – 1000000_{16}

RegisterBaseAddress FUNC0_VIR2 – 5000000_{16}

RegisterBaseAddress FUNC1_MNT – 2000000_{16}

RegisterBaseAddress FUNC1_VIR – 3000000_{16}

RegisterBaseAddress FUNC1_VIR2 – 7000000_{16}

RegisterBaseAddress RESERVED0 – 4000000_{16}

RegisterBaseAddress RESERVED1 –- 6000000_{16}

Note that RESERVED0 and RESERVED1 are both defined and can be accessed by the processor similar to the other management regions. However, software is recommended to use only either FUNC0_MNT or FUNC1_MNT.

TABLE 23-2 summarizes the address assignment.

TABLE 23-2 Virtualization Region Base Address for UltraSPARC T2

27-Bit Base Address ({26:24} Used in Function Selection)	Description
0000000_{16}	Function 0 management region, FUNC0_MNT
1000000_{16}	Function 0 virtualization region, FUNC0_VIR
5000000_{16}	Function 0 virtualization region, FUNC0_VIR1
2000000_{16}	Function 1 management region, FUNC1_MNT
3000000_{16}	Function 1 virtualization region, FUNC1_VIR
7000000_{16}	Function 1 virtualization region, FUNC0_VIR1
4000000_{16}, 6000000_{16}	Defined to be reserved regions, software should not use these areas.

To facilitate the communication among different functions, a shared register is available. Bit 31 provides the Test-And-Set atomic capability, enabling software to implement a lock on the register. In single partition multiple drivers scenario, this register can be used by the first driver, not necessarily the function zero driver to come up first and configure the entire NIU. As an aid to identify the function itself, funcid bits return the value of the address bits 25–26 on read access.

TABLE 23-3 Device Function Shared Register – DEV_FUNC_SR (PIO + 10000_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	tas	0	RW	After read, this bit is always set to 1. Software writes 0 to clear.
30:18	—	0	RO	*Reserved*
17:16	funcid	0	RO	Return bits {26:25} of internal address bits as a hint to function identity.
15:0	sr	0	RW	Shared register.

In one possible virtualization model, a partition manager may own a function and prevent other software entities, via different addresses, from accessing the management region. The intended usage is for the partition manager, which has to own function 0, to first come up and configure the device. After configuration, the partition manager will block the access from other entites by turning off the access from other functions. When the mpc bit, described in TABLE 23-4, is set to 1, the FZC region can only be written to via function 0 address space. This prevents other functions from changing the configuration. All changes must be performed by the partition manager, accessed via function 0 address space. Note that this register is also in FZC space. On power-up or reset, the mpc bit is set to 0, and the register can be accessed (both read and write) via any function's address space. Once a 1 is written to the mpc bit, the register can only be cleared via function 0's address space.

TABLE 23-4 Multipartition Control – MULTI_PART_CTL (FZC_PIO + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	Reserved.
0	mpc	0	RW	Set to 1 to limit write access to FZC region to function zero.

23.2 Virtualization Region

In multiple-partition mode, each partition will be presented with only the virtualization region. This is achieved by binding the virtualization region to the driver by a hyperprivileged management software. Note that the access to the virtualization region is *not* enforced by hardware, and software may choose to access

the device through these regions even in the single partition case. (For example, this may simplify the task of developing both single- and multipartition driver with as much shared code as possible.) Within the virtualization region of each function, access to DMA channels, logical device state vector and interrupt management, and flow classification table are supported. There are two subregions defined for each category. Physically, no registers are implemented in this region. Hardware simply remaps the address, if appropriate, and forward the requests to the individual blocks. Note that some of the requests may be targeted to the PIO block itself. For UltraSPARC T2, TABLE 23-5 summarizes the high-level translation function.

There are certain restrictions when using the virtualization region. (See the next section for the usage of logical device group. Note that there is no restriction when accessing through the management region when operating in single partition mode.) The access to logical device groups is hardwired to each function, with 16 groups per virtualization region. Groups 0 to 15 are for the virtualization region in function 0, groups 16 to 31 are for that of function 1, etc. Similarly, the access to FFLP resources (the interface to external hash table) are also hardwired; function 0 has hardwired access to the first two interfaces, function 1 to the next two interfaces, etc. See the following for clarification. Each virtualization region occupies 32 Kbytes of address space.

TABLE 23-5 UltraSPARC T2 Virtualization Region Address Translation (VRADDR = PIO address issued) *(1 of 2)*

Function	Subregion	Address Range	Remapped Range
0	DMA CSR	FUNC0_VIR: FUNC0_VIR+$1FFF_{16}$	See below for details.
0	Logical Device State Vector and Interrupt Management	FUNC0_VIR+2000_{16}: FUNC0_VIR+$20FF_{16}$	PIO_LDSV+ 8KxVRADDR{7:5} + VRADDR{4:0}
0	Reserved	FUNC0_VIR+2100_{16}, FUNC0_VIR+2108_{16} and FUNC0_VIR+2110_{16}	FFLP+00000_{16}, fFLP+00008_{16}, and FFLP+00010_{16}
0	DMA CSR	FUNC0_VIR+4000_{16}: FUNC0_VIR+5FFF	See below for details.
0	Logical Device State Vector and Interrupt Management	FUNC0_VIR+6000_{16}: FUNC0_VIR+$60FF_{16}$	PIO_LDSV+8x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
0	Reserved	FUNC0_VIR+6100_{16}, FUNC0_VIR+6108_{16} and FUNC0_VIR+6110_{16}	FFLP+8K+00000_{16}, FFLP+8K+00008_{16} and FFLP+8K+00010_{16}
1	DMA CSR	FUNC1_VIR: FUNC1_VIR+$1FFF_{16}$	See below for details.
1	Logical Device State Vector and Interrupt Management	FUNC1_VIR+2000_{16}: FUNC1_VIR+$20FF_{16}$	PIO_LDSV+16x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
1	Reserved	FUNC1_VIR+2100_{16}, FUNC1_VIR+2108_{16} and FUNC1_VIR+2110_{16}	FFLP+16K+00000_{16}, FFLP+16K+00008_{16} and FFLP+16K+00010_{16}
1	DMA CSR	FUNC1_VIR+4000_{16}: FUNC1_VIR+$5FFF_{16}$	See below for details.

TABLE 23-5 UltraSPARC T2 Virtualization Region Address Translation (VRADDR = PIO address issued) *(2 of 2)*

Function	Subregion	Address Range	Remapped Range
1	Logical Device State Vector and Interrupt Management	FUNC1_VIR+6000_{16}: FUNC1_VIR+$60FF_{16}$	PIO_LDSV+24x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
1	Reserved	FUNC1_VIR+6100_{16}, FUNC1_VIR+6108_{16} and FUNC1_VIR+6110_{16}	FFLP+24K+00000_{16}, FFLP+24K+00008_{16} and FFLP+24K+00010_{16}
0	DMA CSR	FUNC0_VIR2: FUNC0_VIR2+$1FFF_{16}$	See below for details.
0	Logical Device State Vector and Interrupt Management	FUNC0_VIR2+2000_{16}: FUNC0_VIR2+$20FF_{16}$	PIO_LDSV+32x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
0	Flow Table	FUNC0_VIR2+2100_{16}, FUNC0_VIR2+2108_{16} and FUNC0_VIR2+2110_{16}	FFLP+32K+00000_{16}, FFLP+32K+00008_{16} and FFLP+32K+00010_{16}
0	DMA CSR	FUNC0_VIR2+4000_{16}: FUNC0_VIR2+$5FFF_{16}$	See below for details.
0	Logical Device State Vector and Interrupt Management	FUNC0_VIR2+6000_{16}: FUNC0_VIR2+$60FF_{16}$	PIO_LDSV+48K+ 8KxVRADDR{7:5} + VRADDR{4:0}
0	Flow Table	FUNC0_VIR2+6100_{16}, FUNC0_VIR2+6108_{16} and FUNC0_VIR2+6110_{16}	FFLP+40K+00000_{16}, FFLP+40K+00008_{16} and FFLP+40K+00010_{16}
1	DMA CSR	FUNC1_VIR2: FUNC1_VIR2+$1FFF_{16}$	See below for details.
1	Logical Device State Vector and Interrupt Management	FUNC1_VIR2+2000_{16}: FUNC1_VIR2+$20FF_{16}$	PIO_LDSV+48x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
1	Flow Table	FUNC1_VIR2+2100_{16}, FUNC1_VIR2+2108_{16} and FUNC1_VIR2+2110_{16}	FFLP+48K+00000_{16}, FFLP+48K+00008_{16} and FFLP+48K+00010_{16}
1	DMA CSR	FUNC1_VIR2+4000_{16}: FUNC1_VIR2+$5FFF_{16}$	See below for details.
1	Logical Device State Vector and Interrupt Management	FUNC1_VIR2+6000_{16}: FUNC1_VIR2+$60FF_{16}$	PIO_LDSV+56x8K+ 8KxVRADDR{7:5} + VRADDR{4:0}
1	Flow Table	FUNC1_VIR2+6100_{16}, FUNC1_VIR2+6108_{16} and FUNC1_VIR2+6110_{16}	FFLP+56K+00000_{16}, FFLP+56K+00008_{16} and FFLP+56K+00010_{16}

The CSRs of a DMA channel fit within a 512-byte address range. When a load/store request reaches the NIU, hardware will check if a device is mapped. If so, the address is translated and the request is forwarded. Otherwise, the request will not be forwarded to the device. Invalid store requests may be silently discarded. Invalid load requests will cause an error read completion. In UltraSPARC T2, This leads to the issuance of a data access exception (*DAE_**) trap through the NCU. TABLE 23-6 shows an example of DMA mapping for an 8-Kbyte range within the virtualization

region. The least significant 9 address bits will be identical to the physical addresses of the CSRs of the DMA channels. The higher-order bits will be translated according to the DMA channel binding registers.

TABLE 23-6 Virtual Page Address Assignment Example

Address	Comment
00_{16} – $1FF_{16}$	CSRs for bound logical transmit DMA channel 0.
200_{16} – $3FF_{16}$	CSRs for bound logical receive DMA channel 0.
400_{16} – $5FF_{16}$	CSRs for bound logical transmit DMA channel 1.
600_{16} –$7FF_{16}$	CSRs for bound logical receive DMA channel 1.
800_{16} – $9FF_{16}$	CSRs for bound logical transmit DMA channel 2.
$A0_{16}$ – BFF_{16}	CSRs for bound logical receive DMA channel 2.
$C0_{16}$ – DFF_{16}	CSRs for bound logical transmit DMA channel 3.
$E0_{16}$ – FFF_{16}	CSRs for bound logical receive DMA channel 3.
100_{16} – $11FF_{16}$	CSRs for bound logical transmit DMA channel 4.
120_{16} – $13FF_{16}$	CSRs for bound logical receive DMA channel 4.
140_{16} – $15FF_{16}$	CSRs for bound logical transmit DMA channel 5.
160_{16} – $17FF_{16}$	CSRs for bound logical receive DMA channel 5.
180_{16} – $19FF_{16}$	CSRs for bound logical transmit DMA channel 6.
$1A0_{16}$ – $1BFF_{16}$	CSRs for bound logical receive DMA channel 6.
$1C0_{16}$ – $1DFF_{16}$	CSRs for bound logical transmit DMA channel 7.
$1E0_{16}$ – $1FFF_{16}$	CSRs for bound logical receive DMA channel 7.

Note that the following register is inside the management region.

TABLE 23-7 DMA Channel Binding – DMA_BIND (FZC_PIO + 10000_{16})

Bit	field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13	tx_bind	0	RW	Set to 1 to bind the following Tx channel.
12:8	tx	0	RW	Transmit DMA channel number.
7:6	—	0	RO	*Reserved*
5	rx_bind	0	RW	Set to 1 to bind the following Rx channel.
4:0	rx	0	RW	Receive DMA channel number.

When a PIO falls into one of the virtualization region for the DMA CSRs, hardware will first use the address bits 26:25, 14, 12:10 to index into the DMA channel binding register array. Bit 9 (0 for transmit, 1 for receive) will determine if it is a receive or a transmit DMA channel PIO. If the corresponding channel is bound, with the rx_bind or tx_bind bit set to 1, then the request is valid and the request will be forwarded to the respective DMA module. When forwarding the request, hardware will translate

the PIO address to the hardwired DMA channel address using the DMA channel number specified in the register, either RX or TX. The actual physical addresses for the Rx and Tx DMA channels are specified in their respective sections.

23.3 System Interrupts

There are four categories of logical devices:

- Receive DMA Channels (16 total)
- Transmit DMA Channels (16 for UltraSPARC T2)
- Ethernet MACs
- MDIO Interface (MIF)
- System Errors

The logical device number assignment is summarized in TABLE 23-8.

TABLE 23-8 Logical Device Number Assignment

Number	Device
0 –15	Receive DMA Channels 0 - 15
16–31	*Reserved*
32–55	Transmit DMA Channels 0–15; device number 48–55 reserved
56–62	*Reserved*
63	MIF
64–65	Ethernet MAC 0, 1
66–67	*Reserved*
68	Device error

Each logical device can belong to a logical device group, sharing a system interrupt. The following register set (TABLE 23-9) defines the LDG number of a logical device. The group number defined in this register set associates the system interrupt to be sent to the CPU system. Note that there are a total of 69 entries, corresponding to the above table. Reserved device number has no logical effect, though the registers exist.

TABLE 23-9 Logical Device Group Number – LDG_NUM (FZC_PIO + 20000_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	num	0	RW	Logical device group number.

A total of 64 logical device groups can be defined by software. Each group may hold an arbitrary number and combination of devices. Each logical device has up to two flag bits. Flag 0 is used for normal datapath events, such as packet arrivals, or

transmission completion. Flag 1 is used for error or noncritical datapath events. (Ethernet MACs and MIF interrupts are in flag 1 category.) These flag bits form the logical device state vector. Each system interrupt service routine, associated with a LDG, may access different register addresses to read the LDSV. Physically, there may not be any registers implemented separately for each LDG. Conceptually, only *one* single LDSV is required.

There are a total of 64 sets of register to read the LDSV. The first set is for LDG 0, the second set is for LDG 1, etc. Note that the addresses of the LDSV registers are 8 Kbytes apart and can be managed by a system manager. Note that since there is only one logical copy of the LDSV, for 32-bit systems, a read will return the appropriate 4 bytes to the CPU.

TABLE 23-10 Logical Device State Vector 0 – LDSV0 (PIO_LDSV + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:0	ldf_0	0	RO	First flag bits of logical devices 0 to 63.

TABLE 23-11 Logical Device State Vector 1 – LDSV1 (PIO_LDSV + 00008_{16})

Bit	Field	Initial Value	R/W	Description
63:0	ldf_1	0	RO	Second flag bits of logical devices 0 to 63.

TABLE 23-12 Logical Device State Vector 2 – LDSV2 (PIO_LDSV + 00010_{16})

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9:5	ldf_1	0	RO	Second flag bits of logical devices 64 to 68.
4:0	ldf_0	0	RO	First flag bits of logical devices 64 to 68.

Though all the flag bits are captured in the registers above, not all flags bits may be accessed at all the addresses. The logical device group number defined for each LD, in TABLE 23-9, is used to generate the membership information for each logical group. Conceptually, one can use a 6-to-64 decoder for each LD to generate the membership information. The *i*-th row of each decoder will indicate if the LD is part of LDG *i*.

Not all devices in a LDG may trigger a system interrupt. The following interrupt masks define which flag(s) of any LD may generate an interrupt. Together with the membership information decoded, one can then decide if a system interrupt should be issued. The first set of registers is for the DMAs, and the second set is for the MACs and system events.

TABLE 23-13 Logical Device Interrupt Mask 0 – LD_IM0 (PIO_IMASK0 + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	ldf_mask	3_{16}	RW	The flag mask bits for a logical device. Set to 0 to select the flag.

TABLE 23-14 Logical Device Interrupt Mask 1 – LD_IM1 (PIO_IMASK1 + 00000_{16}) (count 5 step 8192)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	ldf_mask	3_{16}	RW	The flag mask bits for a logical device. Set to 0 to select the flag.

To eliminate spurious interrupts, each interrupt needs to be re-armed every time to issue one interrupt. Software may also set a timer, and before the timer expires, no interrupt will be issued for that LDGI. This is programmed through the following register.

TABLE 23-15 Logical Device Group Interrupt Management – LDGIMGN (PIO_LDSV + 00018_{16}) (count 64 step 8192)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	arm	0	RW	Set to 1 to arm the LDGI. Cleared by hardware after the interrupt is issued.
30:6	—	0	RO	*Reserved*
5:0	timer	0	RW	Timer. Set by software, count down by hardware.

The resolution of the timer is set by the following register.

TABLE 23-16 Logical Device Group Interrupt Timer Resolution – LDGITMRES (FZC_PIO + 00008_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	res	0	RW	Timer resolution, in number of system clocks. A value of 0 is hardwired to behave the same as a value of 1, i.e., the minimum resolution is one tick, or the same as the system clock.

When issuing an interrupt, hardware will send the System Interrupt Data defined in TABLE 23-17 to the interrupt management unit.

TABLE 23-17 System Interrupt Data – SID (FZC_PIO + 10200_{16})

Bit	Field	Initial Value	R/W	Description
63:7	—	0	RO	*Reserved*
6:0	data	0	RW	The data sent along with an interrupt. 6:5 may be interpreted as function number, 4:0 may be used to indicate a specific interrupt. Bit 6 should always be set to 1. This value is used directly to index into the INT_MAN register set in NCU.

FIGURE 23-1 shows the logic supporting the interrupt muxing. Note that the final signals still need to be qualified by the arm bit and that the timer value equals zero.

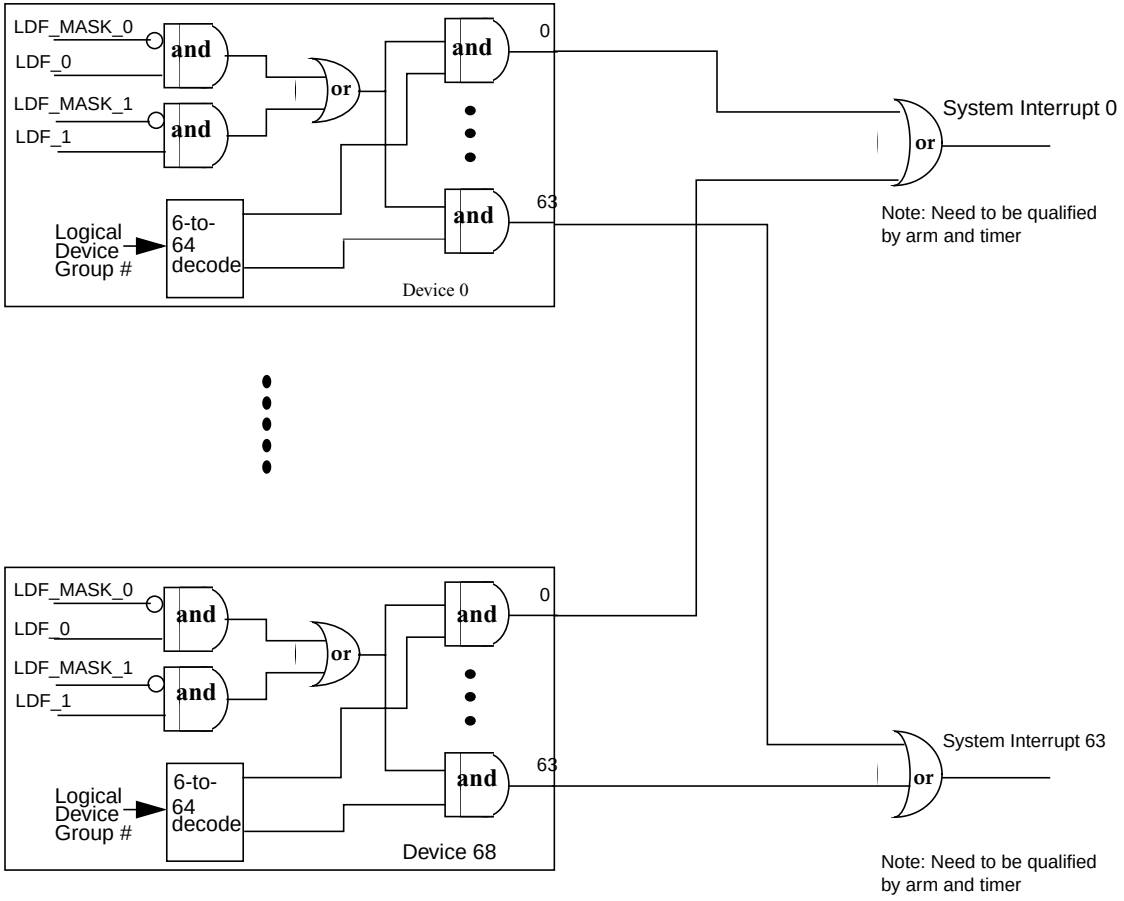


FIGURE 23-1 Interrupt Muxing Logic

The following notes describe the typical usage of interrupts.

It is useful to be able to configure the device to support either polling or interrupt mode dynamically. For example, normally the device is configured in interrupt mode, but when the compute node is under heavy load, then, it may not make much sense to be responsive to packet arrivals. For example, request packets may be queued up or even dropped since the server is already heavily processing the existing requests. Thus, by going to polling mode, less processor cycles are spent on expensive interrupt context switching. Packets will be drained by software when there are free cycles. Because of the asynchronous nature of switching from interrupt mode to polling mode, there may be one in-flight interrupt software has to handle. Also, polling mode does not necessarily mean that the device will not issue interrupts. It typically refers to masking off events associated with packet arrivals. The device may still issue an interrupt when an error occurs.

In the current design, there is a device grouping mechanism to support sharing of interrupts by multiple logical devices forming a group. The group should be viewed conceptually the same as a super device with the flags from different devices as events of the super device that can trigger interrupts. As such, the device flags can be masked or unmasked. The super device may be armed or delayed by setting a non-zero timer.

Because of the asynchronous nature of masking of events through PIOs and the actual event arrivals, interrupt may still occur after software issued a PIO to mask the event. (In some cases, depending on the exact timing of the different events, an erronous interrupt may occur.) To avoid these issues, software should use the following operational rules. 'Interrupt mode' used below refers to the hardware state when certain packet events are unmasked to enable the device to issue interrupts as normal packet handling procedure. The device is armed. 'Polling mode' refers to the hardware state when these events are masked and will not issue interrupts. In polling mode, interrupts may still be issued, when an error occurs, if those events are unmasked, and the group is armed. These states may be different from the software state perceived by higher layer software during the transition period.

1. Switching from interrupt mode to polling mode may be achieved when the armed bit is cleared. This is the case when an interrupt has been issued, and before the driver arm the group again. That is, it is save to mask out events during ISR and before arming the device.

2. When switching from polling mode to interrupt mode, driver may issue a PIO to unmask events that are previously masked.

After a group has been armed, only an interrupt event can clear the arm bit. When switching from interrupt to polling mode, rule 1 implies that either software waits until an event occured or can induce an event to issue an interrupt. When detaching the driver, it is recommanded that the driver induce an interrupt to clear the arm bit. Inducing interrupts may be achieved through the device debug interface.

23.4 Miscellaneous

This section discusses a variety of topics: reset, device error status, the Meta Arbiter, System Mux (SMX), and debug.

23.4.1 Reset

The RST_CTL register resets different parts of the subsystem.

TABLE 23-18 Reset Control – RST_CTL (FZC_PIO + 00038_{16})

Bit	Field	Initial Value	R/W	Description
63:23	—	0	RO	*Reserved*
22	mac_rst3	0	RW	*Reserved*
21	mac_rst2	0	RW	*Reserved*
20	mac_rst1	0	RW	Reset port 1 MAC. It behaves the same way as hardware reset. 0 = MAC reset is deasserted; 1 = MAC reset is asserted.
19	mac_rst0	0	RW	Reset port 0 MAC. It behaves the same way as hardware reset. 0 = MAC reset is deasserted; 1 = MAC reset is asserted.
18:12	—	0	RO	*Reserved*
11	ack_to_en	1	RW	0 = Disable ACK timeout function; 1 = Enable ACK timeout function.
10:1	ack_to_val	$3FF_{16}$	RW	Program ACK time out value for PIO clients who did not ACK any read/write transaction in time.
0	—	0	RO	*Reserved*

23.4.2 Device Error Status

The System Error Mask register collects the error state from the different blocks of the system. The logical **or** of these states equals the ldf_0 of device 68. The state of the block may be masked by the corresponding Mask register.

TABLE 23-19 System Error Mask – SYS_ERR_MASK (FZC_PIO + 00090_{16})

Bit	Field	Initial Value	R/W	Description
63:11	—	0	RO	*Reserved*
10	meta2	1	RW	Set to zero to unmask the interrupt state.
9	meta1	1	RW	Set to zero to unmask the interrupt state
8	peu	1	RW	Set to zero to unmask the interrupt state. (This should always set to 1.)

TABLE 23-19 System Error Mask – SYS_ERR_MASK (FZC_PIO + 00090_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
7	txc	1	RW	Set to zero to unmask the interrupt state.
6	rdmc	1	RW	Set to zero to unmask the interrupt state.
5	tdmc	1	RW	Set to zero to unmask the interrupt state.
4	zcp	1	RW	Set to zero to unmask the interrupt state.
3	fflp	1	RW	Set to zero to unmask the interrupt state.
2	ipp	1	RW	Set to zero to unmask the interrupt state.
1	mac	1	RW	Set to zero to unmask the interrupt state.
0	smx	1	RW	Set to zero to unmask the interrupt state.

TABLE 23-20 System Error State – SYS_ERR_STAT (FZC_PIO + 00098_{16})

Bit	Field	Initial Value	R/W	Description
63:11	—	0	RO	*Reserved*
10	meta2	0	RO	Set to 1 to indicate an error.
9	meta1	0	RO	Set to 1 to indicate an error.
8	peu	0	RO	Set to 1 to indicate an error. (This should always return 0.)
7	txc	0	RO	Set to 1 to indicate an error.
6	rdmc	0	RO	Set to 1 to indicate an error.
5	tdmc	0	RO	Always set to 0, no misc. errors from TDMC module.
4	zcp	0	RO	Set to 1 to indicate an error.
3	fflp	0	RO	Set to 1 to indicate an error.
2	ipp	0	RO	Set to 1 to indicate an error.
1	mac	0	RO	Set to 1 to indicate an error.
0	smx	0	RO	Set to 1 to indicate an error.

23.4.3 Meta Arbiter

The main function of Meta Arbiter is to arbitrate requests from read clients and write clients, assigning a transaction ID to each request. In UltraSPARC T2, requests will be sent to SMX.

There are two arbiters in Meta Arbiter: one for read and one for write. Both of them implement simple round-robin arbitration. The arbiters are independent of each other since read and write requests are two paths. There are total of 64 transaction IDs in Meta Arbiter. The read takes the lower 32 (0-31) IDs, and the write takes upper 32 (32-63). A read/write threshold register can be programmed to limit the number of pending read/write transactions. Default value is 32 for each read and

write. Meta Arbiter snoops the response bus for transaction completion. The corresponding completed transaction ID is reclaimed by Meta Arbiter and reused for new transaction.

In a timeout transaction case, Meta Arbiter puts the corresponding transaction ID to a dirty bin. A "dirty" transaction ID doesn't get reused. Meta Arbiter issues an interrupt when either all read or write transactions are dirty. Reclaim of dirty IDs is done via software.

TABLE 23-21 Dirty TID Control – DIRTY_TID_CTL (FZC_PIO + 0010_{16})

Bit	Field	Initial Value	R/W	Description
63:22	—	0	RO	*Reserved*
21:16	npthred	$1F_{16}$	RW	NP Write Threshold Value, Valid range is 1_{16} through 31_{16}, which will be the value for the number of outstanding nonposted write requests allowed at a given time.
15:10	—	0	RO	*Reserved*
9:4	rdthred	20_{16}	RW	Read Threshold Value. Range of value is from 1_{16} through 32_{16}, which will be the value for the number of outstanding read requests allowed at a given time. 0_{16} results in no outstanding transaction ID read requests); this is illegal.
3:2	—	0	RO	*Reserved*
1	dtidclr	0	RW1C	Clear dirty bin.
0	dtidenable	0	RW	Enable dirty TID algorithm.

TABLE 23-22 Dirty TID Status – DIRTY_TID_STAT (FZC_PIO + 0018_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	dtidnpwstatus	0	RO	Dirty TID status; count of how many NP Write TIDs are dirty.
7:6	—	0	RO	*Reserved*
5:0	dtidrdstatus	0	RO	Dirty TID status; count of how many read TIDs are dirty.

23.4.4 SMX

The main function of SMX is to service meta read and write requests, adjust request start address (64-byte aligned), segment each request to 64-byte chunk, and hand it off to SIU. For each read request, SMX receives and forwards responses to corresponding request client. Flow control between SIU and SMX is maintained via credits. For SMX outbound, there are 16 bypass and 16 orderQ credits. For SMX inbound, there are 4 inboundQ credits. Meta read and write requests are round-robin arbitrated at SIU request (64-byte chunk) boundary.

SMX gathers responses from SIU, and for each client handoff, it generates byte enables and transaction completion accordingly. SMX maintains a lookup table to correlate and track incoming responses.

A timeout occurs if a transaction fails to complete within a programmable timeout value. An interrupt may be generated. Since any transaction timeouts, whether caused by UEs or deadlock cases are considered to be fatal in UltraSPARC T2, NIU does not support any particular recovery mechanism for them.

TABLE 23-23 SMX Configuration Data – SMX_CFIG_DAT (FZC_PIO + 00040_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	ras_det	1	RW	Set to 1 to enable RAS error detection.
30	ras_inj	1	RW	Set to 1 to enable RAS error injection.
29:28	—	0	RO	*Reserved*
27:0	xact_to	$FF0FFFF_{16}$	RW	Transaction timeout value. Default to 50 ms. Bits 20–27 set the timeout count; bits 0–19 set the tick interval. A value of 0 in either of the two fields will disable the timer.

TABLE 23-24 SMX Interrupt Status – SMX_INT_STAT (FZC_PIO + 00048_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	stat	0	RO	Bits 6:31 = SMX internal state machine state. Bits 0:5 are reserved.

TABLE 23-25 SMX Control – SMX_CTL (FZC_PIO + 00050_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ctl	0	RW	Bits 0:2 = debug select: 0*xx* – SMX internal debug signals; 100 – training set 101 – training load. Bits 4:6 = xtb error injection: *xx*1 – inject error in one packet; *x*10 – inject error in alternate packets; 100 – inject error in all packets. Bits 8:10 = resp error injection: *xx*1 – inject error in one packet; *x*10 – inject error in alternate packets; 100 – inject error in all packets. Bits 3, 7 and 11:31 are reserved.

TABLE 23-26 SMX Debug Vector – SMX_DBG_VEC (FZC_PIO + 00058_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	vec	0	RW	Debug training vector.

23.4.5 Debug

TABLE 23-27 PIO Debug Select – PIO_DBG_SEL (FZC_PIO+00060_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	SEL	0	RW	0_{16}: 0_2, rd_ptr{4:0},wr_ptr{4:0}; 1_{16}: 0_2, ig_state{2:0}, pio_rw_state{2:0}, accepted_state{1:0}; 2_{16}: ~pio_debug_port; 3_{16}: training vector.

TABLE 23-28 PIO Training Vector – PIO_TRAIN_VEC (FZC_PIO + 00068_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	vec	0	RW	Used by software to program to set the training vector for testmux calibration. Writing a value of 03_{16} in the debug select register would load this value into testmux logic. Writing a value of 02_{16} in the debug select logic would generate alternating pattern on the debug port once the PIO block is selected.

TABLE 23-29 PIO Arbiter Control – PIO_ARB_CTL (FZC_PIO + 00070_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ctl	0	RW	

TABLE 23-30 PIO Arbiter Debug Vector – PIO_ARB_DBG_VEC (FZC_PIO + 00078_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	vec	0	RW	

23.5 Errata

23.5.1 Spurious Interrupt

If the interrupt usage model as described in 23.3 is not followed, undesirable spurious interrupts may result.

CHAPTER 24

Network Interface Unit: Receive Packet Classification

24.1 Ethernet MAC Subsystem

The register definitions for the MAC are defined in a separate chapter. The following signals are logically available from the MAC controllers.

- Port number
- Address information: The index of the matching address, individual or group; for group addresses, the group filter result. It also indicates if it is a broadcast address.
- Layer 2 error status.

If the received frame has an address match, which can be either individual or group address, the match index and the port number are used to determine the RDC Table number. For frames that do not have an address match but pass the group filter check, an RDC table number and the preference is also specified. This information is also passed forward to the header parser to continue the classification process.

24.2 Layer 2 Classification

The Layer 2 parser will determine the following information from the frame.

1. If the frame is a VLAN packet, the VLAN ID.
2. Ethernet format, whether there is a llc/snap field.

For VLAN frames, the VLAN ID is used to index into a VLAN table, defined by the following register, to determine the RDC table number. Note that hardware does not perform parity check on PIOs. Software is required to initialize this table with proper parity value for hardware to function. See TABLE 24-1 for a description on setting the values.

TABLE 24-1 Ethernet VLAN Table – ENET_VLAN_TBL (FZC_FFLP + 00000_{16}) (count 4096 step 8)

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	parity0	X	RW	Even parity over bits 0 to 7.
15:8	—	X	RO	*Reserved*
7	vpr1	X	RW	Set to 1 to indicate the preference for port 1.
6:4	vlanrdctbln1	X	RW	Receive DMA channel table number for port 1.
3	vpr0	X	RW	Set to 1 to indicate the preference for port 0.
2:0	vlanrdctbln0	X	RW	Receive DMA channel table number for port 0.

Since both the VLAN table and the MAC address table can set the preference, the following logic is used to resolve the final preference. Note that an invalid VLAN should have its vpr bit set to 0. In this case, the MAC preference is used. For example, on power-on, if no VLAN is assigned, the value of all vpr bits should be set to zero. Since we are supporting even parity, this means the entire table should be initialized to zero. If a valid VLAN is assigned, the vpr bit may be set to 1, depending on the classification requirement, and with appropriate DMA table number. The parity should be calculated accordingly.

TABLE 24-2 MAC Address / VLAN Preference Logic

mpr	vpr	Selection
0	0	MACRDCTBLN
0	1	VLANRDCTBLN
1	0	MACRDCTBLN
1	1	VLANRDCTBLN

Note that at the end of Layer 2 parsing, an RDC table number is always associated with the incoming frames the parser recognizes. Frames with errors and frames that do not have a valid MAC address will use the port default Receive DMA channel if they are not discarded. We discuss this in a later section. The Layer 2 result will be passed to the Layer 3 parser.

24.3 Layer 2/3/4 Classification

After the initial classification, hardware continues to determine the class of the received frame. Associated with each class is a Flow Key template and a TCAM Key template if the received frame is recognized by the hardware. First, based on the Ether-Type value, it will determine if the frame is one of two programmed EtherTypes (ARP or RARP), an IP (v4 or v6) frame, or others. For non-IP frames with recognized EtherType, hardware will send the class code, along with the next 11 bytes (from the EtherType) to the TCAM for a match. For IPv4 or IPv6, the parser will determine if further classification is needed, based on the Protocol ID/Next Header field and/or the `tos/dscp` byte. A number of the protocol IDs are hardwired. Software can specify up to four programmable protocol fields hardware can match, and they take precedence over hardwired match. A lower number class has a higher precedence.

TABLE 24-3 specifies the class code and describes the packet type.

TABLE 24-3 Class Code Definition

Class Code	Description	Notes
0_{16}	Hardware does not recognize the packet class.	No Key generated. This encoding indicates an invalid entry.
1_{16}	Dummy Class	Used for in-service testing. No key generated.
2_{16}	Programmable EtherType 1.	Send 11 bytes after EtherType to TCAM. No flow key generated.
3_{16}	Programmable EtherType 2.	Send 11 bytes after EtherType to TCAM. No flow key generated.
4_{16}–7_{16}	User programmable, based on protocol ID/Next Header and TOS/DSCP.	
8_{16}	TCP over IPv4.	
9_{16}	UDP over IPv4.	
A_{16}	AH or ESP over IPv4.	
B_{16}	SCTP over IPv4.	
C_{16}	TCP over IPv6.	
D_{16}	UDP over IPv6.	
E_{16}	IP v6 with Next Header AH or ESP.	
F_{16}	SCTP over IPv6.	
10_{16}	ARP	Send 11 bytes after EtherType to TCAM. No flow key generated.
11_{16}	RARP	Send 11 bytes after EtherType to TCAM. No flow key generated.
12_{16} – $1F_{16}$	Dummy Class	Used for in-service testing. No key generated.

To specify the EtherType values for class 2 and 3, the following registers are used. TABLE 24-4 describes the register for class 2; TABLE 24-5 describes the register for class 3.

TABLE 24-4 Layer 2 Class – L2_CLS (FZC_FFLP + 20000_{16}) (count 2 step 8)

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	vld	0	RW	Set to 1 to indicate the EtherType is valid.
15:0	etype	0	RW	EtherType value.

The following registers specify the Protocol ID (Next Header) and the tos (dscp) fields for the programmable classes 4 to 7. The first register is for class 4, second for class 5 and so on.

TABLE 24-5 Layer 3 Class – L3_CLS (FZC_FFLP + 20010_{16}) (count 4 step 8)

Bit	Field	Initial Value	R/W	Description
63:26	—	0	RO	*Reserved*
25	valid	0	RW	Set to 1 to indicate the entry is valid.
24	ipver	0	RW	0 = v4; 1 = v6.
23:16	pid	0	RW	Protocol ID or Next Header
15:8	tosmask	0	RW	Mask bits for TOS. Set a bit to 1 to specify that the corresponding bit in tos field is required in class matching.
7:0	tos	0	RW	TOS field.

Note that classes 4_{16} to F_{16} always have both a Flow key and a TCAM key associated with them. Hardware will first check if a TCAM search is required (as specified by the tsel bit below). If so, a TCAM match is initiated. If there is a hit, the associated RAM entry will specify how to continue the classification. If TCAM match is not required or there is *no* matching entry, a flow match will be initiated.

For classes 2_{16}, 3_{16}, 10_{16}, and 11_{16}, there is no flow key associated. If there is no matching entry in the TCAM, the classification process will terminate, and the packet will be directed to the default RDC associated with the RDC group.

The following registers specify how to build the TCAM Key for IP packets. If tsel bit is set to zero, hardware will not issue a TCAM match. The first register address is for class 4, and the last one for class F_{16}. For IPv6, a 4-tuple match can be initiated. For IPv4, hardware can initiate a 5-tuple match. Note that for the programmable protocol type, hardware uses the ipver bit in register L3_CLS to determine the criteria for a class. That is, class definition is IP version specific.

A v6 4-tuple plus the TOS byte consists of

- IP source or destination address
- IP protocol ID
- L4 source and destination port number or the SPI number

- TOS byte

A v4 5-tuple plus the TOS byte consists of

- IP source address
- IP destination address
- IP protocol ID
- IP source and destination port number or the SPI number
- TOS byte

TABLE 24-6 TCAM Key – TCAM_KEY (FZC_FFLP + 20030_{16}) (count 12 step 8)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3	disc	0	RW	Discard. Set to 1 to discard all frames of this class. The rest of the fields will not be interpreted.
2	tsel	0	RW	Set to 1 to indicate a TCAM lookup is required.
1	—	0	RO	*Reserved*
0	ipaddr	0	RW	For v6 4-tuple match, set to 1 to select source address. Set to 0 to select destination address. This bit will not be interpreted for 5-tuple match.

The objective of flow classification is to generate an offset into the RDC table to select the Receive DMA channel. The following registers specify how to build the Flow Key. The first register address is for class 4, and the last one for class F_{16}.

TABLE 24-7 Flow Key – FLOW_KEY (FZC_FFLP + 40000_{16}) (count 12 step 8)

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9	port	0	RW	Port number, set to 1 to select.
8	l2da	0	RW	Ethernet destination MAC address. Set to 1 to select.
7	vlan	0	RW	VLAN Tag. Set to 1 to select.
6	ipsa	0	RW	IP source address. Set to 1 to select.
5	ipda	0	RW	IP destination address. Set to 1 to select.
4	proto	0	RW	Protocol IP or Next Header. Set to 1 to select.
3:2	l4_0	0	RW	Select to extract the first and second bytes after the IP header or the 5th and 6th bytes: 00 – Not select; 10 – 1st and 2nd; 11 – 5th and 6th; 01 – *Reserved*.
1:0	l4_1	0	RW	Select to extract the third and forth bytes after the IP header or the 7th and 8th bytes: 00 – Not select; 10 – 3rd and 4th; 11 – 7th and 8th; 01 – *Reserved*.

The key template is shown in TABLE 24-8. Unused fields or fields not available will be filled with zeros. For IPv4, the source/destination addresses are only 32 bits in length and will occupy the least significant 32 bits. The rest of the address field will be filled with zeros. Note that the VLAN valid field is set to 1111_2 for packets with a

VLAN tag. For packets that do not have a VLAN tag, this field is set to zeros. Logically, the first bit that enters the theoretical CRC engine is bit 63 of the last 64-bit block, and the last bit is bit 0 of the first 64 bits or the first “VLAN valid” bit.

TABLE 24-8 Flow Key Template

VLAN valid	L2 DA (48bits)			VLAN ID (12)
IP Source Address {127:96}		IP Source Address {95:64}		
IP Source Address {63:32}		IP Source Address {31:0}		
IP Destination Address {127:96}		IP Destination Address {95:64}		
IP Destination Address {63:32}		IP Destination Address {31:0}		
L4-0 (16)	L4-1(16)	PID (8)	Port (2)	22 Zeros

Note that as long as the class is valid and recognized by hardware, the hash values may be reported to software through the full header if enabled.

24.3.1 TCAM Software Interface

Each TCAM entry has two fields: a key and a mask. The key field stores the pattern to be compared. The mask field, together with the key field, encodes which bit is a don’t care term. Since multiple entries may generate a hit, a priority encoder on the match address(es) is needed to determine a unique hit; the lower the address, the higher the priority.

If there is a match, the match entry address is used to index into a table which stores the classification information.

Physically, the internal TCAM has only a single interface port. This port is shared by the CPU and the receive logic. The current TCAM design is 200 bits wide. 128 entries are supported. The description below is for programming the TCAM directly.

Each entry is composed of a key and a mask. Internally, hardware keeps a temporary register set to hold the TCAM entry. CPU then triggers a read/write/compare access to the TCAM by writing the address to the control register. Note that the same register is used to access the TCAM associated data. The compare command is for testing of the TCAM.

On power-on, software needs to initialize all the TCAM entries to contain class code 0_{16}. The write and compare operations are considered *atomic*, in that the hardware should not send a request to the TCAM from the packet classifier in the middle of a write or a compare operation. This will prevent any inconsistency in the TCAM programming.

TABLE 24-9 TCAM Key 0 – TCAM_KEY_0 (FZC_FFLP + 20090_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	key	0	RW	Bits 192–199 of the TCAM entry.

TABLE 24-10 TCAM Key 1 – TCAM_KEY_1 (FZC_FFLP + 20098_{16})

Bit	Field	Initial Value	R/W	Description
63:0	key	0	RW	Bits 128–191 of the TCAM entry, or the data value in the associated RAM for TCAM results.

TABLE 24-11 TCAM Key 2 – TCAM_KEY_2 (FZC_FFLP + $200A0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	key	0	RW	Bits 64–127 of the TCAM entry.

TABLE 24-12 TCAM Key 3 – TCAM_KEY_3 (FZC_FFLP + $200A8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	key	0	RW	Bits 0–63 of the TCAM entry.

TABLE 24-13 TCAM Key Mask 0 – TCAM_KEY_MASK_0 (FZC_FFLP + $200B0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	key_sel	0	RW	Bits 192–199 of the TCAM mask. Set 1_{16} to the corresponding bit to select the bit to be matched.

TABLE 24-14 TCAM Key Mask 1 – TCAM_KEY_MASK_1 (FZC_FFLP + $200B8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	key_sel	0	RW	Bits 128–191 of the TCAM mask. Set 1_{16} to the corresponding bit to select the bit to be matched.

TABLE 24-15 TCAM Key Mask 2 – TCAM_KEY_MASK_2 (FZC_FFLP + $200C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	key_sel	0	RW	Bits 64–127 of the TCAM mask. Set 1_{16} to the corresponding bit to select the bit to be matched.

TABLE 24-16 TCAM Key Mask 3 – TCAM_KEY_MASK_3 (FZC_FFLP + $200C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:0	key_sel	0	RW	Bits 0–63 of the TCAM mask. Set 1_{16} to the corresponding bit to select the bit to be matched.

When the TCAM Control register is written to, hardware will initiate a read/write/compare to the TCAM or its associated data memory. For writes to TCAM, the data in the TCAM Key and Mask registers will be written to the TCAM at the location specified at loc. For reads from TCAM, the data at TCAM location loc will be retrieved and stored to the TCAM Key and Mask registers. The compare command will initiate a TCAM match with the data stored in the TCAM_KEY registers. The match bit will be set to 1 if there is a TCAM match, and the matching location will be reported in the loc field. To access the TCAM associated RAM, the address of the entry is specified in loc and the TCAM_KEY_1 register is used as the data register. The status of the operation (TCAM or TCAM associated RAM) is signaled through the stat bit of the control register.

TABLE 24-17 TCAM Control – TCAM_CTL (FZC_FFLP + $200D0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:18	rwc	0	RW	000_2 – Specifies a TCAM write; 001_2 – Specifies a TCAM read; 010_2 – For TCAM compare; 011_2 – *Reserved*; 100_2 – TCAM associated RAM write; 101_2 – TCAM associated RAM read; 110_2, 111_2 – *reserved*.
17	stat	1	RW	Status of the read or write operation. When zero is written to this bit, hardware will initiate the action, and this bit will be set to 1 when the operation completes.
16	match	0	RW	Set to 1 if there is a TCAM match for compare command. 0 otherwise.
15	—	0	RO	*Reserved*
14:10	—	0	RO	*Reserved*
9:0	loc	0	RW	TCAM or TCAM associated RAM location. For compare, this is the location of the match.

The KEY/MASK formats for TCAM are described in TABLE 24-18. Note that for IPv4 formats, a noport bit indicates whether the incoming packet is a fragment or whether the hardware cannot capture the L4 port number. The noport condition is defined as follows.

- More fragment bit is set *or* IP Offset is non-zero.

The class code *must* be programmed to be 0 to indicate an invalid entry. The reserved fields *must* be encoded as don't care terms, that is, with the corresponding key_sel bit set to 0.

TABLE 24-18 IP v4 5-Tuple Entry

Bit	Field	Comment
199:195	cls_code	Class code. Zero indicates invalid, set to the appropriate class value for the entry.
194:190	—	*Reserved*, must be set to don't care.
189:187	l2rdc_tbl_num	Layer 2 RDC Table Number
186	noport	No L4 port number.
185:112	—	*Reserved*, must be set to don't care.
111:104	tos	TOS byte.
103:96	pid	Protocol ID.
95:64	l4pt_spi	Either L4 port numbers or SPI.
63:32	ip_addr_sa	IP source address.
31:0	ip_addr_da	IP destination address.
200	*Total*	

For IPv6, the format is described in TABLE 24-19.; the EtherType format is described in TABLE 24-20.

TABLE 24-19 IP v6 4-Tuple Entry

Bit	Field	Comment
199:195	cls_code	Class code. Zero indicates invalid, set to the appropriate class value for the entry.
194:190	—	*Reserved*, must be set to don't care.
189:187	l2rdc_tbl_num	Layer 2 RDC table number.
186:176	—	*Reserved*, must be set to don't care.
175:168	tos	TOS byte.
167:160	nxt_hdr	IP v6 next header.
159:128	l4pt_spi	Source/destination port or SPI.
127:0	ip_addr	IP v6 source or destination address.
200	*Total*	

TABLE 24-20 EtherType Format

Length (bits)	Field	Comment
199:195	cls_code	Class code. Zero indicates invalid, set to the appropriate class value for the entry.
194:192	—	*Reserved*, must be set to don't care.
191:104	eframe	First 11 bits after EtherType.
103:0	—	*Reserved*, must be set to don't care.
200	*Total*	

If there is no TCAM match, the classification process will continue and the L2 RDC table number is used to select the DMA group for flow lookup.

When there is a TCAM match, the match address is used to index into the following table to retrieve the classification result. The tres field indicates whether and how to continue the classification process. Here, software may terminate the classification process and use the result specified or simply override the RDC table number. If the disc bit is set, then all frames resulting in a match in TCAM will be discarded. Hardware will also clear the age bit. (This bit may be written to 1 by software to determine if an entry has not been referenced for a period of time.)

Periodically, software should read the TCAM entries and compare with an image stored in the system memory. In addition, though the soft error probability is very low, software should check the TCAM matching criteria as indicated by the index. If the TCAM entries are used for IPv4 exact matching, software can set the v4_ecc_ck bit and include the syndrome bits in the associated data.

The ECC syndrome is calculated over the following bits.

{ associated_data{25:1}, 0_2, cam_key{103:0} }

The following is the format for the associated data.

TABLE 24-21 Format for TCAM Associated Data

Bit	Field	Description
63:42	—	*Reserved*
41:26	syndrome	ECC syndrome for IPv4 5-tuple, if V4_ECC_CK bit is not set to 1, bits 29:26 contain the even parity for the associated data. Bit 26 for bits 7:0, bit 27 for 15:8, bit 28 for 23:16, bit 29 for 25:24.
25:14	zfid	*Reserved*
13	v4_ecc_ck	Set to 1 to enable ECC error checking on the TCAM key for v4 packets. If errors are detected, the parsing operation will terminate and the RDC table from VLAN/MAC processing will be used. The offset within the table is set to zero. For v6 packets, this bit is ignored, and the syndrome bits are treated as parity. If this bit is set to zero, see description at syndrome.
12	disc	Set to 1 to discard the packet. The classification process will terminate. The value in the rest of the register will be ignored.
11:10	tres	TCAM result: 01_2 – Use the offset specified, and terminate flow lookup; 10_2 – Override the L2 RDC Table number and continue the flow lookup; 11_2 – Override the L2 RDC Table number *and* use the offset specified, terminate flow lookup; 00_2 – Use L2 RDC table number and continue the flow lookup.
9:7	rdctbl	RDC table number.
6:2	offset	Offset queue number.
1	zfvld	*Reserved*.
0	age	Hardware will clear this bit if the entry has a match during a packet lookup. PIO has no effect.

If flow lookup is terminated, zeros will be returned to software in the usr_info field in the packet header.

To filter out fragmented IPv4 packets, software needs to set one single entry in the TCAM, with only the noport bit set to 1 (*only* the respective key_sel bit should be set to 1). This entry should be put at the lowest location where IPv4 classification starts. This forces all fragmented v4 packets to match. At the corresponding TCAM associated memory location, software should specify selecting a specific default offset (with tres set to 01_2). The offset may be set to 0, which is the hardware assumed default offset. This will direct all fragmented IPv4 packets to their respective group default.

If desired, software may program a different behavior for some fragmented packets by creating entries with lower addresses than the above-mentioned location.

24.3.2 Flow Classification and Software Interface

The hashing algorithm is based on Polynomial Hashing with CRC-32C is supported. The last 4 bits of the value is used to index into the RDC table to lookup a DMA channel.

$X^{32} + X^{28} + X^{27} + X^{26} + X^{25} + X^{23} + X^{22} + X^{20} + X^{19} + X^{18} + X^{14} + X^{13} + X^{11} + X^{10} + X^9 + X^8 + X^6 + 1$

The current hardware implementation is based on an optimization that does not require "shifting" an additional *n* (the degree of the polynomial) zeros after the key as in the serial conceptual implementation. The initial value to be programmed into the hardware should be equal to the remainder of theoretical initial value (with *n* zeros appended) divided by the polynomial, using the serial implementation.

TABLE 24-22 H1 Polynomial – H1POLY (FZC_FFLP + 40060_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	Reserved.
31:0	initval	0	RW	Initial value for Hash 1.

The following register set is reserved, and has no effect. The ext bit must be set to zero.

TABLE 24-23 Flow Partition Select – FLW_PRT_SEL (FZC_FFLP + 40070_{16}) (count 8 step 8)

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	ext	0	RW	*Reserved*, set to 0.
15:13	—	0	RO	*Reserved*

TABLE 24-23 Flow Partition Select – FLW_PRT_SEL (FZC_FFLP + 40070_{16}) (count 8 step 8)

Bit	Field	Initial Value	R/W	Description
12:8	mask	0	RW	*Reserved.*
7:5	—	0	RO	*Reserved*
4:0	base	0	RW	*Reserved.*

The following two registers are reserved. Software should not access these registers. Note that the access to external hash table is conditioned by the ext bit in the previous register.

TABLE 24-24 Hash Table Address – HASH_TBL_ADDR (FFLP + 00000_{16}) (count 8 step 8192)

Bit	Field	Initial Value	R/W	Description
63:24	—	0	RO	*Reserved*
23	autoinc	0	RW	*Reserved.*
22:0	addr	0	RW	*Reserved.*

TABLE 24-25 Hash Table Data – HASH_TBL_DATA (FFLP + 00008_{16}) (count 8 step 8192)

Bit	Field	Initial Value	R/W	Description
63:0	data	0	RW	*Reserved.*

24.4 Header Parser RDC Selection (FFLP)

In this section, we summarize the RDC selection by the Hardware Parser (FFLP) based on the packet header. This includes Hardware Parser error behavior and promiscuous mode. Note that this may be changed by the DMA logic if the packet has checksum or L2 CRC error. Hardware Parser does not take these errors into the processing.

In non-promiscuous mode, packets with no MAC-level match will be discarded. For packets that have a MAC-level match, an RDC group will be defined. Normally, the Hardware Parser (FFLP) will further refine the classification. If VLAN table has a hardware soft error, the MAC-level group default (the 0-th entry associated with the group select at the MAC) will be used. If TCAM logic has a hardware error, the

MAC or VLAN group default will be used, depending on the preference programmed by software. In any hardware error case, there will be a signal to the DMA engine to set the completion status of the packet.

In promiscuous mode, packets with MAC level match will be processed the same way as described above. (Hardware parser may mask out the promiscuous bit by the match logic.) All other packets will be sent to the default DMA channel associated with the physical port.

The following pseudocode summarize the hardware classification process.

```
if (!MAC_addr_match) {
   MAC promiscuous mode, use per port RDC; exit;
} else {
   if (!VLAN_tagged) {
      L2_RDC_table_num = MAC_RDC_table_num; goto L2-3-4;
   } else {
      use (port, VLANID) as index to VLAN table to lookup
      VLAN_RDC_table_num;
      if (VLAN_parity_err) {
         L2_RDC_table_num = MAC_RDC_table_num; goto L2-3-4; }
      if (VPR==1) {
         L2_RDC_table_num = VLAN_RDC_table_num;
      } else {
         L2_RDC_table_num = MAC_RDC_table_num; }
      }
   }

label - L2-3-4:

   if (!Class_match) {
      RDC_table_num = L2_RDC_table_num; offset = 0; exit; }
   if (4<=class<=15) {
      if (DISC) { set drop_pkt = 1; exit; }
      if (TCAM_lookup_not_required) goto flow_1;
   }
   start_TCAM_search;
   if (!TCAM_match) goto flow_1;
   read_TCAM_assoc_data;
   if (TCAM_assoc_dat_DISC==1) { set drop_pkt = 1; exit; }
   if (TCAM_assoc_dat_TRES[1]==1) {
      RDC_table_num = L2_RDC_table_num;
   } else { RDC_table_num = TCAM_assoc_dat_table_num; }
   if (TCAM_assoc_dat_TRES[0]==1) {
      offset = TCAM_assoc_dat_RDC_table_offset; exit; }
   if (!(4<=class<=15)) { offset = 0; exit;}
   goto flow_2;

label - flow_1:
```

```
    RDC_table_num = L2_RDC_table_num;
    if (!(4<=class<=15)) { offset = 0; exit;}

label - flow_2:

offset = H1[3:0];exit;

}
```

24.5 Checksum Offload

The hardware supports TCP/UDP payload checksum in the receive datapath for nonfragmented packets.

24.6 Receive Packet Header Format and Alignment

Receive Header Format Classification result will be prepended to the packet and stored in the packet buffer. The formats are shown below. Note that the entire full header is 18 bytes. It is divided into Header 0 and Header 1. Software may select only the two-byte Header 0 to be sent as header, by clearing the full_hdr bit in registers for configuring the receive DMAs. The full_hdr bit setting is per DMA. The following formats are defined in address order. B0 is in low address.

TABLE 24-26 Receive Packet Header 0 Format

Field	Position		Description
	Byte	Bit	
inputport	0	7:6	*Physical* input port number.
maccheck	0	5	Set to 1 to indicate MAC address check passed. For individual address, this indicates that the destination MAC address matches one of the programmable value. For group address, the corresponding multicast filter entry is set.
class	0	4:0	Class value.
vlan	1	7	Set to 1 if vlan field is detected. Set to 0 if l2par = 0.
llcsnap	1	6	Set to 1 if the L2 frame contains an LLC/SNAP header. Set to 0 if l2par = 0.
noport	1	5	Set to 1 to indicate hardware cannot reach the L4 ports or SPI for IPSec.

TABLE 24-26 Receive Packet Header 0 Format *(Continued)*

Field	Position		Description
	Byte	Bit	
badip	1	4	This bit will be forced to zero if class does not indicate an IP packet (see class code description). If this bit is set, it indicates hardware cannot process the IP packet. This includes the following case(s): IHL < 5. In this case, the packet will be sent to the Layer 2 default DMA.
tcamhit	1	3	Set to 1 if a TCAM search is successful.
tres	1	2:1	Same as tres bits in TCAM_RES register. If tcamhit bit is zero, tres bits will be set to zero.
tzfvld	1	0	Same bit as zfvld bit in tcam_res register; always set to zero if tcamhit is 0.

TABLE 24-27 Receive Packet Header 1 Format

Field	Position		Description
	Byte	Bit	
hwrsvd	0		*Reserved* for HW use.
tcammatch	1		TCAM match index. If class ≠ 0 and tcamhit = 1, this is the TCAM match index; set to zero otherwise. Only the least significant 7 bits are valid; higher order bit is set to 0.
—	2	7:6	*Reserved* for hardware.
hashit	2	5	Set to 1 for Hash hit.
exact	2	4	Set to 1 to indicate an exact match in hash table lookup is resulted. Set to 0 if hashit = 0.
hzfvld	2	3	Set to 1 if Hash table entry indicates a zero copy flow ID is available.
hashsidx	2	2:0	Subindex into the Hash Table Entry, set to zero if hashit = zero.
—	3	3	Reserved for zero copy function.
—	4	7:4	*Reserved* for hardware use.
zflowid	4	3:0	Bits 11:8 of zero copy flow ID. Valid if either tzfvld or hzfvld is set. Set to 0 otherwise. For TCAM match, it is the TCAM entry number; for hash table, it is the value stored in the hash table. The unused higher order bits are set to zero.
zflowid	5		Bits 7:0 of zero copy flow ID.
hashval2	6		Bits 15:8 of hash value, H2. Valid when class indicates an IP packet, zero otherwise.
hashval2	7		Bits 7:0 of hash value, H2. Valid when class indicates an IP packet, zero otherwise.
—	8	7:4	*Reserved* for hardware.

TABLE 24-27 Receive Packet Header 1 Format *(Continued)*

Field	Position		Description
	Byte	Bit	
hashval1	8, 9, 10	3:0	Hash value, H1. Valid when class indicates an IP packet, zero otherwise. Bits 19:16 are in byte 8; 15:8 in byte 9; 7:0 in byte 10.
usrdata	11:15		Note: Since external lookup is no longer supported, this field is reserved. (Old text: Bits 39:0 of user data, valid if tres is either 00_2 or 10_2. Set to zero otherwise. 39:32 in byte 11; 31:24 in byte12; 23:16 in byte 13; 15:8 in byte 14; 7:0 in byte 15.)

The following shows an example of a non-jumbo packet format in address order. A packet buffer is at least 64 bytes naturally aligned. There may be multiple of 64-byte blocks reserved by software. The packet header starts after the reserved area and is immediately followed by the packet. As mentioned earlier, the packet header can be 2 bytes or 18 bytes. In either case, the IP packet will always be 4-byte aligned. The multiple 64-byte offset will be discussed in a later section.

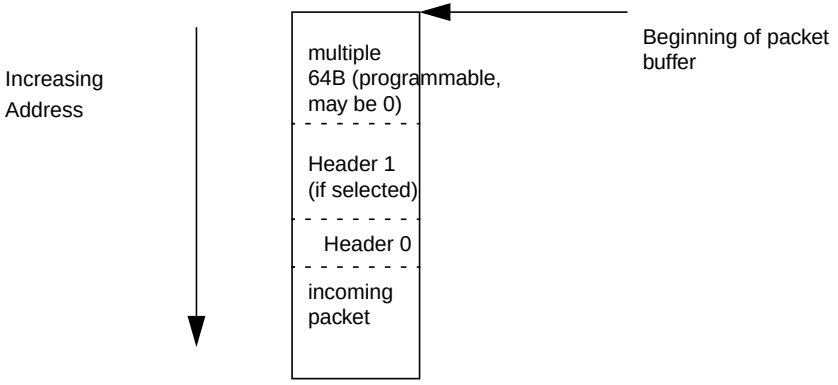


FIGURE 24-1 Packet Buffer Layout

24.7 FFLP Hardware Control Registers

The following registers are used to configure the FFLP.

TABLE 24-28 FFLP Configuration 1 – FFLP_CFG_1 (FZC_FFLP + 20100_{16})

Bit	Field	Initial Value	R/W	Description
63:27	—	0	RO	*Reserved*
26	tcam_dis	0	RW	Set to 1 to disable TCAM, all TCAM search results will be treated as no match.
25:23	pio_dbg_sel	0	RW	000_2 –101_2: debug_data 110_2: debug_training_vector 111_2: ~fflp_debug_port
22	pio_fio_rst	0	RW	*Reserved*.
21:20	pio_fio_lat	0	RW	*Reserved*.
19:16	camlat	0	RW	CAM search latency, Should be set to 4.
15:12	camratio	0	RW	CAM access ratio, similar to bits 11:8.
11:8	fcramratio	0	RW	Reserved.
7:4	fcramoutdr	0	RW	*Reserved.*
3	fcramqs	0	RW	*Reserved.*
2	errordis	0	RW	Set to 1 to disable error check in TCAM.
1	fflpinitdone	0	RW	FFLP initialization done.
0	llcsnap	0	RW	Set to 1 to enable LLCSNAP packets.

TABLE 24-29 FFLP Debug Training Vector – FFLP_DBG_TRAIN_VCT (FZC_FFLP + 20148_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	vector	0	RW	Debug training vector.

TABLE 24-30 TCP Control Flag Mask – TCP_CFLAG_MSK (FZC_FFLP + 20108_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	mask	0	RW	Set to 1 to mask out the correspondent control bit. This field includes 6 control bits and 6 reserved bits in TCP header. This bit selects the bit to be propagated to DMA controller. This should be set to select the sync bit only for the DMA to pick a different WRED parameter.

TABLE 24-31 FCRAM Refresh Timer – FCRAM_REF_TMR (FZC_FFLP + 20110_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	max	0	RW	*Reserved.*
15:0	min	0	RW	*Reserved.*

The following registers are reserved. Access has no effect.

TABLE 24-32 FCRAM Controller Address – FCRAM_FIO_ADDR (FZC_FFLP + 20118_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	addr	0	RW	*Reserved.*

TABLE 24-33 FCRAM Controller Data – FCRAM_FIO_DAT (FZC_FFLP + 20120_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	data	0	RW	*Reserved.*

TABLE 24-34 FCRAM PHY Read Latency – FCRAM_PHY_RD_LAT (FZC_FFLP + 20150_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	lat	0	RO	*Reserved.*

24.8 Error Registers

The FFLP_VLAN_PAR_ERR register logs the VLAN table parity error. Software writes 0's to clear. This register logs errors only during packet lookup.

TABLE 24-35 VLAN Parity Error – FFLP_VLAN_PAR_ERR (FZC_FFLP + 08000_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	err	0	RW	Set to 1 to indicate an error; software writes 0 to clear.
30	m_err	0	RW	Set to 1 to indicate multiple error; software writes 0 to clear.
29:18	addr	0	RW	Address within the block where error occurred.
17:0	data	0	RW	Data read from memory.

The following register logs the TCAM error during packet lookup. The error log needs to be cleared by software.

TABLE 24-36 TCAM Error – TCAM_ERR (FZC_FFLP + $200D8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	err	0	RW	Set to 1 to indicate TCAM error, either in the associated data or TCAM when detected. Software writes 0 to clear.
30	p_ecc	0	RW	Set to 1 to indicate if this is a v4 ECC error; 0 for parity error
29	mult	0	RW	Set to 1 to indicate multiple lookup error
28:24	—	0	RO	*Reserved*
23:16	addr	0	RW	Entry where the error occurs.
15:0	syndrome	0	RW	Hardware-calculated syndrome or parity value for the first error.

The FFLP Error Mask register defines which event will be reported. A logical **or** of the errors with the mask bit cleared will be signaled to the PIO block for error reporting.

TABLE 24-37 FFLP Error Mask – FFLP_ERR_MSK (FZC_FFLP + 20140_{16})

Bit	Field	Initial Value	R/W	Description
63:11	—	0	RO	*Reserved*
10:3	hsh_tbl_dat	FF_{16}	RW	*Reserved*, should always set to FF_{16}.
2	hsh_tbl_lkup	1	RW	*Reserved*, should always set to 1_{16}.
1	tcam	1	RW	TCAM lookup. Set to 0 to enable the event reporting.
0	vlan	1	RW	*Reserved*

The following registers are reserved and should always set at the default value.

TABLE 24-38 Hash Table Data Error log – HASH_TBL_DATA_LOG (FFLP + 00010_{16}) (count 8 step 8192)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	err	0	RW	*Reserved*
30:8	addr	0	RW	*Reserved*
7:0	syndrome	0	RW	*Reserved*

TABLE 24-39 Hash Table Lookup Error Log 1 – HASH_LOOKUP_ERR_LOG1 (FZC_FFLP + $200E0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3	err	0	RW	*Reserved*
2	mult_lk	0	RW	*Reserved*
1	cu	0	RW	*Reserved*
0	mult_bit	0	RW	*Reserved*

TABLE 24-40 Hash Table Lookup Error Log 2 – HASH_LOOKUP_ERR_LOG2 (FZC_FFLP + $200E8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:31	—	0	RO	*Reserved*
30:11	h1	0	RW	*Reserved*
10:8	subarea	0	RW	*Reserved*
7:0	syndrome	0	RW	*Reserved*

TABLE 24-41 FCRAM Error Test 0 – FCRAM_ERR_TST0 (FZC_FFLP + 20128_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	synd	0	RW	*Reserved*

TABLE 24-42 FCRAM Error Test 1 – FCRAM_ERR_TST1 (FZC_FFLP + 20130_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat	0	RW	*Reserved*

TABLE 24-43 FCRAM Error Test 2 – FCRAM_ERR_TST2 (FZC_FFLP + 20138_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat	0	RW	*Reserved*

CHAPTER 25

Network Interface Unit: Receive DMA

There is a total of 16 receive DMA channels in hardware. These channels store incoming packets. Note that the association of PIO control to a DMA, if accessed through the virtualization region, is defined in the PIO block. How addresses are relocated and the association of these requests with a function is defined per DMA.

A receive DMA channel consists of a Receive Block Ring (RBR) and a Receive Completion Ring (RCR). The addresses stored in these structures as well as the definition of the data structure are subjected to a translation defined by the logical pages. From the different per-port logical FIFOs, the port scheduler needs to determine which port with an available packet to service first. A packet is available when the entire packet is in the receive FIFO and the classification result is in the control FIFO. Since the rate may differ from port to port, a DRR scheme is implemented to prevent port starvation. Once the port is determined, the classification result of the first packet in the control FIFO will be processed to determine the DMA channel to queue the packet. Before a packet can be queued, it has to first check if the receive completion ring is congested. If it is congested, the packet will be discarded. If RCR is not congested, a packet buffer will be retrieved from RBR to store the packet, and when the packet is in system memory, the address will be queued to the RCR.

The granularity of the DMA timers is determined by the following counter. This is used to drive the DMA timeout counters. For a 300 MHz system clock, a value of 30000 (decimal) will yield a granularity of 100 µsec. And if the system clock is 375 MHz, a value of 37500 (decimal) will yield a granularity of 100 µsec.

TABLE 25-1 Receive DMA Clock Divider – RX_DMA_CK_DIV (FZC_DMC + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	cnt	0	RW	System clock divider. Determines the granularity of the DMA timeout countdown. The hardware countdown is cnt + 1.

25.1 Receive DMA Channel Selection

Packets with Layer 2 errors or with the promiscuous bit set are sent to the default DMA channel associated with the physical port.

There are eight receive DMA channel groups. For packets that pass through classification successfully, with no L2 CRC error or checksum error, the receive DMA group number and the offset from the hardware parser (FFLP) will be used to select the DMA channel. For packets with checksum errors, the offset will be changed to select the default within the group. The default DMA channel is hardwired to be the zero-th entry of the group.

Note that parser (FFLP) hardware errors are reported as part of the completion status.

The following registers are for the default per port DMA channel.

TABLE 25-2 Default Port 0 RDC – DEF_PT0_RDC (FZC_DMC + 00008_{16})

Bit	Field	Initial Value	R/W	Description
63:5	—	0	RO	*Reserved*
4:0	rdc	0	RW	Default per physical port RDC.

TABLE 25-3 Default Port 1 RDC – DEF_PT1_RDC (FZC_DMC + 00010_{16})

Bit	Field	Initial Value	R/W	Description
63:5	—	0	RO	*Reserved*
4:0	rdc	0	RW	Default per physical port RDC.

The RDC Table register set is used to program the RDC tables. Each table has 16 entries; the first 16 entries are for table 0, next 16 for table 1 and so on. The default value for each group is the zero-th entry of the table.

TABLE 25-4 RDC Table – RDC_TBL (FZC_ZCP 10000_{16}) (count 128 step 8)

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	rdc	0	RW	Receive DMA channel. Since there are only 16 receive DMA channels, only values 0 to 15 are valid. Any value larger than 16 may have unspecified behavior.

The mode32 bit should be set to zero. Also included in the register are control signals for debug and initialization. Setting the ram_acc bit enables software to read from or write to the prefetch or shadow memory.

TABLE 25-5 Receive Addressing Mode – RX_ADDR_MD (FZC_DMC + 00070_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:2	dbg_pt_mux_sel	0	RW	00_2, 01_2: debug_data 10_2: debug_training_vector 11_2: ~rdmc_debug_port
1	ram_acc	0	RW	Set to 1 to enable PIO access to shadow and prefetch memory, used during debug.
0	mode32	0	RW	set this bit to 1 to enable a special mode to spread receive DMA traffic to memory controller 2 and 3. This mode is primarily used in the NDPS framework. See config register RBR_CFIG_B for detail description.

25.2 Port Scheduler

Since a port may be supporting 1 Gbps or 10 Gbps, a rate-based scheduler is needed to guarantee the service rate among different ports on average. We implement a variation of the Deficit Round Robin algorithm. The scheduler only switches port at packet boundary, and only schedules a port when the port FIFO has at least one complete packet. Note that the goal of DRR is to allocate the bandwidth to each port according to the weight for a round. Thus, within a round, any packet from any queue may be serviced, as long as the queue still has a positive credit.

The pseudocode for the algorithm is described below.

```
I :    = {0, 1}
C_i :  = deficit counters of queue i
W_i :  = assigned weight for queue i
```

A queue is eligible if it has a completed packet. The next_queue_in_I operation will return the first queue in I if the last queue is reached.

```
i = last queue in I;
select:i = next_queue_in_I;
                C_i = min [C_i + W_i, W_i];
                if (Queue i is not eligible || C_i <=0) goto select;
loop:  process one packet from Queue i;
   decrement C_i accordingly; // See below on how C_i is decremented.
   if (Queue i not eligible || C_i <= 0) goto select;
   goto loop;
```

C_i is decremented by the number of 16-byte blocks the packet contains. A partial block is considered as one complete block. The following registers program the weight of each port.

TABLE 25-6 Port DRR Weight 0 – PT_DRR_WT0 (FZC_DMC + 00028_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	wt	0400_{16}	RW	Default value is roughly 10 standard maximum Ethernet frame size.

TABLE 25-7 Port DRR Weight 1 – PT_DRR_WT1 (FZC_DMC + 00030_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	wt	0400_{16}	RW	Default value is roughly 10 standard maximum Ethernet frame size.

To monitor the actual bandwidth usage, the transfer issued is kept on a per-queue basis. The granularity of the counters is 16 bytes.

TABLE 25-8 Port FIFO Usage 0 – PT_USE0 (FZC_DMC + 00048_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	cnt	0	RO	Count the number of 16-byte blocks transmitted; clear on read.

TABLE 25-9 Port FIFO Usage 1 – PT_USE1 (FZC_DMC + 00050_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	cnt	0	RO	Count the number of 16-byte blocks transmitted; clear on read.

25.3 Partitioning Support

Each DMA supports two naturally aligned, physically contiguous areas in system memory, each called a logical page (page 0 and 1). The idea is that one area may be used for configuration, the other for buffers. Hardware does not know what the usage is. As long as the address is in one of the defined areas, hardware will apply a

translation function to the address. The logical pages are intended to isolate one DMA's activities from another. Together with the virtualization address space, which isolates the PIOs from different software instances, virtualization may be supported.

An address is a 44-bit value[1] (we discuss how we extend it to 64 bits below.). A logical page is defined by the higher-order 32 bits, with a (value, mask) pair. A mask register define the bits in the value register that define the logical page. The higher order bits of the mask register have to be set with contiguous 1's, if the page is used. The higher-order bits of an address supplied by software will then be **and**ed with the mask to extract the page value. An address is said to be within the page if the bits in the value register selected by the mask equals the page value of the address. Associated with the (value, mask) pair is a relocation register. The translation process will be replacing the bits defined by the mask in the software-supplied address with the bits in the relocation register.

Hardware will check if the address is in page 0 first, then page 1. If it is in page 0, hardware will ignore the checking in page 1 and does not check if the two pages are defined consistently. Note that hardware does not check if the mask is defined with contiguous bits of 1's. So, software may define a page with holes.

To extend the addressing to 64 bits, an address handle register[2] defines the bits 63:44 of a 64-bit address. The higher-order bits 63:44 are concatenated with the 44-bit translated address obtained above to form the DMA address.

Note that software posts the least significant 44-bit addresses, and hardware returns only the least significant 44-bit addresses (the address before the translation) in the receive completion ring. The handle and the translated addresses are only used during the data transfer phase.

The following registers are defined.

TABLE 25-10 Receive Logical Page Valid – RX_LOG_PAGE_VLD (FZC_DMC + 20000_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:2	func	0	RW	Assigned device function number. Only 00_2 and 01_2 are valid.
1	page1	0	RW	Set to 1 to indicate page 1 is defined.
0	page0	0	RW	Set to 1 to indicate page 0 is defined.

1. In UltraSPARC T2, the 44-bit addressing is more than what we implemented, so we may cut down on that to optimize areas. As far as software is concerned, the register can support that width. The driver may need to be aware of the fact that when in UltraSPARC T2, posting more than 37 bits may not make sense.

2. For UltraSPARC T2, this register may have no effect, that is, the higher-order bits are always assumed to be zeros.

TABLE 25-11 Receive Logical Page Mask 1 – RX_LOG_MASK1 (FZC_DMC + 20008_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	mask	0	RW	Logical page mask for bits 43:12. Set the corresponding bit to 1 to select the bit. If the mask bits are all zeros, then the address definition is an automatic pass and there is no need to perform a translation.

TABLE 25-12 Receive Logical Page Value 1 – RX_LOG_VAL1 (FZC_DMC + 20010_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	value	0	RW	Value used to compare with the bits selected by the mask. This corresponds to bits 43:12 of the address. Note that bits not selected by the mask should always be set to 0.

TABLE 25-13 Receive Logical Page Mask 2 – RX_LOG_MASK2 (FZC_DMC + 20018_{16}) (count 16 step 40_{16})

Field	Bit	Initial Value	R/W	Description
—	63:32	0	RO	*Reserved*
mask	31:0	0	RW	Logical page mask for bits 43:12. Set the corresponding bit to 1 select the bit. If the mask bits are all zeros, then the address definition is an automatic pass and there is no need to perform a translation.

TABLE 25-14 Receive Logical Page Value 2 – RX_LOG_VAL2 (FZC_DMC + 20020_{16}) (count 16 step $40)_{16}$

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	value	0	RW	Value used to compare with the bits selected by the mask. This corresponds to bits 43:12 of the address. Note that bits not selected by the mask should always be set to 0.

TABLE 25-15 Receive Logical Page Relocation 1 – RX_LOG_PAGE_RELO1 (FZC_DMC + 20028_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	relo	0	RW	Value used to replace the value of the bits defined by the mask field. This corresponds to bits 43:12 of the address.

TABLE 25-16 Receive Logical Page Relocation 2 – RX_LOG_PAGE_RELO2 (FZC_DMC + 20030_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	relo	0	RW	Value used to replace the value of the bits defined by the mask field. This corresponds to bits 43:12 of the address.

TABLE 25-17 Receive Logical Page Handle – RX_LOG_PAGE_HDL (FZC_DMC + 20038_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	handle	0	RW	Logical page handle, bits 63:44 of a 64-bit address, used to concatenate to a 44-bit address to generate a 64-bit address.

Note that the logical page definition registers have to be configured first before the RBRs and RCRs are configured. An address is considered invalid if it is outside both pages. Thus, software needs to program in at least one valid page. When the mask is programmed with all zeros, then any address is considered valid and none of the bits will be replaced with the value in relo.

25.4 Weighted Random Early Discard (Weighted RED)

The goals of congestion management in our context are to prevent overloading of the CPU and to fence off potential SYN attacks when system resources are low. The control mechanism is to discard packets randomly. A probabilistic algorithm has the benefit of desynchronizing the TCP slow-start behavior and achieving an overall improvement in throughput.

Let Q be defined as follows:

Q = Receive Completion Ring Occupancy Length, measured in number of entries.

Note that Q may be larger than the number of packets received.

WRED is characterized by two parameters: Threshold and Window. If the Q is larger than Threshold, then the packet is subjected to WRED discard. Window determines the range of Q above Threshold where the probabilistic discard is applicable. If Q is

larger than (Threshold + Window), the packet is always discarded. Since we want to protect existing connections and fence off potential SYN attacks, TCP SYN packets are subject to a different set of (Threshold, Window) pair.

The following pseudocode summarizes the algorithm, assuming the appropriate (Threshold, Window) pair is chosen.

```
T = Threshold;
W = Window;
R = Random;
    x = Q-T;
    if (x< 0) exit;
    R= get_random();// get_random returns a value between 0 and 1.
// W is specified as power-of-2, hence can be implemented with shift.
int is the integer function.
    if ( int(R*W) ← x ) discard packet;
    exit;
```

The random number is implemented with a 16-bit LFSR with the following polynomial.

$X^{16} + X^5 + X^3 + X^2 + 1$

The initial value of the polynomial is programmed through the following register. The LFSR is shifted *once* every time a random number is required.

TABLE 25-18 RED Random Number INIT – RED_RAN_INIT (FZC_DMC + 00068_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	opmode	0	RW	Should be set to 1 during normal operation. Set to zero for debug only. When set to 1, WRED is always enabled.
15:0	init	0	RW	Initial value.

The RED parameters of each individual Receive DMA Channel are programmed through the following register. Note that RED is always enabled. The initial value is set to zero, and hardware will discard incoming packets on power-on. Software needs to program a value to be at least 4 x 64 bytes less than the maximum configured completion ring size, or the value (Threshold + window) × 8B < (max configured size) – 4 x 64B.

TABLE 25-19 RDC RED Parameter 2 – RDC_RED_PARA (FZC_DMC + 30000_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	Reserved
31:20	thre_syn	0	RW	Threshold to start WRED for TCP-SYN packets.
19:16	win_syn	0	RW	Window for TCP-SYN packets. 0 is the degenerate case with window = 0; for the other cases, window = $2^n - 1$. The maximum is $2^{15} - 1$.
15:4	thre	0	RW	Threshold to start WRED for non-TCP-SYN packets.
3:0	win	0	RW	Window for non-TCP-SYN packets. 0 is the degenerate case with window = 0; for the other cases, window = $2^n - 1$. The maximum is $2^{15} - 1$.

If the window is zero, the behavior reduces to the tail-drop behavior, that is, packets will be discarded if Q reached the threshold value.

25.5 Receive DMA Datapath Configuration

Before a packet buffer is consumed and reported to software through the completion ring, hardware will check if the DMA channel is enabled. If the DMA channel is not enabled, the packet will be discarded. When an interrupt occurs, hardware writes the status of the DMA to a mailbox to lower the software latency to access the state information. Software may use the rst bit to reset a specific DMA channel. Software should use the following sequence to reset a DMA channel. First, set en bit to zero, and then set rst bit to 1. After the rst bit is cleared by hardware and the qst bit is set to 1, software may then start configuring the DMA channel. After configuration, software may then set the en bit to 1 to enable the DMA. The format of the mailbox is defined in a later section. The en bit can be set or cleared while the DMA is in operation. The state machine of the DMA may not yet have returned to its initial states after the en bit is cleared. This condition is indicated by the value of the qst bit.

Per-Receive DMA channel, software may configure the Receive logic to send minimal header information to software to reduce transfer overhead. By setting the full_hdr bit to 0 in the following register, only the first two bytes will be included in the internal header, that is, incoming frame starts from byte 2.

To facilitate software buffer management, hardware can skip over some number of bytes when storing the internal header and the packet. The offset field specifies the start of the packet buffer. Note that this offset applies to every packet buffer. In particular, if a jumbo frame is received and requires multiple packet buffers (one or

more from the block pool), the offset applies to all the different buffers used; that is, subsequent buffers used also have the same offset area, which does not contain any packet data.

TABLE 25-20 RXDMA Configuration 1 - RXDMA_CFIG1 (DMC + 00000_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
61:32	—	0	RO	*Reserved*
31	en	0	RW	Set to 1 to enable the receive DMA. If set to 0, packets selecting this DMA will be discarded. On fatal errors, this bit will be cleared by hardware.
30	rst	0	RW	Set to 1 to reset the DMA. Hardware will clear this bit after reset is completed. A reset will bring the specific DMA back to the power-on state.
29	qst	1	RO	When set to 1, it indicates all state machines associated with the DMA are in initial state.
28:12	—	0	RO	*Reserved*
11:0	mbaddr_h	0	RW	Bits 43:32 of the Mailbox address.

TABLE 25-21 RXDMA Configuration 2 - RXDMA_CFIG2 (DMC + 00008_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
31:6	mbaddr_l	0	RW	Bits 31:6 of the Mailbox address. Bits 5:0 are assumed to be zero, or 64-byte aligned.
5:3	—	0	RO	*Reserved*
2:1	offset	0	RW	Multiple of 64 bytes. 0 means no offset; 01_2 means 64 bytes; 10_2 means 128 bytes; 11_2 is invalid, hardware behavior not specified.
0	full_hdr	0	RW	Set to 1 to select the entire header of 18 bytes.

The following sections describe how to configure the RBR and RCR. A discussion on the DMA interrupt behavior is presented next.

25.5.1 Receive Block Ring Configuration

Before a packet can be queued, a packet buffer is needed. Hardware maintains 16 pools of buffer blocks, one for each receive channel. The block size of each DMA channel is fixed, say, 4 Kbytes, 8 Kbytes, or 32 Kbytes. The addresses of the buffer blocks of each pool are queued in a receive block ring in system memory. Each ring buffer entry is 4 bytes and stores the higher-order bits of a 44-bit address, aligned to 4-Kbyte boundary. Hardware then partitions each of these blocks into one of three packet buffer sizes programmed by software. Typically, size 0 will be for small packets, size 1 is for larger packets, and size 2 is for jumbo frames. Each block contains packet buffers of the same size. A conceptual view of the hardware and the

format of the entry are shown in FIGURE 25-1. Note that the addresses posted in the RBR will be subjected to the logical page definition check and relocation described earlier.

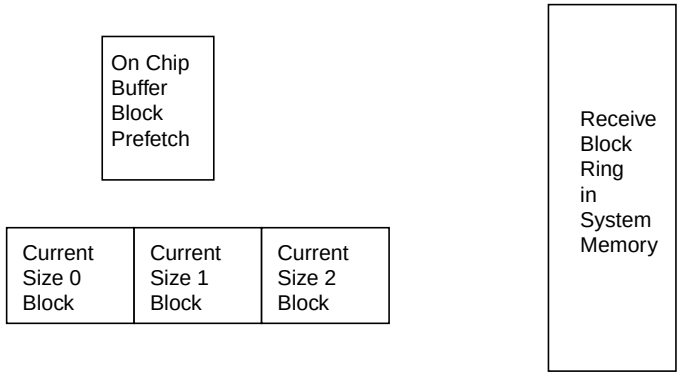


FIGURE 25-1 Buffer Block Ring

The following figure shows two examples of packet buffer layout. The first example shows a block size of 4 Kbytes, a buffer size of 1 Kbyte, an offset value of 128 bytes, and 18 bytes of header. Each packet buffer can only hold an 878-byte packet. FIGURE 25-2 shows the layout of a jumbo frame, using multiple blocks. Note that only the first block has the header. Every block will have the offset.

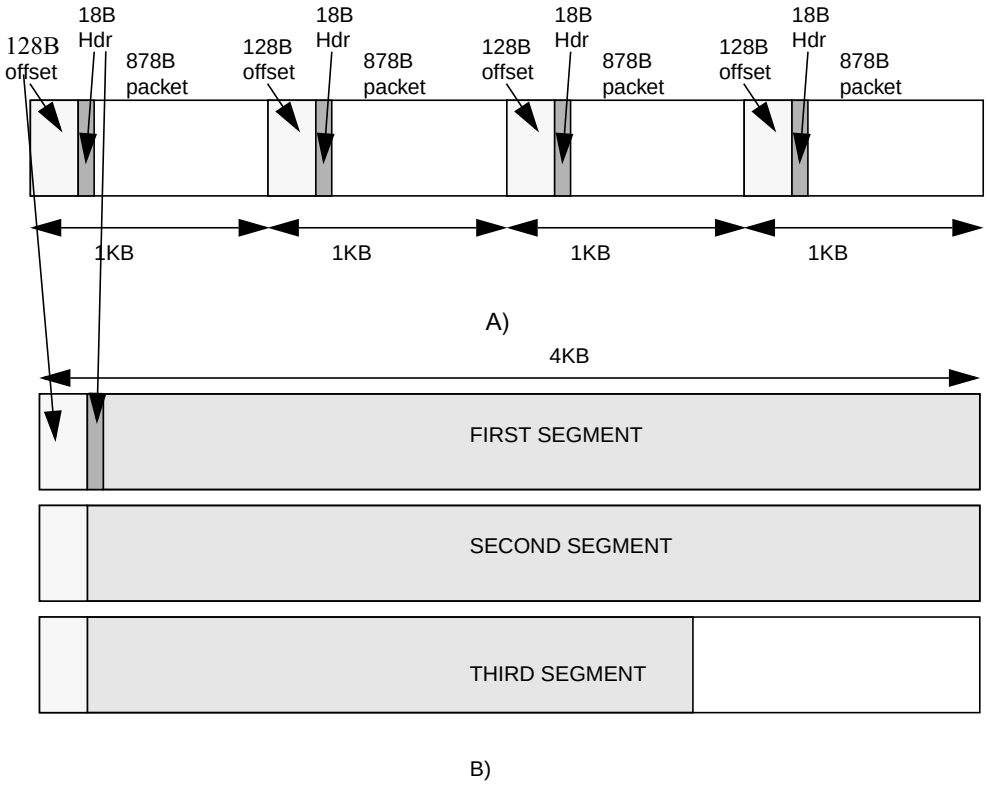


FIGURE 25-2 Packet Buffer Layout

TABLE 25-22 Format: Buffer Block Descriptor

Bit	Field	Description
31:0	blkaddr	Bits 43:12 of buffer block address. The address has to be naturally aligned at block size specified in the RBR_CFG_B register. Hardware will ignore the least significant bits accordingly. That is, for 32 Kbytes, bits 14:12 of the block address are assumed to be zero, and the value specified in the descriptor will be ignored by hardware; for 16 Kbytes, bits 13:12; for 8 Kbytes, bit 12.

There are three hardware registers, one for each size; each holds a block address. These three registers are filled from the prefetch area. The size of each packet received is compared with the programmed buffer sizes, in the order of 0 to 2. The first buffer size that is larger than the packet size will be used to store the packet. If all programmed buffer sizes are smaller than the packet size, the packet will be separated into at most three fragments, each made up of fresh blocks from the prefetch area. Note that the remaining area of the fresh block will not be used for another packet. If packet size is larger than the size of *three* blocks, the packet will be

discarded. A packet will be discarded if there is not enough buffer resource to store the entire packet. For example, the prefetch may be empty, or not enough free blocks for a jumbo packet.

When the current block is used up, another block address will be retrieved from the prefetch area. When the prefetch memory has space, more blocks from the ring buffer may be read into the hardware. Note that hardware might prefetch a memory region that crosses natural alignment. This is crossing 64 bytes. When this happens, two separate responses will be returned and may be out of order. Hardware does not reorder the responses. The consequence of this is that blocks may be consumed out of order. The prefetch buffer is organized as 4 wide by 8 deep. If software post blocks one at a time, it may happen that each 4-wide entry only contains one block. This may potentially degrade the system performance.

If hardware cannot find a buffer (or two buffers) to store the incoming packet, the packet will be discarded. The buffer address(es) will be queued to the RCR when the entire packet is stored in system memory. The packet buffer(s) cannot be logically consumed from the pool until it is clear that the packet will not be discarded by hardware. Note that if the buffer size is the same as the block size, then only one packet will be stored in a block.

The following registers configure and manage the RBRs.

TABLE 25-23 RBR Configuration A – RBR_CFIG_A (DMC + 00010_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:48	len	0	RW	Maximum number of RBBs in the buffer ring. The maximum is $2^{16} - 1$.
47:44	—	0	RO	*Reserved*
43:18	staddr_base	0	RW	Bits 43:18 of the address for the RBR. This value remains fixed and is used as the base address of the ring.
17:6	staddr	0	RW	Bits 17:6 of the address of the RBR. staddr_base concatenated with staddr is the starting address of the RBR.
5:0	—	0	RO	*Reserved*

Note that the entire RBR should stays within the "page" defined by staddr_base. A configuration error will result if the len field is defined such that the last entry is outside the page. In addition, hardware will support the wrap around at the end of the ring buffer defined by len. If the len field is defined in such a way that the last address is not naturally aligned at 64-byte boundary, hardware may read outside the area defined by len during the prefetch but will not use the data read from outside the BRB definition.

The following register configures the block size and the individual packet buffer sizes. Note that bksize has to be larger than all the packet buffer sizes. Hardware behavior is undefined if bksize is smaller than the packet buffer sizes. The vld bits of the three block sizes must be set to 1 in normal operations. These bits may be turned off for debug purpose only.

TABLE 25-24 RBR Configuration B – RBR_CFIG_B (DMC + 00018_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:26	—	0	RO	*Reserved*
25:24	bksize	0	RW	Buffer block size. 00_2 – 4K; 01_2 – 8K; 10_2 – 16K; 11_2 – 32K.
23	vld2	0	RW	Set to 1 to indicate size2 is valid, and enable hardware to allocate buffers of size 2. Always set to 1 in normal operation.
22:18	—	0	RO	*Reserved*
17:16	bufsz2	0	RW	Size 2 of packet buffer. 00_2 – 2K; 01_2 – 4K; 10_2 – 8K; 11_2 – 16K.
15	vld1	0	RW	Set to 1 to indicate size1 is valid; enables hardware to allocate buffers of size 1. Always set to 1 in normal operation.
14:10	—	0	RO	*Reserved*
9:8	bufsz1	0	RW	Size 1 of packet buffer. 00_2 – 1K; 01_2 – 2K; 10_2 – 4K; 11_2 – 8K.
7	vld0	0	RW	Set to 1 to indicate size0 is valid; enables hardware to allocate buffers of size 0. Always set to 1 in normal operation.
6:2	—	0	RO	*Reserved*
1:0	bufsz0	0	RW	Size 0 of packet buffer. 00_2 – 256; 0_2 – 512; 10_2 – 1K; 11_2 – 2K.

When RX_ADDR_MD bit 0 is set to 1 to select the mode for NDPS, one can program RBR_CFIG_B register for DMAs 4, 5, 6, 7, 12, 13, 14, 15 to support two special modes. In the first mode, set the value to 081_0000_{16}. Hardware will add 256B offset to the start of the packet buffer, before adding the selected packet offset. For example, if the DMA is programmed to start with an offset of 64B, in this special mode, the final offset is 256+64 or at 320B. Note that in this mode, only 4KB buffers are supported, and software has to limit packet size by setting MAX_PKT_SIZE in the MAC to guarantee the maximum packet size plus final offset does not exceed 4KB. In the second mode set the register to 283_0000_{16}. The offset calculation is the same as the first mode, except that the packet buffer is 16KB. This is large enough to support the enticipated maximum jumbo frame.

Block buffer addresses are added to the ring buffer by software. When software writes to the RBR_KICK register, indicating the number of descriptors added, hardware will update the internal state of the corresponding buffer pool.

TABLE 25-25 RBR Kick – RBR_KICK (DMC + 00020_{16} (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	bkadd	0	RW	Number of block buffers added by software. Hardware effect will be triggered when the register is written to.

The following registers define the state of the RBR. RBR_STAT will be cleared and RBR_HDH/L will be updated when RBR_CFIG_A is written.

TABLE 25-26 RBR Status – RBR_STAT (DMC + 00028_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	qlen	0	RO	Number of block addresses in the ring buffer. If maximum is reached, the value will be sticky and will be cleared on reset.

TABLE 25-27 RBR Head High – RBR_HDH (DMC + 00030_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	head_h	0	RO	Bits 43:32 of the software posted address, 4-byte aligned. This pointer is updated by hardware after each block buffer is consumed.

TABLE 25-28 RBR Head Low – RBR_HDL (DMC + 00038) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:2	head_l	0	RO	Bits 31:2 of the software posted address, 4-byte aligned. This pointer is updated by hardware after each block buffer is consumed.
1:0	—	0	RO	*Reserved*

25.5.2 Receive Completion Ring (RCR)

After the packet is stored in DRAM, the buffer address will be stored into the receive completion ring selected during the classification process. Each ring is built in system memory. Each entry in the queue is a 64-bit entity with the following format. In the case of jumbo frames, more than one buffer may be required to store the packet, and the multi bit is set to indicate that the next entry is the continuation of the current packet. (The specification allows more than three entries to represent a packet, but the current hardware does not support more than three.) In the current implementation, for packets with multiple entries, only bits multi, dcf_err, and pkt_buf_addr are different. Note that pkt_type is for clasifying non-fragmented packets only. All fragmented packets are marked as others.

TABLE 25-29 Format: Receive Completion Ring Entry

Bit	Field	Description
63	multi	Set to 1 to indicate the next entry is a continuation of the current packet. 0 indicates this is the last entry for a packet.
62:61	pkt_type	10_2 – non-fragmented UDP; 01_2 – non-fragmented TCP; 11_2 – non-fragmented SCTP; 00_2 – all fragmented packets and others.
60	zero_copy	*Reserved*
59	noport	Set to 1 to indicate the packet is an IP fragment; valid if the packet is an IP packet, as indicated by either the pkt_type bits or the class bits in the packet prepend header.
58	promis	In promiscuous mode, this bit will be set to 1, except when a packet is successfully classified by the packet parser, this bit will be cleared.
57:55	error	000_2 – no error; 001_2 – L2 error; 010_2 – reserved; 011_2 – L4 checksum error; 100_2 – FFLP soft error; 101_2 – ZCP soft error; 110_2–111_2 – *Reserved*.
54	dcf_err	Set to 1 to indicate data or control FIFO uncorrectable ECC error detected.
53:40	l2_len	This field returns the length of the incoming packet, as observed by the Ethernet MAC.
39:38	pktbufsz	Packet buffer size; 00_2 – buffer size 0; 01_2 – buffer size 1; 10_2 – buffer size 2; 11_2 – single block.
37:0	pkt_buf_addr	Bits 43:6 of packet buffer address. Bits 5:0 of address are zero, that is, 64-byte aligned. This points to the beginning of the packet buffer; the header will be located after the offset (if any), and the packet will start after the header.

For multiple-entry completion status, error, frag, pkt_type are only reported in the first entry. These fields are cleared to 0 in the second entry.

The following registers configure and manage the RCRs. The (count step) pair below specifies the RCR from 0 to 15. All the addresses specified in the registers are subjected to the translation operation when accesses to system memory are made. pthres and timeout set the condition when an interrupt *may* be issued by hardware. pthres defines the queue level where hardware issues an interrupt if enabled. As discussed in the life of a packet section, to reduce the per-packet overhead of updating the RCR, an on-chip RCR buffer is provided, and software-visible states will be updated after the RCR is updated to system memory. timeout, when enabled,

sets the time limit on how frequently the on-chip RCR buffer is updated and thus on how frequently RCR will be updated. See the next section for a more detailed description of pthres and timeout. Note that timeout needs to be enabled for normal operation.

TABLE 25-30 RCR Configuration A – RCRCFIG_A (DMC + 00040_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:48	len	0	RW	Maximum number of 8-byte entries in RCR.
47:44	—	0	RO	*Reserved*
43:19	staddr_base	0	RW	Bits 43:19 of the start address for the RCR.
18:6	staddr	0	RW	Bits 18:6 of start address for the RCR.
5:0	—	0	RO	*Reserved*

The RCR should be within the "page" defined by the staddr_base, that is, staddr_base concatenated with staddr plus 8 x len should be with the last address of the page defined by staddr_base.

TABLE 25-31 RCR Configuration B - RCRCFIG_B (DMC + 00048_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	pthres	0	RW	Packet threshold. When the number of packets queued in RCR is strictly larger than pthres, the DMA *may* issue an interrupt if enabled.
15	entout	0	RW	Enable timeout. If set to 1, enable the timeout. A timeout will initiate an update of the software-visible states. If interrupt is armed, in addition to the update, an interrupt to CPU will be generated and the interrupt disarmed.
14:6	—	0	RO	*Reserved*
5:0	timeout	0	RW	Timeout value. The system clock is divided down by the value programmed in RX_DMA_CK_DIV register.

The following three registers keep the state of the RCR.

TABLE 25-32 RCR Status A – RCRSTAT_A (DMC + 00050_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	qlen	0	RO	Number of packets queued. Initialize to 0 after RCRCFIG_A is written to.

TABLE 25-33 RCR Status B – RCRSTAT_B (DMC + 00058_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	tlptr_h	0	RO	Address of the RCR Tail Pointer {43:32} (points to the next available location.) Initialized after RCRCFIG_A is written to.

TABLE 25-34 RCR Status C – RCRSTAT_C (DMC + 00060_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:3	tlptr_l	0	RO	Address of the RCR Tail Pointer {31:3} (points to the next available location.) Initialized after RCRCFIG_A is written to.
2:0	—	0	RO	*Reserved*.

25.5.3 Receive DMA Interrupt Behavior

The Receive DMA Channel Event Mask register defines the events that, when they occur, will change the value of LDF, or flagging. There are two LDFs per DMA. The assignment of events to LDF is detailed in the description column. It is important to note that an interrupt is always associated with a mailbox update. The latter cannot be disabled.

TABLE 25-35 Receive DMA Channel Event Mask– RX_DMA_ENT_MSK (DMC + 0006816) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:22	—	0	RO	*Reserved*
21	rbr_tmout	1	RW	Set to 0 to enable flagging when RBR request timeout. Fatal error. Part of LDF 1.
20	rsp_cnt_err	1	RW	Set to 0 to enable flagging when response data count does not match length. Fatal error. Part of LDF 1.
19	byte_en_bus	1	RW	Set to 0 to enable flagging when byte enable bus error. Fatal error. Part of LDF 1.
18	rsp_dat_err	1	RW	Set to 0 to enable flagging when response data occurred on internal bus. Fatal error. Part of LDF 1.
17	rcr_ack_err	1	RW	Set to 0 to enable flagging when internal bus error on ACK or timeout. Fatal error. Part of LDF 1.
16	dc_fifo_err	1	RW	Set to 0 to enable flagging when Data/Control FIFO ECC error. Non-fatal error, status only. DMA will continue operation. Part of LDF 1.
15	—	0	RO	*Reserved*.

TABLE 25-35 Receive DMA Channel Event Mask– RX_DMA_ENT_MSK (DMC + 0006816) (count 16 step 200_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
14	rcrthres	1	RW	Set to 0 to enable flagging when RCR threshold crossed. Part of LDF 0.
13	rcrto	1	RW	Set to 0 to enable flagging when RCR timeout. Part of LDF 0.
12	rcr_sha_par	1	RW	Set to 0 to enable flagging when RCR shadow parity error. Fatal error. Part of LDF 1.
11	rbr_pre_par	1	RW	Set to 0 to enable flagging when RBR prefetch parity error. Fatal error. Part of LDF 1.
10	port_drop_pkt	1	RW	Set to 0 to enable flagging when packet drop indicated from FFLP or IPP. Used mainly for debug. Part of LDF 1.
9	wred_drop	1	RW	Set to 0 to enable flagging when packet drop because of WRED. Used mostly for debug. Part of LDF 1.
8	rbr_pre_emty	1	RW	Set to 0 to enable flagging when Receive Block Ring prefetch is empty when hardware tries to queue a packet. Incoming packets will be discarded. Nonfatal error. Part of LDF 1.
7	rcr_shadow_full	1	RW	Set to 0 to enable flagging when packet discard because of RCR shadow full. Part of LDF 1.
6	config_err	1	RW	Set to 0 to enable either RBR or RCR base page alignment error: last entry outside the base page. Part of LDF 1.
5	rcrincon	1	RW	Set to 0 to enable RCR inconsistent error to flag. When QLEN is decremented to zero, head pointer and tail pointer are not equal. Part of LDF 1.
4	rcrfull	1	RW	Set to 0 to enable flagging when Receive Completion Ring full when hardware tries to queue the completion status of a packet. Part of LDF 1.
3	rbr_empty	1	RW	Set to 0 to enable flagging when RBR empty when hardware attempts to prefetch. Part of LDF 1.
2	rbrfull	1	RW	Set to 0 to enable flagging when Receive Block Ring full when software tries to post more blocks. Part of LDF 1.
1	rbrlogpage	1	RW	Set to 0 to enable flagging when Buffer block logical page violation. Part of LDF 1. Fatal error.
0	cfiglogpage	1	RW	Set to 0 to enable flagging when DMA configuration logical page violation. Part of LDF 1. Fatal error.

The rcrthres and rcrto bits keep track of normal DMA operations, while the remaining bits are primarily used to detect error conditions.

The DMA channels are controlled with the following registers.

TABLE 25-36 Receive DMA Control and Status – RX_DMA_CTL_STAT (DMC + 00070_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:54	—	0	RO	*Reserved*
53	rbr_tmout	0	RO	RBR request timeout. Fatal error.
52	rsp_cnt_err	0	RO	Response data count not match length. Fatal error.
51	byte_en_bus	0	RO	Byte enable bus error. Fatal error.
50	rsp_dat_err	0	RO	Response data occurred on internal bus. Fatal error.
49	rcr_ack_err	0	RO	Internal bus error on ACK or timeout. Fatal error.
48	dc_fifo_err	0	RW1C	Data/Control FIFO ECC error.
47	mex	0	RW	Set to 1 to enable mailbox update and generating the rcrthres and rcrto flags. Hardware will reset to 0 after the mailbox has been updated and either flag has been set. Software needs to set to 1 for each mailbox update and event flag. This bit is also used to keep track of the exclusivity between threshold triggered or timeout triggered interrupt. If this bit is not set, there will be no timer-based interrupt, and threshold-based interrupt will not issue a mailbox update. Software may clear this bit to disable mailbox update and rcrthres and rcrto event flags.
46	rcrthres	0	RW1C	Set to 1 to indicate RCR threshold crossed. This is a level event.
45	rcrto	0	RW1C	Set to 1 to indicate RCR timed out if mex bit is set and the queue length is nonzero when timeout occurs. When software writes 1 to this bit, rcrto will be reset to 0. This is an edge-triggered event.
44	rcr_sha_par	0	RO	RCR shadow parity error. Fatal error
43	rbr_pre_par	0	RO	RBR prefetch parity error. Fatal error.
42	port_drop_pkt	0	RW1C	Set to 1 to indicate packet drop indicated from FFLP or IPP. Used mainly for debug.
41	wred_drop	0	RW1C	Set to 1 to indicate packet drop because of WRED. Used mostly for debug.
40	rbr_pre_emty	0	RW1C	Set to 1 to indicate receive block ring prefetch is empty when hardware tries to queue a packet. Incoming packets will be discarded. Nonfatal error. This is an edge-triggered event.
39	rcr_shadow_full	0	RW1C	Set to 1 to indicate packet discard because of RCR shadow full.
38	config_err	0	RO	Set to 1 to indicate either RBR or RCR base page alignment error: last entry outside the base page.
37	rcrincon	0	RO	Set to 1 to indicate RCR inconsistent; when qlen is decremented to zero, head pointer and tail pointer are not equal. Fatal error. Clear on reset. This is an edge-triggered event.
36	rcrfull	0	RO	Set to 1 to indicate receive completion ring full when hardware tries to queue the completion status of a packet. Incoming packets will be discarded. No buffer consumed. Fatal error. This is an edge-triggered event.

TABLE 25-36 Receive DMA Control and Status – RX_DMA_CTL_STAT (DMC + 00070_{16}) (count 16 step 200_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
35	rbr_empty	0	RW1C	Set to 1 to indicate RBR empty when hardware attempts to prefetch. On reset, since the RBR will be empty, this bit will be set. Software needs to clear this bit after posting the initial buffers.
34	rbrfull	0	RO	Set to 1 to indicate receive block ring full when software tries to post more blocks. Fatal error, clear on reset. This is an edge-triggered event.
33	rbrlogpage	0	RO	Set to 1 to indicate a buffer block has logical page violation. Fatal error, clear on reset. This is an edge-triggered event.
32	cfiglogpage	0	RO	Set to 1 to indicate either RBR or RCR configuration address has a logical page violation. Fatal error. Clear on reset. This is an edge-triggered event.
31:16	ptrread	0	RW	Number of buffer pointers read. Used to advance the RCR head pointer. Since a packet may be stored with multiple pointers, the value may be larger than pktread. *This value must be consistent with* pktread *for hardware to work correctly.*
15:0	pktread	0	RW	Number of packets read; when written to, decrement the qlen counter by pktread. qlen is lower bounded to zero.

Note that since the hardware detect the conditions on RBR_PRE_EMTY, RCR_SHADOW_FULL, PORT_DROP_PKT and WRED_DROP are independent, they may be set for a single packet. The DMA Control And Status Debug register is used primarily to test the interrupt datapath during design verification.

TABLE 25-37 Receive DMA Control and Status Debug – RX_DMA_CTL_STAT_DBG (DMC + 00098_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:54	—	0	RO	Reserved
53	rbr_tmout	0	RW	See description above.
52	rsp_cnt_err	0	RW	See description above.
51	byte_en_bus	0	RW	See description above.
50	rsp_dat_err	0	RW	See description above.
49	rcr_ack_err	0	RW	See description above.
48	dc_fifo_err	0	RW	See description above.
47	mex	0	RW	See description above.
46	rcrthres	0	RW	See description above.
45	rcrto	0	RW	See description above.
44	rcr_sha_par	0	RW	See description above.
43	rbr_pre_par	0	RW	See description above.

TABLE 25-37 Receive DMA Control and Status Debug – RX_DMA_CTL_STAT_DBG (DMC + 00098_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
42	port_drop_pkt	0	RW	See description above.
41	wred_drop	0	RW	See description above.
40	rbr_pre_emty	0	RW	See description above.
39	rcr_shadow_full	0	RW	See description above.
38	config_err	0	RW	See description above.
37	rcrincon	0	RW	See description above.
36	rcrfull	0	RW	See description above.
35	rbr_empty	0	RW	See description above.
34	rbrfull	0	RW	See description above.
33	rbrlogpage	0	RW	See description above.
32	cfiglogpage	0	RW	See description above.
31:16	ptrread	0	RW	See description above.
15:0	pktread	0	RW	See description above.

For each RCR, to reduce the overhead of updating the RCR to the DRAM, the hardware maintains a shadow tail buffer of RCR. After storing the packet to DRAM, the packet pointer(s) will be updated in the shadow tail buffer first. Thus, hardware needs to maintain extra hardware states, including the queue length and the tail pointer. Whenever the shadow tail buffer reaches a cache line, or 64 bytes, hardware will update the RCR in DRAM and update the register states (software visible) with the hardware states.

Either a timeout or threshold crossing may trigger the DMA to raise a flag, which may trigger a system interrupt. Hardware will first update the RCR with the tail buffer (if the CSR state is not up to date) and then may copy the CSRs to the mailbox and raise the LDF if selected.

Timeout (which is enabled by setting the entout bit) and Threshold crossing events are serialized in that if the two events occur simultaneously, hardware may pick one over the other, and if one event has started, the processing of the other will be stalled until the hardware completes the processing associated with the first event. This means that software-visible states are updated before hardware starts processing the second event. It is possible that the first event may disqualify the second event in that the mex bit has been reset, and the second event may not take any action.

Note that shadow and mailbox updates are ordered events, whereas packet payload updates are relax ordered and may bypass earlier requests.

There are two streams of events: packet arrival and timer count down.Two triggers control interrupt: threshold, and timeout. The values are set in the RCRCFG_B register. In addition, a mailbox is associated with the receive DMA, and needs to be

configured when using interrupt based on these two triggers; that is, an interrupt associated with either timeout or threshold is always associated with an mailbox update.

We first describe the timer. Once enabled, the timer starts counting down. When timeout is reached, hardware will check if the shadow state, that is the hardware version of the RCR queue, is different from the software version. For example, there are packets received and their completion statuses have been stored in the shadow buffer but have not been updated to system memory. If so, hardware will update the completion ring. After the update has been acknowledged, hardware will update the software visible completion state, this includes queue length, tail pointers, etc.

If the mex bit is set at timeout, hardware will initiate a mailbox update. After the ACK is received, the rcrto bit will be set and the mex bit will be cleared. If the mex bit is not set at timeout, hardware will not update mailbox and not set the rcrto bit.

When software writes to the RX_DMA_CTL_STAT register with the mex bit set, hardware will initialize the timer. If hardware is in the process of updating the shadow and is waiting for the acknowledgement, hardware will logically apply this action until the ACK is received. Specifically, there will be no mailbox update and no setting of the rcrto bit.

If a shadow update to system memory occurred because the shadow buffer reached a cache line or was triggered by threshold and during that time, a timeout occurred. there will be no additional RCR update.

Next we describe the threshold. There are two versions of queue length: software and hardware. Software version of the queue length reflects the completion ring in system memory. The hardware version reflects the completion ring in system memory and the shadow memory. Note that there may not be an actual hardware queue length computed in hardware. Threshold logic applies to the software version of the completion queue length. software RCR queue length will be updated whenever the shadow is updated to the system memory.

If the mex bit is set and the software version of the RCR queue length is greater than threshold, hardware will initiate the RCR update if shadow is not in sync with system memory, followed by a mailbox update. After the acknowledgement is received, the rcrthres bit will be set.

When a timeout occurs during an update of the shadow to the system memory and software issues a PIO with the mex bit set, then at this point, if the RCR queue length is greater than the threshold, the behavior will be the same as triggered by the threshold.

After reading the packets in memory, the software may write to the ptrread and the pktread fields to update the RCR. The qlen will be decremented by pktread, and the head pointer will be advanced by ptrread. Note that these are not addresses; they are the number of packets read and the number of pointers read, which are 8-byte aligned. The internal hardware state will be updated.

The RCR Flush register will force an update to the RCR in system memory. This function is only applicable when the DMA is enabled.

TABLE 25-38 RCR Flush – RCR_FLSH (DMC + 00078_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	flsh	0	RW	Set to 1 to force the hardware to store the shadow tail block to DRAM if the hardware state (queue length and pointers) is different from the software-visible state. Reset to 0 by hardware when done.

Note that *only* a threshold event or a timeout event may trigger a mailbox update. For fatal errors, the DMA will be in an error state and will require software reset. For nonfatal errors, hardware will continue the receive DMA operation.

No additional logging is supported for logical page violation. For configuration logical page violations, the address written by software will be stored in the respective configuration registers. There is also the head pointer for RBR, or the tail pointer for RCR to indicate the location in violation. For buffer address logical page violation, hardware points to the current prefetch location in system memory. Software needs to examine the values in system memory, at most 128 bytes prior to the current prefetch to determine which entry is the offending location. In addition, software may examine the content prefetched into memory by directly accessing the on-chip memory (see the ram_acc bit in the RX_ADDR_MD register).

The current state of the RDC will be updated to a predefined 64-byte mailbox, specified at mbaddr, when either a timeout or a queue length crosses the threshold. The format of the mailbox is defined below. Note that this is a little-endian data structure and is arranged in address order. The state of the RX_DMA_CTL_STAT register is captured before the software-visible state registers are updated.

TABLE 25-39 Format: Receive DMA Channel Mailbox

Register	Byte	Description
RX_DMA_CTL_STAT	0-7	This field may be written to 0 by software. When an interrupt is issued, at least one of the bits will be non-zero.
RBR_STAT	8-15	
RBR_HDL{31:0}	16-19	
RBR_HDH{31:0}	20-23	
RCRSTAT_C{31:0}	32-35	
RCRSTAT_B{31:0}	36-39	
RCRSTAT_A	40-47	
—	48-63	*Reserved*. Set to 0.

25.5.4 Recommendation for Threshold and Timeout Settings

For low latency operation, it is desirable to have the hardware update the RCR quickly. Thus, the timeout value should be set to a low value, say, of the order of 10 µs. In addition, the threshold value should be set to a low value, say, less than 8. This guarantees that any packet completion information will not stay in the shadow buffer for longer than the programmed value.

Note that whether an actual interrupt will be issued will depend on the arm bit and timer value for the Logical Device Group, which is used to set the interrupt delay for the Logical Device Group. In low latency applications, the delay value for the group needs to be set to a low value as well.

For applications in which latency is not as sensitive, the DMA timeout value may be set to be longer, and the threshold value to be higher. For example, the setting may be used to balance the overhead of context switching. The exact values are application dependent and should be tuned to the platform.

It is possible to have a dynamic scheme where initially the setting is tuned for low latency, and as system load changes, the setting may be adjusted to values closer to latency-insensitive applications.

In general, the setting of the values are application specific. We mention these approaches here for reference only.

25.6 Receive Performance Management and Discard Statistics

If the RBR prefetch buffer is empty, RCR shadow is full or IPP or FFLP indicates discard, the packet will be discarded. The Receive Miscellaneous Discard Count register keeps count of these events.

TABLE 25-40 Receive Miscellaneous Discard Count – RXMISC (DMC + 00090_{16}) (count 16 step 200_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	oflow	Preserved	RW	Sticky count overflow bit. Write 0 to clear.
15:0	count	Preserved	RW	Count of packets dropped because the available buffer pool is empty. Write 0 to clear.

WRED discard counts are kept in the following registers. Note that since hardware determine the discard condition independently, multiple discard criteria may be satisfied. As a result, both counters may be incremented for a single packet.

TABLE 25-41 red Discard Count – RED_DIS_CNT (FZC_DMC + 30008_{16}) (count 16 step 40_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16	oflow	Preserved	RW	Sticky count overflow bit. Write 0 to clear.
15:0	count	Preserved	RW	Count of packets dropped because of red. Write 0 to clear.

25.7 Receive DMA Hardware Registers

The following registers log the parity errors of the prefetch or the shadow memory.

TABLE 25-42 RDMC Prefetch Parity Error – RDMC_PRE_PAR_ERR (FZC_DMC + 00078_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	err	0	RW	Set to 1 to indicate error. Software write 0 to clear.
14	merr	0	RW	Set to 1 to indicate multiple error. Software write 0 to clear.
13:8	—	0	RO	*Reserved*
7:0	addr	0	RW	Error location.

TABLE 25-43 RDMC Shadow Parity Error – RDMC_SHA_PAR_ERR (FZC_DMC + 00080_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	err	0	RW	Set to 1 to indicate error. Software write 0 to clear.
14	merr	0	RW	Set to 1 to indicate multiple error. Software write 0 to clear.
13:8	—	0	RO	*Reserved*
7:0	addr	0	RW	Error location.

The following registers may be used to access the prefetch and shadow memory on chip. The width of the memory is 148 bits. Software first sets up the address register. Memory access is triggered by a read or write access to Data register 4.

TABLE 25-44 RDMC Memory Address – RDMC_MEM_ADDR (FZC_DMC + 00088_{16})

Bit	Field	Initial Value	R/W	Description
63:9	—	0	RO	*Reserved*
8	pre_shad	0	RW	Select prefetch (set to 0) or shadow (set to 1) memory.
7:0	addr	0	RW	Memory address. 0–7 for DM 0; 8–15 for DMA1; etc.

TABLE 25-45 RDMC Memory Data 0 – RDMC_MEM_DAT0 (FZC_DMC + 00090_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Bits 31:0 of memory.

TABLE 25-46 RDMC Memory Data 1 – RDMC_MEM_DAT1 (FZC_DMC + 00098_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Bits 63:32 of memory.

TABLE 25-47 RDMC Memory Data 2 – RDMC_MEM_DAT2 (FZC_DMC + $000A0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Bits 95:64 of memory.

TABLE 25-48 RDMC Memory Data 3 – RDMC_MEM_DAT3 (FZC_DMC + $000A8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Bits 127:96 of memory.

TABLE 25-49 RDMC Memory Data 4 – RDMC_MEM_DAT4 (FZC_DMC + $000B0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	data	0	RW	Bits 147:128 of memory. Bits131:128 are the valid bits for each of the 32-bit blocks, and bits 147:128 are the even parity bits for each 8-bit quantity.

The Receive Control Data FIFO Error Mask register logs the per-port data or control FIFO error status. The errors with the mask bit cleared will be **or**ed into one signal and sent to the PIO block when reporting system errors. These errors are mainly for hardware debug.

TABLE 25-50 Receive Control Data FIFO Error Mask – RX_CTL_DAT_FIFO_MASK (FZC_DMC + $000C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:9	—	0	RO	*Reserved*
8	id_mismatch	1	RW	Set to 0 to enable the interrupt.
7:4	zcp_eop_err	F_{16}	RW	Set to 0 to enable the interrupt.
3:0	ipp_eop_err	F_{16}	RW	Set to 0 to enable the interrupt.

TABLE 25-51 Receive Control Data FIFO Error Status – RX_CTL_DAT_FIFO_STAT (FZC_DMC + $000B8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:9	—	0	RO	*Reserved*
8	id_mismatch	0	RW	Set to 1 to indicate if the packet IDs between data FIFO and control FIFO are different. Chip fatal error.
7:4	zcp_eop_err	0	RW1C	Set to 1 to indicate ZCP data FIFO EOP error. Bits 7 and 6 are always zeros. Bit 5 for port 1 and bit 4 for 0.
3:0	ipp_eop_err	0	RW1C	Set to 1 to indicate IPP data FIFO EOP error. Bits 3 and 2 are always zeros. Bit1 for port 1 and bit 0 for 0.

The Receive Control Data FIFO Error Status Debug register sets the value of the RX_CTL_DAT_FIFO_STAT register. It is used only for hardware debug.

TABLE 25-52 Receive Control Data FIFO Error Status Debug – RX_CTL_DAT_FIFO_STAT_DBG (FZC_DMC + $000D0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:9	—	0	RO	*Reserved*
8	id_mismatch	0	RW	See description above.
7:4	zcp_eop_err	0	RW1C	See description above
3:0	ipp_eop_err	0	RW1C	See description above.

The following is the RDMC training vector.

TABLE 25-53 RDMC Training Vector – RDMC_TRAINING_VECTOR (FZC_DMC + $000C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	training_vector	0	RW	Training vector.

25.8 IPP Data FIFO Hardware Registers

The following are hardware registers related to IPP.

TABLE 25-54 IPP Configuration – IPP_CFIG (FZC_IPP)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	soft_rst	0	WO	IPP soft reset. Writing a zero to this bit has no effect on IPP. Writing a 1 resets all the IPP Finite State Machines and brings IPP back to POR state. It does not reset IPP hardware registers. One application of this bit is when software identified that IPP is in an error state and decided to bring this IPP port back to life. Thus, software disables this IPP first, then issues IPP soft reset.
30:25	—	0	RO	*Reserved*
24:8	ip_max_pkt	$1FFFF_{16}$	RW	If ip_max_pkt_bytes is exceeded, the checksum reports the IP packet is a bad one. Default: 1 $FFFF_{16}$.
7	fflp_cs_pio_w	0	RW	FFLP checksum info PIO write access enable. 0 = Normal operation; the checksum decision comes from FFLP. 1 = Enable checksum decision to be decided by software. It bypasses checksum decision coming from FFLP. Do not set this bit to 1 if MAC is sending packet to IPP, or RDMC is reading out packet from IPP.

TABLE 25-54 IPP Configuration – IPP_CFIG (FZC_IPP) *(Continued)*

Bit	Field	Initial Value	R/W	Description
6	pfifo_pio_w	0	RW	Pre-FIFO PIO write access enable. 0 = Normal operation; software won't be able to write to pre-FIFO and its Rd/Wr pointers. 1 = Enable software to override Pre-FIFO or its Rd/Wr pointer value. Do not set this bit to 1 if MAC is sending packet to IPP, or RDMC is reading out packet from IPP.
5	dfifo_pio_w	0	RW	Data-FIFO PIO write access enable. 0 = Normal operation.;software won't be able to write to Data-FIFO and its Rd/Wr pointers. 1 = Enable software to override Data-FIFO or its Rd/Wr pointer value. Do not set this bit to 1 if MAC is sending packet to IPP, or RDMC is reading out packet from IPP.
4	cksum_en	0	RW	TCP/IP and UDP checksum enable bit. 0 = No TCP/IP & UDP checksum, which is default; IPP operates without checksumming. 1 = IPP does the TCP/IP & UDP checksum.
3	drop_bad_crc	0	RW	Pre-FIFO parity error counted enable bit. 1 = A pre-FIFO's parity error will cause the IPP packet discard counter to be increased.
2	dfifo_ecc_en	0	RW	Data FIFO ECC correction enable bit. 0 = ECC detection is executed, but no correction; 1 = ECC detection is executed, so is the correction
1	debug_bus_ out_en	0	RW	Debug bus out enable bit. Set to 1, this IPP's debug bus goes to the debug pins. No more than one port's debug bus out should be enabled. Multiple IPPs' debug bus enabling will cause IPP debug bus collision.
0	ipp_enable	0	RW	0 = IPP is in the initialization state. IPP does not accept any packets from MAC/XMAC. If PIO changes this bit from 1 to 0, IPP will gracefully transit from its enabled state to `ipp_disabled` state. Once PIO writes a 0, IPP stops accepting any new packets from MAX/XMAC, but finishes processing the current packet. This bit remains a 1 until IPP is ready to enter its "disabled" state. 1 = IPP is in the normal packet operation state. Putting it into enable state upon initialization, IPP is ready to receive packets from MAC/XMAC. This bit remains 0 until IPP is ready to enter its enabled state.

TABLE 25-55 IPP Packet Discard – IPP_PKT_DIS (FZC_IPP + 0020_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:0	count	0	RCW	For statistical purposes, a 14-bit sticky counter in each IPP counts the number of times the IPP asserts the discard packet bit due to any of detected discard conditions. This register will stay at the last counted number until it is read by software. (Reading this register clears the content of this register but will not remove the condition that causes the event to be registered.) A direct write capability provides a way to quickly test this register/counter. Two possibilities of increasing this counter: (1) runt packet from MAC, (2) pre-FIFO parity error if IPP_CFIG{3} is set to 1.

TABLE 25-56 IPP Bad Checksum Count – IPP_BAD_CS_CNT (FZC_IPP + 0028_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:0	count	0	RCW	For statistical purposes, a 14-bit sticky counter in each IPP counts the number of packets dropped due to the bad TCP/IP & UDP checksum. This register will stay at the last counted number until it is read by software. (Reading this register clears the content of this register but will not remove the condition that causes the event to be registered.) A direct write capability provides a way to quickly test this register/counter.

TABLE 25-57 IPP ECC Error Count – IPP_ECC (FZC_IPP + 0030_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	count	0	RCW	For statistical purposes, an 8-bit sticky counter in each IPP counts the number of times an ECC error is detected. This register will stay at the last counted number until it is read by software. (Reading this register clears the content of this register but will not remove the condition which causes the event to be registered.) A direct write capability provides a way to quickly test this register/counter.

TABLE 25-58 IPP Interrupt Status – IPP_INT_STAT (FZC_IPP 0040_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	sop_miss	0	RCW	Data FIFO missed an SOP between two EOPs read by RDMC. (SOP: Start Of a Packet). Fatal error.

TABLE 25-58 IPP Interrupt Status – IPP_INT_STAT (FZC_IPP 0040_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
30	eop_miss	0	RCW	Data FIFO missed an EOP between two SOPs read by RDMC. (EOP: End Of a Packet). Fatal error.
29:28	dfifo_ue	0	RCW	Data FIFO ECC uncorrectable error (msb ← 129:65, lsb ← 64:0). Any ECC error on the SOP or EOP bit is a uncorrectable error. Fatal error.
27:26	dfifo_ce	0	RCW	Data FIFO ECC correctable error (msb ← 129:65, lsb ← 64:0).
25:24	dfifo_ecc	0	RCW	Data FIFO ECC error (msb ← 129:65, lsb ← 64:0).
23	—	0	RO	*Reserved*, 0_{16}.
22:12	dfifo_ecc_idx	0	RCW	Data FIFO ECC error entry index.
11	pfifo_perr	0	RCW	Pre-FIFO parity error occurred. Soft error can happen at the SOP or EOP which dictates a packet's structure/boundary. Fatal error.
10	ecc_err_max	0	RCW	ECC error counter reached maximum value. To set this bit to 1 by PIO, write the maximum value, FF_{16}, to the ECC error counter. To clear this bit, read the ECC error counter.
9:4	pfifo_err_idx	0	RCW	Pre-FIFO parity error entry index.
3	pfifo_over	0	RCW	Pre-FIFO overrun occurred. If 1 and (pfifo_und = 0), the Pre-FIFO is full; IPP will not take MAC's request until Pre-FIFO has free space. Fatal error.
2	pfifo_und	0	RCW	Pre-FIFO underrun occurred. If 1 and (pfifo_over = 0), the Pre-FIFO is empty, but the Data-FIFO is still trying to get data from Pre-FIFO. Fatal error.
1	bad_cs_mx	0	RCW	Bad TCP/IP and UDP checksum Counter reached maximum value. To set this bit to 1 by PIO, write the maximum value, $3FFF_{16}$, to the bad checksum counter. To clear this bit, read the bad checksum counter.
0	pkt_dis_mx	0	RCW	Packet discard counter reached maximum value. To set this bit to 1 by PIO, write the maximum value, $3FFF_{16}$, to the packet discard counter. To clear this bit, read the packet discard counter.

The above interrupt status register is read to clear. The mask register defines which errors will not be masked. All unmasked errors will be **or**ed into one signal. The result will be sent to the PIO block for reporting.

Reading this register clears the content of this register but will not remove the condition that causes the event to be registered.

When a fatal error happens, reboot this port by (1) soft-resetting this individual port by writing ipp_cfig{31} = 1; or (2) niu_core_reset.

If both sop_miss and eop_miss are 1's, IPP is stalled in a fatal error state. Check IPP_SM/c_unload_st{4:0}.

If both pfifo_over and pfifo_und are 1's, IPP is stalled in a fatal error state. Check IPP_SM/{phase_state_xmac{2:0},phase_state{2:0},mac_ack_fsm_curstate{3:0}}.

TABLE 25-59 IPP Interrupt Mask – IPP_MSK (FZC_IPP + 0048_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7	ecc_err_mx	1	RW	ECC error counter reached maximum value. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
6	dfifo_eop_sop	1	RW	Data-FIFO missing an EOP or SOP error occurred. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
5	dfifo_uc	1	RW	Data-FIFO ECC uncorrectable error occurred. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
4	pfifo_par	1	RW	Pre-FIFO parity error occurred. 0 = Enable interrupt generation; l = Disable interrupt generation.
3	pfifo_over	1	RW	Pre-FIFO overrun occurred. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
2	pfifo_und	1	RW	Pre-FIFO under-un occurred. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
1	bad_cs	1	RW	Bad TCP/IP checksum counter reached maximum value. 0 = Enable interrupt generation; 1 = Disable interrupt generation.
0	pkt_dis_cnt	1	RW	Packet discard counter reached maximum value. 0 = Enable interrupt generation.; 1 = Disable interrupt generation.

TABLE 25-60 IPP PreFIFO Read Data 1 – IPP_PFIFO_RD1 (FZC_IPP + 0060_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Pre-FIFO{31:0}. Use Pre-FIFO read pointer to access the desired location.

TABLE 25-61 IPP PreFIFO Read Data 2 – IPP_PFIFO_RD2 (FZC_IPP + 0068_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Pre-FIFO{63:32}. Use Pre-FIFO read pointer to access the desired location.

TABLE 25-62 IPP PreFIFO Read Data 3 – IPP_PFIFO_RD3 (FZC_IPP + 0070_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Pre-FIFO{95:64}. Use Pre-FIFO read pointer to access the desired location.

TABLE 25-63 IPP PreFIFO Read Data 4 – IPP_PFIFO_RD4 (FZC_IPP + 0078_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Pre-FIFO{127:96}. Use Pre-FIFO read pointer to access the desired location.

TABLE 25-64 IPP PreFIFO Read Data 5 – IPP_PFIFO_RD5 (FZC_IPP + 0080_{16})

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:2	data	X	RO	Read this register to get the contents of the last read Pre-FIFO{145:130}. Use Pre-FIFO read pointer to access the desired location.
1:0	data1	X	RO	Read this register to get the contents of the last read Pre-FIFO{129:128}. Use Pre-FIFO read pointer to access the desired location.

TABLE 25-65 IPP PreFIFO Write Data 1 – IPP_PFIFO_WR1 (FZC_IPP + 0088_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you wish to be written to Pre-FIFO{31:0}. Physically, this register is shared with the IPP Data-FIFO Write Data 1.

TABLE 25-66 IPP PreFIFO Write Data 2 – IPP_PFIFO_WR2 (FZC_IPP + 0090_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Pre-FIFO{63:32}. Physically, this register is shared with the IPP Data-FIFO Write Data 2.

TABLE 25-67 IPP PreFIFO Write Data 3 – IPP_PFIFO_WR3 (FZC_IPP + 0098_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Pre-FIFO{95:64}. Physically, this register is shared with the IPP Data-FIFO Write Data 3.

TABLE 25-68 IPP PreFIFO Write Data 4 – IPP_PFIFO_WR4 (FZC_IP + $00A0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Pre-FIFO{127:96}. Physically, this register is shared with the IPP Data-FIFO Write Data 4.

TABLE 25-69 IPP PreFIFO Write Data 5 – IPP_PFIFO_WR5 (FZC_IPP + $00A8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	data	0	RW	Load this register with the value you want to be written to Pre-FIFO{129:128}. If the IPP_CONFIGURATION register's Pre-FIFO PIO write access enable = 1, at the end of writing this register, the full PreFIFO Write Data register{129:0} is written into the Pre-FIFO. Physically, this register is shared with the IPP Data-FIFO Write Data 5.

TABLE 25-70 IPP Pre–FIFO Read Pointer – IPP_PFIFO_RD_PTR (FZC_IPP + $00B0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	ptr	0	RW	Load this 6-bit pointer value to directly control the Pre-FIFO read pointer. This register is write privilege protected. If PIO is writing to this read pointer, preset Pre-FIFO PIO write access enable = 1 in IPP Configuration register. Failure to do so will result in an unpredictable value being written.

TABLE 25-71 IPP PreFIFO Write Pointer – IPP_PFIFO_WR_PTR (FZC_IPP + $00B8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:7	—	0	RO	*Reserved*
6:0	ptr	0	RW	Load this 7-bit pointer value to directly control the pre-FIFO write pointer. This register is write privilege protected. PIO writing to this write pointer presets Pre-FIFO PIO write access enable = 1 in IPP Configuration register. Failure to do so will result in an unpredictable value being written.

TABLE 25-72 IPP Data FIFO Read Data 1 – IPP_DFIFO_RD1 (FZC_IPP + $00C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Data-FIFO{31:0}. Use Data-FIFO read pointer to access the desired location.

TABLE 25-73 IPP Data FIFO Read Data 2 – IPP_DFIFO_RD2 (FZC_IPP + $00C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Data-FIFO{63:32}. Use Data-FIFO read pointer to access the desired location.

TABLE 25-74 IPP Data FIFO Read Data 3 – IPP_DFIFO_RD3 (FZC_IPP + $00D0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Data-FIFO{95:64}. Use Data-FIFO read pointer to access the desired location.

TABLE 25-75 IPP Data FIFO Read Data 4 – IPP_DFIFO_RD4 (FZC_IPP + $00D8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	X	RO	Read this register to get the contents of the last read Data-FIFO{127:96}. Use Data-FIFO read pointer to access the desired location.

TABLE 25-76 IPP Data FIFO Read Data 5 – IPP_DFIFO_RD5 (FZC_IPP + $00E0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	data	X	RO	Read this register to get the contents of the last read Data-FIFO{129:128}. Use Data-FIFO read pointer to access the desired location.

TABLE 25-77 IPP Data FIFO Write Data 1 – IPP_DFIFO_WR1 (FZC_IPP + $00E8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Data-FIFO{31:0}. Physically, this register is shared with the IPP Pre-FIFO Write Data 1.

TABLE 25-78 IPP Data FIFO Write Data 2 – IPP_DFIFO_WR2 (FZC_IPP + $00F0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Data-FIFO{63:32}. Physically, this register is shared with the IPP Pre-FIFO Write Data 2.

TABLE 25-79 IPP Data FIFO Write Data 3 – IPP_DFIFO_WR3 (FZC_IPP + $00F8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Data-FIFO{95:64}. Physically, this register is shared with the IPP Pre-FIFO Write Data 3.

TABLE 25-80 IPP Data FIFO Write Data 4 – IPP_DFIFO_WR4 (FZC_IPP + 0100_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	data	0	RW	Load this register with the value you want to be written to Data-FIFO{127:96}. Physically, this register is shared with the IPP Pre-FIFO Write Data 4.

TABLE 25-81 IPP Data FIFO Write Data 5 – IPP_DFIFO_WR5 (FZC_IPP + 0108_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1:0	data	0	RW	Load this register with the value you want to be written to Data-FIFO{129:128}. If IPP Configuration register's Data-FIFO PIO write access enable = 1, at the end of writing this register, the full Data-FIFO Write Data register{129:0} is written into the Data-FIFO. Physically, this register is shared with the IPP Pre-FIFO Write Data 5.

TABLE 25-82 IPP Data FIFO Read Pointer – IPP_DFIFO_RD_PTR (FZC_IPP + 0110_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	ptr	0	RW	Load this 12-bit pointer value to directly control the Data-FIFO read pointer. This register is write privilege protected. PIO writing to this read pointer presets Data-FIFO PIO write access enable = 1 in the IPP Configuration register. For 2K entry Data-FIFO, load bit 11:0; for 1K entry Data-FIFO, load {0_2, bits 10:0}

TABLE 25-83 IPP Data FIFO Write Pointer – IPP_DFIFO_WR_PTR (FZC_IPP + 0118_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	ptr	001_{16}	RW	Load this 12 bit pointer value to directly control the Data-FIFO write pointer. This register is write privilege protected. PIO writing to this write pointer presets Data-FIFO PIO write access enable = 1 in the IPP Configuration register. For 2K entry Data-FIFO, load bit 11:0; for 1k_entry Data-FIFO, load {0_2, bits 10:0}

TABLE 25-84 IPP State Machine – IPP_SM (FZC_IPP + 0120_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	sm	0	RO	Use this register to observe the state of the IPP internal state machine. Bits 31 through 0 are defined as follows: {0_2, ipp_en_rst_fsm_curstate{1:0}}, {0_2, tag_fsm_curstate{1:0}}, {phase_state_1st_data, phase_state_xmac{2:0}}, {0_2, c_unload_st{4}}, {c_unload_st{3:0}}, {0_2, ipp_ffl_curstate{1:0}}, {mac_ack_fsm_curstate{3:0}}, {0_2, phase_state{2:0}}

TABLE 25-85 IPP Checksum Status – IPP_CS_STAT (FZC_IPP + 0128_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:24	bcyc_cnt	0	RO	Received packet's data bus cycle count (1 = 16 bytes).
23:12	ip_len	0	RO	Extracted IP packet's length field{11:0}.
11	cs_fail	0	RO	If 1, calculated checksum fails the expected value.
10	term	0	RO	If 1, packet is terminated in initial states.
9	bad_num	0	RO	If 1, packet's number of bytes is bad (err_dat_byt).
8:0	cs_state	01_{16}	RO	IPP checksum control state-machine states.

TABLE 25-86 RIPP FFLP Checksum Information – IPP_FFLP_CS_INFO (FZC_IPP + 0130_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:10	pkt_id	0	RW	fflp_pkt_id{3:0}. FFLP issued packet ID.
9:8	l4_proto	0	RW	L4_protocol{1:0}. 00 – unknown; 01 – TCP; 10 – UDP; 11 *Reserved*.
7:4	v4_hd_len	0	RW	IPv4_hd_length{3:0}. IPv4's header length field.
3:2	l3_ver	0	RW	L3_version{1:0}. 00 – unknown; 01 – IPv4; 10 – IPv6; 11 – *Reserved*.
1:0	l2_op	0	RW	L2_option{1:0}. {LLC,VLAN}.

TABLE 25-87 IPP Debug Select – IPP_DBG_SEL (FZC_IPP + 0138_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:0	sel	0	RW	Select debug bus 0 to 8; 9 to 15 default to 0.

TABLE 25-88 IPP Data FIFO ECC Syndrome – IPP_DFIFO_ECC_SYND (FZC_IPP + 0140_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	—	0	RO	*Reserved*, 0_{16}.
15:0	synd	0	RC	Data-FIFO's ECC syndrome bits. (Reading this register clears the content of this register but will not remove the condition that causes the event to be registered.)

TABLE 25-89 IPP Data FIFO EOP Missed Read Pointer – IPP_DFIFO_EOP_RD_PTR (FZC_IPP + 0148_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	ptr	0	RCW	Data-FIFO read pointer when a missing EOP is detected between two SOPs read by RDMC, or a missing SOP is detected between two EOPs read by RDMC. (Reading this register clears the content of this register but will not remove the condition that causes the event to be registered.)

TABLE 25-90 IPP ECC Control – IPP_ECC_CTL (FZC_IPP + 0150_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	dis_dbl	0	RW	If 1, disable double-bit error, that is, suppress the uncorrectable ECC error if it occurred.
30:18	—	0	RO	*Reserved*
17	cor_dbl	0	RW	If 1, corrupt a double-bit error.
16	cor_sng	0	RW	If 1, corrupt a single-bit error.
15:11	—	0	RO	*Reserved*
10	cor_all	0	RW	If 1, corrupt all packets.
9	—	0	RO	*Reserved*
8	cor_1	0	RW	Feature NOT Supported. (If 1, corrupt 1 packet only.)
7:3	—	0	RO	*Reserved*
2	cor_lst	0	RW	If 1, corrupt the last line of a packet.
1	cor_snd	0	RW	If 1, corrupt the second line of a packet.
0	cor_fst	0	RW	If 1, corrupt the first line of a packet.

25.9 ZCP Control FIFO Hardware Registers

The following registers configure the ZCP control FIFO hardware and the interrupts. Note that not all bits are supported in UltraSPARC T2.

TABLE 25-91 ZCP Configuration – ZCP_CFIG (FZC_ZCP + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	Reserved.
24	zcp_32bit_mode	0	RW	program this bit to zero.When reset to 0 = 64-bit address is used.
23:16	zcp_debug_sel	0	RW	See the zcp_debug_sel equation following this table.
15:5		000_{16}	RW	(In UltraSPARC T2, program this field to zero.)
4	ecc_chk_dis	0	RW	ECC check disable. This bit applies to all control FIFOs that use an ECC check scheme. 0 = Enable ECC check (default); 1 = Disable ECC check. The purpose of this bit is to guard against a malfunctioning ECC logic. If a malfunctioning ECC occurred, this bit should be set to 1.
3	par_chk_dis	0	RW	Parity check disable. This bit applies to all translation-table-related RAM that uses a parity check scheme. 0 = Enable parity check (default); 1 = disable parity check.
2	dis_buff_rsp_if	0	RW	program this field to 0.
1	dis_buff_req_if	0	RW	program this field to 0.
0	zc_enable	0	RW	program this field to zero. 0 = The ZCP is in the initialization state. The transition to and out of this state is graceful. All packets are send to regular receive DMA channels.

TABLE 25-92 ZCP Interrupt Status – ZCP_INT_STAT (FZC_ZCP + 00008_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	rrfifo_underrun	0	RW1C	Request response FIFO underrun.
14	rrfifo_overrun	0	RW1C	Request response FIFO overrun.
13	—	0	RO	*Reserved*
12	rspfifo_uncorr_err	0	RW1C	Response interface FIFO uncorrectable ECC errors.
11	buffer_overflow	0	RW1C	FFLP-zcp interface buffer overflows. This happens when FFLP is not following the 7/9 clock rules. This is a catastrophic error. Full chip reset is necessary.
10	stat_tbl_perr	0	RW1C	Static table RAM parity errors.
9	dyn_tbl_perr	0	RW1C	Dynamic table RAM parity errors.
8	buf_tbl_perr	0	RW1C	Buffer pointer RAM (VA) RAM parity errors.
7	tt_program_err	0	RW1C	Software programmed the mode bits with an unsupported value.
6	rsp_tt_index_err	0	RW1C	Zero copy response interface logic detected that tt_index is outside the range that is programmed by software.
5	slv_tt_index_err	0	RW1C	Zero copy PIO logic detected that tt_index is outside the range that is programmed by software.

TABLE 25-92 ZCP Interrupt Status – ZCP_INT_STAT (FZC_ZCP + 00008_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
4	zcp_tt_index_err	0	RW1C	Zero copy engine detected that tt_index is outside the range that is programmed by software.
3	cfifo_ecc3	0	RW1C	Port 3 control FIFO ECC error.
2	cfifo_ecc2	0	RW1C	Port 2 control FIFO ECC error.
1	cfifo_ecc1	0	RW1C	Port 1 control FIFO ECC error.
0	cfifo_ecc0	0	RW1C	Port 0 control FIFO ECC error.

The ZCP Interrupt Status Test register is readable and writeable. The content is the same as ZCP Interrupt Status register.

TABLE 25-93 ZCP Interrupt Status Test – ZCP_INT_STAT_TEST (FZC_ZCP + 00108_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	rrfifo_underrun	0	RW	This field is not meaningful.
14	rrfifo_overrun	0	RW	Request response FIFO overrun.
13	—	0	RO	*Reserved*
12	rspfifo_uncorr_err	0	RW	This field is not meaningful.
11	buffer_overflow	0	RW	This field is not meaningful.
10	stat_tbl_perr	0	RW	This field is not meaningful.
9	dyn_tbl_perr	0	RW	This field is not meaningful.
8	buf_tbl_perr	0	RW	This field is not meaningful.
7	tt_program_err	0	RW	This field is not meaningful.
6	rsp_tt_index_err	0	RW	This field is not meaningful.
5	slv_tt_index_err	0	RW	This field is not meaningful.
4	zcp_tt_index_err	0	RW	This field is not meaningful.
3	cfifo_ecc3	0	RW	This field is not meaningful.
2	cfifo_ecc2	0	RW	This field is not meaningful.
1	cfifo_ecc1	0	RW	Port 1 control FIFO ECC error.
0	cfifo_ecc0	0	RW	Port 0 control FIFO ECC error.

TABLE 25-94 ZCP Interrupt Mask – ZCP_INT_MASK (FZC_ZCP + 00010_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	rrfifo_underrun	1	RW	This field is not meaningful.
14	rrfifo_overrun	1	RW	This field is not meaningful.

TABLE 25-94 ZCP Interrupt Mask – ZCP_INT_MASK (FZC_ZCP + 00010_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
13	loj	1	RW	LOJ
12	rspfifo_uncorr_err	1	RW	This field is not meaningful.
11	buffer_overflow	1	RW	This field is not meaningful.
10	stat_tbl_perr	1	RW	This field is not meaningful.
9	dyn_tbl_perr	1	RW	This field is not meaningful.
8	buf_tbl_perr	1	RW	This field is not meaningful.
7	tt_program_err	1	RW	This field is not meaningful.
6	rsp_tt_index_err	1	RW	This field is not meaningful.
5	slv_tt_index_err	1	RW	This field is not meaningful.
4	zcp_tt_index_err	1	RW	This field is not meaningful.
3	cfifo_ecc3	1	RW	This field is not meaningful.
2	cfifo_ecc2	1	RW	This field is not meaningful.
1	cfifo_ecc1	1	RW	Port 1 control FIFO ECC error. 1 = Enable interrupt generation; 0 = Disable interrupt generation (default). When it equals 0: enable interrupt generation.
0	cfifo_ecc0	1	RW	Port 0 control FIFO ECC error. 1 = Enable interrupt generation; 0 = Disable interrupt generation (default). When it equals 0: enable interrupt generation.

TABLE 25-95 BAM 4 Buffer Region Control – BAM4BUF (FZC_ZCP + 00018_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	loj	0	RW	LOJ. For register RW test.
30	en_ck	0	RW	Set to 1 to enable range checking on zcfid{9:0}.
29:20	idx_end0	0	RW	Last valid value of zcfid{9:0}.
19:10	idx_st0	0	RW	First valid value of zcfid{9:0}
9:0	offset0	0	RW	Row index offset into BAM to define the region.

TABLE 25-96 BAM 8 Buffer Region Control – BAM8BUF (FZC_ZCP + 00020_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	loj	0	RW	LOJ. For register RW test.
30	en_ck	0	RW	Set to 1 to enable range checking on zcfid{9:0}.

TABLE 25-96 BAM 8 Buffer Region Control – BAM8BUF (FZC_ZCP + 00020_{16})

Bit	Field	Initial Value	R/W	Description
29:20	idx_end1	0	RW	Last valid value of zcfid{9:0}.
19:10	idx_st1	0	RW	First valid value of zcfid{9:0}
9:0	offset1	0	RW	Row index offset into BAM to define the region.

TABLE 25-97 BAM 16 Buffer Region Control – BAM16BUF (FZC_ZCP + 00028_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	loj	0	RW	LOJ. For register RW test.
30	en_ck	0	RW	Set to 1 to enable range checking on zcfid{9:0}.
29:20	idx_end2	0	RW	Last valid value of zcfid{9:0}.
19:10	idx_st2	0	RW	First valid value of zcfid{9:0}
9:0	offset2	0	RW	Row index offset into BAM to define the region.

TABLE 25-98 BAM 32 Buffer Region Control – BAM32BUF (FZC_ZCP + 00030_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	loj	0	RW	LOJ. For register RW test.
30	en_ck	0	RW	Set to 1 to enable range checking on zcfid{9:0}.
29:20	idx_end3	0	RW	Last valid value of zcfid{9:0}.
19:10	idx_st3	0	RW	First valid value of zcfid{9:0}
9:0	offset3	0	RW	Row index offset into BAM to define the region.

TABLE 25-99 DynamicStaticTable 4 Buffer Region Control – DST4BUF (FZC_ZCP + 00038_{16})

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9:0	ds_offset0	0	RW	Row index offset into dynamic static table RAM to define the region.

TABLE 25-100 DynamicStaticTable 8 Buffer Region Control – DST8BUF (FZC_ZCP + 00040_{16})

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9:0	ds_offset1	0	RW	Row index offset into dynamic static table RAM to define the region.

TABLE 25-101 DynamicStaticTable 16 Buffer Region Control – DST16BUF (FZC_ZCP 00048_{16})

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9:0	ds_offset2	0	RW	Row index offset into dynamic static table RAM to define the region.

TABLE 25-102 DynamicStaticTable 32 Buffer Region Control – DST32BUF (FZC_ZCP + 00050_{16})

Bit	Field	Initial Value	R/W	Description
63:10	—	0	RO	*Reserved*
9:0	ds_offset3	0	RW	Row index offset into dynamic static table RAM to define the region.

ZCP RAM Access Process

There are two ZCP RAM access mechanisms. The way in which the address is generated differentiates the two. **[Make it clear which is the 2nd mechanism]**

Disable the zero copy processor state machine by resetting configuration register bit 0 (zc_enable) to zero. Except for the Control FIFO, all write operations should come with write enabled.

- Use ZCP RAM Access register bit 11:0 (cfifoaddr) as address to access CFIFOs.
- Use ZCP RAM Access register bit 28:17 (zcfid) as address to access BAM, ST_RAM, DN_RAM.

When accessing BAM, ST_RAM, DN_RAM, the offset value should be reset to zero.

TABLE 25-103 ZCP RAM Access Data 0 – ZCP_RAM_DATA0 (FZC_ZCP + 00058_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat0	0	RW	Load this register with write data bit 31: 0 value. A PIO write operation followed by a PIO read operation will read the contents of this register, not the contents of the internal RAM. This register contains a read data from internal RAM after a PIO read to the internal RAM operation is done.

TABLE 25-104 ZCP RAM Access Data 1 – ZCP_RAM_DATA1 (FZC_ZCP + 00060_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat1	0	RW	Load this register with write data bit 63:32 value. A PIO write operation followed by a PIO read operation will read the contents of this register, not the contents of the internal RAM. This register contains a read data from internal RAM after a PIO read to the internal RAM operation is done.

TABLE 25-105 ZCP RAM Access Data 2 – ZCP_RAM_DATA2 (FZC_ZCP + 00068_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat2	0	RW	Load this register with write data bit 95:64 value. A PIO write operation followed by a PIO read operation will read the contents of this register, not the contents of the internal RAM. This register contains a read data from internal RAM after a PIO read to the internal RAM operation is done.

TABLE 25-106 ZCP RAM Access Data 3 – ZCP_RAM_DATA3 (FZC_ZCP + 00070_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	dat3	0	RW	Load this register with write data bit 127:96 value. A PIO write operation followed by a PIO read operation will read the contents of this register, not the contents of the internal RAM. This register contains a read data from internal RAM after a PIO read to the internal RAM operation is done.

TABLE 25-107 ZCP RAM Access Data 4 – ZCP_RAM_DATA4 (FZC_ZCP + 00078_{16})

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	dat4	0	RW	Load dat4{3:0} of this register with write data bit 131:128 value for accessing CFIFO. Rest of bits dat4{7:4} are not used when accessing CFIFO. Load this register with write data bit 135: 128 when accessing dynamic table RAM, which is 136 bits wide. Static table RAM is 112 bits wide. A PIO write operation followed by a PIO read operation will read the contents of this register, not the contents of the internal RAM. This register contains a read data from internal RAM after a PIO read to the internal RAM operation is done.

The following register is not used for UltraSPARC T2.

TABLE 25-108 ZCP RAM Access Byte Enable – ZCP_RAM_BE0 (FZC_ZCP + 00080_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16:0	be	0	RW	Load this register with byte write enable info. For access to buffer RAM (va_ram), each be bit enables 32 buffer RAM (va_ram) bit enable. For access to buffer RAM (va_ram), each be bit enables 32 buffer RAM (va_ram) bit enable. Only the lower 4 bits are used for accessing one of the 8 banks of buffer RAM (va_ram). For access to dynamic table ram (dn_ram) or static table RAM (st_ram), each be bit enables 8 RAM bit enable. Each bit in ZCP RAM Access Byte Enable register can enable 1 byte of dynamic table RAM/ static table RAM data. For dynamic table RAM, only 17 bits are needed. For static table RAM, only 14 bits are needed. See *RAM_SEL Programming Guide*, on page 747 for details.

TABLE 25-109 ZCP RAM Access – ZCP_RAM_ACC (FZC_ZCP + 00088_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	busy	0	RO	RAM access busy (RAM access not done). 0 = Read/write operation is done (default); 1 = Read/write operation is not done yet. Write to this register will trigger a read/write operation depending on the bit 30 (rdwr).
30	rdwr	0	RW	0 = Write; 1 = Read.
29	loj	0	RW	LOJ function.
28:17	zcfid	0	RW	Zero copy flow ID. It is used as address for accessing va_ram, st_ram, dn_ram. *Reserved*.
16:12	ram_sel	0	RW	RAM access select. Selects one of the RAMs for accessing. Please see *RAM_SEL Programming Guide*, below, for detail.
11:0	cfifoaddr	0	RW	Control FIFO address, used for debugging.

RAM_SEL Programming Guide RAM_access_sel (5b)

Program this value to enable access (r/w) the control FIFO RAM (130 bits wide). Use CFIFOADDR in bit 11:0 as RAM address.

= 5′h10: cfifo0
= 5′h11: cfifo1
All other values reserved.

These registers provide 128 bit (16 byte) wide data access port for user to read or write to the internal ZCP memory. The selection of internal memory can be found in TABLE 25-109 on page 747.

Software should set up ZCP RAM Access Data Register first and then write to the ZCP RAM Access register to kick off the write-to-memory operation.

An internal RAM read operation is accomplished by writing to the ZCP RAM Access register. Software then comes back to read the data register if the RAM Access busy bit (bit 31) of the ZCP RAM Access register is zero.

Address bit 16 is used to divide the regular ZCP registers and receive DMA channel number table.

When *address*{16} = 0, it points to regular ZCP registers.

When address{16} = 1, it points to receive DMA channel number table.

Notes: Unused address bits 18:17, 15:12 should be driven to zero.

25.10 Error Handling Registers

TABLE 25-110 Check Bit Data – CHK_BIT_DATA (FZC_ZCP + 00090_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16:0	chk_bit_data	0	RW	Use this register to inject parity error to translation RAMs. Each bit will corrupt one parity bit.

TABLE 25-111 Reset Control FIFO – RESET_CFIFO (FZC_ZCP + 00098_{16})

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
3	rst_cfifo3	0	RW	This bit is not meaningful in UltraSPARC T2 mode.
2	rst_cfifo2	0	RW	This bit is not meaningful in UltraSPARC T2 mode.
1	rst_cfifo1	0	RW	Write 1 to reset cfifo1. Write 0 to deassert reset cfifo1.
0	rst_cfifo0	0	RW	Write 1 to reset cfifo0. Write 0 to deassert reset cfifo0.

TABLE 25-112 Control FIFO ECC Port 0 – CFIFO_ECC_PORT0 (FZC_ZCP + $000A0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	disable_double_bit_err	0	RW	Disable double bit error. Bypass ECC correction logic.
30:18	—	0	RO	*Reserved*
17	double_bit_err	0	RW	Double-bit error.
16	single_bit_err	0	RW	Single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkt	1	RO	Always set this bit to 1.
9:3	—	0	RO	*Reserved*
2	last_line_of_pkt	0	RW	Inject ECC error on last line of packet.
1	second_line_of_pkt	0	RW	Inject ECC error on second line of packet.
0	first_line_of_pkt	0	RW	Inject ECC error on first line of packet.

TABLE 25-113 Control FIFO ECC Port 1 – CFIFO_ECC_PORT1 (FZC_ZCP + $000A8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	disable_double_bit_err	0	RW	Disable double bit error. Bypass ECC correction logic.
30:18	—	0	RO	*Reserved*
17	double_bit_err	0	RW	Double-bit error.
16	single_bit_err	0	RW	Single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkt	1	RO	Always set this bit to 1.
9:3	—	0	RO	*Reserved*
2	last_line_of_pkt	0	RW	Inject ECC error on last line of packet.
1	second_line_of_pkt	0	RW	Inject ECC error on second line of packet.
0	first_line_of_pkt	0	RW	Inject ECC error on first line of packet.

Control FIFO ECC Port 2 and 3 are not used in UltraSPARC T2.

TABLE 25-114 Control FIFO ECC Port 2 – CFIFO_ECC_PORT2 (FZC_ZCP + $000B0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	disable_double_bit_err	0	RW	Disable double bit error. Bypass ECC correction logic.
30:18	—	0	RO	*Reserved*
17	double_bit_err	0	RW	Double-bit error.

TABLE 25-114 Control FIFO ECC Port 2 – CFIFO_ECC_PORT2 (FZC_ZCP + $000B0_{16}$)

Bit	Field	Initial Value	R/W	Description
16	single_bit_err	0	RW	Single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkt	1	RO	Always set this bit to 1.
9:3	—	0	RO	*Reserved*
2	last_line_of_pkt	0	RW	Inject ECC error on last line of packet.
1	second_line_of_pkt	0	RW	Inject ECC error on second line of packet.
0	first_line_of_pkt	0	RW	Inject ECC error on first line of packet.

TABLE 25-115 Control FIFO ECC Port 3 – CFIFO_ECC_PORT3 (FZC_ZCP + $000B8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	disable_double_bit_err	0	RW	Disable double bit error. Bypass ECC correction logic.
30:18	—	0	RO	*Reserved*
17	double_bit_err	0	RW	Double-bit error.
16	single_bit_err	0	RW	Single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkt	1	RO	Always set this bit to 1.
9:3	—	0	RO	*Reserved*
2	last_line_of_pkt	0	RW	Inject ECC error on last line of packet.
1	second_line_of_pkt	0	RW	Inject ECC error on second line of packet.
0	first_line_of_pkt	0	RW	Inject ECC error on first line of packet.

TABLE 25-116 Training Vector – TRAINING_VECTOR (FZC_ZCP + $000C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	training_vector	0	RW	Training vector

TABLE 25-117 State Machine – STATE_MACHINE (FZC_ZCP + $000C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	state_machine	0	RO	{0_2, kickoff_tt_reg, ififo_state{2:0}, rsp_unload_state{2:0}, rsp_load_state{2:0}, req_unload_state{1:0}, req_load_state{1:0}, tt_state{3:0}};

Note that the following registers have side effects.

- ZCP_INT_STAT
- ZCP_INT_STAT_TEST
- CFIFO_ECC_ERR_CTL_PORT0
- CFIFO_ECC_ERR_CTL_PORT1
- ZCP_RAM_ACC

25.11 Errata

25.11.1 Address Relocation For Jumbo Packet

Receive DMA produces an incorrect address translation for the second or third DMA transection for a Jumbo frame. This happens when:

1. Jumbo frame splited to multiple buffers with second or third transaction using logical page zero;

2. PIO disable DMA comes between two transactions of

   this jumbo frame; and

3. One DMA is shared by two (or more) ports.

Currently, software does not expect to share a single DMA among different physical ports, thus, this bug will not be fixed in the current design.

25.11.2 Minimum RxDMA packet size

Packets have to be at least 64B to be transfered to the host. This is achieved by not removing the Ethernet CRCs from receive packets, and setting the discard option for any runt packets from MAC. The latter will cause the RxDMA to discard the runt packets.

During debug phase, when traffic is light and packets are spaced out, then hardware will be able to handle runt frames.

CHAPTER 26

Network Interface Unit: Transmit DMA Channels

The transmit side supports a total of 16 Transmit DMA Channels, shared among the Ethernet interfaces. Note that the association of PIO control to a DMA, if accessed through the virtualization region, is defined in the PIO block. How addresses are relocated and the association of these requests with a function is defined per DMA.

Each Transmit DMA Channel consists of a transmit ring and a set of status registers. Software creates a list of packet descriptors in the transmit ring in system memory and informs the hardware where the tail is. Hardware will then fetch the descriptors into the chip and transmit the packets in the same order as they are in the ring. Each gather list consists of multiple descriptors.

The bandwidth usage among different transmit queues of a port is allocated according to a deficit round-robin scheduler. The advantage of this scheme is that it supports an allocation based on packet size, as opposed to number of packets.

Transmit logic also supports checksum offload. Checksum is calculated for TCP/UDP payload. The computation is performed along the datapath as the packet is read from memory.

To support partitioning, two logical pages are defined for each DMA channel. DMA control data structures and packet buffer addresses will be translated accordingly. We first present the definition of logical pages.

26.1 Partitioning Support

Each DMA supports two naturally aligned, physically contiguous areas in system memory. The definition is identical to that on the Receive DMA channels. The following registers are defined. *Note that all registers defined in this section can only be changed if the DMA is disabled.*

TABLE 26-1 Transmit Logical Page Valid – TX_LOG_PAGE_VLD (FZC_DMC + 40000_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	0	RO	*Reserved*
3:2	func	0	RW	Assigned device function number. Only 00_2 and 01_2 are valid for UltraSPARC T2.
1	page1	0	RW	Set to 1 to indicate page 1 is defined.
0	page0	0	RW	Set to 1 to indicate page 0 is defined.

TABLE 26-2 Transmit Logical Page Mask 1 – TX_LOG_MASK1 (FZC_DMC + 40008_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	mask	0	RW	Logical page mask for bits 43:12. Set the corresponding bit to 1 to select the bit. If the mask bits are all zeros, then the address definition is an automatic pass, and there is no need to perform a translation.

TABLE 26-3 Transmit Logical Page Value 1 – TX_LOG_VALUE1 (FZC_DMC + 40010_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	value	0	RW	Value used to compare address bits 43:12 selected by mask. Note that bits not selected by the mask register should be set to 0.

TABLE 26-4 Transmit Logical Page Mask 2 – TX_LOG_MASK2 (FZC_DMC + 40018_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	mask	0	RW	Logical page mask for bits 43:12. Set the corresponding bit to 1 to select the bit. If the mask bits are all zeros, then the address definition is an automatic pass, and there is no need to perform a translation.

TABLE 26-5 Transmit Logical Page Value 2 – TX_LOG_VALUE2 (FZC_DMC + 40020_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	value	0	RW	Value used to compare address bits 43:12 selected by mask. Note that bits not selected by the mask register should be set to 0.

TABLE 26-6 Transmit Logical Page Relocation 1 – TX_LOG_PAGE_RELO1 (FZC_DMC + 40028_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	relo	0	RW	Value used to replace bits 43:12 selected by the mask field.

TABLE 26-7 Transmit Logical Page Relocation 2 – TX_LOG_PAGE_RELO2 (FZC_DMC + 40030_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	relo	0	RW	Value used to replace bits 43:12 selected by mask field.

TABLE 26-8 Transmit Logical Page Handle – TX_LOG_PAGE_HDL (FZC_DMC + 40038_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	handle	0	RW	Logical page handle, bits 63:44 of a 64-bit address, used to concatenate to a 44-bit address to generate a 64-bit address.

An address is considered invalid if it is outside both pages. Thus, software needs to program in at least one valid page. When the mask is programmed with all zeros, then any address is considered valid and none of the bits will be replaced with the value in relo. All the addresses supplied by software, data structure configuration addresses, and the packet gather addresses are subjected to translation.

TABLE 26-9 Transmit Address Mode – TX_ADDR_MD (FZC_DMC + 45000_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	mode32	0	RW	This bit musto be set to 0..

26.2 Packet Descriptor Structure

A packet is described as a gather list. Each element of the list is an 8-byte entity, which we called a gather pointer, and has the following format.

TABLE 26-10 Format: Gather Pointer

Bit	Field	Description
63	sop	Start of packet.
62	mark	Only applicable when the sop bit is set. It instructs hardware to issue an interrupt after the packet is transmitted if interrupt is enabled.
61:58	num_ptr	Only applicable if sop is set. It indicates the number of gather pointers in this packet.
57	—	*Reserved*
56:44	tr_len	Transfer length, number of bytes starting from sad that forms part of a packet. The maximum is 4 Kbytes. A zero is interpreted as zero. A value larger than 4 Kbytes is an error. The maximum lengthUltraSPARC T2 is 4096 bytes.
43:0	sad	Start address of the data block.

A packet can be described as a list of pointers shown above. The first pointer of the list should have the sop bit set. num_ptr is the number of gather pointers for the packet. If mark is set, the CSRs associated with the DMA channel may be updated to the mailbox, and an LDF may be raised after the packet is completely transmitted. The mark bit will be ignored if the sop bit is not set. The size of the first segment should be at least 32B, which includes the entire transmit header.

Note that there is a hardware limit on the size of the packet. Since the transmit function is implemented with store and forward, hardware needs to buffer the entire packet, perform the necessary checksum calculation, then transmit the packet. This implies that there is a logical transmit FIFO per port. The transmit FIFO thus determines the maximum size of a packet. Currently, the FIFO is sized to support jumbo frame of size 9 Kbytes, with arbitrary buffer alignment. Hardware will keep

track of the size of the packet when issuing packet requests to the system interface. When the packet size exceeds the limit, the DMA will go into an error state. It will cause an LDF to be raised if selected.

Physically, hardware supports a store-and-forward packet FIFO of 606 entries deep, 16 bytes per entry, or 9696 bytes total. This has to hold the 16-byte internal header (to be described in a later section), padding (2 bytes to 14 bytes), jumbo frame (up to 9 Kbytes, not including the Ethernet CRC, which is generated by the MAC), and potential alignment gaps between gather pointer (15 bytes for the first pointer, plus up to 15 + 15 or 30 bytes per gather pointer for 14 pointers). Thus, if software desires to transmit a frame larger than 9 Kbytes, the absolute maximum may be achieved by copying the packet to a 16-byte naturally aligned buffer, with the only overhead being the internal 16-byte header. This limit is 9680 bytes.

26.3 Transmit Ring Configuration

Transmit descriptors are organized as a FIFO using a ring buffer in main memory. After software created one or more descriptors, it writes the tail pointer to signal hardware that new descriptors have been deposited. Hardware will then read the descriptors into the chip and process the packets automatically until the tail of the queue is reached. Internally, a prefetch buffer is associated with each transmit ring. This reduces the per packet overhead of reading from system memory.

The transmit scheduler processes the gather pointers. A packet is ready for transmission if the gather pointers of the entire packet are in the local on-chip prefetch buffer. Before a request can be issued to the memory system, a buffer from the per-port transmit FIFO is first allocated. This is necessary so that the data returned from memory can be reordered. If the reorder buffer is not available, the port will be stalled.

A 12-bit wrap-around counter, the packet counter (pkt_cnt), keeps track of the packet transmission. This counter is initialized to zero. For every packet transmitted, the counter is incremented. A packet is transmitted when the transmit FIFO received the entire packet.

The programming interface of the transmit rings is as follows. Also included are the registers associated with the states of the transmit ring. TX_RNG_CFIG register cannot be changed while the DMA is enabled. When the DMA is enabled, the address checking will occur.

TABLE 26-11 Transmit Ring Configuration - TX_RNG_CFIG (DMC + 40000_{16})

Bit	Field	Initial Value	R/W	Description
63:61	—	0	RO	*Reserved*
60:48	len	0	RW	Maximum length of the transmit queue ring buffer, measured in number of 64-byte blocks.
47:44	—	0	RO	*Reserved*
43:19	staddr_base	0	RW	Address bits 43:19.
18:6	staddr	0	RW	Bits 18:6 of the start address for the transmit ring buffer. Bits 5:0 are assumed to be zero, or 64-byte aligned.
5:0	—	0	RO	*Reserved*

TABLE 26-12 Transmit Ring Head Low – TX_RING_HDL (DMC + 40010_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19	wrap	0	RO	Hardware will toggle this bit every time the head is wrapped around the configured ring buffer.
18:3	head	0	RO	This is specified as an offset, in number of entries, from the starting address, which is the concatenation of (staddr_base, (staddr<<6)). The actual head pointer is (staddr_base, (staddr<<6)) + head<<3.
2:0	—	0	RO	*Reserved*

After posting transmit descriptors to the transmit ring, the tail pointer is updated by software to start the transmission. *Software can only post descriptors through TX_RNG_KICK when the entire packet is in the ring. Hardware may deadlock if a partial packet is posted.*

TABLE 26-13 Transmit Ring Kick – TX_RING_KICK (DMC + 40018_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19	wrap	0	RW	Software needs to toggle this bit every time the tail is wrapped around the configured ring buffer.
18:3	tail	0	RW	This is specified as an offset, in number of entries, from the starting address, which is the concatenation of (staddr_base, (staddr<<6)). The actual tail pointer is (staddr_base, (staddr<<6)) + tail<<3.
2:0	—	0	RO	*Reserved*

26.4 Transmit DMA Operation

The Transmit Event Mask register lists the events that can raise the LDF. Software needs to set the corresponding bit to 0 to select the event. Similar to the receive side, there are two LDFs. TABLE 26-14 details which event will raise which flag.

TABLE 26-14 Transmit Event Mask – TX_ENT_MSK (DMC + 40020_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	mk	1	RW	Set to 0 to select the event to raise the LDF for packets marked. An LDF 0 event.
14:8	—	0	RO	*Reserved*
7	mbox_err	1	RW	Set to 0 to select the event to raise the LDF when error occurred during mailbox update. An LDF 1 event.
6	pkt_size_err	1	RW	Set to 0 to select the event to raise the LDF when the packet size exceeds hardware limit. An LDF 1 event.
5	tx_ring_oflow	1	RW	Set to 0 to select the event to raise the LDF associated with Transmit Ring overflow. An LDF 1 event.
4	pref_buf_ecc_err	1	RW	Set to 0 to select the event to raise the LDF for nonrecoverable ECC error on prefetch buffer. An LDF 1 event.
3	nack_pref	1	RW	Set to 0 to select the event to raise the LDF for NACK or timeout occurred on prefetch buffer read. An LDF 1 event.
2	nack_pkt_rd	1	RW	Set to 0 to select the event to raise the LDF for NACK or timeout occurred on packet read. An LDF 1 event.
1	conf_part_err	1	RW	Set to 0 to select the event to raise the LDF for partition violation during configuration. This includes the transmit ring and mailbox configuration. An LDF 1 event.
0	pkt_prt_err	1	RW	Set to 0 to select the event to raise the LDF for partition violation during packet read. An LDF 1 event.

TABLE 26-15 on page 761 describes the bits discussed in this section. The transmit DMA is disabled on power-on. After configuring the transmit ring, software needs to clear the rst_state bit to start the operation. Software may also set a mailbox for hardware to update the status registers to system memory after transmitting a marked packet. The DMA channel may be reset by writing a 1 to the rst bit. When reset is completed, hardware will set the rst_state bit to 1. The stop_n_go bit is used to halt the DMA operation and is used primarily for debugging purpose. Hardware will stop the DMA operation at packet boundary. This is necessary since other DMAs may be sharing the same port. The sng_state bit will be set when hardware stopped the DMA. Clearing the stop_n_go bit will enable hardware to resume the DMA operation and start transmitting packets again.

The proper sequence to bring up a DMA should be to first reset the DMA, then configure the transmit ring, and finally clear the rst_state of the DMA.

pkt_cnt is a wrap-around counter indicating the number of packets transmitted. lastmark stores the pkt_cnt value of the packet with the mark bit set. This two counts keep track of how many packets have been transmitted; software may use these values to reclaim packet buffers. This will require software to keep a similar count. Note that the read or write operation to the control register is considered atomic. A read will return the current value of the register to software. mk and mmk bits will be cleared after the read.

An interrupt-driven driver may need to follow the behavior below to properly control the DMA.

If the mailbox feature is enabled, when an interrupt occurs, software may read the status in the mailbox to determine what caused the interrupt and the associated processing required. The hardware status register (TABLE 26-15) will keep the status even if the status has been updated to the mailbox. Before the driver exits the service routine, if the interrupt is associated with packet completion, the driver may simply write 1 to the mb bit to enable the hardware to write to the mailbox when an event occurred and a 1 to the mk bit to reenable the DMA to raise the LDF. Note that the mmk bit keeps track of the transmit queue if more packets with the mark bit set have been completed since the first marked packet. If the mmk bit is set, writing 1 to the mk bit will clear the mmk bit, but the mk bit will remain set. This will cause the LDF to maintain the current value.

Alternatively, at the end of the driver routine, software may read the register to determine if more packets have completed. In this case, hardware will return the state information, along with the lastmark and pkt_cnt information. The mk bit and mmk bits will be cleared on a read. Hardware assumes that software will complete the processing associated with these packets. After the register read, software should *not* write a 1 to the mk bit, since the previous read has already cleared this bit. The mk bit now tracks the completion ring behavior after the last read. Software should simply write a 1 to the mb bit to reenable the update to the mailbox if desired.

If the mailbox feature is not used, software needs to read the status register to determine how many packets have completed. This read will clear the mk and mmk bits.

For polling, software needs to read the status register to determine the progress. The mk and mmk bits are cleared on read.

TABLE 26-15 is the control register for the transmit DMA. For 32-bit systems, read and write to the lower order 32 bits of the Transmit Control and Status register is consistent and can be considered independent of the higher-order 32 bits. Note that setting the rst bit to 1 and rst_state bit to 0 in the same store operation is not recommended. Software should first set the rst bit to 1, wait until hardware has completed the reset operation and then enable the DMA by clearing the rst_state bit.

TABLE 26-15 Transmit Control And Status – TX_CS (DMC + 40028_{16})

Bit	Field	Initial Value	R/W	Description
63:60	—	0	RO	*Reserved*
59:48	pkt_cnt	0	RO	A wrap-around counter to keep track of packets transmitted. Reset to zero when the DMA is reset
47:44	—	0	RO	*Reserved*
43:32	lastmark	0	RO	The pkt_cnt corresponds to the last packet with the mark bit set. Reset to zero when the DMA is reset.
31	rst	0	RW	Set to 1 to reset the DMA. Hardware will clear this bit after reset is completed. A reset will bring the specific DMA back to the clean state, and software is required to reinitialize the DMA before starting. Note that if the DMA is in the process of transmitting a packet or fetching the transmit ring when the rst bit is set, hardware will only reset the DMA after those operations are completed. Note that there may be interrupt in transit to the CPU if the DMA is not in `error_state`. Note that rst_state bit will be set to 1 after reset.
30	rst_state	1	RW	If this bit is equal to 1, the DMA is in a state after reset. Software needs to clear this bit to start the DMA engine. On power-on or after reset, this bit is set to 1. A write of 1 to this bit is ignored.
29	mb	0	RW	Set to 1 to enable HW to update the mailbox. Cleared to zero by hardware after update.
28	stop_n_go	0	RW	Set to 1 to stop the transmit DMA engine. Hardware will complete the processing of the current packet. Clearing this bit will enable hardware to resume the packet transmission.
27	sng_state	0	RO	Set to 1 when the DMA engine is stopped. If hardware is in the process of transmitting a packet, this bit will only be set to 1 after the current packet transmission is completed. When the stop_n_go bit is cleared, this bit is cleared to 0.
26:16	—	0	RO	*Reserved*
15	mk	0	RCW1C	Set to 1 to indicate a packet with the mark bit set is transmitted. Software may read the register to clear the bit. Alternatively, software may write a 1 to clear the mk bit (Write 0 has no effect). In the case of write 1, if mmk bit is set, mk bit will *not* be cleared. This bit is used to generate LDF 0.
14	mmk	0	RC	Set to 1 by hardware if the mk bit is set when a packet with the mark bit set is completed. Software reads to clear. A write 1 to the mk bit will also clear the mmk bit.
13:8	—	0	RO	*Reserved*
7	mbox_err	0	RO	Set to 1 to indicate mailbox update error, either no hardware acknowledgment or data return with error. Fatal error.
6	pkt_size_err	0	RO	Set to 1 when the packet size exceeds hardware limit. Part of LDF 1. Fatal error.

TABLE 26-15 Transmit Control And Status – TX_CS (DMC + 40028_{16}) *(Continued)*

Bit	Field	Initial Value	R/W	Description
5	tx_ring_oflow	0	RO	Set to 1 to indicate transmit ring overflow. Software writes 1 to clear. Fatal error. Clear on reset. Part of LDF 1.
4	pref_buf_par_err	0	RO	Set to 1 to indicate nonrecoverable parity error on prefetch buffer occurred. Write 1 to clear. Fatal error. Clear on reset. Part of LDF 1.
3	nack_pref	0	RO	Set to 1 to indicate NACK or timeout occurred on prefetch buffer read. Write 1 to clear. Fatal error. Clear on reset. Part of LDF 1.
2	nack_pkt_rd	0	RO	Set to 1 to indicate NACK or timeout occurred on packet read. Write 1 to clear. Fatal Error. Clear on reset. Part of LDF 1.
1	conf_part_err	0	RO	Set to 1 to indicate partition violation during configuration. This includes the transmit ring and mailbox configuration. Write 1 to clear. Fatal error. Clear on reset. Part of LDF 1.
0	pkt_prt_err	0	RO	Set to 1 to indicate partition violation during packet read. Write 1 to clear. Fatal error. Clear on reset. Part of LDF 1.

The following two registers are used for debugging.

TABLE 26-16 Transmit DMA Interrupt Debug – TDMC_INTR_DBG (DMC + 40060_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	mk	0	RW	Write 1 to force LDF for packet marked. An LDF0 event
14:8	—	0	RO	*Reserved*
7	mbox_err	0	RW	Write 1 to force LDF when error occurred during mailbox update. An LDF1 event
6	pkt_size_err	0	RW	Write 1 to force LDF when the packet size exceeds hardware limit. An LDF 1 event.
5	tx_ring_oflow	0	RW	Write 1 to force LDF associated with transmit ring overflow. An LDF 1 event.
4	pref_buf_par_err	0	RW	Write 1 to force LDF for nonrecoverable ECC error on prefetch buffer. An LDF 1 event.
3	nack_pref	0	RW	Write 1 to force LDF for NACK or timeout occurred on prefetch buffer read. An LDF 1 event.
2	nack_pkt_rd	0	RW	Write 1 to force LDF for NACK or timeout occurred on packet read. An LDF 1 event.
1	conf_part_err	0	RW	Write 1 to force LDF for partition violation during configuration. This includes the transmit ring and mailbox configuration. An LDF 1 event.
0	pkt_part_err	0	RW	Write 1 to force LDF for partition violation during packet read. An LDF 1 event.

TABLE 26-17 Transmit Control And Status Debug – TX_CS_DBG (DMC + 40068_{16})

Bit	Field	Initial Value	R/W	Description
63:60	—	0	RO	*Reserved*
59:48	pkt_cnt	0	RW	Used to initialize the pkt_cnt value in TX_CS register for debugging.
47:0	—	0	RO	*Reserved*

The following two registers are used to configure the mailbox. The mailbox address is check when the mailbox enable bit (mb) in the TX_CS register is set.

TABLE 26-18 TXDMA Mailbox High - TXDMA_MBH (DMC + 40030_{16})

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11:0	mbaddr	0	RW	Bits 43:32 of the mailbox address.

TABLE 26-19 TXDMA Mailbox Low - TXDMA_MBL (DMC + 40038_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:6	mbaddr	0	RW	Bits 31:6 of the mailbox address. Bits 5:0 are assumed to be zero or 64-byte aligned.
5:0	—	0	RO	*Reserved*.

The following registers describe the internal state associated with the transmit DMA prefetch buffer.

TABLE 26-20 Transmit DMA Prefetch State – TX_DMA_PRE_ST (DMC + 40040_{16})

Bit	Field	Initial Value	R/W	Description
63:19	—	0	RO	*Reserved*
18:0	shadow_hd	0	RO	Indicates how far hardware has prefetched, measured as an offset from the base.

When a fatal error occurs, the DMA channel will be in a halt state and a reset is required to reinitialize the DMA. The following registers log the address associated with a fatal error if applicable.

TABLE 26-21 Transmit Ring Error Log High - TX_RNG_ERR_LOGH (DMC + 40048_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	err	0	RO	Set to 1 if an error occurred. Reset to 0 when the DMA is reset.
30	merr	0	RO	Multiple errors occurred when set.
29:27	errcode	0	RO	The error code of the first error. 000_2 – PKT_PRT_ERR (log logical address); 001_2 – CONF_PART_ERR (log logical address); 010_2 – NACK_PKT_RD (log translated address); 011_2 – NACK_PREF (log translated address); 100_2 – PREF_BUF_PAR_ERR (log prefetch buffer address); 101_2 – TX_RING_OFLOW (zeros, no address log); 110_2 – PKT_SIZE_ERR (log translated address); 110_2 – 111_2: *Reserved*.
26:12	—	0	RO	*Reserved*
11:0	err_addr	0	RO	The address 43:32 that caused the first error if applicable. For internal errors, the address of the entry that caused the error is logged. Note that hardware only logs the relocated address.

TABLE 26-22 Transmit Ring Error Log Low - TX_RNG_ERR_LOGL (DMC + 40050_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	err_addr	0	RO	The address 31:0 that caused the first error if applicable. For internal errors, the address of the entry that cause the error is logged. Note that hardware only logs the relocated address.

The format of the mailbox is shown TABLE 26-23.

TABLE 26-23 Format: Transmit DMA Channel Mailbox

Register	Byte	Description
TX_DMA_PRE_ST	0-7	
TX_CS	8-15	
TX_RING_KICK	16-23	
TX_RING_HDL	24-31	
—	32-39	*Reserved*. Set to 0.
TX_RNG_ERR_LOGL	40-43	
TX_RNG_ERR_LOGH	44-47	
—	48-63	*Reserved*. Set to 0.

26.5 Transmit Ring Scheduler

The binding of transmit ring to physical port is determined by the TXC Port DMA Enable register. Hardware does not check the consistency of the value. If improperly specified, that is, the same queue bound to more than one port, the behavior is not defined.

In UltraSPARC T2, a total of 16 transmit DMAs can be assigned to the two ports.

TABLE 26-24 TXC Port DMA Enable – TXC_PORT_DMA (FZC_TXC + 20028_{16}))

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	port_dma_list	0	RW	Set to 1 to bind DMA to port 0. DMAs are based on bit position. DMA bind must be exclusive, i.e., same DMA cannot be bound to multiple ports.

The scheduling is achieved using the Deficit Round Robin (DRR) Queuing algorithm. First, we give a brief description on the algorithm.

Conceptually, we allocate to each "flow" a certain percentage of capacity. DRR is similar to a round-robin schedule in that every flow will be serviced in a predefined order. Empty queues will be skipped. Each flow maintains a counter as state. On power-up, the counters are initialized to zero. At the beginning of each round, a weight (possibly different for each flow) will be added to the counter of each flow. For every unit (in our design, each unit is up to 64 bytes) of data transmitted, the counter will be decremented by 1. As long as the counter remains positive and the flow is nonempty, the scheduler can process at least one packet from the flow. Thus, it is possible that the counter is negative at the end of a packet. When it is negative, the scheduler will switch to another flow once the current packet is processed.

At the beginning of next round, the weights will be added to the flows' counters again. If a counter's value is larger than the flow's assigned weight, the counter will be reset to the assigned weight. Since the counters keep track of the excess usage at the last round, some flows may receive less allocation for that round. In any case, the value of the counter shall be less than or equal to the programmed weight of the flow.

The pseudocode for the algorithm for one port is described below.

```
I :    = {the set of queues bound to a port}
C_i :  = deficit counters of queue i
W_i :  = assigned weight for queue i
```

A queue is eligible if it is nonempty. The "next queue in I" operation will return the first queue in I if the last queue is reached.

```
i = last queue in I;

select:i = next_queue_in_I;
   C_i = min [C_i + W_i, W_i];
   if (Queue i is not eligible || C_i <=0) goto select;
loop:process one packet from Queue i;
   decrement C_i accordingly;
   if (Queue i not eligible || C_i <= 0) goto select;
   goto loop;
```

The TXC DMA Max Burst register programs the weight of each transmit ring.

TABLE 26-25 TXC DMA Max Burst – TXC_DMA_MAX (FZC_TXC + 00000_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	dma_max_burst	0	RW	Max burst value associated with DMA. Used by DRR engine for computing when DMA has gone into deficit.

26.5.1 DRR Performance Monitoring

The following register is not functional. Software should not access this register. Refer to micro-architecture document for detail if needed. For performance tuning, higher layer software is required. For example, one may observe end to end ftp bandwidth achievable between different end nodes to determine if DRR parameters need to be adjusted.

TABLE 26-26 TXC DMA Max Burst Len – TXC_DMA_MAX_LEN (FZC_TXC + 00008_{16})

Bit	Field	Initial Value	R/W	Description
63:28	—	0	RO	*Reserved*
27:0	dma_length	0	RW	Refer to micro architecture document.

26.6 Internal Transmit Frame Header Format

The hardware supports L4 (only TCP, UDP) partial checksum. Hardware does not process IP header options. Partial checksum is defined as the checksum on the payload only. The value at the checksum field is used in the calculation as well.

The internal 16-byte frame header format is specified TABLE 26-27. The header needs to be prepended to every frame. Hardware *does not* support error checking. This frame header is at the beginning of the transmit packet and has to be naturally 16-byte aligned. Padding bytes may be inserted between the transmit header and the

Ethernet header, in multiples of 2 bytes. This enables the 16-byte transmit header alignment without copying. Hardware will remove these bytes before sending the packet out.

TABLE 26-27 Format: Ethernet Transmit Packet Header

Bit	Field	Description
127:64	—	*Reserved*
63:62	cksum_en_pkt_type	00_2 – No op, packet transmitted as is; 01_2 – TCP; 10_2 – UDP; 11_2 - *Reserved*.
61	ip_ver	0 = v4; 1 = v6.
60:58	—	*Reserved*
57	llc	Set to 1 to indicate if the packet is LLC/SNAP encapsulated.
56	vlan	Set to 1 to indicate if the packet is VLAN packet.
55:52	ihl	IP header length.
51:48	l3start	In multiples of 2 bytes, points to the byte location where the IP packet starts; the offset starts from the first byte of the Ethernet frame.
47:46	—	*Reserved*
45:40	l4start	In multiples of 2 bytes, points to the first byte of L4 payload; the counting starts from the first byte of Ethernet frame.
39:38	—	*Reserved*
37:32	l4stuff	In multiples of 2 bytes, points to the first byte position within L4 payload to put the checksum result; the counting starts from the first byte of Ethernet frame.
31:30	—	*Reserved*
29:16	tot_xfer_len	Total transfer length, the sum of all gather transfers less the 16-byte internal control header.
15:3	—	*Reserved*
2:0	pad	In multiple of 2 bytes, the number of bytes padded between the transmit header and the start of the Ethernet frame.

26.7 Hardware Control and Error Registers

This section contains hardware-specific registers for verification and debug.

The TDMC Inject Parity Error register is to be used for debug purposes only. Setting of any bit would inject a party error for the corresponding DMA. Upon injecting the error, software can detect the effect of this by enabling interrupts or by reading the status register. To reinject the parity, software will have to write a 0 and then write a 1.

For UltraSPARC T2, there are 16 channels.

TABLE 26-28 TDMC Inject Parity Error – TDMC_INJ_PAR_ERR (FZC_DMC + 45040_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	inject_parity_error	0	RW	Setting bit *i* would introduce parity error in the internal cache for the *i*th DMA.

The TDMC Training Vector register is to be used for debug purposes only. Exact encoding and selection of signal descriptions are described in micro-architecture document.

TABLE 26-29 TDMC Debug Select – TDMC_DBG_SEL (FZC_DMC + 45080_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	dbg_sel	0	RW	Setting various values will choose different debug signals for testmux.

TABLE 26-30 TDMC Training Vector – TDMC_TRAINING_VECTOR (FZC_DMC + 45088_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	vec	0	RW	Sets the training vector for testmux calibration. Writing a value of $3F_{16}$ in the debug select register would load this value into testmux logic. Writing a value of $0E_{16}$ in the debug select logic would generate an alternating pattern on the debug port once TDMC is selected.

The following are transmit datapath hardware registers.

For UltraSPARC T2, there are two MACs the TXC needs to enable separately.

TABLE 26-31 TXC Control – TXC_CONTROL (FZC_TXC + 20000_{16})

Bit	Field	Initial Value	R/W	Description
63:5	—	0	RO	*Reserved*
4	txc_enabled	0	RW	Set to 1 to globally enable TX controller. This is a master bit to enable and disable the full TXC controller.
3:2	rscd	0	RO	*Reserved*
1	port1_enabled	0	RW	Set to 1 to enable port 1 engine within TXC.
0	port0_enabled	0	RW	Set to 1 to enable port 0 engine within TXC.

TABLE 26-32 TXC Debug Select – TXC_DEBUG (FZC_TXC + 20010_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	debug_select	0	RW	TX controller debug select; decoding is as follows: Data Fetch Engine – 00_{16}; Port 0 States – 01_{16}; Port 1 States – 02_{16}; Port 2 States – 03_{16}; Port 3 States – 03_{16}; Training Set – $3E_{16}$ – Invert training vector every cycle; Training Load – $3F_{16}$ – Load training vector into debug port.

The TXC Maximum Reorder register sets the per-port reorder resources. Zero is an invalid value and undefined for hardware.

TABLE 26-33 TXC Maximum Reorder – TXC_MAX_REORDER (FZC_TXC + 20018_{16})

Bit	Field	Initial Value	R/W	Description
63:28	—	0	RO	*Reserved*
27:24	port3	7_{16}	RW	*Reserved*.
23:20	—	0	RO	*Reserved*
19:16	port2	7_{16}	RW	*Reserved*.
15:12	—	0	RO	*Reserved*
11:8	port1	F_{16}	RW	Port 1.
7:4	—	0	RO	*Reserved*
3:0	port0	F_{16}	RW	Port 0.

TABLE 26-34 TXC Port Control - TXC_PORT_CTL (FZC_TXC + 20020_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	clr_all_stat	0	WO	Write 1 to clear txc_pkt_stuffed and txc_pkt_xmit.

TABLE 26-35 TXC Packets Stuffed – TXC_PKT_STUFFED (FZC_TXC + 20030_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	packets_processed_reorder	0	RW	Total number of packets processed by reorder engine. This counter will roll over at $FFFF_{16}$.
15:0	packets_processed_packetassy	0	RW	Total number of packets processed by packetAssy engine. This counter will roll over at $FFFF_{16}$.

TABLE 26-36 TXC Packet Transmitted – TXC_PKT_XMIT (FZC_TXC + 20038_{16})

Bit	Field	Initial Value	R/W	Description
63:0	—	0	RO	*Reserved*
31:16	bytes_transmitted	0	RW	Total number of bytes transmitted to the MAC. This counter will roll over at $FFFF_{16}$.
15:0	packets_transmitted	0	RW	Total number of packets transmitted to the MAC. This counter will roll over at $FFFF_{16}$.

The following registers refer to the per-port transmit FIFO. Each consists of two FIFOs: a reorder FIFO (RO), which reorders the returning data segments into the request order, and a store and forward FIFO (SF), which stores the entire packet in FIFO and adds the checksum value if needed before sending the packet to the MAC.

TABLE 26-37 TXC RO ECC Control – TXC_ROECC_CTL (FZC_TXC + 20040_{16})

Bit	Field	Initial Value	R/W	
63:32	—	0	RO	*Reserved*
31	disable_ue	0	RW	Odd parity over bits 15 to 8.
30:18	—	0	RO	*Reserved*
17	double_bit_er	0	RW	Force double-bit error.
16	single_bit_er	0	RW	Force single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkts	0	RW	Corrupt all packets.
9	alt_pkts	0	RW	Corrupt alternate packets.
8	one_pkt_only	0	RW	Corrupt one packet only.
7:3	—	0	RO	*Reserved*
2	last_pkt_line	0	RW	Force error on last line of packet.
1	second_pkt_line	0	RW	Force error on second line of packet.
0	fst_pkt_line	0	RW	Force error on first line of packet.

TABLE 26-38 TXC RO ECC STATE – TXC_ROECC_ST (FZC_TXC + 20048_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	clr_st	0	WO	Write 1 to clear TX_RO_DATA*n* registers, where *n* = 0 to 4, and TXC_ROECC_ST.
30:18	—	0	RO	*Reserved*
17	correct_err	0	RW	Detected correctable error.
16	uncorrect_err	0	RW	Detected uncorrectable error.
15:10	—	0	RO	*Reserved*
9:0	ecc_addr	0	RW	Address of UE or CE error.

TABLE 26-39 TXC RO Data 0 – TXC_RO_DATA0 (FZC_TXC + 20050_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_ecc_data0	0	RW	ro_ecc_data{31:0}.

TABLE 26-40 TXC RO Data 1 – TXC_RO_DATA1 (FZC_TXC + 20058_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_ecc_data1	0	RW	ro_ecc_data{63:32}.

TABLE 26-41 TXC RO Data 2 – TXC_RO_DATA2 (FZC_TXC + 20060_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_ecc_data2	0	RW	ro_ecc_data{95:64}.

TABLE 26-42 TXC RO Data 3 – TXC_RO_DATA3 (FZC_TXC + 20068_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_ecc_data3	0	RW	ro_ecc_data{127:96}.

TABLE 26-43 TXC RO Data 4 – TXC_RO_DATA4 (FZC_TXC + 20070_{16})

Bit	Field	Initial Value	R/W	Description
63:24	—	0	RO	*Reserved*
23:0	ro_ecc_data4	0	RW	ro_ecc_data{151:128}.

TABLE 26-44 TXC SF ECC Control – TXC_SFECC_CTL (FZC_TXC + 20078_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	disable_ue	0	RW	Odd parity over bits 15 to 8.
30:18	—	0	RO	*Reserved*
17	double_bit_err	0	RW	Force double-bit error.
16	single_bit_err	0	RW	Force single-bit error.
15:11	—	0	RO	*Reserved*
10	all_pkts	0	RW	Corrupt all packets.
9	alt_pkts	0	RW	Corrupt alternate packets.
8	one_pkt_only	0	RW	Corrupt one packet only.
7:3	—	0	RO	*Reserved*
2	lst_pkt_line	0	RW	Force error on last line of packet.
1	second_pkt_line	0	RW	Force error on second line of packet.
0	first_pkt_line	0	RW	Force error on first line of packet.

TABLE 26-45 TXC SF ECC STATE – TXC_SFECC_ST (FZC_TXC + 20080_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	clr_st	0	WO	Write 1 to clear tx_sf_data*n* registers, where *n* = 0 to 4, and TXC_SFECC_ST.
30:18	—	0	RO	*Reserved*
17	correct_err	0	RW	Detected correctable error.
16	uncorrect_err	0	RW	Detected uncorrectable error.
15:10	—	0	RO	*Reserved*.
9:0	ecc_addr	0	RW	Address of UE or CE error.

TABLE 26-46 TXC SF Data 0 – TXC_SF_DATA0 (FZC_TXC + 0088_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	sf_ecc_data0	0	RW	sf_ecc_data{31:0}.

TABLE 26-47 TXC SF Data 1 – TXC_SF_DATA1 (FZC_TXC + 20090_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	sf_ecc_data1	0	RW	sf_ecc_data{63:32}.

TABLE 26-48 TXC SF Data 2 – TXC_SF_DATA2 (FZC_TXC + 20098_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	sf_ecc_data2	0	RW	sf_ecc_data{95:64}.

TABLE 26-49 TXC SF Data 3 – TXC_SF_DATA3 (FZC_TXC + $200A0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	sf_ecc_data3	0	RW	sf_ecc_data{127:96}.

TABLE 26-50 TXC SF Data 4 – TXC_SF_DATA4 (FZC_TXC + $200A8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:24	—	0	RO	*Reserved*
23:0	sf_ecc_data4	0	RW	sf_ecc_data{151:128}.

TABLE 26-51 TXC RE-Order TIDs – TXC_RO_TIDS (FZC_TXC + $200B0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—		RO	*Reserved*
31:0	tids_in_use	0	RW	Meta transaction IDs in use, based on bit positions.

TABLE 26-52 TXC RO STATE0 – TXC_RO_STATE0 (FZC_TXC + $200B8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	duplicate_tid	0	RW	Detected duplicate TID in use.

TABLE 26-53 TXC RO STATE1 – TXC_RO_STATE1 (FZC_TXC + $200C0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	unused_tid	0	RW	Uninitialized TID detected.

TABLE 26-54 TXC RO STATE2 – TXC_RO_STATE2 (FZC_TXC + $200C8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	transaction_timeout	0	RW	Timed-out TIDS.

TABLE 26-55 TXC RO STATE3 – TXC_RO_STATE3 (FZC_TXC + $200D0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	enable_spacefilled_ watermark	0	RW	Enable FIFO filled watermark.
30:21	ro_spacefilled_ watermark	0	RO	Highest amount filled in reorder FIFO. Enabled when watermark bit set.
20:11	ro_fifo_ spaceavailable	$3FF_{16}$	RO	Space available in RO FIFO.
10:9	rsv	0	RO	*Reserved*
8	enable_ro_waterma rk	0	RW	Enable reorder state watermark.
7:4	highest_reorder_ used	0	RO	Highest number of reorder states used. Enabled when watermark bit is set.
3:0	num_reorder_used	0	RO	Number of outstanding reorder states.

TABLE 26-56 TXC RO_ST_Control – TXC_RO_CTL (FZC_TXC + $200D8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	clr_fail_state	0	RW	Clear all failure state.
30:28	—	0	RO	*Reserved*
27:24	ro_addr	0	RO	Address of reorder context failure. Note 16 for 10 Gbytes and 8 for 1 Gbyte. Cleared when clr_fail_state is set.
23	—	0	RO	*Reserved*
22	address_failed	0	RO	State valid for Address failure. Cleared when clr_fail_state is set.
21	dma_failed	0	RO	State valid for DMA failure. Cleared when clr_fail_state is set.
20	length_failed	0	RO	State valid for Length failure. Cleared when clr_fail_state is set.
19	—	0	RO	*Reserved*
18	capture_address_fail	0	RW	Capture address failure.
17	capture_dma_fail	0	RW	Capture DMA failure.
16	capture_length_fail	0	RW	Capture length failure.
15:8	—	0	RO	*Reserved*
7	ro_state_rd_done	0	RO	Reorder state read command done; will be cleared on new command. Cleared when clr_fail_state is set.
6	ro_state_wr_done	0	RO	Reorder state write command done; will be cleared on new command. Cleared when clr_fail_state is set.
5	ro_state_rd	0	RW	Execute read on reorder state; will be cleared when done.
4	ro_state_wr	0	RW	Execute write on reorder state; will be cleared when done.
3:0	ro_state_addr	0	RW	Address of reorder context. Note 16 for 10 Gbytes and 8 for 1 Gbyte.

TABLE 26-57 TXC RO_ST_Data0 – TXC_RO_ST_DATA0 (FZC_TXC + $200E0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_st_dat0	0	RW	TBD.

TABLE 26-58 TXC RO_ST_Data1 – TXC_RO_ST_DATA1 (FZC_TXC + $200E8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_st_dat1	0	RW	TBD.

TABLE 26-59 TXC RO_ST_Data2 – TXC_RO_ST_DATA2 (FZC_TXC + $200F0_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_st_dat2	0	RW	TBD.

TABLE 26-60 TXC RO_ST_Data3 – TXC_RO_ST_DATA3 (FZC_TXC + $200F8_{16}$)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	ro_st_dat3	0	RW	TBD.

The following defines the miscellaneous errors in the TXC block. Errors that are not masked are **or**ed and the resulting signal sent to the PIO block for error reporting. The Debug register is used to force error into the interrupt path for design verification only. PacketAssyDead indicates that the state machine associated with the packet assembler dies. This is a fatal error for the associated port. It may be a badly constructed transmit header or an uncorrectable soft error on the FIFOs. The latter case is indicated by the uncorrectable error bit. Reorder error is also a fatal error; it indicates unexpected error encountered. Normally, this error will not occur in deployment. When these two errors occur, software is recommended to core-dump all debug registers associated with the port for further analysis. To recover, the transmit port needs to be reset. In addition, all the DMAs assigned to the port need to be reset.

The following registers store more internal state of the realigner in TXC. The register can be cleared by writing the TXC_PORT_CTL register.

The functionality of TXC_PORT_PACKET_REQ register is no longer supported. Software should not access this register.

TABLE 26-61 Port Packets Request – TXC_PORT_PACKET_REQ (FZC_TXC + 20100_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:28	gather_req	0	RW	Current gather list number in process by realigner.
27:16	packet_req	0	RW	Current packet request count in process by realigner.
15:0	pkterr_abort	0	RW	Number of packets dropped by packetAssy, due to ECC errors on control header and unexpected or premature SOP-EOP tags on reorder FIFO.

For UltraSPARC T2, there are two ports; the transmit DMA datapath logic will issue interrupts separately.

TABLE 26-62 TXC Interrupt Stat_Debug - TXC_INT_STAT_DBG (FZC_TXC + 20420_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	port1_int_status	0	RW	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.
7:6	—	0	RO	*Reserved*
5:0	port0_int_status	0	RW	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.

TABLE 26-63 TXC Interrupt Stat - TXC_INT_STAT (FZC_TXC + 20428_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	port1_int_status	0	RW1C	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.
7:6	—	0	RO	*Reserved*
5:0	port0_int_status	0	RW1C	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.

TABLE 26-64 TXC Interrupt Mask - TXC_INT_MASK (FZC_TXC + 20430_{16})

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:8	port1_int_mask	$3F_{16}$	RW	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.
7:6	—	0	RO	*Reserved*
5:0	port0_int_mask	$3F_{16}$	RW	PacketAssyDead, Reorder error, RO_Uncorrect error, RO_Correct error, SF_Uncorrect error, SF_Correct error.

26.8 Errata

In this section, a list of know bugs are documented. Some of which will be fixed before 2.0 tapeout.

26.8.1 Tx Descriptor Bug: ID 112918

If the programmed Tx descriptor data in the memeory is zero and when the descriptors are fetched by NIU Tx block, it causes hangs in the related Tx port. This bug will be fixed in 2.0 TO.

If a hang is detected in the tx path due to the sop bit not set in the descriptor, the following is the sequence to recover is to shut down all the dmas bound to the port by following the steps below :

0. Wait untill the timer expires (This value need to be provided to the software team). Then check for the head ptr value in tdmc to make sure that it has not moved, can also check the tx registers too, like the pkt_stuffed, pkts_transmitted.

1. Stop all the dmas bound to the port. None of the DMAs will set the stop_done bit in the respective status reg since txc is hung.

To stop the DMA :

```
task DMAChannel :: stop_dma()
{
bit [39:0] address;
bit [63:0] r_data;
bit [63:0] w_data;
address = TX_CS + id*40'h200;
gen_pio_drv.pio_rd(getPIOAddress(address,dis_pio_virt),r_data);
w_data = r_data | 64'h0000_0000_1000_0000;
```

```
  // stop the dma
  SetTxCs(w_data);
  // wait for the dma to go to stop_state
  RdTxCs(3);
}
```

2. We need to create an error on all the dmas bound to the port by writing to the TDMC interrupt debug register(TDMC_INTR_DBG) address is (DMC + 0x400060). This will force each and every dma channel to assert its dma_error signal into the txc engine.

```
task DMAChannel :: SetTDMCIntrDbg(bit [63:0] data)
{
  bit [39:0] address;
  address = TDMC_INTR_DBG + id*40'h200;
  gen_pio_drv.pio_wr(address,data);
}
```

Now the DMA engine causing the hang will return with the stop_done bit first and subsequently all the other dmas will follow. At this point the all the DMAs have been stopped. The software now knows exactly which dma was causing the hang(the first DMA to come back with the stop_done bit).

3. At this point we need to reset all the DMAs bound to the port, wait for the reset done to happen, then re-init the DMAs.

To reset write to the following register :

```
address = TX_CS + id*40'h200;
w_data = 64'h8000_0000;
gen_pio_drv.pio_wr(getPIOAddress(address,dis_pio_virt),w_data);
```

Wait for the reset_done bit to be set and then reinitial the dma.

After this we should be able to setup pkts and kick pkts on all the active DMAs bound to the Port.

26.8.2 Tx Mailbox Address : ID 113524

Whenever software programs a mailbox address which is not within the programmed partition tables, per PRM, TDMC block is supposed to inform software through an interrupt and shut down the DMA. Tx mailbox translation error is not detected by tdmc and doesn't assert dma error. The work around would be for software to always program the mailbox address within the partition tables.

In this case, mailbox would not be updated at all even though the MB enable bit is set in the TX_CS register. Also the head pointer would move past the descriptor in which the mark bit was set. This doesnt affect the data transfer in any of the DMAs/ Ports.

To recover from this condition, re-program the mailbox address to the correct values while this DMA is active. However the recommended solution would be to avoid this by correct programming.

26.8.3 Tx Address Mode: ID 113526

The 32bit mode signal which intention is to tell hw to ignore the msb 32bits of a 64bits address is not used by PIM.PIM treats full 64bits address to be valid at all times regardless whether it is 32b or 64b mode access.

The decision is to enforce SW to not to put 'don't care/junk'in non-used address bits in 32 bit mode and to assure that the unused msb 32bits of the translated address is '0'.

CHAPTER 27

Network Interface Unit: Ethernet Media Adaptation Controller (MAC)

This document describes the programming of the MAC on UltraSPARC T2.

There are two ports in UltraSPARC T2 NIU. The xMAC ports support 10G/1G/100M/10M bit per second data rate.

Each xMAC port is composed of xMAC, XPCS, and PCS. There is only one MIF and one ESR control per chip.

Software needs to independently program each port's xMAC, XPCS, PCS and bMAC.

UltraSPARC T2 supports only XAUI interface.

In the UltraSPARC T2 environment, the two MACs must operate at the same speed.

This chapter is divided into the following subsections:

1. xMAC (10M/100M/1G/10G bps quad speed mac) Programmable Resources
2. XPCS (10 G) Programmable Resources
3. PCS (1 G) Programmable Resources
4. MIF (MDIO STA) Programmable Resources
5. Ethernet SerDes (ESR) Programmable Resources

In this chapter, register attributes are defined as follows:

- WR-AC: Write, Read Auto Clear
- RW-AC: Read, Write Auto Clear
- R-W0C: Read, Write 0 to Clear
- R-W1C: Read, Write 1 to Clear

- RO: Read Only
- RW: Read, Write

27.1 MAC Configuration

The association between internal ports and network interfaces can be described withTABLE 27-1, UltraSPARC T2 MAC Port Configurations, on page 783.

The mode setting column of the table contains the register bit value that is necessary to set up MAC clock source and data path correctly.

The register bit suffix number indicates the port number.

Port 0 and 1 belong to xMAC. Users should look up the corresponding xMAC PRM section for detailed description.

Notes | xgmii_mode0,1, gmii_mode0,1, mii_mode0,1, pcs_bypass0,1 descriptions can be found in TABLE 27-6 on page 794.

atca_ge descriptions can be found in TABLE 27-171 on page 861.

27.1.1 UltraSPARC T2 MAC Normal Configuration

TABLE 27-1 UltraSPARC T2 MAC Port Configurations

Configuration	Port0	Port1	Mode Setting
2x10G	10G (XAUI)	10G (XAUI)	xgmii_mode0 = 1 gmii_mode0 = 0 mii_mode0 = 0 pcs_bypass0 = 0 (XAUI) ---------- xgmii_mode1 = 1 gmii_mode1 = 0 mii_mode1 = 0 pcs_bypass1 = 0 (XAUI)
1x10G + 1x1G (Not tested)	10G (XAUI)	1G (ch0) (optical)	xgmii_mode0 = 1 gmii_mode0 = 0 mii_mode0 = 0 pcs_bypass0 = 0 (optical) ---------- xgmii_mode1 = 0 gmii_mode1 = 1 mii_mode1 = 0 pcs_bypass1 = 0 (optical)
2x1G	1G (optical)	1G (ch0) (optical)	xgmii_mode0 = 0 gmii_mode0 = 1 mii_mode0 = 0 pcs_bypass0 = 0 (optical) ---------- xgmii_mode1 = 0 gmii_mode1 = 1 mii_mode1 = 0 pcs_bypass1 = 0 (optical)

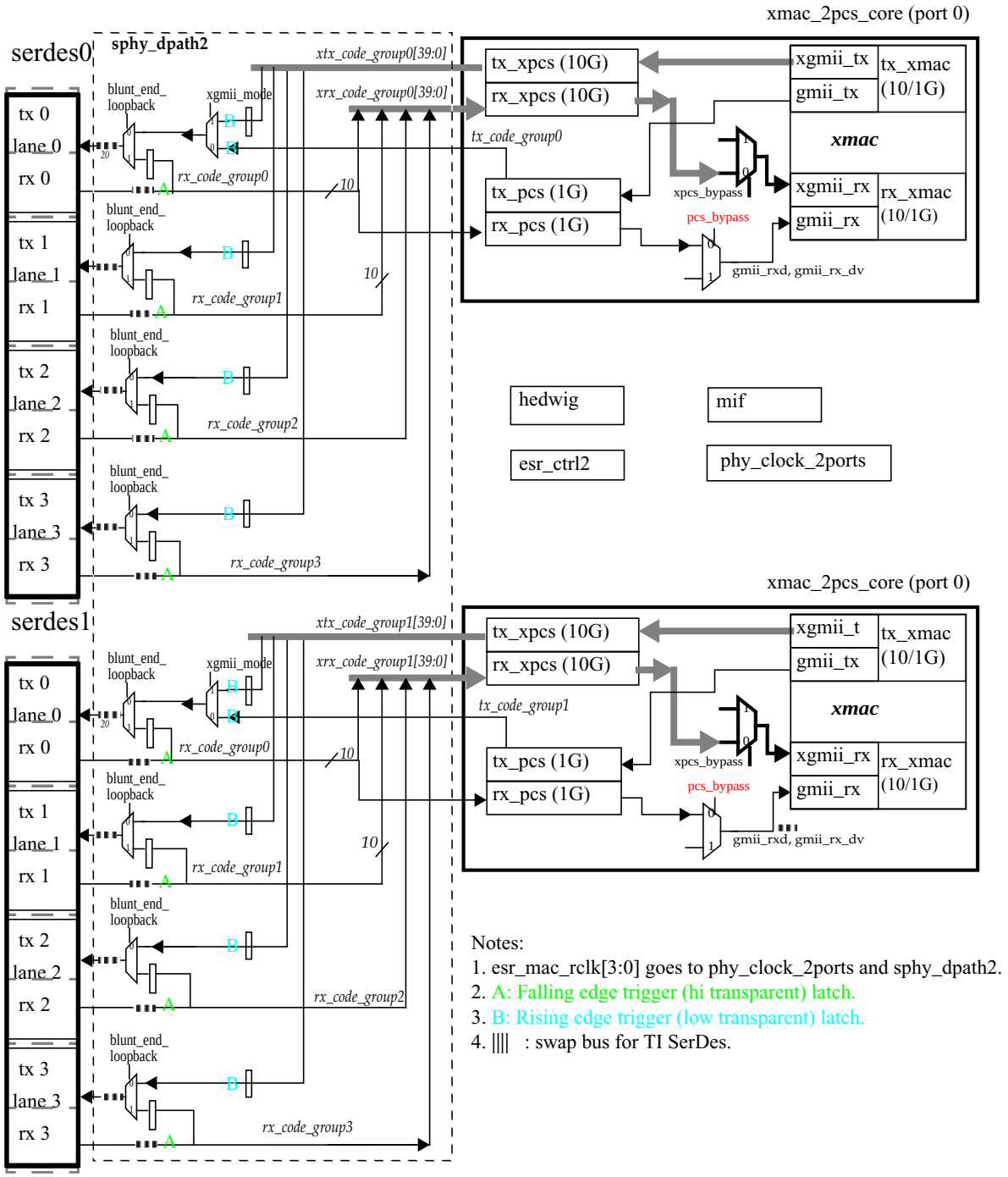
27.1.2 UltraSPARC T2 MAC Loopback Mode

The corresponding UltraSPARC T2 MAC port configuration table should be set up to do the loopback static timing analysis.

TABLE 27-2 UltraSPARC T2 MAC Loopback Mode

Configuration	Port0	Port1	Mode Setting
2x10G XMAC loopback	10G (XAUI)	10G (XAUI)	loopback0 = 1 xpcs_loopback0 = 0 ---------- loopback1 = 1 xpcs_loopback1 = 0
2x10G XPCS loopback	10G (XAUI)	10G (XAUI)	loopback0 = 0 xpcs_loopback0 = 1 ---------- loopback1 = 0 xpcs_loopback1 = 1
10G Blunt loopback	10G (XAUI)	10G (XAUI)	blunt_end_loopback = 1 ========== = loopback0 = 0 xpcs_loopback0 = 0 ---------- loopback1 = 0 xpcs_loopback1 = 0
1G Blunt loopback	1G (optical)	10 (XAUI)	blunt_end_loopback = 1 ========== = loopback0 = 0 xpcs_loopback0 = 0 ---------- loopback1 = 0 xpcs_loopback1 = 0

FIGURE 27-1 UltraSPARC T2 MAC Top-Level Architecture



Notes:
1. esr_mac_rclk[3:0] goes to phy_clock_2ports and sphy_dpath2.
2. A: Falling edge trigger (hi transparent) latch.
3. B: Rising edge trigger (low transparent) latch.
4. |||| : swap bus for TI SerDes.

27.2 MAC Init Sequence

Key:

Symbol	Meaning
o	**Co**nfiguration step
@	Decision step
*	Enable / disable block
-	Reset block
//	Comments

Combined 10G/1G initialization sequence:

o Identify Card/Board type.
o Initialize MIF config register (example, atca_ge).
o Initialize Ethernet Link:
// For consistency between various ASIC vendor's SerDes implementation,
// the BCM8704 is initialized twice.
// The first time is before internal SerDes initialization and the second time after it.
// The following regs MUST be programmed for **BCM8704 user**.
o Device Addr = 3;Register Addr = 0xC800; Data = 0x7FBF;// Control
o Device Addr = 3;Register Addr = 0xC803; Data = 0x0164;
// set XFP_CLKEN bit to enable 161.13MHz clock to XFP.
o Device Addr = 3; Register Addr = 0xC808; bit[6:5]=2'b11;
// Use GPIO[1] for activity LED.
o Device Addr = 3; Register Addr = 0x9000; Data = 0x39;
// RX_ALARM Control
o Device Addr = 3; Register Addr = 0x9001; Data = 0x59;
// TX_ALARM Control
o Device Addr = 3; Register Addr = 0x9002; Data = 0x5;
// LASI_ALARM Control
o Initialize internal SerDes:
@ if (TI SerDes)
o Initialize TI-Hedwig core.
o Program **ALIGN**{1:0} per lane in CFGRX register to 10_2.
o Program Hedwig according to Table 27-186 on page 876.
if (1G) // atca_ge
o Program rate bits of CFGTX,CFGRX = 01 (Half Rate).
o Program mpy bits of CFGPLL = 0100 (8x).
else // 10G
o Use POR default rate of CFGTX,CFGRX = 00_2 (Full rate).
o Use POR default mpy bits of CFGPLL = 0101_2 (10x).
@else
o Initialize LSI-Gigablaze core:
o Program Gigablaze.
//Program ENET_SERDES_CONTROL with 0249 $B6DF_{16}$.

//(P0: FZC_MAC + 14018_{16}, P1: FZC_MAC + 14030_{16})
// Program ENET_SERDES0,1_TEST_CFG
// (P0: FZC_MAC + 14020_{16}, P1: FZC_MAC + 14038_{16})
//if (ewrap)
// value = 055_{16}
//else (pad_loopback)
// value = $0AA_{16}$
//else value = 000_{16} // functional mode

o Program RX_TX_CONTROL_<A-D> registers with
$00A0\ 3001_{16}$.
o Program GLUE_CONTROL0<A-D> registers with
$0300\ 2FFF_{16}$.

@if (1G SerDes) // *atca_GE*
NOTE: Other than POR time, whenever changing speed form 10G to 1G, or vice versa, software must re-program PLL parameters. Software needs to put Gigablaze in "reset" state by writing 1 to serdes_reset0,1 (described in the ESR CTL section, page 866) before modifying PLL parameters.
After changing the PLL value, software should remove Gigablaze from reset state by writing 0 to serdes_reset0,1.
o Program ENET_SERDES0,1_PLL_CFG to 1G speed.
o fbdivt{0:2} = 001_2
o half_ratet_a-d = 1111_2.
o Program atca_ge bit (MIF).
else if (10G SerDes)
o No need to program ENET_SERDES0,1_PLL_CFG registers. Just use the default value which is as follows:
o fbdiv{0:2} = 100_2
o half_ratet_a-d = 0000_2
else (RGMII mode) // Shut down SerDes.
o RX_POWER_CONTROL = {32{1_2}}
o TX_POWER_CONTROL = {32{1_2}}
o MISC_POWER_CONTROL = {32{1_2}}

if (Atlas 2XGF card) { // BCM8704 init routine is done twice.
if (**BCM8704 Port 1**) {
// port 1 XAUI lane 0,1,2 polarity have been swapped
// for SI reason. Program the following three registers.
o Device Addr = 4; Register Addr = $80C6_{16}$; Data = 20_{16}; // RX2
o Device Addr = 4; Register Addr = $80D6_{16}$; Data = 20_{16}; // RX1
o Device Addr = 4; Register Addr = $80E6_{16}$; Data = 20_{16}; // RX0
else (Port 0); // Now polarity swapping on port0.
}
}

o The following registers *must* be programmed for **BCM8704 user**.
Device Addr = 3; Register Addr = $C800_{16}$; Data = $7FBF_{16}$; // Control
Device Addr = 3; Register Addr = $C803_{16}$; Data = 0164_{16};
// set xfp_clken bit to enable 161.13MHz clock to XFP.
Device Addr = 3; Register Addr = $C808_{16}$; bit{6:5} = 11_2;

if (Atlas QGC) {

o Check BCM5464 timing and mode setup against golden values for all four ports.
@case (speed)
1G: No need to reprogram BCM5464 unless something is wrong;
100M: Program BCM5464 speed to 100M.
10M: Program BCM5464 speed to 10M.
}

// Wait for (pll_lock = 1)
@ if (TI SerDes)
poll Hedwig CSR pll_lock bit to be 1 (per SerDes).
(worst case lock time = 2.5 µs)
@ else poll Gigablaze CSR GBT_RXINITDONEZ_<A-D> and GBT_TXINITDONE_<A-D> (per lane)
(min. 60us, max. 480 µs)

o Initialize MAC's XIF sub-block:
@ if xMAC port:
o Set up XMAC XIF Configuration.
o Set up Port Mode (GMII/XGMII).
@ else if BMAC port:
o Set up BMAC XIF Configuration.

o Initialize PCS/XPCS block:
@ if 10G fiber:
- Reset XPCS.
- Initialize XPCS registers.
NOTE: Reset XPCS need not to be performed here. It can be done after MAC initialization since XPCS clock source is from the SerDes, not from MAC. It is done here to have a consistent initialization ordering (initialize PCS/XPCS before MAC) between 10G and 1G initialization sequence.
@ else if 1G Copper:
o Program phy_mode = 1 (PCS bypass) in PCS_DPATH_MODE register.
NOTE: phy_mode must be set here before MAC can be initialized.
Since the phy_mode bit affects clock source selection,
it has to be done here before MAC applies software reset to clean up clock line glitches.
- Reset PCS.
@ else if 1G fiber:
o Program phy_mode = 0 (No PCS bypass) in PCS_DPATH_MODE register.
o Program pcs_enable = 1 in PCS_CONF register.
NOTE: phy_mode must be set here before MAC can be initialized.
Since the phy_mode bit affects clock source selection,
it has to be done here before MAC applies software reset to clean up clock line glitches.
- Reset PCS.
// The following software resets are to clean up the clock glitches introduced by
// changing MAC clock sources in the previous steps.
- Reset TxMAC:
@ if xMAC port:

- Reset xTxMAC.
@ else if BMAC port:
- Reset TxMAC.
o Initialize TxMAC:
@ if xMAC port:
o Initialize xTxMAC.
@ else if BMAC port:
o Initialize TxMAC.
- Reset Rx MAC:
@ if xMAC port:
- Reset xRxMAC.
@ else if BMAC port:
- Reset RxMAC.
o Initialize RxMAC:
@ if xMAC port:
o Initialize xRxMAC.
@ else if BMAC port:
o Initialize RxMAC.
NOTE: Must program PA register.
* Enable Tx MAC:
@ if xMAC port:
* Enable xTxMAC.
@ else if BMAC port:
* Enable TxMAC.
* Enable Rx MAC:
@ if xMAC port:
* Enable xRxMAC.
@ else if BMAC port:
* Enable RxMAC.
if (1G ATCA mode)
o re-initialize serdes again.For detail, please refer to section 14.4, 1G ATCA Mode Serdes Init Sequence Routine, on page 259.

27.2.1 Notes for BCM 8704

2XGF and GQC port numbers are opposite in ordering.

TABLE 27-3 2XGF Card Facts

Port Number	Physical Location on the QGC Card	Intellectual Property	MDIO Port Address
1	Top	xMAC	5 bit 09_{16}
0	Bottom (closest to PCIe link)	xMAC	5 bit 08_{16}

27.3 MAC Warm Reset Sequence

To apply MAC warm reset, software should perform the MAC stop sequence first, followed by the software reset sequence. Per-port reset is a typical application of MAC warm reset sequence.

################################
 # MAC Stop Sequence
################################

=====> **Transmit xMAC Stop Sequence**
Program XMAC_CONFIG register (FZC_MAC + 00060_{16}) bit 0 (tx_enable) to 0.
Software should wait for one max_pkt_size time for xMAC to stop gracefully.
Use xtlm_state = 0 in xMAC State Machines register (FZC_MAC + $001A8_{16}$) as stop_done indicator.

=====> **Receive xMAC Stop Sequence**
Program XMAC_CONFIG register (FZC_MAC + 00060_{16}) bit 8 (rx_enable) to 0.
Software should wait for one max_pkt_size time for xMAC to stop gracefully.
Use xrlm_state = 0 in xMAC State Machines register (FZC_MAC + $001A8_{16}$) as stop_done indicator.

-----> **Transmit bMAC Stop Sequence**
Program TXMAC_CONFIG register (FZC_MAC + $0C060_{16}$) bit 0 (tx_enable) to 0.
Software should wait for one max_pkt_size time for bMAC to stop gracefully.
Use tlm_state = 0 in bMAC State Machines register (FZC_MAC + $0C3A0_{16}$) as stop_done indicator.

-----> **Receive bMAC Stop Sequence**
Program RXMAC_CONFIG register (FZC_MAC + $0C068_{16}$) bit 0 (rx_enable) to 0.
Software should wait for one max_pkt_size time for bMAC to stop gracefully.
Use rlm_state = 0 in bMAC State Machines register (FZC_MAC + $0C3A0_{16}$) as stop_done indicator.

################################
 # MAC Software Reset Sequence
################################

=====> **Transmit xMAC Software Reset Sequence**
Software should write 1 to bit 0 (txmacsoftrst) and bit 1 (txmacregrst) in register XTXMAC_SW_RST (FZC_MAC + 00000_{16}).
Use txmacsoftrst = 0 and txmacregrst = 0 as reset_done indicator.

=====> **Receive xMAC Software Reset Sequence**
Software should write 1 to bit 0 (rxmacsoftrst) and bit 1 (rxmacregrst) in register XRXMAC_SW_RST (FZC_MAC + 00008_{16}).
Use rxmacsoftrst = 0 and rxmacregrst = 0 as reset_done indicator.

-----> **Transmit bMAC Software Reset Sequence**
Software should write 1 to bit 0 (txmac_sw_rst) in register BTXMAC_SW_RST (FZC_MAC + $0C000_{16}$).
Use txmac_sw_rst = 0 as reset_done indicator.

-----> **Receive bMAC Software Reset Sequence**
Software should write 1 to bit 0 (rxmac_sw_rst) in register BRXMAC_SW_RST (FZC_MAC + $0C008_{16}$).
Use rxmac_sw_rst = 0 as reset_done indicator.

27.3.1 Changing Operation Mode Sequence

1. Stop TXDMA first.
2. Execute MAC stop sequence.
3. Execute MAC init sequence.

27.4 1G ATCA Mode SerDes Init Sequence Routine

For UltraSPARC T2-NIU:

Program Hedwig registers with the value defined in the TABLE 27-186 on page 876 1G column.

Programming Notes

Upon changing any XIF registers bits or PCS phy_mode bit or ESR_CTL atca_ge bit, software should issue the MAC local software reset to clean up the potential clock line glitch induces problem.

MAC support full duplex mode only.
MAC does not support padding of transmit packets that are fewer than 60 bytes with hardware assist CRC appending and 64 bytes without hardware assist CRC appending. Failure to program for this will result in an unpredictable packet length sent to the network.

27.5 MAC Interrupts

There are two ports in UltraSPARC T2. Each port can independently generate interrupt requests.

TABLE 27-4 MAC Interrupts Info Table

Port Number	Port Components	Interrupt Request Generation Equation	Logic Device Flag Number
Port 0	xMAC0, XPCS0, PCS0	port_int0 = xmac_int0 | xpcs_int0 | pcs_int0	64
Port 1	xMAC1, XPCS1, PCS1	port_int1 = xmac_int1 | xpcs_int1 | pcs_int1	65
MIF	MIF	mif_int	63

27.5.1 XMAC

XMAC uses the following status and mask registers to generate interrupt requests.

- TABLE 27-7, Tx_xMAC Status – xtxmac_status (fzc_mac + 0002016) (count 2 step 06000), on page 794
- TABLE 27-8, Rx_xMAC Status – xrxmac_status (fzc_mac + 0002816) (count 2 step 06000), on page 795
- TABLE 27-9, xMAC Flow Control Status – xmac_fc_stat (fzc_mac + 0003016) (count 2 step 06000), on page 797
- TABLE 27-10, Tx_xMAC Mask – xtxmac_stat_msk (fzc_mac + 0004016) (count 2 step 06000), on page 797
- TABLE 27-11, Rx_xMAC Mask – xrxmac_stat_msk (fzc_mac + 0004816) (count 2 step 06000), on page 797
- TABLE 27-11, Rx_xMAC Mask – xrxmac_stat_msk (fzc_mac + 0004816) (count 2 step 06000), on page 797

27.5.2 XPCS

XPCS uses the following status and mask registers to generate interrupt requests.

- TABLE 27-138, XPCS Status 1 – base10g_status1 (fzc_mac + 0200816) (count 2 step 06000), on page 843
- TABLE 27-149, XPCS Mask 1 – base10g_mask1 (fzc_mac + 0206016) (count 2 step 06000), on page 849

27.5.3 PCS

PCS uses the following status and mask registers to generate interrupt requests.

- TABLE 27-163, PCS Interrupt Status – pcs_interrupt (fzc_mac + 0403016) (count 2 step 06000), on page 857
- Bit 6 (interrupt mask) of TABLE 27-161, PCS Configuration – pcs_conf (fzc_mac + 0402016) (count 2 step 06000), on page 855

27.5.4 MIF

MIF uses the following status and mask registers to generate interrupt requests.

- TABLE 27-172, MIF Poll Status – mif_poll_status (fzc_mac + 1602816), on page 862
- TABLE 27-173, MIF Poll Mask – mif_poll_mask (fzc_mac + 1603016), on page 863
- TABLE 27-175, MIF Status – mif_status (fzc_mac + 1604016), on page 864
- TABLE 27-176, MIF Mask – mif_mask (fzc_mac + 1604816), on page 864

27.6 xMAC (10G/1G/100M/10M Quad Speed MAC) Programmable Resources

The register XMAC_CONFIG (FZC_MAC + 00060_{16}) bit 26 lfs (Link Fault Signalling) is for 10G only. When XMAC is configured to operate at any other speed, software should program this bit to 1 (disable state).

27.6.1 xMAC Command Registers

These registers are used by the software to instruct the hardware that a certain hardware function is to be executed upon the detection of the command bit. The command bits are set to 1 by a PIO write and are self-cleared after the command execution has completed. A command bit can be polled at any time using a PIO read to verify the command execution.

TABLE 27-5 TxMAC Software Reset Command – XTXMAC_SW_RST (FZC_MAC + 00000_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1	tx_mac_reg_rs	0	RW-AC	Performs a software reset to the TxMAC-related control/ status registers (excluding logic and state machines in the TxMAC). Write 1 to set this bit. It is self-cleared after the command has been executed.
0	tx_mac_soft_rst	0	RW-AC	Performs a software reset to the logic and state machines in the TxMAC (excluding TxMAC-related control/status registers). Write 1 to set this bit This bit is self-cleared after the command has been executed

TABLE 27-6 RxMAC Software Reset Command – XRXMAC_SW_RST (FZC_MAC + 00008_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	rx_mac_reg_rst	0	RW-AC	Performs a software reset to the RxMAC-related control/ status registers (excluding logic and state machines in the RxMAC). Write 1 to set this bit. It is self-cleared after the command has been executed.
0	rx_mac_soft_rst	0	RW-AC	Performs a software reset to the logic and state machines in the RxMAC (excluding RxMAC-related control/status registers). Write 1 to set this bit. This bit is self-cleared after the command has been executed.

27.6.2 xMAC Status and Mask Registers

If a mask bit is cleared to 0, the corresponding status event is enabled to generate an interrupt. Mask register bits default to 1 upon reset.

TABLE 27-7 Tx_xMAC Status – XTXMAC_STATUS (FZC_MAC + 00020_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11	txmac_frame_count_ exp	0		The 25-bit loadable counter (increments when a frame has been transferred successfully) has expired.
10	txmac_byte_count_ exp	0		This 31-bit loadable counter (increments when a byte has been transferred successfully) has expired.
9:5	—	0		*Reserved*

TABLE 27-7 Tx_xMAC Status – XTXMAC_STATUS (FZC_MAC + 00020_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
4	txfifo_xfr_err	0		TxMAC transfer error. It indicated that xMAC enter a unexpected state. This is not fatal. xMAC recovers itself.
3	txmac_overflow	0		The TxMAC FIFO has terminated a frame transmission due to TXC-TXMAC protocol error that causes txfifo to overflow. This is not fatal for xMAC since it can recover itself.
2	txc_max_pkt_size_err	0		A frame that exceeds the maximum allowed length was passed to the TxMAC by the TXC.
1	txmac_underflow	0		The TxMAC FIFO has terminated a frame transmission due to "data starvation" in the transmit data path. For example, if OPP cannot support 10Gbps data rate in 10G mode, MAC will detect this condition and set the txmac_underflow bit. This is not fatal for xMAC since it can recover itself.
0	frame_transmitted	0		Successful transmission of a frame on the network.

TABLE 27-8 Rx_xMAC Status – XRXMAC_STATUS (FZC_MAC + 00028_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19	local_fault_status (link_status_changed)	0	WR-AC	This interrupt status bit is set when link changes status from either up to down or down to up. Software can check the local_fault_oc_sync bit in TABLE 27-36, xMAC Internal Signals 2 – xmac_intern2 (fzc_mac + 001B816) (count 2 step 06000), on page 810 for up/down status. 0 = link status is not changed. 1 = link status has changed. MAC will send remote_fault to link partner when local_fault_oc_sync = 1.
18	remote_fault_detected	0	WR-AC	The RxMAC has detected a remote fault from link partner. MAC will stop sending packets to network and create back pressure toward TXC.
17	link_fault_cnt_exp	0	WR-AC	The Link Fault counter has reached its maximum value plus 1.
16	phy_mdint	0	WR-AC	Per port interrupt input signal. External PHY Interrupt flag from PHY_MDINT0_L pin for port 0 and PHY_MDINT1_L pin for port 1. Software needs to perform MDIO read operation through MIF module to read external PHY status bits to clear this bit. 0 = external PHY has not asserted interrupt. 1= External PHY has asserted interrupt.
15	rx_frag_cnt_exp	0	WR-AC	The Receive Fragments (Runts) (< 64 bytes pkts) counter has reached its maximum value plus 1.

TABLE 27-8 Rx_xMAC Status – XRXMAC_STATUS (FZC_MAC + 00028_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
14	rx_mult_frm_cnt_exp	0	WR-AC	The Receive Multicast Frames counter has reached its maximum value plus 1.
13	rx_broadcast_cnt_exp	0	WR-AC	The Receive Broadcast Frames counter has reached its maximum value plus 1.
12	rx_hist_cnt6_exp	0	WR-AC	The Receive Histogram counter 6 has reached its maximum value plus 1.
11	rx_hist_cnt5_exp	0	WR-AC	The Receive Histogram counter 5 has reached its maximum value plus 1.
10	rx_hist_cnt4_exp	0	WR-AC	The Receive Histogram counter 4 has reached its maximum value plus 1.
9	rx_hist_cnt3_exp	0	WR-AC	The Receive Histogram counter 3 has reached its maximum value plus 1.
8	rx_hist_cnt2_exp	0	WR-AC	The Receive Histogram counter 2 has reached its maximum value plus 1.
7	rx_hist_cnt1_exp	0	WR-AC	The Receive Histogram counter 1 has reached its maximum value plus 1.
6	rx_octets_cnt_exp	0	WR-AC	The Receive Octets counter has reached its maximum value plus 1.
5	code_viol_err_cnt_exp	0	WR-AC	The Code Violation Error counter has reached its maximum value plus 1.
4	length_err_cnt_exp	0	WR-AC	The Maximum packet Length Error counter has reached its maximum value plus 1.
3	crcerr_cnt_exp	0	WR-AC	The CRC Error counter has reached its maximum value plus 1.
2	rx_macunderflow	0	WR-AC	The RxMac rxFIFO is empty while IPP is trying to access more data. This is an IPP-RXMAC protocol error. This is not fatal, for xMAC since it can recover itself.
1	rx_macoverflow	0	WR-AC	The RxMac has dropped a receive frame due to the lack of resources in the receive data path (IPP RX FIFO). This is a normal phenomenon that can happen when rx MAC is under back pressure. When back pressure is removed, xmac recovers itself.
0	framereceived	0	WR-AC	Successful reception of a frame from the network.

TABLE 27-9 xMAC Flow Control Status – XMAC_FC_STAT (FZC_MAC + 00030_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	rx_mac_rcv_pause_ time	0	RO	This information field indicates the value of the "pause_time" operand that was received in the last pause flow control frame.
15:3	—	0	RO	*Reserved*
2	tx_mac_in_not_ pause_st	0	WR-AC	The TxMAC has made a state transition from paused to not-paused.
1	tx_mac_in_pause_st	0	WR-AC	The TxMAC has made a state transition from not-paused to paused.
0	rx_mac_rcv_pause	0	WR-AC	Successful reception of a Pause Flow Control frame from the network.

TABLE 27-10 Tx_xMAC Mask – XTXMAC_STAT_MSK (FZC_MAC + 00040_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:12	—	0	RO	*Reserved*
11	tx_mac_frame_count_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
10	tx_mac_byte_count_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
9:5	—	5 bit $1F_{16}$	RW	*Reserved*
4	txfifo_xfr_err	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
3	tx_mac_overflow	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
2	opp_max_pkt_size_err	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
1	tx_mac_underflow	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
0	frame_transmitted	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.

TABLE 27-11 Rx_xMAC Mask – XRXMAC_STAT_MSK (FZC_MAC + 00048_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19	local_fault_detected	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
18	remote_fault_detected	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
17	link_fault_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
16	mdint _asserted	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.

TABLE 27-11 Rx_xMAC Mask – XRXMAC_STAT_MSK (FZC_MAC + 00048_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
15	rx_frag_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
14	rx_mult_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
13	rx_broadcast_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
12	rx_hist_cnt6_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
11	rx_hist_cnt5_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
10	rx_hist_cnt4_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
9	rx_hist_cnt3_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
8	rx_hist_cnt2_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
7	rx_hist_cnt1_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
6	rx_octets_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
5	viol_err_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
4	length_err_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
3	crc_err_cnt_exp	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
2	rx_mac_underflow	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
1	rx_mac_overflow	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
0	frame_received	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.

TABLE 27-12 xMAC Flow Control Mask – XMAC_FC_MSK (FZC_MAC + 00050_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2	tx_mac_in_not_pause_st	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
1	tx_mac_in_pause_st	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.
0	rx_mac_rcv_pause	1_{16}	RW	0 = enable interrupt; 1 = disable interrupt.

27.6.3 xMAC Configuration Registers

This register contains three sections: TxMac configuration bits, RxMac configuration bits, and Mac transceiver interface (MacXif) configuration bits.

TxMac configuration section controls the TxMac operation.

RxMac configuration section controls the RxMac operation.

XIF configuration bits determine the parameters that control the operation of the transceiver interface. Any XIF bit value changes should be followed by MAC software reset to clean up clock line glitches.

Programming Restrictions:

*To ensure proper operation of the hardware, when changing any of the **xif** bit 25 and bit 27~31, atca_ge in MIF, the TxMAC and RxMAC software reset should be performed.*

TABLE 27-13 xMAC Configuration – XMAC_CONFIG (FZC_MAC + 00060_{16}) (count 2 step 06000) *(1 of 3)*

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	sel_clk_25mhz	0	RW	Valid only when mii_gmii_mode. = 10_2, which means MAC is in mii mode of operation. 0 = Select internal generated 2.5 MHz clock for 10Mbps mode operation. 1 = Select internal generated 25 MHz clock for 100Mbps mode operation.
30	1g_pcs_bypass	0	RW	Valid only when xMAC is programmed to be in 1 Gpbs mode. 0 = Use internal PCS module. Internal SerDes is used. 1 = Internal PCS module is not used (bypassed). No internal SerDes is used. Copper Phy through RGMII interface is used.
29	10g_xpcs_bypass	0	RW	Valid only when the xMAC is in 10Gbps mode. 0 = Use internal XPCS module. 1 = Use external XPCS module. XGMII interface signals are used.
28:27	mii_gmii_mode	0	RW	00_2 = MAC is in xgmii mode of operation. This is equivalent to xgmii_mode = 1_2. There is no explicit xgmii_mode bit. 01_2 = MAC is in gmii mode of operation. This is equivalent to gmii_mode =1_2. There is no explicit gmii_mode bit. 10_2= MAC is in mii mode of operation. This is equivalent to mii_mode = 1_2. There is no explicit mii_mode bit. 1_2 = illegal operation. MAC is in undefined state.
26	lfs_disable	0	RW	0 = Link fault signalling mechanism is enabled. 1 = Link fault signalling mechanism is disabled.
25	loopback	1_{16}	RW	When this bit is set to 1 (default). the xMAC, xgmii transmit data path is internally (on-chip) looped back to the xMAC receive data path (default value). This bit must be low for normal operation.
24	tx_output_en	0	RW	This signals applies to all modes/speeds. 0 = Disables transmit MAC to send out packets. 1 = Enables transmit MAC to send out packets.

TABLE 27-13 xMAC Configuration – XMAC_CONFIG (FZC_MAC + 00060_{16}) (count 2 step 06000) *(2 of 3)*

Bit	Field	Initial Value	R/W	Description
23	sel_por_clk_src	1_{16}	RW	xMAC has two distinct loopback clock sources. The first one is the POR loopback clock source (sys_clk) which should only be used in power on reset time. This is to guarantee a reliable clock source is always available to initialize the state machines and registers. Dependents on configuration, SerDes PLL (tx_clk_312mhz) may not be available for POR loopback clock source. The second one is the normal functional clock sources for doing function loopback test in debug/bring up environment. Software should always reset this be to zero after POR. 0 = Select functional mode loopback clock source. 1 = Select system clock as POR loopback clock (default).
22	led_polarity	1_{16}	RW	0 = LED polarity is negative; 1 = LED polarity is positive.
21	force_led_on	0	RW	0 = Do not force LED to be on; 1 = Force LED to be on all the time. This is to check the health of LED.
20	Pass flow control frames	0	RW	When set to 1, enables the RxMAC to pass valid MAC control packets to the RxDMA. If cleared to 0, the RxMAC will set the abort bit for the received flow control frames, which will cause the IPP to drop them.
19	receive_pause_ enable	0	RW	When set to 1, enables the MAC to start responding to received pause flow control packets. If cleared to 0, the MAC will ignore all received pause flow control packets.
18	mac2ipp_pkt_cnt_en	0	RW	0 = The histo_cntr7 behaves as jumbo packet counter (default). 1 = The histo_cntr7 behaves as MAC to IPP packet counter.
17	strip_crc	1_{16}	RW	Strip off four CRC bytes. 0 = The four CRC bytes are not touched by RxMAC. 1 = Cause the RxMAC to strip the last four bytes of a received frame.
16	addr_filter_en	0	RW	0 = RxMac does not use Alternate Address Filtering registers to perform address matching. 1 = RxMac uses Alternate Address Filtering registers to perform address matching.
15	hash_filter_en	0	RW	0 = RxMAC does not use the Hash Table to perform address matching. 1 = RxMAC use the Hash Table to perform address matching.
14	rx_code_violation_ chk_dis	0	RW	0 = Disable receive code violation (E) checking while receiving MAC frame. 1 = Enable receive code violation (E) checking while receiving MAC frame.
13	reserved_multicast	0	RW	When set to 1, enables the reserved multicast address comparison. The last 4 bits of Mac Unique Address register (Mac Address2 to Address0) will be masked.
12	rx_crc_chk_dis	0	RW	When set to 1, disables RxMac CRC logic. When reset to 0, enables RxMac CRC logic.

TABLE 27-13 xMAC Configuration – XMAC_CONFIG (FZC_MAC + 00060_{16}) (count 2 step 06000) *(3 of 3)*

Bit	Field	Initial Value	R/W	Description
11	err_chk_dis	0	RW	Rx MAC global switch to turn on/off various packet checking mechanism. When set to 1, RxMac stops frame validity checking for CRC, framing or length errors (max, min packet size). The abort bit will not be set if an error was detected. A frame will be accepted or rejected based on address matching only. When reset to 0, Rx Mac performs frame validity checking. This bit will not disable bad_pkt_bit in the status word. IPP should forward this packet to uP when bad_pkt_bit is set. In regular operation, software should reset this bit to 0. Failure to do so can cause down stream to hang in the case of rollover pkt_length when exceeding max_pkt_size of 14 bit wide $3FEF_{16}$.
10	promiscuous_group	0	RW	When set to 1, RxMac accepts all valid multicast frames (frames that have the "group" bit in the da field set to 1).
9	promiscuous	0	RW	When set to 1, RxMac accepts all valid frames from the network, regardless of the contents of the da field of a frame.
8	rx_mac_enable	0	RW	When set to 1, the RxMac state machine is enabled to perform packet receiving. When cleared to 0, it will force the RxMac state machines to enter the idle state at packet boundary and stay there.
7	warning_msg_en	0	RO	Enables printing of warning message for simulation purpose.
6:4	*Used*	0	RO	Used for spare control bits.
3	always_no_crc	0	RW	0 = xMac will act on the CRC generation as indicated by the no_crc bit in the transmit control word from TxMac DMA (OPP). 1 = Force the TxMac not to generate 4 bytes CRC for all transmitted packets.
2	var_min_ipg_en	1_{16}	RW	0 = TxMac generates at least 12 bytes ipg at the end of each packet. 1 = TxMac performs deficit idle count (DIC) algorithm. To maintain the 10Gbps data rate, the minimum ipg is varied from 9 to 15 bytes but will average to 12 bytes.
1	stretch_mode	0	RW	For WAN application requirements a constant 9. 45 Gbps bit rate. When stretch_mode is enabled, var_min_ipg_en is also enabled automatically. 0 = For LAN application. TxMAC transmit data at 10 Gbps rate. 1 = For WAN application. TxMAC transmit data at 9. 45 Gbps rate.
0	tx_enable	0	RW	When set to 1, the TxMac starts requesting packet data from the OPP, and the Ethernet transmit protocol execution begins. When cleared to 0, it will force the TxMac state machines to enter idle state at the packet boundary and stay there.

27.6.4 xMAC Protocol Parameters Registers

TABLE 27-14 xMAC Interpacket Gap – XMAC_IPG (FZC_MAC + 00080_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:24	—	0	RO	*Reserved*
23:21	stretch_constant	1_{16}	RW	Defines the initial Inter-Packet Gap (IPG) value for WAN application when stretch_mode is enabled. When stretch_mode is clear, this value has no effect on extending IPG.
20:16	stretch_ratio	D_{16}	RW	Determines the number of bits in a frame that require one octet of interframe spacing extension when stretch_mode is enabled. This field defines the modulo value for counting a transmitting packet that extends IPG for WAN application when stretch_mode bit is enabled. When stretch_mode bit is clear, this value has no effect on extending IPG.
15:8	ipg_value1	A_{16}	RW	Defines the IPG when MAC is in either the mii or gmii mode of operation.10_{10} – 12 bytes IPG. Set ipg_value{7:0} to 3_{16} and enable the var_min_ipg in TXMAC configuration register bit 2. 11_{10} – 13 bytes IPG. Set ipg_value{7:0} to 4_{16} 12_{10} – 14 bytes IPG. Set ipg_value{7:0} to 4_{16}. 13_{10} – 15 bytes IPG. Set ipg_value{7:0} to 4_{16}. 14_{10} – 16 bytes IPG. Set ipg_value{7:0} to 5_{16}.
7:3	*Used*	0	RW	Used for spare IPG values.
2:0	ipg_value	3_{16}	RW	Defines the IPG when MAC is in xgmii mode of operation (mii_mode = 0 and gmii_mode = 0), which is timed by the TxMAC before each frame's transmission is initiated. This register value is in units of 4-byte time. Example: 1_{16} = illegal value; 2_{16} = illegal value; 3_{16} = 12- to 15-byte IPG (default value); if the var_min_ipg in TXMAC configuration register bit 2 is set, the IPG will vary from 9 to 15 bytes and average 12 bytes; 4_{16} = 16- to 19-byte IPG; 5_{16} = 20~23 byte IPG

TABLE 27-15 xMAC Minimum Frame Size – XMAC_MIN (FZC_MAC + 00088_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
29:20	rx_min_pkt_size	40_{16}	RW	Determines the minimum number of bytes that the RxMAC will receive from the medium. It is in units of 1 byte. A value of 0 disables this function. 0 = no_rx_min_pkt_size_chk; 1 = 1 byte; 2 = 2 bytes; 3 = 3 bytes;... 64 = 64 bytes. (This value should be used for Ethernet packets.)
19:18	—	0	RO	*Reserved*

TABLE 27-15 xMAC Minimum Frame Size – XMAC_MIN (FZC_MAC + 00088_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
17:10	slot_time	08_{16}	RW	Specify the slot time parameter in units of media byte time. This parameter determines the physical span of the network. The value is in units of 8-byte time. To have a 512-bit (64-byte) time, program the value to 8_{16}.
9:3	tx_min_pkt_size	40_{16}	RW	This field is the upper 7 bits of tx_min_pkt_size. The lower three bits are always 0. It determines the minimum number of bytes that the TxMac will transmit to the medium. **A value of 0 disables this function.** For IEEE 802.3 packets, software should use the POR value (64). **{tx_min_pkt_size, 000_2}:** 7'b0 Æ 0000000_000_{10}: no check for tx_min_pkt_size; 7'b1 Æ 000000_000_2 = 8 bytes; 7'b2 Æ 0000010_000_2 = 16 bytes; 7'b3 Æ 0000011_000_2 = 24 bytes; 7'b4 Æ 0000100_000_2 = 32 bytes; 7'b5 Æ 0000101_000_2 = 40 bytes;...; 7'b8 Æ 0001000_000_2 = 64 bytes (this value should be used for Ethernet packets)
2:0	—	0	RO	*Reserved*

TABLE 27-16 xMAC Maximum Frame Size – XMAC_MAX (FZC_MAC + 00090_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:14	—	0	RO	*Reserved*
13:0	max_frame_size	1518_{10}	RW	Specifies the maximum number of bytes that the TxMAC will transmit for any frame on the medium and that the RxMAC will receive from the medium before it recognizes the frame not to be valid. For xMAC TxMAC side: 1. If a packet exceeds max_frame_size with no_crc = 1 option, a worst case packet size of (max_frame_size + 8) bytes can be sent out. 2. If a packet exceeds max_frame_size with no_crc = 0 option, a worst case packet size of (max_frame_size + 8 + 4) bytes can be sent out. For xMAC RxMAC side, the value should be smaller than 14'3FEF.

27.6.5 xMAC Statistics Registers

TABLE 27-17 xMAC Receive Byte Counter – RXMAC_BT_CNT (FZC_MAC + 00100_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	rx_byte_count	0000_{16}	WR-AC	Counts number of bytes received by MAC, including fragments and packets with errors. In case of max_pkt_size violation or back pressure, rx_byte_count will only count the actual data bytes received by MAC. It increments when 8 bytes have been received successfully. Total bytes received = rx_byte_count ¥ 8. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-18 xMAC Receive Broad-Cast Frame Counter – RXMAC_BC_FRM_CNT (FZC_MAC + 00108_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:0	rx_bcast_frame_ count	X	WR-AC	Counts broadcast frames received from the network. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-19 xMAC Receive Multi-Cast Frame Counter – RXMAC_MC_FRM_CNT (FZC_MAC + 00110_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:0	rx_mcast_frame_ count	X	WR-AC	Counts multicast frames received from the network, excluding broadcast packets.

TABLE 27-20 xMAC Receive Fragments Counter – RXMAC_FRAG_CNT (FZC_MAC + 00118_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:0	rx_frag_count	X	WR-AC	This counter is not reliable under max_pkt_size violation or back pressure from RX. This counter is for debug/bring-up. Please either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. It counts frames that are rejected because their length was less than the value programmed in the rxminpacketsize field of TABLE 27-15 on page 802 and regardless of whether they had a good or bad CRC and/or alignment errors. rx_mpk_size = 0 will also disable rx_frag_count. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-21 xMAC Receive 64B Frame Counter – RXMAC_HIST_CNT1 (FZC_MAC + 00120_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:0	rx_hist_count1	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. It counts 64-byte frames received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-22 xMAC Receive 64B-127B Frame Counter – RXMAC_HIST_CNT2 (FZC_MAC + 00128_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:21	—	0	RO	*Reserved*
20:0	rx_hist_count2	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. This 21-bit counter counts the number of frames in the range of 65–127 bytes received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-23 xMAC Receive 128B-255B Frame Counter – RXMAC_HIST_CNT3 (FZC_MAC + 00130_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:0	rx_hist_count3	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. This 20-bit counter counts the number of frames in the range of 128–255 bytes received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-24 xMAC Receive 256B-511B Frame Counter – RXMAC_HIST_CNT4 (FZC_MAC + 00138_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:19	—	0	RO	*Reserved*
18:0	rx_hist_count4	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. This 19-bit counter counts the number of frames in the range of 256–511 bytes received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-25 xMAC Receive 256B-511B Frame Counter – RXMAC_HIST_CNT5 (FZC_MAC + 00140_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:0	rx_hist_count5	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. This 18-bit counter counts the number of frames in the range of 512–1023 bytes received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-26 xMAC Receive 1024B-1522B Frame Counter – RXMAC_HIST_CNT6 (FZC_MAC + 00148_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:17	—	0	RO	*Reserved*
16:0	rx_hist_count6	X	WR-AC	This counter is for debug/bring-up. Either don't use it or use it with this side effect in mind. The max_pkt_size violation or back pressure coming from RX can cause a packet to be truncated to a smaller packet length that falls into this counter's range. This 17-bit counter counts the number of frames in the range of 1024–1522 bytes received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-27 xMAC Receive Packet Count/Receive Jumbo Frame Counter – RXMAC_HIST_CNT7 (FZC_MAC + 00188_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:27	—	0	RO	*Reserved*
26:0	rx_hist_count7	X	WR-AC	Use this counter as a receive packet counter and always set mac2ipp_pkt_cnt_en to 1. Do not use it as a Jumbo Frame Counter. When mac2ipp_pkt_cnt_en is 1, it counts the total number of packets (good and bad) that MAC sent to IPP. When mac2ipp_pkt_cnt_en is 0, it counts the number of frames in the range of minimum 1523 bytes and received from the network (good and bad). The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-28 xMAC Receive Max Packet Length Error Counter – RXMAC_MPSZER_CNT (FZC_MAC + 00150_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	rx_max_pkt_size_ err_count	X	WR-AC	Counts frame whose length is greater than the value that was programmed in the Maximum Frame Size register has been received from the network. These frames include both bad (with CRC and alignment errors) and good ones. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-29 xMAC Receive CRC Error Counter – RXMAC_CRC_ER_CNT (FZC_MAC + 00158_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	rx_crc_err_count	X	WR-AC	Loadable counter increases when a receive packet fails CRC checking. If err_chk_dis is 1, the counter value won't increase. The counter is frozen at maximum value. Software must perform PIO READ to clear it. The back pressure and/or max_pkt_size exceeded truncated packets will not increase this counter.

TABLE 27-30 xMAC Receive Code Violation Error Counter – RXMAC_CD_VIO_CNT (FZC_MAC + 00160_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	rx_code_viol_err_ count	X	WR-AC	Loadable counter increments when a received-packet RX_ER indication is detected over the 10/100/1000M link or E over 10G link. If err_chk_dis is 1, the counter value will not increase. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-31 xMAC Transmit Frame Counter – TXMAC_FRM_CNT (FZC_MAC + 00170_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	tx_frame_count	0000_{16}	WR-AC	Loadable counter increments when a frame has been transferred successfully. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-32 xMAC Transmit Byte Counter – TXMAC_BYTE_CNT (FZC_MAC + 00178_{16}) count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	tx_byte_count	0000_{16}	WR-AC	Loadable counter increments when 8 bytes have been transmitted successfully. Total bytes transmitted = tx_byte_count ¥ 8. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

TABLE 27-33 xMAC Link Fault Counter – LINK_FAULT_CNT (FZC_MAC + 00180_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	link_fault_count	X	WR-AC	Loadable counter increments when a link fault has been detected. The counter is frozen at maximum value. Software must perform PIO READ to clear it.

27.6.6 xMAC Miscellaneous MAC Registers

TABLE 27-34 xMAC State Machines – XMAC_SM_REG (FZC_MAC + $001A8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	state_machines_state	X	RO	This register provides the current state for all the state machines in the MAC. state_machines_state = { rxfifo_empty_clk_reg, rxfifo_full_clk_reg, txfifo_empty_clk_reg, txfifo_full_clk_reg, rxfifo_rd_ptr_clk{4:0}, rxfifo_wr_ptr_clk{4:0}, 00000_2, rxmac_ipp_stat_reg{22:20}, alt_addr_and_hash_func_value, xgmii_only_value, xrlm_state, mgrlm_state, lfs_state{1:0}, sop_state, xtlm_state{2:0}};

TABLE 27-35 xMAC Internal Signals 1 – XMAC_INTERN1 (FZC_MAC + $001B0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	internal_signals1	X	RO	This register contains xMAC-IPP and xMAC-OPP interface protocol signals. It provides a way to look inside the MAC internal signals for diagnostic purposes. internal_signals1 = { rxmac_ipp_ack_reg, ipp_rxmac_req_reg, rxmac_ipp_tag_reg, rxmac_ipp_ctrl_reg, rxmac_ipp_stat_reg{19:0}, txmac_opp_req_reg, opp_txmac_ack_reg, opp_txmac_tag_reg, opp_txmac_abort_reg, opp_txmac_stat_reg{3:0}};

TABLE 27-36 xMAC Internal Signals 2 – XMAC_INTERN2 (FZC_MAC + $001B8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	internal_signals2	X	RO	internal_signals2 = { $00\ 0000_2$, remote_fault_oc_sync, local_fault_oc_sync, xpcs_txc{7:0}, xpcs_rxc{7:0}, rx_heart_beat_timer_reg{3:0}, tx_heart_beat_timer_reg{3:0}}

27.6.7 MAC Station Address Format Programming Guide

TABLE 27-41, MAC Addresses, on page 813 shows the relationship between local 48-bit MAC Address register order construction and the incoming packet 48-bit da field.

The incoming 48-bit da field is composed of 6 octets. From left (MSB) to right (LSB), they are **octet 0, octet 1, octet 2, octet 3, octet 4, octet 5**.

TABLE 27-37 MAC Station Address Format

Station Address →	Octet 0,1 {Octet a,b}	Octet 2, 3 {Octet c,d}	Octet 4, 5 {Octet e,f}
Station unique address	MAC Address 2 register {15:0}	MAC Address 1 register {15:0}	MAC Address 0 register {15:0}
Reserved multicast address (it is the same as MAC unique address)	Mac Address 2 with reserved_multicast bit set (bit 13 of MAC Configuration register)	MAC Address 1 with reserved_multicast bit set (bit 13 of MAC Configuration register)	MAC Address 0 with reserved_multicast bit set (bit 13 of MAC Configuration register)
Alternate MAC address	Alternate MAC Address 2 register {15:0} (most significant 16 bits)	Alternate MAC Address 1 register {15:0} (middle significant 16 bits)	Alternate MAC Address 0 register {15:0} (least significant 16 bits)
Mac control address	hardwired (**01 8016**)	hardwired (**C2 0016**)	hardwired (**00 0116**)

Note | A MAC address of the form **a:b:c:d:e:f** where octets **{a,b,c}** comprise the OUI part of the address and octets **{d,e,f}** are the vendor unique serial number would be stored as shown in the table below. The Group Address bit is the least significant bit of byte **a**.

27.6.8 MAC Unique Address/Reserved Multicast Address

MAC Unique Address register consists of three registers:

- MAC (Unique) Address 0 register
- MAC (Unique) Address 1 register
- MAC (Unique) Address 2 register

MAC Unique Address register is a dual-role register.

When bit 13 (reserved_multicast) of MAC Configuration register is 0, the MAC Unique Address register should be used as a perfect match to da.

When bit 13 (reserved_multicast) of MAC Configuration register is set to 1, MAC Unique Address Register should be used as Reserved Multicast Address register. In this case, the incoming MAC frame da bit {35:32} will not be checked against macuniqueaddr0{3:0}.

MAC unique address, MAC flow control address, MAC hash address checking cannot be disabled. The MAC Unique Address and Reserved Multicast Address registers are nonmaskable. Software should program a valid value.

TABLE 27-38 xMAC Address 0 – XMAC_ADDR0 (FZC_MAC + $000A0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_unique_addr 0	X	RW	Contains the 16 least significant bits of the station's normal priority MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network. The MAC address programmed in this register must be a unicast address.

TABLE 27-39 xMAC Address 1 – XMAC_ADDR1 (FZC_MAC + $000A8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_unique_addr 1	X	RW	Contains the 16 middle bits of the station's normal priority MAC Address. Bits 15:0 will be compared with {octet 2, octet3} of the 6-octet da field for every frame that arrives from the network. The MAC address programmed in this register must be a unicast address.

TABLE 27-40 xMAC Address 2 – XMAC_ADDR2 (FZC_MAC + $000B0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_unique_addr 2	X		Contains the 16 most significant bits of the station's normal priority MAC Address. Bits 15:0 will be compared with {octet 0, octet1} of the 6-octet da field for every frame that arrives from the network. The MAC address programmed in this register must be a unicast address.

27.6.9 MAC Flow Control Frame

The PAUSE operation is used to inhibit transmission of data frames for a specific period of time. The globally assigned, 48-bit, well-known multicast address **01_80_C2_00_00_01** (byte 0, 1, 2, 3, 4, 5) has been reserved for use in MAC Control PAUSE frames in a **full-duplex mode** IEEE 802.3 LAN for flow control purpose. It is implemented in the xMAC as a hardwired value that is compared with the incoming packet da field.

TABLE 27-41 MAC Addresses

	ab	cd	ef
Station Unique Address	MAC Address 2 register	MAC Address 1 register	MAC Address 0 register
Reserved Multicast Address (same as MAC Unique Address)	MAC Address 2 With reserved_multicast bit set (bit 13 of MAC Configuration register)	MAC Address 1 With reserved_multicast bit set (bit 13 of MAC Configuration register)	MAC Address 0 With reserved_multicast bit set (bit 13 of MAC Configuration register)
MAC Control Address	hardwired (01_80_{16})	hardwired ($C2\_00_{16}$)	hardwired (00_01_{16})

27.6.10 xMAC Alternate MAC Address Registers

The Alternate Address Compare Enable register enables the checking of one of the 16 alternate MAC addresses.

There are total 48 (16 MAC address x 3 registers) alternate MAC address registers (48 bits wide) are used to house 16 alternate MAC addresses. Each register contains 16 bits worth of the 48-bit MAC destination address.

Note | hash_hit_en and filter_match are enabled through control bits in the MAC Configuration register bit 16:15. MAC Unique Address and flow control match are always on.

TABLE 27-42 xMAC Alternate Address Compare Enable – XMAC_ADDR_CMPEN (FZC_MAC + 00208_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	alt_addr_cmp_en15	0	RW	0 = disable checking/comparison of alternate address 15. 1 = enable checking/comparison of alternate address 15.
14	alt_addr_cmp_en14	0	RW	0 = disable checking/comparison of alternate address 14. 1 = enable checking/comparison of alternate address 14.
13	alt_addr_cmp_en13	0	RW	0 = disable checking/comparison of alternate address 13. 1 = enable checking/comparison of alternate address 13.
12	alt_addr_cmp_en12	0	RW	0 = disable checking/comparison of alternate address 12. 1 = enable checking/comparison of alternate address 12.
11	alt_addr_cmp_en11	0	RW	0 = disable checking/comparison of alternate address 11. 1 = enable checking/comparison of alternate address 11.
10	alt_addr_cmp_en10	0	RW	0 = disable checking/comparison of alternate address 10. 1 = enable checking/comparison of alternate address 10.
9	alt_addr_cmp_en9	0	RW	0 = disable checking/comparison of alternate address 9. 1 = enable checking/comparison of alternate address 9.

TABLE 27-42 xMAC Alternate Address Compare Enable – XMAC_ADDR_CMPEN (FZC_MAC + 00208_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
8	alt_addr_cmp_en8	0	RW	0 = disable checking/comparison of alternate address 8. 1 = enable checking/comparison of alternate address 8.
7	alt_addr_cmp_en7	0	RW	0 = disable checking/comparison of alternate address 7. 1 = enable checking/comparison of alternate address 7.
6	alt_addr_cmp_en6	0	RW	0 = disable checking/comparison of alternate address 6. 1 = enable checking/comparison of alternate address 6.
5	alt_addr_cmp_en5	0	RW	0 = disable checking/comparison of alternate address 5. 1 = enable checking/comparison of alternate address 5.
4	alt_addr_cmp_en4	0	RW	0: disable checking/comparison of alternate address 4. 1 = enable checking/comparison of alternate address 4.
3	alt_addr_cmp_en3	0	RW	0 = disable checking/comparison of alternate address 3. 1 = enable checking/comparison of alternate address 3.
2	alt_addr_cmp_en2	0	RW	0 = disable checking/comparison of alternate address 2. 1 = enable checking/comparison of alternate address 2.
1	alt_addr_cmp_en1	0	RW	0 = disable checking/comparison of alternate address 1. 1 = enable checking/comparison of alternate address 1.
0	alt_addr_cmp_en0	0	RW	0 = disable checking/comparison of alternate address 0. 1 = enable checking/comparison of alternate address 0.

TABLE 27-43 xMAC Alternate Address 0 LSB – XMAC_ADDR3 (FZC_MAC + 00218_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr0_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-44 xMAC Alternate Address 0 MID – XMAC_ADDR4 (FZC_MAC + 00220_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	X	RO	*Reserved*
15:0	mac_alt_addr0_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-45 xMAC Alternate Address 0 MSB – XMAC_ADDR5 (FZC_MAC + 00228_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	X	RO	*Reserved*
15:0	mac_alt_addr0_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-46 xMAC Alternate Address 1 LSB – XMAC_ADDR6 (FZC_MAC + 00230_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr1_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-47 xMAC Alternate Address 1 MID – XMAC_ADDR7 (FZC_MAC + 00238_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr1_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-48 xMAC Alternate Address 1 MSB – XMAC_ADDR8 (FZC_MAC + 00240_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr1_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-49 xMAC Alternate Address 2 LSB – XMAC_ADDR9 (FZC_MAC + 00248_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr2_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-50 xMAC Alternate Address 2 MID – XMAC_ADDR10 (FZC_MAC + 00250_{16}) (count 2 step06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr2_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-51 xMAC Alternate Address 2 MSB – XMAC_ADDR11 (FZC_MAC + 00258_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr2_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-52 xMAC Alternate Address 3 LSB – XMAC_ADDR12 (FZC_MAC + 00260_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr3_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-53 xMAC Alternate Address 3 MID – XMAC_ADDR13 (FZC_MAC + 00268_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr3_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-54 xMAC Alternate Address 3 MSB – XMAC_ADDR14 (FZC_MAC + 00270_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr3_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-55 xMAC Alternate Address 4 LSB – XMAC_ADDR15 (FZC_MAC + 00278_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr4_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. These bits {15:0} will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-56 xMAC Alternate Address 4 MID – XMAC_ADDR16 (FZC_MAC + 00280_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr4_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-57 xMAC Alternate Address 4 MSB – XMAC_ADDR17 (FZC_MAC + 00288_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr4_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-58 xMAC Alternate Address 5 LSB – XMAC_ADDR18 (FZC_MAC + 00290_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr5_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-59 xMAC Alternate Address 5 MID – XMAC_ADDR19 (FZC_MAC + 00298_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr5_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-60 xMAC Alternate Address 5 MSB – XMAC_ADDR20 (FZC_MAC + $002A0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr5_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-61 xMAC Alternate Address 6 LSB – XMAC_ADDR21 (FZC_MAC + $002A8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr6_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-62 xMAC Alternate Address 6 MID – XMAC_ADDR22 (FZC_MAC + $002B0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr6_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-63 xMAC Alternate Address 6 MSB – XMAC_ADDR23 (FZC_MAC + $002B8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr6_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-64 xMAC Alternate Address 7 LSB – XMAC_ADDR24 (FZC_MAC + $002C0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr7_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-65 xMAC Alternate Address 7 MID – XMAC_ADDR25 (FZC_MAC + $002C8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr7_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-66 xMAC Alternate Address 7 MSB – XMAC_ADDR26 (FZC_MAC + $002D0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr7_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-67 xMAC Alternate Address 8 LSB – XMAC_ADDR27 (FZC_MAC + $002D8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr8_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-68 xMAC Alternate Address 8 MID – XMAC_ADDR28 (FZC_MAC + $002E0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr8_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-69 xMAC Alternate Address 8 MSB – XMAC_ADDR29 (FZC_MAC + $002E8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr8_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-70 xMAC Alternate Address 9 LSB – XMAC_ADDR30 (FZC_MAC + $002F0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr9_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-71 xMAC Alternate Address 9 MID – XMAC_ADDR31 (FZC_MAC + $002F8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr9_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-72 xMAC Alternate Address 9 MSB – XMAC_ADDR32 (FZC_MAC + 00300_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr9_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-73 xMAC Alternate Address 10 LSB – XMAC_ADDR33 (FZC_MAC + 00308_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr10_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-74 xMAC Alternate Address 10 MID – XMAC_ADDR34 (FZC_MAC + 00310_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr10_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-75 xMAC Alternate Address 10 MSB – XMAC_ADDR35 (FZC_MAC + 00318_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr10_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-76 xMAC Alternate Address 11 LSB – XMAC_ADDR36 (FZC_MAC + 00320_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr11_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-77 xMAC Alternate Address 11 MID – XMAC_ADDR37 (FZC_MAC + 00328_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr11_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0} will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-78 xMAC Alternate Address 11 MSB – XMAC_ADDR38 (FZC_MAC + 00330_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr11_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-79 xMAC Alternate Address 12 LSB – XMAC_ADDR39 (FZC_MAC + 00338_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr12_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-80 xMAC Alternate Address 12 MID – XMAC_ADDR40 (FZC_MAC + 00340_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr12_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-81 xMAC Alternate Address 12 MSB – XMAC_ADDR41 (FZC_MAC + 00348_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr12_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-82 xMAC Alternate Address 13 LSB – XMAC_ADDR42 (FZC_MAC + 00350_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr13_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-83 xMAC Alternate Address 13 MID – XMAC_ADDR43 (FZC_MAC + 00358_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr13_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-84 xMAC Alternate Address 13 MSB – XMAC_ADDR44 (FZC_MAC + 00360_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr13_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-85 xMAC Alternate Address 14 LSB – XMAC_ADDR45 (FZC_MAC + 00368_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr14_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-86 xMAC Alternate Address 14 MID – XMAC_ADDR46 (FZC_MAC + 00370_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr14_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-87 xMAC Alternate Address 14 MSB – XMAC_ADDR47 (FZC_MAC + 00378_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr14_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-88 xMAC Alternate Address 15 LSB – XMAC_ADDR48 (FZC_MAC + 00380_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr15_0	X	RW	Contains the 16 least significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-89 xMAC Alternate Address 15 MID – XMAC_ADDR49 (FZC_MAC + 00388_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr15_1	X	RW	Contains the 16 middle significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-90 xMAC Alternate Address 15 MSB – XMAC_ADDR50 (FZC_MAC + 0390_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_alt_addr15_2	X	RW	Contains the 16 most significant bits of the station's alternate MAC address. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

27.6.11 xMAC Address Filter Registers

Address filter works on one byte a time. In other words, one byte comparison only. Rest of the DA octets should be disabled by programming its associated mask bits to logic "1" and program the associated address filter register byte with value that you do not want to match. This is DIFFERENT from the DA octet that you want to compare. Any one byte in Address Filter Register matches its associated DA octet will mark as DA match and packet will be forward to Host.

filter_mask_lsb allows the DA match down to bit granularity with in its associated byte. filter_mask_msb allows the DA match down to nibble granularity with in its associated byte.

xMAC Address Filter usage model is very restrictive. Use it carefully.

TABLE 27-91 xMAC Address Filter LSB – XMAC_ADD_FILT0 (FZC_MAC + 00818_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_addr_filter_0	X	RW	xMAC Address Filter usage model is very restrictive. Use it carefully. Contains the 16 least significant bits of the station's MAC Address Filter. Bits 15:0 will be compared with {octet 4, octet 5} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-92 xMAC Address Filter MID – XMAC_ADD_FILT1 (FZC_MAC + 00820_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_addr_filter_1	X	RW	xMAC Address Filter usage model is very restrictive. Use it carefully. Contains the 16 middle significant bits of the station's MAC Address Filter. Bits 15:0 will be compared with {octet 2, octet 3} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-93 xMAC Address Filter MSB – XMAC_ADD_FILT2 (FZC_MAC + 00828_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_addr_filter_2	X	RW	xMAC Address Filter usage model is very restrictive. Use it carefully. Contains the 16 most significant bits of the station's MAC Address Filter. Bits 15:0 will be compared with {octet 0, octet 1} of the 6-octet da field for every frame that arrives from the network.

TABLE 27-94 xMAC Address Filter Mask MSB – XMAC_ADD_FILT12_MASK (FZC_MAC + 00830_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	mac_addr_filter_ mask_msb	X	RW	xMAC Address Filter usage model is very restrictive. Use it carefully. Contains the 8 most significant bits of the Address Filter Mask. Each bit in this register corresponds to a nibble (4 bits) in the Address Filter registers 2 and 1. Bits 7:6 are used to mask Address Filter register 2 ({octet0, octet1}). Bits 3:0 are used to mask Address Filter register 1 ({octet2, octet3}). This is an active low signal. A bit value of 1 enables comparison of incoming data stream, and 0 disables it.

TABLE 27-95 xMAC Address Filter Mask LSB – XMAC_ADD_FILT00_MASK (FZC_MAC + 00838_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	mac_addr_filter_ mask_lsb	X	RW	xMAC Address Filter usage model is very restrictive. Use it carefully. Contains the 16 least significant bits of the address filter mask. Each bit in this register corresponds to the bits in the Address Filter registers 0.

27.6.12 xMAC Hash Table Registers

Hash Table 0 Register .

TABLE 27-96 xMAC Hash Table 0 – XMAC_HASH_TBL0 (FZC_MAC + 00840_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_0	X	RW	Contains bits 255:240 of the Hash table.

Hash Table 1 Register .

TABLE 27-97 xMAC Hash Table 1 – XMAC_HASH_TBL1 (FZC_MAC + 00848_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_1	X	RW	Contains bits 239:224 of the Hash table.

Hash Table 2 Register .

TABLE 27-98 xMAC Hash Table 2 – XMAC_HASH_TBL2 (FZC_MAC + 00850_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_2	X	RW	Contains bits 223:208 of the Hash table.

Hash Table 3 Register .

TABLE 27-99 xMAC Hash Table 3 – XMAC_HASH_TBL3 (FZC_MAC + 00858_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_3	X	RW	Contains bits 207:192 of the Hash table.

Hash Table 4 Register .

TABLE 27-100 xMAC Hash Table 4 – XMAC_HASH_TBL4 (FZC_MAC + 00860_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_4	X	RW	Contains bits 191:176 of the Hash table.

Hash Table 5 Register .

TABLE 27-101 xMAC Hash Table 5 – XMAC_HASH_TBL5 (FZC_MAC + 00868_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_5	X	RW	Contains bits 175:160 of the Hash table.

Hash Table 6 Register .

TABLE 27-102 xMAC Hash Table 6 – XMAC_HASH_TBL6 (FZC_MAC + 00870_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_6	X	RW	Contains bits 159:144 of the Hash table.

Hash Table 7 Register .

TABLE 27-103 xMAC Hash Table 7 – XMAC_HASH_TBL7 (FZC_MAC + 00878_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_7	X	RW	Contains bits 143:128 of the Hash table.

Hash Table 8 Register .

TABLE 27-104 xMAC Hash Table 8 – XMAC_HASH_TBL8 (FZC_MAC + $0x00880_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_8	X	RW	Contains bits 127:112 of the Hash table.

Hash Table 9 Register .

TABLE 27-105 xMAC Hash Table 9 – XMAC_HASH_TBL9 (FZC_MAC + 00888_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_9	X	RW	Contains bits 111:96 of the Hash table.

Hash Table 10 Register .

TABLE 27-106 xMAC Hash Table 10 – XMAC_HASH_TBL10 (FZC_MAC + 00890_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_10	X	RW	Contains bits 95:80 of the Hash table.

Hash Table 11 Register .

TABLE 27-107 xMAC Hash Table 11 – XMAC_HASH_TBL11 (FZC_MAC + 00898_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_11	X	RW	Contains bits 79:64 of the Hash table.

Hash Table 12 Register .

TABLE 27-108 xMAC Hash Table 12 – XMAC_HASH_TBL12 (FZC_MAC + $008A0_{16}$) (count 2 step 06000

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_12	X	RW	Contains bits 63:48 of the Hash table.

Hash Table 13 Register .

TABLE 27-109 xMAC Hash Table 13 – XMAC_HASH_TBL13 (FZC_MAC + $008A8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_13	X	RW	Contains bits 47:32 of the Hash table.

Hash Table 14 Register .

TABLE 27-110 xMAC Hash Table 14 – XMAC_HASH_TBL14 (FZC_MAC + $008B0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_14	X	RW	Contains bits 31:16 of the Hash table.

Hash Table 15 Register .

TABLE 27-111 xMAC Hash Table 15 – XMAC_HASH_TBL15 (FZC_MAC + $008B8_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:0	xmac_hash_15	X	RW	Contains bits 15:0 of the Hash table.

xMAC Host Info Register.

MAC host info registers contain the Rx DMA channel table number and its associated preference bit that will be further processed by packet classification engine to determine the final Rx DMA channel table number and Rx DMA channel table offset value for the incoming packets.

ZCP uses the final Rx DMA channel table number and Rx DMA channel table offset value to determine the Rx DMA channel number for the packet.

Since both the VLAN table (in FFLP) and the MAC address table can set the preference, the following truth table (TABLE 27-112) is used to resolve the final preference.

Note that an invalid VLAN should have its vpr bit set to 0. In this case, the MAC preference is used.

Note that VPR is located in FFLP.

TABLE 27-112 MAC Address / VLAN Preference Logic

MPR	VPR	Selection	DMA Table Offset Value
0	0	macrdctbln	Always 0.
0	1	vlanrdctbln	FFLP produces the final number.
1	0	macrdctbln	Always 0.
1	1	vlanrdctbln	FFLP produces the final number.

Note that at the end of layer 2 parsing, an RDC Table number that the parser recognizes is always associated with the incoming frame. Frames with errors and frames that do not have a valid MAC address will use the port default Receive DMA channel, if they are not discarded.

There are a total of 20 MAC host info table entries in xMAC. When the alternate MAC address compare bit is enabled and there is a hit or match, the corresponding MAC host info data is forward to FFL (Forwarding Filtering and Learning) via the IPP block.

The following description describes the relationship between each MAC station address and its associated host info register:

```
begin
if (mac_alt_addr0_match)
  mac_info = {mac_host_info0,p_hit,promisc_all};
else if (mac_alt_addr1_match)
  mac_info = {mac_host_info1,p_hit,promisc_all};
else if (mac_alt_addr2_match)
  mac_info = {mac_host_info2,p_hit,promisc_all};
else if (mac_alt_addr3_match)
  mac_info = {mac_host_info3,p_hit,promisc_all};
else if (mac_alt_addr4_match)
  mac_info = {mac_host_info4,p_hit,promisc_all};
else if (mac_alt_addr5_match)
  mac_info = {mac_host_info5,p_hit,promisc_all};
else if (mac_alt_addr6_match)
  mac_info = {mac_host_info6,p_hit,promisc_all};
else if (mac_alt_addr7_match)
  mac_info = {mac_host_info7,p_hit,promisc_all};
else if (mac_alt_addr8_match)
  mac_info = {mac_host_info8,p_hit,promisc_all};
else if (mac_alt_addr9_match)
  mac_info = {mac_host_info9,p_hit,promisc_all};
else if (mac_alt_addr10_match)
  mac_info = {mac_host_info10,p_hit,promisc_all};
else if (mac_alt_addr11_match)
  mac_info = {mac_host_info11,p_hit,promisc_all};
else if (mac_alt_addr12_match)
  mac_info = {mac_host_info12,p_hit,promisc_all};
```

```
    else if (mac_alt_addr13_match)
      mac_info = {mac_host_info13,p_hit,promisc_all};
    else if (mac_alt_addr14_match)
      mac_info = {mac_host_info14,p_hit,promisc_all};
    else if (mac_alt_addr15_match)
      mac_info = {mac_host_info15,p_hit,promisc_all};
    else if (hash_hit_match)//perfect match has higher pri over hash hit
      mac_info = {mac_host_info16,p_hit,promisc_all};
    else if (mac_own_da_match) // station unique address
      mac_info = {mac_host_info17,p_hit,promisc_all};
    else if (addr_filter_match)
      mac_info = {mac_host_info18,p_hit,promisc_all};
    else if (mac_fc_match)
      mac_info = {mac_host_info19,p_hit,promisc_all};
    else
        // no perfect hit or hash hit.
        // This condition will happen only when
        // promiscuous all, promiscuous group, broadcast packets.
      mac_info = {mac_host_info16,1'b0,promisc_all};
end
```

TABLE 27-113 through TABLE 27-132 show the layout of the Host Info register.

TABLE 27-113 xMAC Host Info 0 – XMAC_HOST_INFO0 (FZC_MAC + 00900_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-114 xMAC Host Info 1 – XMAC_HOST_INFO1 (FZC_MAC + 00908_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-115 xMAC Host Info 2 – XMAC_HOST_INFO2 (FZC_MAC + 00910_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-116 xMAC Host Info 3 – XMAC_HOST_INFO3 (FZC_MAC + 00918_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-117 xMAC Host Info 4 – XMAC_HOST_INFO4 (FZC_MAC + 00920_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-118 xMAC Host Info 5 – XMAC_HOST_INFO5 (FZC_MAC + 00928_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive channel table number

TABLE 27-119 xMAC Host Info 6 – XMAC_HOST_INFO6 (FZC_MAC + 00930_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-120 xMAC Host Info 7 – XMAC_HOST_INFO7 (FZC_MAC + 00938_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-121 xMAC Host Info 8 – XMAC_HOST_INFO8 (FZC_MAC + 00940_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-122 xMAC Host Info 9 – XMAC_HOST_INFO9 (FZC_MAC + 00948_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-123 xMAC Host Info 10 – XMAC_HOST_INFO10 (FZC_MAC + 00950_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-124 xMAC Host Info 11 – XMAC_HOST_INFO11 (FZC_MAC + 00958_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-125 xMAC Host Info 12 – XMAC_HOST_INFO12 (FZC_MAC + 00960_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-126 xMAC Host Info 13 – XMAC_HOST_INFO13 (FZC_MAC + 00968_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-127 xMAC Host Info 14 – XMAC_HOST_INFO14 (FZC_MAC + 00970_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-128 xMAC Host Info 15 – XMAC_HOST_INFO15 (FZC_MAC + 00978_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-129 xMAC Host Info 16 – XMAC_HOST_INFO16 (FZC_MAC + 00980_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-130 xMAC Host Info 17 – XMAC_HOST_INFO17 (FZC_MAC + 00988_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-131 xMAC Host Info 18 – XMAC_HOST_INFO18 (FZC_MAC + 00990_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

TABLE 27-132 xMAC Host Info 19 – XMAC_HOST_INFO19 (FZC_MAC + 00998_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:18	—	0	RO	*Reserved*
17:9	*Used*	0	RW	Register RW test.
8	mpr	0	RW	When set to 1, indicates preference for MAC.
7:3	*Used*	0	RW	Register RW test.
2:0	macrdctbln	0	RW	Individual Receive DMA channel table number

Preamble Data Register .

TABLE 27-133 xMAC Preamble Data 0 – XMAC_PA_DATA0 (FZC_MAC + $00B80_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	rx_pa_data0	0	RO	Contains the first 4 bytes of preamble data. rx_pa_data{31:0} = {Byte3, Byte2, Byte1, Byte0};

TABLE 27-134 xMAC Preamble Data 1 – XMAC_PA_DATA1 (FZC_MAC + $00B88_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	rx_pa_data1	X	RO	Contains the last 4 bytes of preamble data. rx_pa_data{63:32} = {Byte7, Byte6, Byte5, Byte4};

xMAC Debug Select Register .

TABLE 27-135 xMAC Debug Select – XMAC_DEBUG_SEL (FZC_MAC+0x00B90) (count 2 step x06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7	*Used*	0	RW	Used for spare bit.

6:3	xmac_debug_sel	0	RW	This 4-bit register selects one of the xMAC debug buses defined in the following case statement to go outside UltraSPARC T2 for debugging/probing. Final debug bus that goes outside UltraSPARC T2 for debugging/probing is defined in the MAC_DEBUG_SEL register.
				case (xmac_debug_sel) // synopsis parallel_case full_case
				4'h0: xmac_debug = state_machine0;
				4'h1: xmac_debug = xpcs_rxd[31:0];
				4'h2: xmac_debug = xpcs_rxd[63:32];
				4'h3: xmac_debug = xpcs_txd[31:0];
				4'h4: xmac_debug = xpcs_txd[63:32];
				4'h5: xmac_debug = internal_signals1;
				4'h6: xmac_debug = internal_signals2; 4'hA: xmac_debug = rx_internal_signals1; 4'hB: xmac_debug = tx_internal_signals1;
				4'hC: xmac_debug = tx_internal_signals2;
				'SEL_mac_training_vector: xmac_debug = training_vector[31:0];
				default: xmac_debug = state_machine0;
				endcase
				Please see the following notes for the definition of each right-hand side signals.

2:0	mac_debug_sel	0	RW	This 3-bit register selects one of the MAC debug buses defined in the following case statement to go outside UltraSPARC T2 for debugging/probing. Only port 0 xmac of this register is used. Port 1 xmac of this register is not used.
				case (mac_debug_sel0) // synopsis parallel_case full_case
				3'h0: mac_debug_port = xmac_debug0; // port 0
				3'h1: mac_debug_port = xpcs_debug0; // port 0
				3'h2: mac_debug_port = xmac_debug1; // port 1
				3'h3: mac_debug_port = xpcs_debug1; // port 1
				3'h4: mac_debug_port = serdes_debug;
				default: mac_debug_port = xmac_debug0;
				endcase // casex(mac_debug_sel)

```
assign state_machine0 = {
                        rxfifo_empty_clk_reg,
                        rxfifo_full_clk_reg,
                        txfifo_empty_clk_reg,
                        txfifo_full_clk_reg,
                        rxfifo_rd_ptr_clk[4:0],
                        rxfifo_wr_ptr_clk[4:0],
                        5'b0,
                        rxmac_ipp_stat_reg[22:20],
```

```
                    ALT_ADDR_AND_HASH_FUNC_value,
                    XGMII_ONLY_value,
                    xrlm_state,
                    mgrlm_state,
                    lfs_state[1:0],
                    sop_state,
                    xtlm_state[2:0]
                    };
assign internal_signals1 = {
                    rxmac_ipp_ack_reg,
                    ipp_rxmac_req_reg,
                    rxmac_ipp_tag_reg,
                    rxmac_ipp_ctrl_reg,
                    rxmac_ipp_stat_reg[19:0],
                    txmac_opp_req_reg,
                    opp_txmac_ack_reg,
                    opp_txmac_tag_reg,
                    opp_txmac_abort_reg,
                    opp_txmac_stat_reg[3:0]
                    };
 assign internal_signals2 = {6'b0,
                        remote_fault_oc_sync,
                        local_fault_oc_sync,
                        xpcs_txc[7:0],
                        xpcs_rxc[7:0],
                        rx_heart_beat_timer_reg,// [3:0]
                        tx_heart_beat_timer_reg // [3:0]
                        };
 assign rx_internal_signals1 = {12'b0,rx_clk_div2,rx_nbclk_div2,
                        S_detected_reg,
                        T_E_detected_at_modified_pkt_reg,
                        END_PKT_ERR_detected_a_at_modified_pkt_reg,
                        END_PKT_ERR_detected_b_at_modified_pkt_reg,
                        S_D_reg,
                        S_I_reg,
                        D_S_reg,
                        I_S_reg,
                        abort_bit_reg,
                        rx_err_reg,
                        crc_error_reg,
                        kill_data_ready_reg,
                        kill_crc_reg,
                        rx_sel_reg[1:0],
                         //
                        last_byte_position[2:0]
                        };
assign tx_internal_signals1 = {tx_clk_div2,tx_nbclk_div2,
```

```
                                txfifo_rd_ptr_clk[4:0],
                                txfifo_wr_ptr_clk[4:0],
                                tx_swap_reg,
                                tx_on_reg,
                                tx_on_half_reg,
                                back2back_swap_reg1,
                                replace_txd_time_reg,
                                adjust2crc_full_case_last_byte_position_reg[2:0],
                                adjust2crc_full_case_last_byte_position_is_3_or_7_reg,
                                stretch_clks_reg[‘BYTE],
                                full_case_last_byte_position_reg[2:0]
                                };
        assign tx_internal_signals2 = {1’b0,
                                stretch_full_case_last_byte_position_reg[3:0],
                                stretch_bytes_reg[2:0],
                                minus_4bytes_reg,
                                B_eop_reg,
                                stretch_1_more_clk_reg,
                                no_wasted_BW_reg,
                                ipg_done_trail_temp_reg,
                                tx_byte0_reg0[7:0],
                                restart_ipg_timer_reg,
                                eop_txclk_reg0,
                                eop_w_fcs_reg0,
                                tx_abort_reg0,
                                eop_w_fcs_reg1,
                                tx_abort_reg1,
                                ipg_done_reg,
                                ipg_done_lead_temp_reg,
                                force_ipg_done_lead_reg,
                                set_back2back_reg,
                                back2back_reg
                                };
```

TABLE 27-136 xMAC Training Vector – XMAC_TRAIN_VEC (FZC_MAC + $00B98_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	training_vector	0	RW	Training vector to calibrate OSC. Load this register with training vector value. If (xmac_debug_sel{3:0} = F_{16}), the training vector starts toggling; else, it is a static value.

27.7 XPCS Programmable Resources

27.7.1 Register Description

The following registers control all modes of (XPCS) operation per IEEE Clause 45.

BASE10G_CONTROL1.

XPCS Control 1 register. This register controls XPCS key modes of operation: loopback vs. functional mode. It also provides the software reset for XPCS. Speed selection is not available, and bits 5 through 2 are hardwired to indicate 10-Gbyte speed only. Additionally, low-power mode is not available, and read/writes to this bit have no effect.

TABLE 27-137 XPCS Control 1 – BASE10G_CONTROL1 (FZC_MAC + 02000_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	reset	0	RW-AC	Self-clearing reset. 0 = Normal operation; 1 = reset.
14	loopback	0	RW	Loopback mode. Loopback is XPCS Tx Æ XPCS Rx datapath. .0 = Disable XPCS loopback mode; 1 = Enable XPCS loopback mode.
13	speed_select3	1_{16}	RO	Reads 1 to indicate 10-Gbyte speed. Writes are ignored.
12	—	0	RO	Always 0. *Reserved* bit per IEEE 802.3ae clause 45.
11	csr_low_power	0	RW	Per IEEE 802.3ae clause 45, XPCS responds to PIOs while in low power mode. Read/writes to this bit have no effect. 0 = Normal operation; 1 = Low power mode.
10:7	—	0	RO	Always 0. *Reserved* bit per IEEE 802.3ae clause 45.
6	csr_speed_select1	1_{16}	RO	Reads 1 to indicate 10-Gbyte speed. Writes are ignored.
5:2	csr_speed_select0	0	RO	Reads 0000_2 to indicate 10 G-byte speed. Writes are ignored.
1:0	—	0	RO	Always 0. *Reserved* bit per IEEE 802.3ae clause 45.

BASE10G_STATUS1.

XPCS Status 1 register. This register provides status information of link and fault. On PCS interrupt, the software Interrupt Status Utility must read this register to check for unmasked link and/or fault status.

TABLE 27-138 XPCS Status 1 – BASE10G_STATUS1 (FZC_MAC + 02008_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:8	—	00_{16}	RO	*Reserved* bits per IEEE 802.3ae clause 45. Ignore when read.
7	csr_fault	0	RO	When read as a one, it indicates that the XPCS has detected a fault condition on either the transmit or receive paths. When read as a zero, it indicates that the XPCS has not detected a fault condition. This bit read as a one when either of the fault bits (bit 10, 11) in TABLE 27-143, XPCS Status 2 – base10g_status2 (fzc_mac + 0203016) (count 2 step 06000), on page 845 are read as a one. Generates a maskable interrupt. Must read STATUS2 register to clear this interrupt. 0 = Fault condition not detected; 1 = Either csr_transmit_fault or csr_receive_fault is set in STATUS2 register.
6:3	—	0_{16}	RO	*Reserved* bits per IEEE 802.3ae clause 45. Ignore when read.
2	csr_receive_link_status	0	RO	It indicates XAUI lane/link status. State change of this bit generates an interrupt. Read BASE10G_STATUS1 (current) register to clear interrupt. The interrupt mask bit is in TABLE 27-149, XPCS Mask 1 – base10g_mask1 (fzc_mac + 0206016) (count2step06000), on page 849 bit 2. 0 = four receive xaui lanes are not aligned/link is down; 1 = four receive xaui lanes are aligned/link is up.
1	csr_low_power_ability	0	RO	0 = XPCS does not supports low power mode. See Control1 register.
0	—	0	RO	*Reserved* bits per IEEE 802.3ae clause 45. Ignore when read.

BASE10G_DEVICE_IDENTIFIER.

This register is unused.

TABLE 27-139 XPCS Device Identifier – BASE10G_DEVICE_IDENTIFIER (FZC_MAC + 02010_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	*unused*	00_{16}	RO	Provides a 32-bit manufacturer's ID. Ignore when read. Returns 0 on all bits.

BASE10G_SPEED_ABILITY.

Only bit 0 provides any meaningful data. Bit 0 is set to 1 to advertise 10-Gbyte capability.

TABLE 27-140 XPCS Speed Ability – BASE10G_SPEED_ABILITY (FZC_MAC + 02018_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:1	—	00_{16}	RO	Always 0. *Reserved* bit per IEEE 802.3ae clause 45 for future speeds.
0	csr_10gig	1	RO	1 = Capable of 10 Gb/s operation.

. .

BASE10G_DEVICES_IN_PACKAGE.

This register advertises XPCS programmable resources available in XPCS. In addition to Clause 45-compliant programmable resources, XPCS has a CSR_VENDOR1 (configuration register) and CSR_VENDOR2 (diagnostic register) available. These supplemental programmable resources are described later in this chapter.

TABLE 27-141 XPCS On-Chip Devices – BASE10G_DEVICES_IN_PACKAGE (FZC_MAC + 02020_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31	csr_vendor2	1	RO	1 = Vendor 2-specific device present on-chip.
30	csr_vendor1	1	RO	1 = Vendor 1-specific device present on-chip.
29:16	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
15:6	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
5	dte_xs	0	RO	1 = DTE XS present on-chip.
4	phy_xs	0	RO	0 = PHY XS not present on-chip.
3	pcs	1	RO	1 = PCS present on-chip.
2	wis	0	RO	0 = WIS not present on-chip.
1	pmd_pma	0	RO	0 = PMA/PMD not present on-chip.
0	clause_22_reg	0	RO	0 = Clause 22 registers are not present on-chip

BASE10G_CONTROL2.

All the bits in the Control 2 register are read-only, and writes to this register have no effect Only bit 0 is meaningful in this register: Type of PCS ability is 8b/10b encoding.

TABLE 27-142 XPCS Control 2 – BASE10G_CONTROL2 (FZC_MAC + 02028_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:2	—	0	RO	*Reserved/unused*. Ignore when read.
1:0	csr_psc_selection	1_{16}	RO	Reads 01 to indicate 10GBASE-X PCS type.

BASE10G_STATUS2.

All the bits in Status 2 register are read-only, and writes to this register have no effect. Bits 10 and 11 advertise remote (receive) or local (transmit) faults. Bit 1 indicates 10-bit encoding

TABLE 27-143 XPCS Status 2 – BASE10G_STATUS2 (FZC_MAC + 02030_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:14	csr_device_present	0_{16}	RO	*Unused*. No configuration device is present. Ignore when read.
13:12	—	0_{16}	RO	*Reserved/unused*. Ignore when read.
11	csr_transmit_fault	0	RO	This bit is set to "1" when csr_receive_link_ status bit is "0" or xMAC sends remote fault to XPCS. Latches high until read. 0 = No fault condition on transmit path; 1 = Fault condition on transmit path.
10	csr_receive_fault	0	RO	This bit is set to "1" when XPCS received local_fault order set. Latches high until read. 0 = No fault condition on receive path; 1 = Fault condition on receive path.
9:3	—	0_{16}	RO	*Reserved/unused*. Ignore when read.
2	ten_gbase_w	0	RO	Reads 0 to indicate that it is not able to support 10G-BASE-W.
1	ten_gbase_x	1	RO	Reads 1 to indicate that it is able to support 10G-BASE-X.
0	ten_gbase_r	0	RO	Reads 0 to indicate that it is not able to support 10G-BASE-R.

BASE10G_PACKAGE_IDENTIFIER.

XPCS is an IP part of a larger device. This package identifier field is unused.

TABLE 27-144 XPCS Package Identifier – BASE10G_PACKAGE_IDENTIFIER (FZC_MAC + 02038_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	*unused*	00_{16}	RO	Ignore when read. Read as 0.

BASE10G_STATUS.

All bits in the XPCS Status register are read-only. Writes to this register have no effect. Bits 3 through 0 indicate status of synchronization. Bit 12 indicates status of lane alignment. Bit 11 advertises capability for test pattern generation.

TABLE 27-145 XPCS Status – BASE10G_STATUS (FZC_MAC + 02040) (count 2 step 06000_{16})

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:13	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
12	csr_lane_align	0	RO	0 = 10GBASE-X PCS receive lanes not aligned. 1 = 10GBASE-X PCS receive lanes aligned.
11	csr_pattern_test_able	1	RO	0 = 10GBASE-X PCS is not able to generate patterns. 1 = 10GBASE-X PCS is able to generate patterns.
10:4	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
3	csr_lane3_sync	0	RO	0 = Lane 3 is not synchronized; 1 = Lane 3 is synchronized.
2	csr_lane2_sync	0	RO	0 = Lane 2 is not synchronized; 1 = Lane 2 is synchronized.
1	csr_lane1_sync	0	RO	0 = Lane 1 is not synchronized; 1 = Lane 1 is synchronized.
0	csr_lane0_sync	0	RO	0 = Lane 0 is not synchronized; 1 = Lane 0 is synchronized.

BASE10G_TEST_CONTROL.

Transmit test pattern for support of TX jitter measurement. IEEE 802.3ae Clause 48 Annex 48A.

TABLE 27-146 XPCS Jitter Test Control – BASE10G_TEST_CONTROL (FZC_MAC + 02048_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:3	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
2	csr_tx_test_enable	0	RW	0 = Tx test pattern not enabled; 1 = Tx test pattern enabled.
1:0	csr_test_pattern_sel	0_{16}	RW	11 = *Reserved*; 10 = Mixed frequency; 01 = Low frequency; 00 = High frequency.

BASE10G_CONFIGURATION.

Programmable features useful for debugging. Bits 6 through 3 select the debug mux to choose which signals to observe in the external debug port.

TABLE 27-147 XPCS Test Config. – BASE10G_CFG_VENDOR1 (FZC_MAC + 02050_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:8	—	00_{16}	RO	*Reserved/unused*. Ignore when read.
7	csr_vendor_debug_io_test	00_{16}	RW	This bit select the following debug signals. assign rx0_debug = csr_vendor_debug_io_test ? {debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1, debug_io_test_0, debug_io_test_1} : {rx_symbols_dft_dummy, 2'b0, rx_xgmii_byte_0, rx_xgmii_special_0, rx_byte_0, rx_special_0, rx_error_0, xrx_code_group_muxout_dbg[9:0]};
6:3	csr_vendor_debug_sel	0_{16}	RW	Debug diagnostic select. 0_{16}: xpcs_debug = rx0_debug; 1_{16}: xpcs_debug = rx1_debug; 2_{16}: xpcs_debug = rx2_debug; 3_{16}: xpcs_debug = rx3_debug; 4_{16}: xpcs_debug = tx0_debug; 5_{16}: xpcs_debug = tx1_debug; 6_{16}: xpcs_debug = tx2_debug; 7_{16}: xpcs_debug = tx3_debug; F_{16}: xpcs_debug = training_vector; default: xpcs_debug = rx0_debug;

TABLE 27-147 XPCS Test Config. – BASE10G_CFG_VENDOR1 (FZC_MAC + 02050_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
2	csr_bypass_signal_ detect	0	RW	Disable signal detect from SerDes. Override of signal detect.
1	csr_enable_tx_ buffers	1	RW	Enables transmit to TX SerDes port. 0 = Disabled. 1 = Enabled (default).
0	csr_xpcs_enable	1	RW	Enables/disables Xpcs RX/TX operation. 0 = Disabled. 1 = Enabled (default).

BASE10G_DIAGNOSTIC.

Debug read-only register. Bit 0 shows the state of the receive state machine. Bits 8 through 1 show the state of the elastic buffer. Bits 24 through 9 show the state of the synchronization state machine. This register provides error and state machine information.

TABLE 27-148 XPCS Diag. – BASE10G_DIAGNOSTIC_VENDOR2 (FZC_MAC + 02058_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:25	—	00_{16}	RO	Ignore when read. *Reserved/unused* bits. All bits are 0.
24:21	sync_sm_lane_3	0_{16}	RO	loss_sync = 0_{16}; in_sync = 1_{16}; one_invalid = 2_{16}; two_invalid = 4_{16}; three_invalid = 8_{16};
20:17	sync_sm_lane_2	0_{16}	RO	loss_sync = 0_{16}; in_sync = 1_{16}; one_invalid = 2_{16}; two_invalid = 4_{16}; three_invalid = 8_{16};
16:13	sync_sm_lane_1	0_{16}	RO	loss_sync = 40_{16}; in_sync = 1_{16}; one_invalid = 2_{16}; two_invalid = 4_{16}; three_invalid = 8_{16};
12:9	sync_sm_lane_0	0_{16}	RO	loss_sync = ,0_{16}; in_sync = 1_{16}; one_invalid = 2_{16}; two_invalid = 4_{16}; three_invalid = 0008_{16};
8:1	elastic_buffer_sm	00_{16}	RO	deskew_loss = 00_{16}; align_det1 = 01_{16}; align_det2 = 02_{16}; align_det3 = 04_{16}; deskew_ok = 08_{16}; align_err1 = 10_{16}; align_err2 = 20_{16}; align_err3 = $840;_{16}$
0	receive_state_machine	0	RO	1 = Receive state; 0 = Fault state.

BASE10G_MASK1.

This register has the mask bit for corresponding status bits in the TABLE 27-138, XPCS Status 1 – base10g_status1 (fzc_mac + 0200816) (count 2 step 06000), on page 843.

TABLE 27-149 XPCS Mask 1 – BASE10G_MASK1 (FZC_MAC + 02060_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15:8	—	00_{16}	RO	Always 0, ignore read/write.
7	csr_fault_mask	1	RW	0 = Do not mask interrupt from csr_fault; 1 = Mask interrupt from csr_fault.
6:3	—	0_{16}	RO	Always 0, ignore read/write.
2	csr_receive_align_link_status_mask	1	RW	0 = Do not mask interrupt generation of csr_receive_lalign_ink_status; 1 = Mask interrupt generation of csr_receive_align_link_status.
1:0	—	0_{16}	RO	Always 0, ignore read/write.

BASE10G_PACKET_COUNTER.

Raw count of received and transmitted packets. Useful for debug of loopback mode. Featured was requested by diagnostics group.

TABLE 27-150 XPCS Packet Counter – BASE10G_PACKET_COUNTER (FZC_MAC + 02068_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	csr_tx_pkt_count	00_{16}	RW	Count of packets transmitted.
15:0	csr_rx_pkt_count	00_{16}	RW	Count of packets received.

BASE10G_TX_STATE_MACHINE.

Read only register providing the state bits of the XPCS transmit state machine.

TABLE 27-151 XPCS Transmit State Machine – BASE10G_TX_STATE_MACHINE (FZC_MAC + 02070_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:4	—	00_{16}	RO	Always 0, ignore read/write.
3:0	csr_tx_state	0_{16}	RO	SEND_DATA = 0_{16}; SEND_SDP = 1_{16}; SEND_A = 2_{16}; SEND_K = 3_{16}; SEND_Q = 4_{16}; SEND_RANDOM_R = 5_{16}; SEND_RANDOM_A = 6_{16}; SEND_RANDOM_K = 7_{16}; SEND_RANDOM_Q = 8_{16}; UNDERRUN = 9_{16};

BASE10G_DESKEW_ERROR_COUNTER.

This counter keeps track of the total number of alignment errors. Software must write to this register to clear and read for poll.

TABLE 27-152 XPCS Deskew Error Counter – BASE10G_DESKEW_ERROR_COUNTER (FZC_MAC + 02078_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:8	—	0	RO	*Reserved*
7:0	deskew_error_count	00_{16}	RW	Count of deskew errors.

BASE10G_SYMBOL_ERROR_COUNTER_LANES_0_AND_1 .

This counter keeps track of symbol errors for Lanes 0 and 1. Software must write to this register to clear and read for poll.

TABLE 27-153 XPCS Symbol Error Counter 01 – BASE10G_SYMBOL_ERROR_COUNTER_01 (FZC_MAC + 02080_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	symbol_error_count_1	00_{16}	RO	Symbol error count on lane 1. Multiply the value of this counter by 16 to obtain the total number of symbol errors for this lane. When the value of symbol errors reaches $FFFF_{16}$, the counter freezes at this max value until read.
15:0	symbol_error_count_0	00_{16}	RO	Symbol error count on lane 0. Multiply the value of this counter by 16 to obtain the total number of symbol errors for this lane. When the value of symbol errors reaches $FFFF_{16}$, the counter freezes at this max value until read.

BASE10G_SYMBOL_ERROR_COUNTER_LANES_2_AND_3.

This counter keeps track of symbol errors in lanes 2 and 3. Software must write to this register to clear and read for poll.

TABLE 27-154 XPCS Symbol Error Counter 23 – BASE10G_SYMBOL_ERROR_COUNTER_23 (FZC_MAC + 02088_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	symbol_error_count_3	00_{16}	RO	Symbol error count on lane 3. Multiply the value of this counter by 16 to obtain the total number of symbol errors for this lane. When the value of symbol errors reaches $FFFF_{16}$, the counter freezes at this max value until read.
15:0	symbol_error_count_2	00_{16}	RO	Symbol error count on lane 2. Multiply the value of this counter by 16 to obtain the total number of symbol errors for this lane. When the value of symbol errors reaches $FFFF_{16}$, the counter freezes at this max value until read.

TABLE 27-155 XPCS Training Vector - XPCS_TRAINING_VECTOR (fzc_mac + 0209016) (count 2 step 0x06000)

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:0	training_vector	00_{16}	RW	Use training vector to calibrate OSC. Load this register with training vector value. If (csr_vendor_debug_sel{3:0} = F_{16}), the training vector starts toggling, else it is a static value.

27.7.2 Programming Guide

27.7.2.1 Initialization Programming Sequence

Per IEEE 802.3ae spec, the XPCS will default to a 10-Gigabit operational mode as reset default. To disable the XPCS, reset the xpcs_enable bit shown in TABLE 27-147, XPCS Test Config. – base10g_cfg_vendor1 (fzc_mac + 0205016) (count 2 step 06000), on page 847.

27.7.2.2 XPCS Link Loss Notification

The XPCS generates a maskable interrupt to the system whenever there is a link status change: link up or down. Upon an interrupt, the software should read the X_10GBASE_STATUS1 and X_10GBASE_STATUS2 registers (See TABLE 27-138, XPCS Status 1 – base10g_status1 (fzc_mac + 0200816) (count 2 step 06000), on page 843 and TABLE 27-143, XPCS Status 2 – base10g_status2 (fzc_mac + 0203016) (count 2 step 06000), on page 845).

27.7.2.3 XPCS SerDes Operational Modes

TABLE 27-156 indicates the most important signals that define the SerDes modes of operation. The bits involved are

- xpcs_enable of the X_10GBASE_CONFIGURATION register
- test_contro of the X_10GBASE_TEST_CONTROL register
- low_power of the X_10GBASE_TEST_CONTROL register
- loopback of the X_10GBASE_TEST_CONTROL register

TABLE 27-156 XPCS Configurations

	xpcs_enable	test_control	low_power	loopback
3.12 Gb/s	1	0	0	0
Test	X	1	X	X
Sleep	1	0	1	0
Loopback	1	0	0	1

27.8 PCS Programmable Resources

Physical Coding Sublayer (PCS) MII Control Register . This register is equivalent to the Control register of the standard MII management register set. Unlike MII management registers, this register is directly mapped in UltraSPARC T2's register space.

TABLE 27-157 PCS MII Control – PCS_MII_CTL (FZC_MAC + 04000_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	reset	0	RW-AC	Resets the PCS when set. Self-clearing when reset process is complete.
14	—	0	RO	*Reserved*
13	10_100_speed_selection	0	RO	Reads back as zero. Ignored on writes.
12	auto-negotiation_enable	1	RW	When set, the PCS goes through an automatic link configuration phase before it can be used. When clear, the link can be used without any link configuration phase. Defaults high.
11	power_down	0	RO	Reads back as zero. Ignored on writes.

TABLE 27-157 PCS MII Control – PCS_MII_CTL (FZC_MAC + 04000_{16}) (count 2 step 06000) *(Continued)*

Bit	Field	Initial Value	R/W	Description
10	isolate	0	RO	Reads back as zero. Ignored on writes.
9	restart_auto-negotiation	0	RW-AC	Set to restart auto-negotiation.
8	duplex_mode	0	RO	Forced low. PCS behavior is similar for half and full duplex.
7	collision_test	0	RW	When set, the COL signal at the PCS-to-MAC interface is activated regardless of activity on the receive inputs.
6	1000mb_speed_select	1	RO	Reads back 1. Ignored on writes.
5:0	—	0	RO	*Reserved*

Programming Note The auto-negotiation_enable bit should be programmed the same at the link partner as in the local device. Otherwise, auto-negotiation will not complete. When this bit is reprogrammed, auto-negotiation/manual configuration is restarted automatically.

PCS MII Status Register (RO). This register is equivalent to the Status register of the standard MII management register set. Unlike MII management registers, this register is directly mapped in UltraSPARC T2's register space.

TABLE 27-158 PCS MII Status – PCS_MII_STAT (FZC_MAC + 04008_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:9	—	0	RO	*Reserved*
8	extended_status	1	RO	Can indicate that this is a 1000 Base-X PHY. Reads back 1. Writes to it are ignored.
7:6	—	0	RO	*Reserved*
5	auto-negotiation_complete	X	RO	0 = Auto-negotiation not complete. 1 = auto-negotiation complete.
4	remote_fault	X	RO	When set, indicates that a remote fault has been detected from the received link code word. Only valid after completion of auto-negotiation.
3	auto-negotiation_ability	1	RO	Always reads 1. Can perform auto-negotiation.
2	link_status	X	RO	Incorporates a latch on 0, such that the 0 is kept until read. The register may be read twice to determine if the link has gone up again. 0 = Link is down; 1 = Link is up.
1	jabber_detect	0	RO	Always reads 0.
0	extended_capability	0	RW	Always reads 0.

PCS MII Advertisement Register . This register is equivalent to the Advertisement register of the standard MII management register set. Unlike MII management registers, this register is directly mapped in UltraSPARC T2's register space. The contents of this register are used during auto-negotiation.

TABLE 27-159 PCS MII Advertisement – PCS_MII_ADVER (FZC_MAC + 04010_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	next_page	0	RO	Forced low.
14	ack	0	RO	Used by the auto-negotiation function to indicate that a device has successfully received its link partners base or next page. Set to logic 1 after the device has successfully received at least three consecutive and matching RX_CONFIG_REG[D15:D0] values (ignoring the Acknowledge bit value), and, for next page exchanges, remains set until the next page information has been loaded into the AN Next Page Transmit register (register 7). After the auto-negotiation process COMPLETE_ACKNOWLEDGE state has been entered, the TX_CONFIG_REG[D15:D0] value is transmitted for the link_timer duration.
13:12	remote_fault	0_{16}	RW	Provides simple fault information to the link partner. Bit 13 is writable by software to optionally indicate to its link partner that it is going off-line. Bit 12 will get set when signal_detect equals FAIL and will remain set until successful negotiation, at which time it will be set to 0.
11:9	—	0	RO	*Reserved*
8	asm_dir	0	RW	Advertises device's capability to perform PAUSE asymmetric to its link partner.
7	pause	1	RW	Advertises device's capability to perform PAUSE symmetrical to its link partner.
6	half_duplex	1	RW	Advertises device's capability to perform half duplex 1000Base-X.
5	full_duplex	1	RW	Advertises device's capability to perform full duplex 1000Base-X.
4:0	—	0	RO	*Reserved*

PCS MII Link Partner Ability Register .

This register is equivalent to the Link Partner Ability Register of the standard MII management register set. The contents of this register are updated as a result of auto-negotiation. The register layout and bit definitions correspond to the PCS MII Advertisement register. Please refer to IEEE 802.3z Clause 37.2.5.1.4 AN link partner ability base page register (register5) for detail.

TABLE 27-160 PCS MII Partner Ability – PCS_MII_PARTNER (FZC_MAC + 04018_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:16	—	0	RO	*Reserved*
15	next_page	X	RO	See IEEE 802.3z Clause 37.2.1.7
14	acknowledge	X	RO	See IEEE 802.3z Clause 37.2.1.6.
13:12	remote_fault	X	RO	See IEEE 802.3z Clause 37.2.1.5.
11:9	—	X	RO	Ignore on read.
8:7	pause	X	RO	See IEEE 802.3z Clause 37.2.1.4.
6	half_duplex	X	RO	See IEEE 802.3z Clause 37.2.1.3.
5	full_duplex	X	RO	See IEEE 802.3z Clause 37.2.1.2.
4:0	—	X	RO	Ignore on read.

PCS Configuration Register .

TABLE 27-161 PCS Configuration – PCS_CONF (FZC_MAC + 04020_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
6	mask	1	RW	PCS global mask bit. 0 = Enable all PCS interrupts; 1 = Mask/disable all PCS interrupts (default value).
5	10ms_timer_override	0	RW	Shortens the 10-20 ms timer used in auto-negotiation and synchronization to a few cycles. The purpose is to shorten ASIC simulation time and vector generation. This should be programmed to 0 for interoperability with other devices.
4:3	jitter_study	0_{16}	RW	Allows detailed jitter measurements to be made to characterize optic parts. A single code group is transmitted repeatedly. Disparity rules are followed. 0_{16} = normal operation; 1_{16} = high frequency test pattern, D21.5; 2_{16} = low frequency test pattern, K28.7; 3_{16} = *Reserved*
2	signal_detect_active_low	0	RW	When set, changes the interpretation of the incoming optic signal so that when the signal is low, signal_detect is OK.
1	signal_detect_override	0	RW	Sets signal_detect internally to OK in case the optic is not able to generate a reliable signal. This bit is non-resetable so that it can mimic a pull-up or pull-down on the board.
0	enable	0	RW	Enables the PCS functions. Must be zero while modifying the PCS MII Advertisement Register.

PCS State Machine Register (RO). This register is mainly for diagnostic purposes. It exposes PCS internal state machine status.

TABLE 27-162 PCS State Machines – PCS_STATE (FZC_MAC + 04028_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:30	—	0	RO	*Reserved*
29	Link not up because partner fails to send idle	0	RO	Set to 1 when at the end of auto-negotiation and the Link Partner continues to send C codes instead of idle symbols or packet data.
28	Link not up because waiting for C codes with ack bits set	0	RO	Set to 1 when waiting for C codes with acknowledge bit set indicating Link partner has received our C codes.
27	Link not up due to loss of sync	0	RO	Set to 1 when word sync has not been achieved. This can occur due to bit errors, receipt of illegal codes or comma characters clocking on RBC0.
26	Link not up due to lack of good C codes	0	RO	Set to 1 when the contents of the C codes is not stable or C codes are not being received.
25	Link not up due to SerDes	0	RO	Set to 1 when the SerDes is being initialized. See SerDes State register. It is possible that the SerDes is not providing SYNC_DET signals indicating that it has achieved byte synchronization.
24	Link not up due to receipt of breaklink C codes from partner	0	RO	Set to 1 when C codes with contents of zero are received triggering start or restart of auto-negotiation. They should be sent for no longer than 20 ms.
23	—	0	RO	*Reserved*
22	Loss of signal_detect	0	R-AC	Set to 1 if signal_detect goes from OK to FAIL. This will cause loss of link and loss of sync. If this bit is set, bit 29 will also be set.
21	Loss of link due to loss of sync	0	R-AC	Set to 1 if loss of sync caused loss of link.
20	Loss of link due to C codes	0	R-AC	Set to 1 if receipt of Configuration codes from the link partner caused loss of link.
19:17	—	0	RO	*Reserved*
16:13	Link configuration state	0	RO	State of B_{16} indicates link is up.
12:11	Sequence detection state	0	RO	Cycling through states 0–3 indicates reception of Config codes. Cycling through states 0–1 indicates reception of idles.
10:8	Word synchronization state	0	RO	State 0 indicates loss of sync. Otherwise, sync has been acquired.
7:4	—	0	RO	State 0 indicates reception of idle. Otherwise, it indicates reception of a packet.
3:0	Link not up because partner fails to send idle	0	RO	States 0 and 1 indicate transmission of idle. Otherwise, it indicates transmission of a packet.

PCS Interrupt Status Register. This register indicates interrupt changes in specific PCS MII status bits. Presently only one bit is implemented, reflecting transitions on the link status. The rationale for a single bit is that all other relevant bits cannot change without altering the link state. There is no mask register at this level. The pcs_int bit may be masked at the Interrupt Status register level.

TABLE 27-163 PCS Interrupt Status – PCS_INTERRUPT (FZC_MAC + 04030_{16}) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:3	—	0	RO	*Reserved*
2	link_status_change	0	R-AC	When set, indicates the Link Status has changed at least once since this register was read.
1:0	—	0	RO	*Reserved*

Datapath Mode Register . This register controls which network interface is used and only working with big MAC ports. For xMAC port, the bypass_pcs function comes from xMAC configuration register bit 30.

TABLE 27-164 PCS Datapath Mode – PCS_DPATH_MODE (FZC_MAC + $040A0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1	phy_mode	0	RW	PCS bypass. 0 = Internal PCS is used. 1 = PCS is bypassed. gmii/rgmii interface is used. This bit is used and only working with big MAC port. For xMAC port, the bypass_pcs function comes from xMAC Configuration register bit 30.
0	link_up_filter_en	0	RW	This filter is to guarantee that at least 1K valid receive symbols have been received before PCS asserts link-up. 0 = Disable link up filter; 1 = Enable link up filter.

PCS Packet Counter (RO). This register indicates the number of packets transmitted or received. Used for diagnostic purposes, it may be ignored by normal software drivers. The counters roll over without generating an interrupt when they reach their terminal count.

TABLE 27-165 PCS Packet Counter – PCS_PKT_CNT (FZC_MAC + $040C0_{16}$) (count 2 step 06000)

Bit	Field	Initial Value	R/W	Description
63:27	—	0	RO	*Reserved*
26:16	rx_pkt_cnt	00_{16}	RO	Number of packets received by the PCS whether they encountered an error or not. Each received preamble is counted as one packet.
15:11	—	0	RO	*Reserved*
10:0	tx_pkt_cnt	00_{16}	RO	Number of packets transmitted by the PCS. Each SFD is counted as one packet.

27.9 MIF Programmable Resources

The MDIO controller (MIF module) is the Station Management (STA)/MDIO Master that is used to access internal (SerDes) as well as external MDIO Manageable Device (MMD)/MDIO slave, for example, off-chip RGMII copper phy, 10/1G optical/ copper phy, etc.

Use clause 45 frame for internal SerDes. Use clause 22 frame only when clause 45 frame is not supported by external MMD.

27.9.1 Bit-Bang Mode Theory of Operation

The Bit-Bang mode provides software a way to directly control MDC and MDIO signals. There is very limited hardware assist involved. Software should use this mode as the last resort.

MIF Bit-Bang Clock . This 1-bit register generates the MDC clock waveform on the MII Management Interface when the MIF is programmed in the Bit-Bang mode. Writing 1 after 0 into this register will create a rising edge on the MDC, while writing 0 after 1 will create a falling edge. For every bit that is transferred on the management interface, both edges must be generated.

TABLE 27-166 MIF Bit-Bang Clock – MIF_BB_MDC (FZC_MAC + 16000_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	bb_mdc	0	RW	Toggles the clock using PIOs.

MIF Bit-Bang Data . This 1-bit register is used to generate the outgoing data (MDO) on the MDIO Management Interface when the MIF is programmed in the Bit-Bang mode.

For incoming Bit-Bang mode read data, please refer to bits 10, 13, 14, 15 of Table 27-174, "MIF State Machine – mif_state_machine (fzc_mac + 1603816)," on page 863.

TABLE 27-167 MIF Bit-Bang Output Data – MIF_BB_MDO (FZC_MAC + 16008_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	bb_mdo	0	RW	Toggles the data using PIOs.

MIF Bit-Bang Output Enable. This 1-bit register enables (1) and disables (0) the bidirectional driver on the MII Management interface when the MIF is programmed in Bit-Bang mode. The MDIO should be enabled when data bits are transferred from the MIF to the transceiver, and it should be disabled when the interface is idle or when data bits are transferred from the transceiver to the MIF (data portion of a read instruction).

TABLE 27-168 MIF Bit-Bang Output Enable – MIF_BB_MDO_EN (FZC_MAC + 16010_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	0	RO	*Reserved*
0	bb_mdo_en	0	RW	0 = Disables the output driver. 1 = Enables the output driver.

27.9.2 Frame Mode Theory of Operation

Frame mode provides the most automated hardware assist function that software can use by simply loading valid IEEE 802.3 CLAUSE 22 or CLAUSE 45 MDIO frame into TABLE 27-170, MIF Frame/Output – mif_frame_output_reg (fzc_mac + 1601816), on page 860. This includes the mixture of Clause 45 and Clause 22 frame with different port address and device address.

Note | Each CLAUSE 45 write/read operation consists of "ADDR_frame + WRITE/READ_frame".

MDIO Frame/Output Register .

This 32-bit register serves as an "Instruction register" when the MIF (Station Management (STA)/MDIO Master) is programmed in the Frame mode. To execute a read/write operation to an MMD (MDIO Manageable Device/MDIO Slave) register, software should load this register with a valid MDIO frame/instruction per IEEE 802.3 CLAUSE 22 or CLAUSE 45 MDIO specification. After issuing an MDIO frame/

instruction, software should poll bit 16 (in the ta field) to be 1 for instruction execution completion status. For a read operation, data field (bit 15:0) contains the 16-bit data that was returned by the MMD.

Frame	Clause 45 MDIO Frame Register Fields					
	st{31:30}	op{29:28}	prtad{27:23}	devad{22:18}	ta{17:16}	address/data{15:0}
Address	00	00	PPPPP	EEEEE	10	AAAAAAAAAAAAAAAA
Write	00	01	PPPPP	EEEEE	10	DDDDDDDDDDDDDDDD
Read	00	11	PPPPP	EEEEE	Z0	DDDDDDDDDDDDDDDD
Post-read-increment-address	00	10	PPPPP	EEEEE	Z0	DDDDDDDDDDDDDDDD

Frame	Clause 22 MDIO Frame Register Fields					
	st{31:30}	op {29:28}	phyad {27:23}	regad {22:18}	ta {17:16}	data{15:0}
Write	01	01	AAAAA	RRRRR	10	DDDDDDDDDDDDDDDD
Read	01	10	AAAAA	RRRRR	Z0	DDDDDDDDDDDDDDDD

See the following table for MDIO slave address assignment.

TABLE 27-169 UltraSPARC T2 XAUI SerDes MDIO Slave Interface Address

Quad SerDes	Port Address	Device Address	Address Offset
hedwig0 (serdes0)	00000_{16}	$0001E_{16}$	8000_{16}
hedwig1 (serdes1)	00001_{16}	$0001E_{16}$	8000_{16}

TABLE 27-170 MIF Frame/Output – MIF_FRAME_OUTPUT_REG (FZC_MAC + 16018_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	frame_msb	0_{16}	RW	Contains the 16 most significant bits of an MDIO frame, as specified in the table above. **Bit 16** (2nd bit of the ta field) is the instruction execution completion status. Software should poll this bit for indication of Frame mode execution completion.
15:0	frame_lsb_output	0_{16}	RW	Contains the 16 least significant bits of an MDIO frame or the result of a read operation, as specified in the table above.

MIF Configuration Register (mixed RW/RO), POR =00_{16}. MIF supports the following different modes to generate the MIF frame.

They are named in order of priority: bb_mode, poll_mode, init_mode, frame_mode.

Port address hex 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F are reserved for internal phy.

TABLE 27-171 MIF Configuration – MIF_CONFIG (FZC_MAC + 16020_{16})

Bit	Field	Initial Value	R/W	Description
63:20	—	0	RO	*Reserved*
19:18	*Used*	0	RW	Register read/write test.
17	*Used*	0	RW	Register read/write test.
16	atca_ge	0_{16}	RW	Advance Telecommunication Architecture Gigabit Ethernet mode (1G mode). When software change the value of this bit, it changes the clock source and may create a glitch in the clock line. Software should follow up with a MAC software reset on each port to clean up the potential "X" in the state machine. 0 = Disable atca_ge. 1 = Enable atca_ge. Port 0 is connected to SerDes 0 lane 0. Port 1 is connected to SerDes 1 lane 0.
15	indirect_mode	1_{16}	RW	Determines if Clause 45 or Clause 22 frame is constructed. It defines the value of the st field. 0 = CLAUSE 22 ST = 01_2; 1 = CLAUSE 45 ST == 00_2 (default)
14:10	poll_prt_phy_addr	0_{16}	RW	Determines the transceiver CLAUSE 45 Port/CLAUSE 22 PHY address to be polled.
9:5	poll_dev_reg_add	00_{16}	RW	Determines the CLAUSE 45 device/CLAUSE 22 register address in the transceiver that will be polled by the polling mechanism in the MIF. It is meaningful only if the poll_enable bit is set to 1.
4	bb_mode	0_{16}	RW	Bit-Bang mode. To enable Bit-Bang mode, set this bit to 1. To use fame mode, reset this bit to 0.
3	poll_en	0_{16}	RW	Poll mode. When set to 1, it enables the polling mechanism. Use this function to check the status bit state changes of the MMD. Poll mode should not be activated when Bit-Bang mode is enabled. poll_en won't take effect until mif_init_done status is set. Poll mode can be enabled/activated along with Frame mode.
2:1	bb_ser_sel	0_{16}	RW	The original function has been removed. Used as spare control bit.
0	manual_mode	0_{16}	RO	The original function has been removed. Used as spare control bit.

27.9.3 Poll Theory of Operation

The mission of poll function is to provide a convenient way for software to examine if there is any bit changed in the target MMD register. Poll mode only supports IEEE standard MMD register (16 bit addressing).

UltraSPARC T2 only supports Clause 22 Poll operation.

MDIO Poll Setup Procedure.

1. For Clause 45 MMD (set indirect_mode bit, which is Config register bit 15, to 1), software sends out address frame (opcode = 00_2) that contains the register address to be polled first. For Clause 22 MMD (reset indirect_mode bit, which is Config register bit 15, to 0), skip this step.
2. Program poll_prt_addr, and poll_dev_addr.
3. Disable poll_mask by setting to 0_2.
4. Set poll_en to 1.
5. Hardware automatically sends out poll frame (read frame) to check the register status.
6. Since the POLL_STATUS register POR value is 32'b0, the first poll-related interrupt will be generated if the read back status is non-zero.
7. If there is any poll status bit changed in the defined register address, the MIF will assert interrupt and Poll mode operation stops automatically. To resume Poll-mode/Frame-mode operation, software should clean up the Poll Status register and MIF Status register.

Poll Status Register . This 32-bit register is used in conjunction with the Poll mode in the MIF. It contains two portions: poll data and poll status. The poll_data field will always contain the latest "image update" of the MMD register that is being polled, and the poll_status field will indicate which bits in the poll_data field have changed since the MIF Status register was last read. The poll_status field is autocleared after being read.

TABLE 27-172 MIF Poll Status – MIF_POLL_STATUS (FZC_MAC + 16028_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:16	poll_data	00_{16}	RO	Contains the latest "image update" of the MMD. Assign new_mif_status{31:16} = reset ? 0_2: (ld_status_reg ? shift_dout{15:0}: mif_status{31:16});
15:0	poll_status	00_{16}	R-AC	Indicates bit-by-bit which bits in poll_data have changed since the last PIO read of this register.

Poll Mask Register . This 16-bit register determines which bits in the poll status portion of the MIF Status register will cause an interrupt. If a mask bit is cleared to 0, the corresponding bit of the poll status will generate the MIF interrupt when set.

TABLE 27-173 MIF Poll Mask – MIF_POLL_MASK (FZC_MAC + 16030_{16})

Bit	Field	Initial Value	R/W	Description
—	63:32	0	RO	*Reserved*
15:0	poll_mask	$FFFF_{16}$	RW	0 = Enable interrupt; 1 = Mask interrupt.

MIF State Machine Register . This 8-bit register provides the current state for all the state machines in the MIF.

TABLE 27-174 MIF State Machine – MIF_STATE_MACHINE (FZC_MAC + 16038_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	0	RO	*Reserved*
31:21		X	RO	{0000_{16}, spc_rdy, spc_req, spc_ack, send_init_fm, mif_init_state{3:0}}.
20:16	port_addr	0	RO	Indicates the port address field that software loaded to the frame register. This bit is useful for Frame mode.
15	—	X	RO	*Reserved.*
14	mdi_1	X	RO	When used in Bit-Bang mode, this bit indicates the incoming bit stream from XAUI SerDes 1 CSR register during a read operation.
13	mdi_0	X	RO	When used in Bit-Bang mode, this bit indicates the incoming bit stream from XAUI SerDes 0 CSR register during a read operation.
12	mdclk	X	RO	Indicates the current MDC clock state. It is useful for Bit-Bang mode.
11	mdo_en	0	RO	Indicates the frame mode MDIO output enable signal. It is useful for Frame mode.
10	mdo	X	RO	Indicates the outgoing bit stream to MDIO interface. This bit is useful for Bit-Bang mode.
9	mdi	1	RO	When used in Bit-Bang mode, this bit indicates the incoming bit stream from external PHY during a read operation.
8:6	ctl_state	0_{16}	RO	Control state machine state.
5:0	ex_state	00_{16}	RO	Execution state machine state.

MIF Status Register (RAC). When any one of the MIF status bits is set and not masked (interrupt enable), an interrupt request will be generated. Frame mode and Poll mode are frozen until the status is read (RAC) by software.

TABLE 27-175 MIF Status – MIF_STATUS (FZC_MAC + 16040_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	X	RO	*Reserved*
5	mdint1	X	RO	This bit should not be used. This bit is replaced by port 1 RxMAC status register bit 16. Please refer to TABLE 27-8 on page 795.
4	mdint0	X	RO	This bit should not be used. This bit is replaced by port 0 RxMAC status register bit 16. Please refer to TABLE 27-8 on page 795.
3	—	0	R-AC	*Reserved. .*
2	—	0	R-AC	*Reserved. .*
1	—	0	R-AC	*Reserved.*
0	—	0	RO	*Reserved.*

MIF Mask Register . This register is not used in UltraSPARC T2.

TABLE 27-176 MIF Mask – MIF_MASK (FZC_MAC + 16048_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5		1	RW	.
4		1	RW	.
3		1	RW	.
2		1	RW	.
1		1	RW	.
0		1	RW	.

27.9.4 MIF Operation Examples

This section gives two examples of performing a SerDes or transceiver register access: one example on write, the other example on write:

Example 1: Write to a SerDes/transceiver register via MIF per CLAUSE 45:

1. Write to MIF Frame/Output register:

```
frame.bits.st = 0;            /* Use Clause 45 */
frame.bits.op = 0;            /* Setup register address */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
frame.bits.data = reg_addr;   /* 16-bit register address */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          set up address*/
```

```
frame.bits.op = 0x1;          /* Write cmd */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.a_msb = 1;         /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
frame.bits.data = value;      /* 16-bit payload value */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          write data */
```

2. Poll for command complete:

```
do { /* Poll for cmd complete */
  PIO_RD(MIF_FRAMEOUT_REG, frame.value);
while (frame.bits.w0.ta_lsh == 0);
/* Cmd complete when ta_lsh is set to 1 */
```

Example 2: Read from a SerDes/transceiver register via MIF per CLAUSE 45:

1. Write to MIF Frame/Output Register:

```
frame.bits.st = 0;            /* Use Clause 45 */
frame.bits.op = 0;           /* Setup register address */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          set up address*/
frame.bits.st = 0;           /* Use Clause 45 */
frame.bits.op = 0x2;          /* Read cmd */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          read data */
```

2. Poll for command complete:

```
do { /* Poll for cmd complete */
  PIO_RD(MIF_FRAMEOUT_REG, frame.value);
while (frame.bits.w0.ta_lsh == 0);
```

3. Obtain result from 16-bit data field of MIF Frame/Output register:

```
value = frame.bits.w0.data;
```

27.10 ESR Control Programmable Resources

27.10.1 Common NIU Registers

Ethernet SerDes Reset Register .

TABLE 27-177 Ethernet SerDes Reset – ENET_SERDES_RESET (FZC_MAC + 14000_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	0	RO	*Reserved*
1	serdes_reset_1	0	RW	Software reset for serdes1. 0 = Deassert reset; 1 = Assert reset.
0	serdes_reset_0	0	RW	Software reset for serdes0. 0 = Deassert reset; 1 = Assert reset.

Ethernet SerDes Configuration Register .

TABLE 27-178 Ethernet SerDes Configuration – ENET_SERDES_CFG (FZC_MAC + 14008_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	blunt_end_loopback	0	RW	Blunt_end_loopback (a.k.a. receive to transmit loopback). In UltraSPARC T2 mode, the SerDes receive data are latched with 10G receive clock and looped back to transmit side SerDes directly. It is only for 10G *not* 1G.
0	force_serdes_rdy	0	RW	0 = Normal operation. 1 = Force SerDes-ready signal.

---------- UltraSPARC T2 SerDes Blunt-End Loopback Test Procedure ----------

1. Tester does a cold reset on UltraSPARC T2 or a niu_reset (warm reset) on niu/MAC/SerDes block.
2. Through JTAG, tester needs to program proper XAUI SerDes Control register to enable each channel and PLL and disable SYNC realignment, etc. Specific programming guide will be provided later.
3. Tester continuously supplies an ID (K28.5 code-group) data to receive side XAUI SerDes.
4. Hardware monitors the PLL status bit stsplli{0} (lock). When lock is 0 (not locked yet), hardware will send all 0 data to tester through SerDes transmit interface.
5. Once XAUI SerDes PLL is locked (stsplli{0} = 1), hardware will connect the receive-side SerDes data bus back to transmit-side SerDes data bus.

Tester will be able to observe any special code-group (ID) sent by Tester on the SerDes transmit-side interface (TX_P/N).

6. When tester observed at least 20 consecutive special ID code-groups coming out of SerDes transmit-side interface (TX_P/N), it can start pumping in real test vectors and do the pattern comparison.

THE END!

P.S. K28.5 can be used as an ID for tester to recognize that PLL is up and running and ready for the real test vector. ID can be any special pattern that is recognizable by tester.

Internal Signals Observation Register.

TABLE 27-179 ESR Internal Signal – ESR_INTERNAL_SIGNALS (FZC_MAC + 14800_{16})

Bit	Field	Initial Value	R/W	Description
63:32	—	X	RO	*Reserved*
31:0	Internal signals	0_{16}	RO	UltraSPARC T2 internal signals = {lock_0, lock_1, serdes_rdy0_0, signal_detect0_0, serdes_rdy0_1, signal_detect0_1, xserdes_rdy_0, xsignal_detect_0{3:0}, xserdes_rdy_1, xsignal_detect_1{3:0}, 8 bits 0, los3_1, los2_1, los1_1, los0_1, los3_0, los2_0, los1_0, los0_0};

Debug Selection Register.

TABLE 27-180 ESR Debug Selection – ESR_DEBUG_SEL (FZC_MAC + 14808_{16})

Bit	Field	Initial Value	R/W	Description
63:6	—	0	RO	*Reserved*
5:0	debug_sel	0_{16}	RO	Select the signals that will be visible at the debug-port pins, following the algorithm specified below.

27.11 MAC Register Maps

MAC occupies 128-Kbyte address space. It requires 17 address bits to address 128-Kbyte address space. These 128 Kbytes are divided into 16 address blocks by using address bits 16:13. Each address block occupies 8-Kbytes address space.

The upper 4 bits, 16:13, are assigned to each MAC sub-block.

All the upper 32 bits (bits 63:32) are reserved.

There are three 16-bit registers used to construct a 48-bit MAC address.

The association between MAC address and MAC host info are documented in the register name column and on the first of the three 16-bit MAC address registers.

27.11.1 xMAC Register Map

There are 120 registers in xMAC.

TABLE 27-181 xMAC Registers *(1 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
000_{16}	XTXMAC_SW_RST (side effect)	
008_{16}	XRXMAC_SW_RST (side effect)	
020_{16}	XTXMAC_STATUS	
028_{16}	XRXMAC_STATUS	
030_{16}	XMAC_FC_STAT	
040_{16}	XTXMAC_STAT_MSK	
048_{16}	XRXMAC_STAT_MSK	
050_{16}	XMAC_FC_MSK	
060_{16}	XMAC_CONFIG	Must be initialized
080_{16}	XMAC_IPG	
088_{16}	XMAC_MIN	
090_{16}	XMAC_MAX	
100_{16}	RXMAC_BT_CNT (rx byte count)	
108_{16}	RXMAC_BC_FRM_CNT	Must perform RAC
110_{16}	RXMAC_MC_FRM_CNT	Must perform RAC
118_{16}	RXMAC_FRAG_CNT	Must perform RAC
120_{16}	RXMAC_HIST_CNT1	Must perform RAC
128_{16}	RXMAC_HIST_CNT2	Must perform RAC
130_{16}	RXMAC_HIST_CNT3	Must perform RAC
138_{16}	RXMAC_HIST_CNT4	Must perform RAC

TABLE 27-181 xMAC Registers *(2 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
140_{16}	RXMAC_HIST_CNT5	Must perform RAC
148_{16}	RXMAC_HIST_CNT6	Must perform RAC
150_{16}	RXMAC_MPSZER_CNT	Must perform RAC
158_{16}	MAC_CRC_ER_CNT	Must perform RAC
160_{16}	MAC_CD_VIO_CNT	Must perform RAC
168_{16}	MAC_AL_ER_CNT	Must perform RAC
170_{16}	TXMAC_FRM_CNT	
178_{16}	TXMAC_BYTE_CNT	
180_{16}	LINK_FAULT_CNT	Must perform RAC
188_{16}	RXMAC_HIST_CNT7	Must perform RAC
$1A8_{16}$	XMAC_SM_REG	
$1B0_{16}$	XMAC_INTERN1	
$1B8_{16}$	XMAC_INTERN2	
$0A0_{16}$	XMAC_ADDR0 (MAC_UNIQUE_ADDR_0) → MAC_HOST_INFO17	Must be initialized
$0A8_{16}$	XMAC_ADDR1 (MAC_UNIQUE_ADDR_1)	Must be initialized
$0B0_{16}$	XMAC_ADDR2 (MAC_UNIQUE_ADDR_2)	Must be initialized
208_{16}	XMAC_ADDR_CMPEN	Must be initialized
218_{16}	XMAC_ADDR3 (MAC_ALT_ADDR0) → MAC_HOST_INFO0	Must be initialized
220_{16}	XMAC_ADDR4 (MAC_ALT_ADDR1)	Must be initialized
228_{16}	XMAC_ADDR5 (MAC_ALT_ADDR2)	Must be initialized
230_{16}	XMAC_ADDR6 (MAC_ALT_ADDR3) → MAC_HOST_INFO1	Must be initialized
238_{16}	XMAC_ADDR7 (MAC_ALT_ADDR4)	Must be initialized
240_{16}	XMAC_ADDR8 (MAC_ALT_ADDR5)	Must be initialized
248_{16}	XMAC_ADDR9 (MAC_ALT_ADDR6) → MAC_HOST_INFO2	Must be initialized
250_{16}	XMAC_ADDR10 (MAC_ALT_ADDR7)	Must be initialized
258_{16}	XMAC_ADDR11 (MAC_ALT_ADDR8)	Must be initialized

TABLE 27-181 xMAC Registers *(3 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
260_{16}	XMAC_ADDR12 (MAC_ALT_ADDR9) → MAC_HOST_INFO3	Must be initialized
268_{16}	XMAC_ADDR13 (MAC_ALT_ADDR10)	Must be initialized
270_{16}	XMAC_ADDR14 (MAC_ALT_ADDR11)	Must be initialized
278_{16}	XMAC_ADDR15 (MAC_ALT_ADDR12) → MAC_HOST_INFO4	Must be initialized
280_{16}	XMAC_ADDR16 (MAC_ALT_ADDR13)	Must be initialized
288_{16}	XMAC_ADDR17 (MAC_ALT_ADDR14)	Must be initialized
290_{16}	XMAC_ADDR18 (MAC_ALT_ADDR15) → MAC_HOST_INFO5	Must be initialized
298_{16}	XMAC_ADDR19 (MAC_ALT_ADDR16)	Must be initialized
$2a0_{16}$	XMAC_ADDR20 (MAC_ALT_ADDR17)	Must be initialized
$2A8_{16}$	XMAC_ADDR21 (MAC_ALT_ADDR18) → MAC_HOST_INFO6	Must be initialized
$2B0_{16}$	XMAC_ADDR22 (MAC_ALT_ADDR19)	Must be initialized
$2B8_{16}$	XMAC_ADDR23 (MAC_ALT_ADDR20)	Must be initialized
$2C0_{16}$	XMAC_ADDR24 (MAC_ALT_ADDR21) → MAC_HOST_INFO7	Must be initialized
$2C8_{16}$	XMAC_ADDR25 (MAC_ALT_ADDR22)	Must be initialized
$2D0_{16}$	XMAC_ADDR26 (MAC_ALT_ADDR23)	Must be initialized
$2D8_{16}$	XMAC_ADDR27 (MAC_ALT_ADDR24) → MAC_HOST_INFO8	Must be initialized
$2E0_{16}$	XMAC_ADDR28 (MAC_ALT_ADDR25)	Must be initialized
$2E8_{16}$	XMAC_ADDR29 (MAC_ALT_ADDR26)	Must be initialized
$2F0_{16}$	XMAC_ADDR30 (MAC_ALT_ADDR27) → MAC_HOST_INFO9	Must be initialized
$2F8_{16}$	XMAC_ADDR31 (MAC_ALT_ADDR28)	Must be initialized
300_{16}	XMAC_ADDR32 (MAC_ALT_ADDR29)	Must be initialized
308_{16}	XMAC_ADDR33 (MAC_ALT_ADDR30) → MAC_HOST_INFO10	Must be initialized
310_{16}	XMAC_ADDR34 (MAC_ALT_ADDR31)	Must be initialized
318_{16}	XMAC_ADDR35 (MAC_ALT_ADDR32)	Must be initialized
320_{16}	XMAC_ADDR36 (MAC_ALT_ADDR33) → MAC_HOST_INFO11	Must be initialized

TABLE 27-181 xMAC Registers *(4 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
328_{16}	XMAC_ADDR37 (MAC_ALT_ADDR34)	Must be initialized
330_{16}	XMAC_ADDR38 (MAC_ALT_ADDR35)	Must be initialized
338_{16}	XMAC_ADDR39 (MAC_ALT_ADDR36) → MAC_HOST_INFO12	Must be initialized
340_{16}	XMAC_ADDR40 (MAC_ALT_ADDR37)	Must be initialized
348_{16}	XMAC_ADDR41 (MAC_ALT_ADDR38)	Must be initialized
350_{16}	XMAC_ADDR42 (MAC_ALT_ADDR39) → MAC_HOST_INFO13	Must be initialized
358_{16}	XMAC_ADDR43 (MAC_ALT_ADDR40)	Must be initialized
360_{16}	XMAC_ADDR44 (MAC_ALT_ADDR41)	Must be initialized
368_{16}	XMAC_ADDR45 (MAC_ALT_ADDR42) → MAC_HOST_INFO14	Must be initialized
370_{16}	XMAC_ADDR46 (MAC_ALT_ADDR43)	Must be initialized
378_{16}	XMAC_ADDR47 (MAC_ALT_ADDR44)	Must be initialized
380_{16}	XMAC_ADDR48 (MAC_ALT_ADDR45) → MAC_HOST_INFO15	Must be initialized
388_{16}	XMAC_ADDR49 (MAC_ALT_ADDR46)	Must be initialized
390_{16}	XMAC_ADDR50 (MAC_ALT_ADDR47)	Must be initialized
	Hash hit uses MAC_HOST_INFO16	Must be initialized
818_{16}	XMAC_ADD_FILT0 → MAC_HOST_INFO18	Must be initialized
820_{16}	XMAC_ADD_FILT1	Must be initialized
828_{16}	XMAC_ADD_FILT2	Must be initialized
830_{16}	XMAC_ADD_FILT12_MASK	Must be initialized
838_{16}	XMAC_ADD_FILT00_MASK	Must be initialized
840_{16}	XMAC_HASH_TBL0	Must be initialized
848_{16}	XMAC_HASH_TBL1	Must be initialized
850_{16}	XMAC_HASH_TBL2	Must be initialized
858_{16}	XMAC_HASH_TBL3	Must be initialized
860_{16}	XMAC_HASH_TBL4	Must be initialized
868_{16}	XMAC_HASH_TBL5	Must be initialized
870_{16}	XMAC_HASH_TBL6	Must be initialized

TABLE 27-181 xMAC Registers *(5 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
878_{16}	XMAC_HASH_TBL7	Must be initialized
880_{16}	XMAC_HASH_TBL8	Must be initialized
888_{16}	XMAC_HASH_TBL9	Must be initialized
890_{16}	XMAC_HASH_TBL10	Must be initialized
898_{16}	XMAC_HASH_TBL11	Must be initialized
$8A0_{16}$	XMAC_HASH_TBL12	Must be initialized
$8A8_{16}$	XMAC_HASH_TBL13	Must be initialized
$8B0_{16}$	XMAC_HASH_TBL14	Must be initialized
$8B8_{16}$	XMAC_HASH_TBL15	Must be initialized
900_{16}	XMAC_HOST_INFO0	
908_{16}	XMAC_HOST_INFO1	
910_{16}	XMAC_HOST_INFO2	
918_{16}	XMAC_HOST_INFO3	
920_{16}	XMAC_HOST_INFO4	
928_{16}	XMAC_HOST_INFO5	
930_{16}	XMAC_HOST_INFO6	
938_{16}	XMAC_HOST_INFO7	
940_{16}	XMAC_HOST_INFO8	
948_{16}	XMAC_HOST_INFO9	
950_{16}	XMAC_HOST_INFO10	
958_{16}	XMAC_HOST_INFO11	
960_{16}	XMAC_HOST_INFO12	
968_{16}	XMAC_HOST_INFO13	
970_{16}	XMAC_HOST_INFO14	
978_{16}	XMAC_HOST_INFO15	
980_{16}	XMAC_HOST_INFO16 (hash hit)	
988_{16}	XMAC_HOST_INFO17 (own da)	
990_{16}	XMAC_HOST_INFO18 (filter hit)	
998_{16}	XMAC_HOST_INFO19 (fc hit)	

TABLE 27-181 xMAC Registers *(6 of 6)*

Address Offset	Register Name	Registers That Must Be Initialized
$B80_{16}$	See section “Preamble Data Register” on page 838	
$B88_{16}$	See TABLE 27-134 on page 838 for Preamble Data Register 1	
$B90_{16}$	See “xMAC Debug Select Register” on page 839	
$B98_{16}$	Table 27-134, “xMAC Preamble Data 1 – xmac_pa_data1 (fzc_mac + 00B8816) (count 2 step 06000),” on page 838	

27.12 XPCS Address Map

There are 19 registers in XPCS.

TABLE 27-182 Address Decode for XPCS

Address Offset	Register Name	Side Effect Register
00_{16}	BASE10G_CONTROL1	Yes
08_{16}	BASE10G_STATUS1	Yes
10_{16}	BASE10G_DEVICE_IDENTIFIER	
18_{16}	BASE10G_SPEED_ABILITY	
20_{16}	BASE10G_DEVICES_IN_PACKAGE	
28_{16}	BASE10G_CONTROL2	
30_{16}	BASE10G_STATUS2	Yes
38_{16}	BASE10G_PACKAGE_IDENTIFIER	
40_{16}	BASE10G_STATUS	Yes
48_{16}	BASE10G_TEST_CONTROL	
50_{16}	BASE10G_CFG_VENDOR1	
58_{16}	BASE10G_DIAGNOSTIC_VENDOR2	Yes
60_{16}	BASE10G_MASK1	
68_{16}	BASE10G_PACKET_COUNTER	
70_{16}	BASE10G_TX_STATE_MACHINE	
78_{16}	BASE10G_DESKEW_ERROR_COUNTER	

TABLE 27-182 Address Decode for XPCS *(Continued)*

Address Offset	Register Name	Side Effect Register
80_{16}	BASE10G_SYMBOL_ERROR_COUNTER_LANE_0_AND_LANE_1	Yes
88_{16}	BASE10G_SYMBOL_ERROR_COUNTER_LANE_2_AND_LANE_3	Yes
90_{16}	Training Vector	

27.13 MIF Register Map

There are 10 registers in MIF.

TABLE 27-183 MIF Registers

Address Offset	Register Name
000_{16}	BB_MDC
008_{16}	BB_MDO
010_{16}	BB_MDO_EN
018_{16}	MIF_FRAME/OUTPUT_REG (side effect)
020_{16}	MIF_CONFIG
028_{16}	POLL_STATUS
030_{16}	POLL_MASK
038_{16}	MIF_STATE_MACHINE
040_{16}	MIF_STATUS
048_{16}	MIF_MASK

27.14 ESR Control Register Map

There are 12 ESR Control registers.

TABLE 27-184 ESR Registers

Address Offset	Register Name
000_{16}	SERDES_RESET
008_{16}	CONFIG

TABLE 27-185 ESR Debug Registers

Address Offset	Register Name
800_{16}	ESR Internal Signals
808_{16}	ESR Debug Selection

27.15 Side Effect Registers

- XTXMAC_SW_RST
- XRXMAC_SW_RST
- BTXMAC_SW_RST
- BRXMAC_SW_RST
- XTXMAC_STATUS (WR-AC)
- XMAC_FC_STAT (WR-AC
- RXMAC_BT_CNT (WR-AC)
- RXMAC_BC_FRM_CNT (WR-AC)
- RXMAC_MC_FRM_CNT (WR-AC)
- RXMAC_FRAG_CNT (WR-AC)
- RXMAC_HIST_CNT1 (WR-AC)
- RXMAC_HIST_CNT2 (WR-AC)
- RXMAC_HIST_CNT3 (WR-AC)
- RXMAC_HIST_CNT4 (WR-AC)
- RXMAC_HIST_CNT5 (WR-AC)
- RXMAC_HIST_CNT6 (WR-AC)
- RXMAC_HIST_CNT7 (WR-AC)
- RXMAC_MPSZER_CNT (WR-AC)
- RXMAC_CRC_ER_CNT (WR-AC)
- RXMAC_CD_VIO_CNT (WR-AC)
- TXMAC_FRM_CNT (WR-AC)
- TXMAC_BYTE_CNT (WR-AC)
- LINK_FAULT_CNT (WR-AC)
- BTXMAC_BYTE_CNT_DEFAULT (WR-AC)
- BTXMAC_FRM_CNT_DEFAULT (WR-AC)
- BRXMAC_BYTE_CNT_DEFAULT (WR-AC)

- PCS_MII_CTL (WR-AC)
- BASE10G_CONTROL1 (XPCS)
- BASE10G_STATUS1 (XPCS)
- BASE10G_STATUS2 (XPCS)
- PCS_ MII_CTL
- MIF_FRAME/OUTPUT_REG

27.15.1 N2_NIU "Hedwig Lite" XAUI SerDes Register Map

TABLE 27-186 Hedwig Lite XAUI SerDes Register Map *(1 of 3)*

MDIO Register Address (hex)	Register Name	POR Default Value	10G Re-program Value	1G ATCA Re-program Value	Comments
0	CFGPLL{15:0}	$000B_{16}$	—	0009_{16}	MPY: 0100_2: 1G 0101_2: 10G
1	CFGPLL{31:16}	0000_{16}	—	—	—
2	STSPLL{15:0}	0001_{16}	—	—	—
3	STSPLL{31:16}	0000_{16}	—	—	—
4	TESTCFG{15:0}	0000_{16}	—	—	—
5	TESTCFG{31:16}	0000_{16}	—	—	For Bump/Pad loopback, set TESTCFG{7:6} = 01_2.
Tx Side Registers					
100	CFGTX0{15:0}	0001_{16}	$0E01_{16}$	$0E21_{16}$	RATE: 00_2: 10G 01_2: 1G
101	CFGTX0{31:16}	0000_{16}	—	—	—
102	STSTX0{15:0}	0000_{16}	—	—	—
103	STSTX0{31:16}	0000_{16}	—	—	—
104	CFGTX1{15:0}	0001_{16}	$0E01_{16}$	$0E20_{16}$ (entx = 0 power down)	RATE: 00_2: 10G 01_2: 1G
105	CFGTX1{31:16}	0000_{16}	—	—	—

TABLE 27-186 Hedwig Lite XAUI SerDes Register Map *(2 of 3)*

MDIO Register Address (hex)	Register Name	POR Default Value	10G Re-program Value	1G ATCA Re-program Value	Comments
106	STSTX1{15:0}	0000_{16}	—	—	—
107	STSTX1{31:16}	0000_{16}	—	—	—
108	CFGTX2{15:0}	0001_{16}	$0E01_{16}$	$0E20_{16}$ (entx = 0 power down)	RATE: 00_2: 10G 01_2: 1G
109	CFGTX2{31:16}	0000_{16}	—	—	—
10A	STSTX2{15:0}	0000_{16}	—	—	—
10B	STSTX2{31:16}	0000_{16}	—	—	—
10C	CFGTX3{15:0}	0001_{16}	$0E01_{16}$	$0E20_{16}$ (entx = 0 power down)	RATE: 00_2: 10G 01_2: 1G
10D	CFGTX3{31:16}	0000_{16}	—	—	—
10E	STSTX3{15:0}	0000_{16}	—	—	—
10F	STSTX3{31:16}	0000_{16}	—	—	—
Rx Side Registers					
120	CFGRX0{15:0}	0101_{16}	9101_{16}	9121_{16}	RATE: 0_20: 10G 01_2: 1G
121	CFGRX0{31:16}	0000_{16}	0008_{16}	0008_{16}	—
122	STSRX0{15:0}	0000_{16}	—	—	—
123	STSRX0{31:16}	0000_{16}	—	—	—
124	CFGRX1{15:0}	0101_{16}	9101_{16}	9120_{16} (enrx = 0 power down)	RATE: 00_2: 10G 01_2: $1G_2$
125	CFGRX1{31:16}	0000_{16}	0008_{16}	0008_{16}	
126	STSRX1{15:0}	0000_{16}	—	—	—
127	STSRX1{31:16}	0000_{16}	—	—	—
128	CFGRX2{15:0}	0101_{16}	9101_{16}	9120_{16} (enrx = 0 power down)	RATE: 00_2: 10G 01_2: 1G

TABLE 27-186 Hedwig Lite XAUI SerDes Register Map *(3 of 3)*

MDIO Register Address (hex)	Register Name	POR Default Value	10G Re-program Value	1G ATCA Re-program Value	Comments
129	CFGRX2{31:16}	0000_{16}	0008_{16}	0008_{16}	—
12A	STSRX2{15:0}	0000_{16}	—	—	—
12B	STSRX2{31:16}	0000_{16}	—	—	—
12C	CFGRX3{15:0}	0101_{16}	9101_{16}	9120_{16} (enrx = 0 power down)	RATE: 00_2: 10G 01_2: 1G
12D	CFGRX3{31:16}	0000_{16}	0008_{16}	0008_{16}	—
12E	STSRX3{15:0}	0000_{16}	—	—	—
12F	STSRX3{31:16}	0000_{16}	—	—	—

CHAPTER 28

Network Interface Unit: Ethernet SerDes

28.1 Overview

Before going into the specifics, this section gives a general overview ofSerDes and transceivers terminology. Generally speaking, depending on different implementations, the functionality of SerDes, transceivers, and PHY are often grouped into one or multiple hardware modules. Some vendors may combine SerDes and transceiver into one chip package. Others may separate them into multiple chips. Therefore, these terms are often interchangeable or overlapped, depending on how their associated functions are grouped and packaged. This section clarifies these terms used.

28.1.1 SerDes

SerDes is the short form for the SERializer/DESerializer. In transmit, the serializer converts parallel data from the MAC (Media Access Control) to serialize bit streams. In receive, the process is reversed.

28.1.1.1 The Role of SerDes in 10G Ethernet

In 10GBase-X 10G Ethernet implementation, the parallel data from the MAC are presented in four differential lane pairs. Each lane pair is operated at data rate of 3.125 Gbps (2.5Gbps + 10-bit/8-bit encoding overhead). This format is named the XAUI Interface.. The SerDes, together with the MAC's XPCS sub-block, forms the XGXS layer of the 10G Ethernet interface:

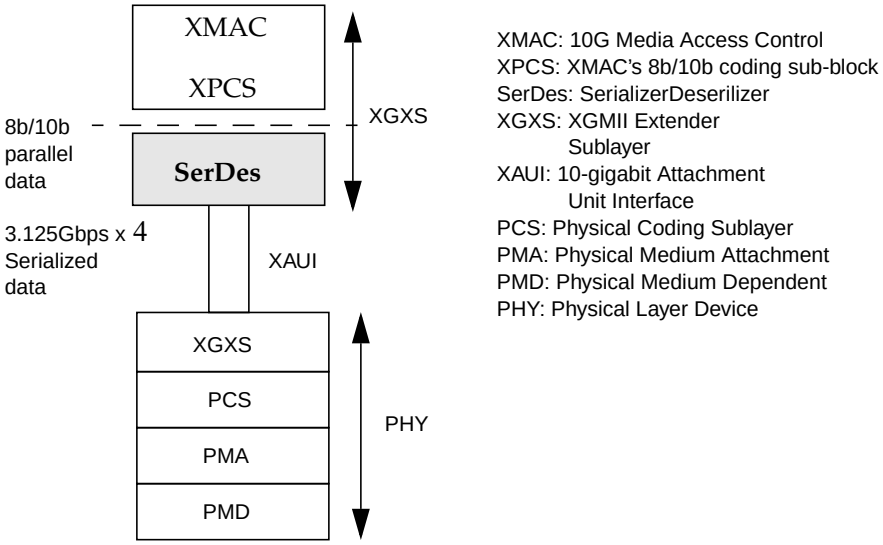


FIGURE 28-1 SerDes' Role in 10G Ethernet

The XGXS is connected to the physical device via the XAUI interface. The deserialization process is the reverse of the serialization process described above. SerDes is outside of the NIU core but is part of the UltraSPARC T2 CPU.

28.1.1.2 The Role of SerDes in 1G Ethernet on Fiber

SerDes also play a role in 1G Fiber mode. However, in 1G Fiber mode, the SerDes' role is the PMA of the physical layer:

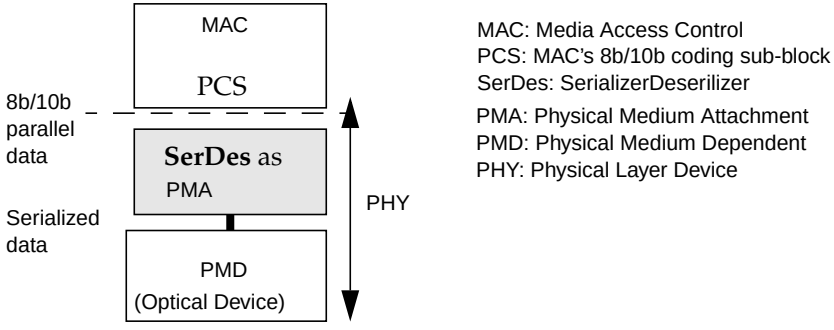


FIGURE 28-2 SerDes' Role in 1G Ethernet on Fiber

28.1.1.3 Transceiver's Role in 1G Copper Ethernet

SerDes play no role in 1G copper mode since data does not need to be serialized.

28.1.2 Transceiver

Transceiver is generally referred to the physical layer device. It consists of the functionality to interface with the Ethernet medium. Therefore, transceivers are associated with the physical medium it is attached to. For example, a fiber transceiver drives an optical fiber medium, and a copper transceiver drives a copper medium.

28.1.2.1 Transceiver's Role in 10G Ethernet

In 10G Ethernet configurations, the transceiver functionality includes the XGXS, PCS, PMA, and the PMD. In transmit, it converts the serial data input groups to a format to be presented to the physical layer medium. In receive, the process is reversed. In 10GBase-X 10G Ethernet implementation, the transceiver interfaces with the MAC layer through the XAUI interface. The transceiver is external to the UltraSPARC T2 CPU. Both 10G copper and fiber transceiver interface to UltraSPARC T2 via XAUI. The following shows the general relationship between the MAC, SerDes, and Transceiver.

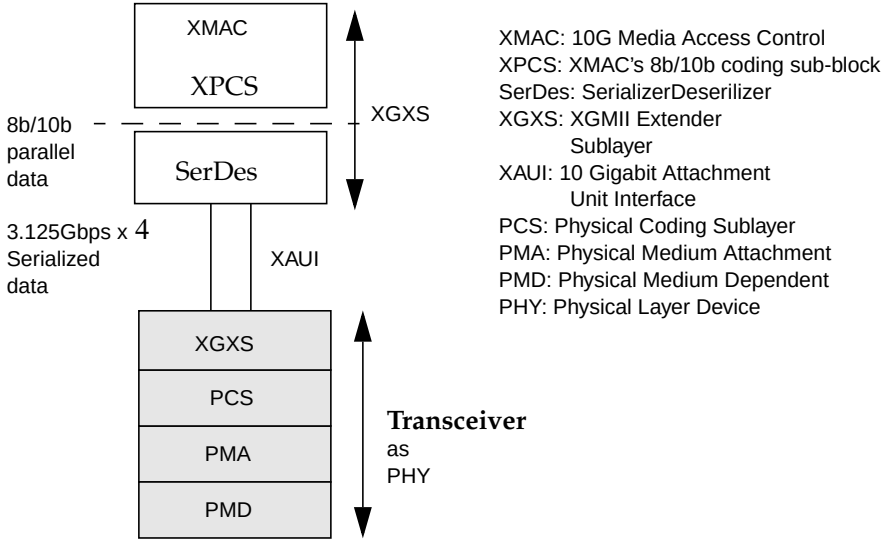


FIGURE 28-3 Transceiver's Role in 10G Ethernet

28.1.2.2 Transceiver's Role in 1G Fiber Ethernet:

In 1G Fiber mode, transceiver is referred to the optical device.

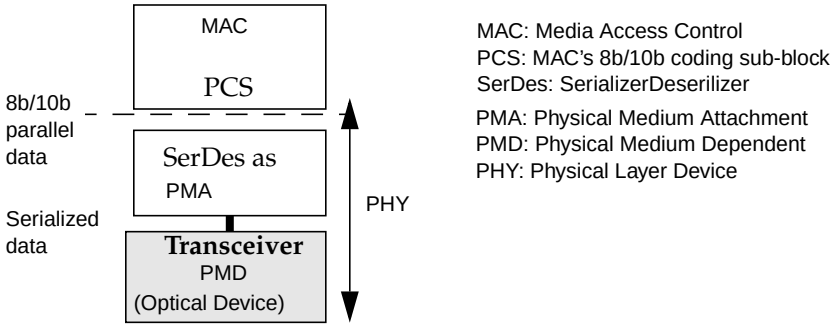


FIGURE 28-4 Transceiver's Role in 1G Fiber Ethernet

28.1.2.3 Transceiver's Role in 1G Copper Ethernet

In 1G Copper mode, transceiver is referred to the 1G copper PHY.

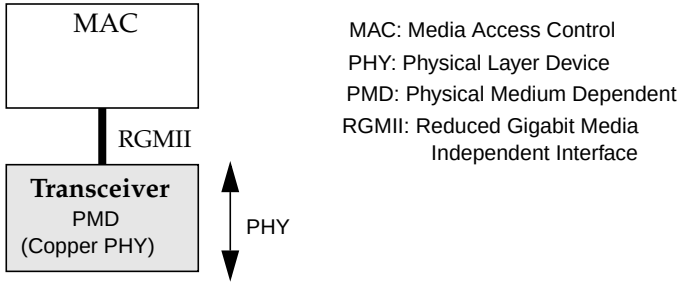


FIGURE 28-5 Transceiver's Role in 1G copper Ethernet

28.1.3 PHY

"PHY" is an abbreviation for the "Physical Layer Device". PHY is most often used to denote a copper-based transceiver. Therefore, Copper transceiver and PHY are often used interchangeably.

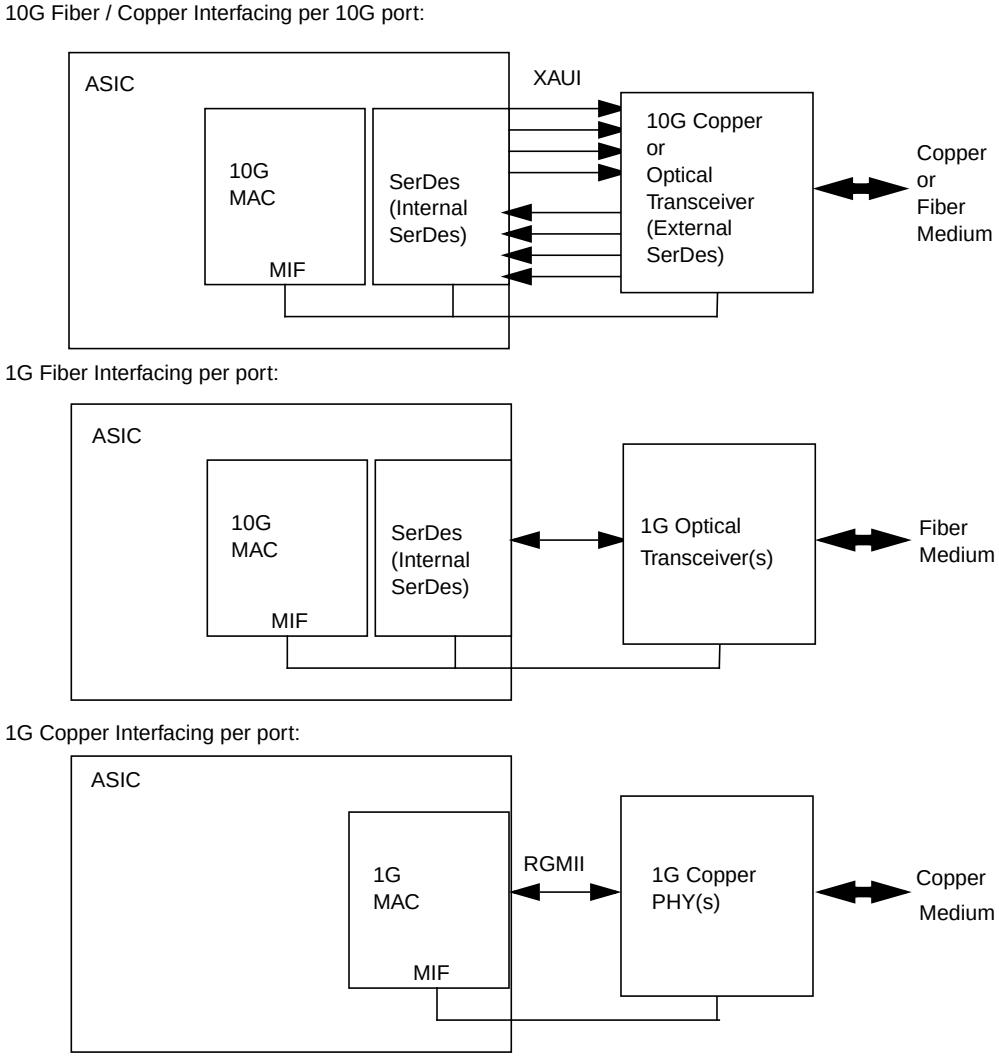


FIGURE 28-6 General Relationship Among MAC, SerDes, Transceiver in 10G and 1G mode

28.2 SerDes and Transceivers

This section describes the SerDes and transceiver configuration of the UltraSPARC T2-based system.

28.2.1 UltraSPARC T2 SerDes

The UltraSPARC T2's networking components consist of the NIU virtual processor interfaces with the SIU on the system side and interfaces with the Ethernet transceiver using the following SerDes: Texas Instrument's WIZ5C2xxN3x0 core.

Please refer to [3] for details of the WIZ5C2xxN3x0 core and the programming reference model. The following diagram shows SerDes configuration in the UltraSPARC T2 environment:

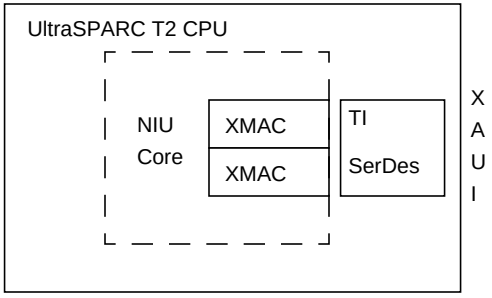


FIGURE 28-7 SerDes Configuration in UltraSPARC T2

28.2.2 10G Optical Transceiver

TBD.

28.3 SerDes / Transceiver Usage

The following matrix summarizes the SerDes/transceiver usage in different UltraSPARC T2 configurations:

TABLE 28-1 SerDes/Transceiver Usage

	Internal (On Chip)		External (On Board)		
	Ethernet SerDes	**PCI-E SerDes**	**10G Optical Transceiver**	**10G Copper Transceiver**	**1G Copper Transceiver**
UltraSPARC T2 NIU	TI WIZ5C2xxN3x0 Core	N/A	TBD	(TBD?)	N/A

28.4 Program I/O Interface

28.4.1 Management Interface

The MAC provide the Management Interface (MIF) to manage the SerDes and the transceivers. The MIF interface is an MII (Media Independent Interface) It uses the MDIO protocol to interface with the physical layer devices. (MII and MDIO are defined in the IEEE 802.3 specification's Clause 22 and Clause 45).

28.4.2 MIF Operations

Please refer to the MIF section of the 's MAC Programming Reference chapter for details of MIF operations. This section gives two examples of performing a SerDes or transceiver register access, one example on write, the other example on read.

Example 1: Write to a SerDes/Transceiver Register via MIF per CLAUSE 45

1. Write to MIF Frame/Output Register:

```
frame.bits.st = 0;            /* Use Clause 45 */
frame.bits.op = 0;            /* Setup register address */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
frame.bits.data = reg_addr;   /* 16-bit register address */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          set up address*/
frame.bits.op = 0x1;          /* Write cmd */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.a_msb = 1;         /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
frame.bits.data = value;      /* 16-bit payload value */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          write data */
```

2. Poll for command complete:

```
do { /* Poll for cmd complete */
        PIO_RD(MIF_FRAMEOUT_REG, frame.value);
while (frame.bits.w0.ta_lsh == 0);
/* Cmd complete when ta_lsh is set to 1 */
```

Example 2: Read from a SerDes/Transceiver Register via MIF per CLAUSE 45

1. Write to MIF Frame/Output Register:

```
frame.bits.st = 0;             /* Use Clause 45 */
frame.bits.op = 0;            /* Setup register address */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          set up address*/
frame.bits.st = 0;            /* Use Clause 45 */
frame.bits.op = 0x2;          /* Read cmd */
frame.bits.phyad = portn;     /* Setup port address */
frame.bits.regad = dev_addr;  /* Setup device address */
frame.bits.ta_msb = 1;        /* Set to 1 to issue a cmd */
frame.bits.ta_lsb = 0;        /* Set to 0 to issue a cmd */
PIO_WR(MIF_FRAMEOUT_REG, frame.value); /* Execute cmd to
                                          read data */
```

2. Poll for command complete:

```
do { /* Poll for cmd complete */
       PIO_RD(MIF_FRAMEOUT_REG, frame.value);
 while (frame.bits.w0.ta_lsh == 0);
```

3. Obtain result from 16-bit data field of MIF Frame/Output register:

```
value = frame.bits.w0.data;
```

Note: Please refer to [1] for port and device address assignment for UltraSPARC T2 SerDes.

28.4.3 Accessing Transceiver Registers

All transceiver registers are accessible through the MDIO interface. Please refer to the "MIF Programmer's Reference Manual" for methods of accessing MIF programming resources.

28.4.4 Accessing SerDes Registers

SerDes registers are accessible through the MDIO interface. Please refer to the "MIF Programmer's Reference Manual" for methods of accessing MIF programming resources.

28.5 Programming Reference Model

The Programming Reference model of all UltraSPARC T2 SerDes/Transceivers are referenced to the individual SerDes/transceiver specifications. This section includes a summary of PRM obtained from each individual specification.

28.5.1 TI SerDes Programming Model

Please refer to the WIZ6C2xxN3x0 Macro Family Datasheet for details of the TI SerDes. The following is a register map summary.

TABLE 28-2 TI SerDes Register Map

Register Address	Register Name
000_{16}	ESR_TI_PLL_CFG_L_REG
001_{16}	ESR_TI_PLL_CFG_H_REG
002_{16}	ESR_TI_PLL_STS_L_REG
003_{16}	ESR_TI_PLL_STS_H_REG
004_{16}	ESR_TI_TEST_CFG_L_REG
005_{16}	ESR_TI_TEST_CFG_H_REG
100_{16} + (chan × 4)	ESR_TI_TX_CFG_L_REG_ADDR (chan 0 ~ 7)
100_{16} + (chan × 4) + 1	ESR_TI_TX_CFG_H_REG_ADDR (chan 0 ~ 7)
100_{16} + (chan × 4) + 2	ESR_TI_TX_STS_L_REG_ADDR (chan 0 ~ 7)
100_{16} + (chan × 4) + 3	ESR_TI_TX_STS_H_REG_ADDR (chan 0 ~ 7)
120_{16} + (chan × 4)	ESR_TI_RX_CFG_L_REG_ADDR (chan 0 ~ 7)
120_{16} + (chan × 4) + 1	ESR_TI_RX_CFG_H_REG_ADDR (chan 0 ~ 7)
120_{16} + (chan ×4) + 2	ESR_TI_RX_STS_L_REG_ADDR (chan 0 ~ 7)
120_{16} + (chan × 4) + 3	ESR_TI_RX_STS_H_REG_ADDR (chan 0 ~ 7)

Each register is 16 bits wide and each register pair, H/L or High/Low, corresponds to an internal bus, with bus width up to 32 bits. Not all register pairs have all 32 bits defined. The following is the definition of each bus. Bits 15:0 corresponds to the L register, and bits 31:16 to the H register. All configuration registers have initial values of 0, with the exception of PLL_CFG. All registers corresponding to STS or status buses are read-only.

TABLE 28-3 PLL_CFG Bus 11:0, Table 8.1 in reference document, p. 30

Bit	Field	Initial Value	Description
0	enpll	1	Enable PLL.
4:1	mpy	101_2	PLL multiply.
7:5	—		*Reserved*
9:8	lb	0	Loop bandwidth.
11:10	std	0	Standard selection.

TABLE 28-4 PLL_STS Bus [3:0], Table 8.2 in reference document, p. 30

Bit	Field	Initial Value	Description
0	lock		PLL lock.
3:1	—	0	*Reserved*

TABLE 28-5 TEST_CFG Bus 15:0, Table 9.1 in reference document, p. 112

Bit	Field	Initial Value	Description
1:0	testpatt		Test pattern.
2	entxpatt		Enable Tx patterns.
3	enrxpatt		Enable Rx patterns.
5:4	clkbyp		Clock bypass.
7:6	loopback		Loopback.
8	enbstx		Transmitter boundary scan.
9	enbsrx		Receiver boundary scan.
10	enbsac		Receiver pulse boundary scan.
11	—		*Reserved*
13:12	rate		Operating rate.
14	invpatt		Invert polarity
15	—		*Reserved*.

TABLE 28-6 TX_CFG Bus 15:0, Table 8.27 in reference document, p. 70

Bit	Field	Initial Value	Description
0	entx		Enable transmitter.
1	entest		Enable test.
4:2	buswidth		Bus width.
6:5	rate		Operating rate.
7	invpair		Invert polarity.
8	cm		Common mode.

TABLE 28-6 TX_CFG Bus 15:0, Table 8.27 in reference document, p. 70 *(Continued)*

Bit	Field	Initial Value	Description
11:9	swing		Output swing.
15:12	de		Deemphasis.
16	enftp		Enable fixed TXBCLKIN[i] phase.
17	bstx		Boundary scan data.
19:18	—		*Reserved*
20	enidl		Idle selection.
22:21	rdtct		Receiver detect control.
23	—		*Reserved*

TABLE 28-7 TX_STS Bus 3:0, Table 8.28 in reference document, p. 71

Bit	Field	Initial Value	Description
0	testfail		Test failure.
1	rdtctip		Receiver detect in progress.
3:2	—	0	*Reserved*

TABLE 28-8 RX_CFG Bus 27:0, Table 8.13 in reference document, p. 42

Bit	Field	Initial Value	Description
0	enrx		Enable receiver.
1	entest		Enable test.
4:2	buswidth		Bus width.
6:5	rate		Operating rate.
7	invpair		Invert polarity.
10:8	term		Termination.
11	—		*Reserved*
13:12	align		Symbol alignment.
15:14	los		Loss of signal.
18:16	cdr		Clock data recovery.
22:19	eq		Equalizer.
23	—		*Reserved*
24	bsinrxp		Boundary scan initialization for RXPi.
25	bsinrxn		Boundary scan initialization for RXNi.
27:26	—		*Reserved*

TABLE 28-9 RX_STS Bus 11:0, Table 8.14 in reference document, p. 43

Bit	Field	Initial Value	Description
0	testfail		Test failure.
1	sync		Symbol alignment.
2	oddcg		Odd code group.
3	losdtct		Loss of signal detect.
4	bsrxp		Boundary scan data.
5	bsrxn		Boundary scan data.
7:6	—		*Reserved*
8	cwdtct		COMWAKE detect.
9	crcidtct		COMRESET/COMINIT detect.
11:10	—		*Reserved*

28.5.2 BCM 10G Optical Transceiver Programming Model

Please refer to “Section 2: Register Summary” (Page 27–66) of the BCM8704 specification for details on all registers.

28.5.3 10G Copper Transceiver Programming Model

TBD.

28.5.4 BCM 1G Copper PHY Programming Model

Please refer to “Section 5: Register Summary” (Page 30–95) of the BCM5464 specification for details on all registers.

28.6 Initialization Sequence

The following shows examples of SerDes initialization. Note that the sequence here is oversimplified. All SerDes timing control steps are excluded from this sequence.

28.6.1 TI SerDes Initialization

Example 1: Sets up the TI SerDes' Port0 Hedwig Lite Core for 10G mode:

Note: In writing to TI SerDes, bit 15 of MIF_FRAMEOUT register should always be set to 1 during the address selecting phase.

```
PIO_WR(MIF_CONFIG_REG, 0x00008008);
                /* Set Poll Enable and choose Clause 45 (st = 00) */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8000); /* Select addr 000h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8001); /* Select addr 001h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8004); /* Select addr 004h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8005); /* Select addr 005h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8100); /* Select addr 100h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8101); /* Select addr 101h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8104); /* Select addr 104h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8105); /* Select addr 105h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8108); /* Select addr 108h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8109); /* Select addr 108h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A810C); /* Select addr 10ch Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A810D); /* Select addr 10dh Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8120); /* Select addr 120h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8121); /* Select addr 121h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8124); /* Select addr 124h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8125); /* Select addr 125h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8128); /* Select addr 128h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5001); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8129); /* Select addr 129h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A812C); /* Select addr 12ch Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5001); /* Write data */
```

```
PIO_WR(MIF_FRAMEOUT_REG, 0x007A812D); /* Select addr 12dh Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8000); /* Select addr 000h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0003); /* Write data */
```

Example 2: Sets up the TI SerDes' Port0 Hedwig Lite Core for 1G mode:

```
PIO_WR(MIF_CONFIG_REG, 0x00008008);
        /* Set Poll Enable and choose Clause 45 (st = 00) */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8000); /* Select addr 000h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8001); /* Select addr 001h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8004); /* Select addr 004h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8005); /* Select addr 005h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8100); /* Select addr 1001h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8101); /* Select addr 101h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8104); /* Select addr 104h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8105); /* Select addr 1051h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8108); /* Select addr 108h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8109); /* Select addr 109h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A810C); /* Select addr 10ch Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A810D); /* Select addr 10dh Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0000); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8120); /* Select addr 120h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8121); /* Select addr 121h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8124); /* Select addr 124h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8125); /* Select addr 125h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8128); /* Select addr 128h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5021); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8129); /* Select addr 129h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A812C); /* Select addr 12ch Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A5021); /* Write data */
```

```
PIO_WR(MIF_FRAMEOUT_REG, 0x007A812D); /* Select addr 12dh Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A0008); /* Write data */
PIO_WR(MIF_FRAMEOUT_REG, 0x007A8000); /* Select addr 000h Port0 */
PIO_WR(MIF_FRAMEOUT_REG, 0x107A000B); /* Write data */
```

28.6.2 BCM 10G Fiber Transceiver Initialization Sequence

Using the BMC8704 10-G Ethernet/Fibre Channel Transceiver with the TI WIZ5C2xxN3x1 SerDes, it is necessary to follow the reset initialization sequence below:

1. Software reset/init of the BMC8704 (see below for detail).
2. Init of the TI WIZ5C2xxN3x1 SerDes. (See above for detail)
3. Software reset/init of the BMC8704 (see below for detail).

On POR (power on reset), we observe the BMC8704 receives garbage characters from the TI SerDes. On POR, while the BMC8704 PLL is trying to lock, the TI SerDes Transmit voltage swing is not adequate to meet the BMC8704 voltage threshold requirements forcing the BMC8404 into an indeterminate state. Consequently, an init/reset of the BMC8704 preceding SerDes init is necessary for recovery. Following SerDes init, with the BMC8704 PLL locked and the SerDes TX Swing properly initialized, a final reset/init of the BMC8704 is necessary for mission mode use.

The Reset/init of the BMC8704 is :

- BMC8704 Reset PHY XS (MDIO write dev = 0x4, write addr = 0x0000, data = 0xA040... poll for 0x2040 for reset clear).
- BMC8704 Reset PMA/PMD (MDIO write dev = 0x1, write addr = 0x0000, data = 0xA040... poll for 0x2040 for reset clear).
- BMC8704 (MDIO write device = 0x3 write addr = 0xc803, data = 0x0164)
- BMC8704 (MDIO write device = 0x3 write addr = 0xc800, data = 0x7fbf)
- BMC8704 (MDIO read device = 0x01, addr 0x0A, bit[0] = 0x1 signal detect of TX diff from UltraSPARC T2 MAC
- BMC8704 (MDIO read device = 0x03, addr 0x20, bit [12] = 0x1 for link up with UltraSPARC T2 MAC
- BMC8704 (MDIO read device= 0x04, addr 0x18, bit [12] = 0x1 for XAUI lane align and bits [3:0] = 0xF for synchronization with UltraSPARC T2 MAC

28.6.3 10G Copper Transceiver Initialization Sequence

TBD.

28.6.4 BCM 1G PHY Initialization Sequence

TBD.

28.7 References

[1] NIU MAC Programming Model.

[2] GigaBlaze Gflx x 4 Core CW000506 Design Manual.

[3] WIZ6C2xxN3x0 Macro Family Datasheet SR60 Wizard Technology 0.5 - 4.25 Gbps SerDes Revision 00.00.w01 4th Jun 2004.

[4] BCM8704 Serial 10-Gigabit Ethernet/Fibre Channel Transceiver with XAUI Interface.

[5] TBD.

[6] BCM5464R Quad 10/100/1000BASE-T Transceiver with HSTL RGMII.

APPENDIX **A**

Programming Guidelines

A.1 Multithreading

In UltraSPARC T2, each physical core contains eight strands. The strands are divided into two thread groups, with strands 0–3 occupying one thread group and strands 4–7 occupying the other.

Within a thread group, among the available strands, the least recently picked strand is selected for execution every cycle. Thus, up to two instructions can be picked each cycle.

Since each physical core has only one load/store unit and one floating-point and graphics unit, only one load/store or FGU instruction may be picked each cycle. One thread group can issue a load/store instruction while the other thread group issues an FGU instruction. Arbitrating between the two thread groups is done with a least-recently-picked mechanism, to ensure fairness.

Since context switching is built into the UltraSPARC T2 pipeline (via the D and P stages), strands are switched each cycle with no pipeline stall penalty (except when resource collisions occur, such as when both thread groups require the load/store unit or the FGU).

In normal operation, UltraSPARC T2 speculates that most control-transfer instructions will be "not taken" and that loads hit in the L1 data cache. An enable bit, accessible to hyperprivileged software, controls whether UltraSPARC T2 speculates on these instructions or not (see `ASI_LSU_CONTROL_REG` description in *`ASI_LSU_CONTROL_REG`* on page 407).

The following instructions change a strand from available to unavailable until hardware determines that their input/execution requirements can be satisfied, assuming speculation is enabled:

- CALL, DONE, RETRY, JMPL
- LDFSR, LDXFSR, STFSR, STXFSR
- All WRPR, WR, WRHPR

- All RDPR, RD, RDHPR
- SAVE(D), RESTORE(D), RETURN, FLUSHW (all register-window management)
- All MUL and DIV
- MULX, UMUL, SMUL, POPC
- MEMBAR#, FLUSH
- FCMP
- All memory operations to/from alternate space
- All atomic (load-store) operations
- PREFETCH

If speculation is not enabled, the following instruction types also change a strand from available to unavailable until hardware determines that their execution requirements can be satisfied:

- All control transfer instructions
- All loads

A.2 Instruction Latency

TABLE A-1 lists the minimum single-strand instruction latencies for UltraSPARC T2. When multiple strands are executing, some or much of the additional latency for multicycle instructions will be overlapped with execution of the additional strands.

TABLE A-1 UltraSPARC T2 Instruction Latencies *(1 of 8)*

Opcode	Description	Latency	Notes
ADD (ADDcc)	Add (and modify condition codes)	1	
ADDC (ADDCcc)	Add with carry (and modify condition codes)	1	
ALIGNADDRESS	Calculate address for misaligned data access	1	
ALIGNADDRESSL	Calculate address for misaligned data access (little-endian)	1	
ALLCLEAN	Mark all windows as clean	25	
AND (ANDcc)	Logical **and** (and modify condition codes)	1	
ANDN (ANDNcc)	Logical **and not** (and modify condition codes)	1	
ARRAY{8,16,32}	3-D address to blocked byte address conversion	6	
Bicc	Branch on integer condition codes	1 not-taken, 6 taken	
BMASK	Write the GSR.mask field	25	

TABLE A-1 UltraSPARC T2 Instruction Latencies *(2 of 8)*

Opcode	Description	Latency	Notes
BPcc	Branch on integer condition codes with prediction	1 not-taken, 6 taken	
BPr	Branch on contents of integer register with prediction	1 not-taken, 6 taken	
BSHUFFLE	Permute bytes as specified by the GSR.mask field	6	
CALL	Call and link	6	
CASA	Compare and swap word in alternate space	20-30	Done in L2 cache
CASXA	Compare and swap doubleword in alternate space	20-30	Done in L2 cache
DONE	Return from trap	6	
EDGE{8,16,32}{L}{N}	Edge boundary processing {little-endian} {non-condition-code altering}	6	
FABS(s,d)	Floating-point absolute value	6	
FADD(s,d)	Floating-point add	6	
FALIGNDATA	Perform data alignment for misaligned data	6	
FANDNOT1{s}	Negated src1 **and** src2 (single precision)	6	
FANDNOT2{s}	src1 **and** negated src2 (single precision)	6	
FAND{s}	Logical **and** (single precision)	6	
FBPfcc	Branch on floating-point condition codes with prediction	1 not-taken, 6 taken	
FBfcc	Branch on floating-point condition codes	1 not-taken, 6 taken	
FCMP(s,d)	Floating-point compare	6	
FCMPE(s,d)	Floating-point compare (exception if unordered)	6	
FCMPEQ{16,32}	Four 16-bit / two 32-bit compare: set integer dest if src1 = src2	6	
FCMPGT{16,32}	Four 16-bit / two 32-bit compare: set integer dest if src1 > src2	6	
FCMPLE{16,32}	Four 16-bit / two 32-bit compare: set integer dest if src1 ≤ src2	6	
FCMPNE{16,32}	Four 16-bit / two 32-bit compare: set integer dest if src1 ≠ src2	6	
FDIV(s,d)	Floating-point divide	19 SP, 33 DP	

TABLE A-1 UltraSPARC T2 Instruction Latencies *(3 of 8)*

Opcode	Description	Latency	Notes
FEXPAND	Four 8-bit to 16-bit expand	6	
FiTO(s,d)	Convert integer to floating-point	6	
FLUSH	Flush instruction memory	variable	
FLUSHW	Flush register windows	25	
FMOV(s,d)	Floating-point move	6	
FMOV(s,d)cc	Move floating-point register if condition is satisfied	6	
FMOV(s,d)R	Move floating-point register if integer register contents satisfy condition	6	
FMUL(s,d)	Floating-point multiply	6	
FMUL8SUx16	Signed upper 8- x 16-bit partitioned product of corresponding components	6	
FMUL8ULx16	Unsigned lower 8- x 16-bit partitioned product of corresponding components	6	
FMUL8x16	8- x 16-bit partitioned product of corresponding components	6	
FMUL8x16AL	Signed lower 8- x 16-bit lower α partitioned product of 4 components	6	
FMUL8x16AU	Signed upper 8- x 16-bit lower α partitioned product of 4 components	6	
FMULD8SUx16	Signed upper 8- x 16-bit multiply → 32-bit partitioned product of components	6	
FMULD8ULx16	Unsigned lower 8- x 16-bit multiply → 32-bit partitioned product of components	6	
FNAND{s}	Logical **nand** (single precision)	6	
FNEG(s,d)	Floating-point negate	6	
FNOR{s}	Logical **nor** (single precision)	6	
FNOT1{s}	Negate (1's complement) src1 (single precision)	6	
FNOT2{s}	Negate (1's complement) src2 (single precision)	6	
FONE{s}	One fill (single precision)	6	
FORNOT1{s}	Negated src1 **or** src2 (single precision)	6	
FORNOT2{s}	src1 **or** negated src2 (single precision)	6	
FOR{s}	Logical **or** (single precision)	6	
FPACKFIX	Two 32-bit to 16-bit fixed pack	6	
FPACK{16,32}	Four 16-bit/two 32-bit pixel pack	6	

TABLE A-1 UltraSPARC T2 Instruction Latencies *(4 of 8)*

Opcode	Description	Latency	Notes
FPADD{16,32}{s}	Four 16-bit/two 32-bit partitioned add (single precision)	6	
FPMERGE	Two 32-bit to 64-bit fixed merge	6	
FPSUB{16,32}{s}	Four 16-bit/two 32-bit partitioned subtract (single precision)	6	
FsMULd	Floating-point multiply single to double	6	
FSQRT(s,d)	Floating-point square root	19 SP, 33 DP	
FSRC1{s}	Copy src1 (single precision)	6	
FSRC2{s}	Copy src2 (single precision)	6	
F(s,d)TO(s,d)	Convert between floating-point formats	6	
F(s,d)TOi	Convert floating point to integer	6	
F(s,d)TOx	Convert floating point to 64-bit integer	6	
FSUB(s,d)	Floating-point subtract	6	
FXNOR{s}	Logical **xnor** (single precision)	6	
FXOR{s}	Logical **xor** (single precision)	6	
FxTO(s,d)	Convert 64-bit integer to floating-point	6	
FZERO{s}	Zero fill (single precision)	6	
ILLTRAP	Illegal instruction		
INVALW	Mark all windows as CANSAVE	6	
JMPL	Jump and link	6	
LDBLOCKF	64-byte block load	32	
LDD	Load doubleword	3	
LDDA	Load doubleword from alternate space	variable	
LDDF	Load double floating-point	3	
LDDFA	Load double floating-point from alternate space	variable	
LDF	Load floating-point	3	
LDFA	Load floating-point from alternate space	variable	
LDFSR	Load floating-point state register lower	1-8	pre-sync to previous FGU op from that thread
LDSB	Load signed byte	3	

TABLE A-1 UltraSPARC T2 Instruction Latencies *(5 of 8)*

Opcode	Description	Latency	Notes
LDSBA	Load signed byte from alternate space	variable	
LDSH	Load signed halfword	3	
LDSHA	Load signed halfword from alternate space	variable	
LDSTUB	Load-store unsigned byte	3	
LDSTUBA	Load-store unsigned byte in alternate space	variable	
LDSW	Load signed word	3	
LDSWA	Load signed word from alternate space	variable	
LDUB	Load unsigned byte	3	
LDUBA	Load unsigned byte from alternate space	variable	
LDUH	Load unsigned halfword	3	
LDUHA	Load unsigned halfword from alternate space	variable	
LDUW	Load unsigned word	3	
LDUWA	Load unsigned word from alternate space	variable	
LDX	Load extended	3	
LDXA	Load extended from alternate space	variable	
LDXFSR	Load extended floating-point state register	1-8	pre-sync to previous FGU op from that thread
MEMBAR	Memory barrier	variable	
MOVcc	Move integer register if condition is satisfied	1	
MOVr	Move integer register on contents of integer register	1	
MULScc	Multiply step (and modify condition codes)		
MULX	Multiply 64-bit integers	5	
NOP	No operation	1	
NORMALW	Mark other windows as restorable	25	
OR (ORcc)	Inclusive-**or** (and modify condition codes)	1	
ORN (ORNcc)	Inclusive-**or not** (and modify condition codes)	1	
OTHERW	Mark restorable windows as other	6	
PDIST	Distance between eight 8-bit components	6	1 per 2 cycles

TABLE A-1 UltraSPARC T2 Instruction Latencies *(6 of 8)*

Opcode	Description	Latency	Notes
POPC	Population count	5	
PREFETCH	Prefetch data	variable	>6
PREFETCHA	Prefetch data from alternate space	variable	>6
RDASI	Read ASI register	variable	
RDASR	Read ancillary state register	variable	
RDCCR	Read condition codes register	variable	
RDFPRS	Read floating-point registers state register	variable	
RDHPR	Read hyperprivileged register	variable	
RDPC	Read program counter	variable	
RDPR	Read privileged register	variable	
RDTICK	Read TICK register	variable	
RDY	Read Y register	variable	
RESTORE	Restore caller's window	6	
RESTORED	Window has been restored	6	
RETRY	Return from trap and retry		
RETURN	Return	7	
SAVE	Save caller's window	6	
SAVED	Window has been saved	6	
SDIV (SDIVcc)	32-bit signed integer divide (and modify condition codes)	12-41	
SDIVX	64-bit signed integer divide	12-41	
SETHI	Set high 22 bits of low word of integer register	1	
SIAM	Set interval arithmetic mode	6	
SIR	Software-initiated reset		
SLL	Shift left logical	1	
SLLX	Shift left logical, extended	1	
SMUL (SMULcc)	Signed integer multiply (and modify condition codes)	5	
SRA	Shift right arithmetic	1	
SRAX	Shift right arithmetic, extended	1	
SRL	Shift right logical	1	
SRLX	Shift right logical, extended	1	

TABLE A-1 UltraSPARC T2 Instruction Latencies *(7 of 8)*

Opcode	Description	Latency	Notes
STB	Store byte	1	
STBA	Store byte into alternate space		
STBAR	Store barrier	variable	
STBLOCKF	64-byte block store	16	Assuming store buffer empty when STBLOCKF decodes
STD	Store doubleword	1	
STDA	Store doubleword into alternate space		
STDF	Store double floating-point	1	
STDFA	Store double floating-point into alternate space		
STF	Store floating-point	1	
STFA	Store floating-point into alternate space		
STFSR	Store floating-point state register	1-8	pre-sync to previous FGU op from that thread
STH	Store halfword	1	
STHA	Store halfword into alternate space		
STPARTIALF	Eight 8-bit/4 16-bit/2 32-bit partial stores	1	
STW	Store word	1	
STWA	Store word into alternate space		
STX	Store extended	1	
STXA	Store extended into alternate space	variable	
STXFSR	Store extended floating-point state register		
SUB (SUBcc)	Subtract (and modify condition codes)	1	
SUBC (SUBCcc)	Subtract with carry (and modify condition codes)	1	
SWAP	Swap integer register with memory	20-30	Done in L2 cache
SWAPA	Swap integer register with memory in alternate space	20-30	Done in L2 cache

TABLE A-1 UltraSPARC T2 Instruction Latencies *(8 of 8)*

Opcode	Description	Latency	Notes
TADDcc (TADDccTV)	Tagged add and modify condition codes (trap on overflow)	1	If no trap, 6 if trap
TSUBcc (TSUBccTV)	Tagged subtract and modify condition codes (trap on overflow)	1	If no trap, 6 if trap
Tcc	Trap on integer condition codes (with 8-bit sw_trap_number, if bit 7 is set, trap to hyperprivileged)	1	If no trap, 6 if trap
UDIV (UDIVcc)	Unsigned integer divide (and modify condition codes)	12-41	
UDIVX	64-bit unsigned integer divide	12-41	
UMUL (UMULcc)	Unsigned integer multiply (and modify condition codes)	5	
WRASI	Write ASI register		
WRASR	Write ancillary state register	variable	
WRCCR	Write condition codes register	25	
WRFPRS	Write floating-point registers state register	25	
WRHPR	Write hyperprivileged register	15	
WRPR	Write privileged register	variable	
WRY	Write Y register	25	
XNOR (XNORcc)	Exclusive-**nor** (and modify condition codes)	1	
XOR (XORcc)	Exclusive-**or** (and modify condition codes)	1	

APPENDIX **B**

IEEE 754 Floating-Point Support

UltraSPARC T2 conforms to the SPARC V9 Appendix B (IEEE Std 754-1985 Requirements for SPARC-V9) recommendation.

Note | UltraSPARC T2 detects tininess before rounding.

B.1 Special Operand Handling

The UltraSPARC T2 FGU follows the UltraSPARC I/UltraSPARC II handling of special operands instead of that used in UltraSPARC T1. While UltraSPARC T1 provides full hardware support for subnormal operands and results, UltraSPARC T2 generates an *fp_exception_other* exception (with FSR.ftt = unfinished_FPop) in some cases. In addition, UltraSPARC T2 implements a nonstandard floating-point mode (enabled when FSR.ns = 1), whereas UltraSPARC T1 does not.

The FGU generates +∞, −∞, +largest number, −smallest number (depending on round mode) for overflow cases for multiply, divide, and add operations.

For higher-to-lower precision conversion instructions FdTOs:

- Overflow, underflow, and inexact exceptions can be raised
- Overflow is treated the same way as an unrounded add result: Depending on the round mode, we will either generate the properly signed infinity or largest number.
- Underflow for subnormal or gross underflow results: (see *Subnormal Handling* on page 914).

For conversion to integer instructions {F<s | d>TOi, F<s | d>TOx}: UltraSPARC T2 follows *The SPARC Architecture Manual-Version 9* (appendix B.5, pg 246.

For NaN's: UltraSPARC T2 follows *The SPARC Architecture Manual-Version 9* appendix B.2 (particularly Table 27) and B.5, pg 244-246.

- Please note that Appendix B applies to those instructions listed in IEEE 754 section 5: "All conforming implementations of this standard shall provide operations to add, subtract, multiply, divide, extract the sqrt, find the remainder, round to integer in fp format, convert between different fp formats, convert between fp and integer formats, convert binary<->decimal, and compare. Whether copying without change of format is considered an operation is an implementation option."
- The instructions involving copying/moving of fp data (FMOV, FABS, and FNEG) will follow earlier UltraSPARC implementations by doing the appropriate sign bit transformation but will not cause an invalid exception nor do a rs2 = SNaN to rd = QNaN transformation.
- Following UltraSPARC II/UltraSPARC III implementations, all Fpops as defined in V9 will update cexc. All other instructions will leave cexc unchanged.

The remainder of this section gives examples of special cases to be aware of that could generate various exceptions.

B.1.1 Infinity Arithmetic

Let "num" be defined as unsigned in the following tables.

B.1.1.1 One Infinity Operand Arithmetic

- Do not generate exceptions.

TABLE B-1 One-Infinity Operations That Do Not Generate Exceptions

Cases
+∞ plus +num = +∞ +∞ plus -num = +∞ -∞ plus +num = –∞ -∞ plus -num = –∞
+∞ minus +num = +∞ +∞ minus -num = +∞ -∞ minus +num = –∞ -∞ minus –num = –∞
+∞ multiplied by +num = +∞ +∞ multiplied by -num = –∞ –∞ multiplied by +num = –∞ –∞ multiplied by –num = +∞
+∞ divided by +num = +∞ +∞ divided by –num = –∞ –∞ divided by +num = –∞ –∞ divided by –num = +∞

TABLE B-1 One-Infinity Operations That Do Not Generate Exceptions *(Continued)*

Cases
+num divided by +∞ = +0 +num divided by -∞ = –0 –num divided by +∞ = –0 –num divided by –∞ = +0
FsTOd, FdTOs (+∞) = +∞ FsTOd, FdTOs (–∞) = –∞
sqrt(+∞) = +∞
+∞ divided by +0 = +∞ +∞ divided by –0 = –∞ –∞ divided by +0 = –∞ –∞ divided by –0 = +∞
Any arithmetic operation involving infinity as one operand and a QNaN as the other operand: V9, B.2.2, Table 27. (± ∞) OPERATOR (QNaN2) = QNaN2 (QNaN1) OPERATOR (± ∞) = QNaN1
Compares when other operand is not a NaN treat infinity just like a regular number: +∞ = +∞, +∞ > anything else; –∞ = –∞, –∞ < anything else. Effects following instructions: V9 fp compares (rs1 and/or rs2 could be ± ∞): * FCMPE * FCMP
Compares when other operand is a QNaN, SPARC V9 A.13, B.2.1; fcc value = unordered = 11_2 FCMP(s,d) (± ∞) with (QNaN2) – no invalid exception FCMP(s,d) (QNaN1) with (±∞) – no invalid exception

- Could generate exceptions

TABLE B-2 One Infinity Operations That Could Generate Exceptions

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, Appendix B.5[1]	IEEE_754 7.1	
F<s | d>TOi (+∞) = invalid	IEEE_754 invalid	$2^{31}-1$
F<s | d>TOx (+∞) = invalid	IEEE_754 invalid	$2^{63}-1$
F<s | d>TOi (–∞) = invalid	IEEE_754 invalid	-2^{31}
F<s | d>TOx (–∞) = invalid	IEEE_754 invalid	-2^{63}
V9, B.2.2	IEEE_754 7.1	(No NaN operand result)
sqrt(–∞) = invalid	IEEE_754 invalid	QNaN
+∞ multiplied by +0 = invalid	IEEE_754 invalid	QNaN
+∞ multiplied by –0 = invalid	IEEE_754 invalid	QNaN
–∞ multiplied by +0 = invalid	IEEE_754 invalid	QNaN
–∞ multiplied by –0 = invalid	IEEE_754 invalid	QNaN
V9, B.2.2, Table 27[2] Any arithmetic operation involving infinity as one operand and SNaN as the other operand except copying/moving data	IEEE_754 7.1	(One operand, a SNaN)
(± ∞) OPERATOR (SNaN2)	IEEE_754 invalid	QSNaN2
(SNaN1) OPERATOR (± ∞)	IEEE_754 invalid	QSNaN1
V9, A.13, B.2.1[2] Any compare operation involving infinity as one operand and a SNaN as the other operand:	IEEE_754 7.1	
FCMP<s | d> (± ∞) with (SNaN2)	IEEE_754 invalid	fcc value = unordered = 11_2
FCMP<s | d> (SNaN1) with (± ∞)	IEEE_754 invalid	fcc value = unordered = 11_2
FCMPE<s | d> (± ∞) with (SNaN2)	IEEE_754 invalid	fcc value = unordered = 11_2
FCMPE<s | d> (SNaN1) with (± ∞)	IEEE_754 invalid	fcc value = unordered = 11_2
V9, A.13[2] Any compare & generate exception operation involving infinity as 1 operand and a QNaN as the other operand:	IEEE_754 7.1	
FCMPE<s | d> (± ∞) with (QNaN2)	IEEE_754 invalid	fcc value = unordered = $2'b11_2$
FCMPE<s | d> (QNaN1) with (± ∞)	IEEE_754 invalid	fcc value = unordered = $2'b11_2$

1. Similar invalid exceptions also included in SPARC V9 B.5 are generated when the source operand is a NaN(QNaN or SNaN) or a resulting number that cannot fit in 32-bit[64-bit] integer format: (large positive argument $\geq 2^{31}[2^{63}]$ or large negative argument $\leq -(2^{31}+1)[-(2^{63}+1)]$

2. Note that in the IEEE 754 standard, infinity is an exact number; so this exception could also applies to non-infinity operands as well. Also note that the invalid exception and SNaN to QNaN transformation does not apply to copying/moving fpops (FMOV, FABS, FNEG).

B.1.1.2 Two Infinity Operand Arithmetic

- Do not generate exceptions

TABLE B-3 Two Infinity Operations That Do Not Generate Exceptions

Cases
+∞ plus +∞ = +∞ –∞ plus –∞ = –∞
+∞ minus –∞ = +∞ –∞ minus +∞ = –∞
+∞ multiplied by +∞ = +∞ +∞ multiplied by –∞ = –∞ –∞ multiplied by +∞ = –∞ –∞ multiplied by –∞ = +∞
Compares treat infinity just like a regular number: +∞ = +∞, +∞ > anything else; –∞ = –∞, –∞ < anything else. Affects following instructions: V9 fp compares (rs1 and/or rs2 could be ± ∞): * FCMPE * FCMP

■ Could generate exceptions

TABLE B-4 Two Infinity Operations That Generate Exceptions

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, B.2.2	IEEE_754 7.1	(No NaN operand result)
+∞ plus –∞ = invalid	IEEE_754 invalid	QNaN
–∞ plus +∞ = invalid	IEEE_754 invalid	QNaN
+∞ minus +∞ = invalid	IEEE_754 invalid	QNaN
–∞ minus –∞ = invalid	IEEE_754 invalid	QNaN
V9, B.2.2	IEEE_754 7.1	(No NaN operand result)
+∞ divided by +∞ = invalid	IEEE_754 invalid	QNaN
+∞ divided by –∞ = invalid	IEEE_754 invalid	QNaN
–∞ divided by +∞ = invalid	IEEE_754 invalid	QNaN
–∞ divided by –∞ = invalid	IEEE_754 invalid	QNaN

B.1.2 Zero Arithmetic

TABLE B-5 Zero Arithmetic Operations that generate exceptions

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, B.2.2 & 5.1.7.10.4	IEEE_754 7.1	(No NaN operand
+0 divided by +0 = invalid	IEEE_754 invalid	result)
+0 divided by –0 = invalid	IEEE_754 invalid	QNaN
–0 divided by +0 = invalid	IEEE_754 invalid	QNaN
–0 divided by –0 = invalid	IEEE_754 invalid	QNaN
		QNaN
V9. 5.1.7.10.4	IEEE_754 7.2	
+num divided by +0 = divide by zero	IEEE_754 div_by_zero	+∞
+num divided by –0 = divide by zero	IEEE_754 div_by_zero	–∞
–num divided by +0 = divide by zero	IEEE_754 div_by_zero	–∞
–num divided by –0 = divide by zero	IEEE_754 div_by_zero	+∞
V9, B.2.2 Table 27[1]	IEEE_754 7.1	(One operand, a SNaN)
Any arithmetic operation involving zero as 1 operand and a SNaN as the other operand except copying/moving data	IEEE_754 invalid	QSNaN2
(± 0) OPERATOR (SNaN2)	IEEE_754 invalid	QSNaN1
(SNaN1) OPERATOR (± 0)		

1.In this context, 0 is again another exact number; so this exception could also applies to non-zero operands as well. Also note that the invalid exception and SNaN to QNaN transformation does not apply to Copying/moving data instructions (FMOV, FABS, FNEG)

TABLE B-6 Interesting Zero Arithmetic Sign Result Case

Cases
+0 plus –0 = +0 for all round modes except round to –infinity where the result is –0.
sqrt (–0) = –0

B.1.3 NaN Arithmetic

- Do not generate exceptions

TABLE B-7 NaN Arithmetic Operations That Do Not Generate Exceptions

Cases
V9, B.2.1: Fp convert to wider NaN transformation FsTOd (QNaN2) = QNaN2 widened FsTOd (7FD1 0000_{16}) = 7FFA 2000 0000 0000_{16} FsTOd (FFD1 0000_{16}) = FFFA 2000 0000 0000_{16}
V9, B.2.1: Fp convert to narrower NaN transformation FdTOs (QNaN2) = QNaN2 narrowed FdTOs (7FFA 2000 0000 0000_{16}) = 7FD 1000_{16} FdTOs (FFFA 2000 0000 0000_{16}) = FFD 1000_{16}
V9, B.2.2 Table 27 Any noncompare arithmetic operations. Result takes sign of QNaN pass through operand. [Note this rule is applicable to sqrt(QNaN2) = QNaN2 as well]. (± num) OPERATOR (QNaN2) = QNaN2 (QNaN1) OPERATOR (± num) = QNaN1 (QNaN1) OPERATOR (QNaN2) = QNaN2

- Could Generate Exceptions

TABLE B-8 NaN Arithmetic Operations That Could Generate Exceptions

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, B.2.1: Fp convert to wider NaN transformation FsTOd (SNaN2) = QSNaN2 widened	IEEE_754 7.1	
FsTOd (7F91 0000_{16}) = 7FFA 2000 0000 0000_{16} FsTOd (FF91 0000_{16}) = FFFA 2000 0000 0000_{16}	IEEE_754 invalid	QSNaN2 widened
V9, B.2.1: Fp convert to narrower NaN transformation FdTOs (SNaN2) = QSNaN2 narrowed	IEEE_754 7.1	
FdTOs (7FF2 2000 0000 0000_{16}) = 7FD 1000_{16} FdTOs (FFF2 2000 0000 0000_{16}) = FFD 1000_{16}	IEEE_754 invalid	QSNaN2 narrowed

TABLE B-8 NaN Arithmetic Operations That Could Generate Exceptions *(Continued)*

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, B.2.2 Table 27 Any noncompare arithmetic operations except copying/moving (FMOV, FABS, FNEG) [Note this rule applies to sqrt(SNaN2) = QNaN2 and invalid exception as well]	IEEE_754 7.1	
(± num) OPERATOR (SNaN2)	IEEE_754 invalid	QSNaN2
(SNaN1) OPERATOR (± num)	IEEE_754 invalid	QSNaN1
(SNaN1) OPERATOR (SNaN2)	IEEE_754 invalid	QSNaN2
(QNaN1) OPERATOR (SNaN2)	IEEE_754 invalid	QSNaN2
(SNaN1) OPERATOR (QNaN2)	IEEE_754 invalid	QSNaN1
V9, Appendix B.5	IEEE_754 7.1	
F<s|d>TOi (+QNaN) = invalid	IEEE_754 invalid	$2^{31}-1$
F<s|d>TOi (+SNaN) = invalid	IEEE_754 invalid	$2^{31}-1$
F<s|d>TOx (+QNaN) = invalid	IEEE_754 invalid	$2^{63}-1$
F<s|d>TOx (+SNaN) = invalid	IEEE_754 invalid	$2^{63}-1$
F<s|d>TOi (–QNaN) = invalid	IEEE_754 invalid	-2^{31}
F<s|d>TOi (–SNaN) = invalid	IEEE_754 invalid	-2^{31}
F<s|d>TOx (–QNaN) = invalid	IEEE_754 invalid	-2^{63}
F<s|d>TOx (–SNaN) = invalid	IEEE_754 invalid	-2^{63}

B.1.4 Special Inexact Exceptions

UltraSPARC T2 follows SPARC V9 5.1.7.10.5 (IEEE_754 Section 7.5) and sets FSR_inexact whenever the rounded result of an operation differs from the infinitely precise unrounded result.

Additionally, there are a few special cases to be aware of:

TABLE B-9 Fp ↔ Int Conversions With Inexact Exceptions

Cases	Possible Exception	Result (in addition to accrued exception) if tem is cleared
V9, A.14: Fp convert to 32-bit integer when source operand lies between $-(2^{31}-1)$ and 2^{31} but is not exactly an integer. FsTOi, FdTOi.	IEEE_754 7.5 IEEE_754 inexact	 An integer number
V9, A.14: Fp convert to 64-bit integer when source operand lies between $-(2^{63}-1)$ and 2^{63} but is not exactly an integer. FsTOx, FdTOx.	IEEE_754 7.5 IEEE_754 inexact	 An integer number
V9, A.15: Convert integer to fp format when 32-bit integer source operand magnitude is not exactly representable in single precision (23-bit mantissa). Note, even if the operand is $> 2^{24}-1$, if enough of its trailing bits are zeros, it may still be exactly representable. FiTOs.	IEEE_754 7.5 IEEE_754 inexact	 An SP number
V9, A.15: Convert integer to fp format when 64-bit integer source operand magnitude is not exactly representable in single precision (23-bit mantissa). Note, even if the operand is $> 2^{24}-1$, if enough of its trailing bits are zeros, it may still be exactly representable. FxTOs.	IEEE_754 7.5 IEEE_754 inexact	 An SP number
V9, A.15: Convert integer to fp format when 64-bit integer source operand magnitude is not exactly representable in double precision (52-bit mantissa). Note, even if the operand is $> 2^{53}-1$, if enough of its trailing bits are zeros, it may still be exactly representable. FxTOd.	IEEE_754 7.5 IEEE_754 inexact	 A DP number

B.2 Subnormal Handling

The UltraSPARC T2 FGU follows the UltraSPARC I/UltraSPARC II subnormal handling instead of that used in UltraSPARC T1. While UltraSPARC T1 provides full hardware support for subnormal operands and results, UltraSPARC T2 generates an unfinished_FPop trap type in some cases. In addition, UltraSPARC T2 implements a nonstandard floating-point mode, whereas UltraSPARC T1 does not.

UltraSPARC T2 provides limited subnormal support in hardware when in standard mode (FSR.ns = 0) or interval arithmetic mode (GSR.im = 1) [Note that when GSR.im = 1, regardless of FSR.ns, UltraSPARC T2 operates in standard mode.]:

- UltraSPARC T2 supports full subnormal operand handling for single and double precision fp compares;
- UltraSPARC T2 supports gross underflow results for fp-to-fp conversions from higher to lower precision (FdTOs);
- UltraSPARC T2 supports gross underflow results in hardware for FMUL(s,d) and FDIV(s,d) which gives 90% of the optimal underflow performance at a fraction of the cost to completely support subnormal operands and results;
- For those instructions without any subnormal support, an unfinished trap is taken.

UltraSPARC T2 supports the following in nonstandard mode ((FSR.ns = 1) and (GSR.im = 0)):

- Subnormal operands and results are flushed to zero with the same sign, and execution is allowed to proceed without incurring the performance cost of an unfinished trap.

TABLE B-11 and TABLE B-12 show how each instruction type is explicitly handled.

Handling of the FMUL<s | d>, FDIV<s | d>, FdTOs instructions requires a few additional definitions:

- Let Signr = sign of result, RP = round to +infinity, RM = round to –infinity. Define RND as round mode bits. In standard mode, these can have two different sources:

  When in typical standard mode ((FSR.ns = 0) and (GSR.im = 0)), RND = FSR.rd

  When in interval arithmetic mode (GSR.im = 1), RND = GSR.irnd
- Let E(rs1) = biased exponent of rs1 operand, and E(rs2) = biased exponent of rs2 operand
- Let Er = unnormalized and unrounded biased exponent result

  For FMUL<s | d>: Er = E(rs1) + E(rs2) – EBIAS(P)

  For FDIV<s | d>: Er = E(rs1) – E(rs2) + EBIAS(P) – 1

  For FdTOs: Er = E(rs2) – EBIAS(P_rs2) + EBIAS(P_rd), where P_rs2 is the larger precision of the source and P_rd is the smaller precision of the destination
- Let Ef = final normalized and rounded biased exponent result

- Define constants dependent on precision type (see TABLE B-10)

TABLE B-10 Subnormal Handling Constants Per Destination Precision Type

Precision (P)	Number of exponent field bits	Exponent Bias (EBIAS)	Exponent Max (EMAX)	Exponent Gross Underflow (EGUF)
Single	8	127	255	–25
Double	11	1023	2047	–54

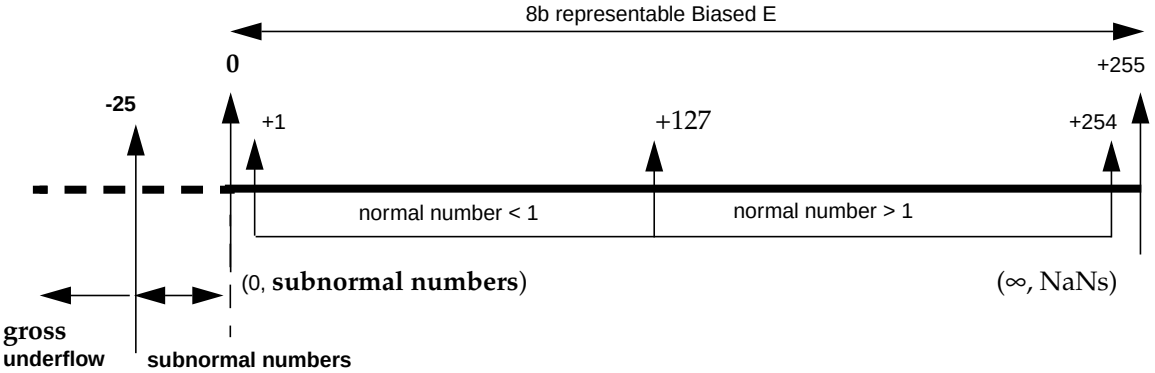


FIGURE B-1 Single Precision Unbiased vs. Biased Subnormals and Gross Underflow

- Note that even though 0 ≤ [E(rs1) or E(rs2)] ≤ 255 for each single precision biased operand exponent, the *computed* Er can be 0 ≤ Er ≤ 255 or can even be negative. For example, for a FMULS instruction:

  If E(rs1) = E(rs2) = +127, then Er = 127 + 127 –127 = +127

  If E(rs1) = E(rs2) = 0, then Er = 0 + 0 - 127 = –127

- Following the sections 5.1.7.6, 5.1.7.8, 5.1.7.9, and figures in 5.1.7.10 of *The SPARC Architecture Manual-Version 9*, Overflow Result is defined as follows:

  > If the appropriate trap enable masks are not set (FSR.ofm = 0 *and* FSR.nxm = 0), then set aexc and cexc overflow and inexact flags: FSR.ofa = 1, FSR.nxa = 1, FSR.ofc = 1, FSR.nxc = 1. No trap is generated.
  >
  > If any or both of the appropriate trap enable masks are set (FSR.ofm = 1 *or* FSR.nxm = 1), then only an IEEE overflow trap is generated: FSR.ftt = 1. The particular cexc bit that is set diverges from previous UltraSPARC I/ UltraSPARC II implementations to follow the SPARC V9 section 5.1.7.9 errata:
  >
  > If FSR.ofm = 0 *and* FSR.nxm = 1, then FSR.nxc = 1.

If FSR.ofm = 1, independent of FSR.nxm, then FSR.ofc = 1 and FSR.nxc = 0.

- Following the sections 5.1.7.6, 5.1.7.8, 5.1.7.9 and figures in 5.1.7.10 of *The SPARC Architecture Manual-Version 9*, Gross Underflow Zero Result is defined as follows:

  Result = 0 (with correct sign)

  If the appropriate trap enable masks are not set (FSR.ufm=0 *and* FSR.nxm = 0), then set aexc and cexc underflow and inexact flags: FSR.ufa = 1, FSR.nxa = 1, FSR.ufc = 1, FSR.nxc = 1. No trap is generated.

  If any or both of the appropriate trap enable masks are set (FSR.ufm = 1 *or* FSR.nxm = 1), then only an IEEE underflow trap is generated: FSR.ftt = 1. The particular cexc bit that is set diverges from UltraSPARC I/ UltraSPARC II implementations to follow the SPARC V9 section 5.1.7.9 errata:

  If FSR.ufm = 0 *and* FSR.nxm = 1, then FSR.nxc = 1.

  If FSR.ufm = 1, independent of FSR.nxm, then FSR.ufc = 1 and FSR.nxc = 0.

- Subnormal handling is overridden for the following cases:
  - Result is a QNaN or SNaN — by *The SPARC Architecture Manual-Version 9* Appendix B.2.2 (Table 27).

    Define "OP_NaN" as instruction uses a SNaN or QNaN operand.

    Examples:

    subnormal + SNaN = QNaN with invalid exception (No unfinished trap in standard mode and no FSR.nx in nonstandard mode)

    subnormal + QNaN = QNaN, no exception (No unfinished trap in standard mode and no FSR.nx in nonstandard mode)

  - Result already generates an exception.

    Define "OP_lt_0" as instruction uses an operand less than zero.

    Examples:

    sqrt(number less than zero) = invalid

  - Result is infinity.

    Define "OP_inf" as instruction uses infinity operand.

    Examples:

    subnormal $+\infty = \infty$ (No unfinished trap in standard mode and no FSR.nx in nonstandard mode)

    subnormal $\times \infty = \infty$ in standard mode; subnormal $\times \infty$ = QNaN with invalid exception in nonstandard mode since subnormal is flushed to zero.

  - Result is zero.

    Define "OP_0" as instruction uses a zero operand

Example:

subnormal × 0 = 0 (No unfinished trap in standard mode, and no FSR.nx in nonstandard mode)

B.2.1 One or Both Subnormal Operands

TABLE B-11 One or Both Subnormal Operands Handling *(1 of 3)*

Instructions	(FSR.ns = 0) *or* (GSR.im = 1) Standard Mode	(FSR.ns = 1) *and* (GSR.im = 0) Nonstandard mode

Single/Double Precision add, subtract [FADD<s|d>, FSUB<s|d>]	**if** (**not** (OP_NaN **or** OP_inf)) {
	generate unfinished trap
	}
		if (**not**(OP_NaN **or** OP_inf)) {
		executes w/subnormal operand flushed to 0 with the same sign
		FSR.nx ← 1
		}

Single/Double Precision FPCOMPARE [FCMP<s|d>, FCMPE<s|d>]	**if** (**not** OP_NaN) {
	execute the compare using the subnormal operand(s)
	}
		if (**not** OP_NaN) {
		executes the compare using the subnormal operand(s)[3]
		}

Single/Double Precision multiply [FMUL<s|d>]	**if** (**not** (OP_NaN **or** OP_inf **or** OP_0)) {
	If ((Er > EGUF(P))
	or
	(Er ≤ EGUF(P) **and** Signr=0 **and** RND=RP)
	or
	(Er ≤ EGUF(P) **and** Signr=1 **and** RND=RM))
	{generate unfinished trap}
	else {
	generate gross underflow zero result[1] }
	}
		if (**not**(OP_NaN **or** OP_0) {
		if (**not**(OP_inf)) {
		executes w/subnormal operand flushed to 0 with the same sign
		FSR.nx ← 1
		} **else** { // 1 op is subnormal, other is ∞
		executes w/subnormal operand flushed to 0 with the same sign
		FSR.nv ← 1 & return QNaN
		}
		}

TABLE B-11 One or Both Subnormal Operands Handling *(Continued)* *(2 of 3)*

Instructions	(FSR.ns = 0) *or* (GSR.im = 1) Standard Mode	(FSR.ns = 1) *and* (GSR.im = 0) Nonstandard mode

Single/double precision divide [FDIV<s|d>]	**if** (**not** (OP_NaN **or** OP_inf **or** OP_0)) {
	  if (**not** (Ef > EMAX(P))) {
	  // not overflow
	  if ((Er > EGUF(P))
	  or
	  (Er ≤ EGUF(P) **and** (Signr=0)
	  and (RND=RP))
	  or
	  (Er ≤ EGUF(P) **and** (Signr=1)
	  and (RND=RM)))
	  {
	  generate unfinished trap
	  } **else** {
	  generate gross-underflow zero result[1]
	  }
	  } **else** {
	  generate overflow result[2]
	  }
	}
		if (**not**(OP_NaN) {
		  if (**not**(OP_inf **or** OP_0) {
		  executes w/subnormal operand flushed to 0 with the same sign
		  if (rs1 and rs2 are flushed to zero) {
		  FSR.nv ← 1 // 0 ÷ 0 is invalid
		  } **else if** (rs2 divisor is flushed to zero) {
		  FSR.dz ← 1
		  } **else** {
		  FSR.nx ← 1
		  }
		  } **else if** (OP_inf) { // 1 op is subnormal,
		  // other = ∞
		  executes w/subnormal operand flushed to 0 with the same sign
		  // 0 ÷ ∞ is 0, and ∞ ÷ 0 is ∞
		  No exceptions are set.
		  Even if divisor is flushed to zero: no need to set FSR.dz
		  } **else if** (OP_0) { // 1 op subnorm, other = 0
		  executes w/subnormal operand flushed to 0 with the same sign
		  // 0 ÷ 0 is invalid exception
		  FSR.nv ← 1 // overrides FSR.dz
		  }
		}

Single to double precision multiply [FSMULD]	**if** (**not** (OP_NaN **or** OP_inf **or** OP_0)) {
	  generate unfinished trap
	}
		if (**not**(OP_NaN **or** OP_0) {
		  if (**not**(OP_inf)) {
		  executes w/subnormal operand flushed to 0 with the same sign
		  Set FSR.nx
		  } **else** { // 1 op is subnormal, other is ∞
		  executes w/subnormal operand flushed to 0 with the same sign
		  Set FSR.nv & return QNaN
		  }
		}

TABLE B-11 One or Both Subnormal Operands Handling *(Continued) (3 of 3)*

Instructions	(FSR.ns = 0) *or* (GSR.im = 1) Standard Mode	(FSR.ns = 1) *and* (GSR.im = 0) Nonstandard mode

Single/Double Precision square root [FSQRT(s,d)] (*note: single operand instruction)	**if** (**not** (OP_NaN **or** OP_inf)) {
	  if (OP_lt_0) {
	  FSR.nv ← 1 & return QNaN
	  } **else** {
	  generate unfinished trap
	  }
	}
		if (**not** (OP_NaN **or** OP_inf)) {
		  executes w/subnormal operand flushed to 0 with the same sign
		  if (OP_lt_0) { //that is, subnormal
		  FSR.nx ← 1; return –0
		  } **else if** (OP_eq_0) { //that is, 0
		  No exception; return 0 with same sign
		  } **else** {
		  FSR.nx ← 1
		  }
		}

Fp to Int and Fp Single to Double conversion [FsTOx, FdTOx, FsTOi, FdTOi, FsTOd] (*note single operand instructions)	**if** (**not** (OP_NaN **or** OP_inf)) {
	  generate unfinished trap
	}
		if (**not** (OP_NaN **or** OP_inf)) {
		  executes w/subnormal operand flushed to 0 with the same sign
		  FSR.nx ← 1
		}

Fp Double to Single conversion [FdTOs] (*note single operand instruction)	**if** (**not** (OP_NaN **or** OP_inf)) {
	  if ((Signr = 0 **and** RND = RP)
	  or
	  (Signr = 1 **and** RND = RM)) {
	  generate unfinished trap
	  } **else** {
	  generate gross underflow zero result[1]
	  }
	}
		if (**not** (OP_NaN **or** OP_inf)) {
		  executes w/subnormal operand flushed to 0 with the same sign
		  FSR.nx ← 1
		}

1. See gross underflow zero result definition on page 917.
2. See overflow definition on page 917.
3. This errata (not flushing subnormal operands to zero for only single and double precision fpcompares) ensures UltraSPARC T2 is backward compatible with UltraSPARC I/UltraSPARC II and UltraSPARC III.

B.2.2 Normal Operand(s) Giving Subnormal Result

TABLE B-12 Subnormal Result Handling for Two Normal Operands

Instructions	(FSR.ns = 0) *or* (GSR.im = 1)	(FSR.ns = 1) *and* (GSR.im = 0)
Single/Double Precision add, subtract [FADD<s|d>), FSUB<s|d>]	Generate unfinished trap	Generate gross underflow zero result[1]

Single/Double Precision multiply [FMUL<s|d>] and divide [FDIV<s|d>], Fp double-to-single conversion [FdTOs] (*note single operand instruction)	**if** (**not** (Ef >= 1)) { // that is, not subnormal intermediate result that rounded to normalized result
	if ((1 > Er > EGUF(P))
	or
	(Er ≤ EGUF(P) **and** Signr = 0 **and** RND = RP)
	or
	(Er ≤ EGUF(P) **and** Signr = 1 **and** RND = RM)) {
	generate unfinished trap
	}
	else {
	generate gross underflow zero result[1]
	}
	} **else** {
	generate normalized result
	}
		Generate gross underflow zero result[1]

1. See gross underflow zero result definition on page 917.

- For those instructions found in TABLE B-11 but not in TABLE B-12, TABLE B-11 is sufficient for their subnormal handling; so the additional rules in TABLE B-12 need not be applied.
- Multiplies that include a conversion from a smaller to larger precision (FSMULD) are not included in TABLE B-12 along with FMUL<s|d> because the larger precision result's exponent range is sufficient to represent a number that would have underflowed in the smaller precision's exponent range.

APPENDIX **C**

Differences From UltraSPARC T1

C.1 General Architectural and Microarchitectural Differences

UltraSPARC T2 follows the CMT philosophy of UltraSPARC T1, but adds more execution capability to each physical core, as well as significant system-on-a-chip components and an enhanced L2 cache. The following lists the microarchitectural differences:

- Physical core consists of two integer execution pipelines and a single floating-point pipeline. UltraSPARC T1 had a single integer execution pipeline and all cores shared a single floating-point pipeline.
- Each physical core in UltraSPARC T2 supports eight strands, which all share the floating-point pipeline. The eight strands are partitioned into two groups of four strands, each of which shares an integer pipeline. UltraSPARC T1 shared the single integer pipeline among four strands.
- Pipeline in UltraSPARC T2 is eight stages, two stages longer than UltraSPARC T1.
- Instruction cache is 8-way associative, compared to 4-way in UltraSPARC T1.
- The L2 cache is 4 Mbyte, 8-banked and 16-way associative, compared to 3 Mbyte, 4-banked and 12-way associative in UltraSPARC T1.
- Data TLB is 128 entries, compared to 64 entries in UltraSPARC T1.
- The memory interface in UltraSPARC T2 supports fully buffered DIMMS (FBDs), providing higher capacity and memory clock rates, as described in Chapter 17.
- The UltraSPARC T2 memory channels support a single-DIMM option for low-cost configurations.
- UltraSPARC T2 supports a pair of on-chip 1-Gbit/10-Gbit Ethernet interfaces, referred to as the network interface unit (NIU). The NIU is described in Chapter 22 through Chapter 28.
- UltraSPARC T2 has an on-chip PCI-Express interface, which replaces the on-chip JBUS interface of UltraSPARC T1. The PCI-Ex Unit is described in Chapter 21.

C.2 ISA Differences

There are a number of ISA differences between UltraSPARC T2 and UltraSPARC T1, as follows:

- UltraSPARC T2 fully supports all VIS 2.0 instructions. UltraSPARC T1 supports a subset of VIS 1.0 plus SIAM (the remainder of VIS 1.0 and 2.0 instructions trap to software for emulation, on UltraSPARC T1).
- UltraSPARC T2 supports the CMP spec, as described in Chapter 14. UltraSPARC T1 has its own version of CMP control/status registers. UltraSPARC T2 consists of eight physical cores, with eight virtual processors per physical core.
- UltraSPARC T2 does not support UltraSPARC T1's `idle` state or its idle, halt, or resume messages. Instead, UltraSPARC T2 supports parking and unparking as specified in the CMP chapter of *UltraSPARC Architecture 2007*. Note that parking is similar to UltraSPARC T1's `idle` state. UltraSPARC T2 does support an equivalent to the `halt` state, which on UltraSPARC T1 was entered via writing to HPR $1E_{16}$. However, UltraSPARC T2 does not support UltraSPARC T1's STRAND_STS_REG ASR, which holds the strand state. Halted state is not software-visible on UltraSPARC T2.
- UltraSPARC T2 does not support the INT_VEC_DIS register (which allowed any UltraSPARC T1 strand to generate an interrupt, reset, idle, or resume message to any strand). Instead, an alias to `ASI_INTR_W` is provided, which allows only the generation of an interrupt to any strand.
- UltraSPARC T2 supports ALLCLEAN, INVALW, NORMALW, OTHERW, POPC, and FSQRT<s|d> in hardware.
- UltraSPARC T2 has a floating-point unit that generates *fp_unfinished_operation* for most denorm cases and supports a nonstandard mode that flushes denorms to zero, as described in Appendix B. UltraSPARC T1 handles denorms in hardware, never generates an *FP_unfinished_operation* trap and does not support a nonstandard mode.
- UltraSPARC T2 generates an *illegal_instruction* trap on any quad-precision FP instruction, while UltraSPARC T1 generates an *fp_exception_other* trap on numeric and move-FP-quad instructions. See Table 5-2 on page 32.
- UltraSPARC T2 generates a *privileged_action* exception upon attempted access to hyperprivileged ASIs by privileged software whereas, in such cases, UltraSPARC T1 takes a *data_access_exception* exception.
- UltraSPARC T2 supports PSTATE.tct; UltraSPARC T1 did not.
- UltraSPARC T2 implements SAVE similar to all previous UltraSPARC processors. UltraSPARC T1 implements a SAVE that updates the locals in the new window to be the same as the locals in the old window, and swaps the *in*s (*out*s) of the old window with the *out*s (*in*s) of the new widow.

- PSTATE.am masking details differ between UltraSPARC T1 and UltraSPARC T2, as described in Section 11.1.8, *Address Masking (Impdep #125)*, on page 95.
- UltraSPARC T2 implements PREFETCH fcn = 18_{16} as a prefetch invalidate cache entry, for efficient software cache flushing, as described in Section 19.15, *L2 Cache Flushing*, on page 437.

C.3 MMU Differences

The UltraSPARC T2 MMU is described in Chapter 12. The UltraSPARC T2 and UltraSPARC T1 MMUs differ as follows:

- UltraSPARC T2 has a 128-entry DTLB, whereas UltraSPARC T1 has a 64-entry DTLB.
- UltraSPARC T2 supports a pair of primary context registers and a pair of secondary context registers. UltraSPARC T1 supports a single primary context and single secondary context register.
- UltraSPARC T2 does not support a locked bit in the TLBs. UltraSPARC T1 supports a locked bit in the TLBs.
- UltraSPARC T2 supports only the sun4v TTE format for I/D-TLB Data-In and Data-Access registers. UltraSPARC T1 supports both the sun4v and the sun4u TTE formats.
- UltraSPARC T2 is compatible with UltraSPARC Architecture 2007 with regard to multiple flavors of data access exception (*DAE_**) and instruction access exception (*IAE_**). UltraSPARC T1 uses the UltraSPARC single flavor of *data_access_exception* and *instruction_access_exception*, indicating the "flavors" in its SFSR register.
- UltraSPARC T2 supports a hardware Table Walker to accelerate ITLB and DTLB miss handling.
- The number and format of TSB configuration and pointer registers differs between UltraSPARC T1 and UltraSPARC T2. UltraSPARC T2 uses physical addresses for TSB pointers, UltraSPARC T1 uses virtual addresses for TSB pointers.
- UltraSPARC T1 and UltraSPARC T2 support the same four page sizes (8 Kbyte, 64 Kbyte, 4 Mbyte, 256 Mbyte). UltraSPARC T2 supports an *unsupported_page_size* trap when an illegal page size is programmed into TSB registers or attempted to be loaded into the TLB. UltraSPARC T1 forces an illegal page size being programmed into TSB registers to be 256 Mbytes and generates a *data_access_exception* trap when a page with an illegal size is loaded into the TLB.
- UltraSPARC T2 adds a demap real operation, which demaps all pages with r = 1 from the TLB.

- UltraSPARC T2 supports an I-TLB probe ASI.
- Autodemapping of pages in the TLBs only demaps pages of the same size or of a larger size in UltraSPARC T2, as described in Section 12.11.3, *I-/D-Demap Context (Type = 1)*, on page 154. In UltraSPARC T1, autodemap demaps pages of the same size, larger size, or smaller size.
- UltraSPARC T2 supports detection of multiple hits in the TLBs as described in Section 16.7.1.1, *ITLB Tag Multiple Hit Error (ITTM)*, on page 221 and Section 16.7.2.1, *DTLB Tag Multiple Hit Error (DTTM)*, on page 224.

C.4 Performance Instrumentation Differences

Both UltraSPARC T1 and UltraSPARC T2 provide access to hardware performance counters through the PIC and PCR registers. However, the events captured by the hardware differ significantly between UltraSPARC T1 and UltraSPARC T2, with UltraSPARC T2 capturing a much larger set of events, as described in Chapter 10. UltraSPARC T2 also has support to count events in hyperprivileged mode; UltraSPARC T1 did not.

In addition, the implementation of *pic_overflow* differs between UltraSPARC T1 and UltraSPARC T2. UltraSPARC T1 provides a disrupting *pic_overflow* on the instruction following the one that caused the overflow event. UltraSPARC T2 provides a disrupting *pic_overflow* on an instruction that generates the event but is within an epsilon number of event-generating instructions from the actual overflow.

Both UltraSPARC T2 and UltraSPARC T1 support DRAM performance counters.

UltraSPARC T2 has performance counters for NIU and PCI-Ex. UltraSPARC T1 has performance counters for JBUS.

C.5 Reset Differences

TBD

C.6 Error Handling Differences

Error handling differs quite a bit between UltraSPARC T2 and UltraSPARC T1. UltraSPARC T1 primarily employs hardware correction of errors, whereas UltraSPARC T2 primarily employs software correction of errors, as described in Chapter 16.

- UltraSPARC T2 uses the *store_error*, *sw_recoverable_error*, *data_access_error*, *instruction_access_error*, *internal_processor_error*, *hw_corrected_error*, *instruction_access_MMU_error*, and *data_access_MMU_error* traps for error handling. UltraSPARC T1 uses the *data_access_error*, *instruction_access_error*, *internal_processor_error*, *hw_corrected_error*, and *data_error* traps.
- UltraSPARC T2 IRF and FRF ECC errors are handled in software. UltraSPARC T1 corrects single-bit transient errors in hardware.
- UltraSPARC T2 has the ability to disable both error reporting and error traps. UltraSPARC T1 only has the ability to disable error traps.
- UltraSPARC T2 takes a deferred *store_error* trap on store buffer uncorrectable errors. UltraSPARC T1 does not have error correction on its store buffers.
- UltraSPARC T2 generates a trap on multiple hits in the ITLB, DTLB, I-cache, or D-cache. UltraSPARC T1 simply uses one of the matching entries.
- UltraSPARC T2 protects its MMU register array with parity, taking a trap if an error is detected during a tablewalk. UltraSPARC T1 MMU registers are not protected by parity. UltraSPARC T2 MMU error handling is described in Section 16.7.1, *ITLB Errors*, on page 221, Section 16.7.2, *DTLB Errors*, on page 223, and Section 16.7.11, *MMU Register Array (MRAU)*, on page 239.
- UltraSPARC T2 protects the TICK (TICK, STICK, HSTICK) compare registers, scratchpad registers, and trap stack registers with SECDED ECC, taking a trap if an error is detected while accessing the registers. UltraSPARC T1 left these registers unprotected by ECC.
- UltraSPARC T2 supports NotData in the L2 cache (NotData is not supported in memory in either UltraSPARC T1 or UltraSPARC T2).
- UltraSPARC T2 protects the vuad bits by SECDED ECC. UltraSPARC T1 protects the vuad bits by parity.
- UltraSPARC T2 has error handling support for the NIU and PCI-Ex units, whereas UltraSPARC T1 has support for error handling from the JBUS unit.

C.7 Power Management Differences

Both UltraSPARC T2 and UltraSPARC T1 support memory access throttling. The mechanisms for supporting CPU throttling differ between UltraSPARC T1 and UltraSPARC T2. UltraSPARC T2 power management is described in Chapter 18.

C.8 Cryptography Unit Differences

Both UltraSPARC T2 and UltraSPARC T1 support a modular arithmetic (MA) unit.

The UltraSPARC T2 MA unit has additional capabilities (not specified in this document) that include a dozen more modular arithmetic operations, stream processing for DES, 3DES, RC4, AES, SHA, MD5, CRC, and TCP/UP/IDP checksumming, and a random number generator.

C.9 Configuration, Diagnostic, and Debug Differences

UltraSPARC T2 configuration and diagnostic support is described in Chapter 28. Debug support is described in Chapter 29. UltraSPARC T2 additions over UltraSPARC T1 include:

- UltraSPARC T2 supports instruction VA watchpoints.
- UltraSPARC T2 supports PA watchpoints.
- UltraSPARC T2 supports the *control_transfer_instruction* trap.
- UltraSPARC T2 implements Prefetch fcn = 18_{16} as a prefetch invalidate cache entry, for efficient software L2 cache flushing. In UltraSPARC T1, flushing of a cache line requires entering "direct-mapped replacement mode," where the L2 LRU is overridden by the address and then forcing out all 12-ways in a set via a displacement with the proper address.
- UltraSPARC T2 supports diagnostic access to the integer register file, store buffers, scratchpad, TICK (TICK, STICK, HSTICK) compare, trap stack, and MMU register arrays.
- UltraSPARC T2 does not require the diagnostic virtual address to match a valid tag for `ASI_DCACHE_DATA`.

APPENDIX **D**

Caches and Cache Coherency

D.1 Cache and Memory Interactions

This appendix describes various interactions between the caches and memory, and the management processes that an operating system must perform to maintain data integrity in these cases. In particular, it discusses the following:

- Invalidation of one or more cache entries—when and how to do it
- Differences between cacheable and noncacheable accesses
- Ordering and synchronization of memory accesses
- Accesses to addresses that cause side effects (I/O accesses)
- Nonfaulting loads
- Cache sizes, associativity, replacement policy, etc.

D.2 Cache Flushing

Data in the level-1 (read-only or writethrough) caches can be flushed by invalidating the entry in the cache (in a way that also leaves the L2 directory in a consistent state). Modified data in the level-2 (writeback) cache must be written back to memory when flushed.

Cache flushing is required in the following cases:

- **I-cache:** Flush is needed before executing code that is modified by a local store instruction. This is done with the FLUSH instruction. Flushing the I-cache with ASI accesses (Section 19.5, *L1 I-Cache Diagnostic Access*, on page 413) does *not* work, because it will leave the I-cache and the L2 directory inconsistent, thus breaking coherency and leading to the possibility of data corruption.

- **D-cache:** Flush is needed when a physical page is changed from (physically) cacheable to (physically) noncacheable. This is done with a displacement flush (*Displacement Flushing*, below).
- **L2 cache:** Flush is needed for stable storage. Examples of stable storage include battery-backed memory and transaction logs. The recommended way to perform this is by using the PrefetchICE instruction (see Section 19.15, *L2 Cache Flushing*, on page 437). Alternatively, this can be done by a displacement flush (see the next section). Flushing the L2 cache flushes the corresponding blocks from the I- and D-caches, because UltraSPARC T2 maintains inclusion between the L2 and L1 caches.

D.2.1 Displacement Flushing

Cache flushing of the L2 cache or the D-cache can be accomplished by a displacement flush. This is done by placing the cache in direct-map mode, and reading a range of read-only addresses that map to the corresponding cache line being flushed, forcing out modified entries in the local cache. Care must be taken to ensure that the range of read-only addresses is mapped in the MMU before starting a displacement flush; otherwise, the TLB miss handler may put new data into the caches. In addition, the range of addresses used to force lines out of the cache must not be present in the cache when starting the displacement flush. (If any of the displacing lines are present before starting the displacement flush, fetching the already present line will *not* cause the proper way in the direct-mapped mode L2 to be loaded; instead, the already present line will stay at its current location in the cache.)

Note | Diagnostic accesses to the L2 cache can be used to invalidate a line, but they are not an alternative to PrefetchICE or displacement flushing. L2 diagnostic accesses do not cause invalidation of L1 lines (breaking L1 inclusion) and modified data in the L2 cache will not be written back to memory using these ASI accesses. See Section 19.16, *L2 Cache Diagnostic Access*, on page 439.

D.2.2 Memory Accesses and Cacheability

Note | Atomic load-store instructions are treated as both a load and a store; they can be performed only in cacheable address spaces.

In UltraSPARC T2, all memory accesses are cached in the L2 cache (as long as the L2 cache is enabled). The cp bit in the TTE corresponding to the access controls whether the memory access will be cached in the primary caches (if cp = 1, the access is cached in the primary caches; if cp = 0 the access is not cached in the primary caches). Atomic operations are always performed at the L2 cache.

D.2.3 Coherence Domains

Two types of memory operations are supported in UltraSPARC T2: cacheable and noncacheable accesses, as indicated by the page translation. Cacheable accesses are inside the coherence domain; noncacheable accesses are outside the coherence domain.

SPARC V9 does not specify memory ordering between cacheable and noncacheable accesses. UltraSPARC T2 maintains TSO ordering, regardless of the cacheability of the accesses, relative to other access by processors. (Ordering of processor accesses relative to DMA accesses roughly follows PCI ordering rules; see *I/O Ordering Rules* on page 942.)

See the *The SPARC Architecture Manual-Version 9* for more information about the SPARC V9 memory models.

On UltraSPARC T2, a MEMBAR `#Lookaside` is effectively a NOP and is not needed for forcing order of stores vs. loads to noncacheable addresses.

D.2.3.1 Cacheable Accesses

Accesses that fall within the coherence domain are called cacheable accesses. They are implemented in UltraSPARC T2 with the following properties:

- Data resides in real memory locations.
- They observe the supported cache coherence protocol.
- The unit of coherence is 64 bytes at the system level (coherence between the virtual processors and I/O), enforced by the L2 cache.
- The unit of coherence for the primary caches (coherence between multiple virtual processors) is the primary cache line size (16 bytes for the data cache, 32 bytes for the instruction cache), enforced by the L2 cache directories.

D.2.3.2 Noncacheable and Side-Effect Accesses

Accesses that are outside the coherence domain are called noncacheable accesses. Accesses of some of these memory (or memory mapped) locations may result in side effects. Noncacheable accesses are implemented in UltraSPARC T2 with the following properties:

- Data may or may not reside in real memory locations.
- Accesses may result in program-visible side effects; for example, memory-mapped I/O control registers in a UART may change state when read.
- Accesses may not observe supported cache coherence protocol.
- The smallest unit in each transaction is a single byte.

Noncacheable accesses are all strongly ordered with respect to other noncacheable accesses (regardless of the e bit). Speculative loads with the e bit set cause a *DAE_so_page* trap.

Note | The side-effect attribute does not imply noncacheability.

D.2.3.3 Global Visibility and Memory Ordering

To ensure the correct ordering between the cacheable and noncacheable domains, explicit memory synchronization is needed in the form of MEMBARs or atomic instructions. CODE EXAMPLE D-1 illustrates the issues involved in mixing cacheable and noncacheable accesses.

CODE EXAMPLE D-1 Memory Ordering and MEMBAR Examples

Assume that all accesses go to non-side-effect memory locations.

```
Process A:
While (1)
{


   Store D1:data produced
1  MEMBAR #StoreStore (needed in PSO, RMO)
   Store F1:set flag
   While F1 is set (spin on flag)
   Load F1
2 MEMBAR #LoadLoad | #LoadStore (needed in RMO)

  Load D2
}

Process B:
While (1)
{

  While F1 is cleared (spin on flag)

      Load F1
2 MEMBAR #LoadLoad | #LoadStore (needed in RMO)

  Load D1

  Store D2
1 MEMBAR #StoreStore (needed in PSO, RMO)

Store F1:clear flag
}
```

Note | A MEMBAR #MemIssue or MEMBAR #Sync is needed if ordering of cacheable accesses following noncacheable accesses must be maintained for RMO cacheable accesses.

Due to load and store buffers implemented in UltraSPARC T2, CODE EXAMPLE D-1 may not work for RMO accesses without the MEMBARs shown in the program segment.

Under TSO, loads and stores (except block stores) cannot pass earlier loads, and stores cannot pass earlier stores; therefore, no MEMBAR is needed.

Under RMO, there is no implicit ordering between memory accesses; therefore, the MEMBARs at both #1 and #2 are needed.

D.2.4 Memory Synchronization: MEMBAR and FLUSH

The MEMBAR (STBAR in SPARC V8) and FLUSH instructions provide for explicit control of memory ordering in program execution. MEMBAR has several variations; their implementations in UltraSPARC T2 are described below. See the references to "Memory Barrier," "The MEMBAR Instruction," and "Programming With the Memory Models," in *The The SPARC Architecture Manual-Version 9* for more information.

D.2.4.1 MEMBAR #LoadLoad

All loads on UltraSPARC T2 switch a strand out until the load completes. Thus, MEMBAR #LoadLoad is treated as a NOP on UltraSPARC T2.

D.2.4.2 MEMBAR #StoreLoad

MEMBAR #StoreLoad forces all loads after the MEMBAR to wait until all stores before the MEMBAR have reached global visibility. MEMBAR #StoreLoad behaves the same as MEMBAR #Sync on UltraSPARC T2.

D.2.4.3 MEMBAR #LoadStore

All loads on UltraSPARC T2 switch a strand out until the load completes. Thus, MEMBAR #LoadStore is treated as a NOP on UltraSPARC T2

D.2.4.4 MEMBAR #StoreStore and STBAR

Stores on UltraSPARC T2 maintain order in the store buffer. Thus Membar #StoreStore is treated as a NOP on UltraSPARC T2.

Notes STBAR has the same semantics as MEMBAR #StoreStore; it is included for SPARC-V8 compatibility.

UltraSPARC T2 block stores and block-init stores are RMO. If a program needs to maintain order between RMO stores to different L2 cache lines, it should use a MEMBAR #Sync.

D.2.4.5 MEMBAR #Lookaside

Loads and stores to noncacheable addresses are "self-synchronizing" on UltraSPARC T2. Thus MEMBAR #Lookaside is treated as a NOP on UltraSPARC T2.

Note For SPARC V9 compatibility, this variation should be used before issuing a load to an address space that cannot be snooped,

D.2.4.6 MEMBAR #MemIssue

MEMBAR #MemIssue forces all outstanding memory accesses to be *completed* before any memory access instruction after the MEMBAR is issued. It must be used to guarantee ordering of cacheable accesses following noncacheable accesses. For example, I/O accesses must be followed by a MEMBAR #MemIssue before subsequent cacheable stores; this ensures that the I/O accesses reach global visibility (as viewed by other strands) before the cacheable stores after the MEMBAR.

Since loads are already self-synchronizing, Membar #MemIssue just needs to drain the store buffer (and receive all the store ACKs) before allowing memory operations to issue again. This is the same operation as UltraSPARC T2's Membar #Sync.

D.2.4.7 MEMBAR #Sync (Issue Barrier)

Membar #Sync forces all outstanding instructions and all deferred errors to be completed before any instructions after the MEMBAR are issued.

Note MEMBAR #Sync is a costly instruction; unnecessary usage may result in substantial performance degradation.

D.2.4.8 Self-Modifying Code (FLUSH)

The SPARC V9 instruction set architecture does not guarantee consistency between code and data spaces. A problem arises when code space is dynamically modified by a program writing to memory locations containing instructions. Dynamic optimizers, LISP programs, and dynamic linking require this behavior. SPARC V9 provides the FLUSH instruction to synchronize instruction and data memory after code space has been modified.

In UltraSPARC T2, FLUSH behaves like a store instruction for the purpose of memory ordering. In addition, all instruction fetch (or prefetch) buffers are invalidated. The issue of the FLUSH instruction is delayed until previous (cacheable) stores are completed. Instruction fetch (or prefetch) resumes at the instruction immediately after the FLUSH.

D.2.5 Atomic Operations

SPARC V9 provides three atomic instructions to support mutual exclusion. These instructions behave like both a load and a store but the operations are carried out indivisibly. Atomic instructions may be used only in the cacheable domain.

An atomic access with a restricted ASI in unprivileged mode (PSTATE.priv = 0) causes a *privileged_action* trap. An atomic access with a noncacheable address causes a *DAE_nc_page* trap. An atomic access with an unsupported ASI causes a *DAE_invalid_ASI* trap. TABLE D-1 lists the ASIs that support atomic accesses.

TABLE D-1 ASIs That Support SWAP, LDSTUB, and CAS

ASI Name
ASI_NUCLEUS{_LITTLE}
ASI_AS_IF_USER_PRIMARY{_LITTLE}
ASI_AS_IF_USER_SECONDARY{_LITTLE}
ASI_AS_IF_PRIV_PRIMARY{_LITTLE}
ASI_AS_IF_PRIV_SECONDARY{_LITTLE}
ASI_AS_IF_PRIV_NUCLEUS{_LITTLE}
ASI_PRIMARY{_LITTLE}
ASI_SECONDARY{_LITTLE}
ASI_REAL{_LITTLE}

Notes Atomic accesses with nonfaulting ASIs are not allowed, because these ASIs have the load-only attribute.

For all atomics, allocation is done to the L2 cache only and will invalidate the L1s.

D.2.5.1 SWAP Instruction

SWAP atomically exchanges the lower 32 bits in an integer register with a word in memory. This instruction is issued only after store buffers are empty. Subsequent loads interlock on earlier SWAPs.

D.2.5.2 LDSTUB Instruction

LDSTUB behaves like SWAP, except that it loads a byte from memory into an integer register and atomically writes all 1's (FF_{16}) into the addressed byte.

D.2.5.3 Compare and Swap (CASX) Instruction

Compare-and-swap combines a load, compare, and store into a single atomic instruction. It compares the value in an integer register to a value in memory; if they are equal, the value in memory is swapped with the contents of a second integer register. All of these operations are carried out atomically; in other words, no other memory operation may be applied to the addressed memory location until the entire compare-and-swap sequence is completed.

D.2.6 Nonfaulting Load

A nonfaulting load behaves like a normal load, except that

- It does not allow side-effect access. An access with the e bit set causes a *DAE_so_page* trap.
- It can be applied to a page with the nfo bit set; other types of accesses will cause a *DAE_NFO_page* trap.

Nonfaulting loads are issued with `ASI_PRIMARY_NO_FAULT{_LITTLE}` or `ASI_SECONDARY_NO_FAULT{_LITTLE}`. A store with a `NO_FAULT` ASI causes a *DAE_invalid_ASI* trap.

When a nonfaulting load encounters a TLB miss, the operating system should attempt to translate the page. If the translation results in an error (for example, address out of range), a 0 is returned and the load completes silently.

Typically, optimizers use nonfaulting loads to move loads before conditional control structures that guard their use. This technique potentially increases the distance between a load of data and the first use of that data, to hide latency; it allows for more flexibility in code scheduling. It also allows for improved performance in certain algorithms by removing address checking from the critical code path.

For example, when following a linked list, nonfaulting loads allow the null pointer to be accessed safely in a read-ahead fashion if the operating system can ensure that the page at virtual address 0_{16} is accessed with no penalty. The nfo (nonfault access only) bit in the MMU marks pages that are mapped for safe access by nonfaulting loads but can still cause a trap by other, normal accesses. This allows programmers to trap on wild pointer references (many programmers count on an exception being generated when accessing address 0_{16} to debug code) while benefitting from the acceleration of nonfaulting access in debugged library routines.

D.3 L1 I-Cache

The L1 Instruction cache is 16 Kbytes, physically tagged and indexed, with 32-byte lines, and 8-way associative with random replacement. The format used to index the cache is shown in TABLE D-2.

TABLE D-2 L1 Instruction Cache Addressing

Bit	Field	Description
39:11	tag	Tag for cache line.
10:5	set	Selects cache set containing the cache line.
4:2	instr	Selects 32-bit instruction in cache line.
1:0	—	Always 0 for access to 32-bit instructions.

D.3.1 LFSR Replacement Algorithm

Details TBD.

D.3.2 Direct-Mapped Mode

The I-cache direct-mapped mode (see Section 19.4.1, *ASI_LSU_DIAG_REG*, on page 412) works by forcing all replacements to the "way" identified by bits [13:11] of the virtual address. Since lines already present are not affected but only new lines brought into the cache are affected, it is safe to turn on (or off) the direct-mapped mode at any time.

D.3.3 I-Cache Disable

Clearing the I-cache enable bit (see Section 19.1, *ASI_LSU_CONTROL_REG*, on page 407) stops all accesses to the I-cache for that strand. All fetches will miss, and the returned data will not fill the I-cache. Invalidates will still be serviced while the I-cache is disabled.

D.4 L1 D-Cache

The L1 Data cache is 8 Kbytes, writethrough, physically tagged and indexed, with 16-byte lines, and 4-way associative with true LRU replacement. The format used to index the cache is shown in TABLE D-3.

TABLE D-3 L1 Data Cache Addressing

Bit	Field	Description
39:11	tag	Tag for cache line.
10:4	set	Selects cache set containing the cache line.
3:0	data	Selects data byte(s) in cache line.

D.4.1 LRU Replacement Algorithm

The D-cache replacement algorithm is true least-recently-used (LRU). Six bits are maintained for each cache index.

D.4.2 Direct-Mapped Mode

In direct-mapped mode, the D-cache (see Section 19.4.1, *ASI_LSU_DIAG_REG*, on page 412) works by changing the replacement algorithm from LRU to instead use two bits of index (address[12:11]) to select the "way." Since lines already present are not affected but only new lines brought into the cache are affected, it is safe to turn on (or off) the direct-mapped mode at any time.

Note that if the D-cache is in direct-mapped mode, and a parity error occurs, the way replaced will be the way which experienced the parity error. This overrides the index selected by the address in direct-mapped mode.

D.4.3 D-Cache Disable

The D-cache may be disabled by setting dc = 0 in the ASI_LSU_CONTROL_REGISTER (see Section 19.1, *ASI_LSU_CONTROL_REG*, on page 407). When disabled, accesses to the D-cache behave as follows. A load which hits in the D-cache ignores the cached data, and fetches the data from L2. A load which misses in the cache fetches the data from L2, but does not allocate the line in the data cache. Stores that miss in the data cache never allocate in the data cache (as normal). Stores that hit in the data cache are performed in the L2, then update the data cache (as normal).

Even if the D-cache is disabled, L2 still keeps the D-cache coherent. Invalidations caused by L2 replacements, stores from other cores, or DMA stores from I/O activity which hit data in the D-cache cause those lines to be invalidated.

To get the D-cache fully disabled, the dc bit must be off on all strands in the virtual processor, and the D-cache must be flushed in a way that doesn't bring new lines back in. This can be done by storing (from a different core) to each line that is in the D-cache, or by displacement flushing the L2 cache so that inclusion will force all D-cache lines to be invalidated.

D.5 L2 Cache

The L2 combined instruction/data cache is 4 Mbytes, writeback, physically tagged and indexed, with 64B lines, 8-banked, and 16-way associative with pseudo-LRU replacement. The format used to index the full cache is shown in TABLE D-4.

TABLE D-4 L2 Cache Addressing (8 banks)

Bit	Field	Description
39:18	tag	Tag for cache line.
17:9	set	Selects cache set containing the cache line. Bit positions listed are used when L2_IDX_HASH_EN.enb_hp = 0. If L2_IDX_HASH_EN.enb_hp = 1, the set selection is done using PA{17:13} **xor** PA{32:28}, PA{19:18C **xor** PA{12:11}, PA{10:9}.
8:6	bank	Selects bank containing the cache line.
5:0	data	Selects data byte(s) in the cache line.

UltraSPARC T2 also supports 4-banked and 2-banked modes to assist in the recovery of partially good die. TABLE D-5 and TABLE D-6 show the format used to index the cache in these reduced modes.

TABLE D-5 L2 Cache Addressing (4 banks)

Bit	Field	Description
38:17	tag	Tag for cache line. Bit positions listed are used when L2_IDX_HASH_EN.enb_hp = 0. If L2_IDX_HASH_EN.enb_hp = 1, the tag stored is PA{38:18}, PA{17} **xor** PA{32}.
16:8	set	Selects cache set containing the cache line. Bit positions listed are used when enb_hp=0. If L2_IDX_HASH_EN.enb_hp = 1, the set selection is done using PA{16:13}^PA{31:28}, PA{19:18} **xor** PA{12:11}, PA{10:8}.
7:6	bank	Selects bank containing the cache line.
5:0	data	Selects data byte(s) in the cache line.

TABLE D-6 L2 Cache Addressing (2 banks)

Bit	Field	Description
37:16	tag	Tag for cache line. Bit positions listed are used when L2_IDX_HASH_EN.enb_hp = 0. If L2_IDX_HASH_EN.enb_hp = 1, the tag stored is PA{37:18}, PA{17:16} **xor** PA{32:31}.
15:7	set	Selects cache set containing the cache line. Bit positions listed are used when L2_IDX_HASH_EN.enb_hp = 0. If L2_IDX_HASH_EN.enb_hp = 1, the set selection is done using PA{15:13} **xor** PA{30:28}, PA{19:18} **xor** PA{12:11}, PA{10:7}.
6	bank	Selects bank containing the cache line.
5:0	data	Selects data byte(s) in the cache line.

D.5.1 NRU Replacement Algorithm

A used-bit scheme is used to implement an NRU (Not Recently Used) replacement. The used bit is set each time a cache line is accessed or when initially fetched from memory. If setting the used-bit causes all used bits (at an index) to be set, the remaining used bits are cleared instead.

In addition, each line has an allocate bit (a), which is set while a line is in a multicycle operation. This can be a cache fill, in which case the a bit gets set when the location is allocated, and the a bit gets cleared when the location is filled with memory data. Alternatively, this could be a multipass operation, either an atomic operation or a subword store (which requires read-modify-write); where the a bit is set on the first pass and cleared on the second/final pass. Any line that has the a bit set is ineligible for replacement.

Each L2 bank has a single rotating replacement pointer, which is the "starting point" to find the "way" to replace. On a miss, the L2 looks for the first line at that index with both u bit and a bit clear, starting with the "way" pointed at by the replacement pointer. If all lines have u bit or a bit set, all u bits are cleared and the scan repeated. The replacement pointer is then rotated forward one "way."

Since the replacement pointer is used by all sets of the L2, replacement is somewhat more random than if each set/index had its own replacement pointer. The replacement pointer is incremented on any L2 miss that causes a cache fill (that is, not DMA reads or full-line DMA writes). The replacement pointer is only reset (put to a known state) by POR, warm reset, or debug reset.

Valid bits do not affect the NRU replacement. In normal use, the only case that creates a line marked Invalid is when a WRI transaction (full-line DMA write) hits in the L2. The WRI will invalidate the L1 and L2, but write directly to memory. This is rare enough that it is not worthwhile to take Valid into account in the replacement algorithm.

D.5.2 Directory Coherence

The L2 cache has a directory of all L1 lines, both I-cache and D-cache, implemented as duplicate tags. Thus, the L2 always knows exactly which lines are in which L1 caches, and in which "way" of each cache. When the L1 requests a line from the L2, the virtual processor specifies whether the line will be allocated (put into the cache), and which "way" it will go into.

The L2 to virtual processor (CPX) protocol allows the L2 to issue invalidates to any/all of the cores simultaneously, but only a single invalidation to each core. For this reason, for a given virtual processor, an L1 line is only allowed to be in either I-cache or the D-cache, but not both. The invalidate transaction includes only index, way, and L1-cache (I or D); it does not include the address.

Since the L2 tracks which lines are in which L1 ways, just invalidating an L1 line via `ASI_ICACHE_TAG` or `ASI_DCACHE_TAG` is not safe and can lead to stale data problems and data corruption. The problem occurs if a line is marked invalid, and a subsequent access to the L1-cache refetches the line, but into a different "way." At this time, the L2 directory has the same line in two places in its directory. Later, when the L2 wants to invalidate that address, it gets a double hit on its CAM access, which the logic does not support. (If an L1 line needs to be invalidated, it can be done by injecting an error into its tag, then accessing it. The hardware error handling will invalidate the line, and inform the L2 directory.)

D.5.3 Direct-Mapped Mode

The L2-cache direct-mapped mode (see Section 19.14.1, *L2 Control Register*, on page 432) works by changing the replacement algorithm from NRU to instead use four bits of index (address{21:18}) to select the "way." Since lines already present are not affected but only new lines brought into the cache are affected, it is safe to turn on (or off) the direct-mapped mode at any time.

D.5.4 L2 Cache Disable

The L2 cache disable (see Section 19.14.1, *L2 Control Register*, on page 432) actually disables an L2 bank. Thus, it is recommended that the L2 be flushed first so that modified lines are written back to memory. While an L2 bank is disabled, the cache effectively has only a single line, which is invalidated or written back at the end of the access. Thus, a store will miss to memory, perform the write into the one-line cache, then flush. Then, the next cache access can be started.

The L2 directory is not used while the L2 cache is disabled. Thus, all L1 caches must be disabled and emptied before disabling any (or all) L2 banks.

D.6 I/O Ordering Rules

UltraSPARC T2 supports the PCI ordering rules to a much higher degree than recent SPARC processors. In particular, UltraSPARC T2 supports maintaining the order of DMA completion notification relative to previous DMA writes, where DMA completion notification can be achieved by (1) receiving an interrupt, (2) reading the device's control register (PIO read return), or (3) reading updated device status in memory (DMA write).

UltraSPARC T2 does not support the following PCI ordering rules:

1. DMA reads may pass previous DMA writes, unless they are to the same address. This rule is not needed for Producer/Consumer, and no need has ever been seen for this.
2. DMA read returns may pass previous PIO writes. This means that if a processor issues a PIO store, followed by a memory store, and wants to guarantee that the I/O device sees the PIO store before seeing the updated memory data, the processor must issue an intervening PIO load to the same device.

The complete rules are described in the following two tables.

TABLE D-7 Outbound (Memory-to-I/O) Transaction Ordering

Row pass Column?	PIO Write	PIO Read Request	DMA Read Completion
PIO Write	No	No	Yes
PIO Read Request	No	No	Yes
DMA Read Compl	Yes	Yes	Yes

Notes:

1. PIO requests are ordered relative to each other, as long as they are to the same "device."
2. DMA read completions follow a different path than PIO requests and are thus unordered relative to PIO requests. So if a processor issues a PIO store followed by a memory store, other processors see them in that order, but an I/O device may see the memory store first.

TABLE D-8 Inbound (I/O-to-Memory) Transaction Ordering

Row pass Column?	DMA Write (Col. 2)	JBUS Mondo Interrupt (Col. 3)	DMA Read Request (Col. 4)	PIO Read Completion (Col. 5)
DMA Write (row A)	No	Yes	Yes	Yes
JBUS Mondo Intr (row B)	No	Yes	Yes	Yes
DMA Read Req (row C)	Yes	Yes	Yes	Yes
PIO Read Compl (row D)	No	Yes	Yes	No

APPENDIX **E**

Glossary

This chapter defines concepts and terminology unique to the UltraSPARC T2 implementation. Definitions of terms common to all UltraSPARC Architecture implementations may be found in the Definitions chapter of *UltraSPARC Architecture 2007*.

ALU Arithmetic Logical Unit

architectural state Software-visible registers and memory (including caches).

ARF Architectural register file.

ASI ring A daisy-chained bus connected in a loop fashion that goes through all of the blocks that have structures with diagnostic path or control registers for ASI access.

blocking ASI An ASI access that accesses its ASI register or array location once all older instructions in that strand have retired, no instructions in the other strand can issue, and the store queue, TSW, and LMB are all empty.

Additionally, the snoop pipeline is stalled before the ASI register/array location is accessed.

branch outcome A reference as to whether or not a branch instruction will alter the flow of execution from the sequential path. A taken branch outcome results in execution proceeding with the instruction at the branch target; a not-taken branch outcome results in execution proceeding with the instruction along the sequential path after the branch.

branch resolution A branch is said to be resolved when the result (that is, the branch outcome and branch target address) has been computed and is known for certain. Branch resolution can take place late in the pipeline.

branch target address The address of the instruction to be executed if the branch is taken.

CAM Content addressable memory.

commit An instruction commits when it modifies architectural state.

complex instruction A complex instruction is an instruction that requires the creation of secondary "helper" instructions for normal operation, excluding trap conditions such as spill/fill traps (which use helpers). Refer to*Instruction Latency* on page 896 for a complete list of all complex instructions and their helper sequences.

consistency *See* **coherence**.

CPU Central Processing Unit. A synonym for **virtual processor**.

CSR Control Status register.

DFT Designed for test.

DTLB Data Cache Translation lookaside buffer.

ECC Error correction code.

EXU Execution Unit.

FP Floating point.

helper A helper instruction is generated by the IRU in response to a complex instruction. Helper instructions are not visible to software. Refer to *Instruction Latency* on page 896 [xref OK?] for a complete list of all complex instructions and their helper sequences.

IFU Instruction Fetch Unit.

ITLB Instruction Cache Translation lookaside buffer.

L2C (or L2$) Level 2 cache.

leaf procedure A procedure that is a leaf in the program's call graph; that is, one that does not call (by using CALL or JMPL) any other procedures.

nonblocking ASI A nonblocking ASI access will access its ASI register/array location once all older instructions in that strand have retired, and there are no instructions in the other strand which can issue.

older instruction Refers to the relative fetch order of instructions. Instruction *i* is older than instruction *j* if instruction *i* was fetched before instruction *j*. Data dependencies flow from older instructions to younger instructions, and an instruction can only be dependent upon older instructions.

one hot An *n*-bit binary signal is one hot if and only if $n - 1$ of the bits are each zero and a single bit is a 1.

Page Table Entry (PTE) Describes the virtual-to-physical translation and page attributes for a specific page. A PTE generally means an entry in the page table or in the TLB, but it is sometimes used as an entry in the TSB (translation storage buffer). In general, a PTE contains fewer fields than does a TTE. *See also* TLB and TSB.

PIO Programmed I/O.

PPN Physical Page Number

pr Processor reset.

PTE Page Table Entry.

quadlet

SFAR Synchronous Fault Address register.

SFSR Synchronous Fault Status register.

SIAM Set interval arithmetic mode instruction.

strand identifier (SID) In a processor implementing 2^n strands, the strand identifier is an n-bit value used to uniquely identify each strand. The strand identifier in UltraSPARC T2 is six bits wide.

VPA Virtual page array.

VPN Virtual page number.

younger instruction *See* **older instruction**.

writeback The process of writing a dirty cache line back to memory before it is refilled.

APPENDIX **F**

Bibliography

[contents of this appendix are TBD]

APPENDIX **G**

ECC Codes

G.1 ECC Summary

TABLE G-1 lists the arrays that are protected by ECC.

TABLE G-1 Error Handling

Array	ECC	Size	Instances	Total
L2 Cache Data	32+7 SEC/DED	936 KB	4	3744 KB
L2 Writeback Buffer	32+7 SEC/DED	624 B	4	2496 B
L2 Fill Buffer	32+7 SEC/DED	624 B	4	2496 B
L2 DMA Input Buffer	32+7 SEC/DED	312 B	4	1248 B
L2 Cache Tag	22+6 SEC	42 KB	4	168 KB
FP Register File	32+7 SEC/DED	2560 B	8	20.8 KB
Integer Register File	64+8 SEC/DED	9 KB	8	72 KB
Trap Stack Array (TSA)	67+8 SEC/DED	320 B	16	5120 B
Tick Compare Array (TCA)	67+8 SEC/DED	288 B	8	2304 B
Scratchpad Array (SCA)	67+8 SEC/DED	288 B	16	4608 B
Store Buffer Data (SBD)	32+7 SEC/DED	640 B	8	5120 B

The L2 Writeback Buffer, L2 Fill Buffer, and L2 DMA Buffer contain data that is in the processor being moved to or from the cache. The ECC generation and check blocks were placed to include these buffers, but errors in these buffers are indistinguishable from L2 cache errors.

G.2 IRF ECC Code

TABLE G-2 IRF Check Bit Generation

Check {7:0}	Data{31:0}																															
	31	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**09**	**08**	**07**	**06**	**05**	**04**	**03**	**02**	**01**	**00**
C0	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1
C1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1
C2	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0
C3	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
C4	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
C5	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	0	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1
	Data{63:32}																															
	63	**62**	**61**	**60**	**59**	**58**	**57**	**56**	**55**	**54**	**53**	**52**	**51**	**50**	**49**	**48**	**47**	**46**	**45**	**44**	**43**	**42**	**41**	**40**	**39**	**38**	**37**	**36**	**35**	**34**	**33**	**32**
C0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
C1	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
C2	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1
C3	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
C4	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
C5	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1

1 – Data bit is **xor**ed into calculation of that check bit.

0 – Data bit is not part of the check bit calculation.

The syndrome calculation is the inverse of the checkbit calculation, for synd{6:0}. synd{7}, however, is simply the **xor** of data{63:0} and check{7:0}.

TABLE G-3 Syndrome Table for IRF ECC Code

SYND {7:4} Value	SYND [3:0} Value															
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	ne	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
3	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
4	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
5	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
6	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
7	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
8	C7	C0	C1	0	C2	1	2	3	C3	4	5	6	7	8	9	10
9	C4	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	C5	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
B	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56
C	C6	57	58	59	60	61	62	63	M	M	M	M	M	M	M	M
D	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
E	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
F	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M

ne — No error

C0–C7 — Single bit error on syndrome/check bit of that number.

0–63 — Single bit error on data bit of that number.

U — Uncorrectable double (or 2*n*) bit error.

M — Triple or worse (2*n* + 1) bit error.

G.3 FRF ECC CODE

The floating-point register files use the same 32+7 SEC/DED codes as the L2 Data, except that the FRF is never intentionally marked with a poison indication, and triple or worse ($2n + 1$) bit errors are reported as correctable errors (FRFC). Software will need to examine the syndrome to determine if a FRFC is a triple or worse error.

G.4 TSA, TCA, and SCA ECC Code

Each of these arrays uses the same ECC code. The TCA and SCA wire the four most-significant bits to 0 as they are only 64 data bits wide.

TABLE G-4 TSA, TCA, SCA 68/8 ECC Check Bit Generation

	Data{31:0}																															
Check {7:0}	**31**	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**09**	**08**	**07**	**06**	**05**	**04**	**03**	**02**	**01**	**00**
C0	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1
C1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1
C2	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0
C3	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
C4	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
C5	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	0	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1

TABLE G-4 TSA, TCA, SCA 68/8 ECC Check Bit Generation

Check {7:0}	Data{31:0}																													
	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00
	Data{63:32}																													
	63	62	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32
C0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
C1	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
C2	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1
C3	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
C4	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
C5	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	1

Check {7:0}	Data[67:64]			
	67	66	65	64
C0	0	0	1	0
C1	1	1	0	0
C2	0	0	0	0
C3	1	1	1	1
C4	0	0	0	0
C5	0	0	0	0
C6	1	1	1	1
C7	1	0	0	1

TABLE G-5 Syndrome Table for TSA, TCA, and SCA Data ECC Code

synd {7:4} Value	synd {3:0} Value															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	ne	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
3	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
4	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
5	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
6	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
7	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
8	M	C0	C1	0	C2	1	2	3	C3	4	5	6	7	8	9	10
9	C4	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	C5	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
B	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56
C	C6	57	58	59	60	61	62	63	64	65	66	67	M	M	M	M
D	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
E	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
F	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M

ne — No error

C0–C6 — Single bit error on syndrome/check bit of that number.

0–67 — Single bit error on data bit of that number.

U — Uncorrectable double (or 2*n*) bit error.

M — Triple or worse (2*n* + 1) bit error

G.5 Store Buffer Data Array (SBD), L2 UA, L2 VD, and L2 Data ECC Code

TABLE G-6 L2 Data Check Bit Generation

	data{31:0}																															
check {6:0}	**31**	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**09**	**08**	**07**	**06**	**05**	**04**	**03**	**02**	**01**	**00**
C0	0	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1
C1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1
C2	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0
C3	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
C4	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
C5	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	0	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1

1 – Data bit is **xor**ed into calculation of that check bit.

0 – Data bit is not part of the check bit calculation.

The syndrome calculation is the inverse of the checkbit calculation, for synd{5:0}. synd{6}, however, is simply the **xor** of data{31:0} and check{6:0}.

TABLE G-7 Syndrome Table for L2 Data ECC Code

synd {6:4} Value	synd {3:0} Value															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	ne	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
1	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
2	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
3	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
4	C6	C0	C1	0	C2	1	2	3	C3	4	5	6	7	8	9	10
5	C4	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
6	C5	26	27	28	29	30	31	M	M	M	M	M	M	M	M	M
7	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	N/M

ne — No Error.

C0–C6 — Single bit error on syndrome/check bit of that number.

0–31 — Single bit error on data bit of that number.

N/M — NotData, or triple or worse ($2n + 1$) bit error.

U — Uncorrectable double (or $2n$) bit error.

M — Triple or worse ($2n + 1$) bit error.

G.6 L2 Tag ECC Code

The syndrome is not captured on L2 Tag ECC errors (LTC), so we only document checkbit calculation for L2 tag.

TABLE G-8 L2 Tag Check Bit Generation

	Tag{21:0}																						
Check {5:0}	**21**	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**09**	**08**	**07**	**06**	**05**	**04**	**03**	**02**	**01**	**00**
C0	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1	1	1
C1	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1
C2	1	1	0	0	1	1	0	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1
C3	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0
C4	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
C5	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0

1 – Data bit is **xor**ed into calculation of that check bit.

0 – Data bit is not part of the check bit calculation.

G.7 Memory Extended ECC Support

UltraSPARC T2 supports Extended ECC error correction (QEC/OED) for main memory, where we can correct any error contained within a single memory nibble (4 bits), and detect as uncorrectable any error that is contained within any two nibbles. The ECC coding scheme uses 4-bit words (16 symbols), 128-bit data (16 words), and a 16-bit syndrome (4 words), and uses Galois Field of (2^4) to implement Add/Multiply Operation that is completely inclusive within its field (Definition of Galois Field). It uses 3×4 bit correction code + 1×4 bit detection code (16 bits total) to correct single-nibble errors and detect double-nibble errors. While the addition is a trivial bitwise **xor**, the multiplication is not as straightforward and involves a Modulo multiplication using its field Primitive Polynomial of value 10011. Also, the syndrome or Parity generated is **xor**ed bitwise with the parity of the address ((**xor_bit_vector**(PA{39:9})) **xor** PA{6}) to that location.

G.7.1 Nomenclature and Nibble Order

The data nibbles and check/syndrome nibbles are numbered in a little-endian fashion, so the order as seen by software is:

C3 C2 C1 C0 N31 N30 … N5 N4 N3 N2 N1 N0

so N0 is made up of data{3:0}, N1 is data{7:4}, etc.

The data for the ECC code is referred to either as check nibbles or as syndrome (nibbles), depending on where it is relative to the ECC generation calculation and to the ECC check calculation. After the ECC generation calculation, it is called “check nibbles,” and this is what is stored in physical DRAM. When memory is accessed, the data nibbles and check nibbles are read out of DRAM and run through an ECC check calculation, which produces the “syndrome nibbles,” synd{15:0}.

The check nibbles are not directly accessible by software.

G.7.1.1 External Hardware Bit Order

External to the chip, the data nibbles are numbered in a little-endian fashion, but the check nibbles are numbered big-endian, so the order as seen by an external logic analyzer is:

TABLE 0-1

C0 C1 C2 C3	N31 N30 N29 N28 . . . . . N3 N2 N1 N0
← dramn_cb{15:0} →	←——— dramn_dq{127:0} ———→

So N0 is made up of dq{3:0}, N1 is dq{7:4}, etc. However, the check nibbles are wired in big-endian order by nibble, but little-endian within each check nibble, so c0{3:0} is made up of cb{15:12}, c1{3:0} is cb{11:8}, c2{3:0} is cb{7:4}, and c3{3:0} is cb{3:0}.

G.7.2 Memory ECC Code Description

The calculation for the check nibbles is as follows:

Check Nibble0 (4 bits) ←
(N0 + 2 × N1 + 3 × N2 + 4 × N3 + 5 × N4 + 6 × N5 + 7 × N6 + 8 × N7 +
9 × N8 + A × N9 + B × N10 + C × N11 + D × N12 + E × N13 + F × N14 +
N15 + 2 × N16 + 3 × N17 + 4 × N18 + 5 × N19 + 6 × N20 + 7 × N21 + 8 × N22 +
9 × N23 + A × N24 + B × N25 + C × N26 + D × N27 + E × N28 + F × N29 + N31)
xor (addr_parity **::** addr_parity **::** addr_parity **::** addr_parity)

Check Nibble1 (4 bits) ←
(N0 + N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10 +
N11 + N12 + N13 + N14 + N30 + N31)
xor (addr_parity **::** addr_parity **::** addr_parity **::** addr_parity)

Check Nibble2 (4 bits) ←
(N15 + N16 + N17 + N18 + N19 + N20 + N21 + N22 + N23 + N24 +
N25 + N26 + N27 + N28 + N29 + N30 + N31)

xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

Check Nibble3 (4 bits) ←
(N0 + 9 × N1+ E × N2 + D × N3 + B × N4 + 7 × N5 + 6 × *N6 + F × N7 + 2 × N8 +
C × N9 + 5 × N10+ A × N11 + 4 × N12 + 3 × N13 + 8 × N14 + N15 + 9 × N16 +
E × N17 + D × N18 + B × N19 + 7 × N20 + 6 × N21 + F × N22 + 2 × N23 + C × N24 +
5 × N25 + A × N26+ 4 × N27 + 3 × N28 + 8 × N29 + N30)
xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

The calculation for the syndrome nibbles is similar, but includes the corresponding check nibble:

Syndrome Nibble0 (4 bits) ←
(C0 + N0 + 2 × N1 + 3 × N2 + 4 × N3 + 5 × N4 + 6 × N5 + 7 × N6 + 8 × N7 +
9 × N8 + A × N9 + B × N10 + C × N11 + D × N12 + E × N13 + F × N14 +
N15 + 2 × N16 + 3 × N17 + 4 × N18 + 5 × N19 + 6 × N20 + 7 × N21 + 8 × N22 +
9 × N23 + A × N24 + B × N25 + C × N26 + D × N27 + E × N28 + F × N29 + N31)
xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

Syndrome Nibble1 (4 bits) ←
(C1 + N0 + N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10 +
N11 + N12 + N13 + N14 + N30 + N31)
xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

Syndrome Nibble2 (4 bits) ←
(C2 + N15 + N16 + N17 + N18 + N19 + N20 + N21 + N22 + N23 + N24 +
N25 + N26 + N27 + N28 + N29 + N30 + N31)
xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

Syndrome Nibble3 (4 bits) ←
(C3 + N0 + 9 × N1+ E × N2 + D × N3 + B × N4 + 7 × N5 + 6 × N6 + F × N7 + 2 ×
N8 + C × N9 + 5 × N10+ A × N11 + 4 × N12 + 3 × N13 + 8 × N14 + N15 + 9 × N16 +
E × N17 + D × N18 + B × N19 + 7 × N20 + 6 × N21 + F × N22 + 2 × N23 + C × N24 +
5 × N25 + A × N26+ 4 × N27 + 3 × N28 + 8 × N29 + N30)
xor (addr_parity :: addr_parity :: addr_parity :: addr_parity)

Error correction is accomplished by following equations. If S0, S1, S2, and S3 are the 4 Syndrome nibbles,

Position 0 – 14 (nibble position) ("/" indicates Galois field division, the inverse of Galois field multiplication operation in TABLE G-9 on page 963)

if (S2 = 0 **and** (S1 ≠ 0) **and** (S0 ≠ 0))
then
nibble_to_correct ← ((S0 / S1) – 1);
corrected_data ← S1 + erred_nibble;
endif

Position 15 – 29 (nibble position)

if (S1 = 0 **and** S0 ≠ 0 **and** S2 ≠ 0), then {
    nibble_to_correct ← ((S0 / S2) + 14);
    corrected_data ← S2 + erred_nibble;
endif

Position 30 (nibble position)

if (S0 = 0 **and** S1 ≠ 0 **and** S2 ≠ 0 **and** S1 = S2)
then
    nibble_to_correct ← N30
    corrected_data ← S1+ erred_nibble;
endif

Position 31 (nibble position)

If (S0 ≠ 0 **and** S1 ≠ 0 **and** S2 ≠ 0 **and** S0 = S1 = S2)
then
    nibble_to_correct ← N31;
    corrected_data ← S1+ erred_nibble;
endif

Notes Nibble S3 is not used in correction but only for multiple error detection. Double errors are detected if (1) exactly two of the check-nibbles are non-zero, or (2) all four of the check-nibbles are non-zero, or (3) the nibble position as indicated by S0/S1 or S0/S2 does not match the nibble position as indicated by S3/S1 or S3/S2, or (4) S1 and S2 are non-zero and the non-zero check-nibbles are not all equal.

This memory ECC scheme assumes that memory is implemented with x4 DRAMs. If x8 parts are used, this is effectively a SEC/DED scheme, with some multibit correction, but no Extended ECC survival capability.

G.7.3 Memory Address Parity Protection

Note that address parity is added (**xor**ed) into all of the check bits. Any normal address parity error will be detected as a multiple-nibble uncorrectable error, since all four syndrome nibbles will be all 1's ($FFFF_{16}$) when the data is read back.

Address parity is defined as the **xor** of all the address bits that specify the bank-specific line address, which is (**xor_bit_vector**(PA{39:9}) **xor** PA{6}) for a four-MCU memory system, (**xor_bit_vector**(PA{39:8}) **xor** PA{6}) for a two-MCU memory system, or (**xor**PA{39:6}) for a one-MCU memory system.

G.7.4 Galois Field Multiplication Table

TABLE G-9 Galois Field Multiplication Table, Polynomial 10011

	Multiplicand															
Multiplier	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2	0	2	4	6	8	A	C	E	3	1	7	5	B	9	F	D
3	0	3	6	5	c	F	A	9	B	8	D	E	7	4	1	2
4	0	4	8	c	3	7	B	F	6	2	E	A	5	1	D	9
5	0	5	A	F	7	2	D	8	E	B	4	1	9	C	3	6
6	0	6	C	A	B	D	7	1	5	3	9	F	E	8	2	4
7	0	7	E	9	F	8	1	6	D	A	3	4	2	5	C	B
8	0	8	3	B	6	E	5	D	C	4	F	7	A	2	9	1
9	0	9	1	8	2	B	3	A	4	D	5	C	6	F	7	E
A	0	A	7	D	E	4	9	3	F	5	8	2	1	B	6	C
B	0	B	5	E	A	1	F	4	7	C	2	9	D	6	8	3
C	0	C	B	7	5	9	E	2	A	6	1	D	F	3	4	8
D	0	D	9	4	1	C	8	5	2	F	B	6	3	E	A	7
E	0	E	F	1	D	3	2	C	9	7	6	8	4	A	B	5
F	0	F	D	2	9	6	4	B	1	E	C	3	8	7	5	A

G.7.5 DRAM Syndrome Interpretation

When examining an UltraSPARC T2 DRAM syndrome, first look at TABLE G-10, to find that pattern of zeros and non-zero nibbles in the 16-bit syndrome. If the pattern of the syndrome nibbles is “a0bc” or “ab0c” (only the second or third nibble is zero), two more tables need to be checked to see which nibble is in error (TABLE G-11 or TABLE G-13), and (2) whether there is a multi-nibble error (TABLE G-12 or TABLE G-14).

The other syndrome nibble patterns are fully described in TABLE G-10.

TABLE G-10 Memory Syndrome Summary

S3	S2	S1	S0	Description
0	0	0	0	No error
a	0	b	c	Possible correctable error in N0-N14. See TABLE G-11 and TABLE G-12.
a	b	0	c	Possible correctable error in N15-N29. See TABLE G-13 on page 967 and TABLE G-14 on page 968.
0	d	d	0	Correctable error in N30. Bits to correct are the value “d” in data nibble N30.
0	d	d	d	Correctable error in N31. Bits to correct are the value “d” in data nibble N31.
0	0	0	e	Error in check nibble C0. Bits to correct are the value “e” in check nibble C0.
0	0	e	0	Error in check nibble C1. Bits to correct are the value “e” in check nibble C1.
0	e	0	0	Error in check nibble C2. Bits to correct are the value “e” in check nibble C2.
e	0	0	0	Error in check nibble C3. Bits to correct are the value “e” in check nibble C3.
8_{16}	2_{16}	2_{16}	1_{16}	Poison Indication, or possibly uncorrectable multiple nibble error
F_{16}	F_{16}	F_{16}	F_{16}	Address Parity Error, or possibly uncorrectable multiple nibble error
Other				Uncorrectable multiple nibble error

a, b, c — Non-zero values for syndrome nibbles, potentially different values, or may be same.

d — Non-zero identical values in these three syndrome nibbles.

e — Non-zero value in a single syndrome nibble.

0 — Syndrome value is zero as part of this pattern.

TABLE G-11 Memory Syndrome, Case a0bc, Contents = Nibble in error, Bits in nibble to correct

	synd{3:0}															
synd{7:4}	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1	*	N0 1_{16}	N1 1_{16}	N2 1_{16}	N3 1_{16}	N4 1_{16}	N5 1_{16}	N6 1_{16}	N7 1_{16}	N8 1_{16}	N9 1_{16}	N10 1_{16}	N11 1_{16}	N12 1_{16}	N13 1_{16}	N14 1_{16}
2	*	N8 2_{16}	N0 2_{16}	N7 2_{16}	N1 2_{16}	N10 2_{16}	N2 2_{16}	N9 2_{16}	N3 2_{16}	N12 2_{16}	N4 2_{16}	N11 2_{16}	N5 2_{16}	N14 2_{16}	N6 2_{16}	N13 2_{16}
3	*	N13 3_{16}	N14 3_{16}	N0 3_{16}	N12 3_{16}	N2 3_{16}	N1 3_{16}	N11 3_{16}	N8 3_{16}	N6 3_{16}	N5 3_{16}	N7 3_{16}	N3 3_{16}	N9 3_{16}	N10 3_{16}	N4 3_{16}
4	*	N12 4_{16}	N8 4_{16}	N3 4_{16}	N0 4_{16}	N11 4_{16}	N7 4_{16}	N4 4_{16}	N1 4_{16}	N14 4_{16}	N10 4_{16}	N5 4_{16}	N2 4_{16}	N13 4_{16}	N9 4_{16}	N6 4_{16}
5	*	N10 5_{16}	N4 5_{16}	N13 5_{16}	N9 5_{16}	N0 5_{16}	N14 5_{16}	N3 5_{16}	N6 5_{16}	N11 5_{16}	N1 5_{16}	N8 5_{16}	N12 5_{16}	N5 5_{16}	N7 5_{16}	N2 5_{16}
6	*	N6 6_{16}	N13 6_{16}	N8 6_{16}	N14 6_{16}	N7 6_{16}	N0 6_{16}	N5 6_{16}	N12 6_{16}	N9 6_{16}	N2 6_{16}	N3 6_{16}	N1 6_{16}	N4 6_{16}	N11 6_{16}	N10 6_{16}
7	*	N5 7_{16}	N11 7_{16}	N9 7_{16}	N10 7_{16}	N12 7_{16}	N6 7_{16}	N0 7_{16}	N4 7_{16}	N2 7_{16}	N8 7_{16}	N14 7_{16}	N13 7_{16}	N7 7_{16}	N1 7_{16}	N3 7_{16}
8	*	N14 8_{16}	N12 8_{16}	N1 8_{16}	N8 8_{16}	N5 8_{16}	N3 8_{16}	N10 8_{16}	N0 8_{16}	N13 8_{16}	N11 8_{16}	N2 8_{16}	N7 8_{16}	N6 8_{16}	N4 8_{16}	N9 8_{16}
9	*	N1 9_{16}	N3 9_{16}	N5 9_{16}	N7 9_{16}	N9 9_{16}	N11 9_{16}	N13 9_{16}	N2 9_{16}	N0 9_{16}	N6 9_{16}	N4 9_{16}	N10 9_{16}	N8 9_{16}	N14 9_{16}	N12 9_{16}
A	*	N11 A_{16}	N10 A_{16}	N6 A_{16}	N4 A_{16}	N8 A_{16}	N13 A_{16}	N1 A_{16}	N9 A_{16}	N5 A_{16}	N0 A_{16}	N12 A_{16}	N14 A_{16}	N2 A_{16}	N3 A_{16}	N7 A_{16}
B	*	N4 B_{16}	N9 B_{16}	N14 B_{16}	N6 B_{16}	N1 B_{16}	N12 B_{16}	N7 B_{16}	N13 B_{16}	N10 B_{16}	N3 B_{16}	N0 B_{16}	N8 B_{16}	N11 B_{16}	N2 B_{16}	N5 B_{16}
C	*	N9 C_{16}	N6 C_{16}	N12 C_{16}	N13 C_{16}	N3 C_{16}	N8 C_{16}	N2 C_{16}	N14 C_{16}	N4 C_{16}	N7 C_{16}	N1 C_{16}	N0 C_{16}	N10 C_{16}	N5 C_{16}	N11 C_{16}
D	*	N3 D_{16}	N7 D_{16}	N11 D_{16}	N2 D_{16}	N6 D_{16}	N10 D_{16}	N14 D_{16}	N5 D_{16}	N1 D_{16}	N13 D_{16}	N9 D_{16}	N4 D_{16}	N0 D_{16}	N12 D_{16}	N8 D_{16}
E	*	N2 E_{16}	N5 E_{16}	N4 E_{16}	N11 E_{16}	N14 E_{16}	N9 E_{16}	N8 E_{16}	N10 E_{16}	N7 E_{16}	N12 E_{16}	N13 E_{16}	N6 E_{16}	N3 E_{16}	N0 E_{16}	N1 E_{16}
F	*	N7 F_{16}	N2 F_{16}	N10 F_{16}	N5 F_{16}	N13 F_{16}	N4 F_{16}	N12 F_{16}	N11 F_{16}	N3 F_{16}	N14 F_{16}	N6 F_{16}	N9 F_{16}	N1 F_{16}	N8 F_{16}	N0 F_{16}

* This table doesn't apply. Look up on Table G-10 on page 964.

Ndd / h_{16} — Top identifies which nibble is in error (in decimal format, dd). Bottom identifies error bits within nibble (in hexadecimal format, h).

TABLE G-12 Memory Syndrome, Case a0bc, Contents = synd{15:12} value

	synd{3:0}															
synd{7:4}	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1	*	1_{16}	9_{16}	E_{16}	D_{16}	B_{16}	7_{16}	6_{16}	F_{16}	2_{16}	C_{16}	5_{16}	A_{16}	4_{16}	3_{16}	8_{16}
2	*	4_{16}	2_{16}	D_{16}	1_{16}	A_{16}	F_{16}	B_{16}	9_{16}	8_{16}	5_{16}	7_{16}	E_{16}	3_{16}	C_{16}	6_{16}
3	*	5_{16}	B_{16}	3_{16}	C_{16}	1_{16}	8_{16}	D_{16}	6_{16}	A_{16}	9_{16}	2_{16}	4_{16}	7_{16}	F_{16}	E_{16}
4	*	3_{16}	8_{16}	1_{16}	4_{16}	E_{16}	9_{16}	A_{16}	2_{16}	6_{16}	7_{16}	F_{16}	D_{16}	C_{16}	5_{16}	B_{16}
5	*	2_{16}	1_{16}	F_{16}	9_{16}	5_{16}	E_{16}	C_{16}	D_{16}	4_{16}	B_{16}	A_{16}	7_{16}	8_{16}	6_{16}	3_{16}
6	*	7_{16}	A_{16}	C_{16}	5_{16}	4_{16}	6_{16}	1_{16}	B_{16}	E_{16}	2_{16}	8_{16}	3_{16}	F_{16}	9_{16}	D_{16}
7	*	6_{16}	3_{16}	2_{16}	8_{16}	F_{16}	1_{16}	7_{16}	4_{16}	C_{16}	E_{16}	D_{16}	9_{16}	B_{16}	A_{16}	5_{16}
8	*	C_{16}	6_{16}	4_{16}	3_{16}	D_{16}	2_{16}	E_{16}	8_{16}	B_{16}	F_{16}	9_{16}	1_{16}	5_{16}	7_{16}	A_{16}
9	*	D_{16}	F_{16}	A_{16}	E_{16}	6_{16}	5_{16}	8_{16}	7_{16}	9_{16}	3_{16}	C_{16}	B_{16}	1_{16}	4_{16}	2_{16}
A	*	8_{16}	4_{16}	9_{16}	2_{16}	7_{16}	D_{16}	5_{16}	1_{16}	3_{16}	A_{16}	E_{16}	F_{16}	6_{16}	B_{16}	C_{16}
B	*	9_{16}	D_{16}	7_{16}	F_{16}	C_{16}	A_{16}	3_{16}	E_{16}	1_{16}	6_{16}	B_{16}	5_{16}	2_{16}	8_{16}	4_{16}
C	*	F_{16}	E_{16}	5_{16}	7_{16}	3_{16}	B_{16}	4_{16}	A_{16}	D_{16}	8_{16}	6_{16}	C_{16}	9_{16}	2_{16}	1_{16}
D	*	E_{16}	7_{16}	B_{16}	A_{16}	8_{16}	C_{16}	2_{16}	5_{16}	F_{16}	4_{16}	3_{16}	6_{16}	D_{16}	1_{16}	9_{16}
E	*	B_{16}	C_{16}	8_{16}	6_{16}	9_{16}	4_{16}	F_{16}	3_{16}	5_{16}	D_{16}	1_{16}	2_{16}	A_{16}	E_{16}	7_{16}
F	*	A_{16}	5_{16}	6_{16}	B_{16}	2_{16}	3_{16}	9_{16}	C_{16}	7_{16}	1_{16}	4_{16}	8_{16}	E_{16}	D_{16}	F_{16}

* This table doesn't apply. Look up on Table G-10 on page 964.

h_{16} — Specifies the value synd{15:12} must have in hexadecimal; otherwise, this is a multi-nibble error.

TABLE G-13 Memory Syndrome, Case ab0c, Contents = Nibble in error, Bits in nibble to correct

	synd{3:0}															
synd{11:8}	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1	*	N15 1_{16}	N16 1_{16}	N17 1_{16}	N18 1_{16}	N19 1_{16}	N20 1_{16}	N21 1_{16}	N22 1_{16}	N23 1_{16}	N24 1_{16}	N25 1_{16}	N26 1_{16}	N27 1_{16}	N28 1_{16}	N29 1_{16}
2	*	N23 2_{16}	N15 2_{16}	N22 2_{16}	N16 2_{16}	N25 2_{16}	N17 2_{16}	N24 2_{16}	N18 2_{16}	N27 2_{16}	N19 2_{16}	N26 2_{16}	N20 2_{16}	N29 2_{16}	N21 2_{16}	N28 2_{16}
3	*	N28 3_{16}	N29 3_{16}	N15 3_{16}	N27 3_{16}	N17 3_{16}	N16 3_{16}	N26 3_{16}	N23 3_{16}	N21 3_{16}	N20 3_{16}	N22 3_{16}	N18 3_{16}	N24 3_{16}	N25 3_{16}	N19 3_{16}
4	*	N27 4_{16}	N23 4_{16}	N18 4_{16}	N15 4_{16}	N26 4_{16}	N22 4_{16}	N19 4_{16}	N16 4_{16}	N29 4_{16}	N25 4_{16}	N20 4_{16}	N17 4_{16}	N28 4_{16}	N24 4_{16}	N21 4_{16}
5	*	N25 5_{16}	N19 5_{16}	N28 5_{16}	N24 5_{16}	N15 5_{16}	N29 5_{16}	N18 5_{16}	N21 5_{16}	N26 5_{16}	N16 5_{16}	N23 5_{16}	N27 5_{16}	N20 5_{16}	N22 5_{16}	N17 5_{16}
6	*	N21 6_{16}	N28 6_{16}	N23 6_{16}	N29 6_{16}	N22 6_{16}	N15 6_{16}	N20 6_{16}	N27 6_{16}	N24 6_{16}	N17 6_{16}	N18 6_{16}	N16 6_{16}	N19 6_{16}	N26 6_{16}	N25 6_{16}
7	*	N20 7_{16}	N26 7_{16}	N24 7_{16}	N25 7_{16}	N27 7_{16}	N21 7_{16}	N15 7_{16}	N19 7_{16}	N17 7_{16}	N23 7_{16}	N29 7_{16}	N28 7_{16}	N22 7_{16}	N16 7_{16}	N18 7_{16}
8	*	N29 8_{16}	N27 8_{16}	N16 8_{16}	N23 8_{16}	N20 8_{16}	N18 8_{16}	N25 8_{16}	N15 8_{16}	N28 8_{16}	N26 8_{16}	N17 8_{16}	N22 8_{16}	N21 8_{16}	N19 8_{16}	N24 8_{16}
9	*	N16 9_{16}	N18 9_{16}	N20 9_{16}	N22 9_{16}	N24 9_{16}	N26 9_{16}	N28 9_{16}	N17 9_{16}	N15 9_{16}	N21 9_{16}	N19 9_{16}	N25 9_{16}	N23 9_{16}	N29 9_{16}	N27 9_{16}
A	*	N26 A_{16}	N25 A_{16}	N21 A_{16}	N19 A_{16}	N23 A_{16}	N28 A_{16}	N16 A_{16}	N24 A_{16}	N20 A_{16}	N15 A_{16}	N27 A_{16}	N29 A_{16}	N17 A_{16}	N18 A_{16}	N22 A_{16}
B	*	N19 B_{16}	N24 B_{16}	N29 B_{16}	N21 B_{16}	N16 B_{16}	N27 B_{16}	N22 B_{16}	N28 B_{16}	N25 B_{16}	N18 B_{16}	N15 B_{16}	N23 B_{16}	N26 B_{16}	N17 B_{16}	N20 B_{16}
C	*	N24 C_{16}	N21 C_{16}	N27 C_{16}	N28 C_{16}	N18 C_{16}	N23 C_{16}	N17 C_{16}	N29 C_{16}	N19 C_{16}	N22 C_{16}	N16 C_{16}	N15 C_{16}	N25 C_{16}	N20 C_{16}	N26 C_{16}
D	*	N18 D_{16}	N22 D_{16}	N26 D_{16}	N17 D_{16}	N21 D_{16}	N25 D_{16}	N29 D_{16}	N20 D_{16}	N16 D_{16}	N28 D_{16}	N24 D_{16}	N19 D_{16}	N15 D_{16}	N27 D_{16}	N23 D_{16}
E	*	N17 E_{16}	N20 E_{16}	N19 E_{16}	N26 E_{16}	N29 E_{16}	N24 E_{16}	N23 E_{16}	N25 E_{16}	N22 E_{16}	N27 E_{16}	N28 E_{16}	N21 E_{16}	N18 E_{16}	N15 E_{16}	N16 E_{16}
F	*	N22 F_{16}	N17 F_{16}	N25 F_{16}	N20 F_{16}	N28 F_{16}	N19 F_{16}	N27 F_{16}	N26 F_{16}	N18 F_{16}	N29 F_{16}	N21 F_{16}	N24 F_{16}	N16 F_{16}	N23 F_{16}	N15 F_{16}

* This table doesn't apply. Look up on Table G-10 on page 964.

Ndd / h_{16} — Top identifies which nibble is in error (in decimal format, dd). Bottom identifies error bits within nibble (in hexadecimal format, h).

TABLE G-14 Memory Syndrome, Case ab0c, Contents = synd{15:12} value

	synd{3:0}															
synd{11:8}	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1	*	1_{16}	9_{16}	E_{16}	D_{16}	B_{16}	7_{16}	6_{16}	F_{16}	2_{16}	C_{16}	5_{16}	A_{16}	4_{16}	3_{16}	8_{16}
2	*	4_{16}	2_{16}	D_{16}	1_{16}	A_{16}	F_{16}	B_{16}	9_{16}	8_{16}	5_{16}	7_{16}	E_{16}	3_{16}	C_{16}	6_{16}
3	*	5_{16}	B_{16}	3_{16}	C_{16}	1_{16}	8_{16}	D_{16}	6_{16}	A_{16}	9_{16}	2_{16}	4_{16}	7_{16}	F_{16}	E_{16}
4	*	3_{16}	8_{16}	1_{16}	4_{16}	E_{16}	9_{16}	A_{16}	2_{16}	6_{16}	7_{16}	F_{16}	D_{16}	C_{16}	5_{16}	B_{16}
5	*	2_{16}	1_{16}	F_{16}	9_{16}	5_{16}	E_{16}	C_{16}	D_{16}	4_{16}	B_{16}	A_{16}	7_{16}	8_{16}	6_{16}	3_{16}
6	*	7_{16}	A_{16}	C_{16}	5_{16}	4_{16}	6_{16}	1_{16}	B_{16}	E_{16}	2_{16}	8_{16}	3_{16}	F_{16}	9_{16}	D_{16}
7	*	6_{16}	3_{16}	2_{16}	8_{16}	F_{16}	1_{16}	7_{16}	4_{16}	C_{16}	E_{16}	D_{16}	9_{16}	B_{16}	A_{16}	5_{16}
8	*	C_{16}	6_{16}	4_{16}	3_{16}	D_{16}	2_{16}	E_{16}	8_{16}	B_{16}	F_{16}	9_{16}	1_{16}	5_{16}	7_{16}	A_{16}
9	*	D_{16}	F_{16}	A_{16}	E_{16}	6_{16}	5_{16}	8_{16}	7_{16}	9_{16}	3_{16}	C_{16}	B_{16}	1_{16}	4_{16}	2_{16}
A	*	8_{16}	4_{16}	9_{16}	2_{16}	7_{16}	D_{16}	5_{16}	1_{16}	3_{16}	A_{16}	E_{16}	F_{16}	6_{16}	B_{16}	C_{16}
B	*	9_{16}	D_{16}	7_{16}	F_{16}	C_{16}	A_{16}	3_{16}	E_{16}	1_{16}	6_{16}	B_{16}	5_{16}	2_{16}	8_{16}	4_{16}
C	*	F_{16}	E_{16}	5_{16}	7_{16}	3_{16}	B_{16}	4_{16}	A_{16}	D_{16}	8_{16}	6_{16}	C_{16}	9_{16}	2_{16}	1_{16}
D	*	E_{16}	7_{16}	B_{16}	A_{16}	8_{16}	C_{16}	2_{16}	5_{16}	F_{16}	4_{16}	3_{16}	6_{16}	D_{16}	1_{16}	9_{16}
E	*	B_{16}	C_{16}	8_{16}	6_{16}	9_{16}	4_{16}	F_{16}	3_{16}	5_{16}	D_{16}	1_{16}	2_{16}	A_{16}	E_{16}	7_{16}
F	*	A_{16}	5_{16}	6_{16}	B_{16}	2_{16}	3_{16}	9_{16}	C_{16}	7_{16}	1_{16}	4_{16}	8_{16}	E_{16}	D_{16}	F_{16}

* This table doesn't apply. Look up on Table G-10 on page 964.

h_{16} — Specifies the value synd{15:12} must have in hexadecimal; otherwise, this is a multi-nibble error.

G.8 Data Poisoning

Data poisoning is the practice of marking known corrupt data with bad ECC so that any later access will get an ECC error. This is normally done when data is transferred from one cache or memory to another, to keep track that the data is

corrupted. If data is not marked with bad ECC, this will lead to silent data corruption when an unsuspecting virtual processor or DMA accesses the old stale copy of the data that still has “good” ECC.

G.8.1 ECC Conversion of UEs as Poison Source

Another source of poison is uncorrectable ECC errors that occur when data is being transferred to a structure that has a different ECC (or parity) encoding. Thus, whenever data is transferred between memory, L2, and the L1 caches, and the source data has an uncorrectable error, the destination will be marked as poison, since the ECC code cannot be transferred intact.

G.8.2 Poisoning L1

The L1 caches (I and D) have only parity protection, so poisoning is implemented by marking the corrupt data (not tag) with a parity error (which is indistinguishable from other parity errors, except that there is also an L2 error with the same address). Since data is always transferred into the L1 in 16-byte chunks, poison in the L1 will always be in aligned 16-byte chunks.

Even though the L1 caches never have dirty data, it is still necessary to receive the corrupt data and install it into the cache, in order to maintain consistency between the L1 tags and the L2 directory.

If the L1 just dropped the corrupt data without installing into the cache, there are scenarios where the L1 tags and L2 directory get out of sync and lose coherency, thus causing silent data corruption (because an invalidate didn’t work).

G.8.3 Poisoning L2

L2 poisoning is implemented by flipping "all" of the 7 checkbits foreach 32bit data word that is corrupt, which means a syndrome of $7F_{16}$ is most likely a poison indication.

L2 ECC (and thus poison) has a 4-byte granularity, so every 4-byte word of uncorrectable data that is written into the cache will have a poison indication. However, all transfers out of the cache are done in 16-byte chunks, so DMA reads, L1 misses, and writebacks to memory will all increase poison/error indications to 16-byte granularity to the recipient, but the L2 is unmodified if it keeps the data.

G.8.3.1 Partial Write Details

Subword writes are read-modify-write, with any ECC correction after the read. If the word read has an uncorrectable error, however, the write is aborted, which leaves the original uncorrectable error intact. Thus, subline writes will not convert uncorrectable errors into poison.

G.8.4 Poisoning Memory

Memory poisoning is implemented by flipping four specific checkbits, specifically C{15, 9, 5, 0} which means that a syndrome of 8221_{16} is most likely a poison indication. This syndrome was chosen because no single nibble error can convert a poison syndrome into a correctable error. A side effect of this is that a minimum of a triple nibble error is needed to "accidentally" generate a failing syndrome of 8221_{16}, so the possibility that a syndrome of 8221_{16} was generated by anything but poison is infinitesimal.

Memory ECC (and thus poison) has a 16-byte granularity, so every 16-byte chunk of uncorrectable data that is written to memory will have a poison indication.

G.8.5 Erasing Poison

Poison can be erased by overwriting it with good data, in such a way that no subword L2 writes occur and no writebacks occur to main memory while the poison is partially erased.

Alternatively, `ASI_BLK_INIT_ST` stores can be used to force poison to be overwritten by causing the entire line to be zeroed out if the line was faulted back to memory.

APPENDIX **H**

JTAG (IEEE 1149.1) Scan Interface

H.1 System JTAG Commands

UltraSPARC T2 supports the following JTAG commands. Note that many of the capabilities that are available through JTAG cannot be used in a running system (for example, accessing normal scan registers).

TABLE H-1 Available System JTAG Commands *(1 of 4)*

Command	Encoding	Description
TAP_EXTEST	00_{16}	Selects BOUNDARY_SCAN_REGISTER.
TAP_IDCODE	01_{16}	Selects IDCODE DR.
TAP_SAMPLE_PRELOAD	02_{16}	Selects BOUNDARY_SCAN_REGISTER.
TAP_HIGHZ	03_{16}	Used for MIO I/O cells.
TAP_CLAMP	04_{16}	Used for MIO I/O cells.
TAP_EXTEST_PULSE	05_{16}	IEEE 1149.6 compliant.
TAP_EXTEST_TRAIN	06_{16}	IEEE 1149.6 compliant.
TAP_CREG_ADDR	08_{16}	Stores address to be used for system access to control register.
TAP_CREG_WDATA	09_{16}	Stores data to be used for system access to control register.
TAP_CREG_RDATA	$0A_{16}$	Captures data from system access.
TAP_NCU_WRITE	$0C_{16}$	Initiates write to system control register.
TAP_NCU_READ	$0D_{16}$	Initiates read from system control register.
TAP_NCU_WADDR	$0E_{16}$	Combination of TAP_CREG_ADDR and TAP_NCU_WRITE.
TAP_NCU_WDATA	$0F_{16}$	Combination of TAP_CREG_WDATA and TAP_NCU_WRITE.
TAP_NCU_RADDR	10_{16}	Combination of TAP_CREG_ADDR and TAP_NCU_READ.
TAP_MBIST_CLKSTPEN	13_{16}	Enables clock stop for MBIST via cycle counter.

TABLE H-1 Available System JTAG Commands *(2 of 4)*

Command	Encoding	Description
TAP_MBIST_BYPASS	14_{16}	Selects engines to be excluded from MBIST operation via mbist_bypass data register.
TAP_MBIST_MODE	15_{16}	Specify serial/parallel, diagnostic mode, or bist/bisi modes via mbist_mode data register.
TAP_MBIST_START	16_{16}	Initiate MBIST.
TAP_MBIST_RESULT	18_{16}	Query 2-bit done/fail register: and/or of all 48 JTAG visible MBIST engines.
TAP_MBIST_DIAG	19_{16}	Run MBIST on one array; MBIST engine and arrays are data register.
TAP_MBIST_GETDONE	$1A_{16}$	Query 48-bit done data register, one bit per JTAG visible MBIST engine.
TAP_MBIST_GETFAIL	$1B_{16}$	Query 48-bit fail data register, one bit per JTAG visible MBIST engine.
TAP_DMO_ACCESS	$1C_{16}$	Set DMO mode; enables DMO logic and package pins.
TAP_DMO_CLEAR	$1D_{16}$	Clears DMO mode.
TAP_DMO_CONFIG	$1E_{16}$	Access 32-bit DMO configuration register.
TAP_MBIST_ABORT	$1F_{16}$	Stop any MBIST activity and reset MBIST controls.
TAP_FUSE_READ	28_{16}	Shift out 32 bits selected by ROW_ADDR; selects eFuse DR.
TAP_FUSE_BYPASS_DATA	29_{16}	Provides user data directly to EFU; selects eFuse DR.
TAP_FUSE_BYPASS	$2A_{16}$	Starts EFU control using bypass data provided by user.
TAP_FUSE_ROW_ADDR	$2B_{16}$	Shift in 7-bit row address for EFU access; selects eFuse ROW ADDRESS DR.
TAP_FUSE_COL_ADDR	$2C_{16}$	Shift in 5-bit column address for EFU programming; selects eFuse COLUMN ADDRESS DR.
TAP_FUSE_READ_MODE	$2D_{16}$	Configures EFU with 3 bits for EFU access; selects eFuse READ MODE DR.
TAP_FUSE_DEST_SAMPLE	$2E_{16}$	Samples EFU destination redundancy value from the destination specified.
TAP_FUSE_RVCLR	$2F_{16}$	Access 6-bit redundancy value clear register.
TAP_SPCTHR0_SHSCAN	30_{16}	Samples thread 0 for all available strands.
TAP_SPCTHR1_SHSCAN	31_{16}	Samples thread 1 for all available strands.
TAP_SPCTHR2_SHSCAN	32_{16}	Samples thread 2 for all available strands.
TAP_SPCTHR3_SHSCAN	33_{16}	Samples thread 3 for all available strands
TAP_SPCTHR4_SHSCAN	34_{16}	Samples thread 4 for all available strands.
TAP_SPCTHR5_SHSCAN	35_{16}	Samples thread 5 for all available strands.

TABLE H-1 Available System JTAG Commands *(3 of 4)*

Command	Encoding	Description
TAP_SPCTHR6_SHSCAN	36_{16}	Samples thread 6 for all available strands.
TAP_SPCTHR7_SHSCAN	37_{16}	Samples thread 7 for all available strands.
TAP_L2T_SHSCAN	38_{16}	Samples specified error registers in the 8 L2 Tags.
TAP_CLOCK_SSTOP	40_{16}	Soft stop of clocks; strands only.
TAP_CLOCK_HSTOP	41_{16}	Hard stop of clocks.
TAP_CLOCK_START	42_{16}	Start clocks.
TAP_CLOCK_DOMAIN	43_{16}	Specify entry clock domain for stopping/starting clocks.
TAP_CLOCK_STATUS	44_{16}	2-bit status indicating if clock stop/start routine finished.
TAP_CLOCK_DELAY	45_{16}	7-bits specifying up to 128 cycle delay between successive clk_stop signals.
TAP_CORE_SEL	46_{16}	8-bit register to specify target SPC cores for clock operations.
TAP_DE_COUNT	48_{16}	Access 32-bit debug event counter.
TAP_CYCLE_COUNT	49_{16}	Access 64-bit reset/cycle counter.
TAP_TCU_DCR	$4A_{16}$	Access 3-bit TCU Debug Control register.
TAP_CORE_RUN_STATUS	$4C_{16}$	Access 64-bit CMP Strand Running Status register.
TAP_DOSS_ENABLE	$4D_{16}$	Access 64-bit disable overlap or single step completion.
TAP_DOSS_MODE	$4E_{16}$	Specify either disable overlap or single step mode; 1 = Enable, 0 = Single step if set to 1, disable overlap if set to 0.
TAP_SS_REQUEST	$4F_{16}$	Pulse single-step request signal.
TAP_DOSS_STATUS	50_{16}	1-bit status for disable overlap or single-step completion.
TAP_CS_MODE	51_{16}	Specify cycle-step mode. 1-bit register set to 1 to enable, uses cycle counter for cycle-step operation.
TAP_CS_STATUS	52_{16}	Read 1-bit status indicating cycle stepping has completed.
TAP_L2_ADDR	58_{16}	Load L2 address (to be written to or read from).
TAP_L2_WRDATA	59_{16}	Load L2 write data.
TAP_L2_WR	$5A_{16}$	Initiate write to L2; wrdata to addr.
TAP_L2_RD	$5B_{16}$	Initiate read from L2 at addr and receive L2 data.
TAP_LBIST_START	60_{16}	Initiate Logic BIST.
TAP_LBIST_BYPASS	61_{16}	Bypass Logic BIST for specified strands; 1-bit per strand.
TAP_LBIST_MODE	62_{16}	Control program mode: parallel/serial modes.
TAP_LBIST_ACCESS	63_{16}	Place one Logic BIST controller between TDI-TDO.

TABLE H-1 Available System JTAG Commands *(4 of 4)*

Command	Encoding	Description
TAP_LBIST_GETDONE	64_{16}	Determine if Logic BIST is done across all selected strands.
TAP_LBIST_ABORT	65_{16}	Abort any Logic BIST currently in progress.
TAP_SERSCAN	80_{16}	Access all or 31 internal scan chains (excluding SerDes scan chain); selects internal scan flops as data register.
TAP_CHAINSEL	81_{16}	Select all or one of the 32 chains for serial scan mode using CHAIN SELECT DR.
TAP_MT_ACCESS	82_{16}	Enables Macro Test mode for JTAG scan.
TAP_MT_CLEAR	83_{16}	Clears Macro Test mode.
TAP_MT_SCAN	84_{16}	Similar to TAP_SERSCAN but drive TCK onto clock tree for pulse capture during RTI state.
TAP_STCI_ACCESS	90_{16}	Enables STCI mode.
TAP_STCI_CLEAR	91_{16}	Clears STCI mode.
TAP_JTPOR_ACCESS	$A0_{16}$	Enables JTAG access window during POR sequence.
TAP_JTPOR_CLEAR	$A1_{16}$	Clears JTAG access window during POR sequence.
TAP_JTPOR_STATUS	$A1_{16}$	JTAG access window status. 1_2 if window is active.
TAP_SCKBYP_ACCESS	$A3_{16}$	Enables bypass for SSI lock time counter in NCU for tester.
TAP_SCKBYP_CLEAR	$A4_{16}$	Clears bypass for SSI lock time counter in NCU for system.
TAP_BYPASS	FF_{16}	Selects Bypass register.

The detailed description of these commands and the exact protocol of how to communicate to the JTAG interface (TAP) plus all the other JTAG commands are documented in the UltraSPARC T2 TCU Microarchitecture Specification.

H.2 JTAG CREG Interface

The CREG interface in the TAP allows access to much of the internal state of UltraSPARC T2 with minimal invasiveness. At a high level, the JTAG CREG interface, represented in TAP_CREG or TAP_NCU type JTAG instructions, allows read and write access to any I/O-addressable location in UltraSPARC T2 except 85_{16} TCU registers. This includes 80_{16} NCU; 81_{16} NIU; 83_{16} CCU; 84_{16} MCU; 86_{16} DBG; 88_{16} DMU; 89_{16} RST; 90_{16} ASI; $A0_{16}$–BF_{16} L2; $C0_{16}$–CF_{16} PCIE; and FF_{16} SSI. All 85_{16} TCU registers can be accessed from NCU via UCB. Note that SPC shscan and L2 accesses are done with different JTAG instructions.

H.2.1 I/O Mapped Register Accesses

The CREG interface can be used to read or write any I/O-addressable location in UltraSPARC T2's address space. All accesses are 8-byte accesses.

The Unit Control Bus interface is a protocol for transmission of packets via the NCU between units. It is implemented inside the TCU and allows access via JTAG to I/O-mapped registers. A register's address and data in the case of writes are loaded via JTAG into holding registers in the TCU. The TCU then uses its UCB interface to communicate to the NCU, which puts the new transaction (packet) into the data flow. The interface allows both reading and writing. On UltraSPARC T2, UCB access through the crossbar to the l2 and strands is not available, so access to the L2 is done via a separate interface between the TCU and the SIU.

For a write, a 40-bit address and 64 bits of data must be provided by JTAG to the UCB. For a read, a 40-bit address is needed, with the data received from the NCU captured into a register in the TCU. To implement a read, a sentinel bit is used since the exact timing of the read return is not deterministic. The system is only allowed to have one read outstanding at one time. There is no protection built in against this; adherence is left to the user.

H.2.1.1 JTAG Instructions Used to Access the UCB

The following descriptions are excerpts from the UltraSPARC T1 DFT specification and the UltraSPARC T1 DFT User's Guide but have been ported to UltraSPARC T2.

TAP_CREG_ADDR

Load System Address: Causes a 40-bit address register to become accessible from TDI. The target system address is loaded during shift-DR. On update-DR a transfer occurs from the TCK domain to a 40-bit holding register in the IO CLK domain.In shift-DR, LSB of CSR `addr` is shifted in first and MSB is shifted in last.

TAP_CREG_WDATA

Load System Write Data: Causes a 64-bit data register to become accessible from TDI, into which the data for the specified system address is loaded during shift-DR. On update-DR a transfer occurs from the TCK domain to a 64-bit holding register in the I/O CLK domain. In shift-DR, LSB of write data is shifted in first and MSB is shifted in last.

TAP_CREG_RDATA

Load System Read Data: Causes a 65-bit data register to become accessible from TDO. The 65th bit is used as a sentinel to allow driver software to synchronize with the read operation. While the read is outstanding, the sentinel bit remains zero. Once the NCU has returned valid data, then the read is complete and the sentinel bit is set to 1. To use this, the JTAG is kept in shift-DR and TCK is clocked until the TDO

reads a 1; this indicates the sentinel bit has been set. When the sentinel bit becomes 1, the next 64 bits shifted out are the valid read data.In shift-DR, LSB of CSR data (that is, CSR content) is shifted out first and MSB is shifted out last.

The TCU can only issue a single access at a given time to the NCU. The user is responsible for ensuring that this is the case. Note too that the TCU does not report erroneous reads made to the NCU. Therefore, the driver software should time out on a read, assuming an error if this occurs.

TAP_NCU_WRITE

Initiate Write Transaction: Causes a write transaction to be initiated on update-IR.

TAP_NCU_READ

Initiate Read Transaction: Causes a read transaction to be initiated on update-IR.

TAP_NCU_WADDR

Load System Address and Initiate Write Transaction: Causes a 40-bit address register to become accessible from TDI. The target system address is loaded during shift-DR. On update-DR a transfer occurs from the TCK domain to a 40-bit holding register in the IO_CLK domain. In the cycle after the transfer is complete, the contents of the address register are forwarded to the UCB interface and a write transaction is initiated. This instruction is a combination of TAP_CREG_ADDR and TAP_NCU_WRITE.

TAP_NCU_WDATA

Load Write Data and Initiate Write Transaction: Causes a 64-bit data register to become accessible from TDI, into which the data for the specified system address is loaded during shift-DR. On update-DR a transfer occurs from the TCK domain to a 64-bit holding register in the IO_CLK domain. In the cycle after the transfer is complete, the contents of the address register and data register are forwarded to the UCB interface to initiate a write transaction. This instruction is a combination of TAP_CREG_WDATA and TAP_NCU_WRITE.

TAP_NCU_RADDR

Load System Address and Initiate Read Transaction: Causes a 40-bit address register to become accessible from TDI. The target system address is loaded during shift-DR. On update-DR a transfer occurs from the TCK domain to a 40-bit holding register in the IO_CLK domain. In the cycle after the transfer is complete, the contents of the address register are forwarded to the UCB interface and a read transaction is initiated. This instruction is a combination of TAP_CREG_ADDR and TAP_NCU_READ.

H.2.1.2 Expected Data and Address Format

The data to be written is 64 bits in length. A 40-bit address is also loaded into the UCB address register.

H.2.1.3 Accesses to Unsupported I/O Addresses

It is possible to specify an illegal or unsupported I/O address in the CREG. Most locations will silently drop writes to unsupported addresses and will return NACK on reads to unsupported addresses, which is ignored by the TAP.

H.2.2 TAP Access to CPU ASI Registers

On UltraSPARC T2, JTAG UCB access to the ASI registers mapped to both CPU and L2 are not available. However, access for all CMP data ASI registers mapped to NCU is available via UCB (TAP_CREG_ and TAP_NCU_ instructions).

Threads in each strand are virtual processors; for those CMP registers specifying physical cores, each physical core is assigned 8 bits in a 64-bit register; allowed values are $1111\ 1111_2$ and $0000\ 0000_2$. The assigned 8 bits are 63:56 = strand 7; 55:48 = strand 6; 47:40 = strand 5; 39:32 = strand 4; 31:24 = strand 3; 23:16 = strand 2; 15:8 = strand 1; 7:0 = strand 0.

Refer to section of 24.4 of the PRM for the ASI registers mapped to NCU and the algorithm for mapping ASI accesses to I/O accesses.

H.3 JTAG Access to Memory

On UltraSPARC T2, JTAG access to L2 is done differently from UCB access through crossbar. Access to the L2 is done via a separate interface between the TCU and the SIU, using TAP_L2 JTAG instructions.

H.3.1 JTAG L2 Access Registers

It is possible to write and read the L2 addresses while the chip is running using JTAG. The L2_ADDR register is accessed via TAP_L2_ADDR; the L2_WRITE_DATA register is accessed via TAP_L2_WRDATA; and the L2_READ_DATA register is accessed via TAP_L2_RD as described below. The L2_WRITE_DATA and L2_READ_DATA registers are the same physical register.

JTAG user should provide physical address that is 8 byte aligned. Irrespective of the original content of bit 2, SIU will force bit 2 = 0.

TABLE H-2 L2 Access Registers

Register	JTAG Instruction	Bits 64:1	Bit 0
L2_ADDR{64:0}	TAP_L2_ADDR	bit 64 =1: JTAG access bits 63:57 = 000 0001 for read request bits 63:57 = 000 0010 for write request bits 56:41 = Unused bits 40:1 = Physical address (8-byte boundary)	Ignored
L2_WRITE_DATA{64:0}	TAP_L2_WRDATA	bits 64:1 = 8-bytes of data to write to L2.	Ignored
L2_READ_DATA{64:0}	TAP_L2_RD	bits 64:1 = 8-bytes of data returned from L2.	1 = Data Valid

H.3.1.1 Memory Write

To write the L2, an address and data must be loaded via JTAG using TAP_L2_ADDR and TAP_L2_WRDATA, followed by TAP_L2_WR. When the TAP_L2_WR instruction is active, the `run-test-idle` state (C_{16}) of the TAP state machine is used to transfer the address and data to the L2 and at least 128 TCK clocks must be cycled while in RTI state for the transfer to complete.

H.3.1.2 Memory Read

A read is accomplished by loading an address using TAP_L2_ADDR followed by a TAP_L2_RD. When the TAP_L2_RD instruction is active, only 64 TCK clocks need be cycled while in RTI to transfer the address to the L2. Then, repeated passes through capture-DR and shift-DR should be used to retrieve the data returned by the L2. Valid data is indicated during TAP_L2_RD at TDO in the shift-DR state by the presence of a leading 1 (bit 0 of the 65-bit L2_READ_DATA register); otherwise, another pass through capture-DR should be implemented without intervening visits to `run-test-idle`. Memory read is nondestructive and has no functional side effect.

Further details on the Addr (Header) and Data (Payload) can be found in the SOC RAS specification. Only one write or read may be outstanding at any time. Also, since non-JTAG logic is used, the POR reset sequence should be performed before using this feature (or at least the POR1 section of the reset sequence).

H.4 JTAG Private Instruction Accessible and Software Accessible Registers

This section describes the JTAG private access registers that are also accessible by software through the NCU. Only the minimal subset of the control registers required for chip debug and testing is accessible by software through the NCU. For joint access between JTAG private and software, JTAG private access will have priority. These registers are local TCU CSRs that the NCU UCB can access.These registers cannot be accessed using JTAG UCB protocol (like TAP_CREG or TAP_NCU type JTAG instructions), but can be accessed using JTAG instructions TAP_DE_COUNT, TAP_CYCLE_COUNT, TAP_TCU_DCR. Note that in PRM section 29.4.4, the trigout_reg bit is set when DCR = 100, 101, 110, or 111, so there is not a separate JTAG instruction to set this bit.

Note | The TCU Debug Control register, TCU Cycle Counter register, TCU Debug Event Counter register, and TCU Trigger Output register also fall in this category of registers and are described in section 29.4 of the PRM.

TABLE H-3 MBIST Mode Register (85 0000 0000_{16})

Bit	Field	Initial Value	R/W	Description
63:4	—	X	RO	*Reserved*
3	loop	0	RW	0 = No loop; 1 = Loop.
2	diagnostic	0	RW	Diagnostic mode if set.
1	bisi	0(1 - See Note below)	RW	BISI if 1; BIST if 0.
0	parallel	0	RW	Parallel mode if 1.

TABLE H-4 MBIST Bypass Register (85 0000 0008_{16}

Bit	Field	Initial Value	R/W	Description
63:48	—	X	RO	*Reserved*
47:0	bypass	0_{16} ($FFFFFFFF FFFF_{16}$ - See Note below)	RW	MBIST bypass.

Note | In the case of the Mbist Mode and Bypass Registers, the default value is overwritten during the power on reset sequence.BISI Enable bit of Mbist Mode Register is written as 1_2 by logic during powere on reset sequence and this bit stays as 1_2 until it is programmed otherwise.The value of the MBIST Bypass register will depend on the core and bank available fuse values after POR1; if there is no partial mode, then this register will be all 1's in bits 47:0

TABLE H-5 MBIST Start Register (85 0000 0010_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	X	W	*Reserved*
0	mbist_ start	0	W	Starts MBIST sequence when written to 1.

TABLE H-6 MBIST Abort Register (85 0000 0018_{16})

Bit	Field	Initial Value	R/W	Description
63:1	—	X	W	*Reserved*
0	mbist_abort	0	W	Aborts MBIST sequence when written to 1.

TABLE H-7 MBIST Result Register (85 0000 0020_{16})

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	mbist_done	0	RO	MBIST done.
0	mbist_fail	0	RO	MBIST fail.

TABLE H-8 MBIST Done Register (85 0000 0028_{16})

Bit	Field	Initial Value	R/W	Description
63:48	—	X	RO	*Reserved*
47:0	mbist_done	0_{16}	RO	MBIST Done bits.

TABLE H-9 MBIST Fail Register (85 0000 0030_{16}))

Bit	Field	Initial Value	R/W	Description
63:48	—	X	RO	*Reserved*
47:0	mbist_fail	0_{16}	RO	MBIST Fail bits.

TABLE H-10 MBIST Start WMR Register ($85\ 0000\ 0038_{16}$)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	W	*Reserved*
0	wrm_start	0_{16}	W	Starts MBIST sequence when written to 1, but delayed until after the next warm reset occurs.

TABLE H-11 LBIST Mode Register ($85\ 0000\ 0040_{16}$)

Bit	Field	Initial Value	R/W	Description
63:2	—	X	RO	*Reserved*
1	program	0	RW	Program mode if 1.
0	parallel	0	RW	Parallel mode if 1.

TABLE H-12 LBIST Bypass Register ($85\ 0000\ 0048_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	—	X	RO	*Reserved*
7:0	lbist_bypass	0_{16}	RW	LBIST bypass.

TABLE H-13 LBIST Start Register ($85\ 0000\ 0050_{16}$)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	lbist_start	0	RW	Starts LBIST sequence when written to 1.

TABLE H-14 LBIST Done Register ($85\ 0000\ 0058_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	—	X	RO	*Reserved*
7:0	lbist_done	0_{16}	RW	LBIST done status.

TABLE H-15 CLKSTOP_DELAY Register ($85\ 0000\ 0120_{16}$)

Bit	Field	Initial Value	R/W	Description
63:8	—	X	RO	*Reserved*
7:0	clkstp_delay	0_{16}	RW	Clock stop delay counter.

TABLE H-16 PEU Testconfig Enable Register ($85\ 0000\ 0180_{16}$)

Bit	Field	Initial Value	R/W	Description
63:1	—	X	RO	*Reserved*
0	peu_testcfg_enable	0	RW	Enables test config bus for PEU.

H.5 Shadow Scan Chains

Shadow scan access is provided in UltraSPARC T2 on SPARC physical core and L2 Error registers. During a shadow-scan operation, JTAG is used to capture the desired values into the shadow scan register. The contents are then scanned-out via TDO. Both the core and L2 tag shadow scan registers can only be read; any value scanned into them will be overwritten. Shadow scan is nondestructive and can happen even when the chip is running in functional mode.

H.5.1 SPARC Shadow Scan

Shadow scan for the strands is controlled via JTAG. The contents to be captured in the shadow scan are shown in TABLE H-1.

Each physical UltraSPARC T2 SPARC virtual processor supports the ability to capture a subset of each strand's state for inspection via a shadow scan facility. The shadow scan is invoked by JTAG commands shown in TABLE H-1.

The TCU continually specifies a strand ID to each physical UltraSPARC T2 SPARC core. In response, the physical core atomically captures the state as described in TABLE 29-1 in a scan string. The TCU then accesses the scan string and captures it in a JTAG-visible register for presentation over the JTAG interface.

VA, HPSTATE, PSTATE, and TL are updated dynamically. Thus, they correctly reflect updates after a trap, a DONE, a RETRY, or a software write to `%tl`.

TPC, TT, and TL_FOR_TT, however, are updated only when the strand takes a trap. They are not updated for DONE, RETRY, or software writes to `%tl`. For example, if the processor traps from TL = 0 to TL = 1 to TL = 2 and then uses DONE and/or RETRY to get back to TL = 0, shadow scan will still reflect TT[2], TPC[2], and TL_FOR_TT will still be 2. Similarly, if the processor traps out to TL = 2 and then software writes TL to 1 or 0, shadow scan will still show TT[2], TPC[2], and TL_FOR_TT will still be 2.

On the JTAG bus, bit 117 appears first, followed by bit 116, sequentially down to and including bit 0.

If multiple traps occur after the state was atomically captured into the shadow scan string but while the shadow scan string is being scanned, only the state belonging to the last trap remains to be (potentially) captured on the next capture point. Due to the length of time to perform a shadow scan relative to the time between traps, sampling via shadow scan can miss several traps.

All eight strand shadow scans are scanned serially as one chain, with strand 0 closest to TDI and strand 7 closest to TDO. Any strand marked unavailable in the CMP STRAND_AVAILABLE register will not be included when scanned via TDI to TDO. The shadow scan chain for a given strand is placed in that strand's second scan chain during ATPG test mode; they are accessible otherwise only via JTAG shadow scan instructions (that is, not during JTAG serial scan).

H.5.2 L2 Shadow Scan

Shadow scan for L2 Error registers is controlled via JTAG. The contents to be captured in the shadow scan are listed in TABLE 20-9 on page 452.

All eight L2 shadow scan contents are captured at the same time and are available at TDO with L2T0 first and L2T7 last (closest to TDO). JTAG instructions to support L2 tag shadow scan are shown in TABLE H-1 on page 971.

On the JTAG bus, bit 141 appears first, followed by bit 140, sequentially down to and including bit 0.

H.6 JTAG Memory BIST

The memory BIST or MBIST engines for UltraSPARC T2 are based on the engine used in UltraSPARC T1. In UltraSPARC T2 there are 80 MBIST engines: 3 per virtual processor (24 total) and 56 distributed throughout the SOC logic. Each MBIST engine will therefore test several arrays. However, these 80 MBIST engines are mapped to 48 JTAG-visible engines. Please refer to TCU Microarchitecture Specification for the mapping.

The MBIST operation may be controlled by the TCU during reset sequencing via the JTAG interface or as invoked via software. The MBIST engines can be operated in a serial mode, a parallel mode, or a diagnostic mode for memory bit-fail mapping. Both the serial and parallel modes run MBIST in a pass/fail mode, where the only information available is whether MBIST passed all of its arrays or failed at least one of them.

For more details on UltraSPARC T2's memory BIST architecture and operation, refer to the TCU Microarchitecture Specification.

H.6.1 MBIST Modes

H.6.1.1 Serial Mode

JTAG will typically be used to run MBIST in the serial mode. When activated in serial mode, the MBIST engines will be started sequentially in the order specified in the TCU Microarchitecture Specification.

To enable the serial MBIST mode via JTAG, the instruction TAP_MBIST_BYPASS must be used to specify which of the 48 JTAG Visible MBIST engines to bypass, if any, via the MBIST_BYPASS register. Next, the TAP_MBIST_MODE is used to clear the parallel mode bit in the MBIST_MODE register. The TAP_MBIST_START instruction is then programmed into JTAG; when JTAG enters the `run-test-idle` state, the MBIST operation will be started; it is not necessary to remain in the `run-test-idle` state. It is up to the user to wait a predetermined number of cycles for the MBIST operation for all arrays to finish. Status can be checked using the TAP_MBIST_RESULT instruction and capturing the MBIST_RESULT register (2 bits) in the capture-DR state and examining them; this can be done repeatedly for polling (via capture-DR, without staying in `run-test-idle`). This allows early truncation of the test (via the TAP_MBIST_ABORT instruction) if the fail bit becomes active before the MBIST operation is done. The done bit must be set to validate a fail bit of 0 indicating a passing condition. A done bit set to 1 and a fail bit set to 0 indicates all arrays for the selected MBIST engines passed MBIST.

The default operation is to run BISI instead of MBIST. To run BISI, the instruction TAP_MBIST_DIAG may be used to program a given MBIST engine's config bits. Selection of BISI or MBIST is done by setting the corresponding bit in the MBIST config register of the MBIST engines via the MBIST_BYPASS register or by setting the bisi bit the TAP_MBIST_MODE register.

H.6.1.2 Parallel Mode

JTAG must be used to run MBIST in the parallel mode. When activated in parallel mode, the MBIST engines will be started in parallel, whereas the arrays controlled by each individual MBIST engine will test their arrays sequentially. Operation of MBIST parallel mode via JTAG is similar to the serial mode, except that the parallel

mode bit of the MBIST_MODE register must instead be set, using the TAP_MBIST_MODE instruction. There is no non-JTAG default method of running MBIST in parallel mode.

Note When the serial or parallel MBIST is determined to be finished (via polling/examination of the done/fail register or a timeout), all that is known is that either all arrays passed or at least one of them failed. To get information on which MBIST engine failed, the TAP_MBIST_GETDONE instruction must be used. This allows capture of all 48 JTAG visible done bits which may then be observed at TDO; the TAP_MBIST_GETFAIL similarly captures all 48 JTAG visible fail bits. If detailed information as to which array failed within a given MBIST engine is needed, then the TAP_MBIST_DIAG instruction must be used to retrieve the contents of the specific MBIST engine that indicated a fail.

H.6.1.3 Diagnostic Mode

In one method to perform bit-fail mapping, the TAP_MBIST_DIAG instruction is used to access the MBIST engine as the target JTAG data register. In this diagnostic mode only one MBIST engine should be selected, by setting the appropriate bits in the MBIST_BYPASS register via the TAP_MBIST_BYPASS instruction; it is up to the user to bypass all but one MBIST engine. Only one array controlled by the selected MBIST engine may be active; this is specified by scanning in (loading) the target MBIST engine registers. After both the MBIST engine and array are specified, the TAP_MBIST_START is programmed, and entering `run-test-idle` will start the MBIST operation on the selected array. After an appropriate wait time, the test should finish. Polling via TAP_MBIST_RESULT can be used to inspect the done/fail JTAG data register, or the TAP_MBIST_GETDONE and TAP_MBIST_GETFAIL can be used to determine the MBIST test results.

To get the detailed information on the target array, the TAP_MBIST_DIAG instruction must be used. This allows the contents of the targeted MBIST engine to be scanned as the MBIST_DIAG register via TDO.

H.6.1.4 Abort Mode

To abort any MBIST activity the TAP_MBIST_ABORT instruction should be used. This will cause all MBIST start signals to be deasserted and any internal JTAG states to be reset. A separate instruction is useful since the JTAG MBIST instructions have memory. Use of TAP_MBIST_ABORT does not clear any of the JTAG data registers used for or during MBIST—only the control states and signals—and does not clear the MBIST engine flops; this allows the TAP_MBIST_DIAG to be used to get data on the failing arrays. Entering `test-logic-reset` state will also stop MBIST.

H.6.2 JTAG MBIST Registers

In UltraSPARC T2, JTAG accessible registers for MBIST are as follows.

TABLE H-17 JTAG MBIST Registers

Register	JTAG Instruction	Fields
RESULT{1:0}	TAP_MBIST_RESULT	Bit 1 – 1 when all 48 Jtag visible mbist engines are done. Bit 0 – 1 if any of 48 Jtag visible mbist engines reports a fail.
BYPASS{47:0}	TAP_MBIST_BYPASS	One bit per JTAG visible mbist engine; to bypass an engine during MBIST testing set its bit to 1.
DONE{47:0}	TAP_MBIST_GETDONE	One bit per JTAG visible mbist engine; a 1 indicates the corresponding engine is done; same order as MBIST_BYPASS register.
FAIL{47:0}	TAP_MBIST_GETFAIL	One bit per JTAG visible mbist engine; a 1 indicates the corresponding engine failed MBIST for one of its arrays.
DIAG{k:0}	TAP_MBIST_DIAG	Includes targeted MBIST engines in a cluster; variable length.
MODE{2:0}	TAP_MBIST_MODE	Bit 2 – user mode if set. Bit 1 – bisi mode if set, else bist mode. Bit 0 – parallel mode if 1, serial mode if 0.
None	TAP_MBIST_CLKSTPEN	Enables mbist controller to begin cycle counter; reset with TLR or TAP_CLOCK_START.

H.6.3 MBIST Clock Stop and Scan Dump

The cycle counter may be used in conjunction with MBIST to stop clocks and perform a scan dump. The instruction TAP_MBIST_CLKSTPEN must be programmed to enable the cycle counter for MBIST. If enabled, the cycle counter will begin decrementing when the MBIST controller begins operation. When the cycle counter reaches zero, a hard clock stop will be issued to the clock sequencer.

All relevant registers—clock domain, clock stop delay—will be recognized in this mode to allow control of the clock stop sequence. The clock stop status may be checked with TAP_CLOCK_STATUS, and when stopped the scan chains can be dumped via TAP_SERSCAN.

Using this feature and repeatedly running MBIST with successively greater cycle count values allows another method of bit-fail mapping arrays. This is sometimes referred to as MBIST Plus. Since the start of MBIST and when the cycle counter begins, decrementing is coordinated and synchronized to the same CMP clock cycle the entire process should be repeatable and cycle accurate

H.6.4 MBIST DMO: Direct Memory Observe

For a complete description of DMO, refer to the UltraSPARC T2 DMO specification. There are three JTAG instructions: TAP_DMO_ACCESS, TAP_DMO_CLEAR, and TAP_DMO_CONFIG, as described in TABLE H-1 on page 971. The TAP_DMO_ACCESS puts the chip in DMO mode, so data from BIST run on L2 Tags or certain SPC or NIU arrays are observable at package pins. TAP_DMO_CLEAR clears this mode. To access and program the DMO control logic inside TCU the TAP_DMO_CONFIG instruction should be used to set the 32-bits as desired.

TABLE H-18 JTAG DMO Configuration Register

Register	JTAG Instruction	Bit	Description
DMO_CONFIG {31:0}	TAP_DMO_ CONFIG	31:16	16-bit shift register.
		15	1 selects CMP clock domain, 0 selects I/O clock domain (NIU arrays are in I/O clock domain)
		14:13	00 selects DMO path to strands 4, 5, 1, or 0. 01 selects DMO path to strands 6, 7, 3, or 2; 10 selects DMO path to L2 tags 4, 5, 1, or 0; 11 selects DMO path to L2 tags 6, 7, 3, or 2.
		12:11	00 selects RTX of NIU; 01 selects RDP of NIU; 10 selects TDS of NIU; 11 is not allowed.
		10:8	000 – rtx_txc_txe0_dmo_dout (default) 001 – rtx_txc_txe1_dmo_dout 010 – rtx_rxc_ipp0_mb3_dmo_dout 011 – rtx_rxc_ipp1_mb3_dmo_dout 100 – rtx_rxc_zcp0_mb7_dmo_dout 101 – rtx_rxc_zcp1_mb7_dmo_dout 110 – rtx_rxc_vlan_mb6_dmo_dout 111 – Not allowed.
		7	Selects data cache.
		6	Selects instruction cache.

TABLE H-18 JTAG DMO Configuration Register *(Continued)*

Register	JTAG Instruction	Bit	Description
		14:13	14:13 5:3 2:0
		5:3	00 xx0 xxx → CORE4
		2:0	10 xx0 xxx → L2T4
			00 x01 xxx → CORE5
			10 x01 xxx → L2T5
			00 011 xxx → CORE1
			10 011 xxx → L2T1
			00 111 xxx → CORE0
			10 111 xxx → L2T0
			01 xxx xx0 → CORE6
			11 xxx xx0 → L2T6
			01 xxx x01 → CORE7
			11 xxx x01 → L2T7
			01 xxx 011 → CORE3
			11 xxx 011 → L2T3
			01 xxx 111 → CORE2
			11 xxx 111 → L2T2

H.7 JTAG Logic BIST

The Logic BIST test function is only applied to the SPC cores in UltraSPARC T2, and one engine is instantiated per core. The control of the Logic BIST engines comes from the TCU via either software or JTAG.

The control logic allows the Logic BIST engines to be run in parallel or in series and gathers the done signals for JTAG to query. There is no pass/fail indication that comes from the Logic BIST engines, so the engine must be scanned to determine the result.

H.7.1 JTAG Logic BIST Registers

In UltraSPARC T2, JTAG accessible registers for Logic BIST are listed in TABLE H-19.

TABLE H-19 JTAG Logic BIST Registers

Register	Jtag Instruction	Fields
BYPASS{7:0}	TAP_LBIST_BYPASS	One bit per Logic BIST engine; to bypass an engine during testing, set its bit to 1.
MODE{1:0}	TAP_LBIST_MODE	Bit 1 – Program access mode selected. Bit 0 – Parallel mode if 1, serial mode if 0.

TABLE H-19 JTAG Logic BIST Registers

Register	Jtag Instruction	Fields
LBIST{k:0}	TAP_LBIST_ACCESS	Includes targeted Logic BIST engines across strands.
DONE{7:0}	TAP_LBIST_GETDONE	One bit per MBIST engine; a 1 indicates the corresponding engine is done; same order as Logic BIST bypass register.

H.7.2 Accessing Pass/Fail Signature

To determine if the Logic BIST engine passed or failed, the signature must be scanned out via JTAG using TAP_LBIST_ACCESS. The signature must be compared to a known-good value.

Index

B

C

D

E

F

G

H

I

J

L

N

O

P

Q

R

S

T

U

V

W

X

Z