OpenSPARC™ T2 System-On-Chip (SoC) Microarchitecture Specification (Part 1 of 2)

Sun Microsystems, Inc.
www.sun.com

Part No. 820-2620-10
May 2008, Revision A

Submit comments about this document at: `http://www.sun.com/hwdocs/feedback`

Contents

Figures

Tables

Preface

This *OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification* includes detailed functional descriptions of the OpenSPARC T2 System-on-Chip I/O components. This manual is divided into two volumes, Part 1 of 2 (P/N 820-2620-10) and Part 2 of 2 (P/N 820-5090-10).

This manual also provides I/O signal list for each component. This processor expands Sun's throughput computing initiative by doubling the number of threads from the OpenSPARC T1 processor and adding support for industry standard I/O interfaces like PCI-Express and 10Gigabit Ethernet.

How This Document Is Organized

This *OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification (Part 1 of 2)*, (P/N 820-2620-10) includes detailed functional descriptions of the following OpenSPARC T2 System-on-Chip I/O components.

Chapter 1 describes the overall OpenSPARC T2

Chapter 2 describes the L2 Cache

Chapter 3 describes the Memory Control Unit (MCU)

Chapter 4 describes the Test Control Unit (TCU)

Chapter 5 describes the Clock Control Unit (CCU)

Chapter 6 describes System Interface Unit (SIU)

Chapter 7 describes the Non-Cacheable Unit (NCU)

Using UNIX Commands

This document might not contain information about basic UNIX® commands and procedures such as shutting down the system, booting the system, and configuring devices. Refer to the following for this information:

- Software documentation that you received with your system
- Solaris™ Operating System documentation, which is at:

  http://docs.sun.com

Shell Prompts

Shell	Prompt
C shell	*machine-name*%
C shell superuser	*machine-name*#
Bourne shell and Korn shell	$
Bourne shell and Korn shell superuser	#

Typographic Conventions

Typeface*	Meaning	Examples
AaBbCc123	The names of commands, files, and directories; on-screen computer output	Edit your `.login` file. Use `ls -a` to list all files. `% You have mail.`
AaBbCc123	What you type, when contrasted with on-screen computer output	`% su` `Password:`
AaBbCc123	Book titles, new words or terms, words to be emphasized. Replace command-line variables with real names or values.	Read Chapter 6 in the *User's Guide*. These are called *class* options. You *must* be superuser to do this. To delete a file, type `rm` *filename*.

* The settings on your browser might differ from these settings.

Related Documentation

The documents listed as online are available at:

http://www.opensparc.net/

Application	Title	Part Number	Format	Location
Documentation	*OpenSPARC T2 Core Microarchitecture Specification*	820-2545	PDF	Online
Documentation	*OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification (Part 1 of 2)*	820-2620	PDF	Online
Documentation	*OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification (Part 2 of 2)*	820-5090	PDF	Online

Documentation, Support, and Training

Sun Function	URL
OpenSPARC T2	http://www.opensparc.net/
Documentation	http://www.sun.com/documentation/
Support	http://www.sun.com/support/
Training	http://www.sun.com/training/

Third-Party Web Sites

Sun is not responsible for the availability of third-party web sites mentioned in this document. Sun does not endorse and is not responsible or liable for any content, advertising, products, or other materials that are available on or through such sites or resources. Sun will not be responsible or liable for any actual or alleged damage or loss caused by or in connection with the use of or reliance on any such content, goods, or services that are available on or through such sites or resources.

Sun Welcomes Your Comments

Sun is interested in improving its documentation and welcomes your comments and suggestions. You can submit your comments by going to:

http://www.sun.com/hwdocs/feedback

Please include the title and part number of your document with your feedback:

OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification (Part 1 of 2), part number 820-2620-10.

OpenSPARC T2 System-On-Chip (SoC) Microarchitecture Specification (Part 2 of 2) , part number 820-5090-10.

CHAPTER 1

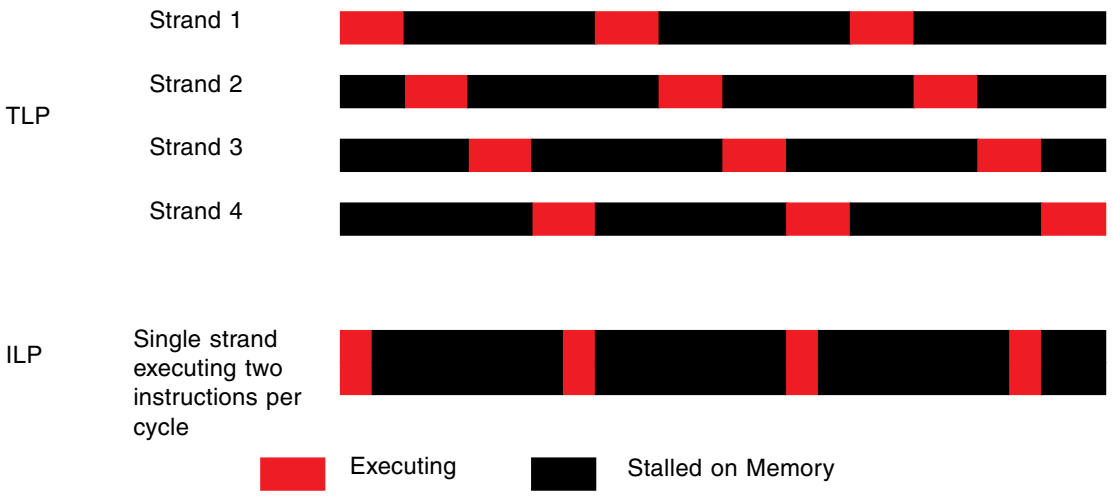
OpenSPARC T2 Basics

1.1 Background

OpenSPARC T2 is the follow-on chip multi-threaded (CMT) processor to the highly successful OpenSPARC T1 processor. The 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 pipe lining 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 non issue 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.

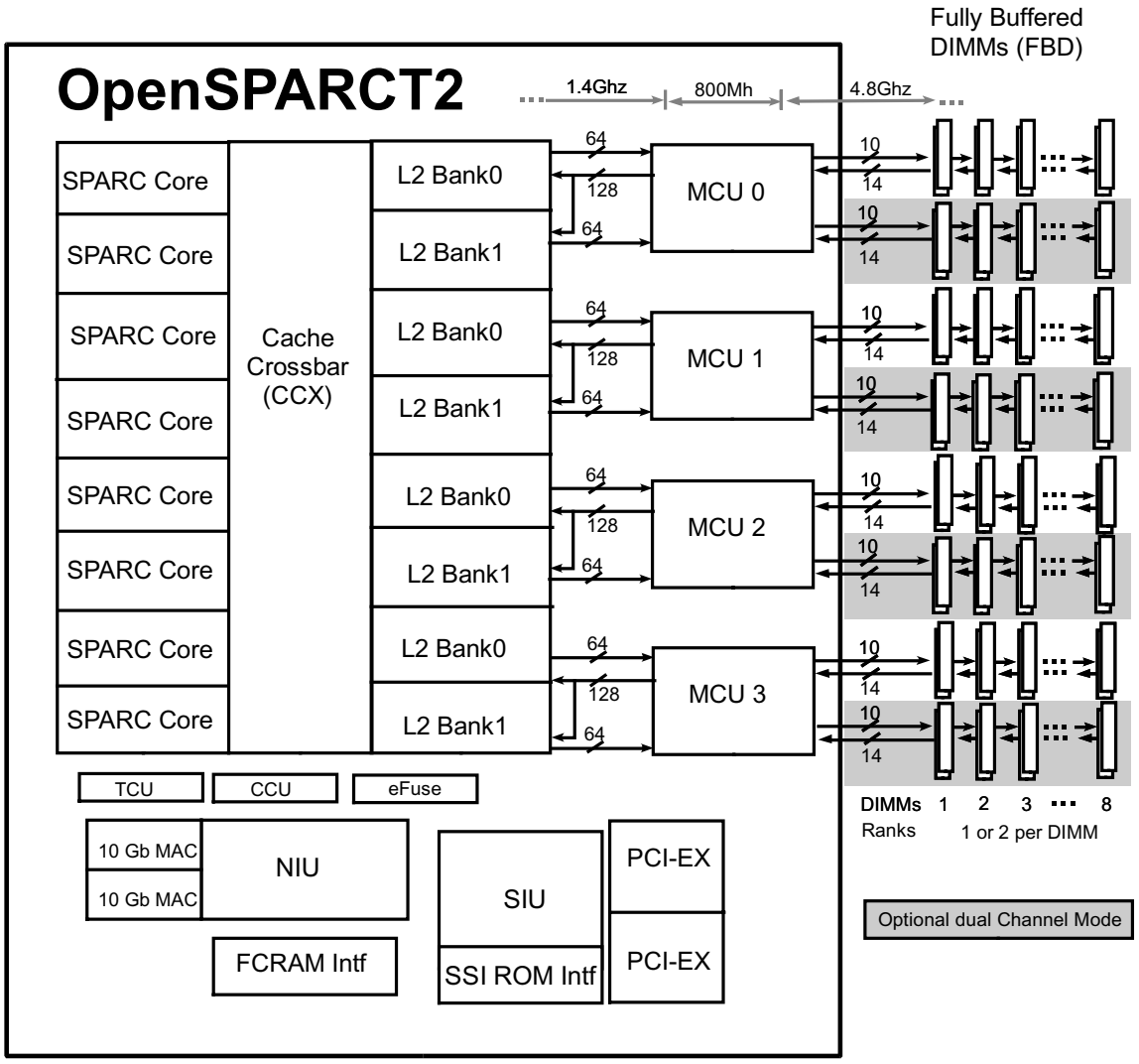
1.2 OpenSPARC T2 Overview

OpenSPARC T2 is a single chip multi-threaded (CMT) processor. OpenSPARC 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 physical core has a 16 KB, eight-way associative instruction cache (32-byte lines), 8 Kilobytes, 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. In addition, each physical core has a cryptography (stream processing) unit that is controlled by processor loads and stores but executes as an independent coprocessor. The eight 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 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 OpenSPARC T2 chip is shown in FIGURE 1-2.

FIGURE 1-2 OpenSPARC T2 Chip Block Diagram



1.3 OpenSPARC T2 Components

This section describes each component in OpenSPARC 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 Translation Lookaside Buffers (TLBs). An auto-demap feature is included with the TLBs to allow the multiple strands to update the TLB without locking.

A single floating-point unit is shared by all eight strands within a SPARC physical core. The shared floating-point unit is sufficient for most commercial applications which typically have less than 1% of their instructions being a floating-point operation.

Each physical core also contains a Stream Processing Unit (SPU) to accelerate cryptography.

Detailed information on the core processor is provided in *OpenSPARC T2 Core Microarchitecture Specification*.

1.3.2 SPARC System-On Chip (SoC)

Each SPARC physical core is supported by system on chip hardware components.

Information on each of the functioning units of the system on chip of OpenSPARC T2 are provided in the following chapters of *OpenSPARC T2 System-On Chip (SoC) Microarchitecture Specification, Part 1 of 2* (this manual) and *OpenSPARC T2 System-On Chip (SoC) Microarchitecture Specification, Part 2 of 2*, P/N 820-5090-10.

1.3.3 L2 Cache

The L2 cache is banked eight ways. To provide for better partial-die recovery, OpenSPARC T2 can also be configured in four-bank and two-bank modes (with 1/2 and 1/4 the total cache size respectively). Bank selection based on physical address

bits 8:6 for eight banks, 7:6 for four banks, and 6 for two banks. The cache is 4 Mbytes, 16-way set associative with pseudo-LRU replacement (replacement is based on a used bit scheme). 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.

1.3.4 Memory Control Unit (MCU)

OpenSPARC T2 has four MCUs, one for each memory branch with a pair of L2 banks interacting with exactly one Dynamic Random-Access Memory (DRAM) branch. The branches are interleaved based on physical address bits 7:6, and support 1–16 Double Data Rate (DDR)2 DIMMs. Each memory branch is two Fully Buffered DIMM (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 data buses, which is twice the DDR frequency. The FBD links run at six times the frequency of the DDR data buses.

1.3.5 Test Control Unit (TCU)

The TCU is the OpenSPARC T2 Test Control Unit and provides access to the chip test logic. It also participates in Reset, eFuse programming, clock stop/start sequencing, and chip debug. The TCU including Joint Test Action Group (JTAG) is completely stuck-fault testable via automatic test pattern generation (ATPG) manufacturing scan

1.3.6 Clock Control Unit (CCU)

The Clock Control Unit encompasses the following functions:

- PLL to drive the core and memory clocks.
- Interfacing with random number generator.
- Unit Control block (UCB) interface for programming the PLLs/RNG and reading RNG data.
- Provide sync pulses for deterministic clock domain crossing.
- Clock stretch and other test clocking mechanisms such as SERDES testing (via DTM) for OpenSPARC T2.

1.3.7 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.

1.3.8 Non-Cacheable 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.

1.3.9 Data Management Unit (DMU)

The DMU manages Transaction Layer Packet (TLP) to/from the PCI-Express Unit (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 Programmable Input/Output (PIO) reads pulling Direct Memory Access (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, reassembles 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).

1.3.10 Miscellaneous Input/Output (MIO)

MIO holds the majority of non-SERDES I/Os of OpenSPARC T2. The I/Os in MIO block fall broadly under the functional categories of clock, reset, test (scan and ramtest), ssi interface, process control PCM), eFuse program enable and debug. Most

of the I/Os in MIO are on Boundary Scan chain under control of TCU. All the functional flops in MIO are connected on regular scan chain with scanin, scanout and flush reset capabilities under the control of TCU.

1.3.10.1 SSI ROM Interface (SSI)

OpenSPARC 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 Programmable Input/Output (PIO) accesses across the SSI, thus supporting optional Control and Status registers (CSRs) or other interfaces within the Field Programmable Gate Array (FPGA).

Note – The SSI microarchitecture description is not included in this document.

1.3.11 Debug

This chapter describes OpenSPARC T2 HW features for post silicon debugability which involves debugging any issues that interfere with early bringup as well as debugging the difficult, complex bugs that eluded pre-silicon verification, and are unexpected or unusual corner cases. The overall goal of implementing these features is to make silicon debug more efficient, shortening the time to root cause complex bugs and thereby reducing time to remove and replace.

1.3.12 eFuse

The eFuse (electronic fuse) unit (EFU) contains an eFuse array macro (EFA), TCU interface and an eFuse controller (FCT). In a broad sense, the eFuseeFuse array is a non-volatile memory used to store information that needs to be programmed at the factory and used in the field.

The eFuse (Electronic Fuse) unit contains configuration information that is electronically burned in as part of manufacturing, including part serial number and Strand_Available information.

1.3.13 Reset

The Reset Unit asserts signals that cause other units to immediately revert to the initial state defined by the *OpenSPARC T2 Programmer's Reference Manual*.

The OpenSPARC T2 team has endeavored to keep OpenSPARC T2 as much the same as OpenSPARC T1 as possible. One major difference is that OpenSPARC T2 conforms to the CMP Programming Model.

1.3.14 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.

CHAPTER 2

Level 2 Cache

This chapter contains the following sections:

2.1 L2 Cache Functional Description

The following sections describe the OpenSPARC T2 processor level 2 cache (L2-cache):

2.1.1 L2 Cache Overview

The OpenSPARC T2 L2 cache is 4 MB in size and is composed of eight symmetrical banks interleaved on a 64 B boundary. Each bank operates independently of all others. Banks are 16 way set associative and 512KB in size. Block (line) size is 64 B. Each L2 bank has 512 sets.

The L2 cache accepts requests from the SPARC cores on the processor to cache crossbar (PCX) and responds on the cache to processor crossbar (CPX). The L2 is also responsible for maintaining on-chip coherency across all L1 caches on the chip by keeping a copy of all L1 tags in a directory structure. Since OpenSPARC T2 implements system on a chip with single memory interface and no L3 cache, there is

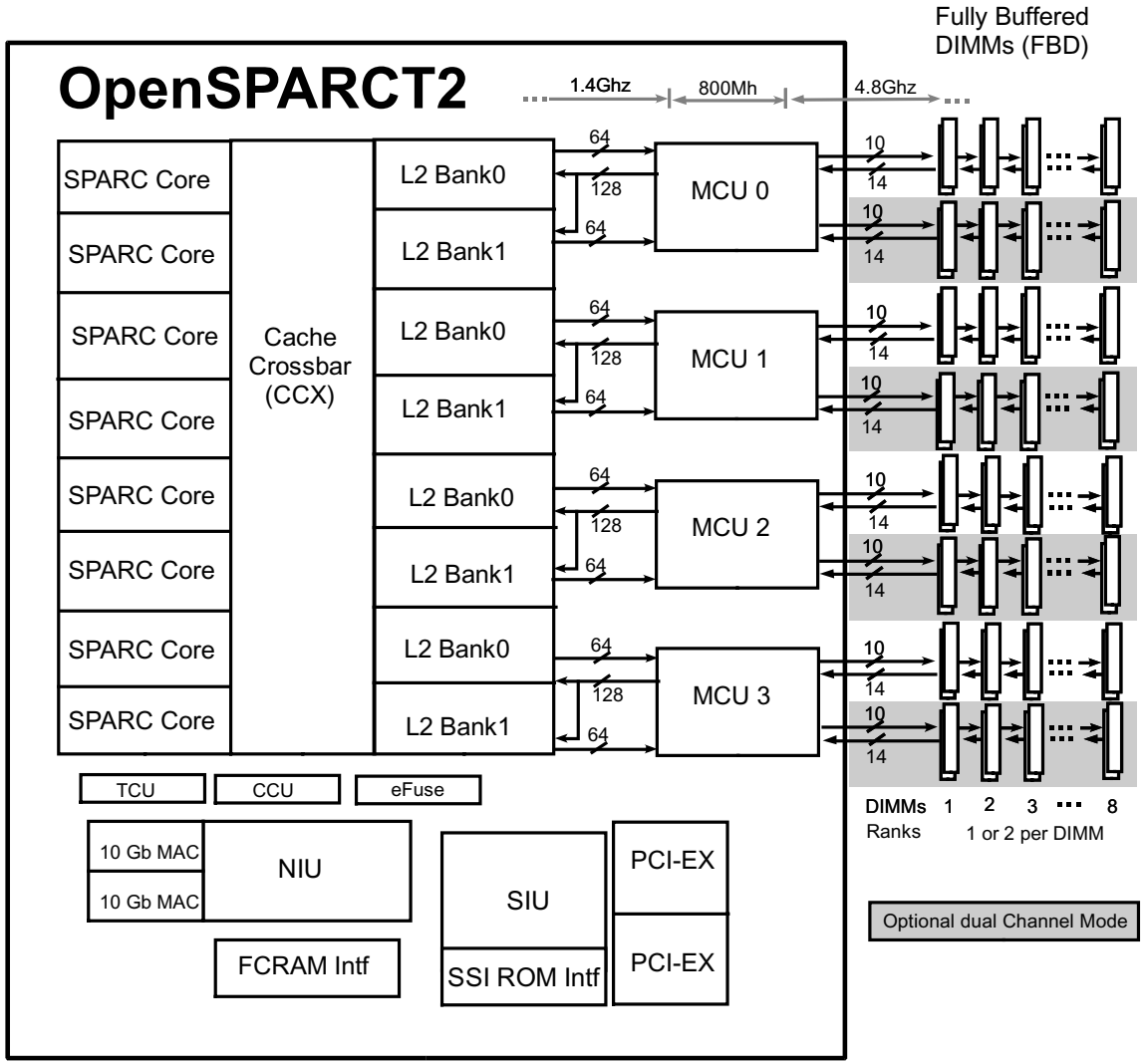
no off-chip coherency requirement for OpenSPARC T2 L2 cache other than being coherent with main memory. The L2 cache is a writeback cache and has lines in one of three states - invalid, clean, or dirty.

Each L2 bank has a 128b Fill interface and a 64b write interface with the dram controller.

Requests arriving on the I/O interface are sent to the L2 from the System Interface Unit.

The L2 cache unit works at the same frequency as the core (1.4 Ghz).

FIGURE 2-1 OpenSPARC T2 Processor Block Diagram



2.1.2 L2 Cache Block Functional Description

The L2 cache is organized into eight identical banks as shown in the FIGURE 2-1. Each bank has its own interface with SIU, MCU and Crossbar.

Each L2 cache bank interfaces with the eight cores through a Processor Cache Crossbar. The crossbar routes the L2 request (loads, ifetches, stores, atomics, asi accesses) from all eight cores to the appropriate L2 bank. The crossbar also accepts read return data, invalidation packets and store ack packets from each L2 bank and forwards them to the appropriate core(s).

Every two L2 cache banks interface with one MCU to issue reads and evictions to DRAM on misses in the L2. Writebacks get issued 64bits at a time to MCU. Fills happen 128 bits at a time from MCU to L2.

For 64 byte I/O writes from SIU, L2 does not allocate, but issues the writes to DRAM through a 64 bit interface with MCU. There is a single 64 bit interface with MCU for writebacks and I/O writes, and hence round robin arbitration is used between the Writeback Buffer and the I/O Write Buffer for access to MCU.

Each L2 cache banks also accepts RDD (read to discard), WRI (block write invalidate) and WR8 (partial write with random byte enables) packets from SIU over a 32 bit interface and queues the packet in the SIU Q. RDD and WRI do not allocate in the L2. On a hit, WRI invalidates in the L2 and issues a 64 B block write to DRAM. On a hit, RDD gets back 64 B of data from L2. On a miss, RDD fetches data from DRAM but does not install in L2, while WRI (on a miss) issues a 64 B block write to DRAM. WR8 packets cause partial stores to happen in L2 like regular CPU stores with random byte enables.

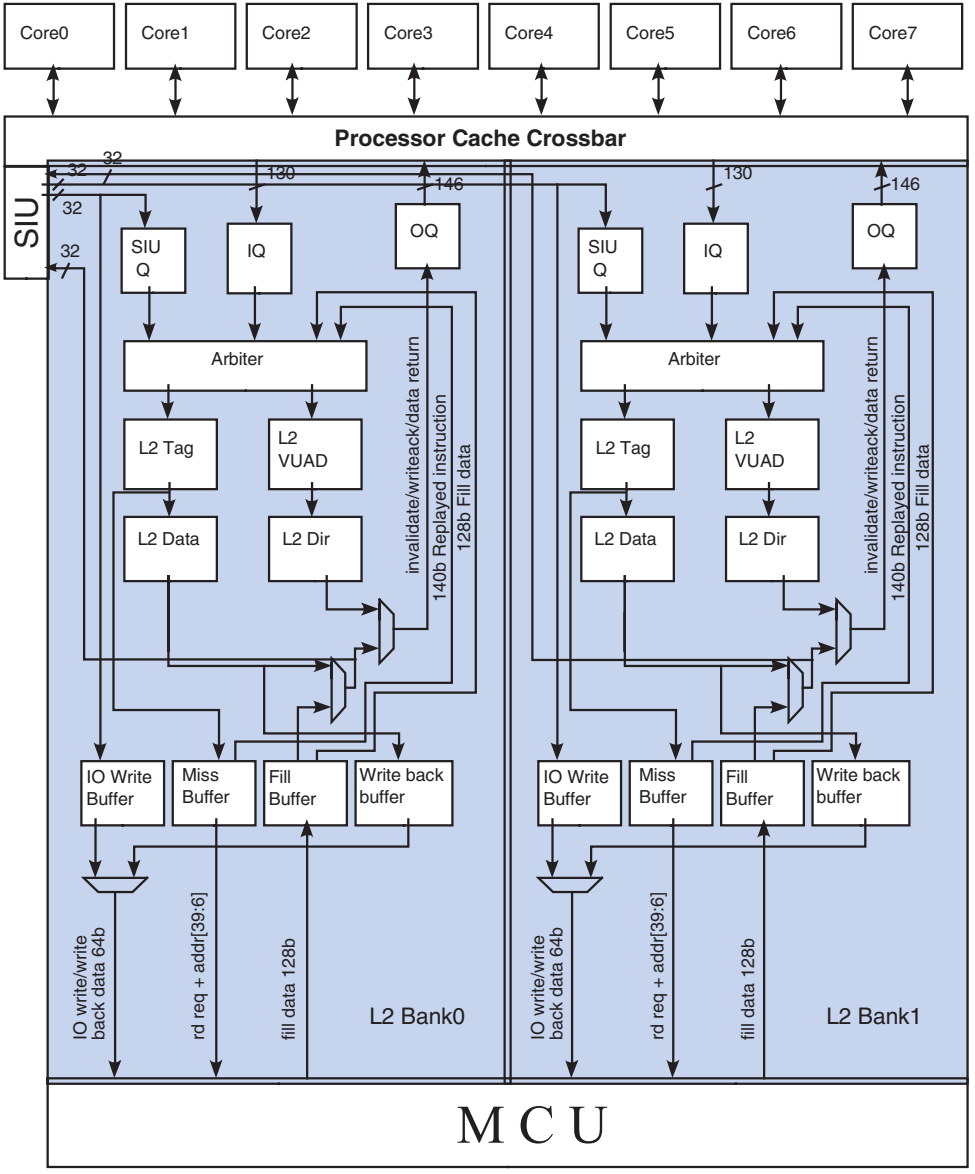
Each L2 cache bank is composed of the following sub-blocks:

- IQ: The input queue is a 16 entry FIFO which queues packets arriving on the PCX when they cannot be immediately accepted into the L2 pipe. Each entry in the IQ is 130 bits wide.
- SIUQ (SIU queue): Accepts RDD,WRI and WR8 packets from the SIU and issues them to the pipe after arbitrating against other requests.
- Arbiter: The arbiter manages access to the L2 pipeline from the various sources which request access. The IQ, MB, SIUQ, FB and stalled instruction in pipe all need access to the L2 pipe.
- L2 Tag: holds the L2 tag array and associated control logic. Tag is protected by SEC ECC.
- L2 VUAD: contains the Valid, Dirty, Used and Allocated bits for the tags in L2 organized in an array structure. There is one array for Valid and Dirty bits and a separate array for Used and Allocate bits. Each array is protected by SEC DED ECC.
- L2 Data: Contains 512 KB of L2 Data storage and associated control logic. Data is protected by SEC DED ECC on a 32/7 boundary.
- L2 Directory: The directory maintains a copy of the L1 tags for coherency management and also ensures that the same line is not resident in both the icache and dcache (across all cores). The directory is split into an icache directory (icdir) and a dcache directory (dcdir), which are similar in size and functionality.

- Miss Buffer: The Miss Buffer (MB) has 32 entries and stores instructions which cannot be processed as a simple cache hit. This includes true L2 cache misses (no tag match), instructions that have the same cache line address as a previous miss or an entry in the Writeback Buffer, instructions requiring multiple passes through the L2 pipeline (atomics and partial stores), unallocated L2 misses, and accesses causing tag ECC errors.
- Fill Buffer: The Fill Buffer is an eight entry buffer used to temporarily store data arriving from DRAM on an L2 miss request. Data arrives from DRAM in four 16B quad-words starting with the critical quad-word.
- Write Back Buffer: The Writeback Buffer is an eight entry buffer used to store dirty evicted data from the L2 on a miss. Evicted lines are streamed out to DRAM opportunistically.
- I/O Write Buffer: The I/O Write Buffer is a four entry buffer which stores incoming data from the PCI-EX interface in the case of a 64 B write operation. Since the PCI-EX interface bus width is only 32 bits wide, the data must be collected over 16 cycles before writing to DRAM

FIGURE 2-2 shows a diagram of the major components of the L2 cache.

FIGURE 2-2 L2 Cache Organization



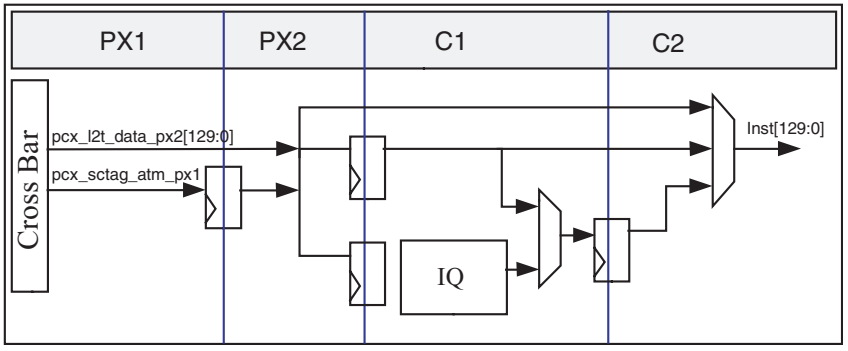
2.1.2.1 L2 Cache Interface Description

L2 cache interfaces with Crossbar, SIU and DRAM.

Crossbar

L2 cache receives requests from the core through the crossbar. These requests are received, decoded and forwarded to the arbiter logic by the Input queue (IQ) depending on the status of the arbiter block. The Input queue pipe line data path diagram is shown in FIGURE 2-3.

FIGURE 2-3 Input Queue Pipeline Data Path Diagram

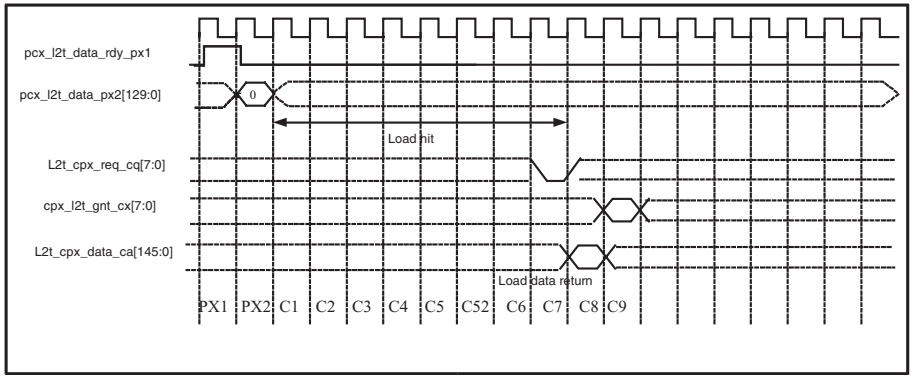


The timing diagram for a single load from PCX is shown in FIGURE 2-4.

The protocol for receiving a request from the crossbar is as follows:

The Input queue receives (pcx_l2t_data_rdy_px1) data valid signal followed by the data (pcx_l2t_data_px2) in the next cycle. Along with the data valid signal, the crossbar also dispatches a signal indicating if the instruction is atomic in nature(pxc_l2t_atm_px1). The request thus received is decoded into address, data and instruction fields in PX2 stage and forwarded to the arbiter logic to request access to L2 cache to process the request. If the arbiter accepts the request, it gets forwarded to L2 in the next clock, at which point the instruction reaches its C1 stage. If the arbiter is busy then it can either be sent after one or two clocks or recorded in the IQ array and dispatched later to the L2 pipe.

FIGURE 2-4 Timing Diagram for a Single Load from PCX

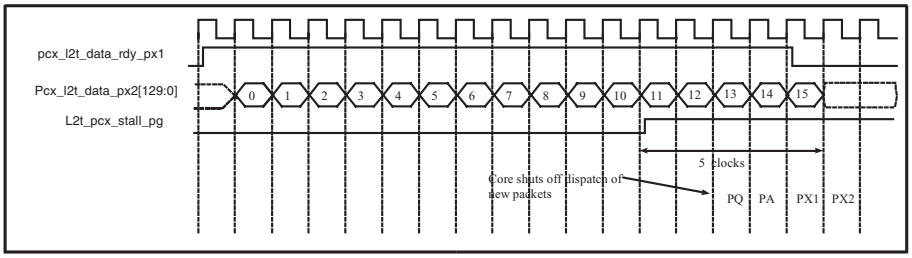


The protocol for L2 cache to send back a packet to the crossbar is as follows:

The L2 cache sends a request (l2t_cpx_req_cq) out in C7 of the pipeline if it has a packet to be dispatched. The packet may be return data for load/ifetch requests, acknowledgments for stores and invalidates for evictions and stores. The packet is dispatched in C8 (through l2t_cpx_data_ca). If the packet is consumed by the crossbar, an (cpx_l2t_gnt_cx) ack is received in C9. If an ack is not received from the crossbar within one or two cycles from C8, it gets retried from the flops at the input and output of the OQ respectively; if the ack gets received after two cycles, it gets retried from the OQ. In case the ack does not come for a long time, the new packets coming from the L2 pipe get accumulated in OQ until OQ fills up at which point the L2 pipe gets stalled.

The Input queue is 16 deep. PCX packets get written to IQ only when the L2 pipeline is stalled or busy and the PX2 arbiter does not accept any new PCX requests. IQ asserts l2t_pcx_stall_pg to crossbar when it is five short of being full. This is shown in FIGURE 2-5. These five cycles covers the packet shut off latency from core assuming the worst case latency of the core shutting off packet dispatch after dispatching an atomic packet.

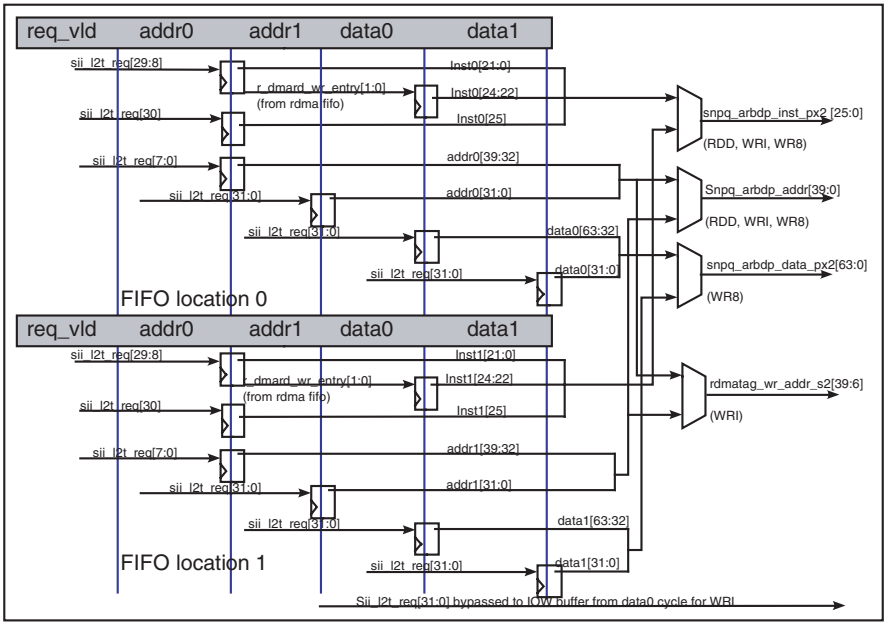
FIGURE 2-5 IQ written from PCX, PCX stall from IQ



SIU Interface

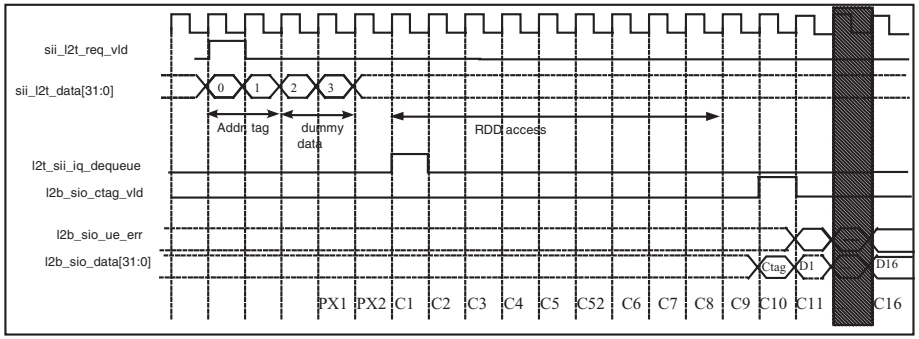
Requests from I/Os are received by L2 cache through SIU Queue block. There are three kinds of requests that can be received from the SIU: RDD (Read 64B), WRI (write 64B) and WR8 (write 8Bytes). FIGURE 2-6 shows the pipeline data path diagram for the SIU Queue Block.

FIGURE 2-6 SIU Queue Pipeline Data path Diagram



SIU dispatches requests to L2 cache through an unified address, data and instruction bus called sii_l2t_req. This bus is 32 bits wide.

FIGURE 2-7 Timing Diagram showing RDD Request and Read Data Return



The FIGURE 2-7 shows a typical RDD hit in L2.

The protocol to receive a request from SIU is as follows:

A valid signal (sii_l2t_req_vld) is sent along with the request to L2 cache. This signal is used to qualify a valid request transfer from the SIU block. Once the request is received by L2 cache, the instruction is registered and decoded into address, data and instruction fields as shown in FIGURE 2-7.

L2 SIU Queue block can record up to two requests in it's two-deep FIFO. Each FIFO entry registers the incoming packet from SIU over four groups of registers as shown in FIGURE 2-7 for WR8 and RDD transactions and into two groups (address and tag) for WRI transaction. For RDD, SIU will issue two dummy (pad) cycles on the sii_l2t_data[31:0] bus, so that the RDD and WR8 pipeline within SIU Queue can stay the same

The requests are received serially. There are two counters on the SIU side for flow control. One counter tracks the number of transaction dispatched to L2 cache and the other tracks the number of WRIs issued to L2 cache. The transaction counter is maintained in the SIU side incrementing on a transaction dispatch to L2 cache and decrementing upon receiving l2t_sii_iq_dequeue. WRI counter is incremented on dispatching a WRI transaction to L2 cache and decremented upon receiving l2t_sii_wib_dequeue signal. I/O Write Buffer can hold up to four cache lines. The transaction counter would block issue of any more transactions that the two-deep FIFO in the L2 SIU queue block can hold, while the WRI counter will keep a track of overall number of WRIs issued (cannot exceed 4). Thus as long as the WRIs are

issued without violating the transaction count specified by the transaction counter, and the WRI count of the WRI counter, there can be four WRIs outstanding to DRAM at any point of time though the SIU queue is two-deep only.

l2t _sii_iq_dequeue signal is asserted when an instruction is issued down the L2 pipe (WR8,WRI & RDD transactions) in C1 stage. l2t_sii_wib_dequeue is asserted when the contents of an I/O Write Buffer entry are streamed to DRAM (only WRI transaction).

1. RDD: 64 byte read request is received by L2 cache over five clocks. During the data cycle, dummy data is driven. The 64byte data from L2 is returned to SIU over 16 cycles with ctag_vld information.

2. WR8: Eight byte writes are received by L2 cache over five clocks. The L2 treats this instruction in exactly the same way as a store. When the write data gets written into the L2 Data Array, an encoded 32 bit ack is sent out to SIU by asserting ctag_vld in the same clock.

3. WRI: 64 byte write is received by L2 cache over 19 clocks. The line being written is not allocated in the L2 cache. However if the write hits in the L2 cache, it invalidates the L2 cache entry and also copies of the line in L1 caches. The WRI packet gets written into the I/O Write Buffer from where the data gets written to DRAM opportunistically. l2t_sii_iq_dequeue signal is asserted when the write instruction is issued down the L2 pipe, and l2t_sii_wib_dequeue is asserted when the contents of an I/O Write Buffer entry are streamed to DRAM. Also an encoded write ack is sent out to SIU on l2b_sio_data[31:0] by asserting ctag_vld indicating completion of the WRI.

Ordering of SIU Transactions in L2 (Data Returns and Write Acks from L2 to SIU):

1. For same address to the same L2 bank, read returns and write acks will be always in transaction order from SIU

2. For different addresses to the same L2 bank, depending on hit or miss, RDDs can send back data out of order.

3. For different addresses to the same L2 bank, WR8s (partial writes with byte masks) can send back acks out of order, depending on hit or miss. (WR8s do read modify writes, and WR8 ack gets sent only in the store update phase of the WR8). So if there are two back to back WR8s, and the first one misses, the second one hits: ack will get sent for the second one before the ack for the first one, while the first one is still waiting to fetch the data from memory.

4. For different addresses to the same L2 bank, WRIs will send back acks in transaction order from SIU as WRIs go straight to memory and do not update L2.

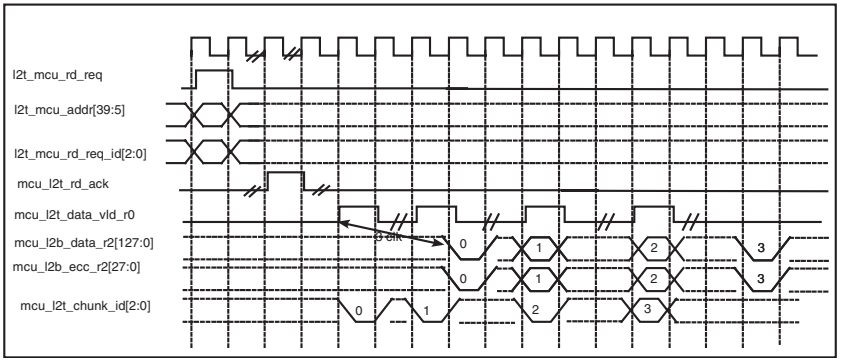
2.1.2.2 MCU Interface:

L2 cache issues read and write requests to MCU. All instructions which do not hit in L2 cache are recorded in the Miss Buffer (MB). Miss Buffer evaluates and sets a (dram_pick) bit if it needs to be issued to MCU.

Reads which need to be dispatched to MCU should satisfy the following criteria:

- Win arbitration among all pending reads (with the dram_pick bit set for reads).
- Should have no pending (read or write) transactions to MCU waiting for an ack.
- Should have enough place in the Fill Buffer for the read data to return.

FIGURE 2-8 Read Request from L2 Cache to MCU and Read Data Return



The protocol for sending out a read request is as follows:

A read request is dispatched to MCU by asserting a a read request (l2t_mcu_rd_req) signal. Along with the read request the address and a read request ID (l2t_mcu_rd_req_id) is dispatched in the same cycle. The read address is also recorded in the Fill Buffer. MCU records and processes the read request. A read ack (mcu_l2t_rd_ack) is sent back indicating that the read request was recorded.

When the data is ready, MCU returns the data to the L2 cache. The data is returned with the data_valid (mcu_l2t_data_vld_r0) being sent first. Data (128 bits wide mcu_l2t_data_r2), the read request id (mcu_l2t_qword_id[1:0]) and ECC information (mcu_l2t_ecc_r2 [27:0]) related to the data is sent after 3 clocks. The read data returned by the MCU gets recorded by the Fill Buffer. Upon receiving the data, the missed load/ifetch from the Miss Buffer gets replayed through the L2 pipe and reads the data from the Fill Buffer itself (critical 16B or 32B first) and sends data to requesting core. After this, the Fill Buffer requests arbiter to complete the fill. The qword data arrives from MCU in four packets. There is no relationship between the dispatch of packets.

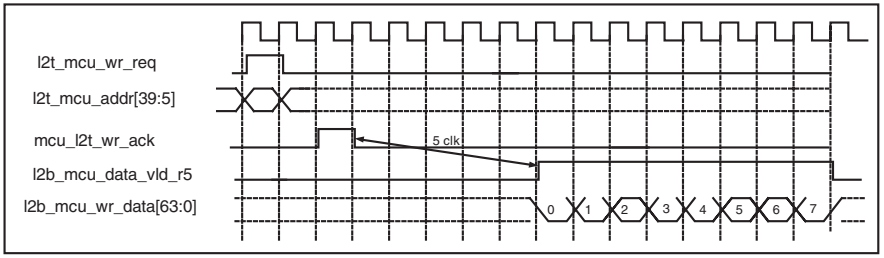
Also in case of a miss in L2 for a Block Init Store with PA[5:0] = 0, L2 will issue a dummy read request to DRAM (l2t_mcu_rd_dummy_req), but MCU will send back all zeros in four packets. L2 will install the line with all zeros in the data.

Writes to the MCU get issued when a request is recorded in the I/O Write Buffer (IOWB) or Write back Buffer (WBB).

The following condition needs to be satisfied for a write to be dispatched to MCU:

- Win arbitration among all pending writes. Writes can be in IOWB (WRI from SIU) or WBB (eviction).
- There should be no pending (read or write) transactions to MCU waiting for an ack.

FIGURE 2-9 MCU Write Transaction



The protocol is similar for writes:

A write request (l2t_mcu_wr_req) is sent to MCU along with the address for the write data. Upon receiving a write request, DRAM sends a acknowledge back indicating it is ready for receiving the write data. L2 cache takes five clocks upon receiving the ack to the time it starts to send data to DRAM in sizes of 8Byte.

2.1.3 L2 Pipeline

The L2 pipeline has 9 stages, the details of which are described below.

- Arbitration (PX2):
  - Mux between PCX, IQ, I/O, MBF and FBF and C1 (stalled) instructions
- Tag Access (C1):
  - Tag access, VUAD Read and Bypass
  - Tag Compare
  - Miss Buffer CAM Operation in Phase1.

- Miss Buffer Hit logic
- Generation of ECC for store data
- Generation & check of ECC for the access address
- WBB and Fb CAM in Phase 2.

- Way Sel Generation (C2):
  - Tag Hit logic
  - Replacement way logic (pseudo LRU)
  - Miss Buffer Hit generation and 2 cyc bypass.
  - Way select logic
  - Set, index, col, way sel, rd/wr, word enables xmit to the data array
  - Stall for mutlicycle operations (For e.g. Eviction, Fill etc.) or column offset collision.
- Way Sel Xmit (C3):
  - Set, index, col, way sel, rd/wr, word enables xmit in the data array
  - Evict way sel generation and eviction logic
  - MB tag write, MB Valid bit setting
  - Fb hit entry xmit to FB Data array (in l2b)
- Data Access cyc1 (C4):
  - Data array read/write cycle 1 (for load/store hit)
  - Fb data buffer read.
  - Way sel transmit to data array for Fill only
  - Setup directory inputs for CAM/write operations.
- Data Access cyc2 (C5):
  - Data array read/write cycle 2 (for load/store hit)
  - Data array read cyc 1 for eviction
  - WB tag write in the case of a dirty eviction
  - VUAD array Write
  - Way sel transmit within data array for Fill only
  - Stage Fb data
  - Write/CAM directory
- Data Access cyc3 (C52):
  - Data array read/write cycle 3 (for load/store hit), 4:1 mux for L2 data
  - FB and L2 data mux
  - Data array read cyc 2 for eviction
  - Data array write cyc 1 for Fill

- Data Return xmit (C6):
  - 16B data xmit to tag block
  - Data array read cyc 3 for eviction
  - Invalidation vector processing.
  - Request vector generation logic
  - Data array write cyc 2 for Fill
- Error Correction (C7):
  - Error Correction/Detection
  - Request vector to OQ/CPX
  - Data array write cyc 3 for Fill
- Data Response (C8):
  - L2 data and Invalidation data MUX
  - Data xmit to OQ/CPX
  - Write WBB data
  - 64b data merge for PSTs and 64b compare for CASX

2.1.4 L2 Interactions with Core

OpenSPARC T2 Cores will use eight bit Byte Mask fields for stores instead of two bit size field that OpenSPARC T1Cores use. The main reason for this is to support VIS partial stores with random byte enables.

This section describes the pipeline flow for a few representative L2 operations.

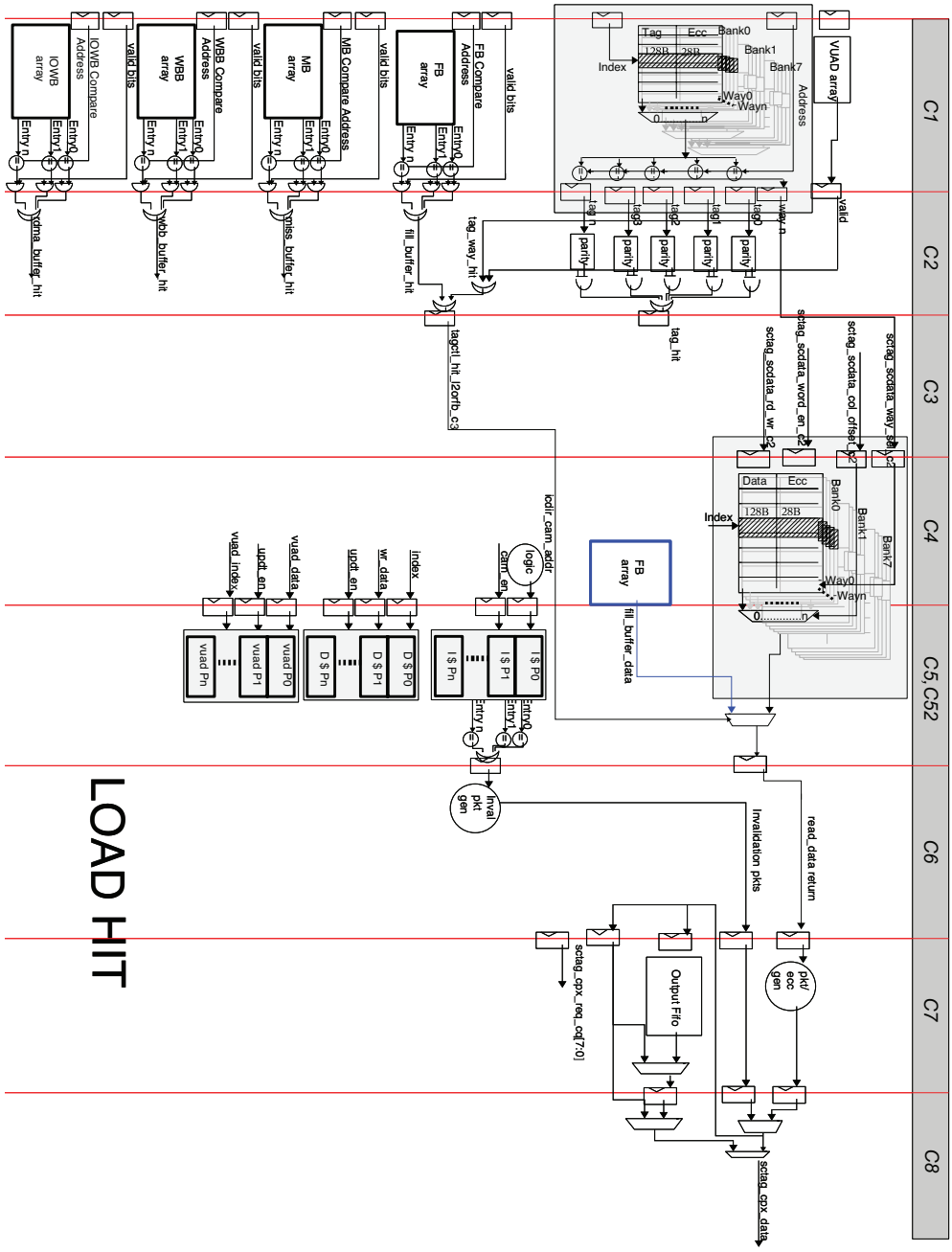
2.1.4.1 Load Hit

TABLE 2-1 Pipeline Diagram: Load Hit

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag, VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! Gen, xmit VUAD ECC check	way sel xmit in l2d	data array read cyc1 FB data read cycle Xmit inputs to directory	data array read cyc2 stage FB data D$ directory write I$ directory CAM VUAD write	data array read cyc3 4:1 mux mux with FB data	data xmit cycle gen inval. vector	request to the dest cpx queue check ECC on data	Mux Data/Inval. Vector data return to dest. cpx

Loads always return 16B of data, and lower address bits are ignored, i.e. if a load request to address 0x13 is presented to the L2, the 16 bytes at 0x10 are returned. The eight-bit byte mask field is ignored for loads. The different instruction types that fall in the category of loads are: *load, prefetch, stream load, mmu load*. Out of these, *prefetch, stream load and mmu load* are non-cacheable (will have NC bit set in the PCX packet). These loads do not cam the I$ directory and do not update the D$ directory.

FIGURE 2-10 Load Hit



2.1.4.2 Store Hit

Eight bytes of store data is always sent to the L2. The LSU will ensure that the data is properly aligned to the 8B boundary. The eight-bit byte mask indicates which bytes are to be stored. Again, the lower address bits are ignored.

(This is different than OpenSPARC T1. OpenSPARC T1 L2 had to use the lower address bits along with the size to determine what to store.)

FIGURE 2-11 shows the timing diagram for a 8/4 byte store with the following combination of Byte Enables:

1111 1111

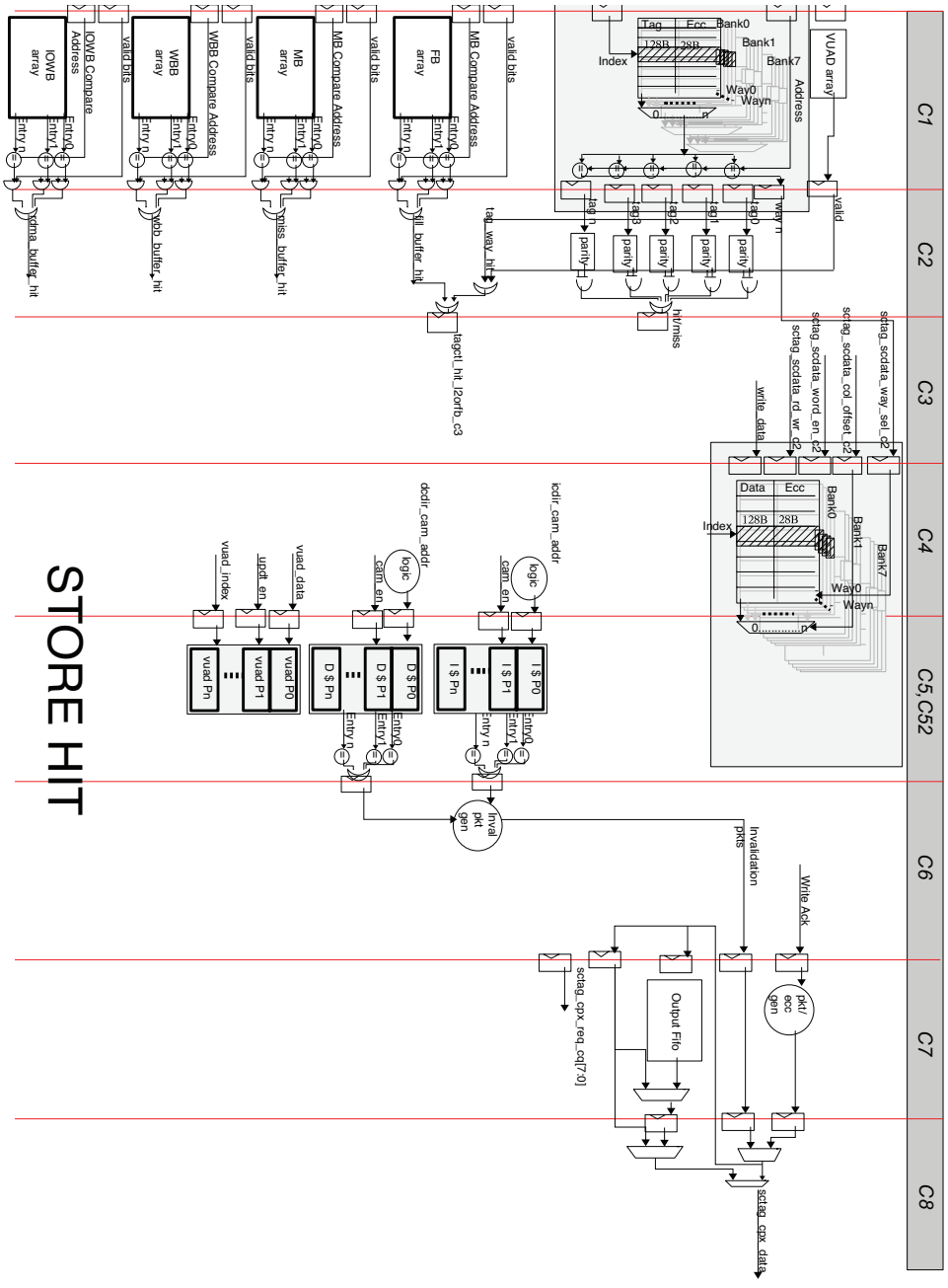
1111 0000

0000 1111

TABLE 2-2 Pipeline Diagram: Store Hit

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag,VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic perform store data ECC FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! gen,xmit VUAD ECC check	way sel xmit in l2d	ddata array write cyc1 Xmit inputs to directory	data array write cyc2 I$ and D$ directory CAM VUAD write	data array wr cyc3	gen inval.vector	request to the dest cpx queue (ack for write) check ECC on data	Mux Data/Inval. Vector

FIGURE 2-11 Store Hit



To improve the performance of stores from L1, L2 cache in OpenSPARC T2 would send back acks to core in case stores from L1 hit to outstanding store miss to the same line in Miss Buffer. This would involve adding a control flop in Miss Buffer control logic to associate a load miss or store miss with the MB entry. However the ack will get sent back if it hits only in store misses. If any one of the addresses it hits is a load miss, the ack will not be generated.

Note – In OpenSPARC T1, stores are ack'ed when they make their first pass through the L2 pipe - hit or miss. The exception to this is when the store hits an entry in the Miss Buffer. The reason for not issuing the ack in this case is that if the entry in the MB were a load, the ack would cause the L1 to update before the load returned data, causing WAR hazard. However, if the entry in the MB was a store, no such hazard exists, and the ack can be issued.

Main reason for wanting to add this earlier ack capability in OpenSPARC T2 is to complement the addition of store pipelining in the L1 for stores going to the same L2 cache line. In this scheme, a store that follows another store to the same L2 line can be issued without waiting for the first store to be ack'ed, however in the absence of acks for these stores, the Store Buffer entries cannot be dequeued. This would stall dequeue of Store Buffer entries due to stores to different lines and different banks also that are behind the stores to the same line. In this particular case (stores to the L2 line are all waiting in the Miss Buffer for the data return from DRAM), this can amount to stalling the drain of the Store Buffer in L1 for ~160 cpu cycles, causing it to be filled up and thread(s) to stall.

The Store Buffer is only eight entries per thread, and once it fills up, the thread stalls, so we want to minimize this. STB stalls cause a noticeable degradation in performance. The decrease in stalls from adding pipelining gained somewhere around 15% on Spec INT, so it's a worthwhile change.

2.1.4.3 Partial Store

Partial stores are stores which have any combination of byte masks other than 00001111,11110000 and 11111111.

Even for partial stores, eight bytes of store data is always sent to the L2. The LSU will ensure that the data is properly aligned to the 8B boundary. The eight-bit byte mask would indicate which bytes are to be stored. Again, the lower address bits (0,1,2) are ignored.

Partial stores are handled as a read-modify-write operation in two passes through the pipe. The first pass is shown in Timing Diagram. The second pass is identical to a store, except that the ack does not get sent again.

TABLE 2-3 Timing Diagram: Partial Store

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag,VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! Gen,xmit VUAD ECC check	way sel xmit in l2d	data array read cyc1 FB data read cycle Xmit inputs to directory	data array read cyc2 stage FB data I$ and D$ directory CAM VUAD write	C52 data array read cyc3 4:1 mux mux with FB data	data xmit cycle gen inval. vector	request to the dest cpx queue (ack for write) check ECC on data	Mux Data/inval vector Merge data

The merged data is written into the Miss Buffer and is readied for reissue in C9.

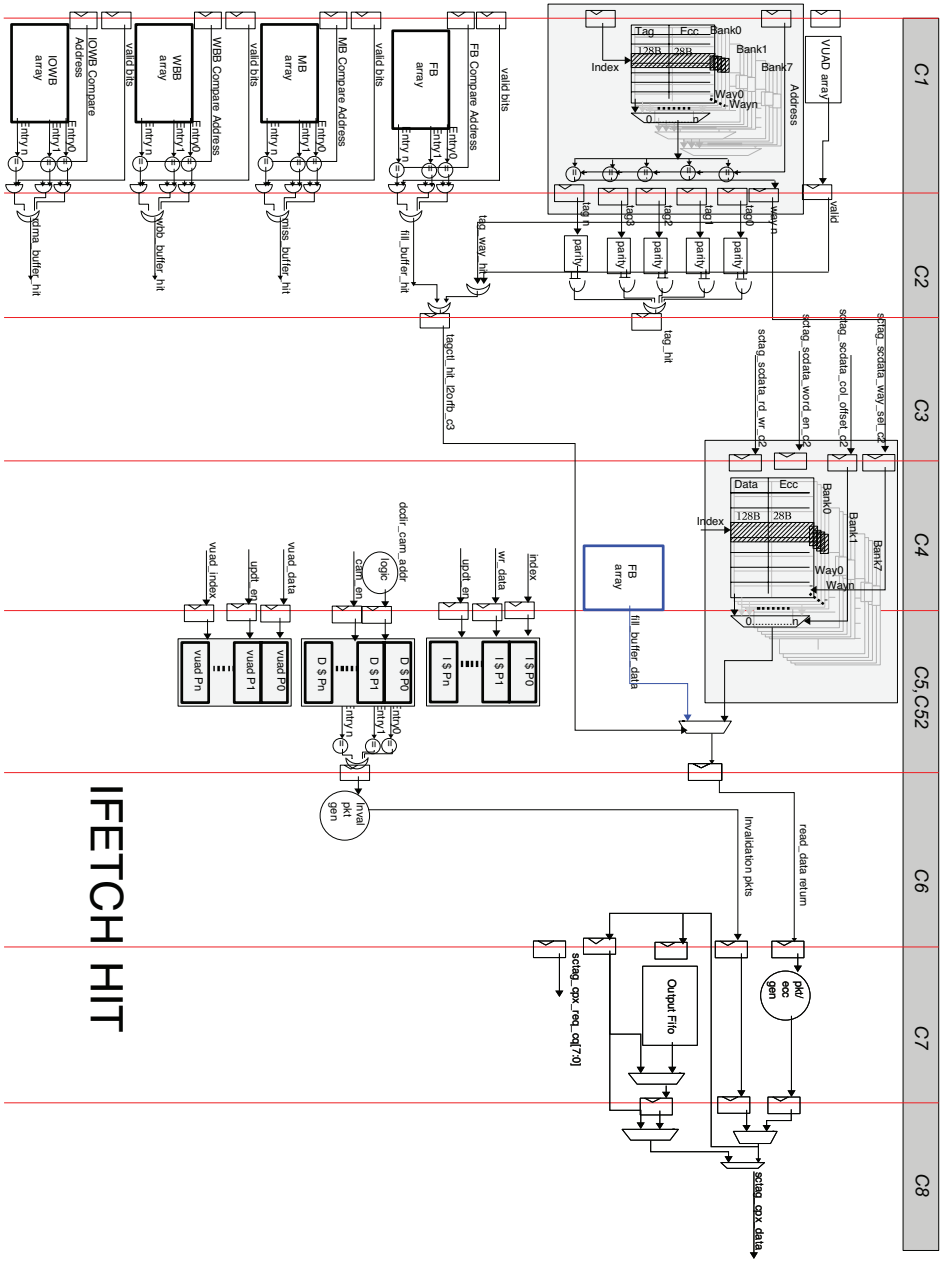
2.1.4.4 Ifetch Hit

ICache line is 32 B, so two data reads are required for an Instruction Fill request.

TABLE 2-4 Timing Diagram: Ifetch Hit

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag,VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! Gen,xmit VUAD ECC check stall next instruction	way sel xmit in l2d	data array read cyc1 FB data read cycle Xmit inputs to directory	data array read cyc2 stage FB data I$ directory write D$ directory CAM VUAD write	data array read cyc3 4:1 mux mux with FB data	data xmit cycle gen inval. vector	request to the dest cpx queue check ECC on data	Mux Data/Inval. Vector data return to dest. cpx

FIGURE 2-12 Ifetch Hit



The following 16B data block is transmitted in C9. Note that Ifetch misses are 32B aligned, and 32 B get returned to the core over two consecutive cycles. For a 32 B ifetch with lower address bits non-zero (unaligned 32B read), the two 16B lines are returned in address order, not critical line first. The eight-bit byte mask field is ignored for Ifetch.

Note that if the NC bit is a 1 for an Ifetch request (L1 I$ is disabled), it will still cam the D$ directory and send an invalidation vector if there is a hit in the D$ directory.

2.1.4.5 Miss

An instruction that does not hit the L2 cache, Fill Buffer or the Writeback Buffer is queued in the Miss Buffer as a "true miss". Eviction is performed during the second pass of the miss operation. This is done to remove the hit/miss determination from the critical C1 stall signal. To improve the performance of stores from L1, L2 cache in OpenSPARC T2 would send back acks to core in case stores from L1 hit to outstanding store miss to the same line in Miss Buffer. This would involve adding a control flop in Miss Buffer control logic to associate a load miss or store miss with the MB entry. However the ack will get sent back if it hits only in store misses. If any one of the addresses it hits is a load miss, the ack will not be generated.

TABLE 2-5 Timing Diagram: Miss

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag,VUAD read tag compare MB CAM perform store data ECC (for store) VUAD bypass FB CAM WBB CAM	way sel logic, check VUAD ECC	write MB tag set MB valid bit					request to the dest cpx queue (ack for write in case it is a store miss but hits in one or more MB store miss entries)	

In C9, EVICT_READY bit gets set in MB. The instruction (along with data for a store) gets written to MB also in C9.

2.1.4.6 Eviction (Clean or Dirty)

The entry in the Miss Buffer is selected for issue in case the EVICT_READY bit gets set indicating it is ready for eviction. An evict instruction gets issued from the Miss Buffer which causes eviction to happen as it makes a pass down the pipe. This also clears the EVICT_READY bit provided evict instruction pass does not encounter a TECC error. Invalidation of L1 cache lines happen for eviction of clean or dirty lines. However only dirty lines are sent to DRAM, clean lines just get overwritten.

TABLE 2-6 Timing Diagram: Eviction

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag,VUAD read tag compare VUAD byp	perform pseudo LRU stall two cycles to avoid collision with instruction after evict instruction for data array access	way sel logic mux out evicted tag way sel Xmit to l2d	way sel xmit in l2d	data array read cyc1 I$ and D$ directory CAM write evicted tag in WBB VUAD Write	data array read cyc2	data array read cyc3 gen inval. vector	xmit inv response to dest cpx queues	inv packet to dest cpx data queues write WBB data	ready for request to DRAM

2.1.4.7 Fill

TABLE 2-7 Timing Diagram: Fill

C1	C2	C3	C4	C5	C52	C6	C7
tag write	stall three cycles to avoid collision with instruction after fill for data array access	Xmit FB entry to l2b	way sel xmit to l2d read FB	VUAD write Xmit way sel inside l2d mux fbdata with l2d	data array wr cyc 1	data array wr cyc2	data array wr cyc3

FIGURE 2-13 Read Miss and Read Data Fill from DRAM

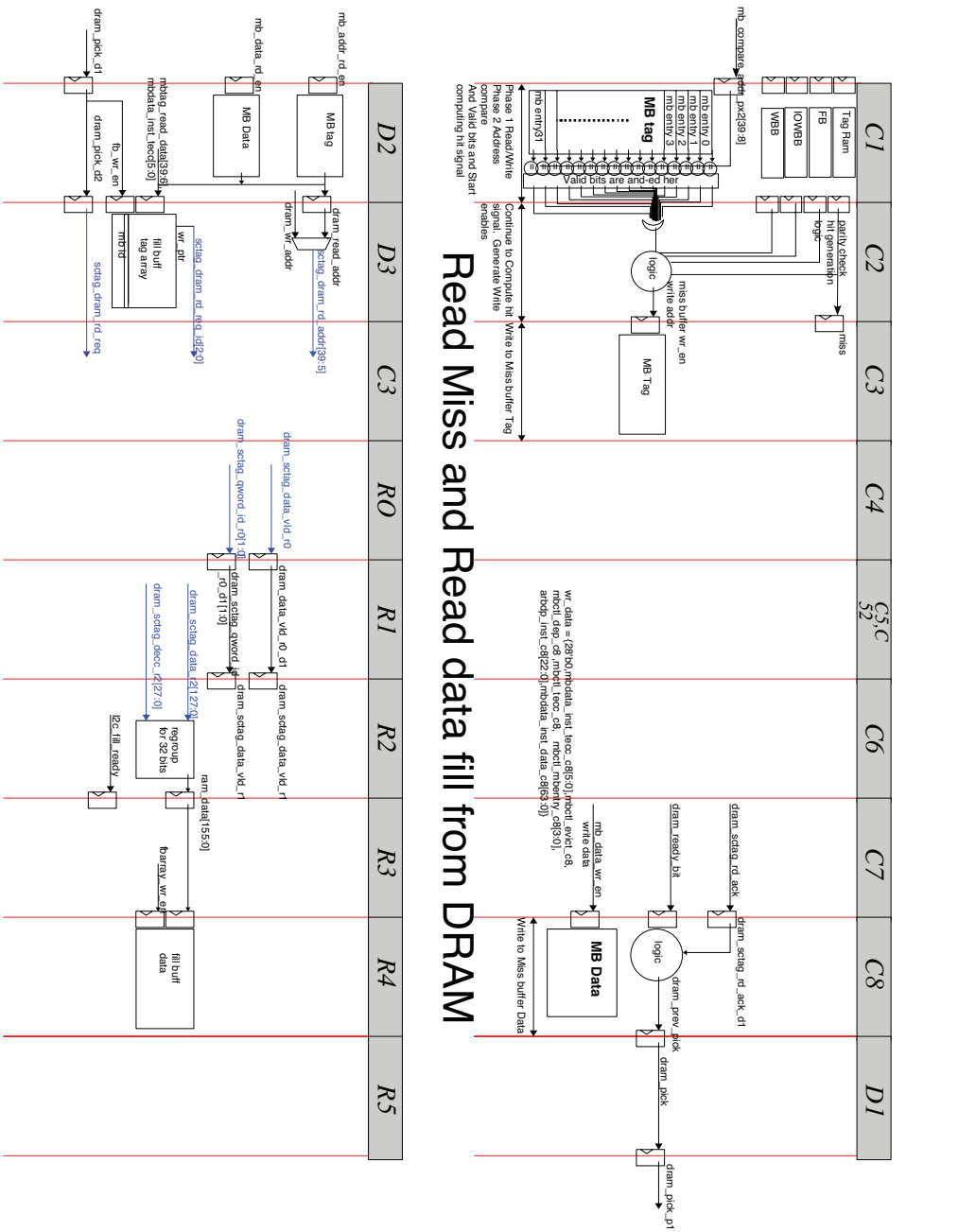


FIGURE 2-14 Evict and Write back to DRAM

2.1.4.8 Atomics LDSTUB/SWAP 1st Pass

Same as a load with a merge in C8.

TABLE 2-8 Timing Diagram: Atomics LDSTUB/SWAP 1st Pass:

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag,VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! Gen, xmit VUAD ECC check	way sel xmit in l2d	data array read cyc1 FB data read cycle Xmit inputs to directory	data array read cyc2 stage FB data D$ directory write I$ directory CAM VUAD write	data array read cyc 3 4:1 mux mux with FB data	data xmit cycle gen inval. vector	request to the dest cpx queue check ECC on data	Mux Data/Inval. Vector data return to dest. Cpx Merge data as in Partial Store

The first pass through the L2 pipe reads 16B of data at the address requested (ignoring the lower bits), returns it to the requesting processor, and merges the swap/UB data. The merged data is written into the Miss Buffer and is readied for reissue in C9. The instruction then goes through a second pass upon which the new data is stored and an acknowledgment is sent to the requesting processor. The second pass of a ldstub/swap is same as that for a store hit.

For SWAP and LDSTUB, the bytes to write in L2 will be picked up from the Byte Mask itself. SWAP is always 32b aligned on four byte boundary and LDSTUB is always 8b.

2.1.4.9 Atomics CAS

CAS{X} instructions are handled as two packets. The first packet (CAS(1)) reads the data from memory, sends the data back to the requesting processor, and performs the comparison in C8. The second packet (CAS(2)) is inserted into the MB as a store.

If the comparison result is true, the second packet proceeds like a normal store. If the result was false, the second pass proceeds to only generate the store acknowledgment. The data arrays are not written.

CASA/CASXA are similar, but with one difference. CASA is 32b, aligned on four byte boundary and CASXA is 64b. The compare and conditional store are assumed to be on an 8B boundary (except the load return which is always 16B). The eight-bit Byte Mask will indicate which bytes to compare and conditionally store.

2.1.4.10 Prefetch Invalidate Cache Entry (ICE)

L2 supports Prefetch ICE instruction which gets used by SW to flush lines in L2 based on an index and a way specified as part of the Physical Address in the instruction itself. Bits [39:37] of the PA has to be driven as 3'b011 by SW and the way, index, bank information would be on PA[21:18], PA[17:9] and PA[8:6] respectively. LSU issues a prefetch instruction over the crossbar to L2 with bit 116 (inv bit) of the cpx packet being 1'b1. On seeing this packet, L2 flips bit 39 to 1'b1 before storing it in IQ array or feeding to the pipe. Thus with PA[39:37] = 3'b111, it is guaranteed that the instruction always misses in L2 irrespective of 8/4/2 bank mode of operation.

The Prefetch ICE gets executed in L2 in two passes. First pass is like a regular prefetch which misses in the L2 tags. On the miss, the instruction gets written to the Miss Buffer and also the Evict bit gets set. However the DRAM read gets suppressed as this is a flush instruction only and no data needs to be read from memory.

In the second pass, an evict instruction gets issued from the Miss Buffer for the Prefetch ICE and it will use the way specified in PA[21:18] of the Prefetch ICE packet itself to pick the Eviction way and L2 Directory Lookup way, instead of the way picked by the LRU logic. Then the eviction proceeds like normal: an eviction invalidation packet gets generated and sent to the crossbar to invalidate all L1 ways for all cores that are included in that line. In case the line is dirty, a writeback happens to DRAM. In the eviction pass of the instruction, it gets deleted from the Miss Buffer. No response packet gets sent to the cores for the instruction itself.

Note that in case the Prefetch ICE instruction encounters a Tag Parity Error or VUAD CE in either the first pass, the error is ignored (not logged and reported) and the Prefetch ICE goes on as normal. However in its second pass if the Prefetch ICE detects a tag parity error, it will be re-inserted into the Miss Buffer, the eviction pass will not happen and a scrub will be issued from the Miss Buffer. After the scrub is complete, the eviction pass of the Prefetch ICE will occur and if this time there are no more tag parity error detected, the eviction pass will complete. This is because the tag parity error could have corrupted any bit of the address, so that unless corrected, the eviction of a dirty line would cause data corruption in memory. However if the eviction pass of the Prefetch ICE encounters a VUAD CE, the error would be ignored and the eviction pass would go through since we know the way already that has to be evicted. The VUAD data would get silently corrected before it gets written to the VUAD array in C5.

Note that since any error in the Dirty bit would have been silently cleaned in the first pass of Prefetch ICE itself, in the second pass of the Prefetch ICE, the way to flush would be identified correctly as clean or dirty and the dirty line would get evicted to DRAM properly. Also if the Dirty bit error gets detected in C2 stage of the second pass of Prefetch ICE (the corruption in the array happened in between the first and second passes), the way to flush would still get correctly identified as clean or dirty in the second pass, as the data would get silently corrected in C2 stage of the second pass itself.

Prefetch ICE First Pass (Miss in L2)

TABLE 2-9 Timing Diagram: Prefetch ICE First Pass (Miss in L2):

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag, VUAD read tag compare MB CAM VUAD bypass FB CAM WBB CAM	miss detected Tag parity error NOT checked VUAD ECC checked, data corrected if SE detected (CE not logged)	write MB tag set MB valid bit		VUAD write					set evict_rdy in MB Prefetch ICE written in MB dram_rdy not set

In C9, EVICT_READY bit gets set in MB. The instruction gets written to MB also in C9. However dram_ready bit does not get set in C9, thereby stopping issue of a DRAM read request.

TABLE 2-10 Timing Diagram: 2nd Pass of Prefetch ICE (Eviction plus Delete from Miss Buffer)

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag, VUAD read tag compare VUAD byp	pick replacement way from PA[21:18], Tag parity error NOT checked VUAD ECC checked, data corrected if SE detected. (CE not logged) stall two cycles to avoid collision with instruction after evict instruction for data array access	way sel logic mux out evicted tag way sel Xmit to l2d delete entry from MB	way sel xmit in l2d	data array read cyc1 I$ and D$ directory CAM write evicted tag in WBB write VUAD array	data array read cyc2	data array read cyc3 gen inval. vector	xmit inv response to dest cpx queues	inv packet to dest cpx data queues write WBB data (if dirty)	ready for WBK request to DRAM (if dirty)

Note that to ensure ordering, after the Prefetch ICE is inserted into the Miss Buffer, requests from crossbar and SIU are blocked from entering into the L2 pipeline, and the second pass of the Prefetch ICE (eviction pass) is not issued from the Miss Buffer until the Miss Buffer count becomes one (the Prefetch ICE is the only instruction in the Miss Buffer). After the second pass of Prefetch ICE completes, the stall of the crossbar and SIU requests are removed, and instructions that accumulated in IQ Array and Snoop queue can go through.

Diagnostic Read of Data Array:

TABLE 2-11 Timing Diagram: Diagnostic Read of Data Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag, VUAD read	disable way sel gen stall c1 instruc.	Gen way sel from decoded address Xmit way sel to l2d	Xmit way sel inside l2d	read data array cyc 1	read data array cyc 2	read data array cyc 3	Xmit 156 bits of data to l2t	Mux out 39 bits. Xmit req on CPX	Xmit 39 bits of data with rest of CPX packet

Diagnostic read access to the L2 data array is done through 64-bit read that access a 32-bit data subblock along with the corresponding 7-bit ECC. The instruction that gets used is Diagnostic Load.

Diagnostic Write of Data Array:

TABLE 2-12 Timing Diagram: Diagnostic Write of Data Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag, VUAD read	disable way sel gen stall c1 instruc.	Gen way sel from decoded address Xmit way sel to l2d	Xmit way sel inside l2d	write data array cyc 1	write data array cyc 2	write data array cyc 3	Gen inv vector = zeros	Xmit write ack to CPX	Mux data/inv vector

Diagnostic write access to the L2 data array is done through 64-bit store that access a 32-bit data subblock along with the corresponding 7-bit ECC.The instruction that gets used is Diagnostic Store.

Diagnostic Read of Tag Array

TABLE 2-13 Timing Diagram: Diagnostic Read of Tag Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
diagnostic decode	mux px2 index	read tag array	prepare way mux selects	mux out tag and flop	stage data	stage data	stage data	xmit req to CPX mux with data from other srcs (diagnostic/VUAD diagnostic/return data/inv vector)	xmit tag data to CPX

Diagnostic read to the L2 tag array is done through 64-bit read that accesses the tag along with the corresponding six-bit ECC. The instruction that gets used is Diagnostic Load.

Diagnostic Write of Tag Array:

TABLE 2-14 Timing Diagram: Diagnostic Write of Tag Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
do nothing	enable write into the tag stall pipe for 3 cycles	write into the tag	do nothing	do nothing	do nothing	do nothing	do nothing	xmit request to cpx	xmit ack to CPX

Diagnostic write to the L2 tag array is done through 64-bit write that accesses the tag along with the corresponding six-bit ECC. The instruction that gets used is Diagnostic Store.

Diagnostic Read of VD/UA Array:

TABLE 2-15 Timing Diagram: Diagnostic Read of VD/UA Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
VUAD array read	stall pipe for four cycles mux out appropriate bits (VD or UA based on address)	flop o/p data	flop o/p data	flop o/p data	flop o/p data	flop o/p data	flop o/p data	xmit req to requesting cpx mux VUAD data with data from other srcs	xmit data to cpx

Diagnostic read to the L2 VD/UA arrays is done through a pair of address access ranges. The first accesses the valid and dirty bits for an entire set plus the parity for each of those bits across the set via 64-bit read. The second range accesses the AU bits for the entire set via 64-bit read. The instruction that gets used is Diagnostic Load.

Diagnostic Write of VD/UA Array:

TABLE 2-16 Timing Diagram: Diagnostic Write of VD/UA Array

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
do nothing	stall pipe for four cycles	stage wr data	stage wr data	vuad write	do nothing	do nothing	do nothing	xmit req to CPX	xmit ack to CPX

Diagnostic write to the L2 VD/UA arrays is done through a pair of address access ranges. The first accesses the valid and dirty bits for an entire set plus the parity for each of those bits across the set via 64-bit write. The second range accesses the used and allocate bits for the entire set via 64-bit write. The instruction that gets used is Diagnostic Store.

Block Loads:

The core issues four 16B loads non-atomically. The L2 treats them as four separate load instructions.

Block Stores:

- Issued by the core with bis bit (bit 114) and bst bit (bit 115) both =1 on the PCX packet.
- On a hit in L2, each store behaves like a normal store.
- On an L2 miss, line is fetched from DRAM, allocated in L2 and updated in L2.
- On directory CAM access, all matching L1's are invalidated. In addition, the entry in the directory that got hit also gets invalidated.

Block Init Stores:

- Issued by the core with the bis bit (bit 114) of PCX packet = 1.
- On a hit in L2, each store behaves like a normal store.
- On a miss in L2, if PA[5:0] = 0, the line is initialized with all zeros by issuing a dummy read request to MCU instead of a regular read request, followed by the store happening.

- On a miss in L2, if PA[5:0]!= 0, it behaves like a regular store miss, in that it fetches the line from DRAM and then writes to it.
- On directory CAM access, all matching L1s are invalidated. In addition, the entry that got hit in the directory also gets invalidated.

Data Array Scrub:

Data array scrubbing refers to recomputing ECC for data across all ways in a particular index, detecting and correcting error in either data or ECC for each way. OpenSPARC T2 L2 data uses SEC/DED ECC (has a single bit error correction and double bit error detection). Through scrubbing single bit errors get corrected and double bit errors get flagged.

The Data Array in L2 gets scrubbed at regular (programmable) intervals after a data fill operation under configuration status register (CSR) control. If the CSR bit enabling the Scrub Mode bit is on, then after a programmed interval of time a bit called TECC gets set by the Scrub controller logic. The scrubber gets called on the first fill with this bit set and starts the scrub operation for an index (which increments with every scrub routine and has no relationship to the fill index) right after the fill completes in the pipe. It scrubs 64-bits of data at a time for each way, so it takes (8 x 16 ways) = 128 back to back scrub operations to complete the data scrub of that particular index in the L2 bank. While the scrub is going on the pipe stays stalled. Timing Diagram shows a typical data scrub operation following a fill.

TABLE 2-17 Timing Diagram: Fill

C1	C2	C3	C4	C5	C52	C6	C7	C8
tag write	stall four cycles	Xmit FB entry to l2b Fill Op with TECC = 1	way sel xmit to l2d read FB	VUAD write Xmit way sel inside l2d mux fbdata with l2d	data array wr cyc1	data array wr cyc2	data array wr cyc3	
	Start scrub FSM. stall pipe (cnt=0)	(cnt=1)	Setup tag read with scrub idx (cnt=2)	Read Tag Read Valid bit (cnt=3)	Gen scrub way (cnt=4)	Xmit Scrub way to l2d (cnt= 5)	Scrub read 1 (cnt = 6)	Scrub Read 2 (cnt=7)

TABLE 2-18 Timing Diagram: Data Scrub

C1	C2	C3	C4	C5	C52	C6	C7	C8
Scrub Read 3 (cnt=8)	Xmit to l2t (cnt=9)	ECC corr. (cnt = 10)	Mux out 64 bit (cnt = 11)	Mux with c1 inst data Perform stecc & gen waysel l2d_wr & col_off (cnt = 12)	way sel xmit to l2d (cnt = 13)	Xmit way sel inside l2d (cnt = 0) Start scrub FSM. stall pipe	Scrub data array wr cyc1 (cnt = 1)	Scrub data array wr cyc2 (cnt = 2) Setup tag read with scrub idx

Tag Array Scrub:

Tag array scrubbing refers to recomputing ECC for tag across all ways in a particular index, detecting and correcting error in either tag or ECC for each way. OpenSPARC T2 L2 tag uses SEC ECC (has a single bit error correction, no double error detection). Through scrubbing single bit errors get corrected.

Once a parity error is detected in C2 on any Tag entry, TECC is marked as 1 in the instruction that gets written to the Miss Buffer in C3. Then the scrub instruction gets issued from the MB and enters the pipe. All 16 ways for the index with parity error gets scrubbed, so the scrub operation takes a total of (8x16) = 128 L2 clocks during which the L2 pipeline stays stalled. Timing diagram TABLE 2-20 shows a typical Tag Scrub operation.

TABLE 2-19 Timing Diagram: Tag Scrub Operation

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tecc inst from MB assert pipe stall	Setup Tag read of corrupted index (cnt=0)	Setup Index (cnt = 1)	Tag Read (cnt = 2)	Setup Muxsel (cnt = 3)	Mux Tag (cnt = 4)	ECC corr. (cnt = 5)	Setup Write Index (cnt = 6)	Tag Write (cnt = 7)	Setup Tag read of corrupted index (cnt=0)

VUAD SBE Correction:

OpenSPARC T1 protects VD and UA arrays with parity. Parity check happens in C2, and if an error is detected, a fatal error trap gets taken. So for a false hit in C1, the read/write happens, but the machine gets fatal error trap and resets. Since this is one of the largest sources of fatal errors, OpenSPARC T2 would protect the VD and UA arrays with SEC DED ECC.

For every set, there would be seven ECC bits with 16 Dirty bits and 16 Valid bits (i.e. 32/7 ECC). Also for every set, there would be seven ECC bits with 16 Allocate bits and 16 Use bits (i.e. 32/7 ECC). Note that the seventh ECC bit is also the parity bit across 32 data bits and six ECC bits to detect double bit error. So a total of (39 + 39) = 78 bits of storage per set. Even though the Used bits need not be ECC protected (since their value is non-critical: any error in the used bits will cause potentially different replacement order, but still functionally correct operation), since VD and UA arrays would be implemented out of the same Register File array, L2 would protect the Use bits and the Allocate bits with ECC.

For any instruction from core or from SIU, the VD and UA arrays get read in C1 stage of the L2 pipe and get muxed with forwarded VD,UA bits from prior instructions in the pipe that are to the same index. The output of the mux gets written to a C2 flop. In case this C1 mux select points to the leg coming from the VD/UA arrays, ECC would get checked in C2 on the data from the arrays. The data will get corrected in C2 stage itself (for a single bit error) and will get written back to the VD,UA arrays with regenerated ECC in C5 stage of the pipe for all instructions other than diagnostic accesses. If a double bit (Uncorrectable) error gets detected in any one of VD or UA arrays, L2 will log LVU in L2 Error Status register which will cause L2 to assert fatal error reset request to the Reset block. The execution recovery and logging mechanism for Correctable VD/UA errors in L2 is as follows:

1. If the instruction was any flavor of load, store, atomic, ifetch from core, L2 would detect the Correctable error in C2 and log it as LVC (VUAD correctable error) with the syndrome in L2 Error Status register in C9 stage of the same pass.The PA[39:0] (index [8:0]) would be captured in the L2 UE/CE address register. The caming of the L2 directories, updates of the L2 directories and dispatch of crossbar packets back to cores would be gated off by the correctable error. If the error resulted in a false hit (tag match but valid = 1 while it should be 0), the data array operation (load or store) would still happen though the crossbar packet would be gated off. There is no memory corruption issue as store to an invalid way does not cause data corruption. Also the instruction would be moved into the Miss Buffer and readied for reissue in C9. However the DRAM ready bit would not be set in the Miss Buffer thereby disabling dispatch of requests to MCU. Once the instruction gets reissued down the L2 pipe, it would see corrected data in the VD/UA arrays and would execute properly based on correct state of the L2 lines at that index. The LVC error that got logged in the first pass would get reported to the Virtual Core specified in the

L2_CONTROL_REG.ERRORSTEER field on bits [139:138] of Error Indication packet of crossbar after the occurrence of the next L2 fill if error reporting is enabled.

2. If the instruction was a RDD or WR8 or WRI from SIU, L2 would detect the Correctable error in C2 and log it as LVC (VUAD correctable error) with the syndrome in L2 Error Status register in C9 stage of the same pass.The PA[39:0] (index [8:0]) would be captured in the L2 UE/CE address register. The caming of the L2 directories, updates of the L2 directories and dispatch of data return and write ack packets in the first pass back to SIU would be gated off by the correctable error. If the error resulted in a false hit (tag match but valid = 1 while it should be 0), the data array operation (load or store) would still happen though the SIU packet would be gated off. There is no memory corruption issue as store to an invalid way does not cause data corruption. Also the instruction would be moved into the Miss Buffer and readied for reissue in C9. However the DRAM ready bit would not be set in the Miss Buffer thereby disabling dispatch of requests to MCU. Once the instruction gets reissued down the L2 pipe, it would see corrected data in the VD/UA arrays, would execute properly based on correct state of the L2 lines at that index and would return packets (read data, wri ack,wib_dequeue) to SIU. The LVC error that got logged in the first pass would get reported to the Virtual Core specified in the L2_CONTROL_REG.ERRORSTEER field on bits [139:138] of Error Indication packet of crossbar after the occurrence of the next L2 fill if error reporting is enabled.

3. If the instruction is of any of the types mentioned in (1) and (2) but issued from the Miss Buffer, all the things mentioned in (1) and (2) happen with one exception: the instruction does not get re-inserted into the Miss Buffer, but the valid bit in the Miss Buffer stays set, and the instruction gets replayed again through the vuad_ce_rdy and vuad_ce_replay bits being set in the Miss Buffer.

4. If the instruction was an evict instruction issued from the Miss Buffer and there is a VUAD CE detected and the instruction is not a Prefetch ICE instruction, the Evict Bit and Evict Ready bit stay set in the Miss Buffer, so that the eviction gets replayed from the Miss Buffer. By this time the VUAD CE has been corrected and the eviction happens as normal. Also in the eviction pass that detected the VUAD CE, the caming of the L2 directories, updates of the L2 directories and dispatch of crossbar packets back to cores and copying of the line to the Write Back Buffer are gated off by the correctable error.

5. If the instruction was any flavor of diagnostic, Icache Invalidate, Dcache Invalidate, Tecc, Fill, Prefetch ICE or replayed instruction from Miss Buffer hitting in the Fill Buffer Tags, L2 would not detect the Correctable error and would not log LVC. Also the instruction will proceed as normal as if nothing happened and would send back responses to the cores as normal. However for all of the above mentioned instructions other than Diagnostics, the corrected data would get written in the VUAD arrays in C5, basically doing a silent correction.

Timing Diagram TABLE 2-20 shows the error detection and correction pass for VUAD single bit correctable error for loads, stores, ifetches, atomics, wr8s, rdds, and wris.

TABLE 2-20 Timing Diagram: VUAD SBE Error Detection and Correction

C1	C2	C3	C4	C5	C52	C6	C7	C8	C9
tag read VUAD read VUAD bypass	VUAD ECC check Single Bit Error Detected	Write Instruction to MB Set DEP bit Disable SIU ack gen logic, IOWB eviction logic for WRIs	Gate of I$,D$ Dir CAM, I$,D$ Dir Update	VUAD write (corrected data			Gate off crossbar request		Ready Instruction for reissue Do not set DRAM_READY

This pass gets followed by the instruction reissue from Miss Buffer by which time the error is already corrected, and the instruction executes normally as shown by earlier pipe diagrams in the document.

Ordering of Future Instructions in L2 in the case of a VUAD SBE:

All instructions other than diagnostic, Icache Invalidate, Dcache Invalidate, Tecc, Fill or Prefetch ICE (and not issued from the Miss Buffer) that would detect VUAD SBE would set the DEP bit when they get inserted into the Miss Buffer. This would guarantee that a future instruction to the same PA would hit in the Miss Buffer and see the DEP bit set and would also get moved into the Miss Buffer. Also, this future instruction would not be issued to the pipe until the offending instruction has been issued to the pipe and its DEP bit cleared. This would maintain ordering and remove hazards. However it is possible that instructions to different PA would send acks and data back to crossbar and SIU out of order with respect to order of arrival to L2. For loads, ifetches, atomic reads and rdds this would not be an issue. However for stores out of order acks may be an issue if loads following the stores return old data instead of new data (this would have caused TSO ordering violation). But since we put the offending instruction in the Miss Buffer and set the DEP bit, loads would always be ordered after stores to the same address and would not complete until the store completes, and hence would return new data.

Also in case the instruction above hits against another address in the Miss Buffer and detects a VUAD CE, the instruction would set its DEP bit and get inserted into the Miss Buffer, but would not get replayed immediately. It would get replayed only when its dependency cleared. In this case the instruction would not set the vuad_ce_rdy and vuad_ce_replay bits in the Miss Buffer.

For any instruction other than diagnostic, Icache Invalidate, Dcache Invalidate, Tecc, Fill or Prefetch ICE and issued from the Miss Buffer that detects VUAD SBE, the DEP bit would not get set, however the Valid bit would not be cleared and the instruction would not get re-inserted into the Miss Buffer. The instruction would get replayed from the Miss Buffer through the vuad_ce_rdy and vuad_ce_replay bits being set in the Miss Buffer.

Tag Parity Error and VUAD Error detected in Single Pass:

Any instruction from core or SIU other than Diagnostics, Prefetch ICE and I$,D$invalidates can detect tag parity error and VUAD SBE. It is architecturally possible for such an instruction to detect a tag parity error and VUAD SBE in the same pass. If that happens, the recovery and correction mechanism will be as follows:

1. Instruction would be inserted into the Miss Buffer in the first pass (in which it detected tag parity error and VUAD ce) and will be readied for reissue not in C9 but by the scrub instruction that is to follow.
2. The VUAD SBE would get corrected in the first pass itself (with the VUAD array getting updated in c5).
3. A scrub instruction will get issued from Miss Buffer which will walk down L2 pipe and do the scrub of the tag array
4. Original instruction will get issued from the Miss Buffer and this time will detect neither tag parity error nor VUAD SBE and will complete as normal.

2.1.4.11 L2 Interactions with SIU (System Interface Unit)

Block Reads (RDDs):

Block Read from SIU goes through L2 pipe like a regular load from the core. On a hit, 64 B of data is returned to SIU. On a miss, L2 does not allocate, but sends a non-allocating read to DRAM. It gets 64 bytes of data from DRAM and sends it back to SIU (read once data only) directly without installing in the L2 cache.

TABLE 2-21 Timing Diagram: Block Reads

C1	C2	C3	C4	C5	C52	C6	C7	C8	
tag,VUAD read VUAD bypass tag compare Check ECC for Tags MB CAM and MB hit logic FB CAM WBB CAM	way sel logic xmit way sel to l2d rd/wr! Gen,xmit VUAD ECC check stall next instruction	way sel xmit in l2d FB data read enable (on a miss)	data array read cyc1 FB data read cycle 1 write 64 B Fbdata to l2d flop	data array read cyc2 VUAD write stage FB data	data array read cyc3 mux with FB data write 64 B data to flop in l2d	data Xmit cyc 1 write 64 B data to flop in l2b	stage 64 B data in flop in l2b	32:1 Mux to get 32-bits of data from 64 bytes (in l2b,critical 32-bits first)	Check ECC on data (32-bits) flop data in l2b

In C10, L2 starts issuing data return to SIU 32-bits per clock from l2b block. While processing a block read from SIU, the L2 arbiter does not accept any other SIU read or write request until the block read is complete. This is because there is not enough queue space within L2 to hold the data that is getting streamed out in case a new block read request comes from SIU.

Write Invalidates (WRIs):

For a 64 B write (write invalidate from SIU), the SIU issues a 64 B write request to L2. The data goes to IOWB and waits there until the write makes it through the pipe after resolving any dependencies with the Miss Buffer entries (resolves ordering issues w.r.t prior accesses from the CPU to the same line). Once this happens, the IOWB empties its contents to DRAM, after arbitrating with the WBB.

When the write progresses through the pipe, it looks up the tags. If tag hit, it invalidates the entry and all L1 entries that match. If tag miss, it does nothing (just comes down the pipe) to maintain order. The only two cases where a WRI gets put into the Miss Buffer are on tag parity error (potential false miss case) or VUAD SBE (can be anything: true miss, false miss, true hit, false hit).

Partial Line Writes (WR8's):

When the SIU issues 8B writes to L2 with random byte enables, the L2 treats them just like 8B stores from core (i.e. does two pass partial store if odd number of byte enables are active or if misaligned access, otherwise regular store). Data gets committed to L2 cache.

2.1.4.12 L2 Pipeline Stalls

- Same column stall - Each column (sub-bank) of the data array (also referred to as subbanks) requires two cycles to access. Therefore, the same column cannot be accessed in consecutive cycles. A one cycle stall is inserted if a collision is detected.
- Ifetch - Ifetch operations require two reads of the data array. A one cycle stall is inserted for any ifetch operation.
- Fill - Fill operation stalls the pipe for stall three cycles
- Eviction - Eviction operation stalls pipe by two cycles
- Diagnostics
- Tag/Data array scrubs

2.1.5 Functional Description of Sub-blocks

The L2 Cache Unit is composed of the following sub-blocks:

- L2 Tags
- L2 VUAD
- L2 Data
- Directory
- Input Queue (IQ)
- IU Queue (SIUQ)
- Output Queue (OQ)
- Arbiter
- Miss Buffer (MB)

- Fill Buffer (FB)
- Writeback Buffer (WBB)
- I/O Write Buffer (IOWB)

2.1.5.1 L2 Tags

TABLE 2-22 shows the physical address mapping for the L2 cache.

TABLE 2-22 Physical Address Mapping for the L2 Cache

39	18	17	9	8	6	5	4	3	0
tag		index		L2 bank		subbank addr		16b offset	

Given a bank of 512 KB with 64 B lines, the tag index is bits <17:9>. Each tag entry contains address<39:18> + six ECC bits. The state of each line is maintained using valid (V), used (U), allocated (A), and dirty (D) bits. These are stored in the VUAD array.

Each 22-bit tag is protected by six-bits of ECC. A 16 way 27-bit compare (with the appropriate bits from the issuing instruction) is performed to generate the way selects for accessing the data array. This approach removes error detection from the "tag to data" critical path.

Thus total Tag Memory per bank of L2 is 28 KB for 64 B line size.

L2 Tag ECC

The L2 tag arrays are protected by ECC. For every 22-bits of tag, there are five SEC ECC bits and one parity bit (which covers all 27-bits). OpenSPARC T1 L2 does not detect Uncorrectable (Double Bit) errors for tag and OpenSPARC T2 L2 won't either (unless we get Epic 9 data which shows high enough Failure In Time (FIT) rate).

In pipe stage C1, {22 tag bits, five ECC bits} get compared with corresponding 27-bits in all of the 16 ways in the set. This prevents a false hit from happening. If {22 tag bits, five ECC bits} match in one entry, then it is a true hit. If a true hit does not happen, then it is miss. As the instruction moves to pipe stage C2, parity is recommitted for each of 16 ways over {22 tag bits, five ECC bits}. If there is a parity error detected in C2, the instruction is moved into Miss Buffer and a scrub instruction is issued from the Miss Buffer to scrub and correct ECC and parity for any entry in error. After the scrub instruction is complete, the original instruction gets re-issued from the Miss Buffer.

Note that if the instruction did not hit in any of the tags in C1, there are four possibilities:

- There is no parity error in C2. In which case it would be a true miss and go to DRAM.
- There is a parity error in C2 and one of the ways differs by one bit only in the tag field with respect to the instruction, then this could be a case of a false miss (if upon scrubbing this bit that is different gets chosen as the bit to be flipped). Then the instruction after getting replayed from the Miss Buffer will hit in L2.
- There is a parity error in C2 and one of the ways differs by one bit only in the tag field with respect to the instruction, but the error is in one of the ECC bits or on another data bit, in which case the address is different, and it will miss in L2 after getting replayed from Miss Buffer and will go to DRAM.
- There is a parity error in C2 and instruction tag mismatches by more than one bit with each way, it will miss in L2 after getting replayed from Miss Buffer and will go to DRAM.

Note – What happens if tag hits in two ways? OpenSPARC T2 L2 response is indeterministic. The SRAM circuits are protected from getting burnt out in such a case. The functional behavior of L2 is not defined.

2.1.5.2 L2 VUAD

This 4.9 KB (including ECC) dual ported array is used to maintain the state of every line in the L2 cache for each bank. The state of each line is maintained using the Valid (V), Used (U), Allocate (A) and Dirty (D) bits. Allocate bit indicates that the marked line has been allocated to a miss. This bit is also used in the processing of some special instruction's such as atomics and "partial" stores (since these do read-modify-writes, which involve two passes through the pipe, the line needs to be locked until the second pass completes; otherwise the line may get replaced before the second pass happens). The Used bit is a reference bit used in the replacement algorithm.

The Allocate bit (per way) gets set when a line gets picked for replacement. For a load or ifetch, it gets cleared when fill happens, and for store when store completes. The Used bit gets set when any store/load hits (1 per way). Used bits get cleared (all 16 at a time) when there are no unused or unallocated entries for that set. The dirty bit (per way) gets set when a stores modifies the line. It gets cleared when the line is invalidated. The valid bit (per way) gets set when a new line is installed in that way. It gets reset when that line gets invalidated.

The L2 uses a Round Robin algorithm to pick replacement candidates. When there is a L2 miss and a line needs to be selected for replacement, in reference to a dynamic pointer which can point to any one of the 16 ways, the first entry with (A = 0 and U = 0) otherwise (A = 0) gets picked for replacement. All 16 ways get looked at in a wrap-around fashion starting from the way that is currently pointed at by this pointer.

The algorithm used to select a way, out of 16 ways, to be evicted out of the L2 cache is not a true LRU algorithm but Round Robin arbitration. Round Robin arbitration is done in two stages by dividing 16 ways in four quads of four ways each. First Round Robin is done within each quads to select one of the four ways and then Round Robin is done to select one of the four quads. A four bit state register is kept for each quad at each level. A one on a bit location corresponding to a way represents highest priority for that way. Every time an eviction takes place, state register is updated by shifting it left by one bit otherwise state of the register does not change. State register is used in C2 for the way selection and it is updated in the C3. On reset state register is initialized to a state such that way0 has the highest priority.

Way selection algorithm depends on the Used and Allocate bit of the VUAD array, read during C1, for the way selection. First priority is given to the ways that has not been Used and has not been Allocated for the eviction in the previous cycle. If there is no Unused and Unallocated way then a way that has not been previously Allocated is given preference. Invalid bit is not used for the way selection as if a way is Invalid then its Used bit will not be set, so checking Invalid bit is redundant.

Note – What happens if all 16 ways have A = 1? No new instructions can enter the pipe until at least one A = 0. This is guaranteed by MB stalling the pipe speculatively. Since MB is 32 entries in OpenSPARC T2, and since there are 16 ways per set, the L2 control logic will detect 12 entries of the same index in MB and speculatively stop accepting requests from PCX and SIUQ (this accounts for four instructions in flight in PX2, C1, C2, and C3 that can take the count to 16 misses to the same index). The stall to PCX/IQ and SIUQ requests lasts for until the number of entries of the same index in MB reaches 11.

2.1.5.3 L2 VUAD ECC

OpenSPARC T1 protects VUAD array with parity. Parity check happens in C2, and if an error is detected, a fatal error trap gets taken. So for a false hit in C1, the read/write happens, but the machine gets fatal error trap and resets. Since this is one of the largest sources of fatal errors, OpenSPARC T2 would protect the VD and UA arrays with SEC DED ECC.

For every set, there would be seven ECC bits with 16 Dirty bits and 16 Valid bits (i.e. 32/7 ECC). Also for every set, there would be seven ECC bits with 16 Allocate bits and 16 Use bits (i.e. 32/7 ECC). Note that the seventh ECC bit is also the parity bit across 32 data bits and six ECC bits to detect double bit error. So a total of (39 + 39) = 78 bits of storage per set. Even though the Used bits need not be ECC protected (since their value is non-critical: any error in the used bits will cause potentially different replacement order, but still functionally correct operation), since VD and UA arrays would be implemented out of the same Register File array, L2 would protect the Use bits and the Allocate bits with ECC.

For any instruction from core or from SIU, the VD and UA arrays get read in C1 stage of the L2 pipe and get muxed with forwarded VD,UA bits from prior instructions in the pipe that are to the same index. The output of the mux gets written to a C2 flop. In case this C1 mux select points to the leg coming from the VD/UA arrays, ECC would get checked in C2 on the data from the arrays. The data will get corrected in C2 stage itself (for a single bit error) and will get written back to the VD,UA arrays with regenerated ECC in C5 stage of the pipe for all instructions other than diagnostic accesses. If a double bit (Uncorrectable) error gets detected in any one of VD or UA arrays, L2 will log LVU in L2 Error Status register which will cause L2 to assert fatal error reset request to the Reset block. If L2 detects Correctable SBE in C2, it will log it as LVC (VUAD correctable error) with the syndrome in L2 Error Status register in C9 stage of the same pass. The index [8:0] would be captured in the L2 UE/CE address register.

2.1.5.4 L2 Data

Each L2 data array bank is a single ported SRAM structure capable of performing the following operations:

- 16B read
- 64B read
- 8B write with any combination of word enables
- 64B write (with any combination of word enables). However fills would update all 64 bytes at a time.

Each L2 bank is 512 KB in size, with each logical line 64 B in size.

Each L2 data array bank is further subdivided into four sub-banks, also referred to as columns, each 16 B in width. These sub-banks are accessed based on bits <5:4> of the physical address. Loads (which are a maximum of 16 B in size) and stores (maximum of 8 B in size) access one subbank. Cache fills and line evictions are 64 B in size, and access four sub-banks per cycle.

Any L2 cache data array access takes two cycles to complete, so no sub-bank can be accessed in consecutive cycles. All access can be pipelined except, back to back accesses to the same sub-bank.

Each 32b word is protected by seven bits of SEC/DED ECC. (Each line is 32 x [32 + 7 ECC] = 1248 bits). All sub-word accesses require a read modify write operation to be performed and are referred to in this document as "partial stores".

2.1.5.5 L2 Directory

The directory maintains a copy of the L1 tags for coherency management and also ensures that the same line is not resident in both the icache and dcache (across all cores). The directory is split into an icache directory (icdir) and a dcache directory (dcdir), which are similar in size and functionality.

The directory is written only when a load is performed. On certain data accesses (loads, stores and evictions), the directory is camed to determine whether the data is resident in L1 caches. The result of this CAM operation is a set of match bits which is encoded to create an invalidation vector to be sent back to the SPARC Cores to invalidate L1 lines.

- Loads - The icdir is camed to maintain I/D exclusivity. The dcdir is updated to reflect the load data that fills the L1 cache.
- IFetch - The dcdir is camed to maintain I/D exclusivity. The icdir is updated to reflect the instruction data that fills the L1 cache.
- Stores - Both directories are camed. This ensures that (i) if the store is to instruction space, the L1 icache invalidates the line and does not pick up stale data; (ii) if a line is shared across SPARC Cores, the L1 dcache invalidates other cores and does not pick up stale data; and (iii) the issuing core has the most current information on the validity of its line.
- Evictions from the L2 cache - Both directories are camed to invalidate any line that is no longer resident in the L2.

2.1.5.6 Directory Organization

D$ Dir:

There are eight cores in OpenSPARC T2. L1 Dcache for each core has 128 sets, each set has four ways. Since L1 Dcache line size is 16 B, this gives a total of (8 x 128 x 4 x 16) bytes = 64 KB or 4K L1 lines in all cores together. Each L1 Dcache is 8KB.

Thus each L1 Dcache will map (128/8) 16 sets to each L2 bank. So for each L2 bank, the DCache directory will consist of (16 x 8) sets of L1 Dcache lines for all cores combined. This gives a total of 512 L1 Dcache lines per bank.

These 512 L1 Dcache line mappings gets organized physically in the DCache directory as follows:

Each Dcache directory has 16 panels arranged as four rows and four columns. Row gets accessed by address[5:4], column by address[10,9]. Each panel has 32 entries indexed by {cpu_id(three-bits), replacement way (two-bits)}.

For an update related to a load, the panel is accessed by address {10,9,5,4} and the entry within the panel to be updated is selected by {cpu_id(three-bits), replacement way (two-bits)} for the load.

For a store or a ICache mutual-exclusivity check on a Ifetch, which can potentially invalidate a maximum eight L1 cache lines (one L1 line per core), the panel gets selected by address {10, 9, 5, 4} and all 32 entries within that panel get camed against the store, based on which a invalidation vector gets generated for a max of eight L1 Dcache line invalidation's (one per core). The way number for each L1 Dcache will be encoded as a two-bit field in the inval vector.

For an eviction, since the L2 cache line is 64 bytes, four panels out of 16 will get camed based on address[10,9] (i.e. one column). This would mean a total of 32x4 = 128 compares to invalidate a max of four cache lines per L1, i.e. a max of 4x8 = 32 L1 Dcache lines for all cores combined. This will come out as 32 D$ L1 lines eviction vector from L2. The way number for each L1 Dcache will be encoded as a two-bit field in the inval vector.

Each entry in the directory will store {L2 index[9-bits], L2 way[4-bits], parity, valid} i.e. a total of 15-bits corresponding to the location in L2 that the L1 line maps to.

For a load hit, the entry gets updated with {L2 index[9-bits], L2 way[4-bits], parity}, while on a store or eviction or a ICache mutual-exclusivity check on a Ifetch, the {L2 index[9-bits], L2 way[4-bits]} gets camed against the stored value of each entry.

I$ Dir:

L1 Icache for each core has 64 sets, each set has eight ways. Since L1 Icache line size is 32 B, this gives a total of (8 x 64 x 8 x 32) bytes = 128 KB or 4K L1 Icache lines in all cores together. Each L1 Icache is 16 KB.

Thus each L1 Icache will map (64/8) = eight sets to each L2 bank. So for each L2 bank, the ICache directory will consist of (8 x 8) sets of L1 Icache lines for all cores combined. This gives a total of 512 L1 Icache lines per bank.

These 512 L1 Icache line mappings gets organized physically in the ICache directory as follows:

Each Icache directory has 16 panels arranged as four rows and four columns. Row gets accessed by {address[5], I$ replacement way[2]}, column by address[10,9]. Each panel has 32 entries indexed by {cpu_id(3-bits), I$ replacement way[1:0]}.

For an update related to a Ifetch hit, the panel is accessed by {address {10,9,5}, I$ replacement way[2]} and the entry within the panel to be updated is selected by {cpu_id(3-bits), I$ replacement way[1:0]} for the Ifetch hit.

For a store or a Dcache mutual-exclusivity check on a load, which can potentially invalidate a maximum eight L1 Icache lines (one line per core), two panels gets selected by address {10,9,5} and all 32 entries within each panel get camed against the store, based on which a invalidation vector gets generated for a max of eight L1 Icache line invalidation's (one per core). The way number for each L1 Icache line will be encoded as a 3-bit field in the inval vector.

For an eviction, since the L2 cache line is 64 bytes, four panels out of 16 will get camed based on address[10,9] (i.e. one column). This would mean a total of 64x2 = 128 compares to invalidate a max of two Icache lines per L1, i.e. a max of 2x8 = 16 L1 Icache lines for all cores combined. This will come out as 16 I$ L1 lines eviction vector from L2. The way number for each L1 Icache line will be encoded as a 3-bit field in the inval vector.

Each entry in the directory will store {L2 index[9-bits],L2 way[4-bits], parity, valid} i.e. a total of 15-bits corresponding to the location in L2 that the L1 line maps to.

For a Ifetch hit, the entry gets updated with {L2 index[9-bits],L2 way[4-bits], parity}, while on a store or eviction or a Dcache mutual-exclusivity check on a load, the {L2 index[9-bits],L2 way[4-bits]} gets camed against the stored value of each entry.

2.1.5.7 SIU Queue (SIUQ)

The SIU Queue accepts RDD,WRI and WR8 packets from the SIU and issues them to the pipe after arbitrating against other requests.L2 SIU Queue block can record up to two requests from SIU in it's two-deep FIFO. The requests are received serially. A counter is maintained in the SIU side incrementing on a transaction dispatch to L2 cache and decrementing upon receiving l2t_siu_iq_dequeue or l2t_siu_wib_dequeue signals from the L2 cache. l2t_siu_iq_dequeue signal is asserted when an instruction is issued down the L2 pipe (RDD, WRI, and WR8 instructions). l2t_siu_wib_dequeue is asserted when the contents of a I/O Write Buffer entry gets streamed to DRAM (WRI).

2.1.5.8 Input Queue (IQ)

The input queue is a 16 entry FIFO which queues packets arriving on the PCX when they cannot be immediately accepted into the L2 pipe. Each entry in the IQ is 130 bits wide. The FIFO is implemented with a dual ported array. The write port is used for writing into the IQ from the PCX interface. The read port is for reading contents for issue into the L2 pipeline. If the IQ is empty when a packet comes on the PCX, the packet can pass around the IQ if it is selected for issue to the L2 pipe.

The IQ asserts a stall to the PCX when 11 entries are used in the FIFO. This allows for packets already in flight as shown in

TABLE 2-23 Input Queue Pipeline

PQ	A	B	C	D	E	F				
PA		A	B	C	D	E	stall			
PX			A	B	C	D	E			
PX?				A	B	C	D	E		
C1					A	B	C	D	E	
C@ (count)						12	13	14	15	16

2.1.5.9 Output Queue (OQ)

The output queue is a 16 entry FIFO which queues operations waiting for access to the CPX. Each entry in the OQ is 146 bits wide. The FIFO is implemented with a dual ported array. The write port is used for writing into the OQ from the L2 pipe. The read port is for reading contents for issue to the CPX. If the OQ is empty when a packet comes from the L2 pipe, the packet can pass around the OQ if it is selected for issue to the CPX.

Multicast requests are dequeued from the FIFO only if all the destination CPX queues can accept the response packet.

When the OQ reaches its high water mark, the L2 pipe stops accepting inputs from the Miss Buffer or the PCX. Fills can happen while the OQ is full since they don't generate CPX traffic. The high water mark is TBD, which accounts for instructions already in the L2 pipe.

2.1.5.10 Arbiter

The arbiter manages access to the L2 pipeline from the various sources which request access. The IQ, MB, IO interface, and FB all need access to the L2 pipe. Access to the pipe is granted based on the following priority:

- Access currently stalled in the pipe
- Second packet of a CAS operation
- SIU instruction from SIU Queue
- Miss Buffer instruction
- Fill Buffer instruction
- Instruction from the IQ

- Background scrub request

2.1.5.11 Miss Buffer (MB)

The Miss Buffer (MB) has 32 entries and stores instructions which cannot be processed as a simple cache hit. This includes true L2 cache misses (no tag match), instructions that have the same cache line address as a previous miss or an entry in the Writeback Buffer, instructions requiring multiple passes through the L2 pipeline (atomics and partial stores), unallocated L2 misses, and accesses causing tag ECC errors.

Miss Buffer in OpenSPARC T2 L2 would be 32 entries instead of 16 as in OpenSPARC T1 L2. This is needed to reduce L2 stalls due to Miss Buffer full which affects the CPI of each of the 64 threads (due to the fact that Miss Buffer going full stalls all accesses to L2; load hits and store hits cannot happen). With DDR 280 Mhz, for TPCC, CPI per thread improves by 8% between 16 and 32 entry Miss Buffer and L2 stall due to MB full reduces from 6.4% to 0.08%. With DDR 333 Mhz, for TPCC, CPI per thread improves by about 7.6% with 32 entry Miss Buffer while L2 stall due to MB full goes closer to zero.

The Miss Buffer is divided into a dual ported RAM portion which holds store data and a CAM portion which contains the address.

A read request is issued to DRAM and the requesting instruction is replayed when the "critical quad-word" of data arrives from DRAM.

All entries in the Miss Buffer that share the same cache line address are linked in the order of insertion to preserve ordering. Instructions to the same address are processed in age order whereas instructions to different addresses are not ordered and exist as a free list.

When a MB entry gets picked for issue to the DRAM (load, store, ifetch miss), the entry gets copied into the Fill Buffer and a valid bit gets set. There can be up to eight reads outstanding from L2 to DRAM at any point of time. Data can come from DRAM to L2 out of order with respect to the address order. When the data comes back out of order, the MB entries get readied for issue in the order of data return. This means that there is no concept of age in the order of data returns to core as these are all independent accesses to different addresses. Thus when a later read gets replayed from the MB down the pipe and invalidates its slot in the MB, a new request from the pipe will take its slot in the MB, even while an older read has not yet returned data from DRAM.

In most cases, when a data return happens, the replayed load from the MB makes it through the pipe before the Fill Request can. Hence the valid bit of the MB entry gets cleared (after the replayed MB instruction execution is complete in the pipe) before the Fill Buffer valid bit. However if there are other prior MB instructions like partial stores that get picked instead of the MB instruction of concern, the fill request can

enter the pipe before the MB instruction and in those cases the valid bit in the Fill Buffer would get cleared prior to the MB valid bit. Thus the MB valid bit and FB valid bits always get set in the order of MB valid first, FB valid later. However they can get cleared in any order.

When the MB reaches its high water mark, the arbiter no longer accepts requests from the IQ or PCX. The high water mark is TBD, which accounts for instructions that are already in the pipe that may be inserted into the MB.

2.1.5.12 Fill Buffer (FB)

The Fill Buffer is an eight entry buffer used to temporarily store data arriving from DRAM on an L2 miss request. Data arrives from DRAM in four 16 B blocks starting with the critical quad-word. A load instruction waiting in the Miss Buffer can enter the pipeline after the critical quad-word arrives from DRAM (critical 16B will arrive first from DRAM) In this case, the data is bypassed. After all four quad-words arrive, the fill instruction enters the pipeline and fills the cache (and the Fill Buffer entry gets invalidated). For a non-allocating read (e.g. I/O read), the data gets drained from the Fill Buffer directly to the I/O Interface when data arrives, and the Fill Buffer entry gets invalidated.

When the FB is full, the Miss Buffer cannot make requests to DRAM.

The Fill Buffer is divided into a RAM portion which stores the data returned from DRAM waiting for a fill to the cache and a CAM portion which contains the address.

2.1.5.13 Writeback Buffer (WBB)

The Writeback Buffer is an eight entry buffer used to store dirty evicted data from the L2 on a miss. Evicted lines are streamed out to DRAM opportunistically. An instruction whose cache line address matches the address of an entry in the WBB is inserted into the Miss Buffer. This instruction must wait for the entry in the WBB to write to DRAM before entering the L2 pipe.

When the WBB reaches its high water mark, the arbiter no longer issues instructions from the Miss Buffer. This stops read requests to DRAM and allow writebacks to proceed. The high water mark is TBD, which accounts for evictions that are already in the pipe.

The Writeback Buffer is divided into a RAM portion which stores the evicted data until it can be written to DRAM and a CAM portion which contains the address.

2.1.5.14 I/O Write Buffer (IOWB)

The I/O Write Buffer is a four entry buffer which stores incoming data from the PCI-EX interface in the case of a 64 B write operation. Since the PCI-EX interface bus width is only 32-bits wide, the data must be collected over 16 cycles before writing to DRAM. An instruction whose cache line address matches the address of an entry in the IOWB is inserted into the Miss Buffer. This instruction waits for the entry in the IOWB to write to DRAM before entering the L2 pipe.

The I/O Write Buffer is divided into a RAM portion which stores the data from the IO interface until it can be written to DRAM and a CAM portion which contains the address.

It is the responsibility of the IO interface to use a handshaking protocol to track the state of the IOW Buffer.

The IO interface must never issue an operation requiring the buffer when the buffer is full.

2.1.6 Unit-level Interface Signals

TABLE 2-24 Unit Level Interface Signals

Signal Name	I/O	Size	From/ To	Timing	Description
Crossbar					
l2t_cpx_req_cq	O	8	CCX		Request to be drained out of L2
l2t_cpx_data_ca	O	146	CCX		Data from L2 cache
cpx_l2t_grant_cx	I	8	CCX		Grant to gain access to crossbar
l2t_cpx_atom_cq	O	1	CCX		First packet of Imiss
l2t_pcx_stall_pq	O	1	PCX		Cannot accept any more requests to L2Cache from core since the Input FIFO is full.
pcx_l2t_data_rdy_px1	I	1	PCX		Cannot accept any more requests to L2Cache from core since the Input FIFO is full.
pcx_l2t_data_px2	I	130	PCX		Data bus from core
pcx_l2t_atm_px1	I	1	PCX		Indicates atomic instruction
SIU					
l2t_sii_iq_dequeue	O	1	SIU		Entry in a IOWBB array has freed. l2t is unloading a request
l2t_sii_wib_dequeue	O	1	SIU		Write invalidate buffer (size= 4x64 B cache lines) is being unloaded
l2b_sio_data	O	32	SIU		Read Data to SIU.
l2b_sio_ue_err	O	1	SIU		UE on read data to SIU
l2b_sio_ctag_vld	O	1	SIU		Ack to SIU from L2
sii_l2t_req_vld	I	1	SIU		SIU request valid.
sii_l2t_req	I	32	SIU		SIU requests L2 cache to be serviced.
sii_l2b_ecc	I	7	SIU		Data ECC
DRAM					
l2t_mcu_rd_req	O				Read request to DRAM
l2t_mcu_rd_dummy_req	O				Flush request to MCU (=COMMIT)

TABLE 2-24 Unit Level Interface Signals *(Continued)*

Signal Name	I/O	Size	From/ To	Timing	Description
l2t_mcu_rd_req_id	O		MCU		Request id
l2t_mcu_addr	O		MCU		Read/write Address
l2t_mcu_wr_req	O		MCU		Write request to mcu
l2b_mcu_wr_data_r5	O		MCU		Write back data to memory
l2b_mcu_data_vld_r5	O		MCU		Writeback Data Valid signal
l2b_mcu_data_mecc_r5	O		MCU		Error signal for mcu
mcu_l2t_rd_ack	I	1	MCU		Read request recorded
mcu_l2t_wr_ack	I	1	MCU		Write request recorded
mcu_l2t_qword_id_r0	I	2	MCU		Quad-word number for a transaction
mcu_l2t_data_vld_r0	I	1	MCU		Valid signal with data
mcu_l2t_rd_req_id_r0	I	3	MCU		Read request ID returned
mcu_l2t_scb_mecc_err	I	1	MCU		Async Scrub Error Signals from mcu
mcu_l2t_scb_secc_err	I	1	MCU		Async Scrub Error Signals from mcu
mcu_l2b_data_r2	I	128	MCU		Fill data from mcu
mcu_l2t_mecc_err_r2	I	1	MCU		Error information
mcu_l2t_secc_err_r2	I	1	MCU		Error information
mcu_l2b_ecc_r2	I	28	MCU		SEC ECC Information

2.1.7 Reliability, Availability, and Serviceability (RAS)

2.1.7.1 General Overview

The Failure In Time (FIT) rates for L2 structures in OpenSPARC T2 in Epic8c are similar to their OpenSPARC T1 counterparts. To improve FIT rates L2, OpenSPARC T2 improves protection on L2 structures already protected on OpenSPARC T1 (e.g. VUAD Array).

L2 consists of the following two major structures that are candidates for protection.

The first type is 6-device, single-ported SRAM cells optimized for density, such as L2 data arrays. These SRAM cells have high Failure In Time (FIT) rates (300-400 FITs per Mb in Epic8c). All L2 SRAMs have ECC protection.

The second type is a CAM cell, whose FIT rate may be 1/2 of a standard SRAM cell. CAM cells are difficult to protect. Adding parity to a CAM cell eliminates false CAM hits due to single-bit errors, but cannot detect false misses.

In L2, only the I$ and D$ Directory CAMs are protected by parity. None of the other CAM structures in L2 are big enough to contribute to the overall FIT rate and hence are not protected by parity.

2.1.7.2 RAS Support in L2 Sub-Blocks

L2 Data Arrays

The L2 data arrays are protected via SEC/DED ECC on a word (32-bit) basis. A correctable error on a core data read or write results in an error being logged in one of the core ESRs and, if enabled, causes a precise or disrupting trap request to the core making the request. On a load, data gets corrected (if a single bit error was detected) when returned to the core but the error still gets reported to the core on ERR bits of the CPX packet for the load return. A core data read results from an instruction cache miss, a data cache miss, an atomic operation, a partial store or a store of less than 32-bits, or an SPU operation. A correctable error on an I/O data read or write results in an error being logged in a global ESR and, if enabled, causes a disrupting trap to the core identified by the ERRORSTEER field of the L2 Control Register. Hardware corrects the error, and rewrites the L2 line with corrected data.

An uncorrectable error on an L2 data read by a core is logged to a global ESR, and, if enabled, causes a disrupting trap to the core on a store. If the core request is due to an instruction fetch or load due to data cache miss or atomic operation, the error is actually a precise trap.

In the case of an I/O read or write with an uncorrectable error, the error is logged in a global ESR, and a disrupting trap is signaled to the core identified in the ASI_CMP_ERROR_STEERING register.

L2 Tag Arrays

The L2 tag arrays are protected by SEC ECC. A correctable error is logged to a global ESR, and, if enabled, signals a disrupting trap request to the core identified in the ERRORSTEER field of the L2 Control Register. Hardware (scrubber) corrects and re-writes the tag. The operation is completed by replaying from Miss Buffer.

OpenSPARC T1 L2 does not detect Uncorrectable double bit errors for tag and OpenSPARC T2 L2 wont either. This is because in OpenSPARC T2 with total FIT rates in the ballpark of 400, the double bit errors in the L2 tag are contributing only 0.1 FITs.

L2 VUAD Arrays

The VUAD array contains valid, used, allocated, and dirty bits. OpenSPARC T1 protects this array with parity. Any single bit error in the valid, allocated, or dirty bits can lead to data corruption and is fatal. OpenSPARC T2 protects the VD and UA arrays via SEC DED ECC since it is one of the largest sources of fatal errors.

A correctable error in VD or UA array (LVC) is logged in L2 ESR. If enabled, the error generates a disrupting trap request to the core identified in the ERRORSTEER field of the L2 control Register. Hardware corrects the entry, and the replayed access is completed.

An uncorrectable error in VD or UA array (LVU) is also logged in L2 ESR. However this would cause OpenSPARC T2 to assert warm reset.

L2 Directories

L2 protects it's I$ and D$ directories with parity. Parity error is fatal. The L2 directory performs a background parity detect which is synchronized with a store issue down the L2 pipe. The entry being checked for parity can be reset using dbginit_l so as to make a test repeatable.

Miss Buffer

The Miss Buffer contains miss requests as well as multi-pass L2 operations. The buffer contains data and address (tag) entries. The tag array is an 32 entry CAM of 40 bits each. A false hit on a tag can result in data corruption. A false miss can also result in data corruption. OpenSPARC T1 does not protect the tags or data. OpenSPARC T2 also does not protect the data or tags due to its small contribution to the FIT rate.

Fill Buffer

The Fill Buffer contains memory read data. This data is either cacheable reads to be written to the L2, or non-allocating cacheable data forwarded to the I/O interface. The buffer contains data and address (tag) entries. The data is protected by SEC/DED ECC and ECC is checked on the way from Fill Buffer to L2 pipe. The tag array is an eight entry CAM of 40 bits each. A false hit on a tag can result in data corruption. A false miss can result in multiple fills for the same line outstanding, reducing performance. OpenSPARC T1 does not protect the tags. OpenSPARC T2 also does not protect the tags due to its small contribution to the FIT rate.

A correctable data ECC error is logged to a global ESR and, if enabled, generates a disrupting trap request to the core identified in the ERRORSTEER field of the L2 Control Register. Hardware corrects the error before writing the data into the L2.

An uncorrectable data ECC error is logged to a global ESR, and, if enabled, generates a disrupting trap request to the core identified in the ERRORSTEER field of the L2 Control Register.

Writeback Buffer

The Writeback Buffer contains modified evicted L2 data to be written back to memory, The data portion is protected by SEC/DED ECC and ECC is checked on the way from WBB to mcu. The tag is implemented as an eight entry CAM with 40 bits per entry. OpenSPARC T1 does not protect the tag. OpenSPARC T2 does not protect the tag due to its small contribution to overall FIT rate.

If a correctable ECC error occurs on the data, the error is logged, and, if enabled, a disrupting trap request is generated to the core identified by the ASI _CMP _ERROR _STEERING register. Hardware corrects the error before writing the data to memory.

If an uncorrectable ECC error occurs on the data, the error is logged, and, if enabled, generates a disrupting trap request to the core identified by the ASI_ CMP_ ERROR_ STEERING register.

I/O Write Buffer

The I/O Write Buffer collects I/O write data prior to writing it to DRAM. The buffer consists of tag and data sections. OpenSPARC T1 and OpenSPARC T2 protect the data with SEC/DED ECC and ECC is checked on the way from IOWB to DRAM. The tag is not protected.

If a correctable ECC error occurs on the data, the error is logged, and, if enabled, a disrupting trap request is generated to the core identified by the ASI_ CMP_ ERROR_ STEERING register. Hardware corrects the error before writing the data to memory.

If an uncorrectable ECC error occurs on the data, the error is logged, and, if enabled, a disrupting trap request is generated to the core identified by the ASI_ CMP_ ERROR_STEERING register. Software could possibly retry the write operation through the device driver.

2.1.7.3 NotDATA in L2 (New Feature in OpenSPARC T2)

Uncorrectable errors are not necessarily unrecoverable. Program execution may or may not be affected depending on the not be affected depending on the condition under which the error occurs. In either case, HW must signal the occurrence of the error when it detects it to the appropriate SW error handler. However the chances for recovery are greatly enhanced if only the offended processor reports the error, and the others do not.

To meet this high level goal, as part of a requirement from SPARC SWG RAS working group to have UE from DRAM stored in L2/L3 caches as Notdata, OpenSPARC T2 L2 would support a scheme for detecting Notdata on UE from DRAM without using extra bits of storage.

The scheme involves flipping the computed ECC bits for the data with UE from DRAM and storing it with the data. The idea is when a subsequent access happens to the line in L2 that would do an ECC check, the generated ECC would be 1's complement of the stored ECC, resulting in check bits[6:0] being all 1's. This would indicate NotData.

System requires that NotData syndrome be protected from single bit errors (SBEs). Single bit error correction is not required. However it is required that the error be reported (even as UE), and that the error is never mistaken as valid data, or data with correctable errors.

The ECC logic in L2 after detecting single bit error in the data or ECC portion of the NotData packet will treat it as an uncorrectable error. A double bit error can be a problem since the ECC logic would potentially treat it as valid data with CE. But the chance of such failure is very remote and double bit error protection on NotData is not a system requirement.

The following sections describe the mechanisms of detecting UE on FIll and storing NotData in L2 and also L2 behavior on subsequent accesses to L2 from core, SIU, scrubber and DRAM (eviction) finding NotData.

Detecting UE on a Fill & Storing NotData in L2:

If MCU detects UE on data return from DRAM, it will indicate UE to L2 on a 16 byte quad-word boundary and also invert all ECC bits associated with the 16 bytes of data. This data with inverted ECC bits will then get written to the Fill Buffer and eventually get written to the L2 data array on the fill. Thus on the UE, MCU will write data marked as Notdata itself into the L2 cache.

However once the data is returned to L2, there are three possibilities:

1. Replayed load/ifetch/atomic reads from the Fill Buffer array before the fill happens and detects UE.

In this case if an UE gets detected on the data read from the Fill Buffer, it would return load/ifetch/atomic data to Core but mark the data as UE in the CPX packet. Also the UE would be logged as DAU error in L2. The Core will take a precise trap. Then the fill would happen and store NotData in the array. Subsequent accesses would see Notdata in the L2.

2. Replayed load/ifetch/atomic reads from the Fill Buffer array before the fill happens and does not detect UE but UE is another 16 byte chunk in the same line.

In that case the load/ifetch/atomic will complete as normal. Later on when the FIll happens, if the UE is on another 16 Byte chunk for the same line (as indicated by a UE bit stored in FIll Buffer from DRAM), the UE would be logged as DAU error in L2 and assert a disrupting trap to CPU. The fill would store NotData in the array so the subsequent accesses see NotData in L2.

3. Replayed load/ifetch/atomic reads from the data array itself after the fill.

In this case if there was UE detected during the fill, the UE would be logged as DAU error in L2 and would cause a disrupting trap to the core. The fill would store NotData in the array. Later on if the replayed load/ifetch/atomic reads the 16 Byte chunk that has NotData in it, it will return data to Core but mark the data as NotData in the CPX packet. This would cause a precise trap. All subsequent accesses that hit would also see NotData.

If the FIll was for a store miss from CPU or SIU, UE would be logged as DAU error in L2 and cause a disrupting trap. The fill would store NotData in the array so the subsequent accesses see NotData in L2.

Detecting UE on a Scrub:

When the L2 data array scrubber detects UE in a line in the data array, it will log the UE as LDSU bit in L2 and will issue a CPX Error packet to the core specified in the ERRORSTEER field of the L2 Control Register.

L2 Behavior on Subsequent Accesses to L2 from Core, SIU, Scrubber and DRAM (eviction) finding NotData.

1. Load/Ifetch/atomic hit from CPU in L2 encountering NotData:

When a load/ifetch/atomic hit detects NotData, it will set NDSP bit in L2 and report NotData to the requesting core on the CPX packet.

2. Partial store hit from CPU encountering Notdata:

When a partial store hit in L2 encounters NotData it will set NDSP bit in L2 and issue a Error Indication packet to the core indicating NotData. A disrupting trap would get issued. The store will not happen in L2.

3. 4/8 byte store from CPU:

The store would complete irrespective of the NotData in the data array. ALthough NotData gets marked for all four byte chunks of a 16 Byte line segment, all four NotData segments may get overwritten with 2/4 back to back stores and would cause the NotData symptom to be lost. In that case, when the disrupting trap caused by the UE on the related Fill gets the issued, the trap handler will just cause an eviction of the dirty line to DRAM (without knowing whether it has Notdata or not). L2 will detect Notdata on the data in the eviction path and if there is no trace of any NotData, MCU will just write it out to DRAM without polluting the ECC on 16 byte boundary. Otherwise L2 will signal UE to MCU and MCU will write it out to DRAM polluting the ECC on 16 byte boundary.

4. SIU load hit in L2 encountering NotData:

SIU load hitting in L2 encountering NotData will return data marked with UE to SIU which will get propagated to PCI_EX /BSC. NDDM bit will be set in L2 and a CPX Error packet will be issued to core specified in the ERRORSTEER field of the L2 Control Register indicating NotData and a disrupting trap will be taken.

5. SIU partial store hit in L2 encountering NotData:

When a partial store from SIU in L2 encounters NotData in L2 it will set NDDM and issue a Error Indication packet to the core specified in the ERRORSTEER field of the L2 Control Register indicating NotData. The store will not happen in L2. A disrupting trap will be taken.

6. 4/8 byte store from SIU:

The store would complete irrespective of the NotData in the data array. ALthough NotData gets marked for all four byte chunks of a 16 Byte line segment, all four NotData segments may get overwritten with 2/4 back to back stores and would cause the NotData symptom to be lost. In that case, when the disrupting trap caused by the UE on the related Fill gets the issued, the trap handler will just cause an eviction of the dirty line to DRAM (without knowing whether it has UE or not). UE will detect UE on the data in the eviction path and if there is no trace of any NotData, it will just write it out to DRAM without polluting the parity. Otherwise it will signal UE to MCU and MCU will write it out to DRAM polluting the data to indicate multi-bit error.

7. Data Ram Scrubber encountering NotData in L2:

When Data Ram Scrubber in L2 detects NotData it will do nothing, and will keep the data and ECC bits unchanged.

8. Eviction from L2 encountering NotData:

When NotData is detected on evicted data from Write Back Data Buffer to DRAM, L2 would indicate to MCU UE for each 16 byte chunk same as today, and the MCU would write the data to DRAM after flipping multiple ECC bits according to Galois Field Hemming code.

L2 Behavior on NotData Reported from Core

1. When uncorrectable errors occur on the store data in the store buffer, core would indicate it to L2 by asserting INV bit (116) of the pcx packet to 1'b1. L2 would then accept the store and do the write, but would mark the data as NotData.
2. To deal with UE on the compare data of a CASA instruction, LSU will assert the INV bit (116) of the pcx packet for both the CAS1 and CAS2 packets. When the L2 sees this bit asserted, it will force the compare result to be "true" so that L2 will be updated. Also instead of storing the swap data, it will write NotData.

2.1.7.4 Error Reporting by L2

L2 cache reports the different errors it detects to the cores through encoded ERR[1:0] field on the CPX packets.

Loads, stream loads, mmu loads, prefetches, ifetches, atomics send back UE/CE/Notdata error on the data read, on ERR[1:0] (139:138) bits of the Load return, Stream Load Return, MMU Load return, Prefetch Return, Ifill Return 1, Ifill Return 2, Swap/Ldstub return and CAS return packets respectively to the requesting virtual core.

LDAU/LDAC/NDSP errors respectively get recorded in L2 Error Status registers (Refer to *OpenSPARC T2 Programmer's Reference Manual*) for UE/CE/Notdata Error with R/W bit being a 0.

In case a UE/Notdata is detected on a CAS1 or swap/ld stub read pass, the store does not happen in L2 in the CAS2 and swap/ldstub writes passes (leaving the data and ECC unchanged). However in the CAS 2 ack packet and the swap/ldstub ack packet, ERR[1:0] (bits 139:138) gets driven as valid to the requesting virtual core with the encoding reflecting CE/UE/Notdata.

On a atomic miss, when the line is returned from DRAM, the fill always happens first and then the load of the atomic is replayed from the Miss Buffer. Now if the line has CE or UE in it, on the fill, an L2 Error indication packet will get sent to the core indicating CE or UE. Note that CE here means that the line is already corrected and error free on account of MCU cleaning it up when writing to L2. If it was a CE, the error would not be visible in the L2 after the fill as the line is already corrected, and the replayed load of the atomic will not see any error and will just return good data.

The store that follows will just complete as normal without any errors. If it was a UE, the error would be still persisting in the line when the fill happens and so the replayed load will see Notdata and will indicate a Notdata on the load return packet to the core. The store that follows will also see the error and indicate Notdata on the store ack packet. The store will not happen in L2.

Stores, Stream Stores on detecting UE/CE/Notdata on the data do not report the error on the ERR[1:0] (139:138) bits of the Store ack, Stream Store ack packets respectively. However the errors get reported on bits 139:138 of Error Indication packet later to the requesting virtual core after the occurrence of the next L2 fill. On the detection of the error, LDAU/LDAC/NDSP errors respectively get recorded in L2 Error Status registers (Refer to *OpenSPARC T2 Programmer's Reference Manual*) for UE/CE/Notdata Error with R/W bit being a 1. Stores do not happen to the L2 in case of UE and Notdata errors leaving the data and ECC unchanged.

Directory parity (LRU) error and VUAD Uncorrectable Error (LVU) cause a fatal error warm reset and does not get reported to the core on the Error Indication Packet.

All other errors (LDWC, LDWU, LDRC, LDRU, LDSC, LDSU, DAC, DAU, DRC, DRU, DSC, DSU, LTC, LVC, and NDDM) would get reported to the Virtual Core specified in the L2_CONTROL_REG.ERRORSTEER field on bits [139:138] of Error Indication packet of crossbar after the occurrence of the next L2 fill.

L2 would drive bit 137 of the CPX packet on an L2 Error indication CPX packet as 1 in case of LDRC error on a SIU RDD. This would indicate to the core that core should take a SW_recoverable trap instead of a HW_corrected error trap. For all other Correctable errors asserted by L2 on the Error Indication packet, this bit will be 0 indicating HW_corrected error. Also for other error types like UE and Notdata, this bit would be driven as 0.

2.1.8 VDFT Features

The following DFT features are supported by L2 cache.

- BIST
- Full scan
- Shadow can

2.1.9 Critical Path Analysis

Following are the critical paths in OpenSPARC T1 L2 cache:

- VUAD access in C1: Memory Access(10g) + 1mm xmit (2g) + 4-1 mux(3.5g) + 4-1 mux(3.5g) + 2-1 mux(2g) = 21 gates.

- Way Sel Generation in C2: c2 compare logic(2g) + 2.32 mm xmit in the tag array (4.7g) + 600U xmit to tagctl(1.2g) + waysel logic(5g) +3mm xmit to middle of the bottom of l2d (6g) = 19 gates.

Transmit from l2t to l2d is based on the existing proposed full chip floor plan.

- Arbitration in PX to tag access: Arb logic to sel PX addr (12g) + 2.7mm xmit to VUAD (6g) + array setup (1g) = 19 gates
- Data cache access in C4, C5 and C52.
- Data return in C8. Xmit from deccdp to oqdp (2g) + 3-1 mux in oqdp (3g) + xmit from oqdp to ccx (5g) + setup in the CCX (10g) = 20 gates.

2.1.10 Performance

The following are the L2 cache performance data for 1.4 Ghz cmp clk and 333 Mhz DRAM clk:

- Load-Use latency seen by core on L1 miss and L2 hit: 21 cmp clks = (0.714x20) = 14.994 nsec

W' |PQ PA PX1 PX2 | C1 C2 C3 C4 C5 C52 C6 C7 C8 |

CQ CA CX1 CX2 | CX3 E M B W

[where the CPU related stages are as follows:

W' - arb for pcx

PQ - send pcx request

PA - send pcx packet

CX3 - packet received from cpx

E - IFU signaled to restart thread

M - data written to L1 cache (if load was cacheable)

B - data sent to EXU/FGU

W - data written to register file or bypassed to dependent instruction in E stage]

- Load-Use latency seen by core on L1 miss and L2 miss: 93 cmp clks= (0.714 x 93) = 66.4 nsec

W' -> C1: 6 cmp clks

C2 -> Req to MCU: 10 cmp clks

MCU -> DRAM address on memory bus: four cmp clks for ack sync + 2 DRAM clks = 12.4 cmp clks

DRAM address -> first DRAM data (critical 16/32 B first) on pins: eight DRAM clocks

= 33.6 cmp clks

MCU -> L2 first critical data: 3 DRAM clks = 12.6 cmp clks

L2 first critical data -> C2 stage of replayed Load: four cmp clks

C3 stage of replayed load -> return critical 16B to core: 9 cmp clks

Data seen by Core -> earliest it gets used in Core pipe = 5 cmp clks

- Peak L2 load bandwidth with back to back hits in different sub-banks: 1.4 Ghz x16 bytes = 22.4 GB/s

(16 bytes of data getting returned from L2 to core every cmp clock over Crossbar)

- L2 store bandwidth with eight byte back to back stores that hit in L2 in different sub-banks: 1.4 Ghz x 8 bytes = 11.2 GB/s

2.2 Appendix

2.2.1 Debug Mode/Initialization Mode

L2 cache comes out disabled after reset. One of the bits in the L2 cache control registers have to be set in order to enable L2 cache. When L2 cache is disabled, there are no accesses to L2 cache tag ram, data ram, VUAD arrays. All the instructions are treated as a miss. However, diagnostic accesses to tag ram, data ram and VUAD arrays are permitted.

The behavior of L2 cache when disabled is as follows:

In this mode the Miss Buffer operates to it's full capacity. However, the Fill Buffer is just one (line) deep. All the loads and stores issued to L2 cache gets recorded in Miss Buffer (MB). Loads recorded in Miss Buffer gets issued to DRAM. The (load) data returned from DRAM is recorded in the Fill Buffer (FB). The instruction gets replayed from the Miss Buffer and data gets returned from the Fill Buffer.

When a Store transaction is encountered in Miss Buffer (MB), the store data gets transferred to the Fill Buffer and a read gets issued to DRAM to fetch the line. When the read data is returned, the store transaction now gets replayed from the Miss Buffer and does a data merge with the Fill data before getting written into the Writeback Buffer (WBB). Stores are issued out of Writeback Buffer to DRAM.

2.2.2 Reset Sequence for L2 cache

In L2 cache, parity bits in the tag array, valid bits in VUAD array and the directories should be initialized before L2 cache is enabled to guarantee coherency and correct functionality.

The directory valid bits are cleared with flash reset during POR_. The reset block drives the flash reset. When the valid bits are cleared (not valid) then the entries are don't care. Hence, the parity bits are not initialized to good parity. Clearing valid bits in the directory informs the L2 cache that there are no valid lines in L1.

BISI or ASIs are used to initialize the VUAD arrays by clearing all the valid bits. This informs L2 cache that there are no valid lines in L2.

BISI or ASIs are used to initialize the tag array with good parity. This eliminates the possibility of any error cases from happening.

CHAPTER 3

Memory Control Unit (MCU)

This chapter contains the following sections:

3.1 Overview

The DRAM memory control unit (MCU) interfaces to external registered DDR2 FBDs through a unidirectional high-speed link to service load and store requests from two L2 cache banks of the on-chip L2 Cache unit. Each load and store request from a L2 cache bank has a data size of one cache line, 64 bytes. There are four physical instantiations of MCU in the OpenSPARC T2 CPU.

The features of the MCU are as follows:

- Maximum memory of 128 GB per MCU branch using 8 GB DDR2 FBDs (assuming 2 Gb DRAM parts).
- Supports DDR2 SDRAM clock frequencies up to 400 MHz (800 MHz double data rate). Internally, the MCU runs at the DDR rate.
- Supports up to 16 ranks of DDR2 FBDs per channel (eight pairs of double sided FBDs).
- 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.
- System peak memory bandwidths (4 branches) with 800 MHz DDR parts: 50 GB/s for reads, 25 GB/s for writes.
- Uses 10-bit Southbound and 14-bit Northbound FBD channel protocols running at 12 times the SDRAM cycle rate.
- Supports DDR2 SDRAM burst length of four with when using both FBD channels in an MCU, burst length of eight when using one FBD channel.
- ECC generation, check, correction.
- Programmable DDR2 SDRAM power throttle control.
- The FBD Hot Plug feature is not supported.

3.1.1 Changes from OpenSPARC T1 MCU design

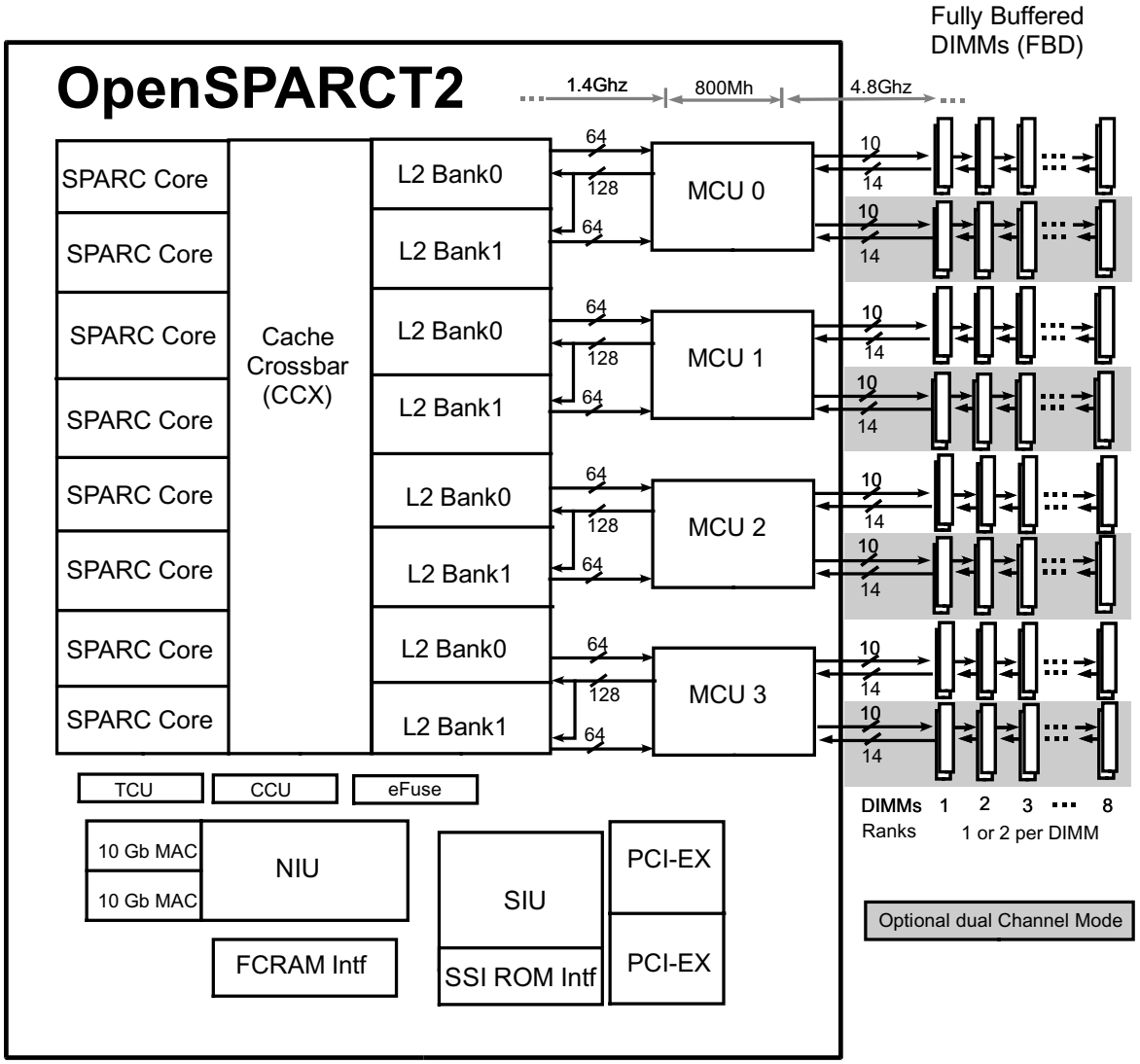
- Use higher DDR2 SDRAM frequency: 266MHz, 333 MHz and 400MHz instead of 166MHz to 200MHz.
- Uses FBDIMM channels to access memories instead of direct DDR2 interface.
- Interface to two L2 cache banks per MCU instead of one or two L2 Cache banks interface per MCU.
- Minimum configuration with one DIMM per MCU branch.

3.1.2 Changes to OpenSPARC T2 MCU to support FBD

- Added new FBD controller with channel initialization, error detection, and frame encode and decode logic.
- Updated address decoding to support up to 16 ranks of DIMMs. Can support either one or two channels per MCU.
- Write data rate reduced to half DDR rate. Data is buffered in AMB to allow more flexibility in issuing write commands.

- Read and write operations to different DIMMs can occur in parallel. Reads and writes to a single FBD must be scheduled so that there are no data collisions on the DIMM's local DDR2 bus. However, since the Northbound and Southbound channels are independent, read data from one DIMM can be returning to the host at the same time that write data is being sent to different DIMM.
- Have separate read and write schedulers that communicate with each other to ensure there are no FBD bus data collisions.
- No dead cycle when switching read or write commands between DIMMs; however, this is still needed when switching access to the other sides of same DIMM.
- Include sync frame generation to AMBs in state machine, at least once every 42 frames.
- Remove read DQS strobe placement support. OCD and ODT support will be programmed through the AMBs.
- Spread transactions over different DIMMs instead of staying in one DIMM as long as possible to keep thermal dissipation better spread across DIMMs.
- Support L0s power saving mode.

FIGURE 3-1 OpenSPARC T2 System Overview



3.2 Terminology and Configuration

3.2.1 DRAM Terminology

- DIMM: Dual Inline Memory Module. Industry standard SDRAM module package. A stick of memory.
- Channel: Port connecting Processor chip to DIMM.
- DRAM chip: single chip inside the DIMM. We differentiate the types by how many bits it outputs and its capacity. (x4 means four bit output, x8 means 8 bit output, x16, x32 etc. and 256Mbit or 512Mbit capacity). Most common ones are the x4, x8 output.
- Bank: Most DDR SDRAM chips are broken up into four or eight logical banks internally to enable full pipelining of memory operations.
- Rank: A group of data 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 15.6uS/7.8uS.
- Single-channel Mode: This is a low-power configuration with one DIMM per memory channel. Only 72 bits of the 144 external IO pins are used, and the memory burst length is 8. While it is possible to support two DIMMs per channel with this configuration, it is only expected to be used with one DIMM.

3.2.2 FBD Terminology

- Advanced Memory Buffer (AMB) - The 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 which operate in lock-step to support error correction. A rank spans a branch. In OpenSPARC T2, a branch will consist of one or two DDR channels.
- DDR Channel - A DDR channel consists of a data channel with 72 bits of data and an addr/cntrl channel.
- DDR Data Channel - a DDR data channel consists of 72 bits of data divided into 18 data groups.
- Frame - Groups of bits containing commands or data sent across the link over 12 cycles.
- FBD - Fully Buffered DIMM.
- Link - High-speed parallel differential point-to-point interface.
- 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.
- 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. Approx. 200 ps for DIMMs running at 800 MHz.

3.2.3 DDR Branch Configuration

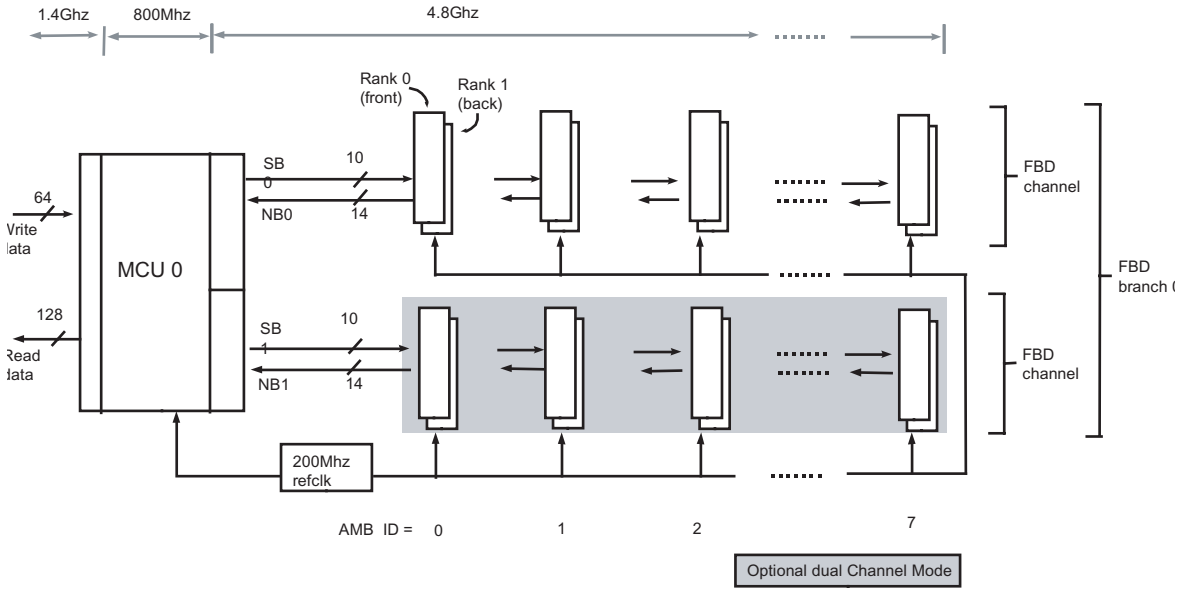
The following are key assumptions made during the design of this controller:

- x4 and x8 DIMMs are supported.
- DIMM capacity, configuration, and timing parameters cannot be different within a memory branch.
- Each DDR branch can have a different memory size and a different kind of DIMM (e.g. a different number of ranks or different CAS latency). Software should not use address space bigger than four times the lowest memory capacity in a branch because the cache lines are interleaved across channels, and using different sized memories can create holes in the address space.
- DRAM banks are always closed after read or write command by issuing an autoprecharge command.

- Burst length is four (BL=4) when using a two channels per DDR branch. Burst length is eight (BL=8) when using a single channel per branch.
- There is a fixed one 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 3-2. 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 eight frames in single channel mode and 288 bits per frame for four 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 3-2 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 four with 16 bytes of

data transferred every half memory clock cycle. In single-channel mode, Reads and Writes will have a burst length of eight with eight bytes of data transferred every half memory cycle.

TABLE 3-1 Supported Memory Organization

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	512 Mb	x4	1	18	1 GB	16 GB
1 GB	256 Mb	x4	2	36	1 GB	16 GB
2 GB	1 Gb	x4	1	18	2 GB	32 GB
2 GB	512 Mb	x4	2	36	2 GB	32 GB
4 GB	2 Gb	x4	1	18	4 GB	64 GB
4 GB	1 Gb	x4	2	36	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

3.2.3.1 Physical Address Mapping

The 40-bit physical memory address PA[39:0] request from the eight L2 banks are decoded and mapped to one of the four MCUs by address bits PA[8:7].

39 - 9	8 - 6	5 - 4	3 - 1
DIMMs Memory Address	L2 Bank Select	L2 Cacheline Sub-Address	16-byte Offset

The L2 memory write requests are 64-byte aligned: PA[5:0] = 6'h00. A partial cache line memory write is not supported by the MCU.

The L2 physical memory read address requests are 16-byte aligned: PA[3:0]=4'h0. The read data returned will be in the following order based on the L2 cache line sub address, PA[5:4] (PA[6:4] for single-channel mode).

TABLE 3-2 Read Data Return Order for BL=8

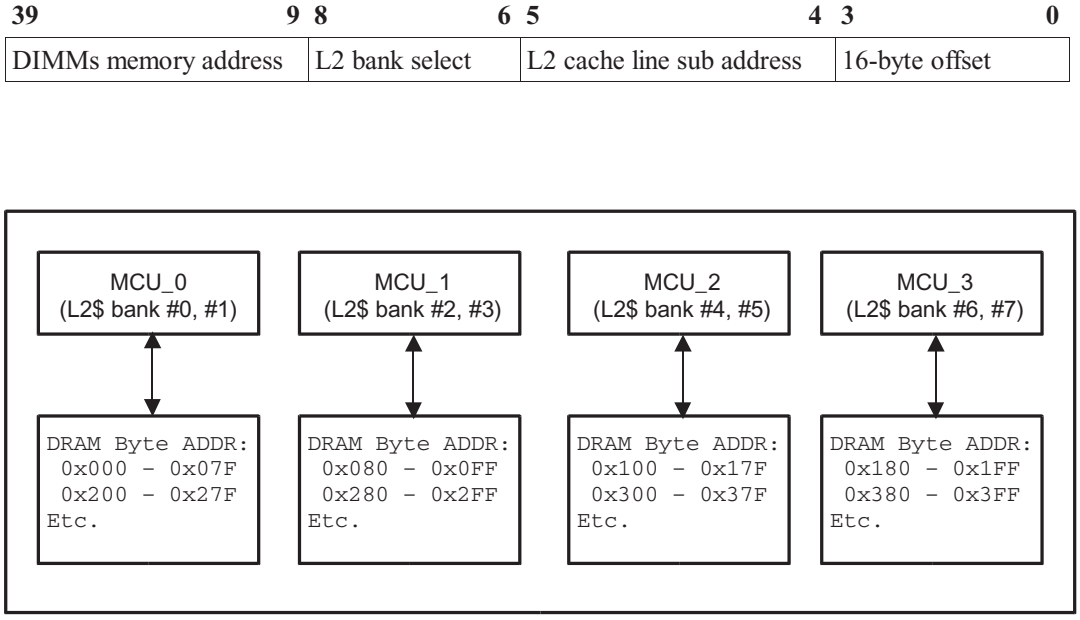
L2 cacheline sub- address, PA[5]	8-byte data return order
0	0,1,2,3,4,5,6,7
1	4,5,6,7,0,1,2,3

TABLE 3-3 Read Data Return Order for BL=4

L2 cacheline sub- address, PA[5]	16-byte data return order
0	0,1,2,3
1	2,3,0,1

The L2 cache bank select (PA[8:6]) is mapped to the four memory branches as shown in FIGURE 3-3.

FIGURE 3-3 L2 Cache Banks Memory Branch Mapping



3.2.4 FBD Channel Configuration

The FBD specification supports two southbound channel configurations and five northbound channel configurations. OpenSPARC 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 bits data and eight bits 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.

3.3 DDR2 FBD Usage

The following sections detail DDR2 FBD information specific to the OpenSPARC T2 MCU.

Note – The OpenSPARC T2 Memory Control Unit (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 OpenSPARC T2 MCU design to review all information assembled by JEDEC and develop their own independent conclusion.

3.3.1 FBD Channel Initialization

The FBD channels must be initialized through a software interface. This allows more flexibility in the initialization over a dedicated hardware state machine.

Software must perform the following sequence of events in order to initialize an FBD channel:

1. 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 Disable (51 frames) before transitioning to Calibrate state.
2. 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 Calibrate time (480K frames).
3. Drive Electrical Idle on SB channel to transition AMBs to Disable state. Remain in Disable state for at least Disable time (51 frames).
4. 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 approx. 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 five.

5. 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 six.

6. 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 to determine how to adjust the Command_to_Data_Incr registers in the AMBs.

7. 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 OpenSPARC T2) and which channel bits are muxed out if using a fail over mode. TS3 patterns are issued until the patterns are correctly received on the NB channel.

8. 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.

3.3.2 FBD Commands

TABLE 3-4 FBD DRAM Commands

DRAM Cmds	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Activate	DS2	DS1	DS0	1	Addr		RS						DRAM Bank and Address											
Write	DS2	DS1	DS0	0	1	1	RS						DRAM Bank and Address											
Read	DS2	DS1	DS0	0	1	0	RS						DRAM Bank and Address											
Precharge All	DS2	DS1	DS0	0	0	1	RS	X	X	X	X	1	1	1	X	X	X	X	X	X	X	X	X	X
Precharge Single	DS2	DS1	DS0	0			RS	DRAM Bank				1	1	0	X	X	X	X	X	X	X	X	X	X
Auto Refresh	DS2	DS1	DS0	0			RS	X	X	X	X	1	0	1	X	X	X	X	X	X	X	X	X	X
Enter Self Refresh	DS2	DS1	DS0	0			RS	X	X	X	X	1	0	0	X	X	X	X	X	X	X	X	X	X
Exit Self Refresh/Exit Power Down	DS2	DS1	DS0	0			RS	X	X	X	X	0	1	1	X	X	X	X	X	X	X	X	X	X
Enter Power Down	DS2	DS1	DS0	0			RS	X	X	X	X	0	1	0	X	X	X	X	X	X	X	X	X	X
reserved	X	X	X	X	X	X	X	X	X	X	X	0	0	X	X	X	X	X	X	X	X	X	X	X

TABLE 3-5 FBD Channel Commands

Channel Cmds	23	22	21	[20:14]	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Debug: In-band Events	EV7	EV6	EV5	7'b0001111	1	EV4	EV3	EV2	EV1	EV0	PV7	PV6	PV5	PV4	PV3	PV2	PV1	PV0
Debug: Relative Timing	PH5	PH4	PH3	7'b0001111	0	PH2	PH1	PH0	RT9	RT8	RT7	RT6	RT5	RT4	RT3	RT2	RT1	RT0
Debug: Exposed Info	EX16	EX15	EX14	7'b0001110	EX13	EX12	EX11	EX10	EX9	EX8	EX7	EX6	EX5	EX4	EX3	EX2	EX1	EX0
reserved	X	X	X	7'b000110x	X	X	X	X	X	X	X	X	X	X	X	X	X	X
reserved				7'b000110xx	X	X	X	X	X	X	X	X	X	X	X	X	X	X

TABLE 3-5 FBD Channel Commands *(Continued)*

Channel Cmds	23	22	21	[20:14]	13	12	11	10	9	8	7	6	5	4	3	2	1	0
DRAM CKE per DIMM	DS2	DS1	DS0	7'b0000111	BC	X	X	X	X	X	DE7	DE6	DE5	DE4	DE3	DE2	DE1	DE0
DRAM CKE per RANK	DS2	DS1	DS0	7'b0000110	BC	X	X	X	X	X	D3R1	D3R0	D2R1	D2R0	D1R1	D1R0	D0R1	D0R0
Write Config Reg	DS2	DS1	DS0	7'b0000101	DS3	TID	X	A10	A9	A8	A7	A6	A5	A4	A3	A2	0	0
Read Config Reg	DS2	DS1	DS0	7'b0000100	DS3	X	X	A10	A9	A8	A7	A6	A5	A4	A3	A2	0	0
reserved	X	X	X	7'b0000011	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Soft Channel Reset	X	X	X	7'b0000010	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sync	X	X	X	7'b0000001	X	SD1	SD0	X	X	X	X	IER	ERC	EL0s	X	X	R1	R0
Channel NOP	X	X	X	7'b0000000	X	X	X	X	X	X	X	X	X	X	X	X	X	X

3.3.2.1 FBD Frame Formats

Data is transmitted across the southbound and northbound channels in frames. For the southbound channel, 10 bits of data are sent per cycle over 12 cycles. For the northbound channel 14 bits of data are sent per cycle over 12 cycles. The next two sections show the format of the frames.

Southbound Frame Formats

The southbound frame format consists of two sections, the "A" command section and the "B"/"C" command or Data section. The 24-bit "A" command section is contained in the first four cycles of the frame. The aC[23:0] bits in TABLE 3-6 shows the location of the "A" command. Bits F[1:0] determines the format of the last eight cycles of the frame as shown in TABLE 3-7. Bits aE[13:0] are the CRC value protecting the aC[23:0] and F[1:0] fields.

The FE[21:0] bits in TABLE 3-6 are CRC bits protecting the 72 bits of command or data in the remaining eight cycles. Bits FE[13:0] are exclusive-ORed with the aE[13:0] field of the following frame.

In failover mode, the data for bit 9 is neither transmitted nor used in CRC calculations.

TABLE 3-6 Common Features of Normal Southbound Frames

	Bit									
Transfer	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
0	aE0	aE7	aE8	F0	aC20	aC16	aC12	aC8	aC4	aC0
1	aE1	aE6	aE9	F1	aC21	aC17	aC13	aC9	aC5	aC1
2	aE2	aE5	aE10	aE13	aC22	aC18	aC14	aC10	aC6	aC2
3	aE3	aE4	aE11	aE12	aC23	aC19	aC15	aC11	aC7	aC3
4	FE21									
5	FE20									
6	FE19									
7	FE18									
8	FE17									
9	FE16									
10	FE15									
11	FE14									
	FE0	FE7	FE8							
	FE1	FE6	FE9							
	FE2	FE5	FE10	FE13						
	FE3	FE4	FE11	FE12						

TABLE 3-7 Southbound Frame Type Encoding

Frame Format	F1	F0	Comments
Command	0	0	Frame contains one or more commands plus optional data
reserved	0	1	
Command + Wdata	1	WSn	Frame contains an "A" command plus 72 bits of Wdata

Command Frame Format

TABLE 3-8 shows the format of a southbound frame with three commands, aC[23:0], bC[23:0], and cC[23:0].

TABLE 3-8 Command Frame Format

Transfer	Bit									
	9	8	7	6	5	4	3	2	1	0
0	aE0	aE7	aE8	F0	aC20	aC16	aC12	aC8	aC4	aC0
1	aE1	aE6	aE9	F1	aC21	aC17	aC13	aC9	aC5	aC1
2	aE2	aE5	aE10	aE13	aC22	aC18	aC14	aC10	aC6	aC2
3	aE3	aE4	aE11	aE12	aC23	aC19	aC15	aC11	aC7	aC3
4	FE21	0	0	0	bC20	bC16	bC12	bC8	bC4	bC0
5	FE20	0	0	0	bC21	bC17	bC13	bC9	bC5	bC1
6	FE19	0	0	0	bC22	bC18	bC14	bC10	bC6	bC2
7	FE18	0	0	0	bC23	bC19	bC15	bC11	bC7	bC3
8	FE17	0	0	0	cC20	cC16	cC12	cC8	cC4	cC0
9	FE16	0	0	0	cC21	cC17	cC13	cC9	cC5	cC1
10	FE15	0	0	0	cC22	cC18	cC14	cC10	cC6	cC2
11	FE14	0	0	0	cC23	cC19	cC15	cC11	cC7	cC3

Command Frame with Data Format

TABLE 3-9 shows the southbound frame format where the "B" command has a 32-bit data payload. The BE[3:0] bits are byte enables for the data. This format is used to write to internal control registers of the AMB.

TABLE 3-9 Command Frame with Data Format

Transfer	Bit									
	9	8	7	6	5	4	3	2	1	0
0	aE0	aE7	aE8	F0	aC20	aC16	aC12	aC8	aC4	aC0
1	aE1	aE6	aE9	F1	aC21	aC17	aC13	aC9	aC5	aC1
2	aE2	aE5	aE10	aE13	aC22	aC18	aC14	aC10	aC6	aC2
3	aE3	aE4	aE11	aE12	aC23	aC19	aC15	aC11	aC7	aC3
4	FE21	0	0	0	bC20	bC16	bC12	bC8	bC4	bC0

TABLE 3-9 Command Frame with Data Format *(Continued)*

Transfer	Bit 9	8	7	6	5	4	3	2	1	0
5	FE20	0	0	0	bC21	bC17	bC13	bC9	bC5	bC1
6	FE19	0	0	0	bC22	bC18	bC14	bC10	bC6	bC2
7	FE18	0	0	0	bC23	bC19	bC15	bC11	bC7	bC3
8	FE17	BE0	D28	D24	D20	D16	D12	D8	D4	D0
9	FE16	BE1	D29	D25	D21	D17	D13	D9	D5	D1
10	FE15	BE2	D30	D26	D22	D18	D14	D10	D6	D2
11	FE14	BE3	D31	D27	D23	D19	D15	D11	D7	D3

Command +WData Frame Format (4-bit Device)

Write data is sent in the southbound frame when F[1] is 1. Write data for a write command is sent to an AMB over four cycles. TABLE 3-10 shows the format of the F[1:0] field for these four frames. The WS[2:0] field identifies the target AMB. Each AMB must speculatively store the write data in an accumulation buffer until it determines that the data is for it. If the data is for that AMB, the data is stored in the Write Data Buffer; otherwise, it is discarded.

TABLE 3-10 WData Address Delivery

Wdata Frame	F1	F0
0	1	WS0
1	1	WS1
2	1	WS2
3	1	0

TABLE 3-11 shows the format of the Command with Wdata frame.

TABLE 3-11 Command+Wdata Frame Format (4-bit Device)

Transfer	Bit 9	8	7	6	5	4	3	2	1	0
0	aE0	aE7	aE8	F0	aC20	aC16	aC12	aC8	aC4	aC0
1	aE1	aE6	aE9	F1	aC21	aC17	aC13	aC9	aC5	aC1
2	aE2	aE5	aE10	aE13	aC22	aC18	aC14	aC10	aC6	aC2

TABLE 3-11 Command+Wdata Frame Format (4-bit Device) *(Continued)*

Transfer	Bit									
	9	8	7	6	5	4	3	2	1	0
3	aE3	aE4	aE11	aE12	aC23	aC19	aC15	aC11	aC7	aC3
4	FE21	C17D0	C15D0	C13D0	C11D0	C9D0	C7D0	C5D0	C3D0	C1D0
5	FE20	C17D1	C15D1	C13D1	C11D1	C9D1	C7D1	C5D1	C3D1	C1D1
6	FE19	C17D2	C15D2	C13D2	C11D2	C9D2	C7D2	C5D2	C3D2	C1D2
7	FE18	C17D3	C15D3	C13D3	C11D3	C9D3	C7D3	C5D3	C3D3	C1D3
8	FE17	C18D0	C16D0	C14D0	C12D0	C10D0	C8D0	C6D0	C4D0	C2D0
9	FE16	C18D1	C16D1	C14D1	C12D1	C10D1	C8D1	C6D1	C4D1	C2D1
10	FE15	C18D2	C16D2	C14D2	C12D2	C10D2	C8D2	C6D2	C4D2	C2D2
11	FE14	C18D3	C16D3	C14D3	C12D3	C10D3	C8D3	C6D3	C4D3	C2D3

Northbound Frame Formats

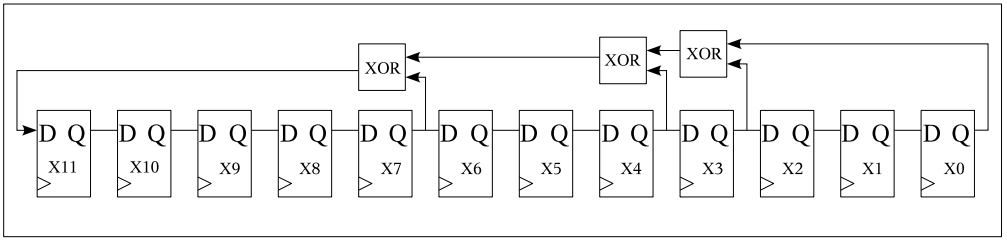
Northbound frames return status information or read data. Alert Frames implicitly tell if an error has occurred on the southbound channel. A Status Frame is an explicit status response to a Sync Frame on the southbound channel. When no Status or Read Data Frame is expected and no Alert Frame is detected on the northbound channel, then an Idle frame is expected.

Bit 13 of the northbound frame is not transmitted or used in CRC calculations when in failover mode.

Northbound Idle Frame Format

The Idle Frames are sent by the AMBs on the northbound channel to indicate that it is still operating correctly. The bits within the frame are determined by a 12-bit LFSR with polynomial x^12 + x^7 + x^4 + x^3 + 1. An example implementation is shown in FIGURE 3-4.

FIGURE 3-4 Idle Frame LFSR Counter



The initial value is 12'b000000000001, and the LFSR cycles through 2^12 - 1 states before repeating. The pattern in the LFSR is mapped to the Idle Frame bit lanes, i.e. X0 maps to bit lane 0, X1 to bit lane 1, etc. The 13th bit lane contains the value of X0 for the first 6 cycles of the frame and the inverse of X0 for the last 6 cycles. The 14th bit lane contains the value of X0 for all 12 cycles of the frame. TABLE 3-13 shows the format of the first Idle Frame.

TABLE 3-12 First Northbound Idle Frame Format

Transfer	Bit										Transfer	Bit		
0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
2	1	1	0	0	0	0	0	0	0	0	0	0	0	1
3	1	1	0	0	0	0	0	0	0	0	0	0	0	1
4	1	1	0	0	0	0	0	0	0	0	0	0	0	1
5	1	1	0	0	0	0	0	0	0	0	0	0	0	1
6	1	0	0	0	0	0	0	0	0	0	0	0	0	1
7	1	0	0	0	0	0	0	0	0	0	0	0	0	1
8	1	0	0	0	0	0	0	0	0	0	0	0	0	1
9	1	0	0	0	0	0	0	0	0	0	0	0	0	1
10	1	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	0	0	0	0	0	0	0	0	0	0	0	0	1

Alert Frame Format

An AMB will begin sending Alert Frames in place of Idle Frames on the northbound channel whenever an error has been detected on the southbound channel and will continue to do so until it receives a Soft Channel Reset command or a channel reset. The Alert Frame format is the inverse of the corresponding Idle Frame. TABLE 3-13 shows the format of the Alert Frame that replaces the Second Idle Frame.

TABLE 3-13 Alert Frame Replacing First Idle Frame

Xfer	Bit													
	13	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
0	0	0	1	1	1	1	1	1	1	1	1	1	1	0
1	0	0	1	1	1	1	1	1	1	1	1	1	1	0
2	0	0	1	1	1	1	1	1	1	1	1	1	1	0
3	0	0	1	1	1	1	1	1	1	1	1	1	1	0
4	0	0	1	1	1	1	1	1	1	1	1	1	1	0
5	0	0	1	1	1	1	1	1	1	1	1	1	1	0
6	0	1	1	1	1	1	1	1	1	1	1	1	1	0
7	0	1	1	1	1	1	1	1	1	1	1	1	1	0
8	0	1	1	1	1	1	1	1	1	1	1	1	1	0
9	0	1	1	1	1	1	1	1	1	1	1	1	1	0
10	0	1	1	1	1	1	1	1	1	1	1	1	1	0
11	0	1	1	1	1	1	1	1	1	1	1	1	1	0

Data Frame Format

TABLE 3-14 shows the format of a Northbound Read Data Frame for x4 devices. It contains two 72-bit data packets, each with 12-bit CRC protection.

TABLE 3-14 Northbound Data Frame Format

Xfer	Bit													
	13	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**	**0**
0	E1.0	E1.11	C17.D2	C16D0	C14.D2	C13.D0	C11.D2	C10.D0	C8.D2	C7.D0	C5.D2	C4.D0	C2.D2	C1.D0
1	E1.1	E1.10	C17.D3	C16D1	C14.D3	C13.D1	C11.D3	C10.D1	C8.D3	C7.D1	C5.D3	C4.D1	C2.D3	C1.D1
2	E1.2	E1.9	C18.D0	C16D2	C15.D0	C13.D2	C12.D0	C10.D2	C9.D0	C7.D2	C6.D0	C4.D2	C3.D0	C1.D2
3	E1.3	E1.8	C18.D1	C16D3	C15.D1	C13.D3	C12.D1	C10.D3	C9.D1	C7.D3	C6.D1	C4.D3	C3.D1	C1.D3

TABLE 3-14 Northbound Data Frame Format *(Continued)*

Xfer	Bit													
4	E1.4	E1.7	C18.20	C17.D0	C15.D2	C14.D0	C12.D2	C11.D0	C9.D2	C8.D0	C6.D2	C5.D0	C3.D2	C2.D1
5	E1.5	E1.6	C18.D3	C17.D1	C15.D3	C14.D1	C12.D3	C11.D1	C9.D3	C8.D1	C6.D3	C5.D1	C3.D3	C2.D0
6	E2.0	E2.11	C17.D2	C16.D0	C14.D2	C13.D0	C11.D2	C10.D0	C8.D2	C7.D0	C5.D2	C4.D0	C2.D2	C1.D0
7	E2.1	E2.10	C17.D3	C16.D1	C14.D3	C13.D1	C11.D3	C10.D1	C8.D3	C7.D1	C5.D3	C4.D1	C2.D3	C1.D1
8	E2.2	E2.0	C18.D0	C16.D2	C15.D0	C13.D2	C10.D0	C10.D2	C9.D0	C7.D2	C6.D0	C4.D2	C3.D0	C1.D2
9	E2.3	E2.8	C18.D1	C16.D3	C15.D1	C13.D3	C10.D1	C10.D3	C9.D1	C7.D3	C6.D1	C4.D3	C3.D1	C1.D3
10	E2.4	E2.7	C18.D2	C17.D0	C15.D2	C14.D0	C12.D2	C11.D0	C9.D2	C8.D0	C6.D2	C5.D0	C3.D2	C2.D0
11	E2.5	E2.6	C18.D3	C17.D1	C15.D3	C14.D1	C12.D3	C11.D1	C9.D3	C8.D1	C6.D3	C5.D1	C3.D3	C2.D1

Northbound Register Data Frame Format

TABLE 3-15 shows the format for an AMB Register Read Data Frame.

TABLE 3-15 Northbound Register Data Frame Format

Xfer	Bit													
	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	E1.0	E1.11	0	0	0	0	0	0	D30	D24	D18	D12	D6	D0
1	E1.1	E1.10	0	0	0	0	0	0	D31	D25	D19	D13	D7	D1
2	E1.2	E1.9	0	0	0	0	0	0	0	D26	D20	D14	D8	D2
3	E1.3	E1.8	0	0	0	0	0	0	0	D27	D21	D15	D9	D3
4	E1.4	E1.7	0	0	0	0	0	0	0	D28	D22	D16	D10	D4
5	E1.5	E1.6	0	0	0	0	0	0	0	D29	D23	D17	D11	D5
6	E2.0	B2.11	0	0	0	0	0	0	0	0	0	0	0	0
7	E2.1	E2.10	0	0	0	0	0	0	0	0	0	0	0	0
8	E2.2	E2.9	0	0	0	0	0	0	0	0	0	0	0	0
9	E2.3	E2.8	0	0	0	0	0	0	0	0	0	0	0	0
10	E2.4	E2.7	0	0	0	0	0	0	0	0	0	0	0	0
11	E2.5	E2.6	0	0	0	0	0	0	0	0	0	0	0	0

Northbound Status Frame Format

TABLE 3-16 shows the format of the northbound Status Frame. The Status Frame is sent in response to Sync Frame on the southbound channel. The Status Frame returns at the time that the first Read Data Frame would return if a read were issued at the time of Sync Frame plus an additional delay determined by the SD[1:0] field of the Sync command. Each AMB sends its status back on the bit lane corresponding to its AMB_ID. The S[3:0] bits are the status information from the configuration register selected in the Sync command (TABLE 3-17). The SP bit is the odd parity of S[3:0].

TABLE 3-16 Status Frame Format

Xfer	Bit													
0	0	0	DBS0	DAS0	D9S0	D8S0	D7S0	D6S0	D5S0	D4S0	D3S0	D2S0	D1S0	D0S0
1	0	0	DBS1	DAS1	D9S1	D8S1	D7S1	D6S1	D5S1	D4S1	D3S1	D2S1	D1S1	D0S1
2	0	0	DBS2	DAS2	D9S2	D8S2	D7S2	D6S2	D5S2	D4S2	D3S2	D2S2	D1S2	D0S2
3	0	0	DBS3	DAS3	D9S3	D8S3	D7S3	D6S3	D5S3	D4S3	D3S3	D2S3	D1S3	D0S3
4	0	0	DBSP	DASP	D9SP	D8SP	D7SP	D6SP	D5SP	D4SP	D3SP	D2SP	D1SP	D0SP
5	0	1	0	1	0	1	0	1	0	1	0	1	0	1
6	1	0	1	0	1	0	1	0	1	0	1	0	1	0
7	0	1	0	1	0	1	0	1	0	1	0	1	0	1
8	1	0	1	0	1	0	1	0	1	0	1	0	1	0
9	0	1	0	1	0	1	0	1	0	1	0	1	0	1
10	1	0	1	0	1	0	1	0	1	0	1	0	1	0
11	0	1	0	1	0	1	0	1	0	1	0	1	0	1

TABLE 3-17 Status Bit Description

Field	Name	Description
FBD Status 0		
SP	Parity	Parity of S[3:0]
S3	NBDE	Northbound Debug Event
S2:S1	Thermal_Trip	AMB thermal information for thermal management
S0	Alert_Asserted	An error has been detected by the AMB.
FBD Status 1		
SP	Parity	Parity of S[3:0]

TABLE 3-17 Status Bit Description *(Continued)*

Field	Name	Description
S[3:1]	reserved	reserved
S[0]	Data_Merge_Error	AMB cannot meet northbound data merge timing requirement.
FBD Status 2		
SP	Parity	Parity of S[3:0]
S[3:0]	reserved	reserved
FBD Status 3		
SP	Parity	Parity of S[3:0]
S[3:0]	reserved	reserved

3.3.3 SDRAM Initialization

The initialization sequence within the MCU for the SDRAMs will still follow the same flow; however, the interface will be different. The MCU will have to initialize the SDRAMs indirectly through registers in the AMBs. The MCU will issue a command to the AMBs and then poll status registers to determine when the AMBs have completed issuing the command to the SDRAMs.

After SDRAM initialization is complete, the MCU will begin scheduling commands directly to the SDRAMs.

The DDR2 SDRAMs must be powered up and initialized in a predefined manner. Operational procedures other than those specified may result in undefined operation. TABLE 3-18 shows the sequence of steps required for POWER UP and Initialization.

TABLE 3-18 SDRAM Power Up and Initialization Sequence

Step	Required Action
1.	Apply power to VDD.
2.	Apply power to VDDQ.
3.	Apply power to VREF and to the system VTT.
4.	Start clock and maintain stable condition for 200 s.
5.	Apply No Operation or Deselect command and take CKE high.
6.	Wait minimum of 400ns, then issue a Precharge-all command.
7.	Issue Extended Mode Register 2 Set (EMRS(2)) command.
8.	Issue Extended Mode Register 3 Set (EMRS(3)) command.

TABLE 3-18 SDRAM Power Up and Initialization Sequence *(Continued)*

Step	Required Action
9.	Issue Extended Mode Register 1 Set (EMRS(1)) command to enable DLL.
10.	Issue Mode Register Set (MRS) command to reset DLL.
11.	Issue Precharge-all command.
12.	Issue two or more Auto-Refresh commands.
13.	Issue MRS command with low on A8 to initialize device operation (i.e. to program operating parameters without resetting the DLL).
14.	At least 200 clocks after step 8, execute OCD Calibration (Off Chip Driver Impedance adjustment). If OCD calibration is not used, EMRS OCD Default command (A9=A8=A7=1) followed by EMRS OCD Calibration Mode Exit command (A9=A8=A7=0) must be issued with other parameters of EMRS.
15.	The DDR2 SDRAM is now ready for normal operation.

3.3.4 DDR2 SDRAM Commands

TABLE 3-19 shows the truth table for the commands supported by the DDR2 SDRAMs.

TABLE 3-19 DDR2 SDRAM Command Truth Table

Function	CKE Previous Cycle	CKE Current Cycle	CS#	RAS#	CAS#	WE#	Bank	Address
Mode/Extended Mode Register Set	H	H	L	L	L	L	BA	Op-Code
Auto-Refresh	H	H	L	L	L	H	X	X
Self-Refresh Entry	H	L	L	L	L	H	X	X
Self-Refresh Exit	L	H	H	X	X	X	X	X
			L	H	H	H		
Single Bank Precharge	H	H	L	L	H	L	BA	A10=L
Precharge All Banks	H	H	L	L	H	L	X	A10=H
Bank Activate	H	H	L	L	H	H	BA	Row Address
Write								Column Address A10=L

TABLE 3-19 DDR2 SDRAM Command Truth Table *(Continued)*

Function	CKE Previous Cycle	CKE Current Cycle	CS#	RAS#	CAS#	WE#	Bank	Address
Write with Auto-Precharge								Column Address A10=H
Read								Column Address A10=L
Read with Auto-Precharge	H	H	L	H	L	H	BA	Column Address A10=H
No Operation	H	X	L	H	H	H	X	X
Device Deselect	H	X	H	X	X	X	X	X
Power Down Entry	H	L	H	X	X	X	X	X
			L	H	H	H		
Power Down Exit	L	H	H	X	X	X	X	X
			L	H	H	H		

3.3.4.1 Commands Supported by OpenSPARC T2

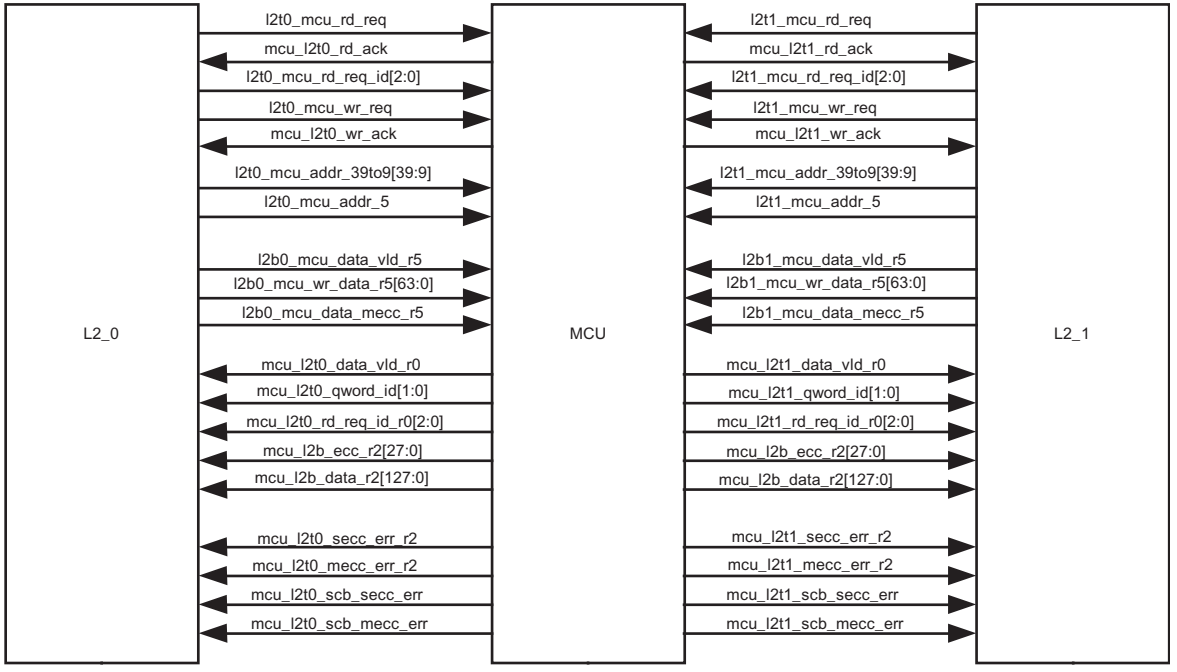
- Mode/Extended Mode Register Set - program Mode or Extended Mode Register which controls operation of SDRAM.
- Auto-Refresh - Refresh all SDRAM banks.
- Self-Refresh - Place SDRAM in refresh mode controlled by a timer within the SDRAM.
- Power Down - Place SDRAM in low power mode.
- Single Bank Precharge - Deactivate row in a particular bank.
- Precharge All Banks - Deactivate rows in all banks.
- Bank Activate - Activate a row within a particular bank.
- Write with Auto-Precharge - Perform a write operation and deactivate the bank after completion.
- Read with Auto-Precharge - Perform a read operation and deactivate the bank after completion.
- No Operation/Device Deselect - No operation.

- (Read and Write without Autoprecharge are not supported in OpenSPARC T2 since a bank is always closed after a transaction.)

3.4 MCU-L2 Cache Interface

An L2 cache bank can send one read or write request at a time to the MCU. Once it sends a request it must wait for the appropriate acknowledge before sending the next request. There is a delay of three cycles from the completion of a transaction (acknowledge for a read, last data word for a write) until the next request can be made. There can be a total of eight outstanding read requests and eight outstanding write requests from each L2 cache bank at any time.

FIGURE 3-5 MCU-L2 Cache Interface Signals



3.4.1 MCU Read Transaction

The L2 Cache makes a read request by asserting l2t_mcu_rd_req to the MCU at the same time that l2t_mcu_addr[39:4] and l2t_mcu_rd_req_id[2:0], the index into the L2 cache's fill buffer for returning data, are valid. The request is registered in the MCU and held in the l2clk domain until l2if_cmp_ddr_sync_en is high. (l2if_cmp_ddr_sync_en is a registered version of cmp_ddr_sync_en, the top-level clock synchronization signal.) This signal is used to synchronize the signals crossing from the l2clk (1.4 GHz) domain to the drl2clk (800 MHz) domain and indicates that the request will be stored in the Read Request Queue on the next drl2clk cycle. FIGURE 3-6 shows the timing of the read request and the internal acknowledge. There is a two-deep FIFO on the read request port. When a read request comes in, it is placed in the FIFO. An acknowledge for a transaction is sent to the L2 cache when that transaction reaches the head of the FIFO, either when a transaction is placed into an empty FIFO or when both entries are full and then an entry is dequeued.

No flow control is needed for the returning read data because the L2 cache will guarantee space to receive the data. When mcu_l2t_data_vld_r0 signal is asserted, the signals mcu_l2t_qword_id_r0[1:0] and mcu_l2t_read_req_id[2:0] are driven at the same time, and mcu_l2b_data_r3[127:0] and mcu_l2b_ecc_r3[27:0] are driven three cycles later. Also, mcu_l2t_secc_err_r3 or mcu_l2t_mecc_err_r3 will be asserted at the same time as the data if a correctable or uncorrectable error, respectively, occurred in the corresponding data beat. The data is returned to the L2 cache over several cycles because of the difference between l2clk and drl2clk. FIGURE 3-6 and FIGURE 3-7 show an example of a six to one l2clk to drl2clk ratio for a read request. The order in which the data beats are returned to the L2 cache depends on the bit PA[5].

Since an L2 bank can only have eight outstanding read requests at a time, and the MCU can handle eight outstanding reads per L2 bank, the MCU does not have to keep track of the number of outstanding reads.

Reads are not necessarily serviced in the order they are received from the L2 cache. Transactions are scheduled in order to limit the amount of dead data cycles on the bus to the DIMMs.

FIGURE 3-6 Read Request Timing

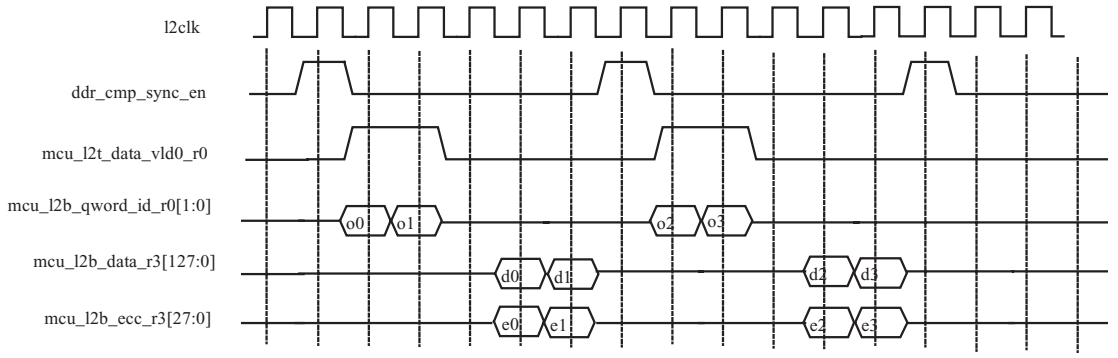
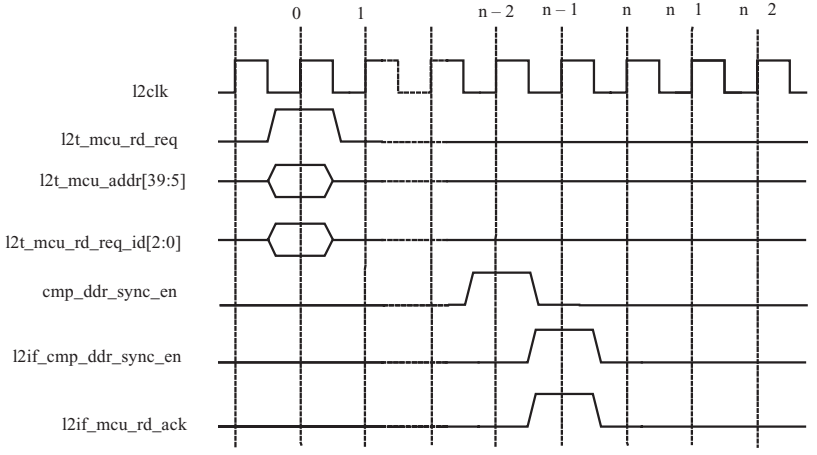


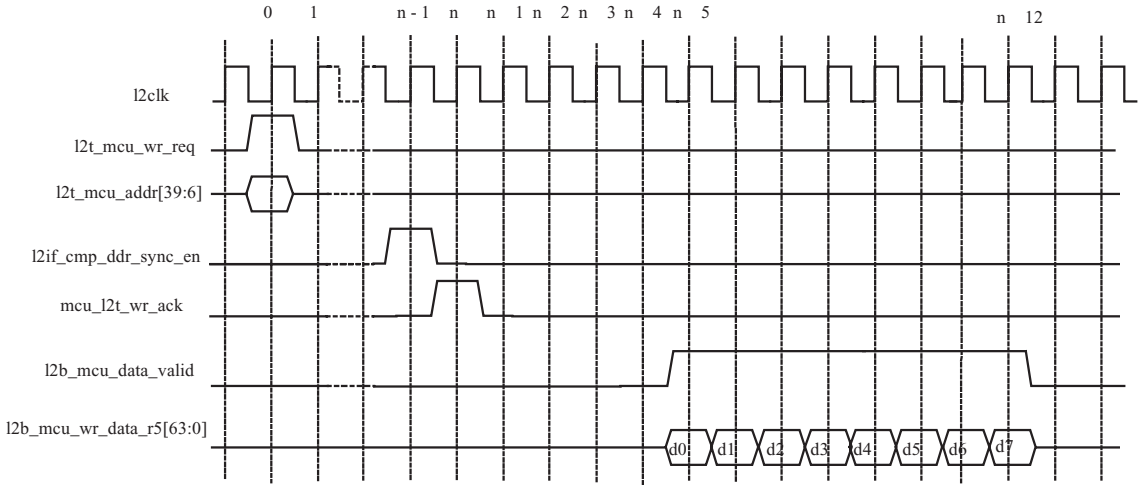
FIGURE 3-7 Read Data Return Timing



3.4.2 MCU Write Transaction

The L2 Cache makes a write request by asserting l2t_mcu_wr_req to the MCU at the same time that l2t_mcu_addr[39:6] is valid. Once the transaction is placed into the Write Request Queue, mcu_l2t_wr_ack is sent back to the L2 cache to indicate that it can now send another write transaction. l2b_mcu_data_valid and the write data l2b_mcu_wr_data_r5[63:0] are asserted five cycles after the acknowledge. The write data is sent to the MCU over eight cycles for a total of 64 bytes of data.

FIGURE 3-8 Write Request Timing



3.5 DDR2 SDRAM Transaction Timing

3.5.1 Memory Read

The MCU reads data from the external memory by:

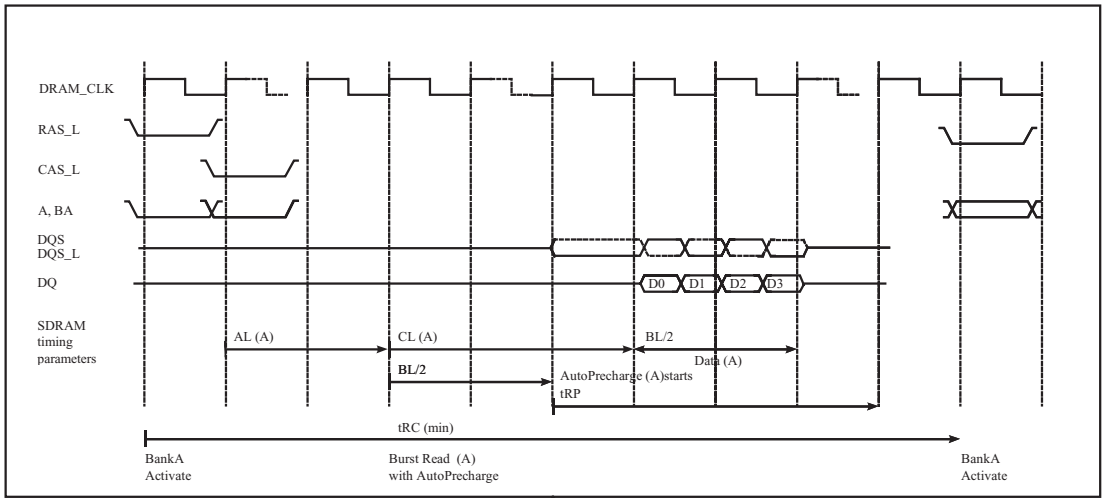
- Issuing a bank activate by assert RAS and driving the row address and the bank select.
- Issuing a Posted CAS Burst Read with AutoPrecharge command by asserting CAS and driving the column address.
- Waiting for a delay of AL (Additive Latency) + CL (CAS latency) before sampling the read data.
- Sampling data returning in a burst length of four in two drl2clk cycles.

A new Bank Activate command may be issued to the same bank if the following conditions are satisfied:

- The RAS precharge time (tRP) has been satisfied from the clock cycle at which the AutoPrecharge begins.
- The RAS cycle time (tRC) from the current Bank Activate has been satisfied.

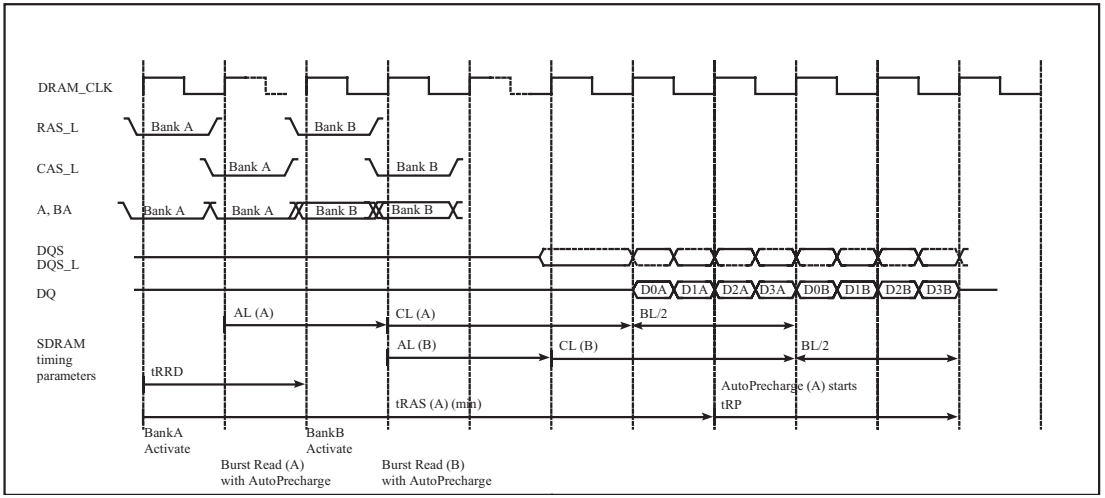
FIGURE 3-9 shows a single burst read of length four (BL=4) with AutoPrecharge, followed by a reactivation of the same bank.

FIGURE 3-9 Memory Burst Read with AutoPrecharge, Same Bank Reactivated



For seamless back-to-back memory reads, a different bank (B) can be activated during the memory read of bank (A). Bank (B) can start a burst read with AutoPrecharge command after a delay of BL/2 = 2cycles from the bank (A) burst read with AutoPrecharge command. FIGURE 3-10 shows a seamless burst read with AutoPrecharge command of two different banks.

FIGURE 3-10 Memory Burst Read with AutoPrecharge with Multiple Banks Activated



Note – There is an additional cycle delay when switching from one rank to a different rank.

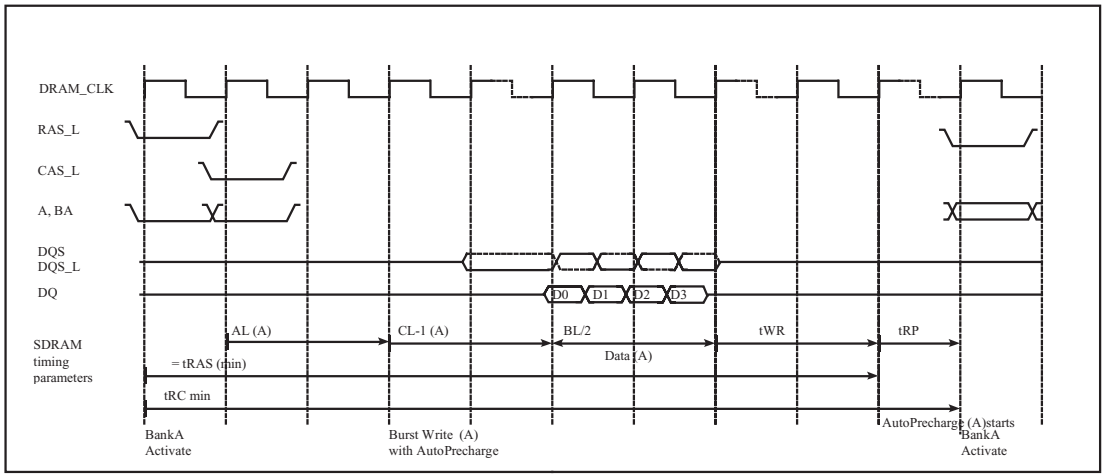
3.5.2 Memory Write

The MCU writes data to the external memory by:

- Issuing a bank activate by asserting RAS and the row address and the bank selects.
- Issuing a Posted CAS Burst Write with AutoPrecharge command by asserting CAS and driving the column address.
- Waiting for delay of (AL + CL-2) cycles from the CAS assertion, deasserting the data strobe (DQS/DQS_L) for one cycle.
- Waiting for a delay of (AL + CL-1) cycles from the CAS cycle before driving the data strobe and the write data.

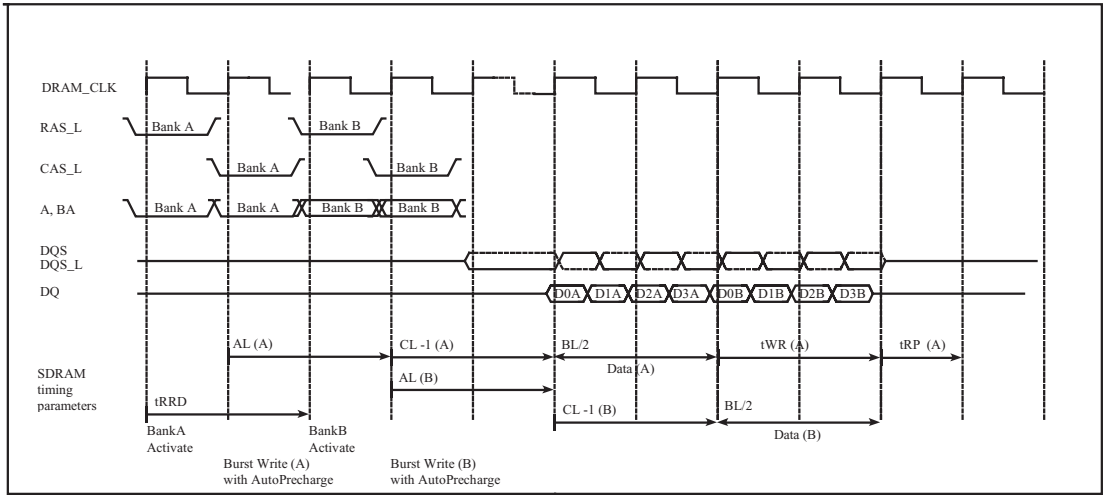
The same bank can be reactivated after a total delay of (1 + AL + CL-1 + BL/2 + tWR + tRP) cycles for reactivating the same bank. FIGURE 3-11 shows a single burst write with AutoPrecharge command.

FIGURE 3-11 Memory Burst Write with AutoPrecharge and Same Bank Activate



For a seamless memory write, a different bank (B) can be activated during the memory write of bank (A). The bank (B) burst write with AutoPrecharge command after a delay of BL/2 from the bank (A) burst write with AutoPrecharge command. FIGURE 3-12 shows a seamless burst write with AutoPrecharge command of two different banks.

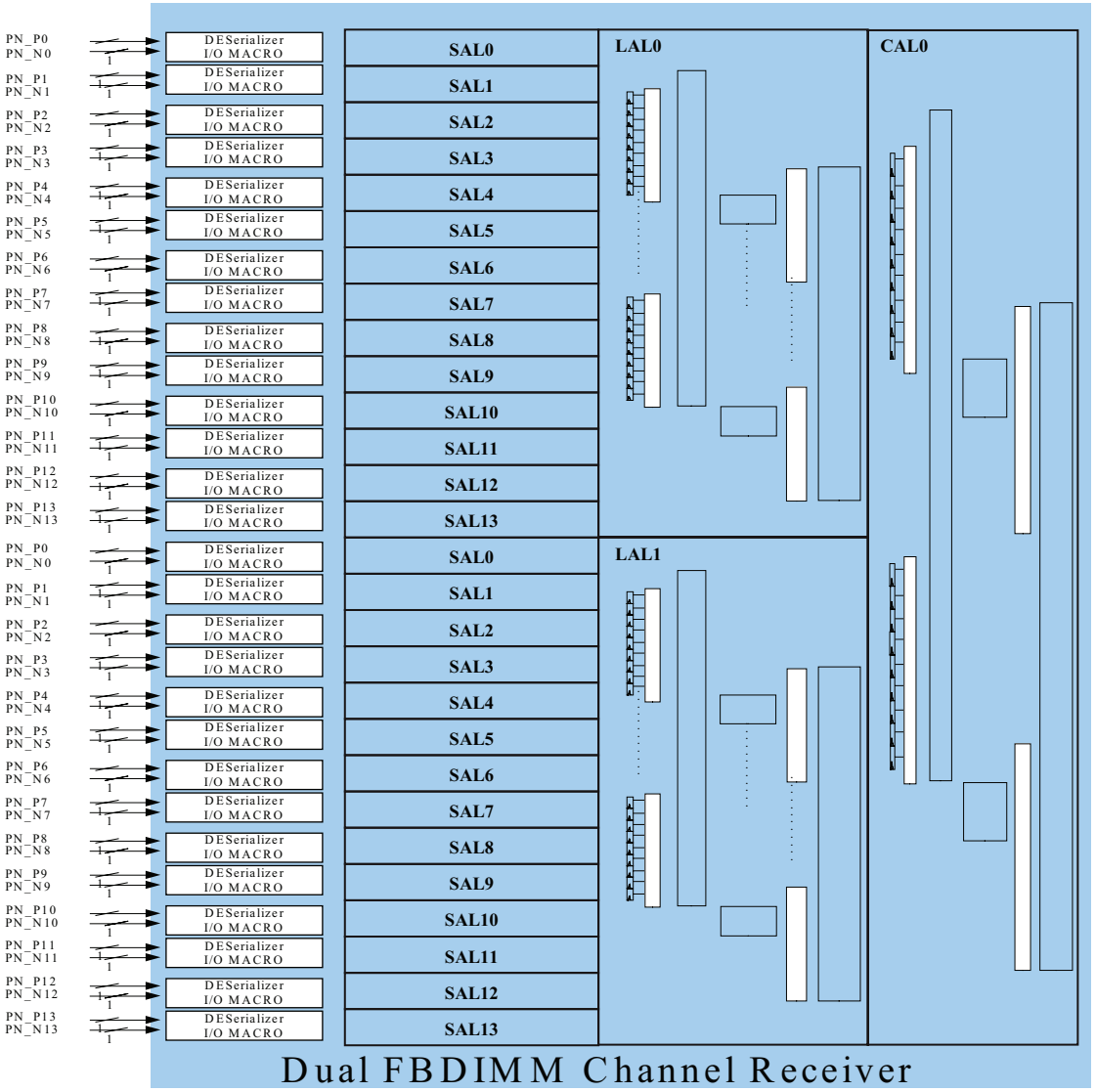
FIGURE 3-12 Memory Burst Write with AutoPrecharge and Multiple Banks Activated



3.5.3 SERDES (I/O) Timing

The SERDES handles the physical layer of the FBD channel. The packets from MCU are converted into frames of bits sent out on the serial link. On the northbound serial lanes, the data returns from the memory link. Each frame going southbound is divided into 10 bit lanes and 12 bit times or Unit Interval (UI). The maximum bit lane frequency is 4.8GHz. Each frame going northbound is divided into 14 bit lanes and 12 bit times. The DRAM clock is 1/12 the link's frequency.

FIGURE 3-13 Dual FBDIMM Channel Receiver



On the receive side, each SERDES macro recover its high speed clock, convert the electrical signals into a bit stream of logical 1s and 0s, and outputs a group (1 symbol) of 12 bits every 400MHz max. There are three fundamental microarchitectural issues for the host receiver:

- Frame/symbol alignment logic (SAL) within a single bit lane

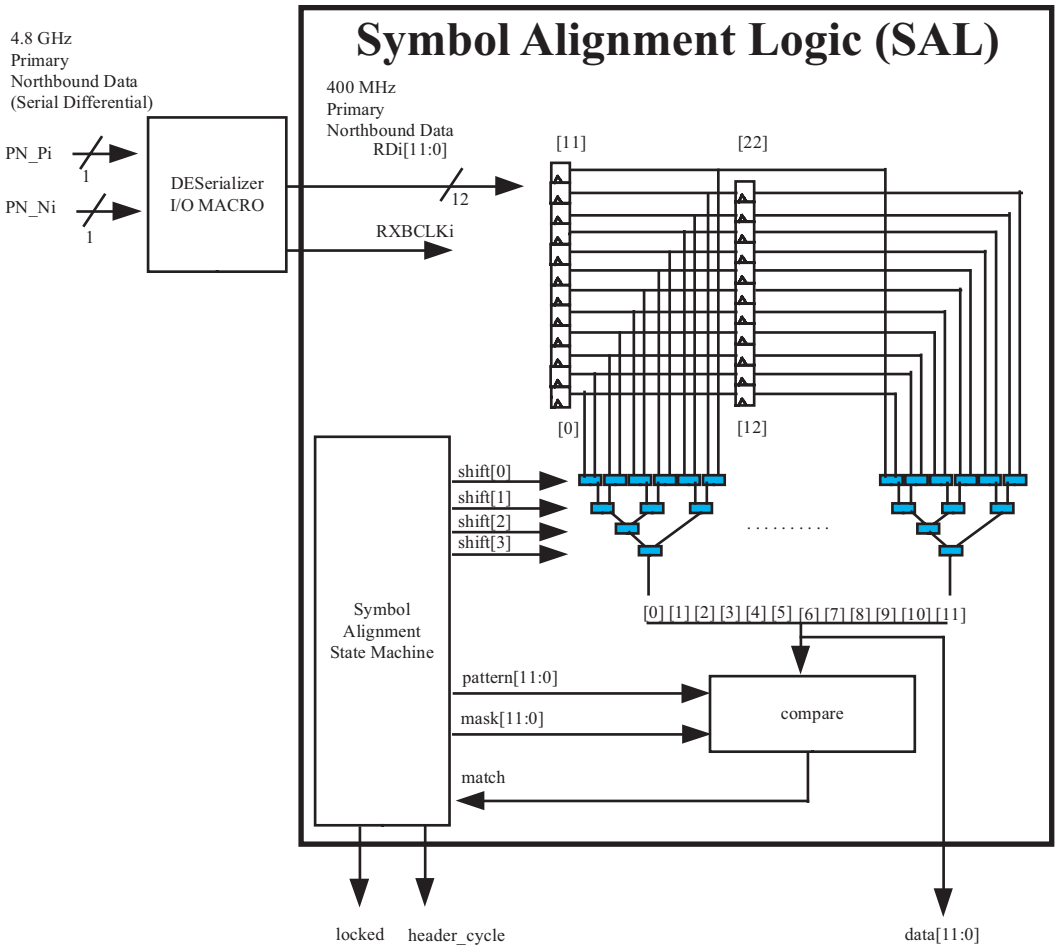
- Frame/lane alignment logic (LAL) across all 14 northbound lanes needed so MCU data layer does not see the physical aspects of a single channel
- Frame/channel alignment logic (CAL) across two northbound channels (across potentially all 28 northbound lanes) needed so MCU data layer does not see the physical aspects of two independent channels but sees instead dual channels operating logically in lockstep.

To minimize the number of wires between MCU and the SERDES logic, especially for supporting the dual channel case, half a frame is transferred every MCU's drl2clk clock cycle, which implies that drl2clk frequency must be at least 1/6 rather than 1/12 the link frequency.

3.5.3.1 Single Lane Symbol Alignment Logic

FIGURE 3-14 shows the SAL block needed if the macro doesn't symbol lock. The bit order for the 12 bit data is bit position 0 corresponds to oldest received bit on the serial link.

FIGURE 3-14 Symbol Alignment Logic

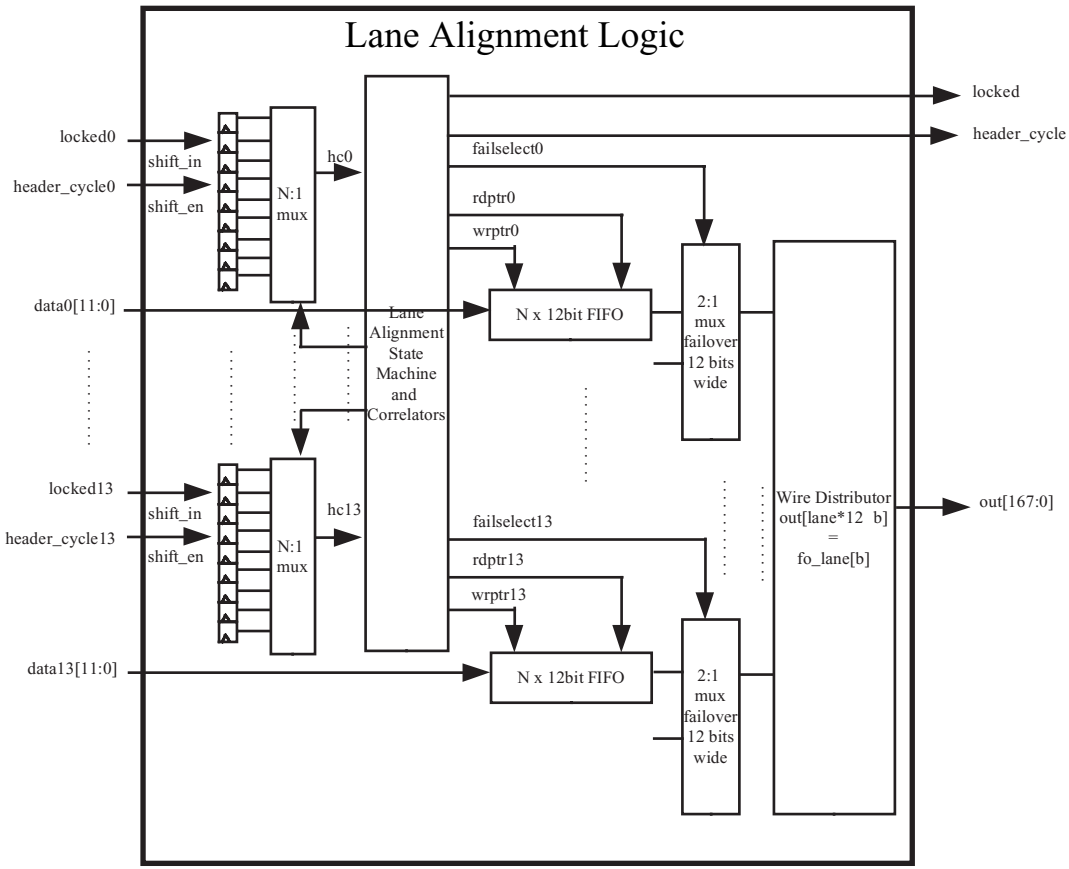


3.5.3.2 Frame Lane Alignment Logic across all 14 Northbound Lanes

Each bit lane has a static skew relative to another bit lane due to differences in drivers, receivers and traces between bit lanes. The FBDIMM link specification requires only the host perform the deskew on the northbound direction. Each AMB is required to deskew in the southbound direction.

Even when an AMB sends out the same symbol on all of the northbound lanes all initially aligned, a symbol from one SAL instance is not guaranteed to arrive at the same cycle as the same symbol from another SAL instance. A full frame must be collected and all 14 northbound lanes must be aligned to meaningfully interpret the original content. This lane alignment is done during link training. During training state 0 (TS0), identical frames are repeated many times and all lanes carry the same sequence of symbols. The header cycle of the frame is a uniquely identifiable symbol compared to the rest of the symbols in the frame. During TS0, the Lane Alignment Logic (LAL) queues up the symbols from each lane and search for the header cycle symbol. Once the skew between lanes have been determined consistently across multiple TS0 patterns, the LAL locks onto the wavefront and begins outputting frame aligned data for the full channel.

FIGURE 3-15 Lane Alignment Logic



Algorithm

```
lal_locked =0, lal_hc = 0, out=0, wrptr[0]=wrptr[1]=wrptr[2]=...=
wrptr[13] = 0, fastest_found = 14, lockcount=0, master_offset_counter
= 0; master_TS0pattern_counter = 0;

foreach cycle {

while (TS0state)  {

    if (lockcount<13) {

    foreach lane (i=0 to i=13) {

        if sal_locked[lane] {

           if sal_hc[lane] {

              if !fastest_found() {

          /* assume lane is fastest */

          /* set its rdptr to the deepest entry of the symbol fifo */

          rdptr[lane] = N;

          set_lockseen[lane] = 1;

          start_master_offset_counter(); /* counter increments each
cycle */

          start_master_TS0pattern_counter(); /* counter increments
every TS0 pattern */

          set_fastest_found(lane);

          }

               else if !relationship_exists_to_fastest() {

          if (master_offset_counter_value()  > N) /* something is
wrong... */

              error_handle(SKEW_TOO_LARGE);

          } else {

              rdptr[lane] =
rdptr[fastest_lane]-master_offset_counter_value();

              increment_lock_count();

          }
```

```
                else if !same_relationship_to_fastest() {

               error_handle(DYNAMIC_SKEW_TO_FASTEST);

          }

             } /* if sal_hc[lane]

        } else { /* sal_locked is false */

          if (lockseen[lane]) {

             error_handle(SYMBOL_LOCK_UNLOCKED);

          }

         } /* foreach lane */

      } /* if lockcount < 13 */

      else {

          zero_offset_to_slowest(); /* for all i {rdptr[i]=
rdptr[i]-rdptr[slowest]}

              }

} /* while TS0 state */


lal_locked = (lock_count>=13) && (master_TS0pattern_counter > 2);

lal_header_cycle = lal_locked && (symbol(fastest_found,
rdptr[fastest_found]) == TS0_HEADER);

lal_out[167:0] = concat( failsymbol(0, rdptr[0], rdptr[1]),
failsymbol(1, rdptr[1], rdptr[2]), ... , symbol13(1, rdptr[13]));
```

The Lane alignment state machine has direct control of the write pointers and read pointers of 14 FIFOs. The SAL's lock signal is the write and shift enable. The state machine tracks the procession of each SAL's header_cycle signal over many cycle to determine the relative delays between the lanes. The first lane to have its SAL header_cycle assert is declared the fastest lane. It's symbol read pointer is set to the deepest entry (N). Each time a header cycle is detected on a locked lane, its symbol read pointer is set to the appropriate distance 'earlier' from the fastest's symbol read pointer. Once 13 out of the 14 lanes (enough to support failover) have achieved lock, the read pointer for all the lanes are subtracted such that the slowest lane's symbol read pointer is 0, while all the other lane remains the same distance away. This is for latency purpose only. A separate counter makes sure lane lock is not declared until multiple TS0 patterns have been elapsed with all 13 or 14 lanes locked.

For efficiency, each SAL's header_cycle signal is accumulated using flops to allow for arbitrary access for correlation purpose. Until lane alignment lock has occurred, the 168 bit data out is indeterministic. The lane alignment state machine also handles masking out the failed lane during training state 3.

Architecturally Complete Implementation

Architecturally, FBDIMM allows for up to eight DIMMs and four logic analyzers on a single FBDIMM channel. All intermediate DIMMs transmitter are allowed to introduce 100ps + 2UI of skew. The last (southernmost) DIMM is allowed to introduce a longer lane skew for its northbound driver (100ps + 3UI). All receiver skew component have the same identical maximum (6UI).

Max lane-skew is

= (# AMBs -1) x (L_txskew2 + L_rxskew) + (L_txskew1 + L_rxskew)

= (# AMBs - 1) x (100ps + 2UI + 6UI) + (100ps + 3UI + 6UI)

= 11 x (100ps + 8UI) + 100ps + 9UI

= 1200ps + 97UI

= Worse case UI max lane skew @ 4.8Gbps, 1UI=208ps

= ~6UI + 97UI = ~9 frames

= Worse case UI max lane skew @ 4.8Gbps, 1UI=208ps

= ~6UI + 97UI = ~9 frames

The TS0 pattern is 12 frames long, so with a 9 frame lane-skew, and 9 frame deep buffer, there would not be an alias problem of erroneously locking one lane's TS0 pattern to a prior or next TS0 pattern on another lane. IF ANY TS0 patterns were dropped on any lanes by the designated AMB while it lane-deskews and the host locks the northbound lanes prior to that AMB locking its southbound lanes, then the true maximum allowable skew would be 6 frames long to prevent the situation of the fastest lane aliasing into a slower lane. A more robust (but much more costly) mechanism for the host would use TS1's unique end delimiters for the n-3, n-2, n-1, and n frames to correct any alias problem. After symbol lock, a budget of two or three TS0 patterns should be sufficient to assert lane-locked. If lock is not acquired after the total of 13 TSO patterns, the AMB_ID returned will be incorrect and MCU will need to look at lane-locked signal. If not locked, then it'll need to transition back to EI and do TS0 again.

Nine frames of lane skew can tolerate approximately 22 to 34 ns of delay due to trace mismatch. At about 61.5 ps/cm of trace velocity, that's about a mismatch of 50+ cm.

Having a 9 deep x 12 bit wide FIFO per lane seems excessive for the type of board design and blade system OpenSPARC T2 would go into. FBDIMM architecturally allows for very inexpensive board design and board manufacturing costs - relaxed board routing rules between lanes using cheap FR4 dielectric material - while allowing for a long latency system topology with DIMMs plugged on riser card and long chains.

PC Board/System Dependent OpenSPARC T2 Implementation

If a more realistic board design and system topology is available, the cost of the deep buffer and logic to quickly search exhaustively through theoretically 914 permutation of ordering could be significantly reduced. A typical blade server does not have that much cubic space of freedom nor expected to need to access more memory on another blade via the backplane.

A sample analysis from slides from Intel Spring'03 IDF FBDIMM and RawCard simulation results suggests a better upper bound for the lane trace mismatch.

The short answer is:

1st order approximation of frames needed

frames = roundup (N * (7 * D + L)/40)

N = settling time (# of wave propagations) of a tx/rcr pair.

D = max trace mismatch between a pair of DIMMs, in cm

L = max trace mismatch between host and northernmost DIMM.

Intel's "8 DIMM Layout" slide has the following:

- 0.4" DIMM to DIMM,

- shortest lead-in 1.2"

  (southbound channel 0)

- Longest lead-in 7.3"

  (northbound channel 3)

Using this as a recommended layout, .4" is practically 1cm.

Assuming OpenSPARC T2's board has at least a 6 layer so there can be a clean back current return path and sufficient shielding against radiation to the outside and edges of the PCB; Assuming these critical' signals avoid high impedance areas (vias,

gaps, and zones in ground planes) and takes the path of least inductance (like avoid going vertical between planes as much as possible) and also assuming 45 degree turns instead of 90 degree turns to avoid changing the 'w' parameter of the wave guide.

The FR4 board dielectric has a permittivity of 4.5 at 1MHz (number used for DIMM card), but is 3.4 for the GHz operation in the channel (using a higher channel impedance of 85 ohm)

Propagation delay calculation:

Assuming the simple stripline transmission line model where the

inner metal traces are sandwiched by the pair of ground or VCC planes,

separated by a FR4 dielectric (er = ~3.4), wave velocity

ignoring the higher order effects of the full RLGC,

v = c/sqrt (er)

= 300um/ps/sqrt (3.4)

= 0.5423 * 300 um/ps

= 1627 um/ps (or 61.5 ps/cm)

Relative to the UI of 208ps (4.8GHz), v ~= 30% of a UI

I'm also going to assume a very slow driver to minimize the spike in driver-current and ringing. FBDIMM does not explicitly specify a maximum slew rate but they're implied by the edge transition rate and eye diagrams. So setting N to be a simple 1.

1 * .3 UI/cm = 0.3 UI skew per cm of trace mismatch.

7 channels between the eight DIMMs @ 1 cm separate. It shouldn't be too difficult to lay out these 48 southbound traces on the OpenSPARC T2 board, all next to each other in parallel, and the DIMM connectors minimally spaced for a field technician to install the DIMMs. Bear in mind that requirement for each set of two differential conductors have very stringent skew mismatch requirement to avoid the collapse of the eye at the next receiver.

But let's suppose it's D=4 cm of trace mismatch each time we go from one DIMM to the next DIMM. This allows for going through the connector, through the AMB package pins to the actual transceivers and associated delay between the AMB secondary northbound receiver and AMB primary northbound driver) =>

4 * 7 * 0.3 UI = 8.4 UI

Now, the long leg -- between the northernmost DIMM and the host-- contributes the most skew. potentially 6" mismatch in Intel's case (although they were from different channels and opposite directions)

Call it L=16 cm trace mismatch

16 * 0.3 UI = 4.8 UI

Add them up and we are 13.2 UI. This is two frames only.

1st order approximation of frames needed

frames = roundup (N * (7 * D + L)/40)

N = settling time (# of wave propagations) of a tx/rcr pair.

D = max trace mismatch between a pair of DIMMs, in cm

L = max trace mismatch between host and northernmost DIMM.

3.5.3.3 Channel Alignment Logic across all Two FBDIMM Channels.

The FBDIMM Link specification does not specify the skews between two different channels, nor does it specify that the lane skew numbers are strictly for the same channel. If it can be assumed that the numbers for lane deskewing applies also regardless of number of channels as long as the host guarantees that it locks the two southbound channels then deskewing across two northbound channels can be done in two ways:

1. Expand the Lane alignment logic to support 28 simultaneous channels. This approach introduces less memory latency and alignment buffers but makes the search algorithm more complex (search within 928 permutations)

2. Use the two independent LAL outputs and add two sets of 9 deep FIFO and a delay counter or set of flops to record the delay (maximum 9 cycles apart) between one channel achieving lane lock and the other channel achieving lane lock.

To minimize bus width for a dual channel option, rather than route 168x2 (334) data bits per direction between the IO and MCU when both channels are used, a 168 bit data bus per direction is used and MCU runs at 2x dram speed (800MHz max). In both dual and single channel mode, fbd_mcu_data[83:0] always contain the primary channel's 1st half-frame the 1st cycle and the primary channel's 2nd half-frame the 2nd cycle. In dual channel mode, fbd_mcu_data[167:84] always contain the secondary channel's 1st half-frame the 1st cycle and the secondary channel's 2nd half-frame the 2nd cycle. In single channel mode, fbd_mcu_data[167:84] will always contain the 2nd half-frame on both cycles.

3.6 Memory Latencies

3.6.1 Read Latency

TABLE 3-20 shows the stages in the memory read pipeline and their approximate latencies for a 4-4-4 800 MHz DDR SDRAM with worst-case bit-lane deskewing.

The total latency does not include the time in the l2clk domain or the time to synchronize the data between the l2clk and drl2clk domains (A and I). These latencies assume that the MCU is idle when it receives the read request and that the request is going to the last of eight FBDs in the channel.

TABLE 3-20 Memory Read Pipeline and Latency

Stage	Clock Domain	Latency
A. L2 issues read request and MCU acknowledges	l2clk (1.4 GHz)	
B. MCU schedules read command	drl2clk (800 MHz)	7.5 ns (6 cycles)
C. Read request transmitted on SB FBD channel	fbdclk (4.8 GHz)	7.0 ns (1 ns per DIMM)
D. Bit-lane deskew	sdram clock (drl2clk/2 == 400 MHz)	22.5 ns (9 frames worst case)
E. Read command issued to SDRAMs	sdram clock	22.5 ns (9 cycles: CL + AL + 1 for frame alignment)

TABLE 3-20 Memory Read Pipeline and Latency

Stage	Clock Domain	Latency
F. Read data returned to MCU on NB FBD channel	fbdclk	7.0 ns (1 ns per DIMM)
G. Bit-lane deskew	drl2clk	22.5 ns (9 frames worst case)
H. MCU checks for CRC and ECC errors	drl2clk	3.75 ns (3 cycles)
I. Read data returned to L2	l2clk	
Total		92.75 ns

The read requests from the L2 cache are placed in the Read Request Queue (RRQ) which is checked every drl2clk for a transaction. When a transaction enters the RRQ, there are two cycles for arbitration, two to issue the Activate command, and two to issue the Read command. (two drl2clk cycles == one sdram cycle). The Activate and Read commands are transmitted on successive cycles on the high-speed FBD channel which runs at 12 times the sdram speed. The Activate and Read commands are driven to the SDRAMs one cycle after each reaches the FBD. After the data returns from the SDRAMs, the AMB places the data in the appropriate NB frame, 144 bits per frame, to return to the MCU. Once the MCU receives the read data, it checks for CRC errors in the frame and ECC errors in the data. If no errors are found, the data is returned to the L2 cache 128 bits per drl2clk cycle.

3.6.2 Write Latency

TABLE 3-21 shows the stages latencies in the memory write pipeline. The total latency does not include the time in the l2clk domain or the time to synchronize the data between the l2clk and drl2clk domains (A and B). These latencies assume that the MCU is idle when it receives the write request and that the request is going to the last of eight FBDs in the channel.

TABLE 3-21 Memory Write Pipeline and Latency

Stage	Clock Domain	Latency
A. L2 issues write request and MCU acknowledges	l2clk (1.4 GHz)	
B. L2 sends write data	l2clk	
C. MCU schedules write data and write command	drl2clk (800 MHz)	7.5 ns (6 cycles)
D. Write data and command transmitted on SB FBD channel	fbdclk (4.8 GHz)	7 ns (1 cycle per DIMM)

TABLE 3-21 Memory Write Pipeline and Latency

Stage	Clock Domain	Latency
E. Bit-lane deskew	sdram clock (drl2clk/2 == 400 MHz)	22.5 ns (9 frames worst case)
F. Write command issued to SDRAMs	sdram clock	2.5 ns (1 cycle)
G. AMB issues Idle frame on NB FBD channel	fbdclk	7 ns (1 cycle per DIMM)
H. Bit-lane deskew	drl2clk	22.5 ns (9 frames worst case)
I. MCU checks for Alert frame	drl2clk	1.25 ns (1 cycle)
Total		70.25 ns

When the L2 cache issues a write request, the MCU transfers the request from the l2clk domain to the drl2clk domain, places it in the Write Request Queue (WRQ), and sends an acknowledge back to the L2 cache. Once the L2 receives the acknowledge, it transmits the write data to the MCU 64-bits per cycle over eight cycles, and the MCU stores the write data in the Write Data Queue. The MCU sends the write data and write command independently on the FBD channel; however, it must ensure that the write data reaches the write data fifo within the AMB early enough that the AMB can write command the timing requirements to the SDRAMs. If the write command and data are received with no CRC errors, an Idle frame (as opposed to an Alert frame) is sent on the NB channel. Once the MCU sees there is no Alert frame, and thus no error on the write, it can the release the WRQ entry for the write.

3.7 Multiple Clock Domains

The MCU has three clock domains - l2clk, drl2clk, and iol2clk - and also interfaces to high-speed SERDES IOs for the DDR channels. The l2clk is the main cpu clock whose frequency is a multiple of the system clock, iol2clk. The l2clk and drl2clk are synchronous but do not have an integer ratio between them. The drl2clk will run at the same frequency as the SDRAM. The l2clk frequency target is 1.4 GHz, and the system clock target is 350 MHz. The SERDES IOs have a data rate of 12x the DDR rate - 3.2 GHz for 266 MHz DDR FBDs, 4.0 GHz for 333 MHz DDR FBDs, and 4.8 GHz for 400 MHz FBDs. The SERDES must run at one of these rates within +/- 5%.

The clock inputs to MCU are from the Clock Control Unit (CCU). The transmitting (l2if_cmp_ddr_sync_en, l2if_cmp_io_sync_en) and receiving (l2if_ddr_cmp_sync_en, l2if_io_cmp_sync_en) synchronization pulses are delayed versions of outputs from

the CCU which act as clock enable for synchronizing signals between two clock domains. The CCU will generate one of each of these enable pulses per MCU clock cycle.

Example waveforms for two clock ratios and the synchronizing signals across two clock domains are shown in FIGURE 3-16, FIGURE 3-17, and FIGURE 3-18. More detail on clock domain synchronization and the supported clock ratios can be found in the CCU specification.

FIGURE 3-16 Odd Ratio (13:2) Clock from the On-chip PLL Block

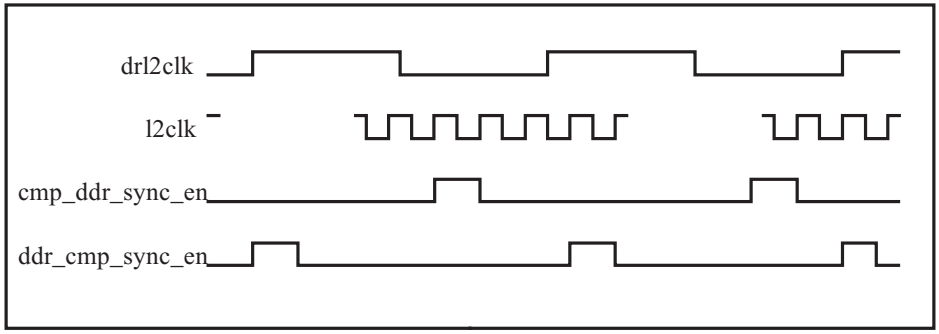


FIGURE 3-17 Even Ratio (12:2) Clock from the On-chip PLL Block

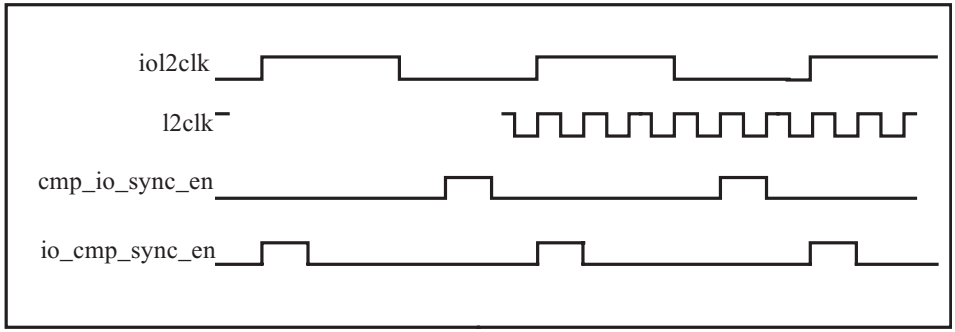
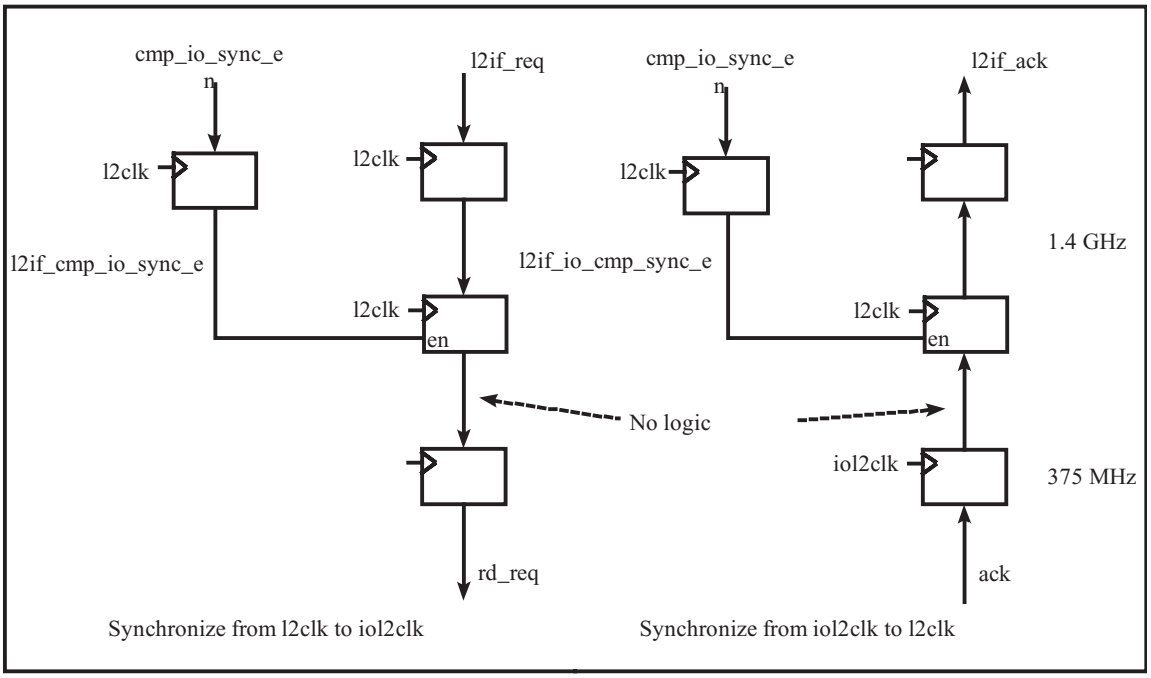


FIGURE 3-18 Example of Synchronizing between l2clk and iol2cls



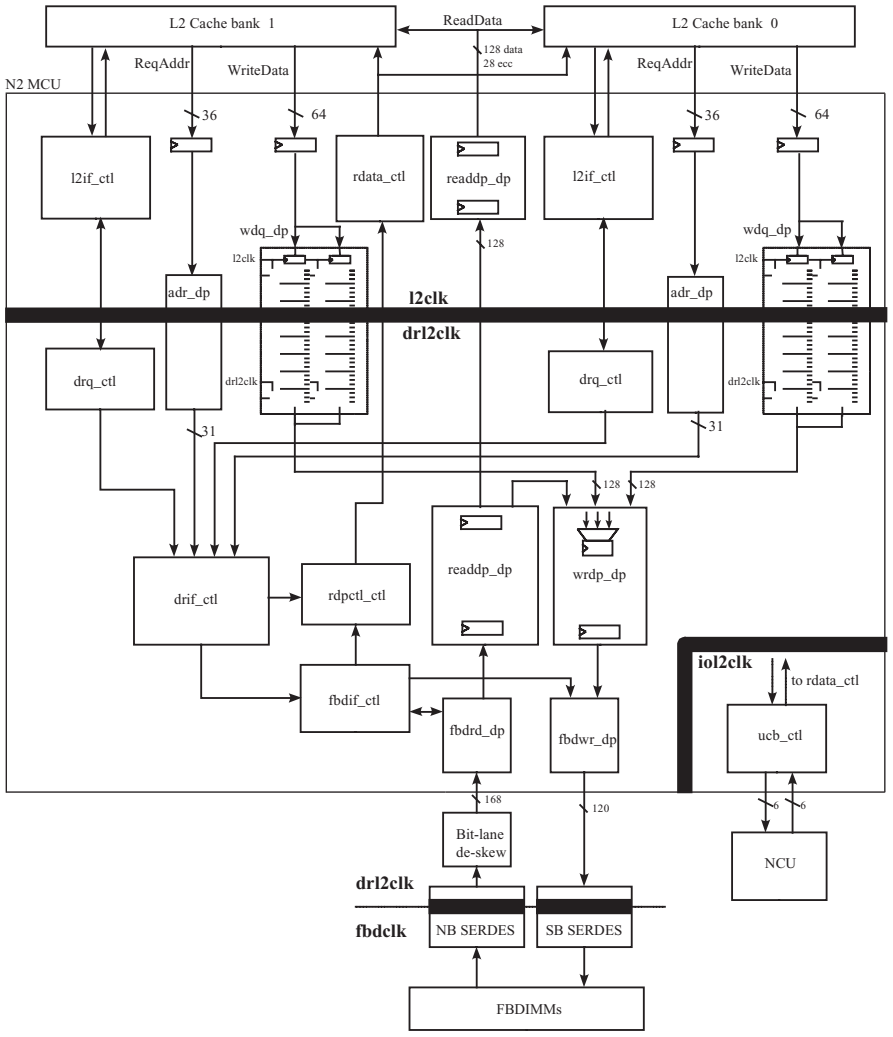
3.8 Functional Description

The MCU top level block diagram consists of the control logic and the datapaths to interface to the two L2 banks and the external DDR2 memory channel. The top level block diagram of the MCU is shown in FIGURE 3-19 with its three clock domains.

In FIGURE 3-18, requests come from the L2 cache banks to the L2 Cache Interface Control (L2IF_CTL) block. The incoming physical addresses are converted into DIMM addresses and stored in the Address Datapath (ADR_DP) block. For write requests, write data is stored in the Write Data Queue Datapath (WDQ_DP). The DRAM Request Queue Control (DRQ_CTL) block determines the order in which read and write transactions will go out to the DIMMs. The transactions are forwarded to the DRAM Interface Control (DRIF_CTL) block which generates the control signals for the transactions going out to the DIMMs. The DRAM Write Datapath (WRDP_DP) block generates ECC for the write data going to the DIMMs. Read data returning from the DIMMs goes through the Read Data Datapath (READDP_DP) which checks ECC and corrects single-bit errors and regenerates

ECC for the data before it is returned to the L2 cache. The Unit Control Block (UCB) logic unit provides the CSR interface for the MCU. Details of each of these blocks in given in the following sections.

FIGURE 3-19 MCU Block Diagram



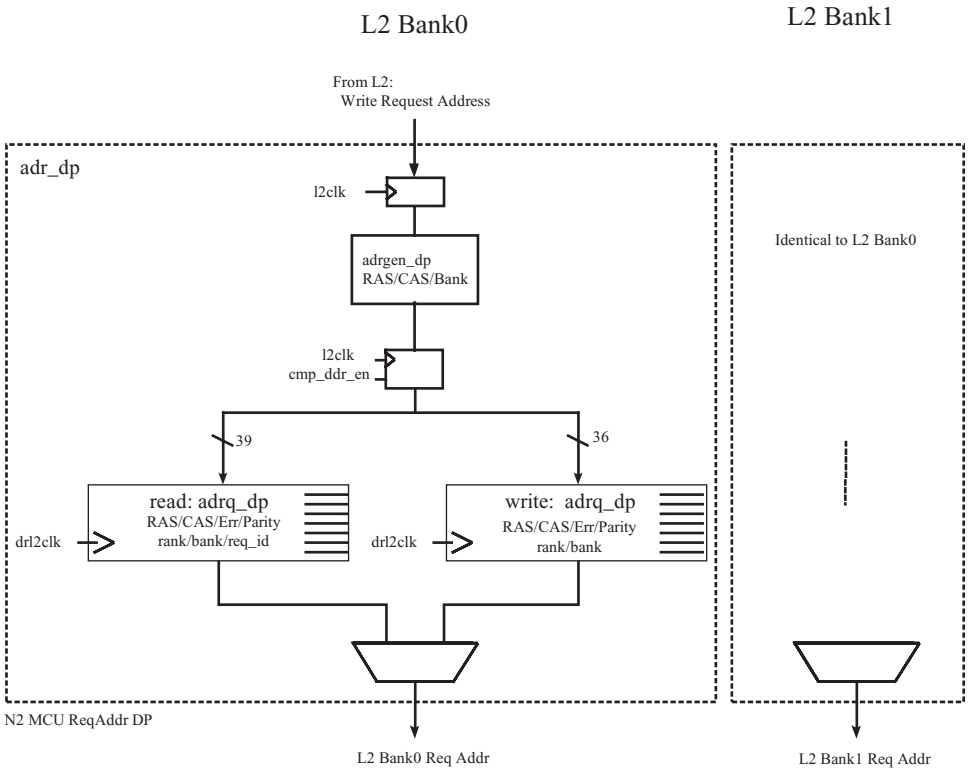
3.8.1 MCU Datapaths

3.8.1.1 Request Address Datapath

The physical memory address of the L2 bank read requests and write requests from the L2 cache are converted to the row and column addresses for the DIMMs in the system. The converted addresses are queued in the read request address queue or the write request address queue. These queues participate in arbitration for the Activate command cycle.

The block diagram of the request address datapath for the two L2 banks is shown FIGURE 3-20:

FIGURE 3-20 MCU Request Address Queue Datapath



DDR2 Address Generation (ADRGEN_DP)

The Address Generation block converts the incoming physical address to an SDRAM address consisting of the row, column, and bank address bits, and, if they exist, rank selects. Also, parity is generated for the address, and an address error bit is generated if an address is accessed that is out of range for the installed DIMMs.

Inputs to the block are the physical address and the MCU configuration registers which give the number of row and column address bits, number of ranks, number of internal banks, and whether the rank selects should come from upper or lower physical address bits. All of the address information is stored in the Read or Write Request Address Queue, and the rank and internal bank information will also be sent to the MCU Request Queue Control (drq_ctl) module.

The converted DIMM address consists of the following 36 bits:

- RAS address (15b)
- CAS address (14b)
- SDRAM internal bank address (3b)
- DIMM rank select (2b)
- Request address error (1b)
- Address parity (1b)

Refer to DDR2 Address Generation (ADRGEN_DP) for a description of how SDRAM address is converted from the physical address.

Read and Write Request Address Queues (ADRQ_DP)

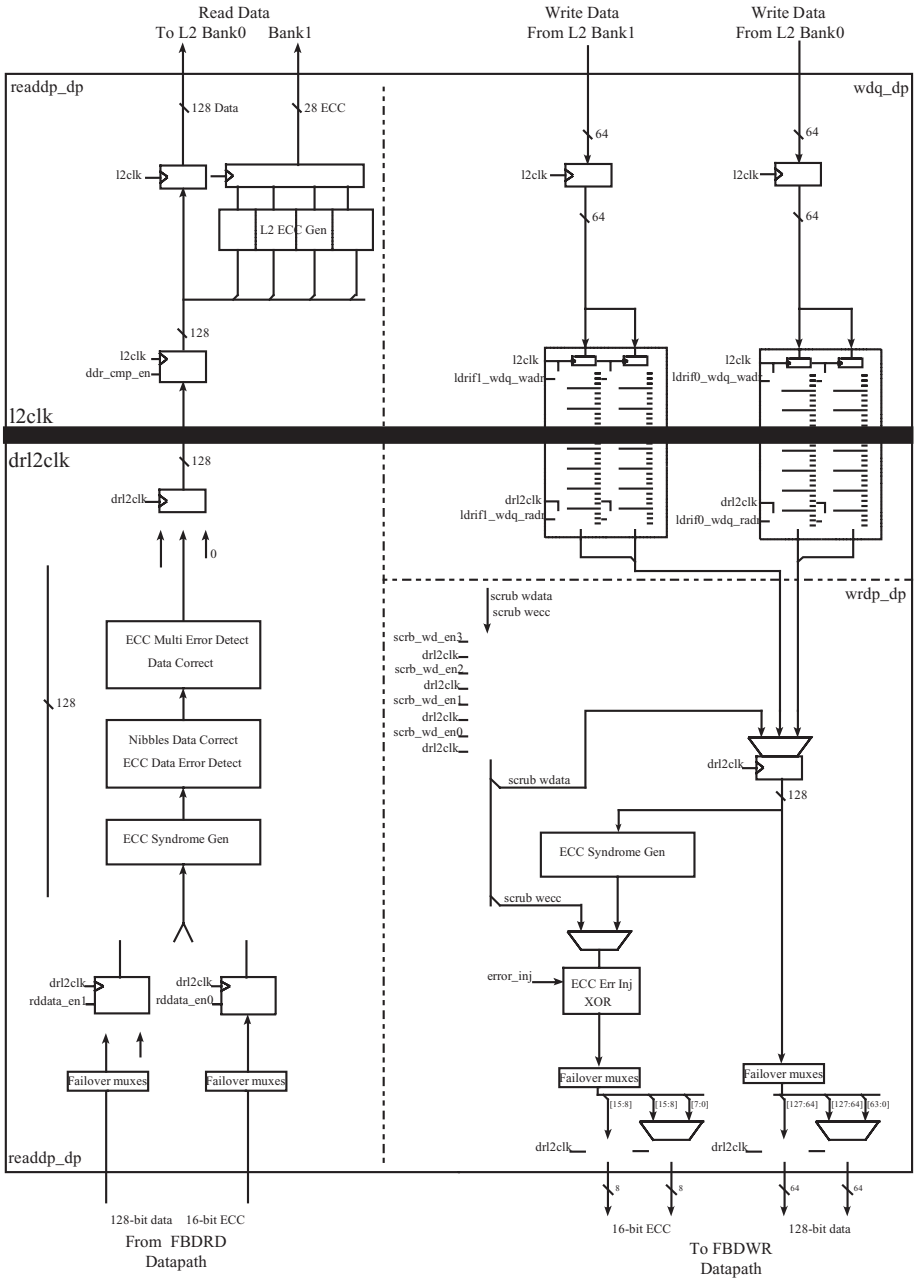
The Read and Write Request Address Queues store the converted DIMM addresses of the MCU memory requests. Each queue has eight entries of 36 bits. The 36 bits are the same as the converted DIMM address bits described in DDR2 Address Generation (ADRGEN_DP).

3.8.1.2 Read and Write Data Datapaths

The MCU read and write data datapaths block diagram is shown in FIGURE 3-21. These datapaths queue write data from the L2 cache to the DIMMs and read data returning from the DIMMs to the L2 cache.

The MCU read and write data datapath between the two L2 banks and the external memory channel consists of a write data queue per L2 bank (WDQ_DP), the write data datapath to the DRAM (WRDP_DP), and the read data return datapath to the L2 banks (READDP_DP). The MCU read and write data datapath supports two databus size interface modes to external memory - full size memory databus (128 bits data and 16 bits ECC), and half size memory databus (64 bits data and eight bits ECC).

FIGURE 3-21 Read and Write Datapaths Block Diagram



Write Data Queue Datapath (WDQ_DP)

The Write Data Queue stores the L2 write data to be written to the external memory. There are two Write Data Queues in each MCU module, one for each L2 cache bank that it communicates with. Each queue is composed of two dual port 1R/1W register files, each of which is 32 entries by 66 bits wide.

The L2 Cache Interface Control (L2IF_CTL) module controls write access to the WDQ_DP. The L2 write data is written to the write data queue at the rate of 64bits/cycle for eight l2clk cycles to complete one L2 cache line (64 bytes). Write enables are used to enable writing to one register file at a time.

Read access to the WDQ is controlled by the DRAM Interface Control (DRIF_CTL) module in the drl2clk domain. The output ports of the register files are concatenated and read as a single 128-bit bus.

Write Data Datapath (WRDP_DP)

The Write Data datapath block sends write data to the IO pads. A multiplexer controlled by the DRIF_CTRL module selects between the two Write Data Queues and scrubbed data coming from the Read Data Return datapath.

16 bits of ECC is generated for each 128 bits of data, and 144 bits of data and ECC are sent to the SDRAM for four cycles in the normal configuration. In single-channel mode, the 144 bits data are multiplexed into two 72-bit data packets and sent to the SDRAM over eight cycles.

Read Data Datapath (READDP_DP)

The 128 bits of data and 16 bits of ECC from the SDRAM are flopped in the drl2clk domain for ECC error detection and correction. In the single-channel mode, two 64-bit data packets are combined before ECC error detection and correction.

ECC is regenerated based on the 128 bits of data read from the SDRAMs and is compared with the ECC bits read from the SDRAMs. If a single bit error is detected, it is corrected in the data correction logic; however, if multiple errors are detected, no correction is done. The data are transferred from the dr2clk domain to the l2clk domain upon the assertion of the ddr_cmp_sync_en signal, or if the data is from a scrubbing request, it is sent to the Write Data datapath module.

All ECC errors on L2 cache reads and scrubbing reads are signaled to the L2 cache (mcu_l2t_scb_secc_err, mcu_l2t_scb_mecc_err, mcu_l2t_secc_err_r2, and mcu_l2t_mecc_err_r2) as well as being flagged in the MCU Error Status Register.

In the l2clk domain, the L2 cache ECC is generated on the 128 bits received from the drl2clk domain. The L2 cache uses a 7-bit Hamming code for to protect each 32-bit word, allowing single-bit error correction and double-bit error detection per 32 bits. After ECC is generated, the 128 bits of data and 28 ECC bits are then sent back to the L2 cache.

3.8.1.3 FBD Write and Read Datapaths (FBDWR_DP, FBDRD_DP)

The datapath portion of the FBD Controller will stage write data or the B and C command portions of the frame and generate all of the CRC values. There are six CRC generators required for each FBD channel, four in the FBDWR_DP and two in the FBDRD_DP.

FIGURE 3-22 shows the FBD Write Datapath and the following four CRC generation blocks:

1. CRC[13:0] for command A portion of 10-bit mode SB frame, generated from 26 bits of data.
2. CRC[9:0] for command A portion of 10-bit failover mode SB frame, generated from 26 bits of data.
3. CRC[21:0] for command BC/data portion of 10-bit mode SB frame, generated from 72 bits of data.
4. CRC[9:0] for command BC/data portion of 10-bit failover mode SB frame, generated from 72 bits of data.

FIGURE 3-22 FBD Write Datapath

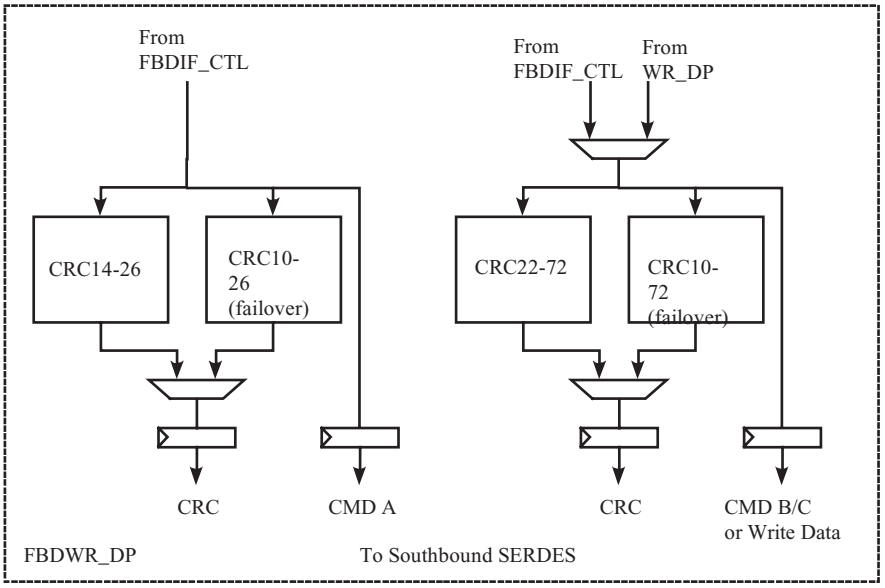
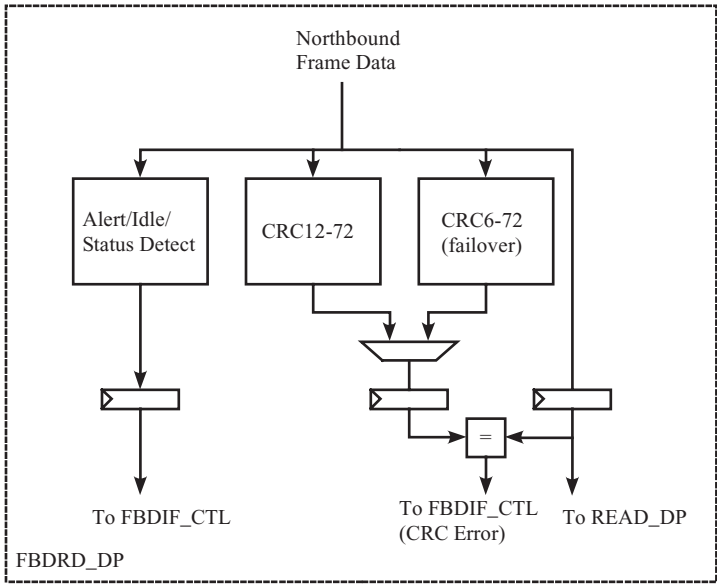


FIGURE 3-23 shows the FBD Read Datapath and the following two CRC generation blocks:

1. CRC[11:0] for read data of 14-bit mode NB frame, generated from 72-bits of data, compared to transmitted CRC.
2. CRC[5:0] for read data of 14-bit failover mode NB frame, generated from 72-bits of data, compared to transmitted CRC.

If there is a CRC error on a read frame, the FBDRD_DP signals the DRIF_CTL and FBDIF_CTL blocks for error reporting and to try to recover from the error.

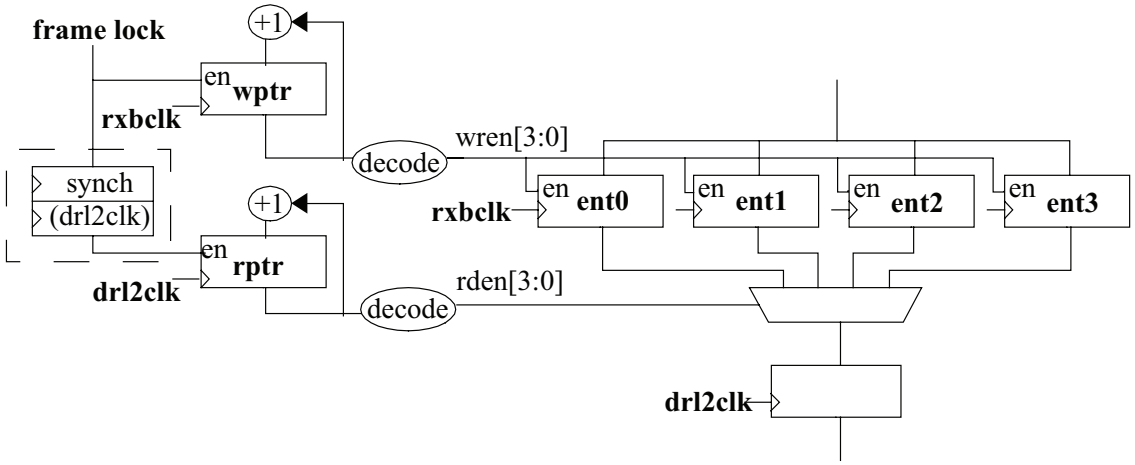
FIGURE 3-23 FBD Read Datapath



3.8.1.4 FSR to MCU Cross-Domain FIFO (FBD_DP)

The recovered clock from the SERDES runs at the same speed as the drl2clk; however, the two clocks are asynchronous. Therefore, an asynchronous FIFO is needed to pass data between these clock domains. A FIFO is required for each northbound bit lane since each generates its own clock. The FIFOs are implemented in flip-flops. The FIFOs are 12 bits wide and four entries deep. Once enabled, data will be written to the write port every cycle and read from the read port every cycle. The write pointer is in the recovered clock domain and is enabled once the MCU has seen 16 sync signals from the SERDES receiver. The read pointer is in the drl2clk domain and is enabled by the same write pointer enable signal after being sent through a synchronizer. After both the read and write pointers are enabled, data placed into the FIFO should be stable for approximately two clocks drl2clk's before being read out. As long as the phase difference between the two clocks does not drift by +- 1 period, no data will be lost. If the phase drift is greater that +-1 period, data may get dropped in which case the MCU will detect an error and may have to retrain the link to reestablish the cross-domain relationship.

FIGURE 3-24 FBD Cross Domain Logic

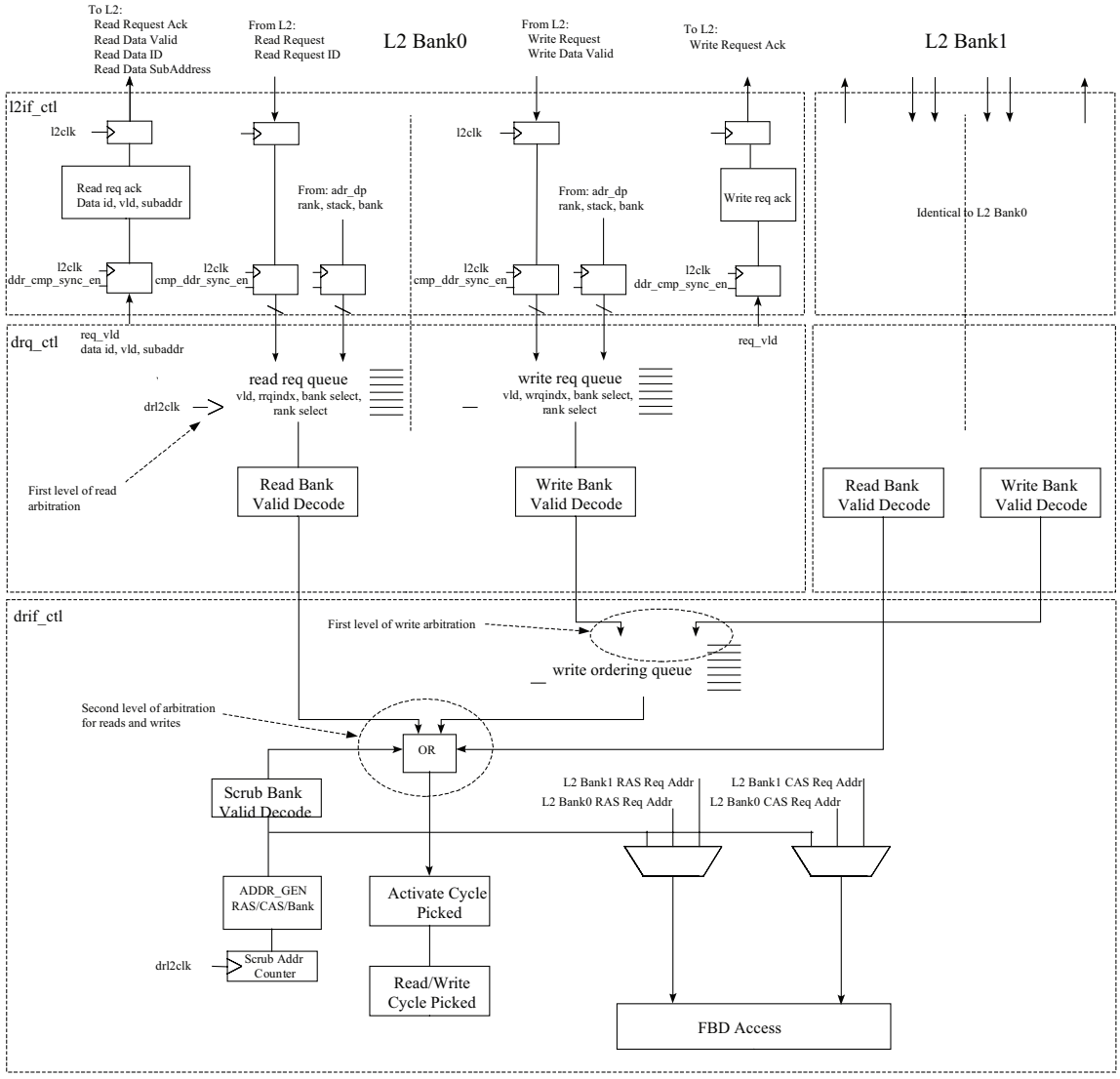


3.8.2 MCU Control Logic

The MCU control logic provides the interface between the two L2 banks and the external memory channel. The MCU control logic block diagram is shown in FIGURE 3-25:

Requests enter the L2IF_CTL blocks from the L2 cache banks. The requests are passed to the DRQ_CTL blocks with the cmp_ddr_sync_en signals where they are placed in the Read or Write Request Queues. Write arbitration occurs between the two Write Request Queues, and the requests are then placed in the Write ordering queue. Read request arbitration occurs between the requests in one Read Request Queue. The second level of arbitration selects a request from the Write Ordering Queue or one of the Read Request Queues to issue to the FBDs. Details of the control flow are given in the following sections.

FIGURE 3-25 MCU Control Logic Block Diagram



3.8.2.1 MCU - L2 Cache Interface Control (L2IF_CTL)

The L2 Cache Interface Control module handles incoming read and write requests from the L2 cache. It controls the synchronization of these requests to the drl2clk domain, the writing of data to the Write Data Queues, and the return of read data to the L2 cache or scrubbed data back to the SDRAMs. MCU-L2 Cache Interface discusses the protocol the MCU uses for communicating with the L2 cache.

When a read request comes in from the L2 cache, it is held in a staging register in the L2IF_CTL block and loaded with the signal l2if_cmp_ddr_sync_en. The acknowledge is always sent back one cycle after l2if_cmp_ddr_sync_en, and the request information is moved to the MCU Request Queue Control block at the rising edge of the next drl2clk.

Write requests are handled similarly except that the L2 cache does not limit the number of write requests it issues, so the MCU must handle this. The L2IF_CTL block receives information from the DRQ_CTL block telling how many Write Request Queue entries are free. If all eight entries are used, the ninth request is held in the L2IF_CTL block until one frees up, at which time the write request information is passed to the DRQ_CTL. Once the acknowledge is sent to the L2 cache, the L2 cache will start sending the write data 64 bits at a time, and the L2IF_CTL will control the loading of data into the Write Data Queue.

3.8.2.2 MCU Request Queue Control (DRQ_CTL)

Upon receiving a read or write request from an L2 bank, the MCU Request Queue Control (DRQ_CTL) logic generates the write enables to enqueue the incoming request in the read or write request address queue. The information required for request arbitration is stored in the read or write request queue within the DRQ_CTL block. Every drl2clk cycle, the arbitration logic must look at all entries in the read and write request queues in parallel to determine the next request to be scheduled. A request's entry in the read or write request queue will be invalidated upon the completion of the memory access.

The Read and Write Request Queues are implemented in registers as collapsing queues. The newest entry is placed at the tail of the queue; however, the oldest entry is not necessarily the first to be removed. The arbitration algorithm decides which entry to select, and when it is removed, the remaining queue entries which entered after it collapse to fill the empty entry.

Read Request Queue (RRQ)

The following information from the incoming L2 read request is stored in the 8-entry read request queue:

1. L2 read request valid - 1b

2. Index of DIMM address in the read request address queue - 3b
3. DIMM rank select - 2b
4. DIMM internal bank select - 3

This information is used by the arbitration logic for request scheduling.

Write Request Queue (WRQ)

The following information from the incoming L2 write request is stored in the 8-entry write request queue:

1. L2 write request valid - 1b
2. Index of DIMM address in the write request address and write data queues - 3b
3. DIMM rank select - 2b
4. DIMM bank select - 3b

This information is used by the arbitration logic for request scheduling.

3.8.2.3 Write Ordering Queue (WOQ)

The Write Ordering Queue controls the issuing of write data to the FBDs. The WOQ uses round-robin arbitration to choose between the two write request queues when selecting a request to issue. Within a WRQ, the WOQ looks for a transaction going to a different FBD than where the previous two transactions were issued. (This helps the MCU be able to schedule multiple write requests later.) If a transaction for a different FBD is not found, the first request in the WRQ is selected. The data corresponding to the selected request is sent to be buffered in the AMB and the request is placed in the WOQ. This ensures that write requests will be sent to the FBDs in the same order that the write data was sent.

When a write request is issued to an FBD, the read pointer for the WOQ is incremented; however, the information for the request remains in the WOQ data structure. Another pointer (the outstanding write transaction (OWT) pointer) points to the head of all requests that have been sent to the FBDs but have not been verified as having completed. When a write is issued, the MCU waits that time that a read would take. Once this time elapses, the OWT pointer is incremented; however, the transaction is not considered complete until an Idle Frame or a Status Frame has been seen after this point. A third pointer, the WOQ error pointer, points to the head of requests that should have completed but for which the MCU has not yet seen an Idle or Status Frame. If an error occurs, all writes from this WOQ error pointer to the tail of the WOQ will be reissued. If no error occurs, the WOQ error pointer is updated to the location of the OWT pointer once an Idle or Status Frame is received.

The WOQ has 16 entries and holds the same information as the WRQ; however, each entry has an additional bit that points to the source L2 bank.

3.8.2.4 MCU - DDR2 Interface Control (DRIF_CTL)

The MCU arbitrates among three memory access request sources - memory refresh requests, memory scrubbing requests, and L2 cache memory access requests. The priority among the three request types is as follows:

1. Memory refresh
2. Memory scrubbing requests
3. L2 cache read and write requests

Memory Refresh Request

DDR2 SDRAMs require a refresh of all rows within the memory every 7.8s. The MCU has a programmable refresh counter, clocked with drl2clk, to keep track of the refresh time interval. At every refresh interval, a memory refresh request is issued successively to each rank present.

With a drl2clk of 400 MHz, the programmable refresh counter value is calculated as follows:

refresh counter value = 7.8s/2.5ns = 3120 (0xC30)

Memory Scrubbing Request

At intervals defined by the DRAM Scrub Frequency Register, scrub read requests are issued to the SDRAMs. The purpose of these requests is to detect and correct transient memory errors. More detail on the scrubbing procedure is given in RAS Features.

First-level Write Request Arbitration

Because write data is buffered in the AMBs on the FBDs, it is advantageous to send the write data to the DIMMs early and arbitrate for the write requests later. When write requests are available in the write request queue, the write scheduler will issue the write data in frames to the FBDs whenever free time slots are available. When a write request is selected, its data is sent out to the FBDs, and the request is placed in the Write Ordering Queue for the second level of arbitration. When the write

commands are eventually issued, the ordering of write transactions to a given FBD must be maintained; however, write transactions to different FBDs can bypass each other.

First-level Read Arbitration

Each Request Queue Control block (DRQ_CTL) sends read memory requests to the DRIF_CTL arbiter, and the Write Ordering Queue within the DRIF_CTL provides write requests to the arbiter. The DRIF_CTL uses a first-come-first-served algorithm at the first level of arbitration for reads, selecting the oldest read request in the Read Request Queue whose bank is available.

Second-level Read and Write Arbitration

At the second level of arbitration, the DRQ_CTL does a round-robin selection of requests from the two DRQ_CTL read request queues. Reads have highest priority and will be scheduled when ready. A maximum of one read per frame can be scheduled. One or two write requests from the write ordering queue can be scheduled simultaneously with the read if they target different FBDs than the read. If there are no read requests ready, up to three write requests can be scheduled if they all target different FBDs.

Each read or write transaction must be issued over two consecutive cycles. On the first cycle, an Activate command is issued to activate a bank and row within an FBD. On the following cycle, the read or write command is issued with the bank and column addresses. The MCU uses the DDR2 SDRAM Posted CAS feature, so the read or write commands are delayed internal to the SDRAMs by the Additive Latency (AL) value programmed in the SDRAMs' extended mode register.

An Auto-Refresh command is issued when the MCU state machine transitions to the refresh state. Transactions to the rank being refreshed are blocked until the refresh completes.

Requests issued to the DIMMs are arbitrated every drl2clk cycle and are prioritized as follows, in order of descending priority:

Command slot A:

1. Scrub Read Activate requests
2. Read Activate requests from read request queues
3. Write Activate requests from write ordering queue

Command Slot B:

1. Auto refresh

2. Write data

3. Write Activate request

Command Slot C:

1. Power Down mode exit

2. Power Down mode enter

3. Write data

4. Write Activate request

Write Starvation Prevention

A write starvation counter ensures that writes do not get starved out. There is a separate write starvation counter in each DRQ_CTL block. Initially, reads have priority, and each write starvation counter is incremented whenever there are eight pending write requests total (between the DRQ and the WOQ). It is reset whenever there are less than eight pending writes or when a write request is issued from the WOQ. If either starvation counter reaches 64 (meaning there has been eight pending writes for 64 consecutive cycles), then starvation mode is entered. Once in starvation mode, write activation requests from the write ordering queue are given priority over read activate requests (i.e. the priorities of 2 and 3 for Command Slot A above are reversed). During starvation mode, the starvation counter that reached 64 is decremented each cycle, and once it reaches 0, starvation mode is exited, and reads are again given priority.

Scheduling Writes in Command Slots B and C

When there are no higher priority requests for command slots B and C, write requests can be sent out. These write requests can only be sent to DIMMs whose read-to-write delay has been satisfied, and they must be scheduled so that no write data collisions occur on the FBD data buses. Also, in order to simplify scheduling, these requests cannot be selected on the same cycle as a read or write in command slot A.

Read-after-Write Hazards

It is possible for the MCU to receive a read request to a physical address that matches the address of a write request pending in the write request queue. If this occurs, the MCU must ensure that the write completes first. (Whenever there is an address match between a read and a write, it is guaranteed that the write has preceded the read. Write-after-read hazards are handled by the L2 cache itself.)

When a read transaction wins arbitration, its address is compared against all of the valid entries in the write request queue for the L2 bank from which the read request came. If there is a match, the read request is not sent out, and the write request queue entry that matches the read is flagged. Writes are given priority until the flagged write request queue entry is issued to the FBDs. After the write request is issued, reads are given priority.

Scheduling State Machines

In a fully populated FBD channel, there can be eight dual rank dimms for a total of 16 ranks. Each rank supports DIMMs with up to eight banks. Thus, there can be 128 banks per channel.

There are 16 state machines to keep track of available banks. Each is dedicated to a set of DRAM banks depending on the configuration. The equations below show which state machine will be used for a given address:

4-bank DRAMs, double-sided DIMMs: state_machine = {dimm[0], rank, bank[1:0]};

4-bank DRAMs, single-sided DIMMs: state_machine = {dimm[1:0], bank[1:0]};

8-bank DRAMs, double-sided DIMMs: state_machine = {rank, bank[2:0]};

8-bank DRAMs, single-sided DIMMs: state_machine = {dimm[0], bank[2:0]};

ECC Error Handling

The DRIF_CTL block handles errors that have occurred. Information on the read request that caused the error is received from the RDPCTL_CTL block and stored in a register FIFO. The FIFO has eight entries to hold all outstanding reads that may generate ECC errors. Once an error is detected, the DRIF_CTL stops issuing requests, and after all outstanding requests have completed, it begins retrying each transaction from the error FIFO. Refer to ECC Error Handling for an explanation of the error handling scheme.

3.8.2.5 FBD Interface Control (FBDIC_CTL)

The FBD Interface Control block is responsible for FBD channel initialization, channel error detection, and frame encoding and decoding. All of the channel configuration registers also reside in this block.

At power-on, software is responsible for sequencing the FBD controller through the initialization sequence. The FBD channel initialization sequence is described in FBD Channel Initialization.

After initialization is complete and the AMBs are in the L0 state, software must program various registers in the AMBs using channel commands, and when it is ready to accept SDRAM commands, the FBDIC_CTL signals the DRIF_CTL. The FBDIC_CTL will encode the channel and SDRAM commands using the frame formats described in SDRAM Initialization using the CRC data provided by the FBD_DP block. This information is sent to the FBD SERDES IO block to be transmitted to the FBDs.

The Latency Queue (LATQ) informs the FBDIC_CTL when read data is returning. When a request is issued, a timestamp is placed in the LATQ (current time plus channel latency). When the current time matches the timestamp at the head of the LATQ, the read data (or other response, such as a Status Frame) is expected. If a CRC error occurs or an Alert or Idle Frame is detected, the DRIF_CTL is signaled so that it can retry the read. If an error is detected on the second read, the FBDIC_CTL module tries to recover with a Soft Channel Reset or a Fast Reset, if necessary. If neither of these revives the FBD channel, software is signalled to handle the recovery.

If a Status Frame shows an error has occurred or if an Alert Frame is received on the NB channel, this indicates an error has occurred on the SB channel. Again, a Soft Channel Reset, and possibly a Fast Reset, will be issued, after which software will signal if the condition persists.

3.8.2.6 MCU Read Datapath Control (RDPCTL_CTL)

The Read Data Path Control module controls the portion of the READDP_DP within the drl2clk domain and prepares the data valid and error signals to be returned to the L2 cache banks. Error logging is also performed in this block.

The following information is received from the DRIF_CTL block to keep track of the read requests outstanding to the FBDs:

- L2 bank - 1 bit
- Read/write - 1 bit
- Starting quadword - 2 bits
- Read request id - 3 bits
- Location in read or write request queue - 3 bits
- Out-of-bound address error on read - 1 bit

The information is stored in a FIFO implemented in registers. The FIFO depth is 16 and only holds outstanding reads. Reads are freed from this FIFO when a read transaction's data returns correctly.

If a CRC error occurs on the Southbound channel, all transactions in the RDPCTL_CTL FIFO must be retried in their original order after the channel is reset. If a CRC or ECC error occurs on the Northbound channel, only the transaction with an error must be retried.

If a CRC error occurs on the Southbound channel, all transactions in the RDPCTL_CTL FIFO must be retried in their original order after the channel is reset. If a CRC or ECC error occurs on the Northbound channel, only the transaction with an error must be retried.

3.8.2.7 MCU Read Data Control (RDATA_CTL)

The Read Data Control block sends the data valid, qword id and read request id signals to the L2 cache banks, and responds to the L2 cache dummy read requests. It also generates the address generation control signals for the ADRGEN_DP blocks in the Address Datapath block and acts as a bridge between the IO and MCU clock domains for CSR reads and writes.

3.8.3 Unit Control Block (UCB) Configuration Status Register (CSR) Interface

The Unit Control Block (UCB) provides the Configuration Status Register (CSR) interface to the MCU. The NCU communicates with the UCB module through a 4-bit bus. For register writes, the UCB assembles the 4-bit packets received into a 32-bit address and a 64-bit data word, and conversely for reads, breaks the 64-bit read data into 4-bit packets to send back to the NCU.

3.8.4 Interconnect Built-In Self Test (IBIST) Engine

The FBDIMM standard requires an IBIST engine within the MCU that will stress the FBDIMM channel electrical connections. When the FBDIMM channel initialization reaches the Testing stage.

The IBIST engine will take control of the channel after the TS1 header is issued.

The following registers are implemented by the MCU:

SBFIBPORTCTL: 0x84_0000_0E80

Since the MCU will always be a master on the southbound port, bit 1 and bits 6 through 23 of this register are not used by the MCU and will be read only. Bit [1] will be 1'b1, and bits [23:6] will be 18'h00000.

SBFIBPGCTL: 0x84_0000_0E84

SBFIBPATTBUF1: 0x84_0000_0E88

SBFIBTXMSK: 0x84_0000_0E8C

SBFIBTXSHFT: 0x84_0000_0E94

SBFIBPATTBUF2: 0x84_0000_0EA0

SBFIBPATT2EN: 0x84_0000_0EA4

SBFIBINIT: 0x84_0000_0EB0

SBIBISTMISC: 0x84_0000_0EB4

NBFIBPORTCTL: 0x84_0000_0EC0

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 1'b0, and bits [23:22] will be 2'h0.

NBFIBPGCTL: 0x84_0000_0EC4

NBFIBPATTBUF1: 0x84_0000_0EC8

NBFIBRXMSK: 0x84_0000_0ED0

NBFIBRXSHFT: 0x84_0000_0ED8

NBFIBRXLNERR: 0x84_0000_0EDC

NBFIBPATTBUF2: 0x84_0000_0EE0

NBFIBPATT2EN: 0x84_0000_0EE4

The following registers are not implemented:

SBFIBRXMSK

SBFIBRXSHFT

SBFIBRXLNERR

NBFIBTXMSK

NBFIBTXSHFT

NBFIBINIT

NBIBISTMISC

3.9 SDRAM Power Reduction and Reduced-Configuration Operating Modes

The SDRAMs in a system consume a large portion of the power budget and ways to limit the power consumption in certain configurations or at certain times. A single-channel mode is available that allows one DIMM per MCU channel, power throttling limits the number of SDRAM transactions over a period of time, and the SDRAMs can also be put in self refresh modes.

3.9.1 Single Channel Mode

Normally, the memory will be configured in dual-channel mode. In order to reduce system power, a single-channel mode has been added which supports one DIMM per channel. In this mode, 72 bits of data and ECC are driven externally per memory cycle. The burst length for this mode is eight to maintain the 64-byte cache line size per memory transaction.

To enable single channel mode, the single channel mode register, address 0x84_0000_0148 must be set to 1. Also, for proper operation, the Trrd (0x84_0000_0080) and Trc (0x84_0000_0080) must be increased by 2.

3.9.2 MCU Programmable Power Throttle

There are two registers per controller that control power throttling. The DRAM Open Bank Max Register designates the maximum number of DRAM bank openings that can occur in a time period. The time period is determined by the DRAM Programmable Time Counter Register whose value is a count of DRAM clock cycles. There is a counter that counts the number of DRAM banks that are opened. If this counter exceeds the maximum number of open banks, the DRAM controller is blocked from issuing anymore DRAM accesses until the counter is reset. A second counter counts DRAM clock cycles. When this counter is greater than or equal to the programmable time counter value, both this counter and the DRAM open bank counter are reset to zero.

The registers in the four controllers should be programmed to the same values. The chip will still operate correctly if they are programmed differently, but there may performance penalties (e.g. if one controller stops much earlier than the others) and power may not be as effectively controlled. Also, a mechanism is needed to ensure

that all of the controllers are working with in the same time window of DRAM clock cycles. This will ensure that all controllers stop and start at approximately the same time. When any of the Programmable Time Counter Registers is written, a reset signal will be sent to the other three controllers to reset their DRAM clock cycle counters.

3.9.3 SDRAM Self-Refresh Mode

The DDR2 SDRAMs support Self-Refresh mode allows the OpenSPARC T2 MCU to be reset without data loss in the SDRAMs.

For Self-Refresh mode, when the clock control unit signals the MCU to enter this mode, the MCU waits until all requests have completed and then issues a Self-Refresh Entry command to the FBDs. In order to leave this mode, the clock and other external control must be stable for at least one clock cycle. The MCU issues a Self-Refresh Exit command and waits 200 cycles before returning to normal operation.

3.9.4 FBD L0s State

Some AMBs provide a low-power state. When an AMB receives a Sync frame with the 'Enter L0s' bit set, it transitions to the L0s state for a time period determined by its L0s_Duration register. This register value is between 32 and 42, and must be less than the minimum interval between Sync frames defined in the AMB's Sync Train Interval register. Once the timer for the L0s state expires, the AMB transitions back to the L0 state. After exiting the L0s state, the first command that the host must issue is another Sync frame in order to ensure that the AMB clocks remain locked. With this Sync frame, it is also possible to put the AMBs back into the L0s state for another low-power interval.

The MCU must be programmed to decide when to transition to this state. This is enabled by setting bit [6] of the L0s Duration Register. When this mode is enabled and there are no pending transactions when a Sync frame is being sent out, the el0s bit will be set in the Sync frame.

There is Thermal_Trip information returned in NB status frames which indicates that a thermal threshold has been exceeded and that power throttling may be required. This information is held in the Thermal Trip Status Register that software can check to decide when to enable L0s mode.

3.9.5 Power Down Mode

Power Down mode is a low power mode for the DRAMs. The MCU can optionally use this mode. When a transaction enters the MCU, a counter for the destination DIMM is incremented. When the transaction completes, DIMM counter is decremented. When any counter goes from 0 to 1, an Exit Power Down command will be sent to the corresponding DIMM. When a counter goes from 1 to 0, an Enter Power Down command will be sent to the DIMM.

3.9.6 Partial Bank Mode

Partial bank mode is a mode where less than eight L2 banks are used in the system. In four-bank mode, two MCUs are used and in two-bank mode, one MCU is used. In these modes, the addressing to the DIMMs is changed. In four-bank mode, the MCU left-shifts address bits [39:7] by one bit before applying the normal address decoding. In two-bank mode, the MCU left-shifts address bits [39:7] by two bits before applying the normal address decoding.

When in partial bank mode, the MCU will detect out-of-bound errors on a smaller address range for a given configuration as compared to a system with all MCUs enabled. In four-bank mode, the system will have half of the memory of a full system with the same configuration and one-fourth the memory when in two-bank mode. The MCUs must be configured with twice the memory in four-bank mode and four times the memory in two-bank mode to provide the same address space as a full system. If the full system uses the maximum memory configuration, then the tests must reside in the lower half or quarter of memory in order to execute properly in a four-bank or two-bank configured system, respectively.

3.10 RAS Features

3.10.1 SDRAM 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 in this block for each 128 bits of data.

3.10.2 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, at the end of the time interval defined by the DRAM Scrub Frequency Register, 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. 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, a single-bit and double-bit error is flagged in the DRAM Error Status Register as well as being signalled to the L2 cache; then the MCU generates additional requests to the SDRAMs to collect more information on the error which is 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.

3.10.3 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 0x8221 which, when encountered on a read, will most likely indicate poisoned data. The L2 cache asserts l2b_mcu_data_mecc which causes the MCU to corrupt the ECC for the corresponding 64-bit data word.

3.10.4 ECC Error Handling

When an error occurs on a scrub read or an L2 cache read request, the MCU will flag the error and then try to determine if the error is a hard error or a transient error. After the error occurs, the MCU will first perform another Read and log its ECC status. If the second read does not have an uncorrectable error, the corrected read data is written back to the SDRAM, and a third read is issued, and its status is also logged. Only the status from the first read will be sent to the L2 cache bank. One of the cores must perform a register read to check the status of the subsequent reads.

3.10.5 FBD Channel Errors

There are several ways that errors may be detected on the SB or NB channels. Alerts and Status Frames show when CRC errors have occurred on the SB channel. CRC errors on data frames and corrupted Idle or Status Frames show errors on the NB channel. When a channel error occurs on the SB channel, all transactions not guaranteed to have completed before the problem was detected must be reissued. When an error occurs on the NB channel, only the transaction with an error must be reissued.

1. Alert Frame: The MCU will stop issuing transactions and will issue a Soft Channel Reset (SCR) frame to attempt to reset the state of the AMBs and try to determine which AMB has detected an error. If errors persist, the MCU will issue a Fast Reset. If errors still persist after the Fast Reset, the MCU will log an Unrecoverable error in the MCU ESR, log an Alert Frame error in the MCU Syndrome Register and assert mcu_l2t0_scb_mecc_err to the L2. If at any point in the error processing the MCU is able to recover from the error condition, the MCU will instead assert mcu_l2t0_scb_secc_err to the L2 and set the Recoverable error bit in the ESR. Any outstanding reads or subsequent reads must still be returned to the L2 after the error processing is completed. If the channels are not working, these reads will be seen as Unrecoverable CRC errors.

2. Status Frame with Alert asserted: This indicates that the asserting AMB has detected an error on the SB channel. If this is the only error detected, the MCU will wait for the next Status Frame. If no other error occurs before or in the next Status Frame, the MCU will flag an Alert Asserted error in the syndrome register, set the Recoverable error bit in the MCU ESR, and assert mcu_l2t0_scb_secc_err to the L2. If any other error type occurs before or in the next Status Frame, processing proceeds as for that error condition.

3. Status Frame Parity Error: If there is a parity error in a Status Frame, the MCU will wait for the next Status frame. If the error persists, the MCU will attempt a Soft Channel Reset and if necessary, a Fast Reset. If the error persists the MCU will flag a Status Frame Error in the syndrome register, log a Status Parity error in the MCU Syndrome Register and assert mcu_l2t0_scb_mecc_err to the L2. If at any point in the error processing the MCU is able to recover from the error condition, the MCU will instead assert mcu_l2t0_scb_secc_err to the L2 and set the Recoverable error bit in the ESR. Any outstanding reads or subsequent reads must still be returned to the L2 after the error processing is completed. If the channels are not working, these reads will be seen as Unrecoverable CRC errors.

4. CRC Error on Read Data: The MCU will discard the data and retry the transaction. If it fails again, the MCU will issue a Fast Reset of the FBD channel. If the error persists after the Fast Reset, the MCU will flag an Unrecoverable error in the ESR, flag a CRC error in the Syndrome register and assert mcu_l2t_mecc_err_r3 to the L2 along with the bad data. If there is no CRC error on any of the retries, the MCU will flag a Recoverable Error in the ESR and assert mcu_l2t0_scb_secc_err once to the L2 and then send the correct read data.

When a Soft Channel Reset command is issued to the FBDIMMs, it will be followed by a CKE command (to enable all CKEs) and a precharge all command to put the DRAMs in a legal operating state. In order for the CKE commands to be issued correctly, the FBD Per Rank CKE Register must be set correctly and the CKE bit in the DIMM Initialization Register must be set.

When two channels are operating in lock-step, the MCU will perform error handling as if the same error occurred on both channels.

3.10.6 Interrupts

The MCU has two interrupt types that it can send to the NCU based on certain MCU errors. The error types are Correctable ECC Error Count and Recoverable FBD Channel Error Count. When one of these errors is generated, the MCU will send a single cycle pulse to the NCU in the iol2clk domain. Either of FBD errors will also generate a syndrome which is stored in the MCU Syndrome register if no other FBD error is pending.

The NCU also has a mechanism for generating these error types and an Unrecoverable FBD Channel Error. When one of these signals is asserted, the MCU will inject that type of error within its logic to verify that the error detection and reporting logic is operating correctly.

3.10.6.1 Correctable ECC Error Count Interrupt

When the MCU Error Count Register reaches zero, the MCU will generate an interrupt on mcu_ncu_ecc, asserting it for one iol2clk cycle. No syndrome is reported, and no more of this type of interrupt will be generated until the software writes to the Error Count Register to enable interrupt generation.

If ncu_mcu_ecci is asserted, the MCU will inject a single correctable error on the lowest ECC bit on the next read packet. If the Error Count Register is already zero, nothing happens. Otherwise, the error count will be decrement by one. If the MCU Error Count Register goes to zero, then the Correctable ECC Error Interrupt will be generated. This ECC error should also be reported to the L2 regardless of the value of the Error Count Register.

3.10.6.2 Recoverable FBD Channel Error Count Interrupt

A Recoverable FBD Channel Error Interrupt will be generated whenever the Recoverable FBD Channel Error Count Register reaches zero. The MCU will assert mcu_ncu_fbr to the NCU for one iol2clk cycle. Once the count value is zero, no more interrupts of this type will be generated until software writes a non-zero value to the count register.

If ncu_mcu_fbri is asserted from the NCU to the MCU, the MCU will inject an error within the FBD channel. The source of the error will be determined by the Injected Error Source Register. The MCU will handle the error as if it had actually occurred in hardware. If the Recoverable FBD Channel Error Count Register reaches zero, then mcu_ncu_fbr will be asserted to the NCU.

3.10.6.3 Unrecoverable FBD Channel Error Interrupt Injection

There is no error generated to the NCU for Unrecoverable FBD Channel Errors. These types of errors will only be indicated through the L2 cache.

If ncu_mcu_fbui is asserted from the NCU to the MCU, the MCU will inject an error within the FBD channel. The source of the error will be determined by the Injected Error Source Register. The MCU will handle the error as if it had actually occurred in hardware.

3.11 Test Features

3.11.1 DFT Features

3.11.1.1 Debug Reset

During Debug Reset, the MCU needs to keep the FBD links active so that runs can be reproducible. By keeping them active, the channel latencies and the SERDES to MCU asynchronous crossings will not change between test runs.

In order to achieve this, a small part of the MCU must be protected from warm reset, and the clock to it needs to remain active during reset. For each MCU, there is an additional clock stop signal, tcu_mcu*_fbd_clk_stop, and a common test mode signal, tcu_mcu_testmode. The subblock mcu_fdout_ctl contains the logic that remains active during the Debug Reset. Its clock is taken directly from dr_gclk, not

from the output of a cluster header. The tcu_mcu*_fbd_clk_stop signal goes directly to the L1 headers in mcu_fdout_ctl, and these clock stop signals are not asserted during the Debug Reset. The tcu_mcu_testmode signal is used to qualify the scan and reset signals. tcu_scan_en, tcu_aclk and tcu_bclk are ANDed with tcu_mcu_testmode in mcu_fdout_ctl. tcu_mcu_testmode is asserted for Power-On Reset and for a normal Warm Reset, but not for Debug Reset. tcu_mcu_testmode is also used to control a mux which bypasses the protected flops on the scan chain when it is not asserted but includes them when not asserted.

3.11.2 Deterministic Test Mode (DTM)

Deterministic Test Mode is a mode in which all of the IO units operate at the same speed so that test runs may be reproducible on a tester. The mode is entered when the ccu_serdes_dtm is set to 1'b1. In this mode different data will be sent by the MCU to the debug bus. Also, since the tester will be "sourcing" transactions, the MCU will act as a slave device during channel initialization and must respond to transactions on the NB channel in order to enter a predictable state.

3.11.2.1 Debug Signals

For DTM, the MCU will provide CRC data from the southbound FBD frames to the Debug unit. Each frame contains 22 bits of CRC, so since there are two channels per MCU, there will be 44 bits of CRC data per MCU. The MCU will take the CRC from the two channels and XOR them; however, either channel's CRC can be masked off before the XOR by setting the appropriate bit in the Debug Trigger Enable Register. This data will then be sent to the UCB module to be multiplexed with the normal-mode debug signals. One additional bit needs to be sent to the Debug unit since the current debug bus signal count is 21 bits. The MCU will use the ccu_serdes_dtm signal to select the CRC data for sending to the Debug unit.

These debug signals will be taken from the drl2clk domain to the iol2clk domain; however, they will not need to be synchronized since they will only be used for DTM where the drl2clk and iol2clk will be driven from the same source.

3.11.2.2 Initialization for Testing

When in DTM mode, the northbound FBD channel still needs to be initialized by northbound TS0 patterns. Therefore, the tester must send enough TS0 patterns to achieve bit lock, frame lock and lane deskew within the MCU. After these are achieved, the tester must then cause the MCU to transition to the L0 state in order to enable southbound transactions and to enable recognition of northbound transactions.

After reset, the MCU's FBD initialization state machine will be in the Disable state, and the DTM state machine will begin in the IDLE state. The MCU's initialization state machine must be in the Disable state before DTM testing begins. The tester will send TS0 patterns to train the northbound link. The DTM state machine will transition to the TS0 state once it sees the TS0 patterns on the NB channel. This will also cause the MCU state machine to transition to the TS0 state, and the MCU will begin sending southbound TS0 patterns. Once in the TS0 state, the DTM state machine will wait until it sees at least four northbound frames containing all 0's. At this point, the MCU's unitization state machine will be transitioned to the L0 state, and the DTM state machine will return to IDLE. Once the MCU enters the L0 state, operation will proceed as in normal system mode.

Since the Polling state will be bypassed during the initialization for DTM, the channel latency register must be programmed to match the channel latency used for generating the test vectors.

In order to achieve operating rates that the tester can support, the RXTX_RATE field of the SERDES Configuration Bus Register must be set to Half Rate (2'b01) or Quarter Rate (2'b10) as required. The required link rates for supported operating frequencies are given in SERDES Deterministic Test Mode (DTM).

3.11.3 SERDES Blunt-End Loopback

SERDES Blunt-End Loopback within the MCU is controlled by the Loopback Mode Control Register. This register controls both links within an MCU. When bit 1 is set, data received on a northbound FBD channel will be placed on the corresponding southbound channel. Since there are 14 northbound lanes and only 10 southbound lanes, bit 0 selects which northbound lanes map to which southbound lanes. If bit 0 is 1'b0, then northbound lanes 0 through 9 are mapped to the southbound lanes. If bit 0 is 1'b1, then northbound lanes four through 13 are mapped to the southbound lanes.

Since the data on the northbound channel must be synchronized from the recovered clock domain into the MCU's drl2clk domain, TS0 patterns must be sent on the northbound channel in order to achieve frame lock. Once frame lock is achieved, the cross domain FIFO will be enabled, and data will be forwarded from the northbound channel to the southbound channel.

This increase in size would change the MCU's Y dimension to about 1750?m due to the addition of the FBDWR_DP and FBDRD_DP. (The size of the MCU as of June 10, 2004 is 1744?m x 847?m.) The X dimension would have to be reduced and the aspect ratios of the control blocks would have to change in order to use up any whitespace created.

A break down of the area changes per block is given in the following sections.

3.12 MCU Level I/O

TABLE 3-22 MCU Level I/O

Signal Name	I/O	Description
Clocks, Reset, Etc.		
iol2clk	I	System bus clock
drl2clk	I	DRAM clock
l2clk	I	CPU domain clock
mcu_ce	I	DRAM module clock enable
ccu_mcu_ddr_cmp_sync_en	I	Dram to Cmp clock synchronization
ccu_mcu_cmp_ddr_sync_en	I	Cmp to Dram clock synchronization
ccu_mcu_io_cmp_sync_en	I	Sys to Cmp clock synchronization
ccu_mcu_cmp_io_sync_en	I	Cmp to Sys clock synchronization
clspine_mcu_selfrsh	I	Enter hardware self-refresh mode
rst_por_	I	Power-on reset signal
rst_wmr_	I	Warm reset signal
mcu_pt_sync_in0 mcu_pt_sync_in1 mcu_pt_sync_in2	I	Incoming power throttling counter synchronizing signals
mcu_pt_sync_out	O	Outgoing power throttling counter synchronizing signal
mcu_id[1:0]	I	MCU ID for error reporting
mcu_clk_en	I	Clock enable to synchronize MCU clock domain to external DRAM clock
Test		
scan_in	I	Scan in
scan_out	O	Scan out
tcu_aclk	I	
tcu_bclk	I	
tcu_soc_cmp_clk_stop	I	Clock stop signal for l2clk domain
tcu_soc5dr_clk_stop	I	Clock stop signal for drl2clk domain
tcu_soc6io_clk_stop	I	Clock stop signal for iol2clk domain
tcu_pce_ov	I	Clock enable override signal

TABLE 3-22 MCU Level I/O *(Continued)*

Signal Name	I/O	Description
tcu_dectest	I	
tcu_scan_en	I	Scan enable
tcu_se_scancollar_in	I	Scan enable for memory input flops
tcu_se_scancollar_out	I	Scan enable for memory output flops
tcu_array_wr_inhibit	I	Inhibit memory array updates
tcu_array_bypass	I	Bypass memory array
tcu_mbist_bisi_en	I	Enable MBIST engine
tcu_mcu_mbist_start	I	Start MBIST sequence
mcu_tcu_mbist_done	O	MBIST done
mcu_tcu_mbist_fail	O	MBIST fail
tcu_mcu_mbist_scan_in	I	MBIST module scan in
mcu_tcu_mbist_scan_out	O	MBIST module scan
Debug		
mcu_dbg1_rd_req_in_0[3:0]	O	Read request received from L2 bank 0
mcu_dbg1_rd_req_in_1[3:0]	O	Read request received from L2 bank 1
mcu_dbg1_rd_req_out[4:0]	O	Read data returned to L2 bank
mcu_dbg1_wr_req_in_0	O	Write request received from L2 bank 0
mcu_dbg1_wr_req_in_1	O	Write request received from L2 bank 1
mcu_dbg1_wr_req_out[1:0]	O	Number of writes retired
mcu_dbg1_mecc_err	O	Multiple nibble ECC error
mcu_dbg1_secc_err	O	Single nibble ECC error
mcu_dbg1_fbd_err	O	FBD channel error
mcu_dbg1_err_mode	O	MCU in error processing mode
mcu_dbg1_err_event	O	Debug error event when debug trigger is enabled
NCU Interface		
ncu_mcu_data[3:0]	I	NCU to MCU module CSR bus
ncu_mcu_stall	I	Stall signal from NCU for outgoing transactions
ncu_mcu_vld	I	Incoming CSR data valid
ncu_mcu_ecci	I	Inject Correctable Error Count
ncu_mcu_fbri	I	Inject FBDIMM Channel Recoverable Error

TABLE 3-22 MCU Level I/O *(Continued)*

Signal Name	I/O	Description
ncu_mcu_fbui	I	Inject FBDIMM Channel Unrecoverable Error
mcu_ncu_data[3:0]	O	MCU module to NCU CSR bus
mcu_ncu_stall	O	Stall incoming transactions
mcu_ncu_vld	O	Outgoing CSR data valid
mcu_ncu_ecc	O	Correctable Error Count Interrupt
mcu_ncu_fbr	O	FBDIMM Channel Recoverable Error Interrupt
ncu_mcu_pm	I	Enables partial-bank mode
ncu_mcu_ba01	I	L2 banks 0 and 1 are enabled in partial-bank mode
ncu_mcu_ba23	I	L2 banks 2 and 3 are enabled in partial-bank mode
ncu_mcu_ba45	I	L2 banks 4 and 5 are enabled in partial-bank mode
ncu_mcu_ba67	I	L2 banks 6 and 7 are enabled in partial-bank mode
MCU-L2 Interface		
l2b0_mcu_data_mecc_r5 l2b1_mcu_data_mecc_r5	I	Signal to inject ECC errors in write data
l2b0_mcu_data_vld_r5 l2b1_mcu_data_vld_r5	I	Data valid signal from L2 cache
l2b0_mcu_wr_data_r5[63:0] l2b1_mcu_wr_data_r5[63:0]	I	Data from L2 cache
l2t0_mcu_addr_39to9[39:7] l2t0_mcu_addr_5 l2t1_mcu_addr_39to9[39:7] l2t1_mcu_addr_5	I	L2 cache transaction address
l2t0_mcu_rd_dummy_req l2t1_mcu_rd_dummy_req	I	Dummy read request from L2
l2t0_mcu_rd_req l2t1_mcu_rd_req	I	L2 cache read request signal
l2t0_mcu_rd_req_id[2:0] l2t1_mcu_rd_req_id[2:0]	I	L2 cache read request ID
l2t0_mcu_wr_req l2t1_mcu_wr_req	I	L2 cache write request signal
mcu_l2t0_data_vld_r0 mcu_l2t1_data_vld_r0	O	L2 cache read data valid signal
mcu_l2t0_rd_ack mcu_l2t1_rd_ack	O	Read request acknowledge signal to L2 cache
mcu_l2t0_scb_mecc_err mcu_l2t1_scb_mecc_err	O	MCU scrub multiple ECC error indication to L2 cache

TABLE 3-22 MCU Level I/O *(Continued)*

Signal Name	I/O	Description
mcu_l2t0_scb_secc_err mcu_l2t1_scb_secc_err	O	MCU scrub single ECC error indication to L2 cache
mcu_l2t0_wr_ack mcu_l2t1_wr_ack	O	Write request acknowledge signal to L2 cache
mcu_l2t0_wr_addr_err mcu_l2t1_wr_addr_err	O	Write address error signal to L2 cache
mcu_l2t0_qword_id[1:0] mcu_l2t1_qword_id[1:0]	O	Quadword of data being returned from MCU
mcu_l2t0_mecc_err_r3 mcu_l2t1_mecc_err_r3	O	MCU multiple ECC error indication to L2 cache
mcu_l2t0_rd_req_id_r0[2:0] mcu_l2t1_rd_req_id_r0[2:0]	O	Read request ID for L2 cache read data
mcu_l2t0_secc_err_r3 mcu_l2t1_secc_err_r3	O	DRAM single ECC error indication to L2 cache
mcu_l2b_data_r3[127:0]	O	Read data to L2 cache
mcu_l2b_ecc_r3[27:0]	O	ECC data for read data to L2 cache
MCU-FBD IO Interface		
mcu_fsr0_data[119:0] mcu_fsr1_data[119:0]	O	Southbound FBD Channel Data
mcu_fsr0_cfgpll_enpll mcu_fsr1_cfgpll_enpll	O	Enable PLLs for FBD Channels
mcu_fsr01_cfgpll_lb[1:0]	O	PLL Loopback for Channels 0 and 1
mcu_fsr01_cfgpll_mpy[3:0]	O	PLL Multiplier for Channels 0 and 1
mcu_fsr0_cfgrx_enrx mcu_fsr1_cfgrx_enrx	O	Enable SERDES receivers
mcu_fsr0_cfgrx_align mcu_fsr1_cfgrx_align	O	Enable Alignment detection for FBD SERDES
mcu_fsr0_cfgrx_los[1:0] mcu_fsr1_cfgrx_los[1:0]	O	Enable Loss-of-Signal (Electrical Idle) detection
mcu_fsr0_cfgrx_invpair[13:0] mcu_fsr0_cfgrx_invpair[13:0]	O	Invert RXP and RXN per bit
mcu_fsr01_cfgrx_eq[3:0]	O	Enable and configure adaptive equalizer
mcu_fsr01_cfgrx_cdr[2:0]	O	Configure clock/data recovery algorithm
mcu_fsr01_cfgrx_term[2:0]	O	Set input termination
mcu_fsr0_cfgtx_entx mcu_fsr1_cfgtx_entx	O	Enable SERDES transmitters

TABLE 3-22 MCU Level I/O *(Continued)*

Signal Name	I/O	Description
mcu_fsr0_cfgtx_enidl mcu_fsr1_cfgtx_enidl	O	Enable Electrical Idle on Transmitter
mcu_fsr0_cfgtx_invpair[9:0] mcu_fsr1_cfgtx_invpair[9:0]	O	Invert TXP and TXN per bit
mcu_fsr01_cfgtx_enftp	O	Enable fixed phose on TXBCLKIN
mcu_fsr01_cfgtx_de[3:0]	O	Set transmitter output de-emphasis
mcu_fsr01_cfgtx_swing[2:0]	O	Set transmitter output swing
mcu_fsr01_cfgtx_cm	O	Adjust common mode.
fsr0_mcu_rxbclk[13:0] fsr1_mcu_rxbclk[13:0]	I	Clocks for Northbound FBD Channels
fsr0_mcu_data[167:0] fsr1_mcu_data[167:0]	I	Northbound FBD Channel Data
fsr0_mcu_stspll_lock fsr1_mcu_stspll_lock	I	SERDES PLLs are locked
fsr0_mcu_stsrx_sync[13:0] fsr1_mcu_stsrx_sync[13:0]	I	Header alignment signal from SERDES Receivers
fsr0_mcu_stsrx_losdtct[13:0] fsr1_mcu_stsrx_losdtct[13:0]	I	Electrical Idle signal from SERDES Receivers

3.13 MCU Registers

MCU register definitions are detailed in the *OpenSPARC T2 Programmer's Reference Manual*. This document will only provide a list of registers.

3.13.1 Control and Status Registers (CSR)

TABLE 3-23 Control and Status Registers

Register Offset	Register Name
0x00000000	CAS Address Width Register
0x00000008	RAS Address Width Register
0x00000010	CAS Latency Register
0x00000018	Scrub Frequency Register
0x00000020	Refresh Frequency Register
0x00000038	Refresh Counter Register
0x00000040	Scrub Enable Register
0x00000080	RAS to RAS Different Bank Delay Register
0x00000088	RAS to RAS Same Bank Delay Register
0x00000090	RAS to CAS Delay Register
0x00000098	Write to Read CAS Delay Register
0x000000A0	Read to Write CAS Delay Register
0x000000A8	Internal Read to Precharge Delay Register
0x000000B0	Active to Precharge Delay Register
0x000000B8	Precharge Command Period Register
0x000000C0	Write Recovery Period Register
0x000000C8	Auto refresh to Active Period Register
0x000000D0	Mode Register Set Command Period Register
0x000000E0	Internal Write to Read Command Delay Register
0x000000E8	Precharge Wait Register During Power Up
0x00000108	DIMM Stacked Register
0x00000110	Extended Mode 2 Register
0x00000118	Extended Mode 1 Register
0x00000120	Extended Mode 3 Register
0x00000128	8 Bank Mode Register
0x00000138	Branch Disabled Register
0x00000140	Select Low Order Address Bits Register
0x000001A0	DIMM Initialization Register

TABLE 3-23 Control and Status Registers *(Continued)*

Register Offset	Register Name
0x00000208	Mode Register Write Status Register
0x00000210	Initialization Status Register
0x00000218	DIMMs Present Register
0x00000220	Fail-Over Status Register
0x00000228	Fail-Over Mask Register

3.13.1.1 Changes to DIMM Initialization Register- 0x84_0000_01A0

The DIMM Initialization Register for OpenSPARC T2 has the following format:

TABLE 3-24 DRAM Initialization Register

Field	Bit position	Initial value	R/W	Description
RSVD	[62:2]	0x0	RO	Reserved
CKE	[1]	0x0	RW	Enabled CKE to DIMMs
INIT	[0]	0x1	RW	Set to 1 during software initialization of DRAMs; cleared by software when done.

3.13.1.2 Single Channel Mode Register - 0x84_0000_0148

TABLE 3-25 Single Channel Mode Register

Field	Bit position	Initial value	R/W	Description
RSVD	[62:1]	0x0	RO	Reserved
MODE	[0]	0x0	RW	Enable use of one FBD channel for memory transactions. Burst length becomes 8.

3.13.1.3 Four Activate Window Register

TABLE 3-26 Four Activate Window Register

Field	Bit position	Initial value	R/W	Description
RSVD	[62:5]	0x0	RO	Reserved
MODE	[4:0]	0xA	RW	tFAW. Number of cycles in which activate commands may be issued to a DIMM. Preserved on warm reset.

3.13.2 Error Registers

TABLE 3-27 Error Registers

Register Offset	Register Name
0x00000280	Error Status Register
0x00000288	Error Address Register
0x00000290	Error Injection Register
0x00000298	Error Counter Register
0x000002A0	Error Location Register

3.13.2.1 Changes to Error Status Register - 0x84_0000_0280

The MCU Error Status Register has three additional bits:

TABLE 3-28 MCU Error Status Register

Field	Bit Position	Initial Value	R/W	Description
MEB	[56]	0x0	R/W1C	Multiple Out-of-Bound Errors
FBU	[55]	0x0	R/W1C	FBDIMM Channel Unrecoverable Error
FBR	[54]	0x0	R/W1C	FBDIMM Channel Recoverable Error

3.13.2.2 Error Retry Register - 0x84_0000_02a8

TABLE 3-29 Error Entry Register

Field	Bit Position	Initial Value	R/W	Description
VALID	[63]	0x0	RW	Error Retry Register is valid
RSVD	[62:50]	0x0	RO	Reserved
SYNDROME2	[49:34]	0x0	RW	Syndrome from second retry read
TYPE2	[33:32]	0x0	RW	Result of second retry read
RSVD	[31:18]	0x0	RO	Reserved
SYNDROME1	[17:2]	0x0	RW	Syndrome from second retry read
TYPE1	[1:0]	0x0	RW	Result of second retry read: 00: No read 01: No error 10: Correctable Error 11: Uncorrectable Error

3.13.3 Power Management Registers

TABLE 3-30 Power Management Registers

Register Offset	Register Name
0x00000028	Open Bank Max Register
0x00000048	Programmable Time Counter Register

3.13.3.1 Power Down Mode Register - 0x84_0000_0238

This register enables the use of Power Down mode for power savings. When enabled, an FBD will be placed in Power Down mode when there are no pending or outstanding transactions to that FBD.

TABLE 3-31 Power Down Mode Register

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:1]	0x0	RO	Reserved
ENABLE	[0]	0x0	RW	Enable use of Power Down mode if 1.

3.13.4 Performance Registers

TABLE 3-32 Performance Registers

Register Offset	Register Name
0x00000400	Performance Control Register
0x00000408	Performance Counter Register

3.13.5 Changes to Debug Trigger Enable Register

TABLE 3-33 Debug Trigger Enable Register

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:6]	0x0	RO	Reserved
DTM_ATSPEED	[5]	0x0	RW	If set, Debug bus sends normal mode data in DTM mode
DTM_MASK1	[4]	0x0	RW	If set, mask off CRC data from Channel 1 going to Debug bus
DTM_MASK0	[3]	0x0	RW	If set, mask off CRC data from Channel 0 going to Debug bus
DBG_EN	[2]	0x0	RW	Enable error events to Debug unit
MASK_ERR	[1]	0x0	RW	Mask LFSR related errors on NB FBD links
KP_LNK_UP	[0]	0x0	RW	Keep FBD links up during Warm Reset

3.13.6 State Registers for FBD Branch

FBD controller register address space will be a subset of the MCU's address space, differentiated by some upper bits.

3.13.6.1 Channel State Register - 0x84_0000_0800

TABLE 3-34 Channel State Register

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:8]	0x0	RO	Reserved
MDISABLE	[7]	0x0	RW	Disable AMB data merging for TS2 patterns
AMBID	[6:3]	0x0	RW	Target AMB for training sequences
STATE	[2:0]	0x0	RW	State in initialization sequence 0x0 = Disable, 0x1 = Calibrate, 0x2 = Training, 0x3 = Testing, 0x4 = Polling, 0x5 = Config, 0x6 = L0

3.13.6.2 Fast Reset Flag - 0x84_0000_0808

TABLE 3-35 Fast Reset Flag

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:4]	0x0	RO	Reserved
SYNC_IER	[3]	0x0	RW	Enables use of IER bit in Sync command. IER will be issued in last Sync frame before a Channel Reset.
SYNC_R	[2:1]	0x0	RW	Indicates which status register will be received from AMBs
FASTRESET	[0]	0x0	RW	Causes MCU to enter use Fast Reset sequence for channel initialization

3.13.6.3 Channel Reset (Initialization) Flag - 0x84_0000_0810

TABLE 3-36 Channel Reset (Initialization) Flag

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
FBDINITERR	[1]	0x0	RW	Set to 1 if an error occurred during the FBD Channel initialization.
FBDINIT	[0]	0x0	RW	Causes FBD Channel to be initialized. Uses fast reset sequence if Fast Reset Flag is set, otherwise performs full initialization (including Calibration). Reset to 0 when initialization is complete.

3.13.6.4 TS1 Southbound to Northbound Mapping Register - 0x84_0000_0818

TABLE 3-37 TS1 Southbound to Northbound Mapping Register

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:4]	0x0	RO	Reserved
IBRX_CHNL	[3]	0x0	RW	Selects which NB channel will be checked by the IBIST Receive engine
MAPPING	[2:0]	0x0	RW	Determines how targeted AMB maps data from SB bit lanes to NB bit lanes

3.13.6.5 TS1 Test Parameter Register - 0x84_0000_0820

TABLE 3-38 TS1 Test Parameter Register

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:24]	0x0	RO	Reserved
PARAM	[23:0]	0x0	RW	AMB test parameters for TS1 sequence

3.13.6.6 TS3 Failover Configuration Register - 0x84_0000_0828

TABLE 3-39 TS3 Failover Configuration Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:16]	0x0	RO	Reserved
SBCONFIG1	[15:12]	0xF	RW	Indicates which southbound lanes for channel 1 will be used.
NBCONFIG1	[11:8]	0xF	RW	Indicates which northbound lanes for channel 1 will be used.
SBCONFIG0	[7:4]	0xF	RW	Indicates which southbound lanes for channel 0 will be used.
NBCONFIG0	[3:0]	0xF	RW	Indicates which northbound lanes for channel 0 will be used.

3.13.6.7 Electrical Idle Detected Register - 0x84_0000_0830

TABLE 3-40 Electrical Idle Detected Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:28]	0x0	RO	Reserved
ELECTIDLE1	[27:14]	0x3ff	RO	Electrical Idle detected from bit lanes
ELECTIDLE0	[13:0]	0x3ff	RO	Electrical Idle detected from bit lanes

3.13.6.8 Disable State Period Register - 0x84_0000_0838

TABLE 3-41 Disable State Period Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:10]	0x0	RO	Reserved
COUNT	[9:0]	0xFF	RW	Counter value for Disable state. Once Disable state is entered, a counter will count to this value. Once it reaches it, the Disable_Done bit in the Disable State Period Done Register will be set.

3.13.6.9 Disable State Period Done Register - 0x84_0000_0840

TABLE 3-42 Disable State Period Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:1]	0x0	RO	Reserved
DONE	[0]	0x0	RW	Indicates that counter for Disable state period has completed counting.

3.13.6.10 Calibrate State Period Register - 0x84_0000_0848

TABLE 3-43 Calibrate State Period Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:20]	0x0	RO	Reserved
COUNT	[19:0]	0x0	RW	Counter value for Calibrate state. Once Calibrate state is entered, a counter will count to this value. Once it reaches it, the Calibrate_Done bit in the FBD Status Register will be set.

3.13.6.11 Calibrate State Period Done Register - 0x84_0000_0850

TABLE 3-44 Calibrate State Period Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:1]	0x0	RO	Reserved
DONE	[0]	0x0	RW	Indicates that counter for Disable state period has completed counting.

3.13.6.12 Training State Minimum Time Register - 0x84_0000_0858

TABLE 3-45 Training State Minimum Time Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:16]	0x0	RO	Reserved
COUNT	[15:0]	0xFF	RW	Minimum number of frames for Training state before starting to check for Done.

3.13.6.13 Training State Done Register - 0x84_0000_0860

TABLE 3-46 Training State Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
TIMEOUT	[1]	0x0	RW	Set when timeout period has elapsed before Done has been asserted.
DONE	[0]	0x0	RW	Set when Training state has completed.

3.13.6.14 Training State Timeout Register - 0x84_0000_0868

TABLE 3-47 Training State Timeout Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:8]	0x0	RO	Reserved
PERIOD	[7:0]	0xFF	RW	Number of frames for Training state to complete after minimum number of frames have elapsed.

3.13.6.15 Testing State Done Register - 0x84_0000_0870

TABLE 3-48 Testing State Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
TIMEOUT	[1]	0x0	RW	Set when timeout period has elapsed before Done has been asserted.
DONE	[0]	0x0	RW	Set when Testing state has completed.

3.13.6.16 Testing State Timeout Register - 0x84_0000_0878

TABLE 3-49 Testing State Timeout Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:8]	0x0	RO	Reserved
PERIOD	[7:0]	0xFF	RW	Number of frames for Testing state to complete.

3.13.6.17 Polling State Done Register - 0x84_0000_0880

TABLE 3-50 Polling State Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
TIMEOUT	[1]	0x0	RW	Set when timeout period has elapsed before Done has been asserted.
DONE	[0]	0x0	RW	Set when Polling state has completed.

3.13.6.18 Polling State Timeout Register - 0x84_0000_0888

TABLE 3-51 Polling State Timeout Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:8]	0x0	RO	Reserved
PERIOD	[7:0]	0xFF	RW	Number of frames for Polling state to complete.

3.13.6.19 Config State Done Register - 0x84_0000_0890

TABLE 3-52 Config State Done Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
TIMEOUT	[1]	0x0	RW	Set when timeout period has elapsed before Done has been asserted.
DONE	[0]	0x0	RW	Set when Config state has completed.

3.13.6.20 Config State Timeout Period Register - 0x84_0000_0898

TABLE 3-53 Config State Timeout Period Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:8]	0x0	RO	Reserved
PERIOD	[7:0]	0xFF	RW	Number of frames for Config state to complete.

3.13.6.21 Per Rank CKE Register - 0x84_0000_08A0

Writing this register or the CKE bit of register 0x84_0000_01A0 sends a CKE command to the FBDIMMs. Each bit corresponds to a rank in a fully populated FBDIMM system. Bit 0 is for DIMM0, rank0; bit one is for DIMM0, rank1; etc. The enable bits are qualified by the number of DIMMs and whether they are stacked before the CKE command is issued to the DIMMs.

TABLE 3-54 Per Rank CKE Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:16]	0x0	RO	Reserved
D7R1	[15]	0x1	RW	CKE enable for DIMM 7 Rank 1
D7R0	[14]	0x1	RW	CKE enable for DIMM 7 Rank 0
D6R1	[13]	0x1	RW	CKE enable for DIMM 6 Rank 1
D6R0	[12]	0x1	RW	CKE enable for DIMM 6 Rank 0
D5R1	[11]	0x1	RW	CKE enable for DIMM 5 Rank 1
D5R0	[10]	0x1	RW	CKE enable for DIMM 5 Rank 0
D4R1	[9]	0x1	RW	CKE enable for DIMM 4 Rank 1
D4R0	[8]	0x1	RW	CKE enable for DIMM 4 Rank 0
D3R1	[7]	0x1	RW	CKE enable for DIMM 3 Rank 1
D3R0	[6]	0x1	RW	CKE enable for DIMM 3 Rank 0
D2R1	[5]	0x1	RW	CKE enable for DIMM 2 Rank 1
D2R0	[4]	0x1	RW	CKE enable for DIMM 2 Rank 0
D1R1	[3]	0x1	RW	CKE enable for DIMM 1 Rank 1
D1R0	[2]	0x1	RW	CKE enable for DIMM 1 Rank 0
D0R1	[1]	0x1	RW	CKE enable for DIMM 0 Rank 1
D0R0	[0]	0x1	RW	CKE enable for DIMM 0 Rank 0

3.13.6.22 L0s Duration - 0x84_0000_08A8

TABLE 3-55 L0s Duration

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:7]	0x0	RO	Reserved
ENABLE	[6]	0x0	RW	Enables use of L0s mode when MCU is idle.
COUNT	[5:0]	0x2A	RW	Determines the number of frames that the branch will be in the L0s state. Legal values are 0x20 to 0x2A. Values below 0x20 will be treated as 0x20 and values above 0x2A will be treated as 0x2A.

3.13.6.23 Sync Frame Frequency Register - 0x84_0000_08B0

TABLE 3-56 Sync Frame Frequency Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:6]	0x0	RO	Reserved
FREQ	[5:0]	0x2A	RW	Frequency at which sync frames are issued on the FBDIMM channels

3.13.6.24 Channel Read Latency Register - 0x84_0000_08B8

TABLE 3-57 Channel Read Latency Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:16]	0x0	RO	Reserved
LATENCY1	[15:8]	0xFF	RW	Read latency for channel 1. Determined during Polling state.
LATENCY0	[7:0]	0xFF	RW	Read latency for channel 0. Determined during Polling state.

3.13.6.25 Channel Capability Register - 0x84_0000_08C0

TABLE 3-58 Channel Capability Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:10]	0x0	RO	Reserved
CAPABIL1	[9:5]	0x0	RO	Channel capabilities for selected AMB in channel 1. Only valid during Polling state.
CAPABIL0	[4:0]	0x0	RO	Channel capabilities for selected AMB in channel 0. Only valid during Polling state.

3.13.6.26 Loopback Mode Control Register - 0x84_0000_08C8

TABLE 3-59 Loopback Mode Control Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
MODE	[1:0]	0x0	RW	Loopback Mode: 0x: Loopback Mode disabled 10: Place low-order NB data on SB bus 11: Place high-order NB data on SB bus

3.13.6.27 SERDES Configuration Bus Register - 0x84_0000_08D0

TABLE 3-60 SERDES Configuration Bus Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:25]	0x0	RO	Reserved
RXTX_RATE	[29:28]	0x0	RW	Receiver/Transmitter Operating Rate
TX_CM	[27]	0x0	RW	Transmitter Common Mode
TX_SWING	[26:24]	0x1	RW	Transmitter Output Swing
TX_DE	[23:20]	0x0	RW	Transmitter De-emphasis
TX_ENFTP	[19]	0x0	RW	Transmitter Enable
RX_TERM	[18:16]	0x0	RW	Receiver Termination
RSVD	[15]	0x0	RO	Reserved
RX_CDR	[14:12]	0x0	RW	Receiver Clock/Data Recovery Algorithm
RX_EQ	[11:8]	0x0	RW	Receiver Adaptive Equalizer Configuration
RSVD	[7:6]	0x0	RO	Reserved
PLL_MPY	[5:2]	0x0	RW	PLL Multiplier
PLL_LB	[1:0]	0x0	RW	Loop bandwidth

3.13.6.28 SERDES Transmitter and Receiver Differential Pair Inversion Register - 0x84_0000_08D8

TABLE 3-61 SERDES Transmitter and Receiver Differential Pair Inversion Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:48]	0x0	RO	Reserved
TX1_INVPAIR	[47:38]	0x0	RW	Invert Channel 1 TXPi/TXNi if bit is 1.
TX0_INVPAIR	[37:28]	0x0	RW	Invert Channel 0 TXPi/TXNi if bit is 1.
RX1_INVPAIR	[27:14]	0x0	RW	Invert Channel 1 RXPi/RXNi if bit is 1.
RX0_INVPAIR	[13:0]	0x0	RW	Invert Channel 0 RXPi/RXNi if bit is 1.

3.13.6.29 SERDES Test Configuration Bus Register - 0x84_0000_08E0

TABLE 3-62 SERDES Test Configuration Bus Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:32]	0x0	RO	Reserved
FSR1_TX_ENTEST	[31]	0x0	RW	Enable testing of FSR1 Transmit ports
FSR0_TX_ENTEST	[30]	0x0	RW	Enable testing of FSR0 Transmit ports
FSR1_RX_ENTEST	[29]	0x0	RW	Enable testing of FSR1 Receive ports
FSR0_RX_ENTEST	[28]	0x0	RW	Enable testing of FSR0 Receive ports
FSR1_INVPATT	[27]	0x0	RW	FSR1 Invert Polarity
FSR1_RATE	[26:25]	0x0	RW	FSR1 Operating Rate
FSR1_ENBSPLS	[24]	0x0	RW	FSR1 Receiver pulse boundary scan
FSR1_ENBSRX	[23]	0x0	RW	FSR1 Receiver boundary scan
FSR1_ENBSTX	[22]	0x0	RW	FSR1 Transmitter boundary scan
FSR1_LOOPBACK	[21:20]	0x0	RW	FSR1 Loopback
FSR1_CLKBYP	[19:18]	0x0	RW	FSR1 Clock bypass
FSR1_ENRXPATT	[17]	0x0	RW	FSR1 Enable Rx patterns
FSR1_ENTXPATT	[16]	0x0	RW	FSR1 Enable Tx patterns
FSR1_TESTPATT	[15:14]	0x0	RW	FSR1 Test pattern
FSR0_INVPATT	[13]	0x0	RW	FSR0 Invert Polarity
FSR0_RATE	[12:11]	0x0	RW	FSR0 Operating Rate
FSR0_ENBSPLS	[10]	0x0	RW	FSR0 Receiver pulse boundary scan
FSR0_ENBSRX	[9]	0x0	RW	FSR0 Receiver boundary scan
FSR0_ENBSTX	[8]	0x0	RW	FSR0 Transmitter boundary scan
FSR0_LOOPBACK	[7:6]	0x0	RW	FSR0 Loopback
FSR0_CLKBYP	[5:4]	0x0	RW	FSR0 Clock bypass
FSR0_ENRXPATT	[3]	0x0	RW	FSR0 Enable Rx patterns
FSR0_ENRXPATT	[2]	0x0	RW	FSR0 Enable Tx patterns
FSR0_TESTPATT	[1:0]	0x3	RW	FSR0 Test pattern

3.13.6.30 SERDES PLL Status Register - 0x84_0000_08E8

TABLE 3-63 SERDES PLL Status Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:6]	0x0	RO	Reserved
FSR1_STSPLL	[5:3]	0x0	RO	PLL Lock Status for FSR1 macros
FSR0_STSPLL	[2:0]	0x0	RO	PLL Lock Status for FSR0 macros

3.13.6.31 SERDES Test Status Register - 0x84_0000_08F0

TABLE 3-64 SERDES Test Status Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:48]	0x0	RO	Reserved
FSR1_TX_TESTFAIL	[47:38]	0x0	RO	Test Status for FSR1 Transmit ports
FSR0_TX_TESTFAIL	[37:28]	0x0	RO	Test Status for FSR0 Transmit ports
FSR1_RX_TESTFAIL	[27:14]	0x0	RO	Test Status for FSR1 Receive ports
FSR0_RX_TESTFAIL	[13:0]	0x0	RO	Test Status for FSR0 Receive ports

3.13.6.32 Configuration Register Access Address Register - 0x84_0000_0900

TABLE 3-65 Configuration Register Access Address Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:16]	0x0	RO	Reserved
CHANNEL	[15]	0x0	RW	Channel of Configuration Register Access.
AMB	[14:11]	0x0	RW	AMB ID of Configuration Register Access.
DATA	[10:2]	0x0	RW	Address for Configuration Register read or write.
RSVD	[1:0]	0x0	RO	Reserved

3.13.6.33 Configuration Register Access Data Register - 0x84_0000_0908

TABLE 3-66 Configuration Register Access Data Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:32]	0x0	RO	Reserved
DATA	[31:0]	0x0	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.

3.13.6.34 FBD Thermal Trip Status Register - 0x84_0000_0A00

TABLE 3-67 FBD Thermal Trip Status Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:48]	0x0	RO	Reserved
TTRIP1_11	[47:46]	0x0	RO	Thermal Trip information for AMB 11, Channel 1
TTRIP1_10	[45:44]	0x0	RO	Thermal Trip information for AMB 10, Channel 1
TTRIP1_9	[43:42]	0x0	RO	Thermal Trip information for AMB 9, Channel 1
TTRIP1_8	[41:40]	0x0	RO	Thermal Trip information for AMB 8, Channel 1
TTRIP1_7	[39:38]	0x0	RO	Thermal Trip information for AMB 7, Channel 1
TTRIP1_6	[37:36]	0x0	RO	Thermal Trip information for AMB 6, Channel 1
TTRIP1_5	[35:34]	0x0	RO	Thermal Trip information for AMB 5, Channel 1
TTRIP1_4	[33:32]	0x0	RO	Thermal Trip information for AMB 4, Channel 1
TTRIP1_3	[31:30]	0x0	RO	Thermal Trip information for AMB 3, Channel 1
TTRIP1_2	[29:28]	0x0	RO	Thermal Trip information for AMB 2, Channel 1
TTRIP1_1	[27:26]	0x0	RO	Thermal Trip information for AMB 1, Channel 1
TTRIP1_0	[25:24]	0x0	RO	Thermal Trip information for AMB 0, Channel 1
TTRIP0_11	[23:22]	0x0	RO	Thermal Trip information for AMB 11, Channel 0
TTRIP0_10	[21:20]	0x0	RO	Thermal Trip information for AMB 10, Channel 0
TTRIP0_9	[19:18]	0x0	RO	Thermal Trip information for AMB 9, Channel 0
TTRIP0_8	[17:16]	0x0	RO	Thermal Trip information for AMB 8, Channel 0
TTRIP0_7	[15:14]	0x0	RO	Thermal Trip information for AMB 7, Channel 0

TABLE 3-67 FBD Thermal Trip Status Registers *(Continued)*

Field	Bit Position	Initial Value	R/W	Description
TTRIP0_6	[13:12]	0x0	RO	Thermal Trip information for AMB 6, Channel 0
TTRIP0_5	[11:10]	0x0	RO	Thermal Trip information for AMB 5, Channel 0
TTRIP0_4	[9:8]	0x0	RO	Thermal Trip information for AMB 4, Channel 0
TTRIP0_3	[7:6]	0x0	RO	Thermal Trip information for AMB 3, Channel 0
TTRIP0_2	[5:4]	0x0	RO	Thermal Trip information for AMB 2, Channel 0
TTRIP0_1	[3:2]	0x0	RO	Thermal Trip information for AMB 1, Channel 0
TTRIP0_0	[1:0]	0x0	RO	Thermal Trip information for AMB 0, Channel 0

3.13.6.35 MCU Syndrome Register - 0x84_0000_0C0

TABLE 3-68 MCU Syndrome Registers

Field	Bit Position	Initial Value	R/W	Description
VALID	[63]	0x0	RW	Error Status is Valid
RSVD	[62:30]	0x0	RO	Reserved
ALERT1	[29:18]	0x0	RW	AMB's on Channel 1 with Status Alert Asserted
ALERT0	[17:6]	0x0	RW	AMB's on Channel 0 with Status Alert Asserted
SOFTRESET	[5]	0x0	RW	Soft Channel Reset Performed on Channel
FASTRESET	[4]	0x0	RW	Fast Reset Performed on Channel
SOURCE	[3:0]	0x0	RW	Source(s) of error (multiple may be asserted: 0xXXX1: CRC Error 0xXX1X: Alert Frame 0xX1XX: Status Alert Asserted 0x1XXX: Status Frame Parity Error

3.13.6.36 Injected Error Source Register - 0x84_0000_0C08

TABLE 3-69 Injected Error Source Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:2]	0x0	RO	Reserved
SOURCE	[1:0]	0x0	RW	Source for Injected Error: 0x0 - CRC Error 0x1 - Alert Frame 0x2 - Status Alert Asserted 0x3 - Status Frame Parity Error

3.13.6.37 MCU FBR Count Register - 0x84_0000_0C10

TABLE 3-70 MCU FBR Count Registers

Field	Bit Position	Initial Value	R/W	Description
RSVD	[63:17]	0x0	RO	Reserved
COUNTONE	[16]	0x0	RW	Hardware behaves as if count was always one, i.e. it will always generate an interrupt
COUNT	[15:0]	0x0	RW	Number of recoverable errors before a recoverable error interrupt will be sent to NCU

3.14 Other Registers

3.14.1 Self-Refresh Registers

The Clock Control Register in the Clock Control Unit (CCU), described in the *OpenSPARC T2 Programmer's Reference Manual*, controls self-refresh mode for the MCU upon assertion of warm reset.

CHAPTER 4

Test Control Unit (TCU)

This document defines the architecture for the Test Control Unit (TCU) for OpenSPARC T2. It is intended for RTL design engineers of other OpenSPARC T2 blocks, verification engineers, and DFT engineers. The document contains the functional description, and some level of implementation detail for the TCU, test control unit.

This chapter contains the following sections:

4.1 Introduction

The TCU is the OpenSPARC T2 Test Control Unit and provides access to the chip test logic. It also participates in Reset, eFuse programming, clock stop/start sequencing, and chip debug. The TCU including Joint Test Action Group (JTAG) is completely stuck-fault testable via Automatic Test Pattern Generation (ATPG) manufacturing scan.

4.1.1 Features

The features available for debug or test, implemented in OpenSPARC T2 and supported by the TCU are as follows:

- ATPG or Manufacturing scan for stuck-fault testing.
- TAP and Boundary Scan (JTAG) - support for IEEE 1149.1 and 1149.6
- JTAG scan for scan chain loading and unloading.
- JTAG shadow scan - allows for inspection of specific registers while part is running in system.
- Support for Macrotest.
- JTAG UCB - Allows CREG access via instructions sent to the NCU which will then intermix the transaction with normal requests. The NCU will then take the results and pass them back to the TCU which can then send out TDO.
- eFuse - Control and programming
- Transition fault testing - This is done on the tester while PLLs are locked; slower domains may be driven via pins directly.
- MBIST - Memory Built-in Self Test; tests Array bit cells and write/read mechanisms. BISI (Built-in Self Initialization) allows arrays to be initialized.
- LBIST - Logic BIST, implemented in cores.
- Reset - Handshaking with RST unit to control scan flop reset and clock stop/start sequencing.
- L2 Access - via JTAG through the SIU.
- Debug Support.
- Support for SERDES - ATPG, STCI, boundary scan

4.2 Joint Action Test Group (JTAG)

The JTAG block resides in the TCU. This interface will be used to access not only standard JTAG services but also implementation of specific debug features. The JTAG architecture is designed to be compliant with the IEEE 1149.1 Standard.

JTAG provides these features:

1. Access to JTAG ID code
2. Implementation of JTAG public instructions (see Note)
3. Ability to load/unload chip scan chains as a single chain, or individually
4. Initiation and control of Shadow scan
5. Initiation and control of Boundary scan
6. Control of MBIST or BISI
7. Control of LBIST
8. Interface to Chip UCB
9. Interface to E-Fuse Unit
10. Interface to various debug features
11. Control of clock domains (starting, stopping)
12. Write/read access to L2

The following JTAG pins are implemented for OpenSPARC T2:

- TDI
- TDO
- TMS
- TRST_L
- TCK

Note – The JTAG unit implements all instructions specified as mandatory in the 1149.1 and 1149.6 standards along with a number of private instructions that help the debugger to access specific debug features. However, not all I/O on OpenSPARC T2 support the HIGHZ and SAMPLE instructions. HIGHZ and SAMPLE are not supported on SERDES I/O. In addition, some non-SERDES I/O do not implement HIGHZ correctly.

4.2.1 Instruction Register

The instruction register provides eight bits to access up to 256 instructions. On the rising edge of TCK in the capture-IR state, the instruction register shift portion is updated to the IDCODE instruction. The instruction register update portion loads the IDCODE instruction on the falling edge of TCK in the reset state, or when TRST_L goes low.

TABLE 4-1 JTAG Instruction Register

Instruction Decode	Value on Reset
7:0	IDCODE Instr: 8b 0000 0001

4.2.2 Reset State and TRST_L

The TRST_L pin provides an asynchronous reset for the JTAG state machine and associated registers. When TRST_L is activated (low), the TMS pin should be held high and it is recommended that TCK be held off. When TRST_L goes low:

- The TAP state machine is put in the test-logic-reset state.
- The Instruction Register is set to the IDCODE instruction.
- All data registers internal to the JTAG block are reset to their default states.

After TRST_L is deasserted it is recommended to keep TCK off until JTAG is to be used, and then allow TCK to run with TMS be held high for a few cycles to allow the reset state inside JTAG to stabilize before entering the Run-Test-Idle state.

Synchronous resetting of the TAP is done by entering the test-logic-reset state via control of TMS and TCK. This does not necessarily reset private data registers to their default states.

4.2.3 Instruction Summary

Unimplemented or undefined instructions will default to the BYPASS instruction.

TABLE 4-2 JTAG Public Instructions

Instruction	Encoding	Description
TAP_BYPASS	0xFF	Mandatory; selects BYPASS REGISTER
TAP_EXTEST	0x00	Mandatory; selects BOUNDARY SCAN REGISTER
TAP_IDCODE	0x01	Optional per standard; selects IDCODE DR

TABLE 4-2 JTAG Public Instructions *(Continued)*

Instruction	Encoding	Description
TAP_CLAMP	0x04	Optional per standard
TAP_EXTEST_PULSE	0x05	Mandatory for 1149.6
TAP_EXTEST_TRAIN	0x06	Mandatory for 1149.6

TABLE 4-3 JTAG Private Instructions

Instruction	Encoding	Description
TAP_SAMPLE_PRELOAD	0x02	Mandatory; shared encoding allowed per standard; selects BOUNDARY SCAN REGISTER - SERDES I/O do not support SAMPLE part of this instruction
TAP_HIGHZ	0x03	Optional per standard - SERDES I/O do not support HIGHZ, and some DBG_DQ I/O have weak pullup/down resistors.
UNDEFINED	0x07	--
TAP_CREG_ADDR	0x08	Stores address to be used for system access to control reg (ASI/IO mapped)
TAP_CREG_WDATA	0x09	Stores data to be used for system access to control reg
TAP_CREG_RDATA	0x0A	Captures data from system access
UNDEFINED	0x0B	--
TAP_NCU_WRITE	0x0C	Initiates write to system control register
TAP_NCU_READ	0x0D	Initiates read from system control register
TAP_NCU_WADDR	0x0E	Combination of TAP_CREG_ADDR and TAP_NCU_WRITE
TAP_NCU_WDATA	0x0F	Combination of TAP_CREG_WDATA and TAP_NCU_WRITE
TAP_NCU_RADDR	0x10	Combination of TAP_CREG_ADDR and TAP_NCU_READ
UNDEFINED	0x11-0x12	--
TAP_MBIST_CLKSTPEN	0x13	Enables clock stop for mbist via cycle counter
TAP_MBIST_BYPASS	0x14	Select engines to by excluded from MBIST operation; using mbist_bypass data register
TAP_MBIST_MODE	0x15	Specify serial/parallel, diag. mode or bist/bisi modes via mbist_mode data reg
TAP_MBIST_START	0x16	Initiate MBIST
UNDEFINED	0x17	--
TAP_MBIST_RESULT	0x18	Query 2-bit done/fail register: and/or of all MBIST engines
TAP_MBIST_DIAG	0x19	Run MBIST on one array; MBIST engine & arrays are data reg
TAP_MBIST_GETDONE	0x1A	Query 48-bit done data register, one bit per MBIST engine

TABLE 4-3 JTAG Private Instructions *(Continued)*

Instruction	Encoding	Description
TAP_MBIST_GETFAIL	0x1B	Query 48-bit fail data register, one bit per MBIST engine
TAP_DMO_ACCESS	0x1C	Set DMO Mode - enables DMO logic and package pins
TAP_DMO_CLEAR	0x1D	Clears DMO Mode
TAP_DMO_CONFIG	0x1E	Access 48-bit DMO configuration register
TAP_MBIST_ABORT	0x1F	Stop any MBIST activity and reset MBIST controls
UNDEFINED	0x20-0x27	--
TAP_FUSE_READ	0x28	Shift out 32 bits selected by ROW_ADDR; selects EFUSE DR
TAP_FUSE_BYPASS_DATA	0x29	Provides user-data directly to EFU; selects EFUSE DR
TAP_FUSE_BYPASS	0x2A	Starts EFU control using bypass data provided by user
TAP_FUSE_ROW_ADDR	0x2B	Shift in 7-bit row address for EFU access; selects EFU ROW ADDRESS DR
TAP_FUSE_COL_ADDR	0x2C	Shift in 5-bit column address for EFU programming; selects EFU COLUMN ADDRESS DR
TAP_FUSE_READ_MODE	0x2D	Configures EFU with three bits for EFU access; selects EFU READ MODE DR
TAP_FUSE_DEST_SAMPLE	0x2E	Samples EFU destination redundancy value from the destination specified
TAP_FUSE_RVCLR	0x2F	Access 7-bit redundancy value clear register
TAP_SPCTHR0_SHSCAN	0x30	Samples thread 0 for all available cores
TAP_SPCTHR1_SHSCAN	0x31	Samples thread 1 for all available cores
TAP_SPCTHR2_SHSCAN	0x32	Samples thread 2 for all available cores
TAP_SPCTHR3_SHSCAN	0x33	Samples thread 3 for all available cores
TAP_SPCTHR4_SHSCAN	0x34	Samples thread 4 for all available cores
TAP_SPCTHR5_SHSCAN	0x35	Samples thread 5 for all available cores
TAP_SPCTHR6_SHSCAN	0x36	Samples thread 6 for all available cores
TAP_SPCTHR7_SHSCAN	0x37	Samples thread 7 for all available cores
TAP_L2T_SHSCAN	0x38	Samples specified error registers in the eight L2 Tags
UNDEFINED	0x39-0x3F	--
TAP_CLOCK_SSTOP	0x40	Soft Stop of clocks; cores only
TAP_CLOCK_HSTOP	0x41	Hard Stop of clocks
TAP_CLOCK_START	0x42	Start clocks
TAP_CLOCK_DOMAIN	0x43	Specify entry clock domain for stopping/starting clocks

TABLE 4-3 JTAG Private Instructions *(Continued)*

Instruction	Encoding	Description
TAP_CLOCK_STATUS	0x44	2-bit status indicating if clock stop/start routine finished
TAP_CLKSTP_DELAY	0x45	7-bits; Specify up to 128 cycle delay between successive clk_stop signals
TAP_CORE_SEL	0x46	8-bit register to specify target SPC cores for clock operations.
UNDEFINED	0x47	--
TAP_DE_COUNT	0x48	Access 32-bit Debug Event Counter
TAP_CYCLE_COUNT	0x49	Access 64-bit Reset/Cycle Counter
TAP_TCU_DCR	0x4A	Access 4-bit TCU Debug (event) Control Register
UNDEFINED	0x4B	--
TAP_CORE_RUN_STATUS	0x4C	Access 64-bit CMP core-running-status reg.
TAP_DOSS_ENABLE	0x4D	Access 64-bit disable overlap/single step mode enable register
TAP_DOSS_MODE	0x4E	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	0x4F	Pulse single step request signal; need to go through update-dr
TAP_DOSS_STATUS	0x50	8-bit status for disable overlap or single step completion
TAP_CS_MODE	0x51	Specify cycle-step mode. 1-bit register. set to '1' to enable; uses Cycle Counter for cycle-step operation.
TAP_CS_STATUS	0x52	Read 1-bit status indicating cycle stepping has completed.
UNDEFINED	0x53-0x57	--
TAP_L2_ADDR	0x58	Load L2 Address (to be written to or read from)
TAP_L2_WRDATA	0x59	Load L2 Write Data
TAP_L2_WR	0x5A	Initiate write to L2: WRDATA to ADDR
TAP_L2_RD	0x5B	Initiate read from L2 at ADDR and receive L2 data
UNDEFINED	0x5C-0x5F	--
TAP_LBIST_START	0x60	Initiate Logic BIST
TAP_LBIST_BYPASS	0x61	Bypass Logic BIST for specified cores; 1 bit per core
TAP_LBIST_MODE	0x62	Control program mode; parallel/serial modes
TAP_LBIST_ACCESS	0x63	Place one Logic BIST controller between TDI-TDO
TAP_LBIST_GETDONE	0x64	Determine if Logic BIST is done across all selected cores
TAP_LBIST_ABORT	0x65	Abort any Logic BIST currently in progress
UNDEFINED	0x66-0x7F	--

TABLE 4-3 JTAG Private Instructions *(Continued)*

Instruction	Encoding	Description
TAP_SERSCAN	0x80	Access internal scan chains; selects INTERNAL SCAN FLOPS as DATA REGISTER
TAP_CHAINSEL	0x81	Select all or one of 32 chains for serial scan mode using CHAIN SELECT DR
TAP_MT_ACCESS	0x82	Enables Macro Test mode for JTAG Scan
TAP_MT_CLEAR	0x83	Clears Macro Test mode
TAP_MT_SCAN	0x84	Similar to TAP_SERSCAN but drives TCK onto clock tree for capture pulses during RTI state
UNDEFINED	0x85-0x87	--
TAP_TP_ACCESS	0x88	Enables Test Protect mode to block inputs to TCU and other blocks such as RST, CCU, and DMU
TAP_TP_CLEAR	0x89	Clears Test Protect mode
UNDEFINED	0x8A-0x8F	--
TAP_STCI_ACCESS	0x90	Enables STCI mode for SERDES Test Configuration Interface Bus
TAP_STCI_CLEAR	0x91	Clears STCI mode for SERDES Test Configuration Interface Bus
UNDEFINED	0x92-0x9F	--
TAP_JTPOR_ACCESS	0xA0	Enables JTAG access window during POR sequence
TAP_JTPOR_CLEAR	0xA1	Clears JTAG access window during POR sequence
TAP_JTPOR_STATUS	0xA2	JTAG access window status: returns ‘1’ if window is active
TAP_SCKBYP_ACCESS	0xA3	Enables Bypass for SCK counter in NCU
TAP_SCKBYP_CLEAR	0xA4	Clears Bypass for SCK counter in NCU
UNDEFINED	0xA5-0xFE	--

4.2.4 Data Registers

The data registers accessible via JTAG are listed in TABLE 4-4. The least significant bit (lsb: bit 0, the rightmost bit) is always closest to tdo.

TABLE 4-4 JTAG Data Registers

Data Register	Width	Capture	Update	Description & JTAG Instructions
Boundary Scan	~ 3 x I/Os	Yes	Yes	I/O pin boundary scan cells; see private instructions
Bypass	1	Yes	No	Selected via bypass instr. or any undefined instr.
ID Code	32	Yes	No	See following section; TAP_IDCODE
Chain Select	6	No	Yes	Active during JTAG Serial Scan; TAP_CHAINSEL
Serial (Internal) Scan	~1200K	No	via scan	Chip scan chains; TAP_SERSCAN, TAP_MT_SCAN
Macrotest Enable	1	No	Yes	Enables write/read control of arrays during serial scan; set with TAP_MT_ACCESS, clear with TAP_MT_CLEAR
Test Protect Enable	1	No	Yes	Enables Test Protect mode; set with TAP_TP_ACCESS, clear with TAP_TP_CLEAR
EFUSE	32	Yes	No	Contents of 1 row in eFuse array, per Row Addr.; TAP_FUSE_READ
EFUSE Bypass_Data	32	No	Yes	Data for BYPASSING eFuse Array; TAP_FUSE_BYPASS_DATA
EFUSE Row_Address	7	Yes	Yes	Select one row in the eFuse array; TAP_FUSE_ROW_ADDR
EFUSE Column_Address	5	Yes	Yes	eFuse Column, only for programming; TAP_FUSE_COL_ADDR
EFUSE Read_Mode	3	Yes	Yes	See eFuse document; TAP_FUSE_READ_MODE
EFUSE Dest_Sample	32	Yes	No	See eFuse Document; TAP_FUSE_DEST_SAMPLE
EFUSE RVCLEAR	7	No	Yes	TAP_FUSE_RVCLR bit[6]=enable; bits[5:0]=RV_ID
MBIST Result	2	Yes (1)	No	Bit 1=1 when MBIST Done, Bit 0=1 if MBIST failed; TAP_MBIST_RESULT
MBIST Bypass	48	Yes	Yes	MBIST: Specify engines to bypass during MBIST; TAP_MBIST_BYPASS
MBIST Done	48	Yes (1)	No	MBIST engine Done status bits; TAP_MBIST_GETDONE
MBIST Fail	48	Yes (1)	No	MBIST engine Fail status bits; TAP_MBIST_GETFAIL
MBIST Diag	Variable	No	via scan	All registers in an MBIST engine. Updated via scan only. TAP_MBIST_DIAG

TABLE 4-4 JTAG Data Registers *(Continued)*

Data Register	Width	Capture	Update	Description & JTAG Instructions
MBIST Mode	4	No	via scan	Select serial or parallel modes; bisi; user mode TAP_MBIST_MODE
CREG Address	40	No	Yes	40-bit address for system control register; TAP_CREG_ADDR, TAP_NCU_WADDR, TAP_NCU_RADDR
CREG Write_Data	64	No	Yes	64-bit data to be written to system control register; TAP_CREG_WDATA, TAP_NCU_WDATA
CREG Read_Data	65	Yes	No	65-bit data read from system control register; TAP_CREG_RDATA; due to sentinel bit, scan-out data is blocked to TDO during shiftDR
Core Shadow_Scan	8*len	Yes	via scan	Shadow scan for all available cores concatenated, spc0 to spc7; TAP_SPCTHR0_SHSCAN-TAP_SPCTHR7_SHSCAN
L2TAG Shadow_Scan	8*len	Yes	via scan	Shadow scan for all l2 tags concatenated, l2t0 to l2t7; TAP_L2T_SHSCAN
Clock Domain	32	Yes (1)	Yes	Specify starting points for turning clocks on or off; TAP_CLOCK_DOMAIN; bits [31:24] reserved, should be loaded to zeros.
Clock Status	2	Yes (1)	No	TAP_CLOCK_STATUS bits = 00 --> clock sequencer is running bits = 01 --> clock sequencer has started all clocks bits = 10 --> clock sequencer has stopped all clocks bits = 11 --> should not happen; indeterminate
Clock Stop Delay	7	Yes (1)	Yes	Delay between successive clk_stop's to clk domains; TAP_CLKSTP_DELAY
Core Select	8	Yes (1)	Yes	Enables clock stop to target cores; TAP_CORE_SEL
Debug Event Counter	32	Yes (2)	Yes	Decrementing counter to delay debug action by counting debug events; TAP_DE_COUNT
Cycle Counter	64	Yes (2)	Yes	Decrementing counter triggered by debug event; upper word is Reset Counter; TAP_CYCLE_COUNT
TCU Debug Control	4	Yes (1)	Yes	Control reg. for TCU debug events; TAP_TCU_DCR
(CMP) Core Run Status	64	Yes (1)	No	Thread (CMP) running status register; TAP_CORE_RUN_STATUS
DOSS Enable	64	Yes	Yes	Disable Overlap (do) and Single Step (ss) enable bits, per thread; TAP_DOSS_ENABLE
DOSS Mode	2	Yes	Yes	Controls disable overlap or single step modes; TAP_DOSS_MODE

TABLE 4-4 JTAG Data Registers *(Continued)*

Data Register	Width	Capture	Update	Description & JTAG Instructions
DOSS Status	8	Yes (1)	No	Indicates completion of disable overlap or single step per SPC; TAP_DOSS_STATUS
Cycle Step Enable	1	Yes (1)	Yes	Enable cycle step mode; TAP_CS_MODE
Cycle Step Status	1	Yes (1)	No	Indicates cycle step is complete; TAP_CS_STATUS
L2_Addr	65	No (3)	Yes	64-bit address to L2 for write or read; bit 0 ignored; TAP_L2_ADDR
L2_Write_Data	65	No (3)	Yes	64-bit data to write to L2; bit 0 ignored; TAP_L2_WRDATA
L2_Read_Data	65	Yes (1)	No	64-bit data received from L2 per ADDR; bit 0 indicates data is valid, bits 64:1 are data; TAP_L2_RD
LBIST Bypass	8	Yes	Yes	One bit per core, to bypass an engine set to '1'; TAP_LBIST_BYPASS
LBIST Mode	2	No	via Scan	bit[1]: enable user (program) mode bit[0]: 0=serial, 1=parallel TAP_LBIST_MODE
LBIST Access	tbd	No	via Scan	Place one Logic BIST controller between TDI-TDO; TAP_LBIST_ACCESS
LBIST Done	8	Yes (1)	No	Read status of all enabled Logic BIST controllers; TAP_LBIST_GETDONE
DMO Config	48	Yes (2)	Yes	Access 48-bit DMO Configuration register; TAP_DMO_CONFIG
JTAG POR Status	1	Yes (1)	No	Access 1-bit status for JTAG POR Access; TAP_JTPOR_STATUS

An update of "Yes" means there is an update register loaded during UpdateDR; an update of "via scan" means there is no separate update register.

For Notes (1) and (2) in the Capture field of TABLE 4-4, see JTAG Errata.

Note (3): L2_Addr and L2_Write_Data registers cannot be observed during shiftDR due to a bug in the RTL - see errata for details.

4.2.4.1 Boundary Scan

The boundary scan data register is selected by EXTEST, SAMPLE/PRELOAD, HIGHZ and CLAMP and is defined by the BSDL (Boundary Scan Description Language) file for OpenSPARC T2. It is also selected by the 1149.6 instructions TAP_EXTEST_TRAIN and TAP_EXTEST_PULSE and is part of the internal scan register when selected by manufacturing scan.

The HIGHZ instruction is not supported by the SERDES I/O, and also some of the DBG_DQ pins have weak pullup or pulldown resistors. So the HIGHZ instruction is not fully supported by OpenSPARC T2. See JTAG Errata for details.

The SAMPLE instruction (encoded with PRELOAD) is not supported by the SERDES I/O, but the PRELOAD instruction is supported.

4.2.4.2 Bypass Register

This is a one-bit register. The bypass register loads "0" on the rising edge of TCK in the capture-DR state when the bypass register is selected. All non-specified instructions cause the bypass register to be selected, so that it is placed between TDI and TDO.

4.2.4.3 ID Code Register

The ID Code register is a 32-bit read-only register defined as:

TABLE 4-5 ID Code Register

Version Field	Part Number Field	Manufacturing ID Field	lsb
[31:28]	[27:12]	[11:1]	[0]
Initially 0x0; updated per BSDL change	0x2aaa	0x03e	1

The ID Code register is always placed between TDI and TDO when the select-DR state is reached directly after the test-logic-reset state with no intervening instruction register programming. The lsb is closest to the TDO as required by the standard.

4.2.4.4 CMP Data Registers

Access for all CMP registers will be via UCB (TAP_CREG_ and TAP_NCU_ instructions).

Threads in each core are virtual cores; for those CMP registers specifying physical cores each physical core is assigned eight bits in a 64-bit register; allowed values are 8'b11111111 and 8'b00000000 The assigned eight bits are [63:56] = core 7; [55:48] = core 6; [47:40] = core 5; [39:32] = core 4; [31:24] = core 3; [23:16] = core 2; [15:8] = core 1; [7:0] = core 0.

4.2.5 JTAG SCK Bypass

To get around the SSI lock loss issue in OpenSPARC T2 (see bug 97461), NCU implements a down counter that would decrement to zero starting from the first iol2clk after flush reset completion (part of POR and Warm resets). After this counter expires (reaches 0 count), NCU will send out the first fetch on SSI interface by asserting SSI_MOSI. By this time the FPGA would have relocked against the SSI_SCK coming from OpenSPARC T2.

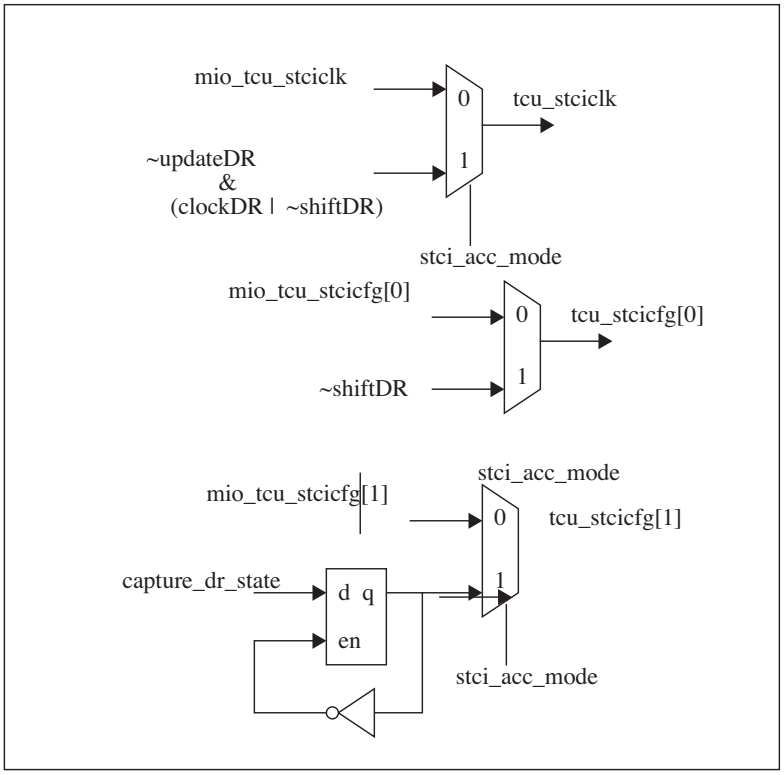
To remove this 5-6 msec wait on the tester, TCU supports a bit that is programmable through JTAG on the tester between deassertion of TRST_L and assertion of POR2 by the reset block. This bit preserves its value on POR2 and WARM resets and is sent by the TCU to the NCU as a signal tcu_sck_bypass. When this signal is a 1, (bit programmed to 1'b1 for the tester), the NCU bypasses the counter to assert SSI_MOSI (thereby eliminating the 5-6 msec wait time on the tester). If this signal is a 0, (POR1 reset state of the bit in TCU) then the NCU honors the counter and waits till the counter expires before asserting SSI_MOSI. This would be the configuration in the system.

The TAP_SCKBYP_ACCESS and TAP_SCKBYP_CLEAR instructions are used to set and reset, respectively, the tcu_sck_bypass signal sent from TCU to NCU.

4.2.6 JTAG Access to SERDES STCI

JTAG provides access to the SERDES STCI bus. There are four inputs, STCICLK, STCICFG[1:0], and STCID, and one output STCIQ. To enable JTAG access to STCI, the JTAG instruction TAP_STCI_ACCESS should be executed. During STCI JTAG Access, STCICLK and STCICFG are driven as shown in FIGURE 4-1, while STCID is connected to TDI and STCIQ to TDO. To clear JTAG access to STCI, use TAP_STCI_CLEAR or reset the TAP state machine. The updateDR, clockDR and shiftDR are as specified in the IEEE 1149.1 spec., while capture_dr_state is active during the Capture DR TAP state.

FIGURE 4-1 SERDES STCI Bus Control



4.2.7 JTAG Errata

4.2.7.1 JTAG Accesses to some Registers in CMP Clock Domain may result in Erratic Read Values.

Symptom: JTAG reads of some registers which exist in CMP clock domain may result in erratic read data, if the registers are being updated when the read occurs.

Description: There are several JTAG instructions that can sample data from registers contained in the CMP clock domain. These are sampled with TCK in the JTAG block without synchronization. Consequently, if those registers are changing in the cmp domain, the read results will be indeterminate. The indeterminism results from the

asynchronous nature of TCK with respect to the CMP clock. It is permissible to read these registers directly after they are written with JTAG, before their corresponding activity is initiated, to verify contents were correctly written.

For the first class of registers the data will generally be stable when reading, or at least changing very infrequently. Thus, it is recommended that when reading these registers, back to back reads be performed so that the contents can be compared; if the same, then the read is successful.

JTAG Registers in this class are indicated with a (1) in the Capture column of TABLE 4-4:

- Clock Domain[31:0]
- Core Run Status[63:0]
- MBIST Done[47:0]
- MBIST Fail[47:0]
- MBIST Result[1:0]
- LBIST Done[7:0]
- L2_Read_Data[64:0]
- Clock Stop Delay[6:0]
- Clock Status[1:0]
- Core Select[7:0]
- TCU Debug Control[3:0]
- DOSS Status[7:0]
- Cycle Step Enable[0]
- Cycle Step Status[0]
- JTAG POR Status[0]

For the second class of registers, the data may be changing every cycle. In this case it is recommended that the user refrain from reading these until the contents are stable - this can be determined by examination of the associated status registers.

Registers in this class are indicated with a (2) in the Capture column of TABLE 4-4

- Debug Event Counter[31:0]
- Cycle Counter[63:0]
- DMO Config[47:0]

Workaround: JTAG users should either 1) read the registers in the first group multiple times until two stable values are obtained, or 2) read the status registers associated with the registers in the second group to make sure hardware is not updating them.

4.2.7.2 JTAG View of some CSR Registers is Not Correct

Symptom: When accessing certain JTAG registers that are also accessible via SW as CSRs, there is a possibility that the JTAG view of the register can be inconsistent with the SW view. The SW view will always be correct, but the JTAG view may be incorrect.

Description: The TCU does not properly synchronize the update signal for writes to certain JTAG registers when written by SW. There are two copies of certain JTAG registers, one in the TCK clock domain inside JTAG, and one in the IO clock domain which services UCB access by SW. These registers are supposed to be coherent, but if SW writes to one of these registers it may be possible for the JTAG register to miss the data and be incoherent, or even to update with indeterminate data.

The registers affected by this condition (with their bit positions relative to JTAG) are:

- MBIST Mode[3:0]
- MBIST Bypass[47:0]
- MBIST Abort[0]
- LBIST Mode[1:0]
- LBIST Bypass[7:0]

One way that these registers can become incoherent is when SW writes its copy, the logic in JTAG also tries to write the same data into the JTAG register. But, the write pulse is not synchronized and may be missed depending on the relative frequencies of TCK and the IO clock. Hence, it is possible that the JTAG register will not be updated, or it may be updated (corrupted) with indeterminate data.

A second way that these registers can become incoherent is if the TCK is not running. In order to maintain coherency, the TCK clock must be running. However, the JTAG clock - TCK - is not required to run for functional operation, so there is no guarantee that TCK is even active when SW writes its register. So the JTAG version of the register is not updated.

Workaround: Given these problems, and since it is always the JTAG view which may be incorrect, the JTAG user should be aware that if SW writes these registers then the JTAG view may be incorrect.

4.2.7.3 HIGH-Z Boundary Scan Instruction is Not Supported

Symptom: Pins associated with DBG_DQ bus in the MIO have weak pullups or pulldowns, and thus do not go to a high-z state when instructed to do so. In addition, SERDES pins do not support the high-z state.

Description: The JTAG TAP_HIGHZ instruction is intended to put all output pins in a high-z state (tristate) and is typically used during manufacturing board test to prevent overdriving signals. Due to the use of SERDES macros which do not support

the high-z capability, and the use of pullups and pulldowns in the DBG_DQ bus pins, many of the OpenSPARC T2 output pins can not go to a high-z state even when instructed to do so by the TCU.

Workaround: Do not use the TAP_HIGHZ instruction. Support for the JTAG TAP_HIGHZ instruction has been moved from the public to the private section of the boundary scan description language file (BSDL); this is legal since HIGHZ is optional per the IEEE 1149.1 standard. The actual JTAG instruction, TAP_HIGHZ, has been kept in place but is now a private instruction. This was done since the RTL still supports the TAP_HIGHZ instruction. Moving it to private means it is not a supported public instruction and should not be used.

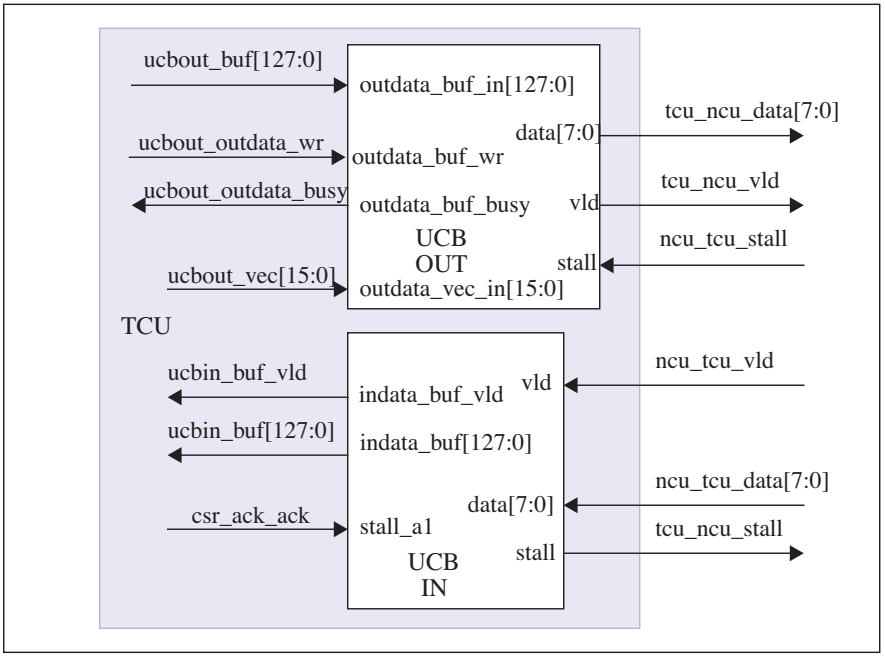
4.3 UCB Interface

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 IO mapped registers, and some ASI 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 OpenSPARC T2, UCB access through the crossbar to the l2 and cores is not available so access to the L2 is done via a separate interace 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.

4.3.1 UCB Simple Block Diagram

FIGURE 4-2 UCB Interface Inside the TCU



4.3.2 JTAG Instructions used to Access the UCB

The following descriptions are excerpts from the *OpenSPARC T1 DFT Specification* and the *OpenSPARC T1 DFT User's Guide* but have been ported to OpenSPARC 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.

4.3.2.1 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 IO CLK domain.

4.3.2.2 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 one. To use this, the JTAG is kept in ShiftDR and TCK is clocked until the TDO reads a "1", this indicates the sentinel bit has been set. When the sentinel bit becomes one, the next 64 bits shifted out are the valid read data.

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.

4.3.2.3 TAP_NCU_WRITE

Initiate Write Transaction: Causes a write transaction to be initiated on Update-IR.

4.3.2.4 TAP_NCU_READ

Initiate Read Transaction: Causes a read transaction to be initiated on Update-IR

4.3.2.5 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 is forwarded to the UCB interface and a write transaction is initiated. This instruction is a combination of TAP_CREG_ADDR and TAP_NCU_WRITE.

4.3.2.6 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.

4.3.2.7 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 is forwarded to the UCB interface and a read transaction is initiated. This instruction is a combination of TAP_CREG_ADDR and TAP_NCU_READ.

4.3.3 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.

4.3.4 TCU as a Slave for UCB

The OpenSPARC T1 implementation provided only that TCU be a master for UCB interactions. To support debug requirements for OpenSPARC T2, the TCU will also act as a slave for UCB. The interface remains the same, the only changes will be in the TCU. For joint access between JTAG and UCB the result is indeterminate. The list of registers accessible via SW inside the TCU is provided in TCU Local CSR Assignments.

Reading local TCU CSRs via JTAG UCB protocol is not supported; local TCU CSRs should be accessed directly via the appropriate JTAG instructions. Note that the register bit ordering may not be consistent between both methods.

TCU is not designed to handle burst read requests, that is, a read request cannot be followed immediately by another 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 read request after the data for the first

one has returned. In the case of multiple threads accessing TCU CSR registers, some mechanism (such as a semaphore lock) should be used to guarantee only one thread accesses any TCU CSR register at a given time. See UCB Erratum.

Note – Read requests to internal non-existent TCU CSRs (base address 85_0000_0000) cause TCU to respond with a READ_ACK instead of a READ_NACK. This means that TCU responds with garbage data. The requesting thread doesn't hang. Writes to undefined CSR addresses within TCU appear to complete but do not write to any real TCU registers; this is expected behavior. Reads appear to complete also but the READ_ACK is not correct behavior since TCU should respond with a READ_NACK. This behavior will not be fixed. A workaround is to never allow software to request data from undefined TCU CSRs. Software should take care to access only valid TCU register addresses.

4.3.5 UCB Erratum

4.3.5.1 TCU UCB Hangs on Reads from SPARC Core

Symptom: A thread hangs while reading a TCU CSR.

Description: Multiple threads can access CSRs inside the TCU and cause the NCU to send back-to-back reads to the TCU. As the TCU sends data back in response to these requests, the NCU may become overloaded and stall the TCU. If the stall lasts for more than one cycle, this will cause the TCU to drop any request that the NCU is sending at the time of the stall.

The TCU is not designed to handle burst read requests where the NCU has to stall the TCU for more than one cycle to receive the data the TCU is returning.

The problem is that the TCU does not provide a buffer for catching incoming requests when stalled, and drops the incoming request. The TCU does not hang, but the thread that issued the request will hang since its request is not serviced by TCU.

This issue only appears with multiple threads. A single thread cannot issue back-to-back read requests since it will always wait for return data before issuing the next request. Only multiple threads can send read requests which appear back-to-back to the NCU and TCU.

Also, this issue requires several back-to-back reads to cause enough activity for the NCU to stall the TCU. Simulation shows that for four back-to-back reads, the NCU stalls the TCU on the third request. Consequently, the TCU drops the fourth request.

Workaround: Users should program the second TCU CSR read request to wait until after the data for the first read request has returned. In the case of multiple threads accessing TCU CSR registers, some mechanism (such as a semaphore lock) should be used to guarantee only one thread accesses any TCU CSR register at a given time.

4.4 L2 Access via SIU

4.4.1 JTAG L2 Access Registers

It is possible to write and read the L2 addresses while the chip is running using JTAG.

TABLE 4-6 L2 Access Registers

Register	JTAG Instr.	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 L2a	Ignored
L2_Read_Data[64:0]	TAP_L2_RD	bits[64:1] = 8-bytes of Data returned from L2	1 = Data Valid

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. The L2_Write_data and L2_Read_Data registers are the same physical register.

The TCU to L2 interface thru SIU is 4B aligned. The SIU will force bit 2 = 0.

4.4.2 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 (0xC) 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. The RTI state should be avoided except for the actual transfer of data, and once entered should not be reentered during the write operation.

4.4.3 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. Note: bit 0 is not the same as the "sentinel bit" of the creg access. Note: The RTI state should be avoided except for the actual transfer of data, and once entered should not be reentered during the read operation.

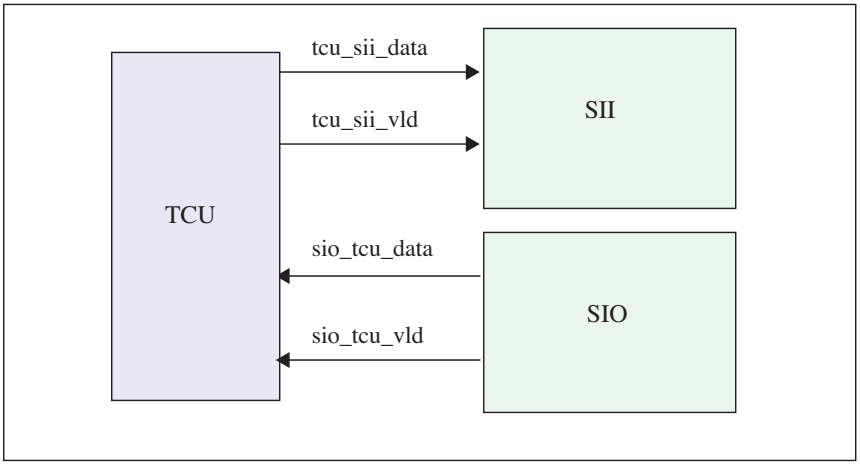
Only one write or read may be outstanding at any time. Also, since non-JTAG logic is used the POR reset sequence should be performed before using this feature (or at least the POR1 section of the reset sequence).

4.4.4 Diagram

The signals used between TCU and SIU are:

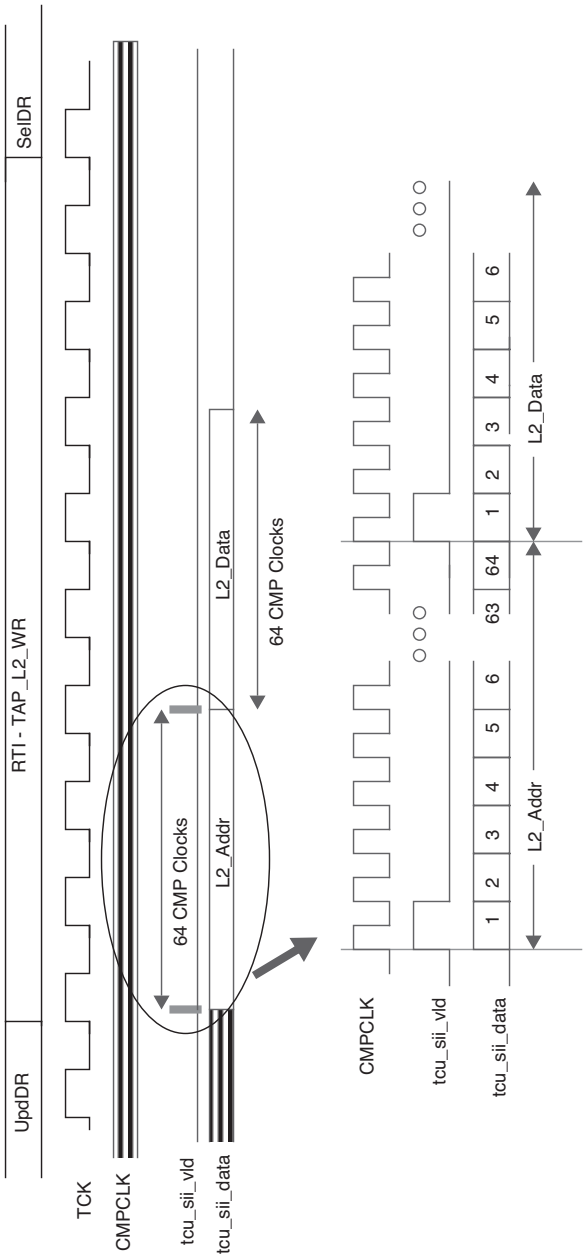
- tcu_sii_data - Sends L2_Addr[64:1] as address followed by L2_Write_Data[64:1] as data to SIU (Data only present for Write, absent for Read). Bit 1 is sent first for both address and data. Output from TCU.
- tcu_sii_vld - Pulsed when bit 1 of L2_Addr or L2_Write_Data goes onto tcu_sii_data. Output from TCU.
- sio_tcu_data - Input to TCU containing data returned from a Read request, bit 0 first.
- sio_tcu_vld - Input to TCU, pulsed when bit 0 is on sio_tcu_data

FIGURE 4-3 TCU Interface with SIU



A sample waveform is shown in FIGURE 4-4.

FIGURE 4-4 JTAG Write to L2 via SIU - Waveform



4.5 Scan

4.5.1 Manufacturing Scan

There are 32 scan chains in the chip available in parallel during manufacturing (automatic test pattern generation (ATPG) scan mode. In manufacturing scan mode they will be accessed through the 32 scan in and 32 scan out pins. These 64 pins will be shared with functional pins; the dedicated testmode pin is used to configure the pins as scan inputs and scan outputs.

Clocks in this mode are provided by the tester (the PLL is not used) but will be multiplexed onto the clock domain trees outside of the TCU, under control of the pll_bypass and testmode pins. In order to allow tester control of clock domains individually during ATPG test, most clock domains will have separate pin control. SERDES is a special issue; the SERDES logic will have some of their configuration signals sourced from the pins. Also, even though the TCU might provide a scan clock the SERDES will still need two test clocks from the pins for transmit and receive timing information in the manufacturing test modes that TI requires.

TCU participates passively in manufacturing scan; during manufacturing scan the JTAG logic and the TCU itself is included in one of the 32 scan chains and is testable via ATPG patterns.

TABLE 4-7 Manufacturing Parallel Scan Chains

Chain	Contents	TCU Input	TCU Output
0	SPC 0 internal chain 0	spc0_tcu_scan_in[0]	tcu_spc0_scan_out[0]
1	SPC 0 internal chain 1	spc0_tcu_scan_in[1]	tcu_spc0_scan_out[1]
2	SPC 1 internal chain 0	spc1_tcu_scan_in[0]	tcu_spc1_scan_out[0]
3	SPC 1 internal chain 1	spc1_tcu_scan_in[1]	tcu_spc1_scan_out[1]
4	SPC 2 internal chain 0	spc2_tcu_scan_in[0]	tcu_spc2_scan_out[0]
5	SPC 2 internal chain 1	spc2_tcu_scan_in[1]	tcu_spc2_scan_out[1]
6	SPC 3 internal chain 0	spc3_tcu_scan_in[0]	tcu_spc3_scan_out[0]
7	SPC 3 internal chain 1	spc3_tcu_scan_in[1]	tcu_spc3_scan_out[1]
8	SPC 4 internal chain 0	spc4_tcu_scan_in[0]	tcu_spc4_scan_out[0]
9	SPC 4 internal chain 1	spc4_tcu_scan_in[1]	tcu_spc4_scan_out[1]
10	SPC 5 internal chain 0	spc5_tcu_scan_in[0]	tcu_spc5_scan_out[0]

TABLE 4-7 Manufacturing Parallel Scan Chains *(Continued)*

Chain	Contents	TCU Input	TCU Output
11	SPC 5 internal chain 1	spc5_tcu_scan_in[1]	tcu_spc5_scan_out[1]
12	SPC 6 internal chain 0	spc6_tcu_scan_in[0]	tcu_spc6_scan_out[0]
13	SPC 6 internal chain 1	spc6_tcu_scan_in[1]	tcu_sp6_scan_out[1]
14	SPC 7 internal chain 0	spc7_tcu_scan_in[0]	tcu_spc7_scan_out[0]
15	SPC 7 internal chain 1	spc7_tcu_scan_in[1]	tcu_spc7_scan_out[1]
16	CCX[0], SII	soca_tcu_scan_in	tcu_soca_scan_out
17	CCX[1], MCU0	socb_tcu_scan_in	tcu_socb_scan_out
18	MCU 1:2, SIO	socc_tcu_scan_in	tcu_socc_scan_out
19	DMU	socd_tcu_scan_in	tcu_socd_scan_out
22	NCU, MCU3	socg_tcu_scan_in	tcu_socg_scan_out
23	L2B 0:7	soch_tcu_scan_in	tcu_soch_scan_out
24	L2T 0:1, L2D 0:1	soc0_tcu_scan_in	tcu_soc0_scan_out
25	L2T 2:3, L2D 2:3	soc1_tcu_scan_in	tcu_soc1_scan_out
26	L2T 4:5, L2D 4:5	soc2_tcu_scan_in	tcu_soc2_scan_out
27	L2T 6:7, L2D 6:7	soc3_tcu_scan_in	tcu_soc3_scan_out
30	TCU, DB1, DB0, MIO, EFU, RST, CCU, BScan	soc6_tcu_scan_in	tcu_soc6_scan_out
31	SERDES Macros	srd_tcu_atpgq	tcu_srd_atpgd

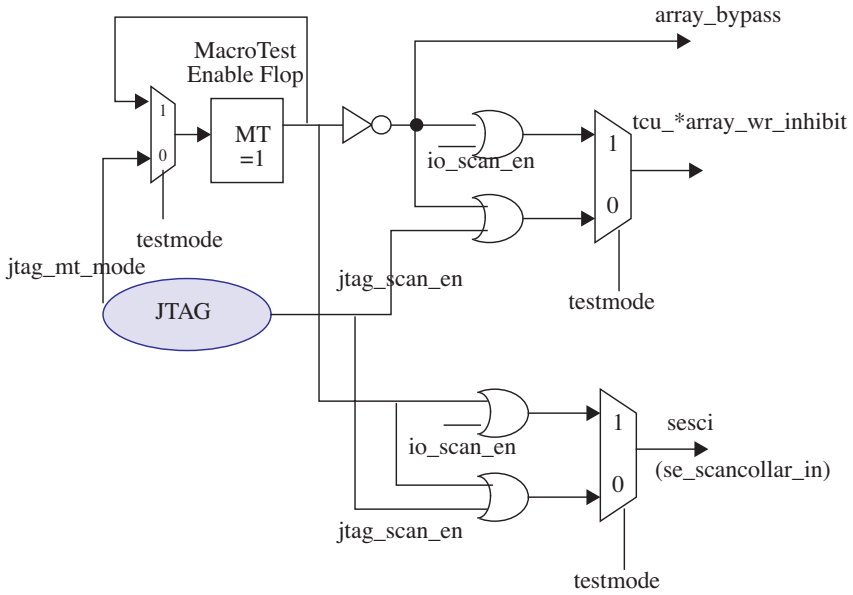
Clusters are ordered as shown with scan-in feeding the left-most block. Boundary scan cells inside the MIO and MCU clusters are included in chain 30. The BScan (boundary scan) chain is ordered from scan-in (TDI) to scan-out (TDO) as MCU0-MCU1-MCU3-MCU2-MIO.

4.5.2 MacroTest Scan

MacroTest on OpenSPARC T2 is primarily a subset of manufacturing scan. During MacroTest mode, control of the array write inhibit signal, scan enables for scan collar input (sesci) flops for arrays, and the array_bypass signals is different than non-macrotest scan. This additional control allows arrays to be accessed and written to or read from via scan. To enable MacroTest mode, a control flop for array write inhibit, sesci and array_bypass must be set. For manufacturing scan this may be controlled with cell constraints.

MacroTest on OpenSPARC T2 is also be a subset of serial (JTAG) scan. To set JTAG MacroTest mode, the instruction TAP_MT_ACCESS should be programmed; this will set the MacroTest enable flop (default is off). The instruction TAP_MT_SCAN can then be used to perform MacroTest scan accesses. To clear the MacroTest enable flop, use TAP_MT_CLEAR. This mode exists solely to satisfy debug requirements for scan access to arrays.

FIGURE 4-5 Signals Controlled for Macrotest (in TCU)



JTAG MacroTest is used extensively in debug to access the arrays, and to allow control using JTAG. The PLL is bypassed to allow TCK to be placed onto the clock tree during MacroTest mode. Before entering JTAG MacroTest mode the clocks to the chip should be stopped via a hard stop since TCK will need to be routed onto the gclk distribution. JTAG MacroTest will access all clock domains, there is no user control over individual domains.

4.5.2.1 Procedure for Entering JTAG MacroTest

Because the JTAG MacroTest must be run with the PLL locked, a special sequence is used to enter this mode. This sequence puts the chip in JTAG MacroTest mode while not disrupting the CCU (PLL), TCU or RST blocks.

1. Lock PLL: POR sequence (optional; can run in pll bypass mode with slow clock)
2. Run Diag (optional)
3. Stop Clocks: TAP_CLOCK_HSTOP or via Debug Event
4. Set Test Protect: TAP_TP_ACCESS
5. Set Macro Test Mode: TAP_MT_ACCESS
6. Set Chain Select if desired: TAP_CHAINSEL
7. Perform JTAG SCAN: TAP_MT_SCAN

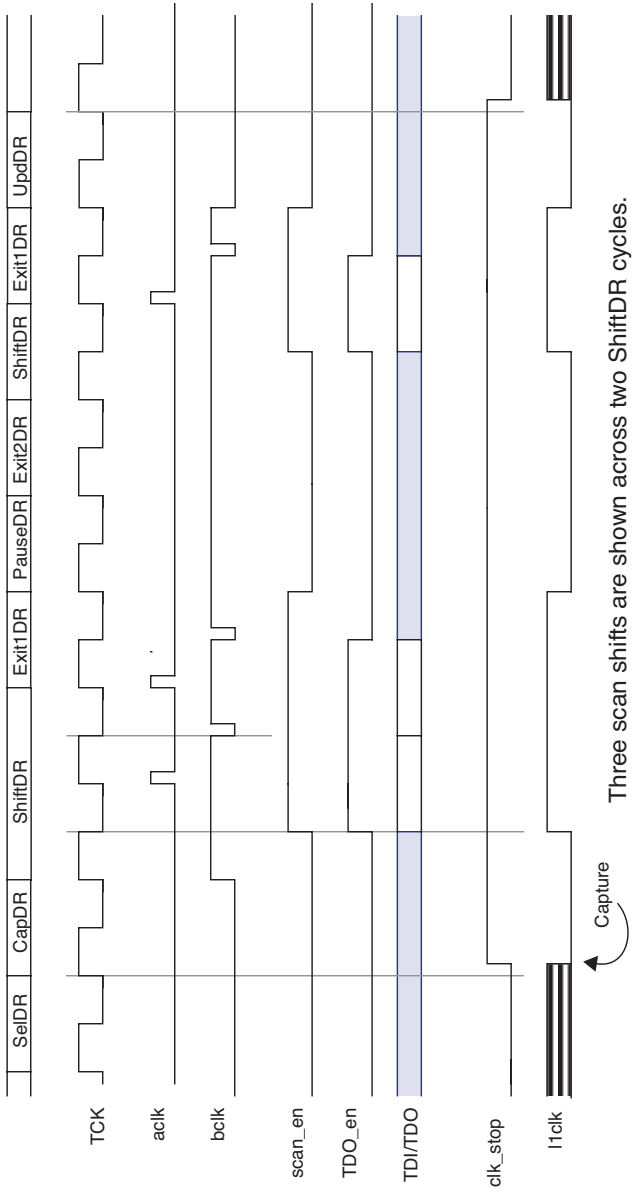
4.5.3 Serial Scan

Serial scan refers to concatenating the 32 internal scan chains into a single chain primarily for observation during debug. Serial scan is initiated via JTAG instructions, where the scan chains are configured into a single long chain and placed between TDI and TDO.

Two JTAG instructions are available to control serial scan; the TAP_SERSCAN instruction places all 32 scan chains between TDI and TDO excluding the TCU, CCU and RST blocks. The chains will be concatenated in the order specified in TABLE 4-7. One of the 32 scan chains may also be selected via TAP_CHAINSEL;

During serial scan the scan clocks (aclk, bclk) are generated from the leading and trailing edges of TCK during ShiftDR. The scan enable signal drives the l1clk to '1'; prior to unloading a scan chain with JTAG the clock should be stopped to that chain's clock domain(s) using the JTAG clock stop instructions. The PLL is locked and running typically during serial scan but serial scan does not rely on the PLL. There is no ability to perform capture clocks during serial scan. A sample waveform is shown in FIGURE 4-6.

FIGURE 4-6 JTAG Serial Scan Sample Waveform



4.5.3.1 Chain Select Register

In order to observe and update the state of OpenSPARC T2, parallel scan chains 0 to 31 can be put between TDI and TDO by programming the TAP_SERSCAN instruction. The behavior is qualified by the Chain Select register. The internal scan register consists of the chains specified in Table 6-1 on page 22. This is a 6-bit register selected via the TAP_CHAINSEL instruction. It is only recognized when the TAP_SERSCAN (or TAP_MT_SCAN) instruction is programmed and allows one of 32 scan chains in OpenSPARC T2 to be selected with all others bypassed if the msb is set to '1'. If the msb (bit 5) is 0 then the chain selection field is ignored and chains 0 to 30 are concatenated during the serial scan operation.

This register has no effect on chain 31 which is reserved for SERDES scan flops, which means that chain 31 is not accessible via JTAG serial scan. Scan chains for Boundary Scan, TCU, CCU and RST units are not accessible via JTAG serial scan; when chain_sel[4:0]=30 only the peu, mio, efu, db0 and db1 scan flops will be returned.

The Chain Select register is reset with TRST_L or entering TLR. The chain selection field is directly decoded to specify a chain from 0 to 30. Chain 31 is not selectable. Selecting either chain in a SPC core will result in both SPC scan chains being concatenated. The MBIST and shadow scan chains in each SPC will be concatenated to that SPC's chain[0] and chain[1], respectively.

Note – The length of the scan chain during JTAG serial scan will change between POR1/POR2 and WMR1/2. The chain will be longer during POR1/POR2 due to inclusion of MCU logic as shown in the next subsection "Logic Included in JTAG Serial Scan". Use of JTAG POR (See JTAG Access During POR) pauses the POR sequence during POR2 and the MCU logic will be included during JTAG serial scan. A subsequent warm reset will move the chip out of POR2 and cause the MCU FBD logic to be excluded from JTAG serial scan.

Note – Unpredictable behavior will result if a JTAG (tap) reset is initiated after scandumping any of the soc chains (this includes the "all chain" scan dump mode) since the TAP_TP_ACCESS command will be reset (See Protecting TCU During Serial Scan: Test Protect Mode). This will expose the logic protected by TAP_TP_ACCESS to any random data scanned during the SOC scan process, with indeterminate results.

TABLE 4-8 Chain Select Register

Enable Bit	Chain Selection Field
bit [5]	bits [4:0]
Enables the chain selection field when set to '1'	If enabled, specifies one of 31 chains to be placed between TDI and TDO
	x_xxxx -> Selects chains 0-30 concatenated
0	
	0_0000 -> Selects chain 0 & 1 of SPC0
1	0_0001 -> Selects chain 0 & 1 of SPC0
1	0_0010 -> Selects chain 0 & 1 of SPC1
1	0_0011 -> Selects chain 0 & 1 of SPC1
1	0_0100 -> Selects chain 0 & 1 of SPC2
1	0_0101 -> Selects chain 0 & 1 of SPC2
1	0_0110 -> Selects chain 0 & 1 of SPC3
1	0_0111 -> Selects chain 0 & 1 of SPC3
1	0_1000 -> Selects chain 0 & 1 of SPC4
1	0_1001 -> Selects chain 0 & 1 of SPC4
1	0_1010 -> Selects chain 0 & 1 of SPC5
1	0_1011 -> Selects chain 0 & 1 of SPC5
1	0_1100 -> Selects chain 0 & 1 of SPC6
1	0_1101 -> Selects chain 0 & 1 of SPC6
1	0_1110 -> Selects chain 0 & 1 of SPC7
	0_1111 -> Selects chain 0 & 1 of SPC7
1	
1	1_0000 -> Selects chain 16
1	1_0001 -> Selects chain 17
1	1_0010 -> Selects chain 18
1	1_0011 -> Selects chain 19
1	1_0100 -> Selects chain 20
1	1_0101 -> Selects chain 21
1	1_0110 -> Selects chain 22
1	1_0111 -> Selects chain 23
1	1_1000 -> Selects chain 24
1	1_1001 -> Selects chain 25
1	1_1010 -> Selects chain 26
1	1_1011 -> Selects chain 27
1	1_1100 -> Selects chain 28
1	1_1101 -> Selects chain 29
1	1_1110 -> Selects chain 30
	1_1111 -> Ignored

4.5.3.2 Logic Included in JTAG Serial Scan

During JTAG serial scan this logic is included:

- MBIST engines
- MCU FBD logic (during JT POR access in POR2)
- Shadow scan - both spc & l2t
- Unavailable SPC cores and Banks

This logic is NOT included during JTAG serial scan:

- Cluster headers
- MCU FBD logic (when not in POR2 - i.e., during diag scan dumps)
- Boundary scan flops
- Any flops non-scannable for manufacturing scan
- CCU, RST and TCU
- SERDES (chain 31)

4.5.3.3 Protecting TCU During Serial Scan: Test Protect Mode

When JTAG serial scan is performed, random signals can be generated to TCU inputs. If the TCU responds to these they can disrupt the JTAG serial scan; for example random debug requests from a SPC while it is being scanned can disrupt the TCU JTAG scan process. To protect against this it is up to the user to tell TCU to protect itself. Two JTAG instructions are available for this: TAP_TP_ACCESS to set the Test Protect mode, and TAP_TP_CLEAR to clear it. Setting the Test Protect mode will cause TCU to assert a signal tcu_test_protect which will block incoming SPC debug requests and incoming UCB requests. This signal also goes to RST and CCU and other blocks which need to block random UCB requests which may occur when scanning the SOC blocks (specifically NCU). This mode is also needed during MBIST scan operations, and possibly LBIST scan operations. The expected usage is to set the Test Protect mode before performing the test operations, and then clear it when done. For Transition Test and Macro Test the mode should be set via scan operations if needed.

Setting Test Protect mode should not interfere with PLL lock, but it may interfere with diags trying to change clock frequency or generate resets.

Note – When accessing any scan chains via JTAG mode, the TAP_TP_ACCESS protocol should be followed. This includes TAP_MBIST_DIAG and TAP_LBIST_ACCESS.instructions as well as variations of TAP_SERSCAN or TAP_MT_SCAN.

4.5.4 SERDES Scan

TCU supports scan for SERDES by connecting the SERDES macros onto chain 31. Package pin SCAN_IN31 becomes tcu_srd_atpgd and connects to ATPGD of fsr0, and then ATPGQ and ATPGD are connected to daisy-chain the SERDES macros, with ATPGQ of fsr4 connecting through TCU tsrd_tcu_atpgq to SCAN_OUT31. The scan-enable signal for SERDES is tcu_srd_atpgse and is driven by TCU from package pin io_scan_en. The following outlines the mode under which SERDES operates; bits [1:0] are accessible only via scan and bit [2] is driven directly from io_test_mode. The ATPG mode bus tcu_srd_atpgmode[2:0] ultimately connects to the TESTCFG[18:16] of the SERDES macros.

- tcu_srd_atpgmode[2:0]
  - 000: for normal operation
  - 001, 010, 011: reserved
  - 100: for stuck-at ATPG
  - 101: to select 2-clock transition test
  - 110: to select 3-clock transition test
  - 111: to select 4-clock transition test

4.6 Clock Stop

On OpenSPARC T2 the ability to stop clocks to various sections of the chip is provided via the TCU. Clocks can be stopped via JTAG directly or as a result of a debug or other event. Clocks are also stopped before any flush reset and then restarted after the flush reset is finished.

There are two modes of clock stops: a hard clock stop and a soft clock stop. The purpose of the hard clock stop is to stop as fast as possible, without waiting for the chip to become quiet. The 2nd method, soft clock stop, only applies to the cores and upon receiving a request the TCU will wait for the requesting core to settle into a quiescent state (via the core_running register) before stopping the clock to that core. Multiple cores may also be stopped this way. This allows the core(s) the possibility to restart after clocks are restarted. Clocks for the chip can be stopped either in parallel or serially across clock domains. After a clock stop, data can then be shifted out for debug via JTAG which allows the user to determine the state of the chip.

4.6.1 Serial and Parallel Clock Stop Modes

Stopping all clock domains in parallel may not be advisable due to excessive current fluctuations across the chip. Because of these di/dt concerns there is a serial clock stop mode where the clocks are stopped over several predefined clock domains with 128 cpu clock cycles between each clock stop activation. Stopping the clocks in such a staggered fashion with intervening delays is expected to lessen the di/dt concern. In the serial mode, via JTAG or software the user can update a clock domain register to specify which clock domain should be stopped first. Subsequent domains will then be stopped in a predetermined order, but the order is fixed.

During a parallel clock stop, the clocks will all be stopped relative to the same cpu clock cycle from the TCU. For both the serial and parallel clock stop methods, due to division ratios between the cpu and other clock domains, the actual cpu clock cycle at which a non cpu clock domain stops may vary between those domains, although it should be repeatable. To specify a parallel stop, all bits in the clock domain register should be set to 1, signifying they should all stop first.

Specification of serial or parallel is controlled by setting the 32-bit Clock Domain register with JTAG TAP_CLOCK_DOMAIN instruction, ordered as specified TABLE 4-9 (bit == stop number). Setting only one bit indicates the starting point for serial stopping. If serial and parallel clock stop modes are mixed, that is multiple bits are set in the clock_domain register, the clocks will stop in both serial and parallel across the specified bit fields. Originally tcu only supported either sequential or parallel without mixed modes but flexibility was given to allow the modes to work together. This results in a mixed behavior, with all bits set to'1' stopping in parallel, but with sequential stop behavior across the remaining fields of'0' bits. The user should consider this when programming mixed serial and parallel clock stopping.

Because the ability to stop selected domains in parallel would mainly be used for scan dump purposes, it doesn't matter if the remaining bits stop sequentially as described since the object of the scan dump should be in the domains that are set to stop in parallel.

4.6.2 Hard Clock Stop

A hard clock stop request will result in the clocks being stopped without waiting for the chip to acquiesce. The clocks may be stopped either in serial or parallel mode. In all cases the clocks will be stopped over all the chip as specified by the Clock Domain register, except the RST, CCU and TCU clocks will not be stopped.

A hard clock stop may be initiated in response to a flush request, a specific JTAG request via TAP_CLOCK_HSTOP, or in response to a debug event. A status register can be polled via TAP_CLOCK_STATUS to determine the state of the clock

sequencer. When the status indicates clocks are stopped, a scan dump via TAP_SERSCAN can be done. To restart clocks, the JTAG TAP_CLOCK_START is used.

4.6.3 Soft Clock Stop

A soft clock stop request will not be serviced until the core requesting the soft clock stop is acquiesced. The cores are the only clusters that can request a soft clock stop, and only the clocks to the target cores will be stopped by any soft stop request. A soft clock stop may be initiated in response to a JTAG request via TAP_CLOCK_SSTOP or in response to a debug event. Via JTAG, multiple cores can be soft-stopped. A debug event can only stop a single core as a default, however, setting bit 3 of the TCU DCR causes the TCU to soft stop all enabled SPCs if any requests a soft stop; See TCU Debug Control Register.

To request a soft clock stop via JTAG the target core(s) should first be parked with UCB access to the 64-bit Core Run register; the user is responsible for setting all eight bits per each core to be parked to '0'. When status is indicated via TAP_CORE_RUN_STATUS that the target threads are parked (64-bits), the clocks may be stopped via TAP_CLOCK_SSTOP.

For example, to acquiesce and stop SPC cores two and 5 only, first check with TAP_CORE_RUN_STATUS and then set the Core Run register to 64'hkkkk 00kk kk00 kkkk. This sets bits 23:16 for core 2, and 47:40 for core 5 to all 0's; 'k' means keep previous value read with TAP_CORE_RUN_STATUS. Check that the target threads are parked with TAP_CORE_RUN_STATUS. Set the target cores that should respond to a clock stop using TAP_CORE_SEL (in this case, cores two and 5) to set the core select register; any subsequent TAP_CLOCK_SSTOP will only stop clocks to cores two and 5. Note that TAP_CLOCK_START will clear the core select register.

In response to a debug event the requesting core's clock will be stopped similarly.

When the clocks are stopped a status register is set indicating the clocks are stopped. Polling of this status register can be done with TAP_CLOCK_STATUS to determine when it is safe to do a subsequent scan dump of the stopped cores. To restart the clock to the target core, the JTAG TAP_CLOCK_START is used.

The TAP_CORE_SEL instruction allows the user to enable cores to respond to a soft clock stop JTAG request using TAP_CLOCK_SSTOP and assumes all cores were already acquiesced. If it is used without acquiescing the cores you will get in effect a hard stop across only the cores. See Cycle Step Mode for usage of TAP_CORE_SEL.

When using the soft stop mode, the Clock Domain register should be all 0s if TCU DCR bit [2] is '1'. This applies to JTAG TAP_CLOCK_SSTOP and a request for Soft Stop by spc debug event when tcu_dcr[3] is set to 1 to stop all cores. In general, TCU DCR bit [2] should be '0' if any Soft Stop is used, otherwise the interaction between the Clock Domain and Core Select registers is complex.

The TCU clock sequencer - described in the next section - is controlled directly by the Clock Domain register. The Core Select register has no effect on the clock sequencer inside TCU, it will sequence independently of Core Select. The sequencer always runs through all 24 clock domains.

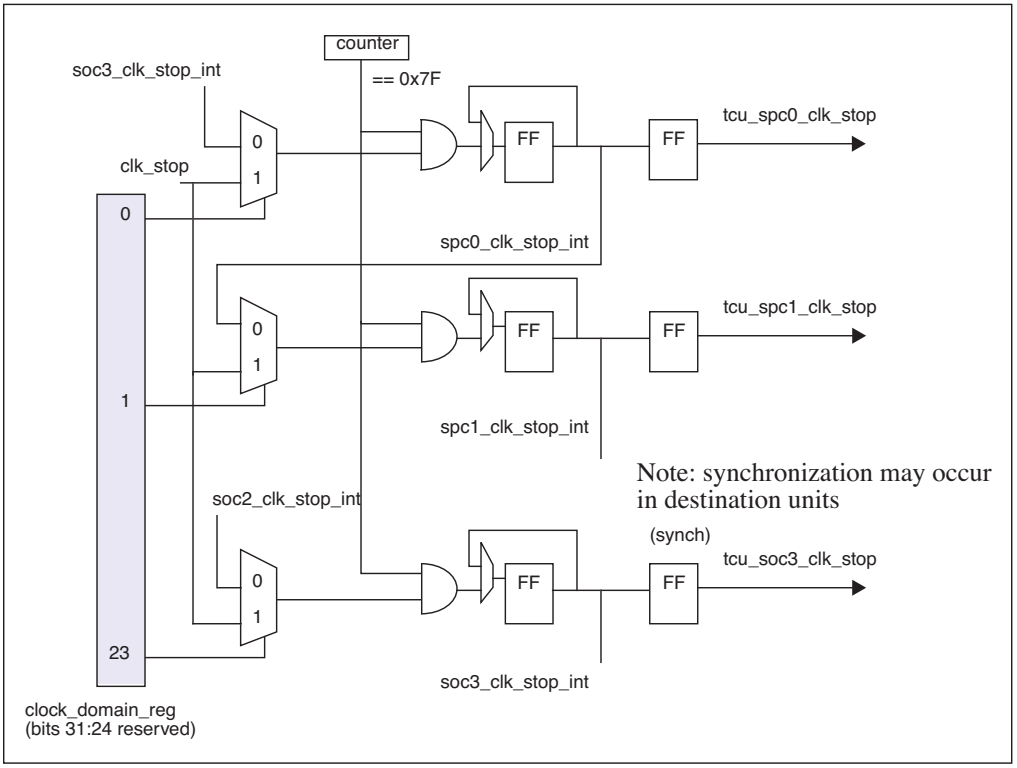
In soft stop mode, only the eight outputs associated with the spc cores are allowed to propagate. When Core Select is set and TAP_CLOCK_SSTOP issues, this begins the clock sequencer and it should be in a default mode starting with spc0 - this is either with Clock Domain all zeros or 24'b1. If the Clock Domain is something else, then it will tell the clock sequencer to begin at a different starting point, or if multiple bits are set the sequencer will stop clocks in parallel for those bits. The cores will still be stopped, but in an unexpected order.

For example, if clock_domain = 24'h000003 then spc 1 and spc 2 will be stopped in parallel and spc2 will be the starting point of the sequence - so spc0 will be stopped last. If clock_domain = 24'h000083, then spc7, spc1 and spc2 will stop in parallel.

4.6.4 Stop Domains

Clock domains are partitioned so that control is achieved for disabling sections of the chip with respect to the L2 and Core Enable/Available registers, and to minimize di/dt. The sequence of stopping the clocks serially will always be the same given a specific start point and defaults to the order given in TABLE 4-9. The user can program the starting point, but then the domains will stop in the predetermined order and wrap around until reaching the first domain stopped. For instance, stopping with spc1 first will result in spc0 being stopped last.

FIGURE 4-7 TCU Clock Sequencer



A seven-bit Clock Stop Delay counter (programmable via TAP_CLKSTP_DELAY) provides a delay of up to 128 (default) cmp clock cycles between generation of successive clock stop signals from the TCU. Setting this value to zero results in a one-cmp clock cycle delay between clock stop signals. This may be bypassed by setting bits [23:0] in the Clock Domain register via JTAG, so that all clocks stop in parallel.

The general structure of the clock sequencer control logic in the TCU is shown in FIGURE 4-7. To stop clocks starting with spc0, bit 0 of Clock Domain register is set to '1' with the remaining bits all to '0'. The clock sequencer state machine is in an initial state because clk_stop is low '0', then the clk_stop signal is activated and held to '1' to begin the sequence. When the Clock Stop Delay counter reaches 0xFF the tcu_spc0_clk_stop is set to '1' and held; when the counter next reaches 0xFF the tcu_spc1_clk_stop is set to '1' and the machine continues this sequence until all clock domains are disabled. When the clk_stop signal is driven to '0' the clk_stop signals are sequenced off starting with the domain specified in Clock Domain register.

The Clock Domain register is shown in TABLE 4-9.

TABLE 4-9 Clock Domain Register

Stop Number	Clock Domain Controlled	Stop Number	Clock Domain Controlled
0	SPC 0: cmp clock domain	12	Bank 4: L2 T, D, B: cmp clock domain
1	SPC 1: cmp clock domain	13	Bank 5: L2 T, D, B: cmp clock domain
2	SPC 2: cmp clock domain	14	Bank 6: L2 T, D, B: cmp clock domain
3	SPC 3: cmp clock domain	15	Bank 7: L2 T, D, B: cmp clock domain
4	SPC 4: cmp clock domain	16	MCU 0: cmp and io clock domains
5	SPC 5: cmp clock domain	17	MCU 1: cmp and io clock domains
6	SPC 6: cmp clock domain	18	MCU 2: cmp and io clock domains
7	SPC 7: cmp clock domain	19	MCU 3: cmp and io clock domains
8	Bank 0: L2 T, D, B: cmp clock domain	20	SOC0: sii, sio, ncu, efu: cmp and io clock domains. ccx; cmp clock domain. db0, db1, mio: io clock domain
9	Bank 1: L2 T, D, B: cmp clock domain	21	SOC1: rdp, mac, rtx, tds io and io2x clock domains
10	Bank 2: L2 T, D, B: cmp clock domain	22	SOC2: dmu: io clock domain
11	Bank 3: L2 T, D, B: cmp clock domain	23	SOC3: peu: io and pc clock domains

All clock stop control logic in the TCU is in the cmp clock domain. The outgoing signals, tcu_*_clk_stop, are sent from the cmp clock domain and are staged at the cpu level before reaching the cluster headers. The cluster header synchronizes the clk_stop into the corresponding clock domain. For clusters with io or io2x clock domains, the tcu_*_clk_stop is synchronized to the io clock domain before leaving TCU. This is done to provide transition test the capability of controlling the clock stop relative to the target domain. The dr clock domain clock stops are synchronized into the dr clock domain before leaving TCU, to mesh with the top-level dr staging flops.

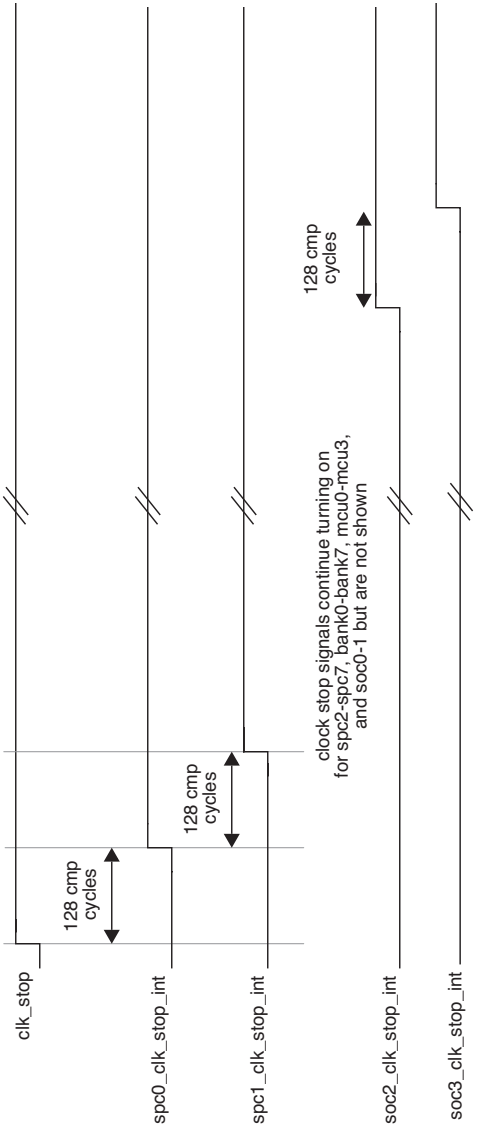
Most clusters with both cmp and io clock domains have separate clock stop signals from TCU, one for each domain. In some cases clusters with multiple clock domains share a single clock stop. The DB0 and DB1 clusters have both cmp and io domains, but share a clock stop synchronized to the io clock domain. The same holds for MIO. The RDP, RTX and TDP clusters have both io and io2x but share a clock stop synchronized to the io clock domain. The effect of this is that for clock stopping, the logic in these clusters will stop at different clock cycles. For example, in MIO during a clock stop the cmp logic will stop 3 (cmp) cycles later than other cmp logic.

- tcu_db0_clk_stop and tcu_db1_clk_stop => each is connected to cmp and io headers

- tcu_rdp/rtx/tdp_io_clk_stop => each is connected to io and io2x headers
- tcu_mio_clk_stop => connected to four cmp headers and one IO header

Clocks are restarted by turning off clk_stop signals. When started serially, the 128 cmp cycle delay is used again to reduce di/dt concerns.

FIGURE 4-8 Clock Stop Sequencing through Clock Domains



4.6.5 FBD Logic in MCU

The FBD logic in the MCUs is handled differently from other SOC logic. A separate clock stop signal is provided to each MCU, tcu_mcu[0123]_fbd_clk_stop, which is activated only during POR1 and POR2 (to facilitate flush reset of the FBD logic) or if the MCU is disabled via bank available or bank enable. During JTAG serial scan, the FBD logic is bypassed and left running. This is achieved with a second shared signal to all four MCUs, tcu_mcu_testmode, which is '1' only during POR1, POR2, or manufacturing ATPG testing.

If JTAG serial scan is performed while in POR2 then the MCU FBD logic will be included, so during the JTPOR access window (See JTAG Access During POR) the JTAG serial scan length will be longer than after POR2 completes.

4.6.6 Clock Stopping and Core/L2 Available and Disable Controls

4.6.6.1 Core and L2 Available Control

SPC cores are made unavailable if these signals are not asserted after transfer from EFU after POR1 or POR2:

- ncu_spc0_core_available
- ncu_spc1_core_available
- ncu_spc2_core_available
- ncu_spc3_core_available
- ncu_spc4_core_available
- ncu_spc5_core_available
- ncu_spc6_core_available
- ncu_spc7_core_available

L2 Logic (L2 Tag, L2 Data, L2 Buffer) can be made unavailable if the corresponding bits in this bus are not asserted after transfer from EFU after POR1 or POR2. Note that the L2 Tags will not have their clocks stopped even if listed as unavailable or disabled.

- ncu_tcu_bank_avail[7:0]

4.6.6.2 Core and L2 Disabling Control

SPC Cores can be disabled via Software after a warm reset with these signals:

- ncu_spc0_core_enable_status
- ncu_spc1_core_enable_status
- ncu_spc2_core_enable_status
- ncu_spc3_core_enable_status
- ncu_spc4_core_enable_status
- ncu_spc5_core_enable_status
- ncu_spc6_core_enable_status
- ncu_spc7_core_enable_status

L2 Banks (Two L2 Data and L2 Buffers along with associated MCU) can be disabled via Software after a warm reset with these signals:

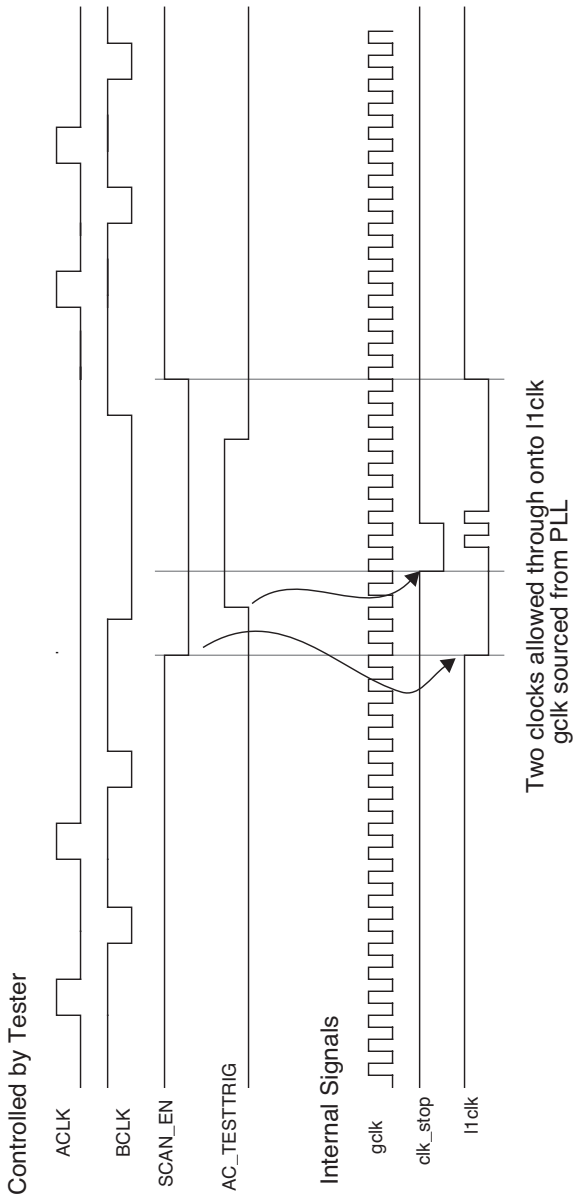
- ncu_spc_pm
- ncu_spc_ba01
- ncu_spc_ba23
- ncu_spc_ba45
- ncu_spc_ba67

There are certain legal combinations for the signals controlling disabling of SPCs and L2 Banks, for details refer to the *OpenSPARC T2 Programmer's Reference Manual*. Also, JTAG can be used to overwrite these values in certain cases such as using the JTAG POR access window or via JTAG UCB access. TCU looks at the available signals after POR1 or POR2 and deasserts clocks to any unavailable SPC or L2 (except L2 Tag clocks are not stopped since they contain top-level staging flops). This is shown in FIGURE 4-22. The disabling signals are observed by TCU after any WMR2, and clocks will be deasserted by TCU for any disabled SPC core or L2 bank and associated MCU (except for L2 Tags).

4.7 Transition Testing

Transition test on OpenSPARC T2 is designed to be run with the PLL locked to allow testing of the clock domains for transition faults. The patterns will be generated via the ATPG tools and applied on the wafer and chip testers. The testmode pin and ac_testmode pins must both be driven to '1' by the tester to enable transition test mode. The AC_TESTTRIG pin is used to tell the internal logic to allow a programmed number of l1clk cycles to reach the target flops.

FIGURE 4-9 Transition Test Sample Vector



4.7.1 Operation and Constraints During Transition Test

The transition test logic inside TCU is programmed via scan with a setup routine, or via cell constraints. A counter specifies how many system clock pulses are issued. This counter is based on the clock for the domain that is under test and is loaded with the number of clock pulses that will be issued to the functional logic. Only one clock domain may be active at one time since transition test across clock domains introduces non-determinism, and only the cmp and io clock domains are supported. During transition test the array_wr_inhibit signal from TCU is driven high into the clusters and overrides the write inhibit generated by the cluster headers from clock stop transitions.

When the mio_tcu_io_ac_testtrig input to TCU (package pin AC_TESTTRIG) is driven to '1' by the tester it is synchronized and used to enable the clock stop counter. The clk_stop signal to the target clock domain and cluster(s) is generated by TCU and pipelined out to the target cluster headers, and held to '0' for the programmed number of clock cycles. The counter is 8-bits and allows up to 255 clock pulses to be issued. The transition test control bit used to select the clock domain (cmp or IO) is one bit.

- tcusig_ttclksel_reg
  - 0 => selects cmp clock domain (default)
  - 1 => selects io clock domain

The control register, the 8-bit counter, and the flops driving the clock_stop signals all must be set to the appropriate values before each transition test capture cycle. The 8-bit counter values are true binary representations for cmp clock domain, and should be set to a multiple of four for the io clock domain. So to get two io clock cycles during the capture phase, the counter should be set to binary 0000_1000. The counter will start counting aligned to the io sync enable pulses during both cmp and io clock domain testing to achieve accurate clock counting during the io clock domain tests. A counter value of zero is not supported for transition test.

To set the clock stop flops in the tcu, the user should scan in values of zero to all clock stop flops, except set a one to flop _0 of all targeted domains. The values are inverted onto the clock stop signals. There are 24 clock domains as shown in TABLE 4-9. Bit 0 is closest to scan-in. Each clock domain has two flops associated with it and MCU and SOC have extra flops for io and dr clock domains. Flops that can be set to activate clock pulses to a domain are indicated. To select an io clock domain, set the corresponding _0 flop and also set the transition test control bit to '1'.

- sync_ff_clk_stop_spc0_0Set to '1' for SPC0 cmp clock domain
- sync_ff_clk_stop_spc0_1
- ...
- sync_ff_clk_stop_spc7_0Set to '1' for SPC7 cmp clock domain
- sync_ff_clk_stop_spc7_1

- sync_ff_clk_stop_bnk0_0Set to ‘1’ for BNK0 cmp clock domain
- sync_ff_clk_stop_bnk0_1
- sync_ff_clk_stop_l2t0_0Set to ‘1’ for L2T0 cmp clock domain
- sync_ff_clk_stop_l2t0_1
- ...
- sync_ff_clk_stop_bnk7_0Set to ‘1’ for BNK7 cmp clock domain
- sync_ff_clk_stop_bnk7_1
- sync_ff_clk_stop_l2t7_0Set to ‘1’ for L2T7 cmp clock domain
- sync_ff_clk_stop_l2t7_1
- sync_ff_clk_stop_mcu0_0Set to ‘1’ for MCU0 cmp or io clock domain
- sync_ff_clk_stop_mcu0_1
- sync_ff_ioclk_stop_mcu0_1
- sync_ff_drclk_stop_mcu0_1
- ...
- sync_ff_clk_stop_mcu3_0Set to ‘1’ for MCU3 cmp or io clock domain
- sync_ff_clk_stop_mcu3_1
- sync_ff_ioclk_stop_mcu3_1
- sync_ff_drclk_stop_mcu3_1
- sync_ff_clk_stop_soc0_0Set to ‘1’ for SOC0 cmp or io clock domain
- sync_ff_clk_stop_soc0_1
- sync_ff_ioclk_stop_soc0_1
- sync_ff_clk_stop_soc1_0Set to ‘1’ for SOC1 io clock domain
- sync_ff_ioclk_stop_soc1_1
- sync_ff_clk_stop_soc2_0Set to ‘1’ for SOC2 io clock domain
- sync_ff_ioclk_stop_soc2_1
- sync_ff_clk_stop_soc3_0Set to ‘1’ for SOC3 io clock domain
- sync_ff_clk_stop_soc3_1(cmp goes to pc clock domain, not supported)
- sync_ff_ioclk_stop_soc3_1

The transition test counter flops are:

- tcuregs_ttcounter_reg[7:0]Set to binary count as described above

In addition, these synchronizer flops should be scanned to ‘00’ so that they do not interfere with the clock stop logic when in transition test mode.

- cpu.tcu.sigmux_ctl.tap_spc7_mb_cs_sync_reg
- ...

- cpu.tcu.sigmux_ctl.jtag_l2t0_ss_cs_sync_reg

These flops should be scanned to 0 also. They are the first stage of pipeline flops on the clock stop signals as they leave TCU.

- cpu.tcu.clkstp_ctl.clkstp_spc0stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc1stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc2stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc3stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc4stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc5stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc6stop_reg
- cpu.tcu.clkstp_ctl.clkstp_spc7stop_reg
- cpu.tcu.clkstp_ctl.clkstp_bnkstop_reg
- cpu.tcu.clkstp_ctl.clkstp_l2tstop_reg
- cpu.tcu.clkstp_ctl.clkstp_mcustop_reg
- cpu.tcu.clkstp_ctl.clkstp_mcuiostop_reg
- cpu.tcu.clkstp_ctl.clkstp_mcudrstop_reg
- cpu.tcu.clkstp_ctl.clkstp_mcufbdstop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc0stop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc0iostop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc1iostop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc2iostop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc3stop_reg
- cpu.tcu.clkstp_ctl.clkstp_soc3iostop_reg

Note – This listing of flops is intended to be used as a guide only and may not include all flops necessary to implement all variations of transition test.

For IO clock domain, the counter should be set to reflect the desired io clock pulses. For example, to get two io clock pulses the counter should be set to eight, for three io clock pulses the counter should be set to 12, etc. One or more CMP clock domains can be tested with transition test at the same time; the same is true for IO clock domains. However, CMP and IO clock domains cannot be tested together during transition test, one or the other must be specified. Finally, the dr clock domain cannot be tested with transition test since the dr clock is asynchronous to the logic in TCU.

In mio, db0 and db1 there are both cmp and IO clock domains but a single clock stop is sent as soc0_io_clk_stop, so in transition test these blocks cannot be tested since they use the same clock stop for both cmp and IO cluster headers.

4.7.2 Procedure for Entering Transition Test

Because the transition test must be run with the PLL locked, a special sequence is used to enter the transition test mode. This sequence makes use of the dedicated TDI package pin to put the chip in transition test mode while not disrupting the CCU (PLL), TCU, or RST blocks.

1. Lock PLL: POR sequence
2. Stop Clocks: TAP_CLOCK_HSTOP
3. Set Test Protect: TAP_TP_ACCESS
4. Drive TDI to'1'
5. Drive TEST_MODE to'1'
6. Drive AC_TEST_MODE to'1' (active)
7. SCAN_EN to'0' (inactive)
8. ACLK to'0' (inactive)
9. BCLK to'0' (active)
10. AC_TESTTRIG to'0' (inactive)
11. Drive TDI to'0' and hold it
    - This allows (6) signals to propagate but with values unchanged
12. Enter ATPG sequence - scan_en, aclk, bclk, shifting, etc.
    - Hold TCK low and load JTAG_CTL with safe (all 0) state

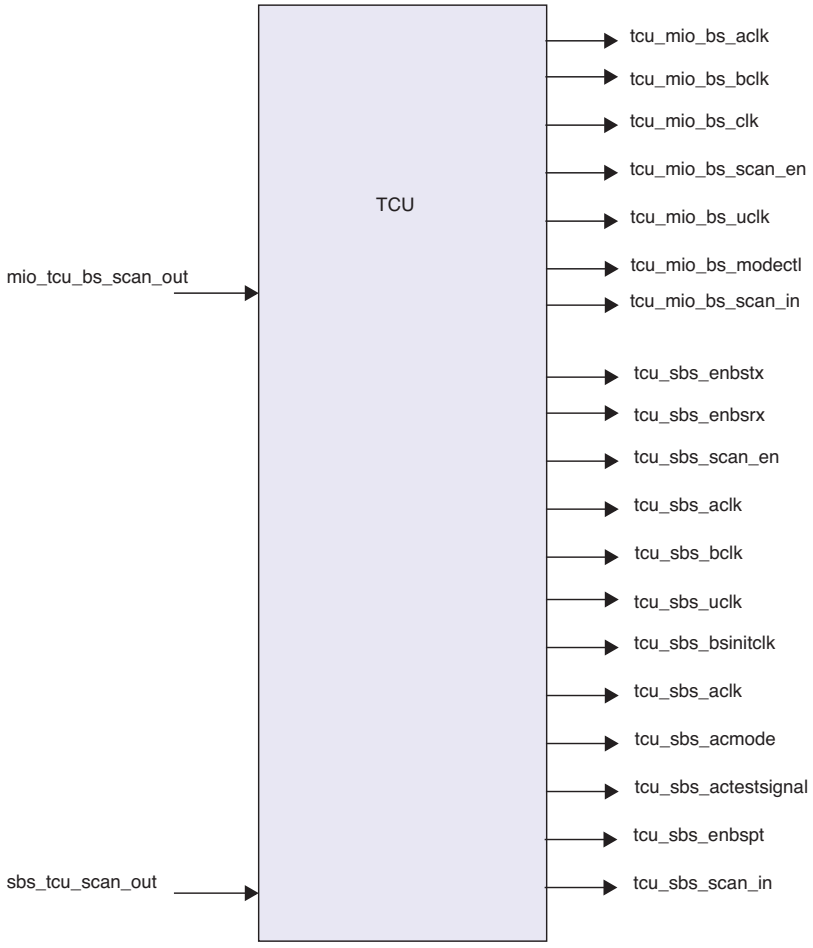
Before entering TT mode, test_protect is asserted with TAP_TP_ACCESS so that it doesn't change when AC_TEST_MODE goes high. Test_protect is OR'd with AC_TEST_MODE so that it is always high during TT, even during shifting, to prevent external signals from affecting TCU, RST and CCU during TT.

4.8 Boundary Scan

The TCU has logic to support boundary scan testing, through the use of JTAG instructions. The interface will provide the following JTAG instructions: Sample/Preload, Extest, HighZ, and Clamp for 1149.1 support in the MIO and some SERDES, and Extest_Pulse and Extest_Train for 1149.6 support for SERDES. The boundary scan cells have also been designed such that they will be included as part of the scan chain.

Timing for boundary scan will be similar to JTAG serial scan as shown in FIGURE 4-6 with an additional update clock occurring in the Update-DR state. During manufacturing scan the boundary scan control signals will be driven from the package pins through multiplexors.

FIGURE 4-10 TCU to Boundary Scan Interface

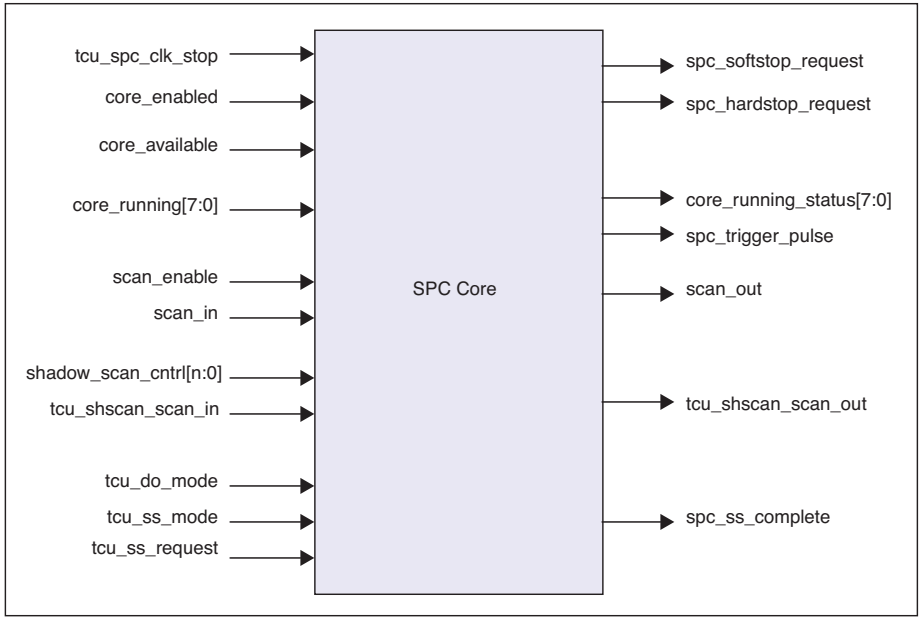


SERDES boundary scan signals are indicated with _sbs_ in the signal name and connect to the clusters which interface to the SERDES macros since the SERDES boundary scan cells are located in the MCU clusters.

4.9 TCU Debug Interface to SPC Cores

The TCU interfaces with the SPC Cores to support debug as shown in FIGURE 4-11

FIGURE 4-11 TCU to SPC Core Interface



4.9.1 Clock Interface

The TCU provides a clock stop signal to the flop headers in the core, and drives this signal active when the core is unavailable.

4.9.1.1 Tcu_spc_clk_stop

This signal is deasserted to allow the OpenSPARC T2 Cores' clocks to run. This is the main signal the TCU uses to control the OpenSPARC T2 Cores' clocks. This signal can be set to'1' at any time to stop the OpenSPARC T2 Cores' clocks. Sourced from TCU as tcu_spc7_clk_stop... tcu_spc0_clk_stop.

4.9.1.2 Core_available & Core_enabled

Core_available is set via eFuse at manufacturing time and determines whether the physical core can be used in normal operation. It serves as a clock gate and if'0' will result in the clk_stop being asserted to the core (this happens in the TCU). Core_enabled is driven from the ASI_CMP_CORE_ENABLED register.

4.9.1.3 Core_running & Core_running_status

The core_running bus is an input from the NCU by which the TCU requests the core to perform a soft-stop. When the core sees the core_running signals transition from a'1' to a'0', it will stop issuing instructions, and wait for all pending core and SPU operations to complete. Once all core-initiated memory operations have been globally performed, the core will raise the Soft_stop_req signal to allow the TCU to stop the clocks to the core and also raise the core_running_status bus to indicate to the CMP logic that the core is parked. When the soft_stop_req is a'1', the OpenSPARC T2 Core will not issue instructions or initiate any activity, until the TCU drives core_running to a'1'.

4.9.1.4 Scan_enable

Besides configuring the scan chains for scanning, this signal also gates off OpenSPARC T2 Core's interface signals so that other SOC units do not respond to spurious OpenSPARC T2 Core interface activity during scanning. At least the crossbar PCX interface is protected in this way.

4.9.1.5 Hardstop_request & Softstop_request

These signals are outputs to the TCU which indicates that the core has reached either a hard-stop or a soft-stop condition and wants to request service from TCU. When a hardstop_request is received, the TCU should disable the clocks to the core using the TCU_spc_clk_stop signal. When the softstop is received, the TCU should request a soft-stop of the entire core via the core_running bus and upon receiving the softstop_request acknowledgement from all eight threads in the core, the TCU will then stop the clocks via TCU_spc_clk_stop. In both cases, the TCU should begin

decrementing the Cycle Counter when the stop_request is received; when the Cycle Counter reaches zero, the clock_stopping sequence is initiated by the TCU. Note that the type of stop (hard vs. soft) is determined either by the Soft_stop_req signal, or which configured event in the DECR has occurred.

Setting bit three of the TCU DCR causes the TCU to soft stop all enabled SPCs if any requests a soft stop; See TCU Debug Control Register

These debug event requests will be honored when the Cycle Counter and Debug Event counters reach zero, and the Reset Counter is not enabled.

4.9.2 Debug Event Interface

This group of core outputs signals that either an error or a debug trigger event has occurred. These debug event requests will be honored when the Cycle Counter and Debug Event counters reach zero, and the Reset Counter is not enabled.

4.9.2.1 Trigger_event

This is a signal from the core to TCU. If the OpenSPARC T2 Core is configured to trigger on an event in the DECR, and the associated event occurs, this signal transitions from a'0' to a'1'. It then transitions back to'0', unless another enabled DECR trigger event occurred that cycle. The TCU will pass this signal to a package pin as the OR of the (64) bits from all cores. The trigger_event signal will be synchronized to the I/O clock frequency.

4.9.3 Scan Interface

Not all signals relevant to the scan interface are detailed here (e.g., not all the scan clocks and controls are listed).

4.9.3.1 Scan_in & Scan_out

There are three scan chains in each core. All flops on this scan string are reset both at POR and during warm reset unless protected via use of the "warm_reset_flop_header".

There are three external scan-out signals per core; each corresponds to a scan-in signal. During JTAG access via scan an entire physical core may be scanned; in this mode the TCU will concatenate the three scan chains in the core, in addition to any JTAG private scan chains such as for shadow scan or memory bist.

4.9.3.2 Shadow_scan_in

This is the scan-in for the shadow-scan string.

4.9.3.3 Shadow_scan_cntrl[n:0]

This bus controls shadow scan operation and identifies which thread's state will be sampled to the shadow scan string.

When the TCU wants to do a shadow scan on a particular core, it first sends the command to the OpenSPARC T2 Core on this bus. At some time later, OpenSPARC T2 Core will capture the state requested by the TCU on the internal shadow scan flops. At that point the TCU can scan out the state by accessing the shadow scan scan string. For details of operation See Shadow Scan Operation.

The signals included in this bus are:

- tcu_shscanid[2:0]: selects one of eight threads
- tcu_shscan_pce_ov: provides a capture signal to the shadow scan reg.
- tcu_shscan_clk_stop: stops the clock to the shadow scan register to allow it to be scanned via JTAG
- tcu_shscan_aclk & tcu_shscan_bclk: shift clocks to perform the scan operation
- tcu_shscan_scan_en: a separate scan_enable for the shadow scan register

4.9.3.4 Shadow_scan_out

This is the scan-out of the core's shadow-scan string.

4.9.4 Single Step Mode

Individual threads can be placed in single step mode via JTAG. To place threads in single step mode the following sequence is used. The user must keep one physical core (SPC) in functional mode.

1. Specify which threads to be in single step mode via TAP_DOSS_ENABLE by setting the corresponding bits in the 64-bit disable overlap/single step enable register.
2. Park all threads by deasserting core_running[7:0] to the target SPCs via the TAP_CREG_ or TAP_NCU_ instructions and accessing the corresponding 8-bit fields in the 64-bit core run register. For any SPC to be operated in single-step mode, all of its threads should first be parked (turned off in core run register).

3. Wait until all threads from the targeted SPCs indicate they are parked via core_running_status[7:0]. This is done by reading the 64-bit core run status register via TAP_CORE_RUN_STATUS. Each bit corresponds to a thread.

4. When all targeted cores are parked, set the DOSS_MODE register to '11' using TAP_DOSS_MODE. Bit [0] indicates single step mode and bit [1] enables the mode. At this stage, the tcu_ss_mode signal is asserted to the targeted physical cores.

5. Assert core_running to the threads that will be single-stepped, via TAP_CREG or TAP_NCU; these threads should correspond to those set in DOSS_ENABLE to maintain compatibility with future enhancements.

6. Pulse the tcu_ss_request signal by executing a TAP_SS_REQUEST (the pulse is generated by going through the update-DR tap state); each running thread in a physical core enabled with tcu_ss_mode will fetch/execute a single instruction.

7. When a SPC's threads have all finished the single-step operation, then that SPC will pulse spc_ss_complete. The TAP_DOSS_STATUS is used to check the spc_ss_complete bit and returns eight bits, one for each SPC. The status is held until the next TAP_SS_REQUEST. When all SPC's indicate they have completed, another single-step can be requested via TAP_SS_REQUEST.

8. Steps six and seven can be repeated to execute a string of 'n' instructions.

9. To exit single step mode, park all threads in the SPCs being single-stepped using TAP_CREG_ or TAP_NCU_. After TAP_CORE_RUN_STATUS indicates all threads are parked, disable the mode using TAP_DOSS_MODE to set the mode to '00'. The DOSS_ENABLE register should also be cleared. Then unpark the desired threads by asserting the respective bits in the core run register using TAP_CREG_ or TAP_NCU_.

Note – Single stepping for multiple threads can be executed independently by control of the targeted threads' respective bits in core_running[7:0] in the above actions.

4.9.5 Disable Overlap Mode

Placing a SPC in disable overlap mode is similar to that for single step mode:

1. Specify which threads to be in disable overlap mode via TAP_DOSS_ENABLE by setting the corresponding bits in the 64-bit disable overlap/single step enable register.

2. Park all threads by deasserting core_running[7:0] to the target SPCs via the TAP_CREG_ or TAP_NCU_ instructions and accessing the corresponding 8-bit fields in the 64-bit core run register. For any SPC to be operated in disable-overlap mode, all of its threads should first be parked (turned off in core run register).
3. Wait until all threads from the targeted SPCs indicate they are parked via core_running_status[7:0]. This is done by reading the 64-bit core run status register via TAP_CORE_RUN_STATUS. Each bit corresponds to a thread.
4. Set the number of cycles to run during disable overlap mode using TAP_CYCLE_COUNT to set the cycle counter.
5. When all targeted cores are parked, set the DOSS_MODE register to '10' using TAP_DOSS_MODE. Bit [0] indicates single step mode and bit [1] enables the mode. At this stage, the TCU will automatically:
6. Assert tcu_do_mode to the target SPCs
7. Unpark the targeted threads
8. Start counting down the cycle counter, waiting until it reaches zero
9. Park the targeted SPCs
10. Set the DOSS_STATUS register
11. Status can be checked with TAP_DOSS_STATUS; bits will be set corresponding to the SPCs which have completed running in disable overlap and are parked.
12. To exit disable overlap mode, park all threads in the target SPCs using TAP_CREG_ or TAP_NCU_. After TAP_CORE_RUN_STATUS indicates all threads are parked, disable the mode using TAP_DOSS_MODE to set the mode to '00'. The DOSS_ENABLE register should also be cleared. Then unpark the desired threads by asserting the respective bits in the core run register using TAP_CREG_ or TAP_NCU_.

Note – The latency of parking and unparking the threads via UCB should be considered when setting the cycle counter.

4.9.6 Cycle Step Mode

Cycle step refers to stopping the clocks to a SPC and then stepping a number of clock cycles to that SPC. Individual SPC cores can be placed in cycle step mode via JTAG. In this mode, a SPC core is stopped via a hard clock stop, then a predefined number of clock pulses is allowed through. After this, the SPC flop contents can be

examined. In this mode the clock domains must be controlled since the clocks are stopped to the target SPCs; this is done by performing a hard stop across only the target SPCs, not the entire chip as in the default hard clock stop.

1. Use TAP_CORE_SEL to set the corresponding bits of the SPCs targeted for cycle stepping.
2. Use TAP_CLOCK_SSTOP to stop the clocks to the SPCs - note this will perform a hard stop on the target SPCs since the TAP_CORE_SEL is active. No SPCs will be parked.
3. Program the Cycle Counter using TAP_CYCLE_COUNT; this can be done before steps one and two also.
4. Verify that the clocks are stopped via TAP_CLOCK_STATUS, the value should be '10' indicating the clock stop operation is finished.
5. Issue a cycle step command via TAP_CS_MODE and loading a '1'. This begins the Cycle Counter operation and allows the number of clocks specified in the Cycle Counter to be sent to the target cores. When the Cycle Counter reaches zero, the clocks will again be stopped to the target SPCs.
6. Check status using TAP_CS_STATUS. This returns a 1-bit value that will be set when the Cycle Counter has finished. It does not indicate if the clocks have yet been stopped, the TAP_CLOCK_STATUS must be used for this.
7. When the status indicates the cycle step has completed, further actions may be taken such as dumping the core contents.
8. To turn the clocks to the SPCs back on, use TAP_CLOCK_START to turn clocks on to the target cores. Note it is impractical to expect the cores to resume operation as a hard stop was in effect.

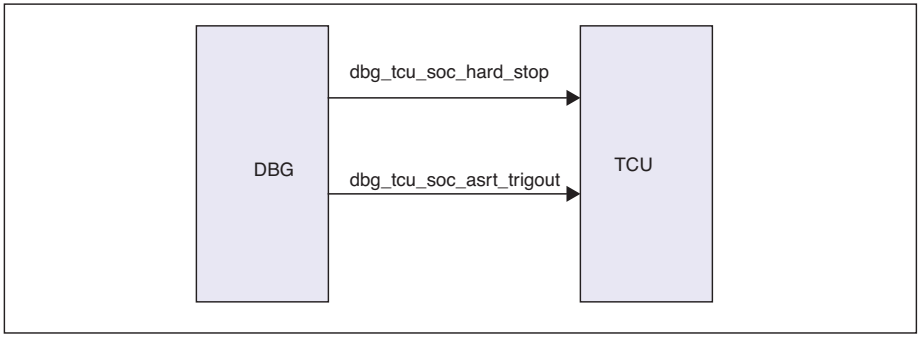
4.9.7 JTAG Priority for Debug

In general, any JTAG instructions related to core debug will take priority over other debug functionality in the TCU. This means that TCU responses to debug events could be blocked if another JTAG instruction is active. Results for JTAG debug operations concurrent with SPC debug service requests are unpredictable.

4.10 TCU Debug Interface to SOC Logic

The TCU interfaces with the SOC logic via the DBG unit as shown in FIGURE 4-12.

FIGURE 4-12 TCU to SPC Core Debug Interface



4.10.1 Clock Interface

4.10.1.1 Hardstop_request

The DBG unit drives a signal dbg_tcu_soc_hard_stop to request a hard stop of the clocks to the chip. This signal is pulsed in io clock domain and when received, TCU will count down the cycle_counter and then begin stopping the chip's clocks. The manner in which the clocks are stopped - serial or parallel - is determined by the contents of the clock_domain register. There is no soft-clock-stop for the SOC logic. The TRIGIN pin is treated as an SOC hard stop request.

4.10.2 Debug Event Interface

4.10.2.1 Trigger_event

To send a watchpoint to the external trig_out pin, the DBG unit pulses dbg_tcu_soc_asrt_trigout high (‘1’) for one io clock cycle. TCU will pass this out to the I/O pins.

4.11 TCU Debug Registers

The TCU handles debug events requests from the SPC cores directly, or from the SOC via the DBG unit, as described in Clock Stop, TCU Debug Interface to SPC Cores and TCU Debug Interface to SOC Logic. 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.

4.11.1 Cycle Counter

This is a 64-bit counter that 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 then the hard-stop will be performed. All debug event requests from the SPC cores or a hard-stop request from SOC logic will be delayed by the Cycle Counter. The Clock Domain register is ignored in this mode: an SOC hard stop will start with clock domain [8]; a SPC hard stop will start with the requesting SPC, and a SPC soft stop will start with SPC0 or all SPCs in parallel if TCU_DCR[3] = 1 (also see Soft Clock Stop).

These actions are only valid when TCU_DCR[2] = 0. For behavior when TCU_DCR[2] = 1, see TCU Debug Control Register. The Cycle Counter is loaded with JTAG instruction TAP_CYCLE_COUNT; default is zero.

4.11.2 TCU Debug Event Counter

This ia a 32-bit counter that must be zero before the Cycle Counter is enabled. If it is non-zero, then each debug event request received at the TCU will decrement it and 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 insure only one type of debug event is enabled when using the Debug Event Counter. Debug event requests consist of these requests from SPC: spc_softstop_request, spc_hardstop_request, spc_trigger_pulse, or these requests from SOC: dbg_tcu_soc_hard_stop, dbg_tcu_soc_arst_trigout, and the trigin package pin. The events are counted per cmp clock cycle.

The Debug Event Counter is only recognized when TCU_DCR[2] = 0. When TCU_DCR[2] = 1, the Debug Event Counter is disabled. The Debug Event Counter is accessed with JTAG instruction TAP_DE_COUNT; default is zero.

4.11.3 TCU Debug Control Register

The TCU has a 4-bit register to control responses to debug events, the TCU DCR (Debug Control Register). When bit two of TCU DCR is '0' the Cycle Counter and Debug Event Counters perform as described TCU Debug Event Counter.

When bit two 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 begins decrementing with each cmp clock cycle after the Power-on Reset (POR) sequence ends (when tcu_rst_flush_stop_ack goes high at the end of flush reset sequencing in WMR2, see WMR2. Once zero is reached either 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 two of TCU DCR = 1) the Debug Event counter will be ignored.

The behavior of the Debug Event and Cycle Counters is determined by the values in the TCU DCR as specified in the following table. The TCU DCR is loaded with JTAG instruction TAP_TCU_DCR; default is zero ('0000').

TABLE 4-10 TCU Debug Control Register Field Definitions

Soft Stop [3]	Enable [2]	[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 second Watchpoint

The following actions are valid when bit two of the TCU DCR is set to '1':

4.11.3.1 Watchpoint

If the TCU DCR is set to '100', then a single pulse of an external chip pin (TRIGOUT) will occur when the Reset Counter reaches zero. The pulse will be synchronized to the io clock domain. The upper 32-bits of the Cycle Counter are ignored.

4.11.3.2 Hard Stop

A hard clock stop will be performed if the TCU DCR is set to '101', as specified in "Hard Clock Stop" on page 220, and a watchpoint pulse generated, when the Reset Counter reaches zero. The upper 32-bits of the Cycle Counter are ignored.

4.11.3.3 Clock Stretch

If the TCU DCR is 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.

4.11.3.4 Clock Stretch then Hard Stop

If the TCU DCR is set to '111', when the Reset Counter reaches zero a clock stretch will be triggered and a watchpoint pulse will also be generated, and 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.

Note – The Soft Stop bit 3 when set will cause TCU to Soft Stop across all enabled SPCs when any SPC requests a soft stop. It should only be active when Enable bit two is '0'; if Soft Stop bit 3 is set when Enable bit two is set, the Clock Domain and Core Select registers will interact with each other - see Soft Clock Stop for details.

For JTAG access to the clock stop logic, the TCU DCR should be in a reset condition so that bits 3 and two are both inactive (0). If either of these bits is set, the interaction between JTAG and the TCU DCR can become unpredictable.

4.11.4 TRIGOUT (Watchpoint) Events

A watchpoint is also referred to as a Trigout event and pulses the TRIGOUT package pin. Watchpoints can be generated by the SPC cores, the SOC logic, SW, or as defined when TCU DCR bit [2] is set, see TABLE 4-10.

TCU will forward a trigout request from SOC or SPC when the cycle counter and debug events counters are zero, but it will only forward that one request. For clock stop/scan dump usage where a trigout is used, TCU generates the trigout based upon the debug event.

For clock stop/scan dump, one trigout may be sufficient. For soc/spc generated trigout requests, it may be desirable to have the ability for TCU to send out every one it receives. However, TCU was designed to only forward/create out one trigout. Two questions came up in regard to this:

1. After the first trigout request is recognized by TCU and forwarded, how do we reset TCU to forward the next trigout request?

2. How do we get multiple trigout events forwarded by TCU to the output pin?

For (1), this was not intended usage. A work-around is to wait until the first debug event has initiated a TRIGOUT request, and then to re-program the debug_event counter (the count does not matter, should be non-zero) and upon the next trigout request the TCU will forward it, and then block subsequent trigout requests. Re-programming the debug_event_counter can be done via SW, no JTAG is needed.

Alternatively, after the initial debug event, programming a TAP_CLOCK_START and then putting JTAG in test-logic-reset will allow TCU to recognize the next TRIGOUT request.

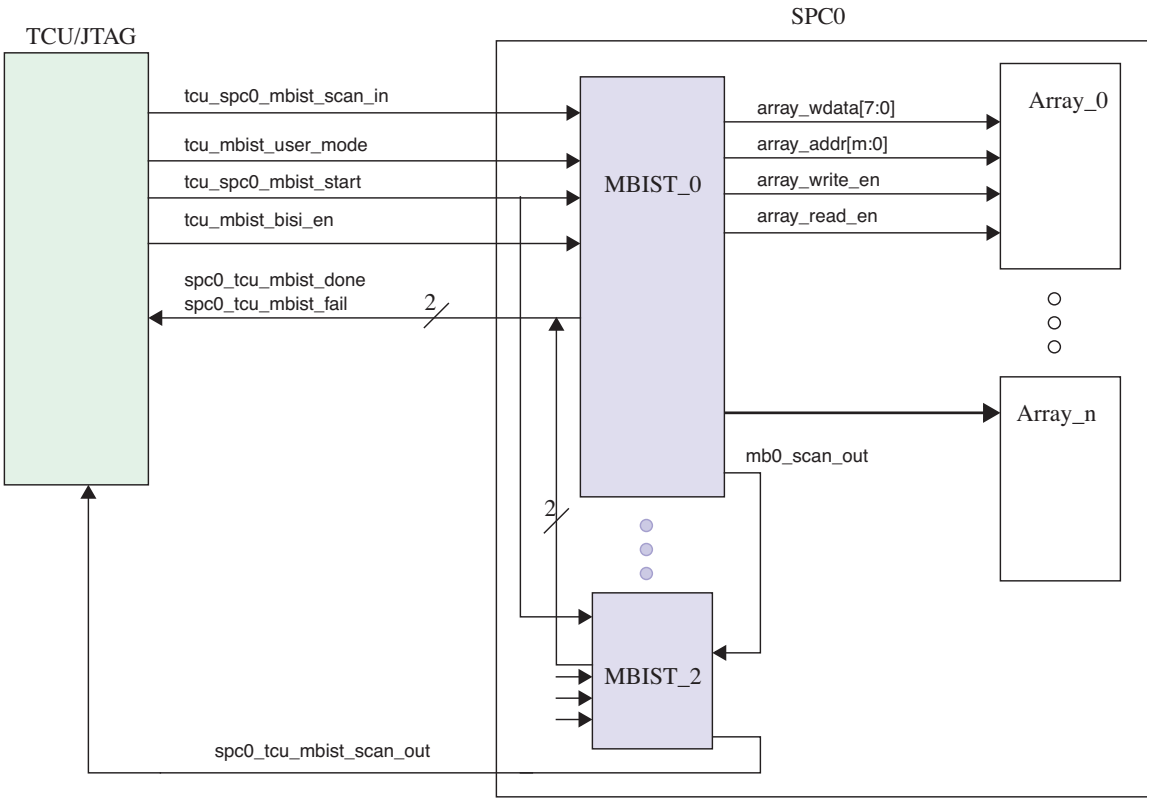
For (2), again this was not the intended usage. A work-around is to wait until the initial debug event occurs and the first TRIGOUT is forwarded by TCU. Then, program the TAP_CLOCK_START and keep it active - do not go to test-logic-reset, instead go to run-test-idle state. Then, all trigout requests received by TCU will be forwarded to the package TRIGOUT pin as long as the TAP_CLOCK_START is active.

4.12 Memory BIST Control

4.12.1 Overview

The memory BIST or MBIST engines for OpenSPARC T2 are based on the engine used in OpenSPARC T1. The general organization between TCU/JTAG, a single MBIST engine in SPC0, and its associated arrays is shown in FIGURE 4-13. In OpenSPARC T2 there are 80 MBIST engines: three per core (24 total) and 56 distributed throughout the SOC logic. Each MBIST engine will therefore test several arrays. Note that even though there are 80 engines, only 48 are visible from the TCU MBIST controller as explained in MBIST Engine Ordering.

FIGURE 4-13 Overview of MBIST Control via TCU/JTAG



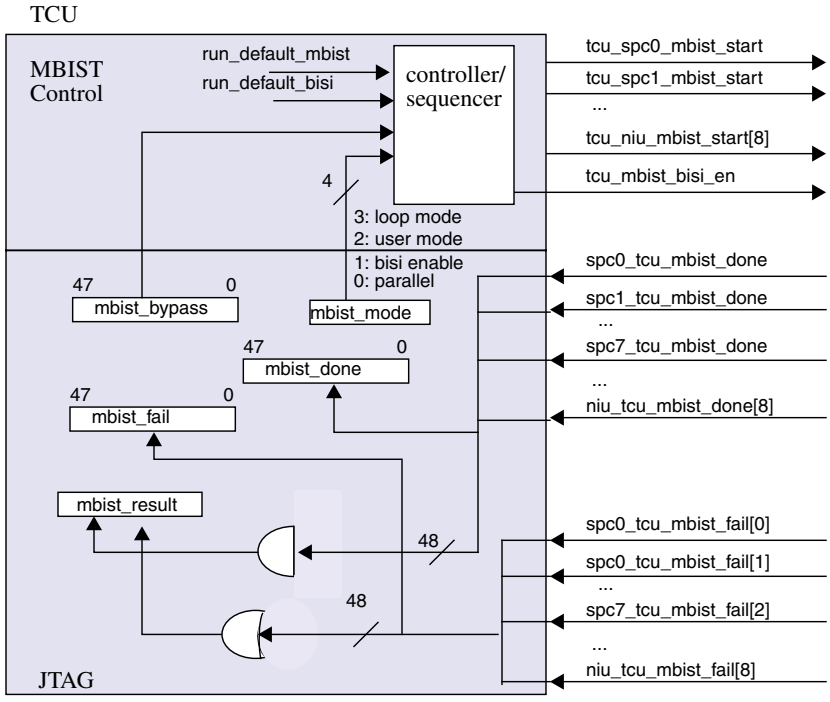
The MBIST operation may be controlled by the TCU during reset sequencing via the JTAG interface or as invoked via SW. The MBIST engines can either be operated in a serial mode, a parallel mode or a diagnostic mode for memory bit-fail mapping. Both the serial and parallel modes run MBIST in a pass/fail mode, where the only information available is whether MBIST passed all of its arrays, or failed at least one of them.

4.12.2 Memory BIST Operation

TCU controls operation of MBIST through an MBIST controller which can be programmed either via JTAG or SW. Operation of the MBIST controller is described in the following sections.

The signal mbist_parallel tells a sequencer in the TCU to begin either serial activation of the start signals if low, or parallel activation if high, based on the contents of the mbist_bypass register. The signal run_default_mbist is activated if enabled during the reset sequence to perform MBIST over all the arrays. The signal run_default_bisi activates BISI on all arrays, in which case the sequencer would set tcu_mbist_bisi_en which goes to all arrays, along with sequencing start signals either serially or in parallel. To run BISI on a given array, the TAP_MBIST_DIAG instruction must 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, or by setting the BISI bit in the mbist_mode register.

FIGURE 4-14 Conceptual Look at TCU/JTAG MBIST Control



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 finished, the TAP_MBIST_GETDONE instruction must be used. This allows capture of 48 done bits which may then be observed at TDO; the TAP_MBIST_GETFAIL similarly captures 48 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.

4.12.3 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 following order.

Serial MBIST ordering:

- SPC0: MBIST 0, 1, 2
- SPC1: MBIST 0, 1, 2
- ...
- SPC7: MBIST 0, 1, 2
- SII: MBIST 0
- ...

To enable the serial MBIST mode via JTAG the instruction TAP_MBIST_BYPASS must be used to specify which of the 48 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 rti 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 CaptureDR state and examining them; this can be done repeatedly for polling (via captureDR, 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 access the MBIST engine specified via the mbist_bypass register, and setting the bisi/bist bit to 1 in the config register if provided in the BIST engine. Another option for running BISI is to set the BISI bit in the TAP_MBIST_MODE register. Optionally, the memory BIST can be run in the serial mode without programming JTAG; this is done via the run_default_mbist signal.

4.12.4 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, while 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.

4.12.5 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. This architecture is depicted in the following figure for spc0 where all three MBIST engines in the core are on the same chain. Similarly, an individual scan chain is provided for each spc and soc cluster as listed in TABLE 4-11.

4.12.6 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. Note: Entering test-logic-reset state does not stop MBIST.

4.12.7 MBIST Engine Ordering

The MBIST engines are ordered as in TABLE 4-11 for the 48-bit JTAG done, fail and bypass registers:.

TABLE 4-11 MBIST Engine Ordering

Cluster	# of Engines	JTAG Reg.	Cluster	# of Engines	JTAG Reg.
SPC0	3	bits[0]	L2B_2	1	bit[20]
SPC1	3	bit[1]	L2B_3	1	bit[21]
SPC2	3	bit[2]	L2B_4	1	bit[22]
SPC3	3	bit[3]	L2B_5	1	bit[23]
SPC4	3	bit[4]	L2B_6	1	bit[24]
SPC5	3	bit[5]	L2B_7	1	bit[25]
SPC6	3	bit[6]	L2T_0	3	bit[26]
SPC7	3	bit[7]	L2T_1	3	bit[27]
SII	2	bits[9:8]	L2T_2	3	bit[28]
SIO	2	bits[11:10]	L2T_3	3	bit[29]
NCU	2	bits[13:12]	L2T_4	3	bit[30]
MCU0	1	bit[14]	L2T_5	3	bit[31]
MCU1	1	bit[15]	L2T_6	3	bit[32]
MCU2	1	bit[16]	L2T_7	3	bit[33]
MCU3	1	bit[17]	DMU	2	bits[35:34]
L2B_0	1	bit[18]			
L2B_1	1	bit[19]			

There are 80 MBIST engines in OpenSPARC T2 but from a JTAG perspective only 48 are visible. Three engines in each SPC are visible as one engine by JTAG, and similarly for the L2 Tags. Since there are eight SPCs and eight L2Ts, this is a reduction of 32 (24 to eight for SPCs, 24 to eight for L2Ts) for a total reduction of 80-32=48 engines visible by JTAG.

4.12.8 Notes

The TCU sources the MBIST engine register scan controls over the scan controls for the holding cluster/core. When an MBIST engine is accessed via a JTAG MBIST instruction the other scan chains in the cluster will also scan, but the data will be lost in those chains.

To use TAP_MBIST_DIAG, the user must bypass all engines (using TAP_MBIST_BYPASS) except the one desired, else the result is indeterminate.

All three core MBIST engines and associated array information (if any) selected via the JTAG instructions are placed in that core's first scan chain for ATPG test mode.

JTAG instructions to support MBIST:

TAP_MBIST_BYPASS 48-bit mbist_bypass register

TAP_MBIST_MODE4-bit mbist_mode register

TAP_MBIST_STARTno user data register

TAP_MBIST_RESULT2-bit mbist_result register

TAP_MGIST_DIAGx-bit mbist_diag Reg: Engine + array flops

TAP_MBIST_GETDONE48-bit mbist_done register

TAP_MBIST_GETFAIL48-bit mbist_fail register

TAP_MBIST_ABORTno user data register

TAP_MBIST_CLKSTPENno user dr; enables clock stop via cycle counter

4.12.9 JTAG MBIST Data Registers

JTAG accessible registers for MBIST are presented in TABLE 4-12.

TABLE 4-12 JTAG MBIST Registers

Register	JTAG Instr.	Fields
Result[1:0]	TAP_MBIST_RESULT	bit[1]: 1 when all 80 mbist engines are done bit[0]: 1 if any of 80 mbist engines reports a fail
Bypass[47:0]	TAP_MBIST_BYPASS	One bit per mbist engine; to bypass an engine during MBIST testing set its bit to 1
Done[47:0]	TAP_MBIST_GETDONE	One bit per mbist engine; a 1 indicates the corresponding engine is done; same order as mbist_bypass register

TABLE 4-12 JTAG MBIST Registers *(Continued)*

Register	JTAG Instr.	Fields
Fail[47:0]	TAP_MBIST_GETFAIL	One bit per 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[3:0]	TAP_MBIST_MODE	bit[3]: user loop mode bit[2]: user mode bit[1]: bisi mode if 1, bist mode if 0 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

4.12.10 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 - such as clock domain and 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.

4.12.11 MBIST DMO - Direct Memory Observe

The basic operation as implemented in TCU is described here. There are three JTAG instructions, TAP_DMO_ACCESS, TAP_DMO_CLEAR and TAP_DMO_CONFIG, as described in TABLE 4-3. The TAP_DMO_ACCESS puts the chip in DMO mode, so that read data from L2 Tags and some SPC or NIU arrays will be observable at package pins during MBIST operation, in addition to done and fail information.

TAP_DMO_CLEAR clears this mode. To access and program the dmo control logic inside TCU the TAP_DMO_CONFIG should be used to set the 48-bits as desired.

TABLE 4-13 JTAG DMO Configuration Register accessed via TAP_DMO_CONFIG

Register	Field	Description
DMO_Config[47:0]	[47:16]	32-bit shift register; bit 47 is used to sample dmo data
	[15]	1 selects CMP clock domain (SPC/L2T) 0 selects IO clock domain
	[14:13]	00 selects dmo path to cores 4, 5, 1 or 0 01 selects dmo path to cores 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]	Not defined
	[10:8]	Not defined
	[7]	selects SPC data cache upper/lower word
	[6]	1 selects SPC instr. cache and L2 Tag output 0 selects SPC data cache and L2 Data output
	[14:13] & [5:0]	[14:13] [5:3][2:0] ==> Cluster Selected 00 xx0 xxx ==> CORE4 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

TAP_MBIST_ABORT does not clear DMO mode.

Note – As in most single-access JTAG instructions in OpenSPARC T2, reading the dmo config register using TAP_DMO_CONFIG is done via the capture-DR state and is always followed by writing the register when update-DR is passed through. Thus, the dmo config register should not be accessed while dmo is actively running as the shift register contents will be disturbed during update-DR.

In DMO mode, TCU will pass the data, done and fail information for the mbist array under test to the package pins, along with a synchronization pulse.

4.12.11.1 MBIST Done and Fail Observe Ability at Pins

The capability exists to have any MBIST engine pass its done and fail signals to the package pins and is enabled with TAP_DMO_ACCESS. Due to synchronization it is not guaranteed that all fail pulses will be seen at the package pin. The done and fail information is intended to be observed only in user mode with one array selected. In non-user mode the fail pin is indeterminate if failures occur (if no failures occur, the fail pin will pulse only at MBIST sequencer initiation).

When the MBIST sequencer is initiated, TCU will pulse the mbist done and fail pins for one io2x clock cycle. While MBIST is running TCU will pass the fail information received from the MBIST engine which is running to the pins in user mode only (non-user mode is indeterminate). When the MBIST sequencer finishes, TCU will assert done to the pins and assert fail to the pins if there was any failure recorded from MBIST. This means that there will be at least one pulse generated by TCU on the fail pin even if no fail occurred.

Upon initiation of the MBIST sequencer the pulses on the done and fail pins will be coincident. The reason for the initial pulses on the done and fail pins is so the user can determine that the MBIST controller in TCU has started and the pins are capable of toggling. The done and fail will only be pulsed upon MBIST start when the MBIST controller begins operation. This means that the done and fail will not pulse each time a successive MBIST engine is started if more than one engine is left in non-bypass mode.

FIGURE 4-15 Sample: MBIST DMO Data coming from CMP Clock Domain

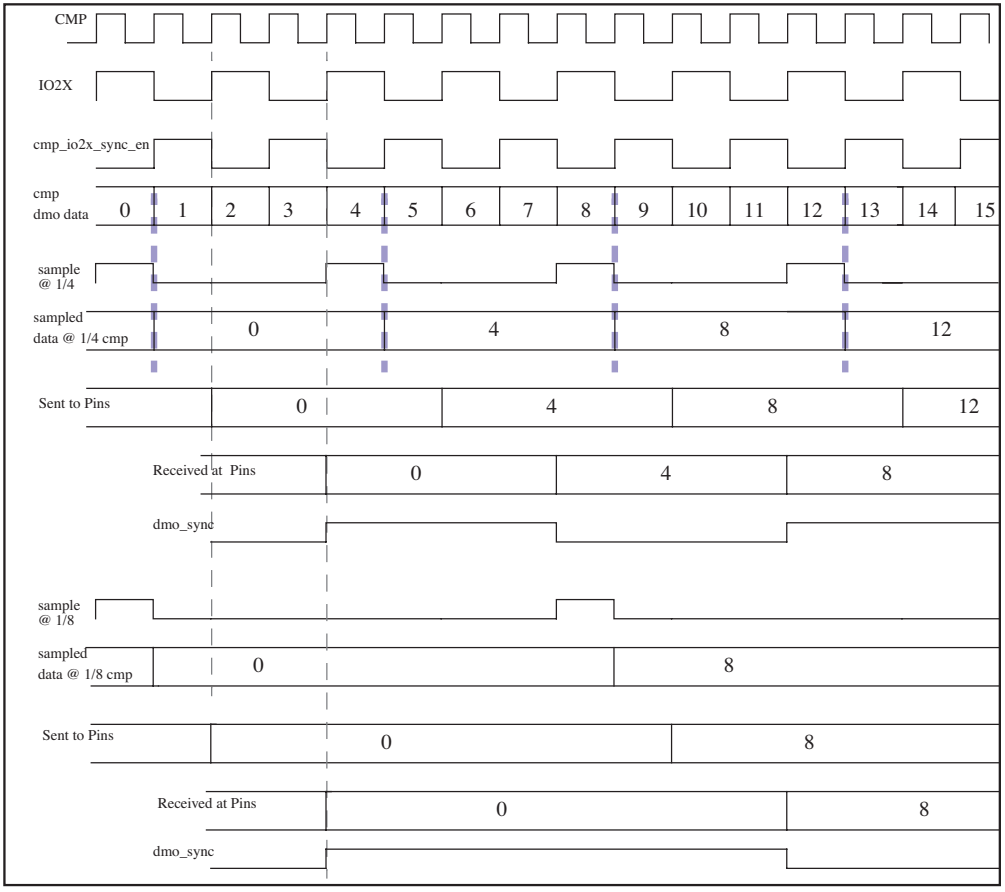


FIGURE 4-15 is representative only, and shows 1/4 sampling (32-bit shift register set to 32'b00010001000100010001000100010001) and sampling at 1/8 (32-bit shift register set to 32'b00000001000000010000000100000001). Note that the msb (bit 47 of the dmo config register, bit 31 of the shift register inside the dmo config register) is used to sample the dmo data.

4.12.12 Scanning of MBIST Engines via JTAG

When scanning MBIST engines using the TAP_MBIST_DIAG instruction or TAP_SERSCAN it may be that flops outside the targeted scan chain may be disturbed. The effects for SPC and SOC engines are different.

When scanning a SPC MBIST engine the controls from TCU will not be seen by any logic outside the target SPC. However, if TAP_MBIST_DIAG is used to obtain a short chain between TDI and TDO the flops outside this chain in the target SPC will be affected by aclk, bclk and scan enable. It may be advantageous to use TAP_SERSCAN with only the targeted SPC selected so that all values in the SPC scan chain can be controlled.

When scanning an SOC MBIST engine there is no individual scan control. So all SOC logic except for RST, TCU and CCU will be affected by the scanning of any SOC MBIST engine.

Also see Protecting TCU During Serial Scan: Test Protect Mode for proper use of TAP_TP_ACCESS during MBIST serial scanning.

4.12.13 Effect of Unavailable or Disabled Cores and Banks

The MBIST sequencer in TCU observes the available and disabling signals as described in Clock Stopping and Core/L2 Available and Disable Controls. When either BIST or BISI is run the MBIST sequencer in TCU will automatically bypass any MBIST engines in an unavailable or disabled SPC or L2 array including MCU and the associated L2 Tag. If only one MBIST engine is selected, with all others bypassed, and that MBIST engine is in an unavailable or disabled SPC or L2/MCU, the MBIST sequencer will bypass it and effectively do no MBIST testing. This is an illegal state and MBIST sequencer operation is not deterministic.

4.12.14 BIST During Reset

During the POR sequence as described in the chapter on reset see Reset Sequencing. TCU will run a BISI sequence after POR1 and optionally either run BISI or BIST between WMR1 and WMR2. The BISI run after POR1 is in parallel mode by default and has a timeout counter of 32 bits. The signal tcu_rst_bisx_done will be asserted when all non-bypassed engines return their done signals to TCU or the timeout counter expires. BISI will use the bypass register to select which engines to run and will not expect done signals from bypassed engines. The BIST mode register is not applicable except for changing from parallel to serial mode since the BISI run is "hard-coded" to run after POR1.

An optional BIST or BISI run is available if programmed by Software and will be recognized by TCU after the next WMR1 and will be serviced between WMR1 and WMR2. This optional run will recognize the bypass and mode registers.

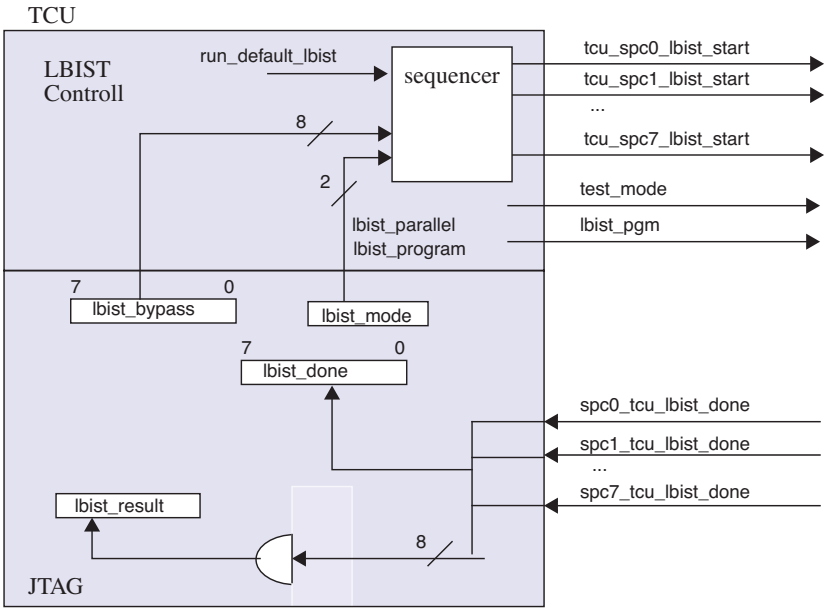
Both the POR1 BISI and the optional Software-requested BIST/BISI will obey the SPC and L2 available and disable criteria as specified in Effect of Unavailable or Disabled Cores and Banks. Note, the BISI enable is written by logic during the power-on reset sequence, and once written it will remain high until it is programmed otherwise.

4.13 Logic BIST Control

The Logic BIST test function is only applied to the SPC cores in OpenSPARC T2, and one engine is instantiated per core. The control of the Logic BIST engines comes from the TCU either via SW 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.

FIGURE 4-16 Conceptual Look at TCU/JTAG Logic BIST Control



To start a single Logic BIST engine, the TCU will drive the lbist_start signal high and hold it until the lbist_done signal is received. TCU will also source a test_mode signal to all Logic BIST engines to control them during manufacturing scan. In parallel mode, the non-bypassed engines will be sent lbist_start signals in the same cycle. In serial mode, the first non-bypassed engine (counting from 0) will be sent an lbist_start indication, then when its lbist_done is received the next non-bypassed engine will be sent an lbist_start indication. All lbist_start signals will be held until the sequencer has received done indications from all engines.

4.13.1 JTAG Logic BIST Instructions

The instruction TAP_LBIST_START can be used to begin Logic BIST sequencing, and TAP_LBIST_ABORT can be used to stop sequencing. The JTAG accessible registers for Logic BIST are:

TABLE 4-14 JTAG Logic BIST Registers

Register	JTAG Instr.	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
Lbist[k:0]	TAP_LBIST_ACCESS	Includes targeted Logic BIST engines across cores
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

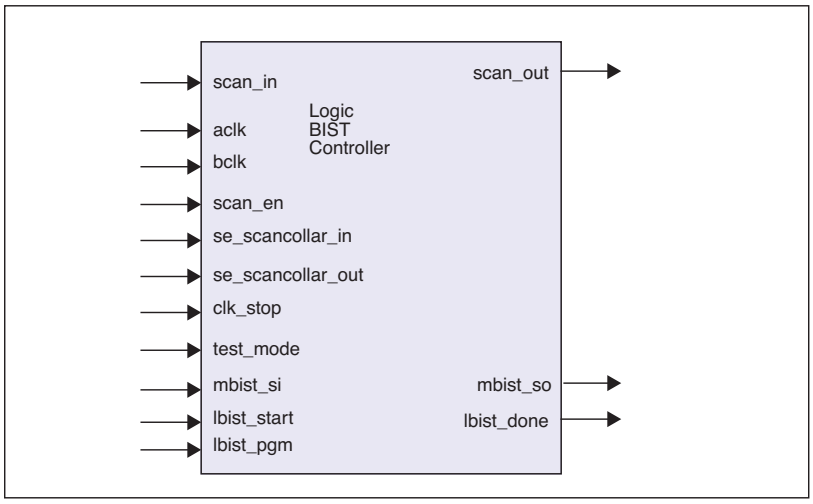
4.13.2 Accessing Pass/Fail Signature

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 against a known-good value. Alternatively, Software may access the signature. This mechanism will be implemented in the SPC core.

Also see Protecting TCU During Serial Scan: Test Protect Mode for proper use of TAP_TP_ACCESS during LBIST serial scanning with TAP_LBIST_ACCESS.

4.13.3 Logic BIST Interface

FIGURE 4-17 Logic BIST Controller Interface with TCU

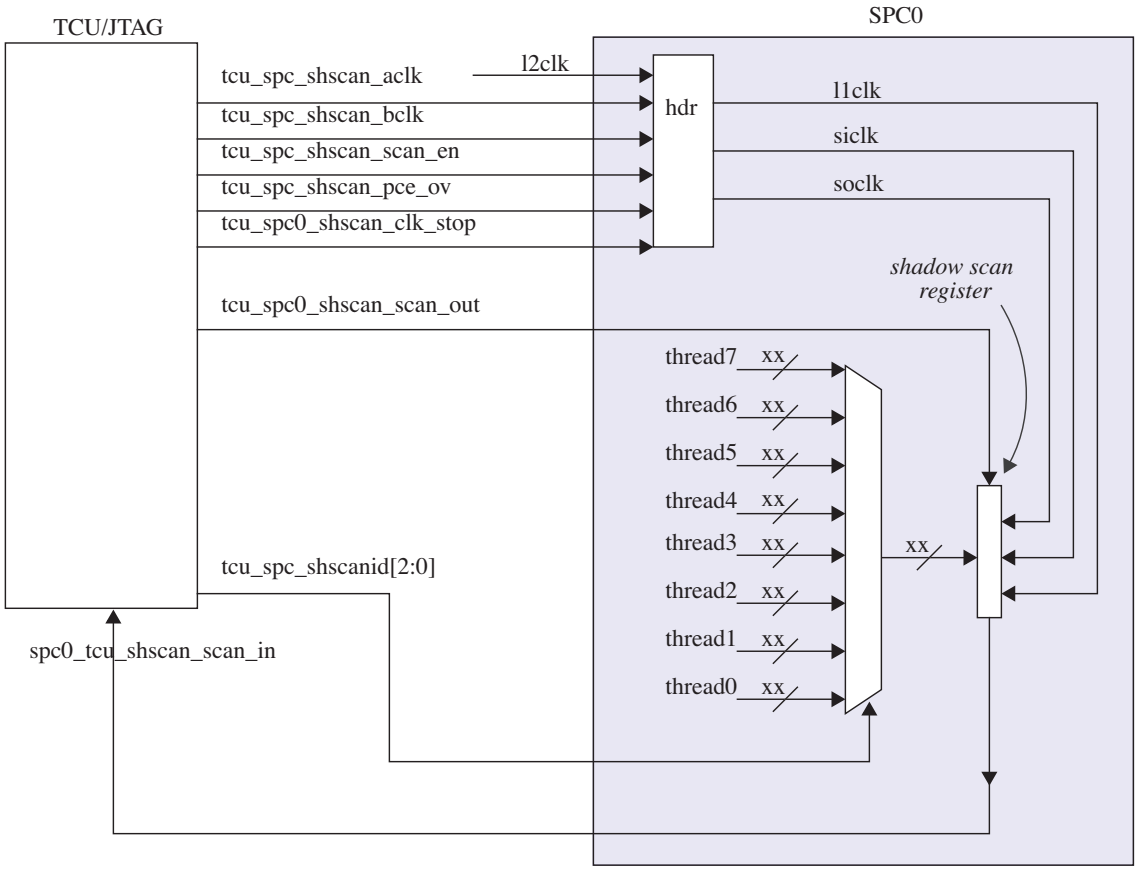


4.14 Shadow Scan

4.14.1 Core Shadow Scan

Shadow scan for the cores is controlled via JTAG. The architecture is shown FIGURE 4-18; the header is a conceptual view of both the cluster and flop headers combined. Each core shadow scan will be contained in a separate scan chain, with its own clock headers and controls coming from the TCU. The contents to be captured in the shadow scan are in the *OpenSPARC T2 Programmer's Reference Manual*. If a core is disabled then its shadow scan contents will be excluded and the number of TCK clocks should be reduced to reflect the unavailable core(s).

FIGURE 4-18 Core Shadow Scan Architecture



4.14.2 SOC Shadow Scan

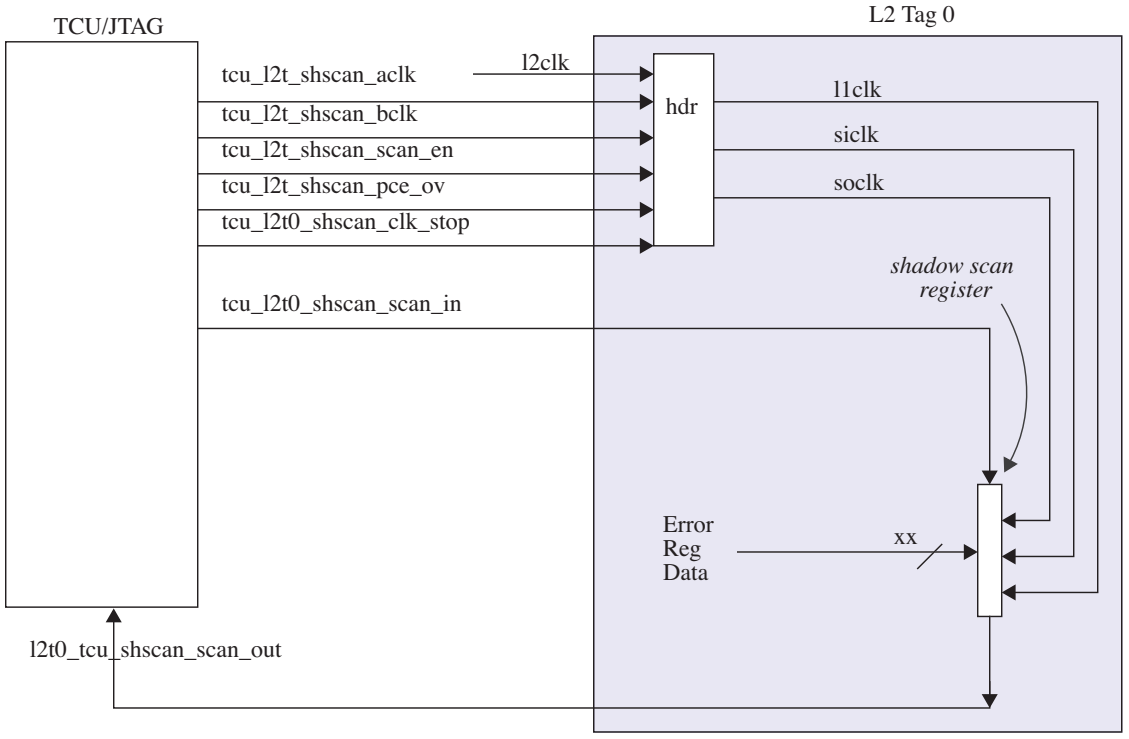
Shadow scan for the SOC consists solely of L2Tag error registers, and is controlled via JTAG. The architecture is shown in FIGURE 4-19; the header is a conceptual view of both the cluster and flop headers combined. Each L2 Tag shadow scan will be

contained in a separate scan chain, with its own clock headers and controls coming from the TCU. The contents to be captured in the shadow scan are in the *OpenSPARC T2 Programmer's Reference Manual*.

TABLE 4-15 Shadow Scan Registers

siisyn_data[61]	"000001"	"000001"	"000001"	"000001"	"000001"	"000001"
Etag[2:0]	"001"	"111"	"101"	"000"	"100"	"110"

FIGURE 4-19 L2 Tag Shadow Scan Architecture

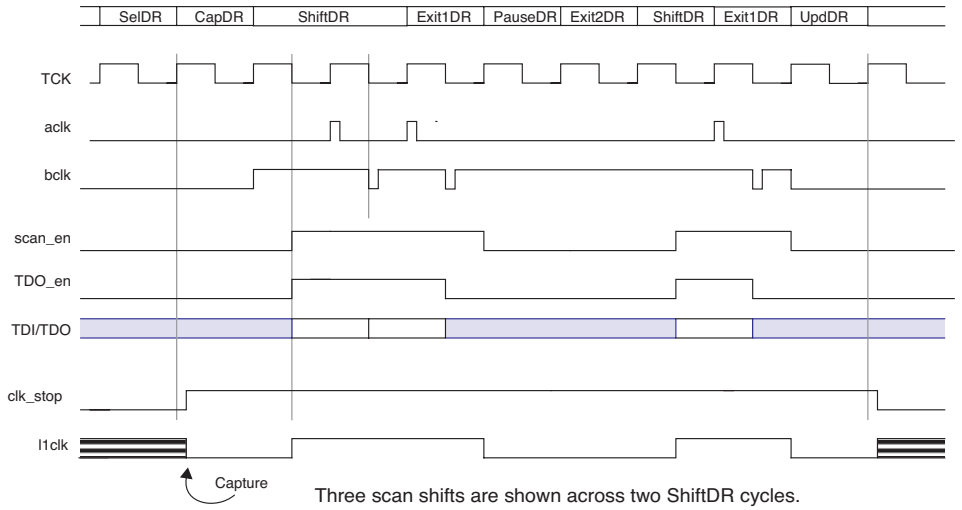


4.14.3 Shadow Scan Operation

During a shadow-scan operation, the PLL is running and 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. Because TCK is specified to be at a much slower frequency than any cpu clock, any cpu clock cycles required for synchronization from TCK to cpu clock domains will not cause overlapping.

FIGURE 4-20 JTAG Shadow Scan Sample Waveform



NOTES:

1. All eight core shadow scans are scanned serially as one chain, with core 0 closest to TDI and core 7 closest to TDO.
2. Assignment of shadow scan contents is in the *OpenSPARC T2 Programmer's Reference Manual*.
3. Any core marked unavailable in the CMP core_available register will not be included when scanned via TDI to TDO.
4. The shadow scan chain for a given core is placed in that core's second scan chain during ATPG test mode; they are accessible via JTAG shadow scan instructions and during JTAG serial scan.
5. All eight L2 Tag shadow scan contents are captured at the same time, and are available at TDO with L2T0 first and L2T7 last (closest to TDO).
6. JTAG instructions to support Core Shadow Scan:
   - TAP_SPCTHR0_SHSCAN Thread 0 contents for all available cores

- TAP_SPCTHR1_SHSCAN Thread 1 contents for all available cores
- TAP_SPCTHR2_SHSCAN Thread 2 contents for all available cores
- TAP_SPCTHR3_SHSCAN Thread 3 contents for all available cores
- TAP_SPCTHR4_SHSCAN Thread 4 contents for all available cores
- TAP_SPCTHR5_SHSCAN Thread 5 contents for all available cores
- TAP_SPCTHR6_SHSCAN Thread 6 contents for all available cores
- TAP_SPCTHR7_SHSCAN Thread 7 contents for all available cores

7. JTAG instructions to support L2 Tag Shadow Scan
   - TAP_L2T_SHSCAN

4.15 Array Guidelines to Support Scan Test

To facilitate scan test the arrays should be configured so that they can be inhibited during scan load and unload, and surrounded with scan collars. There are several different scan modes used on OpenSPARC T2 and this section outlines the use and requirements for the scan mode control signals. The different scan modes consist of Manufacturing or pin-based scan, also known as ATPG scan, MacroTest, Logic BIST (LBIST), Transition Test, JTAG scan, and Flush scan.

The TCU sources four signals for scan control specifically related to arrays:

- *tcu_se_scancollar_in* - connect to "se" port of flop headers for memory "input" flops
- *tcu_se_scancollar_out* - connect to "se" port of flop headers for memory "output" flops
- *tcu_array_wr_inhibit*
- tcu_array_bypass

4.15.1 Flop (Clock) Headers

To control the clocks to arrays during the various scan modes, clock headers are needed with specific se (scan enable) signals. The se signals from TCU to arrays are *tcu_se_scancollar_in* and *tcu_se_scancollar_out*.

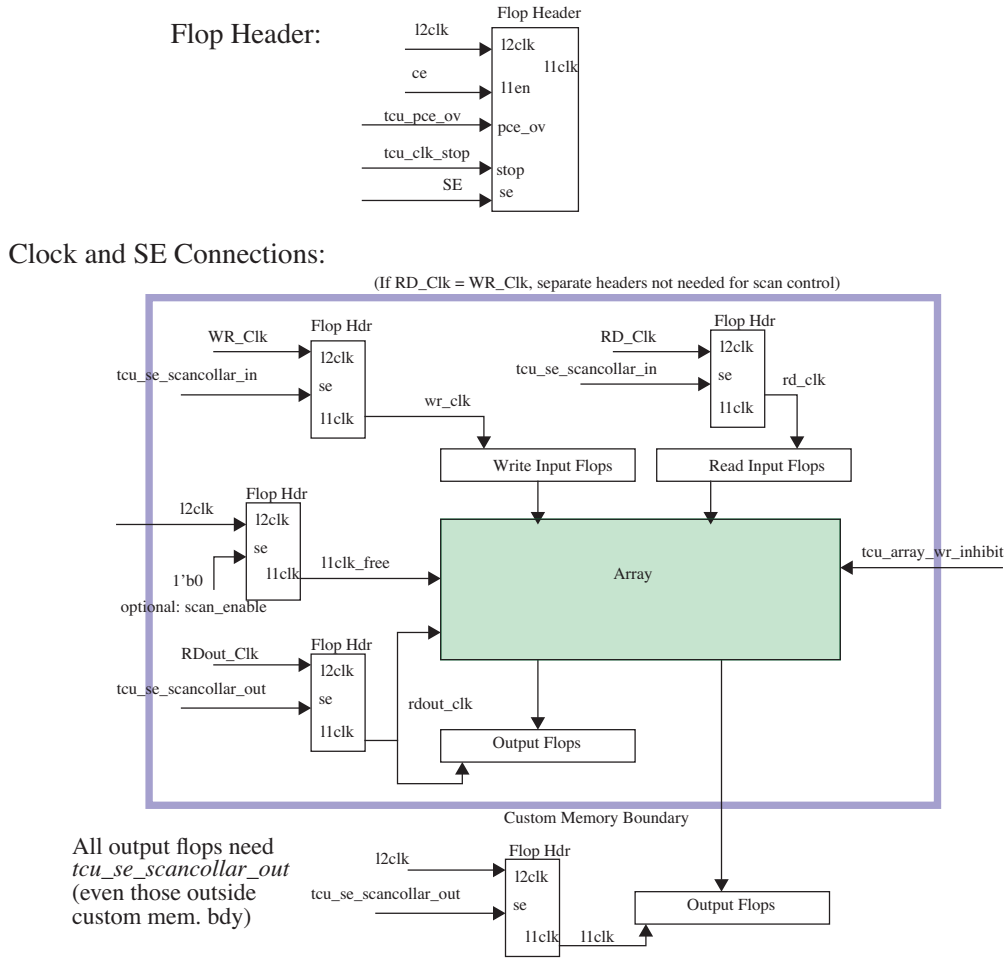
For any input flops including write address/data, read address/data, write/read enable, and inputs related to lookup for CAMS, the flop headers for these flops should use *tcu_se_scancollar_in*. For any output flops, such as read data, a second flop header is required and should use *tcu_se_scancollar_out*. The various flop

headers in an array should share the *tcu_pce_ov* and *tcu_clk_stop* signals and all flops can share *tcu_aclk* and *tcu_bclk*. If the "se" port of a flop header is tied low, then during scan operations the l1clk will track the l2clk - this is sometimes referred to as making the l1clk free-running. If the power-savings function of a flop header is not needed then the "l1en" (ce) can be tied high.

4.15.1.1 Write Inhibit and Bypass

To inhibit writing to the arrays the TCU generates a signal *tcu_array_wr_inhibit*. When active, this signal should protect the array from updates and can also be used to turn off read logic and CAM compare logic if desired. If it is determined by test coverage analysis that there is logic not tested, such as shadow logic, a bypass mux may be needed. In general it is expected that most arrays will not need a bypass mux; this decision will be made for each array individually. The bypass mux will be controlled via *tcu_array_bypass*. Plan-of-record for OpenSPARC T2 arrays is to place the bypass mux outside the custom memory boundary (in RTL logic), except for CAMS that have already placed bypass muxes in the custom memory.

FIGURE 4-21 Array Flop Header Guidelines



Note – If WR_Clk and RD_Clk are different clock domains then separate stop signals should be used as provided by the TCU.

4.15.2 Scan Modes

The values for the various test control signals and the clocks are given in the following table for the different scan modes (*tcu_clk_stop*=0 and *tcu_pce_ov*=1). The value of *l1clk_free* tracks *l2clk*.

TABLE 4-16 Array Control Signals During Scan Modes

Scan Mode	Phase	l2clk	se_scancollar_in	se_scancollar_out	array_wr_inhibit	l1clk
ATPG	Scan Shift	Tester drives to 1	1	1	1	1
	Capture	Tester toggles	0	0	1	l2clk
MacroTest	Scan Shift	Tester drives to 1	1	1	1	1
	Capture	Tester toggles	1	0	0	in: 1 out: l2clk
Logic BIST	Scan Shift	PLL Locked	1	1	1	1
	Capture	PLL Locked	0	1	1	in: l2clk out: 1
Trans. Test	Scan Shift	PLL Locked	1	1	1	1
	Capture	PLL Locked	0	0	1	l2clk
JTAG	Scan Shift	PLL Locked	1	1	1	1
Flush	Scan Shift	PLL Locked	1	1	1	1

4.15.3 Scan Cell Ordering Guidelines

Scan Cell Ordering: There is no specific requirement for ordering of scan cells in the scan chain, although it is desirable for all flops of the same function to be grouped in the chain to facilitate macrotest pattern development. Lockup latches are not needed since the scan clocks are always non-overlapping.

It is required that the ordering of scan cells in the circuit match that in the RTL.

4.15.4 Reset

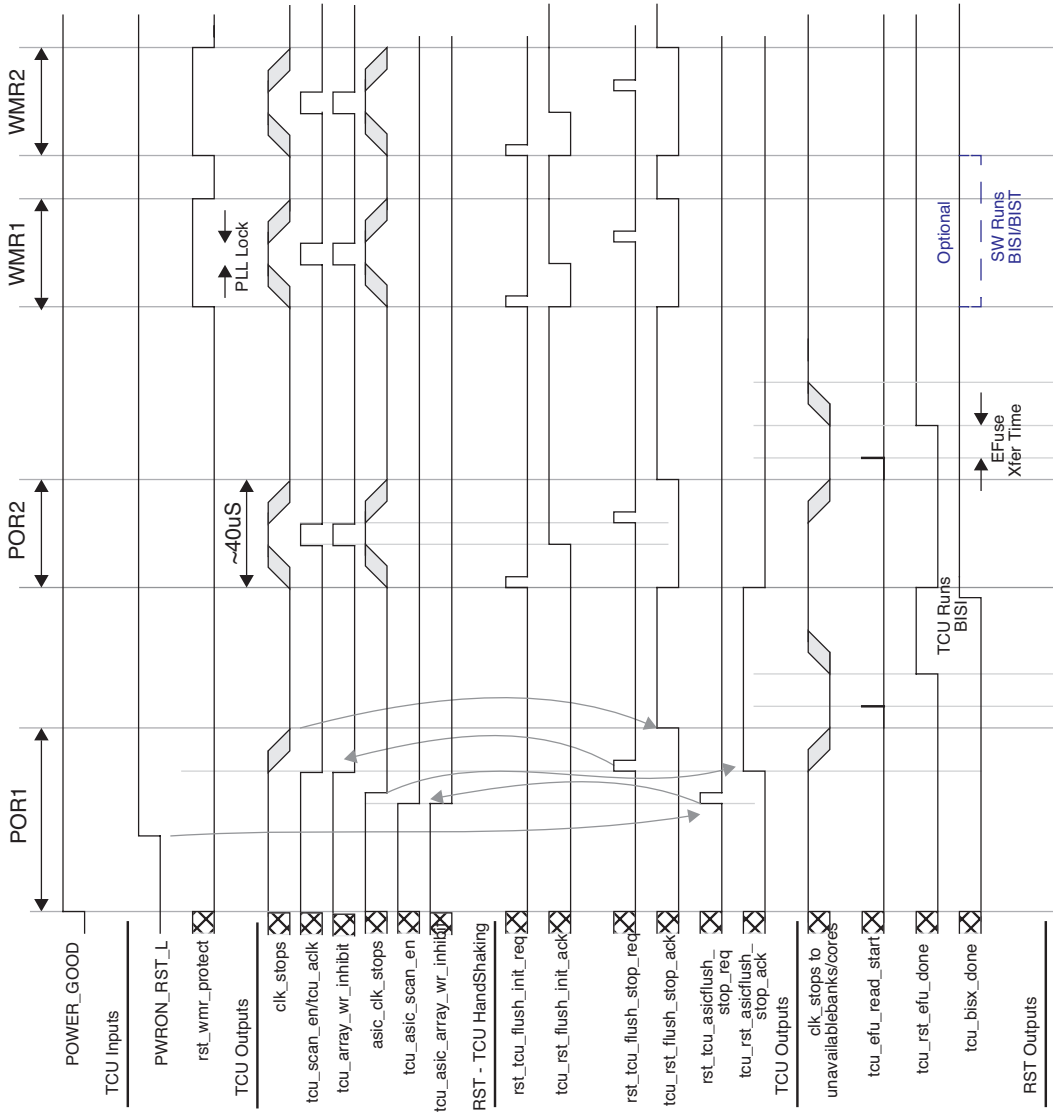
During portions of the power-on-reset sequence, such as before the PLLs lock or during flush scan, tcu_array_wr_inhibit will be driven active to protect the arrays.

ASIC arrays do not participate in flush scan, so aclk/bclk would be inhibited during this time to those arrays.

4.16 Reset Sequencing

The TCU participates in Power-On Reset by interfacing with the RST unit and providing flush reset. The waveforms with respect to TCU during power-on reset are shown FIGURE 4-22.

FIGURE 4-22 Power-On Reset Sequence



There are four phases to the Power-On Reset sequence:

Two Power-On Resets - POR1 and POR2 - and two Warm Resets - WMR1 and WMR2.

4.16.1 POR1

At time zero, PWR_ON_RST_L is driven to '0'; the TCU sees this and then asserts all clk_stop signals and resets itself and all chip logic (except for JTAG, RST, CCU, and DMU which get their own reset signals) via flush reset.

The TCU is released from reset and becomes active when the PWR_ON_RST_L has deasserted (goes to '1') and the RST block drives rst_tcu_asicflush_stop_req high. Once the TCU is active it will wait for rst_tcu_flush_stop_req before ending the flush state and starting the clocks by turning off clk_stops to the 24 clock domains. Note that this does not apply to the ASICs as described in ASIC Reset. When it has turned on all clock domains, it will drive tcu_rst_flush_stop_ack to '1' and then assert tcu_efu_rvclr after a delay of 32 cmp clock cycles, hold tcu_rvclr for eight cmp clock cycles, deassert and wait eight cmp clock cycles, then pulse tcu_efu_read_start for eight cmp clock cycles. The hold times and delays allow the eFuse signals to be synchronized into the io clock domain.

After the EFU is done, the TCU will start the default BISI sequence. This will complete before POR2 is entered. From a TCU perspective, POR1 ends when POR2 begins.

Note – Since the PLL is locking during POR1 reset, the flush will be held until the PLL is stable before exiting flush_POR1. This is controlled by the RST unit.

Note – The clocks to unavailable SPC cores and L2 Banks will be stopped; for details. See Clock Stop.

4.16.2 POR2

The POR2 state is recognized by TCU when the rst_tcu_flush_init_req goes to '1' and the rst_wmr_protect is low. The TCU responds by stopping all clocks, and drives tcu_rst_flush_init_ack high when it begins to flush the scan chains after clocks have been stopped.

When the rst_tcu_flush_stop_req is received the TCU will cease flushing the scan chains, and turn all clock domains back on, followed by driving tcu_rst_flush_stop_ack high to tell RST the clocks are running. TCU also pulses tcu_efu_read_start to again signal the EFU to begin operation. The timing of the eFuse handshaking is the same as in POR1.

The DMU is excluded from flush resets and clock stopping (the RST, CCU and TCU are excluded from POR2). The start of WMR1 indicates the end of POR2 from a TCU perspective.

4.16.3 WMR1

The WMR1 state is recognized by rst_wmr_protect being active during the receipt of rst_tcu_flush_init_req to begin this phase. The actions are similar to POR2 between TCU and RST as shown, except no eFuse start is sent by TCU. When TCU drives the tcu_rst_flush_stop_ack high this indicates the end of WMR1 from a TCU perspective.

Before leaving WMR1 TCU will start any pending BIST/BISI sequence which may have been requested via SW; this will be allowed to complete before WMR2 is entered.

4.16.4 WMR2

The WMR2 state is recognized only after WMR1 has occurred by rst_wmr_protect being active during the receipt of rst_tcu_flush_init_req; TCU responds as in WMR1. At the end of WMR2, though, the Reset Counter will be allowed to decrement as specified in the section on debug. See TCU Debug Registers. The Reset Counter should be programmed before WMR2 ends.

When TCU drives the tcu_rst_flush_stop_ack high this indicates the end of WMR2 from a TCU perspective.

4.16.5 JTAG Access During POR

JTAG is operational after TRST_L goes inactive during POR1; however, to access registers that are outside of JTAG the clocks must be running and the targeted areas should be capable of receiving the JTAG actions. During the POR sequence there are specific times when such access via JTAG is possible. To insure the user performs JTAG accesses during a safe period in the POR sequence, three JTAG instructions have been provided that create a window in the POR sequence so that JTAG instructions can be safely performed. By executing TAP_JTPOR_ACCESS a signal

inside TCU will be set that will cause TCU to pause after eFuse two transfer completes by delaying activation of the tcu_rst_efu_done signal. The user should execute TAP_JTPOR_ACCESS after releasing TRST_L in the POR sequence. The status of TCU can then be checked with TAP_JTPOR_STATUS; when the status is '1' this indicates that the TCU is paused and the JTAG programming window is active. Clocks will be running and JTAG instructions can be executed during this window. To continue with POR the user should execute TAP_JTPOR_CLEAR, which will cause TCU to continue with the POR sequence.

Note that when setting the JT POR access via TAP_JTPOR_ACCESS, it is possible to hold the chip input pin PWRON_RST_L low to allow enough time for the JTAG programming to be completed. The sequence then would be to begin the POR sequence, release TRST_L but hold PWRON_RST_L low, complete the TAP_JTPOR_ACCESS programming, and then release PWRON_RST_L to allow the reset sequence to continue.

To shorten the bisi sequence in POR1 the JTAG TAP_MBIST_ABORT can be used. The execution of this instruction would need to be controlled by counting sys clock pulses and choosing the appropriate cycle to abort. If the BISI sequencer in TCU receives an MBIST-Abort request it will simulate a BISI timeout and the sequence will continue as if an actual timeout had occurred. This is useful during testing on ATE to shorten the POR sequence. It is up to the user to determine when to execute the TAP_MBIST_ABORT, and this does not require using TAP_JTPOR_ACCESS as the abort needs to occur before the JTPOR pause occurs.

One of the primary uses of JTAG access during POR is to bypass the eFuse. One important point to keep in mind is that anything that has already happened in the POR sequence before the pause won't see the bypassed eFuse values. In particular, BISI will have already completed during POR1 using data from the first eFuse transfer. Thus, if a bank available row in the eFuse array is bypassed during JTPOR to make an unavailable bank available, the bank that was marked as unavailable during POR1 will not have been initialized by BISI and will contain garbage data. It is up to software to tolerate this.

4.16.6 ASIC Reset

The ASIC blocks in OpenSPARC T2 DMU are treated differently from other clusters during the reset sequence and warm or debug resets.

During POR1 the DMU has its clocks stopped until the RST unit tells TCU to release them with rst_tcu_asicflush_stop_req; this signal comes earlier than rst_tcu_flush_stop_req. When the asicflush_stop_req is received, TCU releases itself from its own flush reset and turns off the clock stops to the ASICs and deasserts tcu_asic_scan_en. The tcu_asic_aclk is not asserted at all during POR1, preventing a flush state to the ASICs. During subsequent resets (WMR1, WMR2) the ASIC clock stops are allowed to activate in the normal clock stop sequence but the ASICs are not

flushed. During debug resets (DBR) the signal rst_tcu_dbr_gen is active and TCU does not activate the clock stops to the ASICs to allow them to continue running. During JTAG clock stop operations, these blocks behave as other SOC blocks. During POR2 the ASIC clock stops will be held deasserted.

4.17 EFuse

The interface between the TCU and the E-Fuse Unit (EFU) is similar to that from OpenSPARC T1. This section only describes the TCU to EFU interface including the JTAG instructions used. For information about the EFU refer to O*penSPARC T2 SoC Microarchitecture Specification, Part 2 of 2*.

There are five modes of operation which the TCU recognizes for interfacing with the TCU. All except the POR mode require JTAG instructions and user intervention. The POR mode is handled directly by the TCU during the power-on reset sequence.

Note – Bypass data register is in EFU, and has reversed bit ordering from other JTAG registers (msb is closest to TDO). Bit ordering may also apply to other EFU register values as interpreted by the EFU.

4.17.1 POR Mode

During the power-on reset sequence the EFU needs to send data to the chip. It does this when activated by a signal from TCU called tcu_efu_read_start (ioclk domain). This signal is pulsed and released to start the EFU, and may be activated multiple times during the POR sequence.

4.17.2 JTAG Read Access

This allows the user to read each row of the E-fuse Array (EFA) inside the EFU. The EFA is 64 rows by 32 columns. The JTAG instructions in the suggested order of application are:

- TAP_FUSE_READ_MODE Set the mode bits
- TAP_FUSE_ROW_ADDR Specify the row address
- TAP_FUSE_READ Read the specified row

4.17.3 Program Mode

This allows the user to program the EFA one bit at a time in conjunction with proper application of the package pins required for EFA programming (fct_efa_prog_en). The JTAG instructions in the suggested order of application are:

- TAP_FUSE_ROW_ADDR Specify the row address
- TAP_FUSE_COL_ADDR Specify the column address
- Assert appropriate package pins per TI specifications

These steps may be repeated; the addresses will remain active until changed so once a row address is set it need not be changed until all columns have been traversed.

4.17.4 Bypass Mode

This allows the user to bypass the EFA, so that the EFU will treat user-supplied data as if it came from the EFA. This is useful for sending user data to SRAM redundancy registers to verify repairability. The JTAG instructions in the suggested order of application are:

- TAP_FUSE_BYPASS_DATA Send the data to EFU that will be used in place of EFA
- TAP_FUSE_BYPASS Tell the EFU to use the bypass_data and send it out

4.17.5 Sample Mode

This allows the user to sample a redundancy value from an array. The JTAG instructions in the suggested order of application are:

- TAP_FUSE_BYPASS_DATA Specify register to be sampled
- TAP_FUSE_DEST_SAMPLE Request EFU to get data and return it to JTAG

4.17.6 Redundancy Value Clear

To provide a means of clearing redundancy values, the TAP_ FUSE_RVCLR instruction is provided. This allows the user to clear all or specific redundancy values via the eFuse unit. The same mechanism is used by TCU to tell eFuse to clear all redundancy values before initiating an eFuse start sequence during POR.

TABLE 4-17 eFuse Redundancy Value Clear Register

Register	JTAG Instr.	Fields
efu_rvclr[6:0]	TAP_FUSE_RVCLR	efu_rvclr[6] = 1 enables a clear efu_rvclr[5:0] = block_id per eFuse spec.; selects Redundancy Value to clear efu_rvclr[5:0] = 11_1111 will tell eFuse to clear all RVs

4.18 TCU Local CSR Assignments

The base address for TCU is 0x85_0000_0000.

Devices can access the following registers in TCU via the UCB protocol with the offset addresses listed.

In the case of the MBIST Mode and Bypass registers the default value is over-written by logic during the power on reset sequence: the BISI enable bit 1 of the MBIST mode register is written by logic during the power-on reset sequence, and once written it will remain high 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 the MBIST Bypass register will be all 1's in bits 47:0.

4.18.1 Memory BIST Registers

These registers are protected during warm resets unless modified via JTAG.

TABLE 4-18 MBIST Mode Register (0x00)

Bits	Name	Initial Value	R/W	Description
[63:4]	Reserved	X	RW	Reserved
[3]	Loop	0	RW	Loop mode if '1'

TABLE 4-18 MBIST Mode Register (0x00) *(Continued)*

Bits	Name	Initial Value	R/W	Description
[2]	User	0	RW	Diagnostic (user) mode if '1'
[1]	BISI	0 (1 - note 1)	RW	BISI if '1', BIST if '0'
[0]	Parallel	0	RW	Parallel mode if '1'

TABLE 4-19 MBIST Bypass Register (0x08)

Bits	Name	Initial Value	R/W	Description
[63:48]	Reserved	X	RW	Reserved
[47:0]	Bypass	0 (48'hFFFFFFFFFFFF - note 1)	RW	MBIST Bypass

TABLE 4-20 MBIST Start Register (0x10)

Bits	Name	Initial Value	R/W	Description
[63:1]	Reserved	X	W	Reserved
[0]	Start	0	W	Starts MBIST Sequence when written to '1'

TABLE 4-21 MBIST Abort Register (0x18)

Bits	Name	Initial Value	R/W	Description
[63:1]	Reserved	X	W	Reserved
[0]	Abort	0	W	Aborts MBIST Sequence when written to '1'

TABLE 4-22 MBIST Result Register (0x20)

Bits	Name	Initial Value	R/W	Description
[63:2]	Reserved	X	R	Reserved
[1]	Result	0	R	MBIST Done (bit 1)
[0]	Result	0	R	MBIST Fail (bit 0)

TABLE 4-23 MBIST Done Register (0x28)

Bits	Name	Initial Value	R/W	Description
[63:48]	Reserved	X	R	Reserved
[47:0]	Done	0	R	MBIST Done

TABLE 4-24 MBIST Fail Register (0x30)

Bits	Name	Initial Value	R/W	Description
[63:48]	Reserved	X	R	Reserved
[47:0]	Fail	0	R	MBIST Fail

TABLE 4-25 MBIST Start WMR Register (0x38)

Bits	Name	Initial Value	R/W	Description
[63:1]	Reserved	X	W	Reserved
[0]	Start	0	W	Starts MBIST Sequence when written to '1', but delayed until after the next warm reset occurs

4.18.2 Logic BIST Registers

These registers are protected during warm resets unless modified via JTAG.

TABLE 4-26 LBIST Mode Register (0x40)

Bits	Name	Initial Value	R/W	Description
[63:2]	Reserved	X	RW	Reserved
[1]	Program	0	RW	Program mode if '1'
[0]	Parallel	0	RW	Parallel mode if '1'

TABLE 4-27 LBIST Bypass Register (0x48)

Bits	Name	Initial Value	R/W	Description
[63:8]	Reserved	X	R/W	Reserved
[7:0]	Bypass	0	R/W	LBIST Bypass

TABLE 4-28 LBIST Start Register (0x50)

Bits	Name	Initial Value	R/W	Description
[63:1]	Reserved	X	W	Reserved
[0]	Start	0	W	Starts LBIST Sequence when written to '1'

TABLE 4-29 LBIST Done Register (0x60)

Bits	Name	Initial Value	R/W	Description
[63:8]	Reserved	X	R	Reserved
[7:0]	Done	0	R	LBIST Done Status

4.18.3 Debug Control Register

This register can operate during warm resets if enabled by TCU DCR.

TABLE 4-30 Cycle Counter Register (0x100)

Bits	Name	Initial Value	R/W	Description
[63:0]	Cycle Counter	0	RW	See “Cycle Counter” on page 244

CHAPTER 5

Clock Control Unit (CCU)

This chapter contains the following sections:

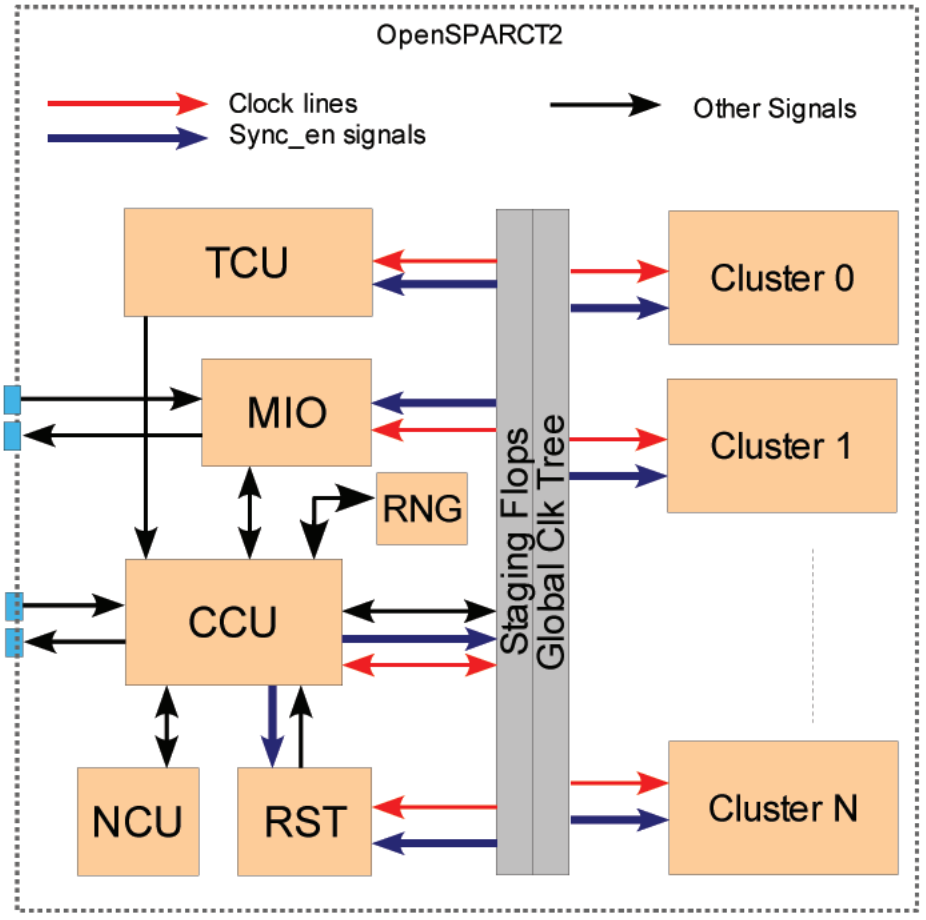
5.1 Overview

This is the microarchitecture specification for the CCU block. It encompasses the following functionality

- PLL to drive the core and memory clocks
- Interfacing with random number generator
- UCB interface for programming the PLLs/MG and reading RMG data
- Provide sync pulses for deterministic clock domain crossing
- Clock stretch and other test clocking mechanisms such as SERDES testing (via DTM) for OpenSPARC T2

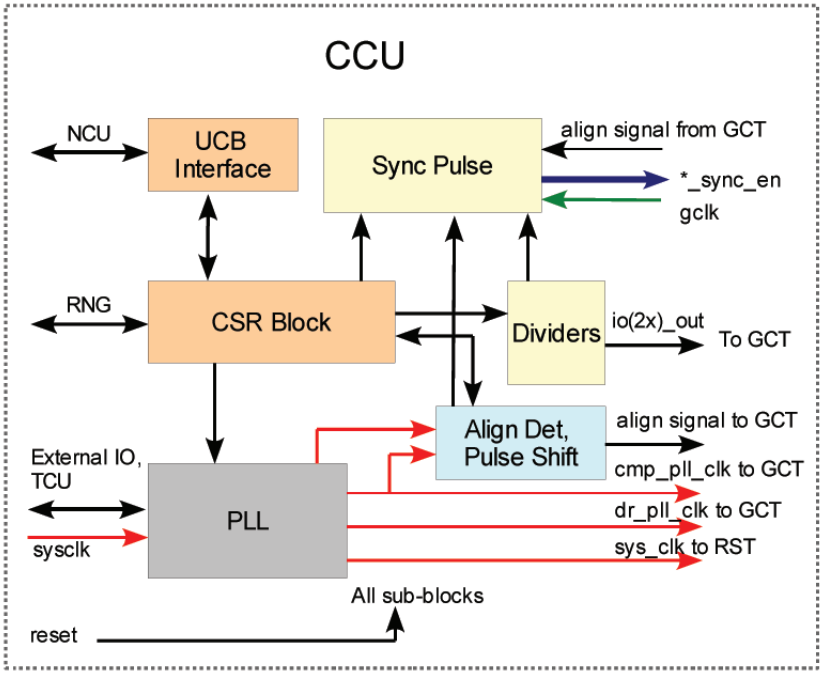
5.1.1 System Block Diagram

FIGURE 5-1 System Block Diagram



5.1.2 CCU Block Diagram and Description

FIGURE 5-2 CCU Block Diagram



Currently, the PLL is the only hard-macro in the CCU. The PLL generates clocks and also performs clock stretching. This is described in Clock and Reset Inside CCU and Sync Pulses, they also includes information on the clock dividers, global clock tree and staging flops.

Interfacing with the RNG block involves an LFSR that can be accessed as a CSR register. Details are given in RNG Description.

UCB interface for programming CSRs is reused from the RST cluster. Its protocol is described in System Block Diagram.

The actual registers resides in the CSR block. It also includes logic for interfacing to the other side of UCB, as described in the UCB interface document. Register addresses and fields are defined in CSR Block.

Sync pulse generation is described in Sync Pulses, with the detailed analysis in the appendices.

5.2 CCU Ports List

TABLE 5-1 CCU Ports List.

Name	Dir	Width	Domain	Description
CLK Input/Outputs				
gclk	I	1	N/A	Input to CCU cluster headers from global clk tree
dr_pll_clk	O	1	N/A	Connect to global clock tree input
cmp_pll_clk	O	1	N/A	Connect to global clock tree input
CCU-NCU Interface				
ccu_ncu_stall	O	1	io	UCB interface between CCU <-> NCU
ncu_ccu_vld	I	1	io	UCB interface between CCU <-> NCU
ncu_ccu_data	I	[3:0]	io	UCB interface between CCU <-> NCU
ncu_ccu_stall	I	1	io	UCB interface between CCU <-> NCU
ccu_ncu_vld	O	1	io	UCB interface between CCU <-> NCU
ccu_ncu_data	O	[3:0]	io	UCB interface between CCU <-> NCU
PLL-Bump Interface				
pll_sys_clk_p	I	1	N/A	Differential input reference to PLL
pll_sys_clk_n	I	1	N/A	Differential input reference to PLL
pll_vdd	I	1	static	PLL VDD – static tie high
mio_ccu_pll_char_in	I	1	async	Direct bump input to PLL – selects internal PLL signal during characterization active when mio_pll_testmode== 1. Also CSR programmable
CCU-RNG Interface				
rng_arst_l	O	1	async	Asynchronous reset of rng, also used to precharge voltage of large caps of the RC filters
rng_data	I	1	async	Input bit stream of random data (combined from up to three noise cells). Loaded into LFSR which is accessible via CSR address RNG_DAT
rng_bypass	O	1	async	Relates to generation of entropy in noise cells *vco control voltage = (bypass) ? output of bias generator : output of feedback amplifier* -- CSR programmable
rng_vcoctrl_sel	O	[1:0]	async	PMOS diode D/A setting bus -- CSR programmable

TABLE 5-1 CCU Ports List. *(Continued)*

Name	Dir	Width	Domain	Description
rng_ch_sel	O	[1:0]	async	Channel select for using entropy from 1,2 or 3 noise cells -- CSR programmable
rng_anlg_sel	O	[1:0]	async	Selects internal analog signal for characterization -- CSR programmable
SCAN/Test Related				
scan_in	I	1	aclk	Scan chain input – (currently hooked up to DMU scan_out output)
tcu_scan_en	I	1	async	Scan enable from TCU
tcu_aclk	I	1	N/A	aclk input to clkgen module. Connect to TCU
tcu_bclk	I	1	N/A	bclk input to clkgen module. Connect to TCU
scan_out	O	1	aclk	Scan chain output – (currently drives RST scan_in port)
tcu_atpg_mode	I	1	async	Puts the CCU in test mode for ATPG testing. Unless this signal is asserted, aclk, bclk and scan inputs into the CCU are all held low, and the scan chain is shorted.
ccu_dbg1_serdes_dtm	O	1	io	Sets DBG1 mux controls for DTM
ccu_mio_serdes_dtm	O	1	io	Sets MIO mux controls for DTM
Global Clock Tree Interface				
ccu_cmp_io_sync_en	O	1	cmp	Sync pulse for cmp -> io clk domain
ccu_io_cmp_sync_en	O	1	cmp	Sync pulse for io -> cmp clk domain
ccu_dr_sync_en	O	1	cmp	Sync pulse for cmp -> dr clk domain
ccu_io2x_sync_en	O	1	cmp	Sync pulse for cmp -> io2x clk domain
ccu_io2x_out	O	1	cmp	Divider phase signal output – rate of CMP clk. Connect to ccu_div_ph of clkgen module in other clusters as needed
ccu_io_out	O	1	cmp	Divider phase signal output – rate of CMP clk. Connect to ccu_div_ph of clkgen module in other clusters as needed
gl_ccu_io_out	I	1	cmp	Divider phase input; similar to ccu_io_out inputs for other clusters.
ccu_vco_aligned	O	1	vco	Align signal tightly coupled to PLL clock domain
gclk_aligned	I	1	cmp	Align signal tightly coupled to (cmp) gclk domain
ccu_serdes_dtm	O	1	async	Places chip in DTM mode where dr_clk == io_clk, and cmp, dr, io clock phases are deterministic. Currently unused in cluster headers.

TABLE 5-1 CCU Ports List. *(Continued)*

Name	Dir	Width	Domain	Description
CCU-MIO Interface				
mio_pll_testmode	I	1	async	Dedicated. Input from external IO through MIO – Used to place PLL in test mode (active high)
ccu_mio_pll_char_out	O	[1:0]	async	Dedicated. Digital characterization output of PLL. Connect to external IO through MIO. Valid when mio_pll_testmode==1
mio_ccu_vreg_selbg_l	I	1	static	Dedicated. Input from external IO through MIO – controls VREG input of PLL and RNG
mio_ccu_pll_clamp_fltr	I	1	static	Shared. Input from external IO through MIO – Used to control clamp filter input of PLL in test mode (mio_pll_testmode==1)
mio_ccu_pll_div2	I	[5:0]	async	Shared. Input from external IO through MIO – Used to program D2 of PLL in test mode (mio_pll_testmode==1)
mio_ccu_pll_div4	I	[7:0]	async	Shared. Input from external IO through MIO – Used to program D4 of PLL in test mode (mio_pll_testmode==1)
mio_ccu_pll_trst_l	I	1	async	Shared. Input from external IO through MIO – Used to reset PLL in test mode (mio_pll_testmode==1)
CCU-TCU Interface				
gl_ccu_clk_stop	I	1	cmp	Clock stop for cmp domain (provisional signal. as of now, no application for it)
gl_ccu_io_clk_stop	I	1	cmp	Clock stop for io domain (provisional signal. as of now, no application for it)
tcu_pce_ov	I	1	async	Overrides clock stop assertion (provisional signal. as of now, no application for it)
tcu_ccu_mux_sel	I	[1:0]	cmp	Controls PLL muxes from TCU – one of four signals to gclk tree inputs: PLL VCO, sysclk, bypass clock, or stretched clock.
tcu_ccu_ext_cmp_clk	I	1	N/A	Bypass clock input for CMP clk (muxed with TCK) from TCU
tcu_ccu_ext_dr_clk	I	1	N/A	Bypass clock input for DR clk (muxed with TCK) from TCU
tcu_ccu_clk_stretch	I	1	cmp	Controls clock stretch in PLL
CCU-RST Interface				
rst_ccu_pll_	I	1	sys	Active low PLL reset – once de-asserted, PLL will start to lock
rst_ccu_	I	1	sys	Active low reset for CCU logic – staggered with respect to rst_ccu_pll_

TABLE 5-1 CCU Ports List. *(Continued)*

Name	Dir	Width	Domain	Description
ccu_sys_cmp_sync_en	O	1	cmp	Special sync pulse for sys -> cmp clk domain – ONLY for RST cluster
ccu_cmp_sys_sync_en	O	1	cmp	Special sync pulse for cmp -> sys clk domain – ONLY for RST cluster
ccu_rst_sys_clk	O	1	N/A	Provides buffered version of sysclk that the PLL is running off
ccu_rst_sync_stable	O	1	cmp	When asserted after PLL has finished locking, indicates to RST block that all clocks and sync pulses are stable
ccu_rst_change	O	1	io	When asserted, indicates to the RST block that the pll divider values WILL change. NOTE. CCU does NOT PERFORM an actual check of old and new divider values. It relies solely on the value of CHANGE field in PLL_REG for the RST to determine if PLL lock required
rst_wmr_protect	I	1	async	Prepares CCU for a warm reset
cluster_arst_l	I	1	async	Holds cluster header output clock low

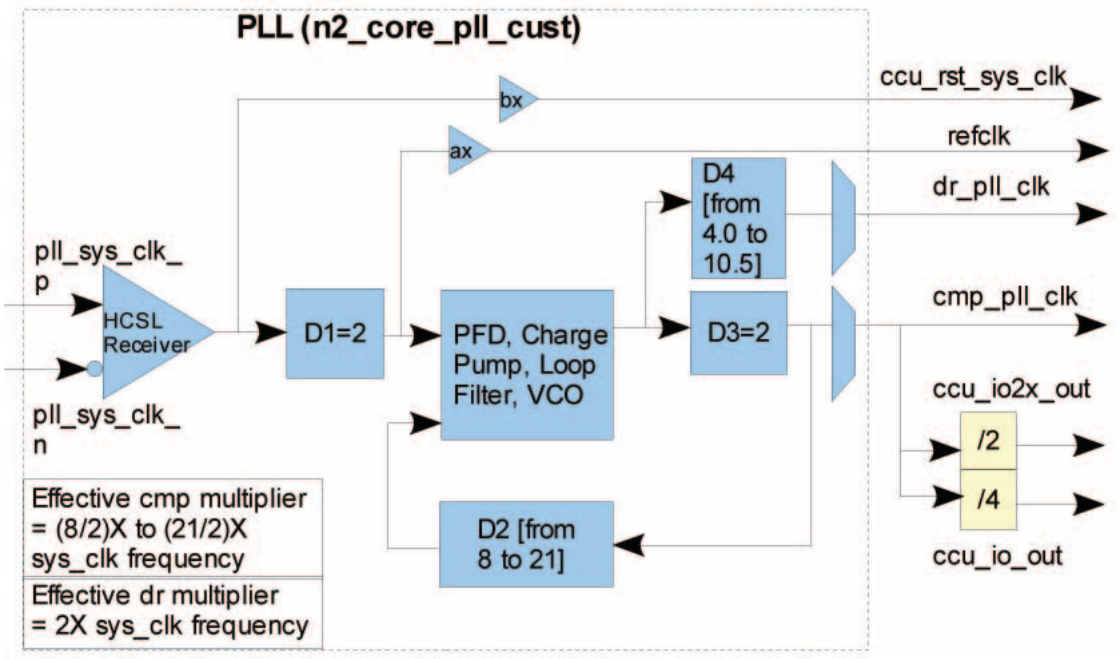
5.2.1 Clock Generation and Distribution

5.2.1.1 Generation

There is one PLL in the CCU for generating both core and memory clocks. The SPARC Cores, CCX and L2 cache operate at the cmp frequency. Parts of the chip are on the io and io2x domains, both of them derived directly from the cmp clock. The other clock that goes into the MCU for interfacing with FBDIMMs is the dr clock.

The same system reference clock is pll_sys_clk_p/n. This is a differential input that is fed from the bumps directly into the core pll. There are two independent dividers that generate cmp and dr clocks, which are rational multiples of each other. Hence, they are *ratioed synchronous*. FIGURE 5-3 gives a simplistic representation of how the PLL generates these clocks.

FIGURE 5-3 PLL Clock Generation during Mission Mode



The other clocks, io_clk and io2x_clk, are derived from the cmp clock, by generating divided down (by four and two respectively), phase signals. They are distributed as clocks from *cluster header* outputs. The distribution is discussed in more detail in Distribution.

Note that the PLL feedback loop is entirely self-contained within the PLL. Thus there is arbitrary phase difference between the rising edge of a sys_clk cycle and a rising edge of a clock into the CLK input of a flop in any cluster.

5.2.2 PLL Programming

The PLL is programmed through a combination of registers (CSR fields), direct chip-level pin control and combinational logic. The CSR based controls are covered in CSR Block. External pin-level control is applicable typically in test mode, and is covered in CCU Testability. This sub-section focuses on divider configuration (CSR programmable) and combinational mux controls from the TCU.

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 OpenSPARC T2 since there is no operation on the low phase.

The divider values are summarized in TABLE 5-2 with information on both effective and actual bits.

TABLE 5-2 PLL Divider Program 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 two. Putting a divide by three 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 TABLE 5-3.

The clock frequency multiplication equations with respect to the external oscillator output (sys_clk) are shown.

fvco = (D2 X D3 /D1) fsys

fcmp = (1 /D3) fvco = (D2 /D1) fsys

fdr = (1 /D4) fvco = (D2 X D3)/(D1 X D4) fsys

fio = fcmp = (D2 /4D1) fsys

fio2x = fcmp = (D2 /2D1) fsys

The first row in any of the three sets in TABLE 5-3 holds the default divider ratio during power-on-reset. The rows in blue (14, 10 and 7) of the three sets refer to the targeted operating frequencies.

TABLE 5-3 Clock Frequency Table in Mission Mode

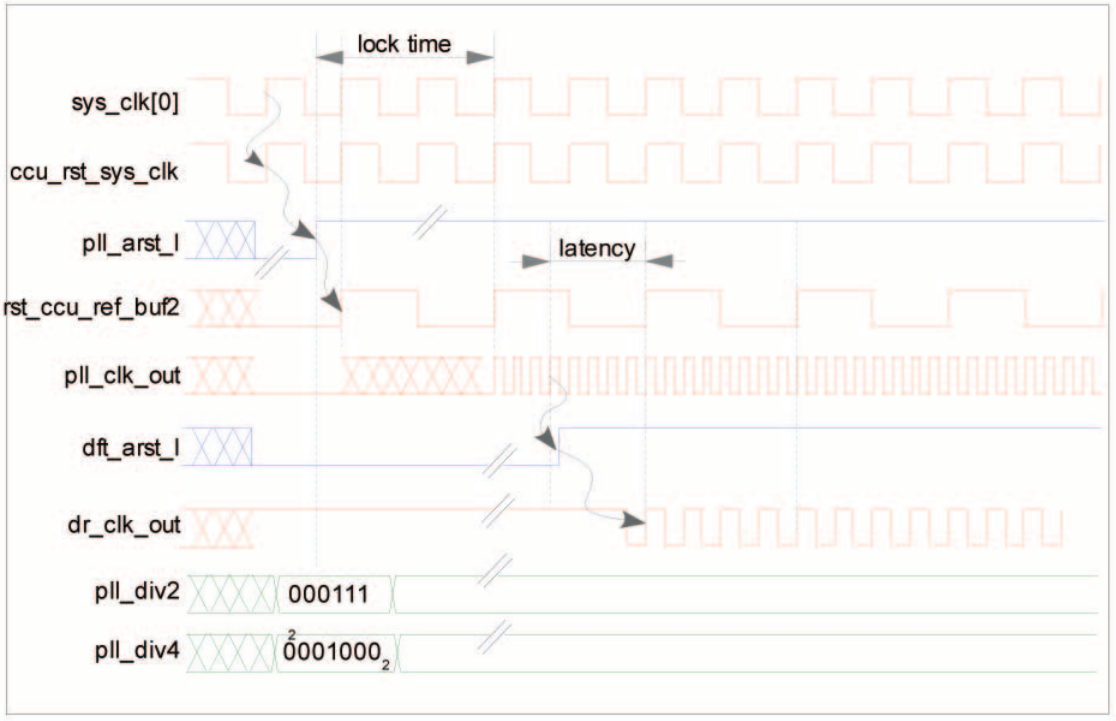
No.	sys_clk (MHz)	D1	D2	D3	D4	D2*D3	VCO (MHz)	cmp_clk (MHz)	io_clk (MHz)	io2x_clk (MHz)	r_clk (MHz)	cmp:dr (ratio)
1	133.33	2	8	2	4.00	16.00	1066.67	533.33	133.33	266.67	266.67	2.00
2	133.33	2	9	2	4.50	18.00	1200.00	600.00	150.00	300.00	266.67	2.25
3	133.33	2	10	2	5.00	20.00	1333.33	666.67	166.67	333.33	266.67	2.50
4	133.33	2	11	2	5.50	22.00	1466.67	733.33	183.33	366.67	266.67	2.75
6	133.33	2	12	2	6.00	24.00	1600.00	800.00	200.00	400.00	266.67	3.00
6	133.33	2	13	2	6.50	26.00	1733.33	866.67	216.67	433.33	266.67	3.25
7	133.33	2	14	2	7.00	28.00	1866.67	933.33	233.33	466.67	266.67	3.50
8	133.33	2	15	2	7.50	30.00	2000.00	1000.00	250.00	500.00	266.67	3.75
8	133.33	2	16	2	8.00	32.00	2133.33	1066.67	266.67	533.33	266.67	4.00
10	133.33	2	17	2	8.50	34.00	2266.67	1133.33	283.33	566.67	266.67	4.25
11	133.33	2	18	2	9.00	36.00	2400.00	1200.00	300.00	600.00	266.67	4.50
12	133.33	2	19	2	9.50	38.00	2533.33	1266.67	316.67	633.33	266.67	4.75
13	133.33	2	20	2	10.00	40.00	2666.67	1333.33	333.33	666.67	266.67	5.00
14	133.33	2	21	2	10.50	42.00	2800.00	1400.00	350.00	700.00	266.67	5.25
1	166.67	2	8	2	4.00	16.00	1333.33	666.67	166.67	333.33	333.33	2.00
2	166.67	2	9	2	4.50	18.00	1500.00	750.00	187.50	375.00	333.33	2.25
3	166.67	2	10	2	5.00	20.00	1666.67	833.33	208.33	416.67	333.33	2.50
4	166.67	2	11	2	5.50	22.00	1833.33	916.67	229.17	458.33	333.33	2.75
6	166.67	2	12	2	6.00	24.00	2000.00	1000.00	250.00	500.00	333.33	3.00
6	166.67	2	13	2	6.50	26.00	2166.67	1083.33	270.83	541.67	333.33	3.25
7	166.67	2	14	2	7.00	28.00	2333.33	1166.67	291.67	583.33	333.33	3.50
8	166.67	2	15	2	7.50	30.00	2500.00	1250.00	312.50	625.00	333.33	3.75
8	166.67	2	16	2	8.00	32.00	2666.67	1333.33	333.33	666.67	333.33	4.00
10	166.67	2	17	2	8.50	34.00	2833.33	1416.67	354.17	708.33	333.33	4.25
11	166.67	2	18	2	9.00	36.00	3000.00	1500.00	375.00	750.00	333.33	4.50
12	166.67	2	19	2	9.50	38.00	3166.67	1583.33	395.83	791.67	333.33	4.75

TABLE 5-3 Clock Frequency Table in Mission Mode *(Continued)*

No.	sys_clk (MHz)	D1	D2	D3	D4	D2*D3	VCO (MHz)	cmp_clk (MHz)	io_clk (MHz)	io2x_clk (MHz)	r_clk (MHz)	cmp:dr (ratio)
13	166.67	2	20	2	10.00	40.00	3333.33	1666.67	416.67	833.33	333.33	5.00
14	166.67	2	21	2	10.50	42.00	3500.00	1750.00	437.50	875.00	333.33	5.25
1	200	2	8	2	4.00	16.00	1600.00	800.00	200.00	400.00	400	2.00
2	200	2	9	2	4.50	18.00	1800.00	900.00	225.00	450.00	400	2.25
3	200	2	10	2	5.00	20.00	2000.00	1000.00	250.00	500.00	400	2.50
4	200	2	11	2	5.50	22.00	2200.00	1100.00	275.00	550.00	400	2.75
6	200	2	12	2	6.00	24.00	2400.00	1200.00	300.00	600.00	400	3.00
6	200	2	13	2	6.50	26.00	2600.00	1300.00	325.00	650.00	400	3.25
7	200	2	14	2	7.00	28.00	2800.00	1400.00	350.00	700.00	400	3.50
8	200	2	15	2	7.50	30.00	3000.00	1500.00	375.00	750.00	400	3.75
8	200	2	16	2	8.00	32.00	3200.00	1600.00	400.00	800.00	400	4.00
10	200	2	17	2	8.50	34.00	3400.00	1700.00	425.00	850.00	400	4.25
11	200	2	18	2	9.00	36.00	3600.00	1800.00	450.00	900.00	400	4.50
12	200	2	19	2	9.50	38.00	3800.00	1900.00	475.00	950.00	400	4.75
13	200	2	20	2	10.00	40.00	4000.00	2000.00	500.00	1000.00	400	5.00

PLL output clock behavior with respect to its reset signals and sys_clk are shown. They are produced independently of the PLL specification. The example FIGURE 5-4 assumes D1 and D3 are left in their default states (effective value of two for both).

FIGURE 5-4 PLL Clocking Waveforms



5.2.3 PLL Mux Control

All functional clock muxing in OpenSPARC T2 is performed in custom blocks – the PLL, cluster headers and specialized L1 headers. However, the PLL clock mux control logic is divided between the CCU and the PLL hard macro, with the CCU often exercising additional constraints.

The combination of CCU inputs, and equivalent mappings to PLL inputs are tabulated in TABLE 5-4. For actual usage related to various modes of operation, refer to CCU Testability of this MAS.

TABLE 5-4 CCU and PLL Mapping

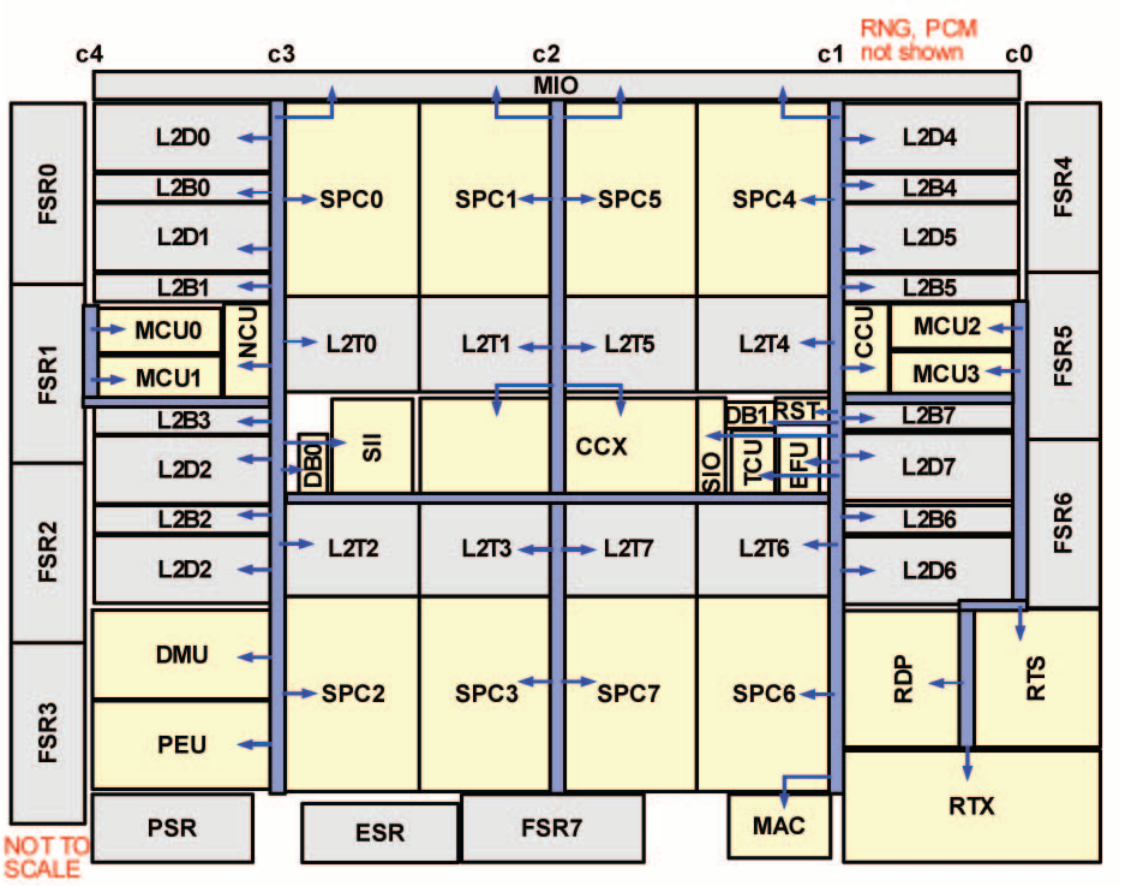
CCU or CSR inputs				PLL inputs				PLL outputs		
Mode	dtm[1,2]	atpg_mode	mux_sel	pll_arst_l	pll_bypass	pll_sel_a	pll_dtm	dr_sel_a	pll_clk_out	dr_clk_out
Func	0	0	00	rst_ccu_pll_	0	00	0	00	sys_clk x N	sys_clk x M
Str	0	0	01	rst_ccu_pll_	0	01	0	01	sys_clk x N (str)	sys_clk x M (str)
ATPG	0	1	10	0	1	10	1	10	ext_cmp_clk	ext_dr_clk
Byp	0	1	11	0	1	11	1	11	sys_clk_p	sys_clk_p
DTM	1	0	00	rst_ccu_pll_	0	00	1	11	sys_clk x N	sys_clk_p
MTest	0	0	11	0	1	10	1	10	ext_cmp_clk	ext_dr_clk

FIGURE 5-3 and FIGURE 5-19 complement the information in the table. Note that the PLL input and output signals above will be allowed to change based on any sequential logic within the CCU or PLL. For example, if DTM 1 or 2 mode is programmed into the CSRs, during PLL reset, *pll_bypass* will be held high to avoid internal PLL clock mux contention, and then set to 0 upon reset release. Under all other conditions, *pll_bypass* will simply be assigned to *tcu_ccu_mux_sel[1]*.

5.2.4 Distribution

As shown in FIGURE 5-2, cmp_clk and dr_clk are distributed via a global clock tree to gclk inputs of various clusters. Each cluster receives the same phase of gclk. The CCU also sends out a few control signals that are distributed closely with gclks and pipelined on various tap points of the global clock tree. A high level diagram is shown in FIGURE 5-5.

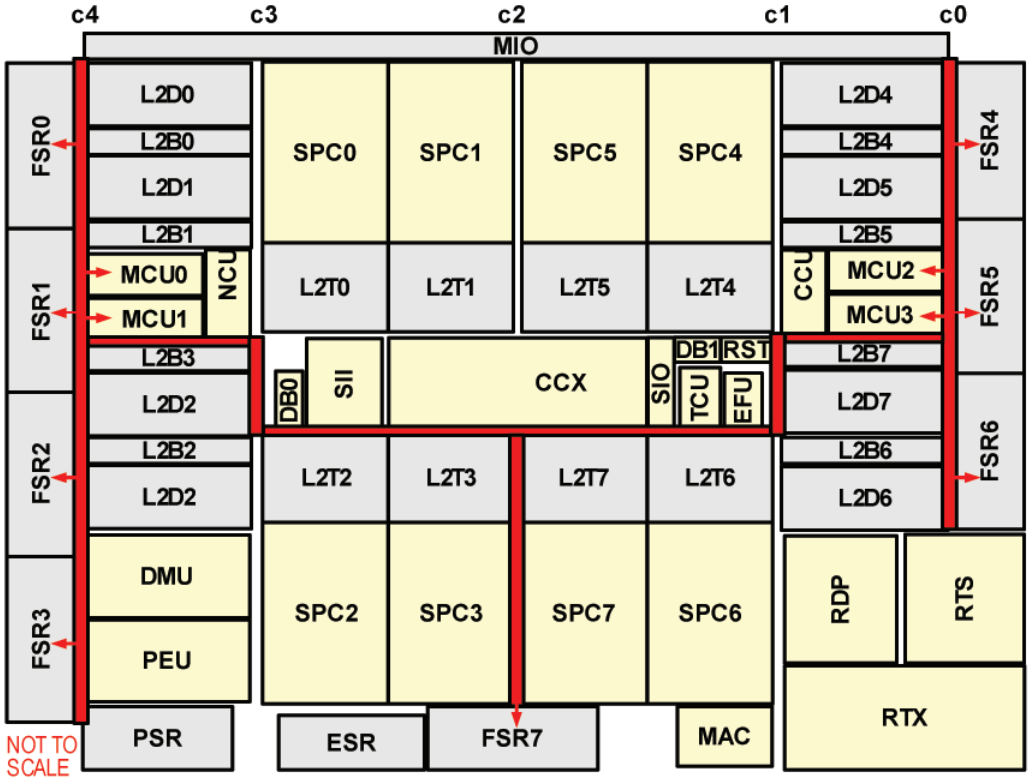
FIGURE 5-5 Simplified Global Distribution of CMP Clock



There are a set of other global signals from the CCU such as the divider phase signals, and sync pulses that are staged in the global clock tree, along with reset lines from the RST and clock stops from the TCU. All these staged signals are sent out on the CMP domain, or in some cases on the DR domain. Synchronization to the clock outputs is performed in the cluster header, as described in the Usage document.

For completeness and references to it in other sections, DR distribution is shown in FIGURE 5-6.

FIGURE 5-6 Global Distribution of the DR Clock



5.3 Clock and Reset Inside CCU

5.3.1 Clock Domains

There are three clock domains within the CCU: L2, IOL2, and CMP_PLL. There is a distinction between CMP_PLL and L2 domains. The L2 domain is synchronous to all other clusters. CMP_PLL domain is the result of using the PLL output clock at CMP rate prior to distributing the clock through the GCLK macro. L2 clock is a phase shifted version of CMP_PLL clock; the phase shift due to distribution can vary from process to process from 0.5 to ~1.5 CMP periods.

Note – There is temporarily a fourth clock domain (SYSCLK) to work around missing latch models in the std cell library. Once the latch becomes available, SYSCLK domain will be removed.

A breakdown of the CCU by clock domain and approximate functionality appears in FIGURE 5-7. FIGURE 5-8 details the clock align detection logic since it uses non-conventional clocking and has a few special constraints:

The first pair of flops act as synchronizers should the negative edge of the reference clock be sampled by the flops (the outcome of the align detection is immaterial since it is zero in this case).

There is a half-cycle, or 2X clocking path where data from negative clocked flops gets transferred to positive edge-triggered flops.

5.3.2 Reset Scheme

The CCU relies on the RST block for explicit reset signals, and does not operate via flush reset. Also, it needs to be released from reset before all other blocks on the chip. One reset is solely for the PLL, and the other for the remaining CCU logic, loosely speaking. The CCU itself needs to generate one or two staggered resets. These resets work in a domino like fashion to ultimately provide a signal to the RST unit that indicates the CCU is done with initialization, and that the RST block may release the rest of the chip from reset. This signal is ccu_rst_sync_stable. When the signal goes high, all clocks from the CCU are valid, at the correct frequency, and all sync pulses are operating in their proper positions.

Depending on whether clocks may be stable or not, the CCU needs to use either asynchronous or synchronous reset. However, all resets within the CCU are released synchronously. Emphasis has been placed on determinism and repeatability, so even where brute-force synchronization is used, additional signals ensure determinism.

There is only one CSR register in the CCU that is warm reset protected. All clock generating and pll programmation bits are warm reset protected. The rest are not.

5.3.3 Initialization Sequence

The Power-On-Reset scheme in the CCU is highlighted by the waveforms in FIGURE 5-9. For functional operation, the CCU is activated in a very simple manner. There are two resets to the CCU, ccu_rst_pll_ and ccu_rst_ that need to be applied in a sequence. Testmode, and divider_bypass pins need to be held low.

An explanation of the various numbered parameters is given in TABLE 5-5.

TABLE 5-5 Key Parameters in Initialization Sequence

Parm #	Description	Duration
1	Time taken for first rising edge of refclk to appear from release of rst_ccu_pll_	<1 sys_clk cycle
2	Deassertion of rst_ccu_pll_ to rising edge of stable CMP PLL clock output	LOCK TIME
3	Clock distribution delay of global clock tree from PLL output to gclk input of cluster header	~0.5 – ~1.3 CMP cycle
4	Deassertion of rst_ccu_ to gclk_rst_n (requires use of brute force synchronizer)	1 to 2 CMP cycles
5	Rising edge of refclk to assertion of aligned_shift pulse.	3 CMP cycles
6	Shift of aligned_shift pulse to create VCO aligned	4 to 17 CMP cycles depending on pll_div2[5:0]
7	Transfer of aligned signal from CMP PLL domain to CMP_GCLK domain.	Tracks parameter #3
8	From first aligned pulse to aligned_rst_n signal for internal CCU blocks for coherent reset release.	1 CMP cycle
9	Deassertion of aligned_rst_n to first rising edge of ccu_io2x_out	2 CMP cycles
10	Deassertion of aligned_rst_n to first rising edge of ccu_io_out	4 CMP cycles
11	Time when aligned == 1 to deassertion of divider for generating DR clock within PLL	2-3 CMP cycles depending on pll_div4[6:0]
12	Deassertion of dft_a_rst_l to first rising edge of dr_clk	5-6 CMP cycles depending on pll_div4[6:0]

FIGURE 5-7 CCU Clock Domains and Function

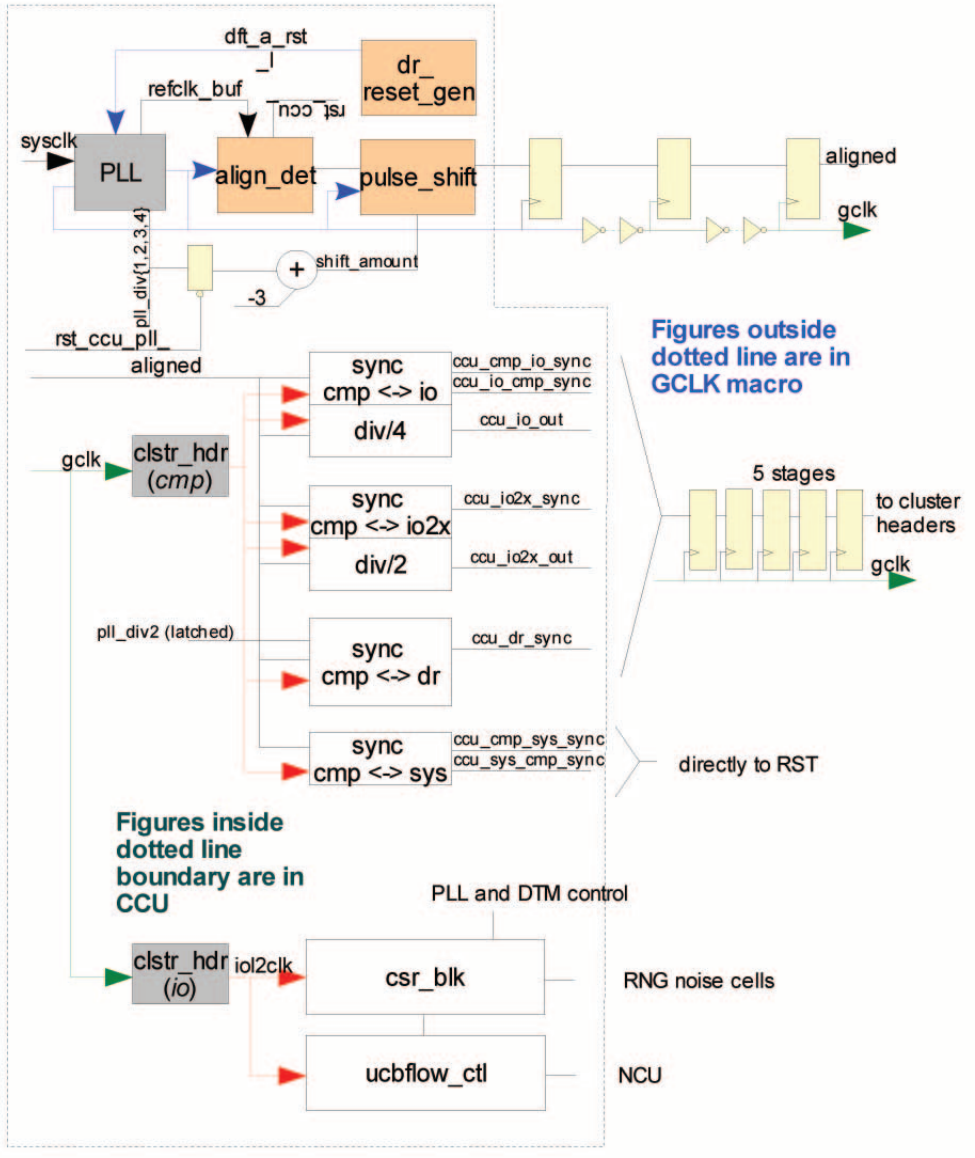


FIGURE 5-8 Align Detection Circuitry

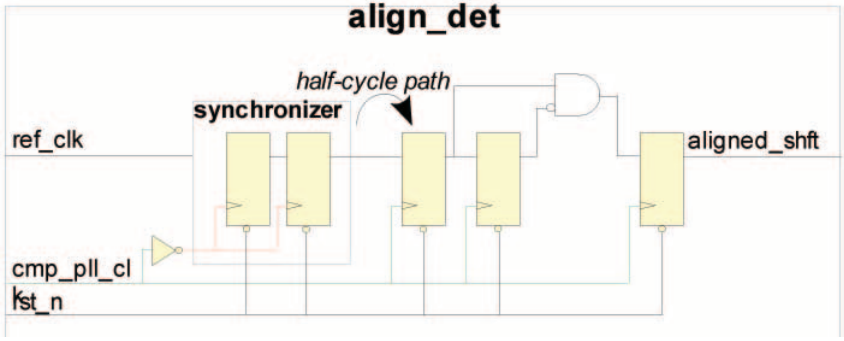
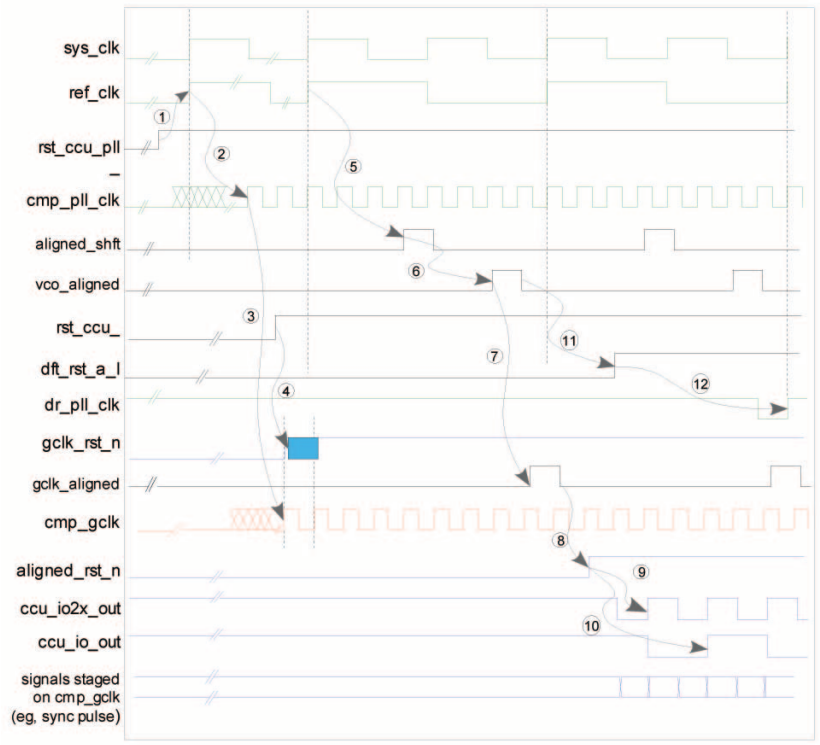

align_det
half-cycle path
synchronizer
ref_clk
aligned_shft
cmp_pll_cl
rst_n

FIGURE 5-9 Initialization Sequence for CCU Clocks



5.4 Sync Pulses

The main application of generating synchronization pulses in OpenSPARC T2 is to allow low latency, deterministic data transfer between *ratioed synchronous* clock domains. The key requirements for this scheme to work are:

A single reference clock source.

PLLs that have similar behavior, in particular a known input-output phase relationship.

The clock frequencies need to be rational multiples of each other, or *ratioed synchronous*

Jitter, skew, and other PVT mismatches are taken into account to ensure setup and hold requirements are met during domain crossing.

Clock domains that are of primary concern are the CMP and DR domains. Synchronization between cmp and IO, or IO2X domains is a simpler problem, but handled similarly.

5.4.1 Proposed Scheme

The following circuit shows the proposed scheme for clock domain transfers.

FIGURE 5-10 CMP to DR Synchronization

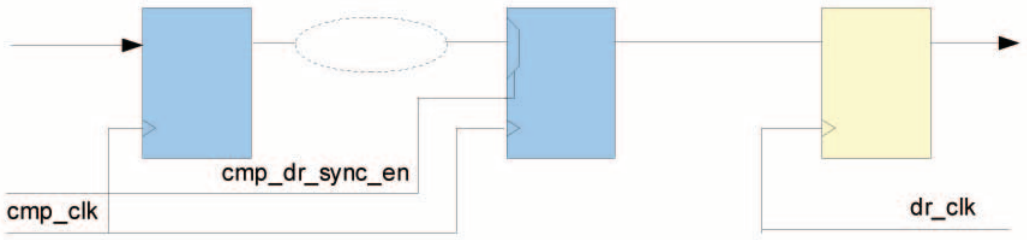
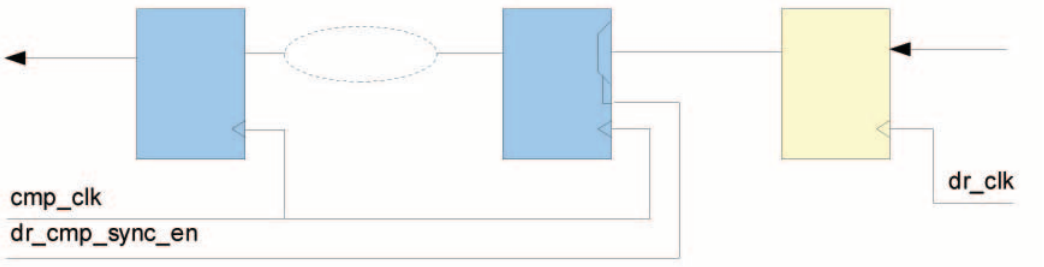


FIGURE 5-11 DR to CMP Synchronization



It has been borrowed from past designs and modified. All it does is allow data to cross one domain to another during a safe interval, avoiding setup and hold problems. The mechanism for operation for fast clock (e.g., cmp) to slow clock (e.g., dr) domain is as follows:

Mux enable to launch flip-flop is generated on cmp_clk.

Next cmp rising edge, data is launched.

Data is captured on dr_clk.

For slow clock to fast clock transfers, the procedure is:

Data is launched on rising edge of dr_clk.

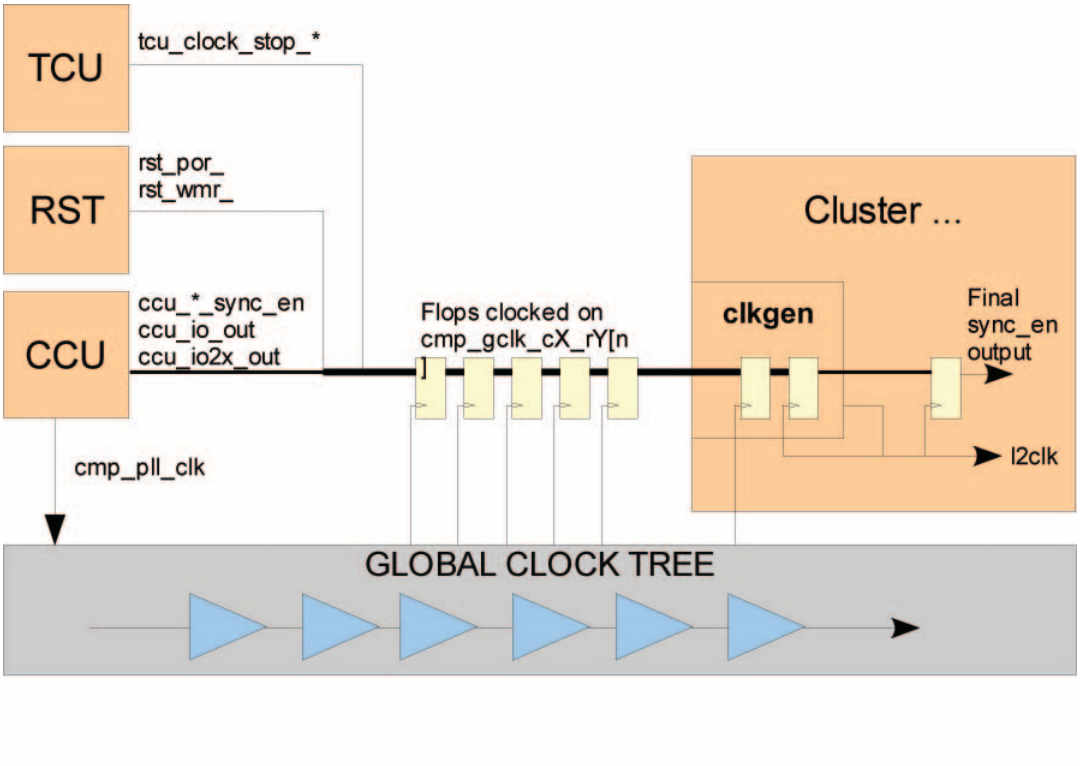
Mux enable to capture flip-flop is generated on cmp_clk.

Next cmp rising edge, data is captured.

In both cases, the rate of communication is limited by the slower clock frequency, so the enable is generated once every slow clock cycle. The main challenge is to determine the ideal intervals between pulse generation for robust operation. For a discussion on determining the positions, refer to Appendix A.1 – Sync Pulse Design Procedure.

5.4.2 Sync Pulse Distribution

FIGURE 5-12 Logical Representation of Sync Pulse Global Distribution



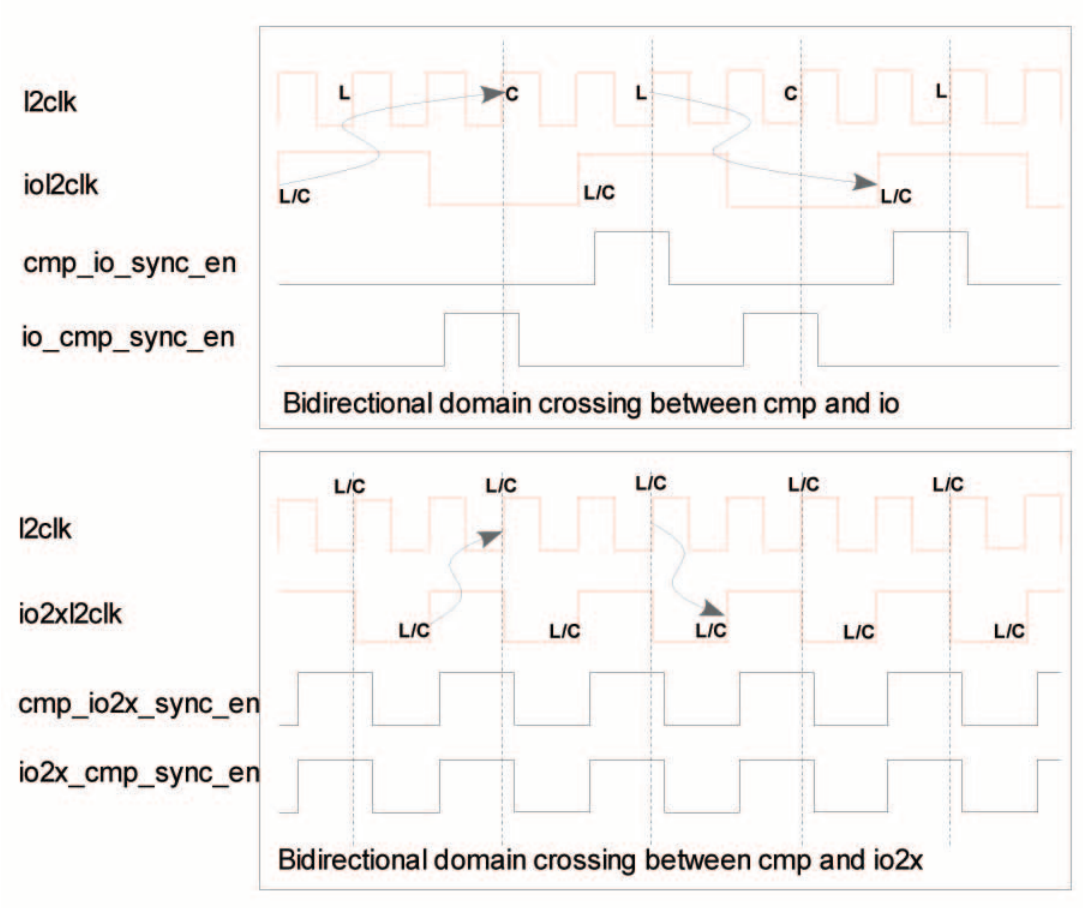
Sync pulses will be generated in the CCU on the cmp_gclk domain, and be distributed (along with other control signals) in five stages of pipeline in mini-clusters to each cluster header. In the cluster headers, there will be one more stage of latching the data on the gclk domain. From there, each cluster will flop the enables on the l2clk domains before local distribution. In effect, there will be seven stages of cmp_cycle before sync pulses are output from cluster headers, and then flopped one last time within clusters.

5.4.3 CMP to IO/IO2X Waveforms

Domain crossing between CMP and IO/IO2X domains is a special, and simpler case of CMP to DR communication because cmp_clk is an integer multiple of io_clk and io2x_clk, and both io_clk and io2x_clk are directly derived from cmp_clk

FIGURE 5-13 shows the actual usage, i.e., the final sync_en output (refer to FIGURE 5-9).

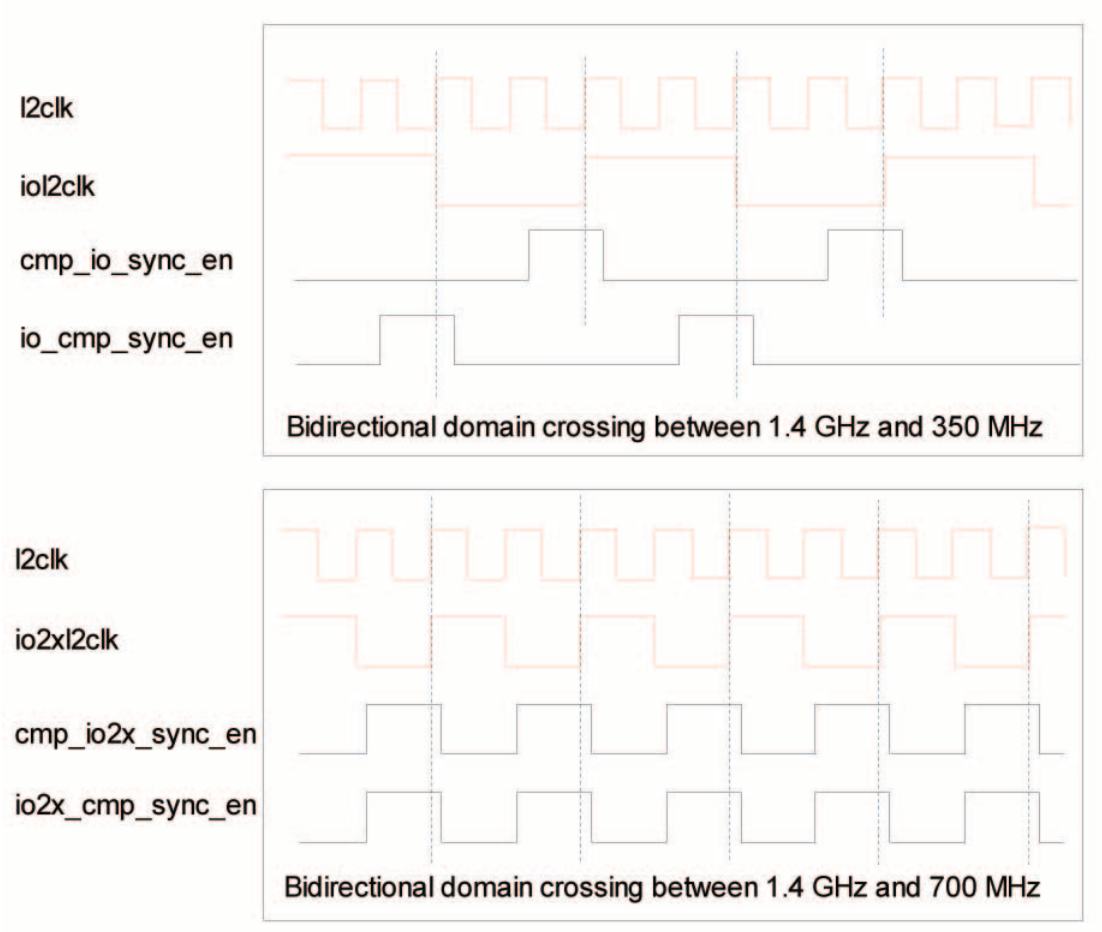
FIGURE 5-13 Actual Usage of Sync Pulses at Enable Pin of Transfer Flops



Note – Since *cmp_io2x_sync_en* and *io2x_cmp_sync_en* are shown at the point of usage; however, they would both be driven by a single source – *cluster header->io2x_sync_en ->flop output*.

For clarity, the outputs of cluster headers are also shown. These are, as expected from FIGURE 5-9, one l2clk cycle early.

FIGURE 5-14 Sync Enable Positions at the Outputs of Cluster Headers prior to being latched.

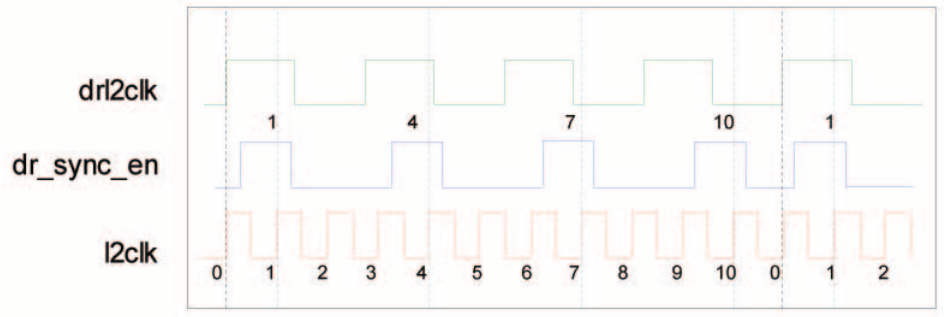


5.4.4 CMP/DR Pulses

CMP to DR pulse positions are determined by the amount of uncertainty that can exist between cmp_clk and dr_clks. A discussion on the procedure of determining the positions appears in the AAppendix A.1 – Sync Pulse Design Procedure. There

are several documents detailing the sync pulse schemes and timing budgets that have been created to ensure robustness. An example of the positions of the dr sync pulses is shown in FIGURE 5-15.

FIGURE 5-15 Sync Pulse Example for fCMP:fDR = 11:4



The convention is to describe the sync pulse position in terms of cmp clk phases, with phase 0 being set to the nominal alignment of cmp and dr clocks. The sync pulse positions at the point of domain crossing are given in TABLE 5-6.

TABLE 5-6 DR<->CMP Sync Pulse Positions

CMP<->DR Transfer Edge								Transfer phase (normalized for four dr=2pi)			
				K - > clk cycles				K - > clk cycles			
N	M	Meff	N/M	0	1	2	3	0	1	2	3
8	4	1	2.00	1	1	1	1	1	3	5	7
9	4	4	2.25	1	3	6	8	1	3	6	8
10	4	2	2.50	1	4	1	4	1	4	6	9
11	4	4	2.75	1	4	7	10	1	4	7	10
12	4	1	3.00	1	1	1	1	1	4	7	10
13	4	4	3.25	2	5	8	11	2	5	8	11
14	4	2	3.50	2	5	2	5	2	5	9	12
15	4	4	3.75	2	6	9	13	2	6	9	13
16	4	1	4.00	2	2	2	2	2	6	10	14
17	4	4	4.25	2	6	11	15	2	6	11	15
18	4	2	4.50	2	7	2	7	2	7	11	16

TABLE 5-6 DR<->CMP Sync Pulse Positions *(Continued)*

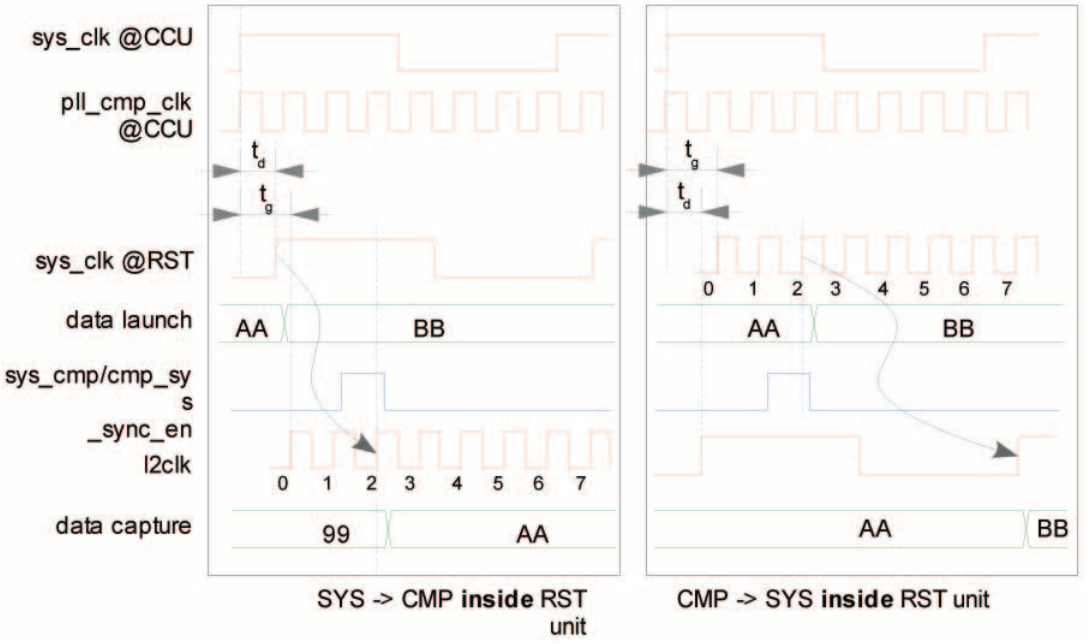
			CMP<->DR Transfer Edge					Transfer phase (normalized for four dr=2pi)			
				K - > clk cycles				K - > clk cycles			
19	4	4	4.75	2	7	12	17	2	7	12	17
20	4	1	5.00	2	2	2	2	2	7	12	17
21	4	4	5.25	3	8	13	18	3	8	13	18

5.4.5 CMP/SYS Pulses

There are a pair of sync pulses between CMP and SYS_CLK strictly for the RST unit. These pulses are not staged on the global clock tree, and not taken in through cluster headers. However, to account for fanout, the signals are flopped twice inside the RST cluster. The scheme relies on the RST block being placed close to the CCU; there is tolerance built in for skew between the CMP and SYS_CLK up to a couple of CMP cycles.

The active position of the sync pulse (“1” on rising edge of cmp_clk) will be on phase two of l2clk. This will provide ample margin, > 1 *fast* cmp cycle for setup or hold. Illustrations of data transfers in both directions are shown in FIGURE 5-16. For quantification of the amount of margin available, refer to Appendix A.1 – Sync Pulse Design Procedure.

FIGURE 5-16 Domain Crossing using Sync Pulses in RST



5.5 RNG Description

The random number generator (RNG) generates random numbers from three noise cells. There is one RNG block (and LFSR) to be shared amongst 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 5-7 Encoding for Noise Cell Selection

CTL3	CTL2	CTL1	Effect
0	0	0	Deselect all noise cells (feeds 0 into LFSR)
0	0	1	Select noise cell 1

TABLE 5-7 Encoding for Noise Cell Selection *(Continued)*

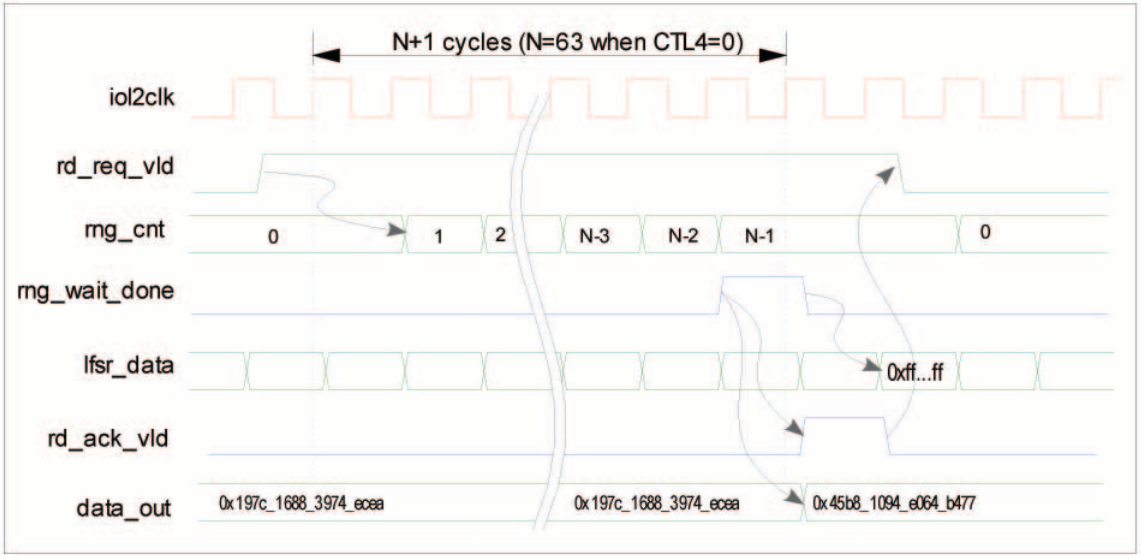
CTL3	CTL2	CTL1	Effect
0	1	0	Select noise cell 2
1	0	0	Select noise cell3
011, 101, 110, 111			Select all 3 noise cells

Every clock cycle, the XOR of the outputs of the selected noise cells is fed into a 64-bit register. Under functional mode, the register generates data by implementing the CRC-polynomial

$$P(x) = x^{64} + x^{61} + x^{57} + x^{56} + x^{52} + x^{51} + x^{50} + x^{48} + x^{47} + x^{46} + x^{43} + x^{42} + x^{41} + x^{39} + x^{38} + x^{37} + x^{35} + x^{32} + x^{28} + x^{25} + x^{22} + x^{21} + x^{17} + x^{15} + x^{13} + x^{12} + x^{11} + x^{7} + x^{5} + x + 1$$

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 0xFFFF_FFFF_FFFF_FFFF. 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 CSR register.

FIGURE 5-17 Read Access Operation of rng_data via Memory Mapped Address

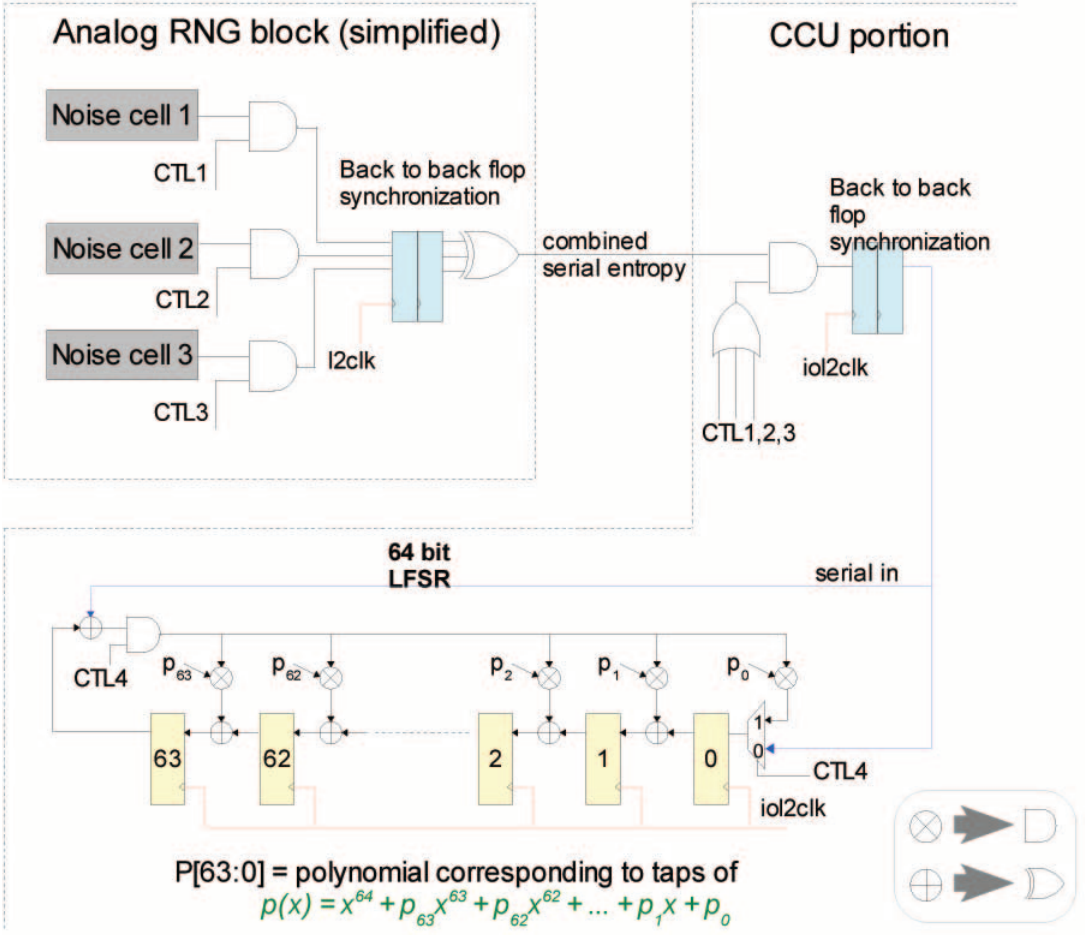


In diagnostic mode (CTL4 = 0), the LFSR acts as a simple shift register capturing the noise cell output directly, determined independently by CTL1 CTL2 and CTL3 as per encoding. The additional constraint in this mode is that successive read requests for the rng_data will be delayed by 64 iol2clk cycles. Also, flushing the LFSR after every read will be disabled in this mode.

The nominal frequency of the oscillator in each noise cell can be set independently by programming the rng_vco_ctrl[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 desired configuration: noise cell1 -> 00 setting, cell2 -> 10 setting, cell 3 --> 01 setting, and observe all three 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

FIGURE 5-18 Entropy Generator Design



5.6 CSR Block

The CSR block consists of registers that can be used for programming other CCU blocks, and for accessing information. This includes both functional and test related data. The other part of the CSR block communicates with the standard UCB interface.

Note that values written into the PLL_CTL register will not take effect immediately (even though reading them back will show the new values). A warm reset needs to be applied to affect clocks.

5.6.1 PLL_CTL (*0x83_0000_0000)*

TABLE 5-8 PLL Control Register

Field Name	Bits	Default	WMR Protected	R/W	Description
Reserved	63:37	0x0	N/A	R	Reserved
pll_clamp_fltr	36	0x0	YES	R/W	PLL clamp filter setting
st_delay_dr	34:35	0x0	YES	R/W	DR stretch delay setting (40ps intervals) 00 –> 40 , 01 -> 80 , 10 -> 120 , 11 -> 160
pll_char_in	33	0x0	YES	R/W	PLL characterization test input
change	32	0x1	YES	R/W	PLL frequency to be changed
align_shift	30:31	0x0	YES	R/W	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.
serdes_dtm2	29	0x0	YES	R/W	Mode 2 – causes ccu_serdes_dtm to be asserted during reset. io, io2x set to DR rate. Used by DBG1/MIO for selecting setting mux controls.
serdes_dtm1	28	0x0	YES	R/W	Mode 1 – causes ccu_serdes_dtm to be asserted during reset. io, io2x set to DR rate. Used by DBG1/MIO for selecting setting mux controls.
st_delay_cmp	27:26	0x0	YES	R/W	CMP stretch delay setting (40ps intervals) 00 –> 40 , 01 -> 80 , 10 -> 120 , 11 -> 160
st_phase_hi	25	0x0	YES	R/W	High or low phase of clk to be stretched 0 indicates low phase.
pll_div4	24:18	0x8	YES	R/W	PLL VCO divider (D4) for dr. Refer to PLL Programming section.
pll_div3	17:12	0x1	YES	R/W	PLL VCO divider (D3) for cmp. Refer to PLL Programming section.
pll_div2	11:6	0x7	YES	R/W	PLL feedback divider (D2). Refer to PLL Programming section.
pll_div1	5:0	0x1	YES	R/W	PLL pre-scalar (D1). Refer to PLL Programming section.

5.6.2 RNG_CTL (0x83_0000_0020)

TABLE 5-9 RNG Control Register

Field Name	Bits	Default	WMR Protected	R/W	Description
Reserved	63:25	0x0	N/A	R	Reserved
rng_wait_cnt	24:9	0x003E	NO	R/W	Minimum wait time before successive RNG data is sent
rng_bypass	8	0x0	NO	R/W	rng_bypass=0 sets noise cell vco control voltage = output of feedback amplifier rng_bypass=1, sets noise cell vco control voltage = output of bias generator
rng_vcoctrl_sel	7:6	0x0	NO	R/W	pmos diode D/A setting bus. Controls VCO rate for each noise cell. Refer to RNG Description for programming.
rng_anlg_sel	5:4	0x0	NO	R/W	Analog mux select for characterization
rng_ctl4	3	0x1	NO	R/W	Enables using LFSR or plain shift register. Set to LFSR mode by default.
rng_ctl3	2	0x1	NO	R/W	Control for using noise cell 3. Refer to RNG Description for programming.
rng_ctl2	1	0x1	NO	R/W	Control for using noise cell 2. Refer to RNG Description for programming.
rng_ctl1	0	0x1	NO	R/W	Control for using noise cell 1. Refer to RNG Description for programming.

5.6.3 RN

G_DATA (0x83_0000_0030)

TABLE 5-10 RNG Data Register

Field Name	Bits	Default	WMR Protected	R/W	Description
rng_data	63:0	x	N/A	R	64 bits of rng data

5.7 CCU Testability

This section deals with the testability of the CCU logic and the PLL.

5.7.1 CCU ATPG

The CCU logic is scannable only during ATPG testing. In mission mode, the scan chain input to the CCU is short-circuited to scan_out, and the following signals are set to zero: tcu_aclk, tcu_bclk, tcu_scan_en.

When testmode is set to 1, the TCU gets full control of the CCU, and treats the CCU just like any other logic on the chip being scanned. The TCU also controls the clock muxes within the PLL via tcu_ccu_mux_sel (refer to FIGURE 5-15). When tcu_ccu_mux_sel == 2b'10, the external clocks tcu_ccu_ext_cmp_clk and tcu_ccu_ext_dr_clk are muxed into the cmp_gclk and dr_gclk buffer trees respectively. These external clocks in turn are muxed inside the TCU with TCK, such that TCK can be forced onto both lines, or be controlled by the tester independently.

The only portion of the CCU that is not scannable is logic that does not run on the regular l2clk or iol2clk. This results in about a dozen flops that are kept out of the scan chain at all times.

The CCU PLL may be put in testmode, asserting the signal mio_pll_testmode via an external pin. This signal is independent of the testmode signal used for ATPG. However, tcu_atpg_mode has higher priority than mio_pll_testmode**.** For example, when both are asserted, ccu_pll**.** pll_arst_l will be set to 0.

With pll_testmode == 1, the CCU provides access to the PLL directly through the following signals from the MIO:

```
mio_ccu_vreg_selbg_l

mio_ccu_pll_clamp_fltr

mio_ccu_pll_div2

mio_ccu_pll_div4

mio_ccu_pll_trst_l

mio_ccu_pll_char_in
```

Likewise, it is possible to observe the internal signals from the PLL through a pair of muxed (internal to PLL) outputs ccu_mio_pll_char_out.[1:0].

When pll_testmode is active, rst_ccu_pll_ has no effect on resetting the PLL, and the CSR values for PLL control are overridden. The exception is the pll_char_in signal which is OR'ed with the pll_reg bit 33 output.

5.8 Full Chip Testability

Clock Control Unit (CCU) describes the role of the CCU, and its features for supporting full chip testability.

5.8.1 Full Chip ATPG

The support provided by the CCU for full chip ATPG is no different from setting the CCU itself in ATPG mode. The same procedure for setting the CCU for ATPG mode is followed, while the CCU becomes merely a conduit for forwarding clocks.

The custom global clock tree does not perform any clock gating, so test clocks injected onto the main line are never blocked. Within the cluster header, in testmode, the clock stop signal is permanently disabled, ensuring that the test clocks into any cluster are free running with direct tester control. In addition, parts of the cluster header are fully scannable, while tcu_clk_stop and ccu_div_ph inputs of the header are observable.

Note that the sync pulses between all ratio'ed synchronous domains in each cluster would have to be set to a logic 1 to allow scan capture to take place consistently. This control is outside the scope of the CCU.

5.8.2 Transition Fault Test

During transition fault testing, the CCU needs to be fully functional, as do the cluster headers, since it would not be possible to apply at-speed scan capture pulses through the external clock ports from the tester.

They will be kept out of the scan chain by ensuring the external signal testmode = 0. The operation of tcu_clock_stop will be critical in ensuring t-fault testing is programmable to provide two or more high-speed pulses. The staging flops in the global clock tree macro, of course will be free running on gclk, and have no scope of blocking clock stop. TCU will have full control of the cluster and domain that will be tested.

5.8.3 Clock Stretch

5.8.3.1 Clock Stretch Requirements

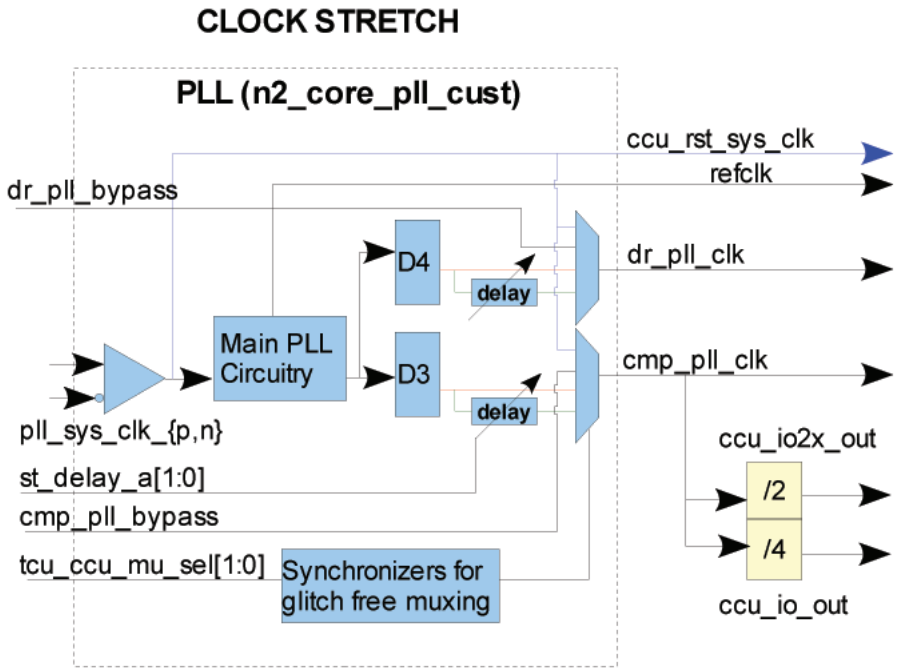
OpenSPARC T2 clock stretch operates within the following guidelines and requirements as defined by SPTE.

- Clock period stretch is needed only on the “cmp” domain.
- Frequency modulation due to stretch is restricted to one cycle.
- The amount of shift is in the interval (0, $T_{VCO}/2$)
- This actual shift is implemented using an RC delay line with reasonable granularity.
- Assertion of clock stretch is controlled by test registers programmed through the JTAG interface. These registers also control the amount of shift in RC delay line.
- There is no latency requirement, measured from the time clock stretch is asserted to the time clock shift occurs.
- Core clocks can then be stopped and state element values can be shifted out via scan chains.

5.8.3.2 PLL Support for Pulse Stretching

Clock stretching capability is built into the PLL because of the analog RC delay line. The mechanism for shifting the positive or negative edge is simple. The VCO output is muxed with a delayed version of itself as shown in the simplified FIGURE 5-19.

FIGURE 5-19 Clock Stretching Capability in PLL



Ports on the CCU that are relevant to clock stretch are tcu_ccu_mux_sel[1:0]. It determines which leg of the mux in the PLL is selected for output. 2'b11will select the leg for clock stretch.

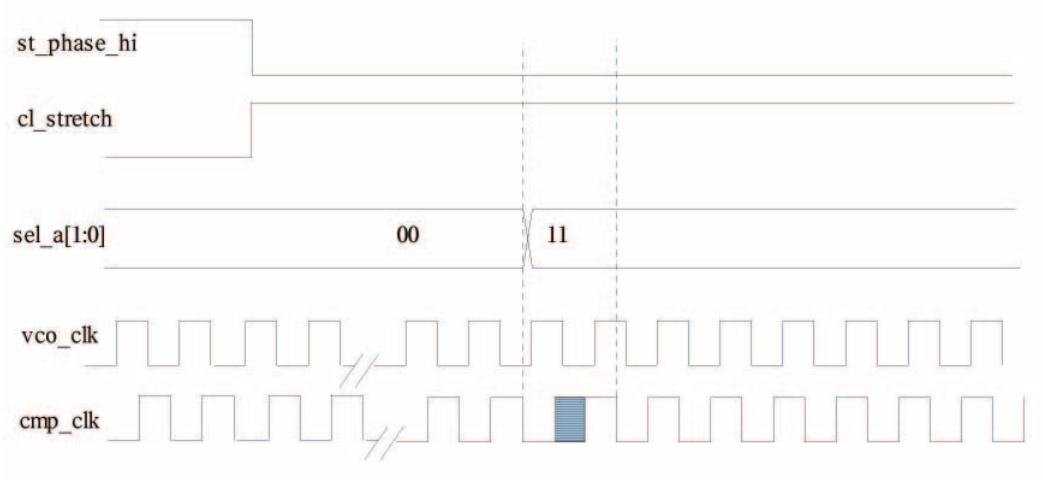
The amount of pulse shift (st_delay_a), and the phase to be stretched (st_phase_hi), are programmed in the CSR prior to the actual clock stretch event. The other two inputs to the CCU activate the stretch mux.

These signals can be asynchronous to the cmp_clk domain. They are synchronized appropriately in the PLL, depending on whether the high phase or the low phase is stretched. Stretching on the low phase (shifting the positive clock edge) requires synchronizing the mux selects to the negative edge of VCO clock. Conversely, stretching on the high phase requires changing the mux selects on the positive edge of VCO clock. This is illustrated in the next section.

5.8.3.3 Timing Diagram

FIGURE 5-20 illustrate the operation of the pulse stretching circuitry for shift during the low phase (i.e., rising edge).

FIGURE 5-20 Clock Stretch Timing Events



Note the synchronization of cl_stretch internally to generate the inputs to sel_a[1:0] on the falling edge of cmp_clk. This ensures the select lines to the mux change in the low phase.

The approach is similar for st_phase_hi = 1, the main difference being that sel_a[1:0] inputs are generated on the rising edge of cmp_clk.

5.8.3.4 Programmability

In PLL_REG, five bits can be programmed for the following clock stretch fields.

TABLE 5-11 Clock Stretch Fields in CSR Block

CSR Field	Bits	Description
st_phase_hi	25	Stretches high phase if true, else stretches low phase
st_delay_cmp	27:26	cmp clock stretch delay settings [00, 01, 10, 11] => [40, 80, 120, 160] ps under nominal PVT
st_delay_dr	35:34	dr clock stretch delay settings [00, 01, 10, 11] => [40, 80, 120, 160] ps under nominal PVT

5.8.4 SERDES Deterministic Test Mode (DTM)

5.8.4.1 Basic Requirements

DTM is a strategy for running tests for SERDES in a repeatable, deterministic manner. It allows testers to sweep the SPARC Core clock frequencies without breaking PLL lock, and perform traditional functional testing using the serial link interface.

In a nutshell, the tester goes through an initialization process to calibrate the RX lanes, and place data such that the outcomes on the blunt side are known and controllable. However, the TX data cannot be observed deterministically, so a workaround is to observe this TX data via the debug interface. The basic requirements are:

- All reference clocks to PLL inputs should come from the same source
- This applies to the core PLL, PSR, and FSR. ESR is excluded from DTM testing
- Convert clock domains from mission mode as follows:
  - IO -> DR
  - PC -> DR
  - CMP and DR domains unchanged
- Sync pulses between IO <-> CMP now are equivalent to DR <-> CMP
- Clock rates changed as follows
  - ref1 = ref2 = ~75-100 MHz
  - cmp = ~600-1500 MHz
  - dr = io = pc = ~75-100 MHz
  - cmp:dr ratio = 1:8, 1:11 or 1:15

5.8.4.2 Supported Clock Frequencies

The ideal scenario is to be able to perform a schmoo of OpenSPARC T2 across the entire operating range of core frequencies, i.e., from 600 Mhz – 1.5 Ghz. However, because of PLL characteristics, no single divider setting will allow this, and a minimum of three *gear ratios* is needed.

A gear ratio corresponds to the CCU core divider configuration (in this scheme, affected only by one divider, D2). Link rate for serial links indicates the data transfer rate which may be equal to or fractional multiples of internal clock speeds.

5.8.4.3 Clocking Scheme

FIGURE 5-21 CCU PLL Configuration for DTM

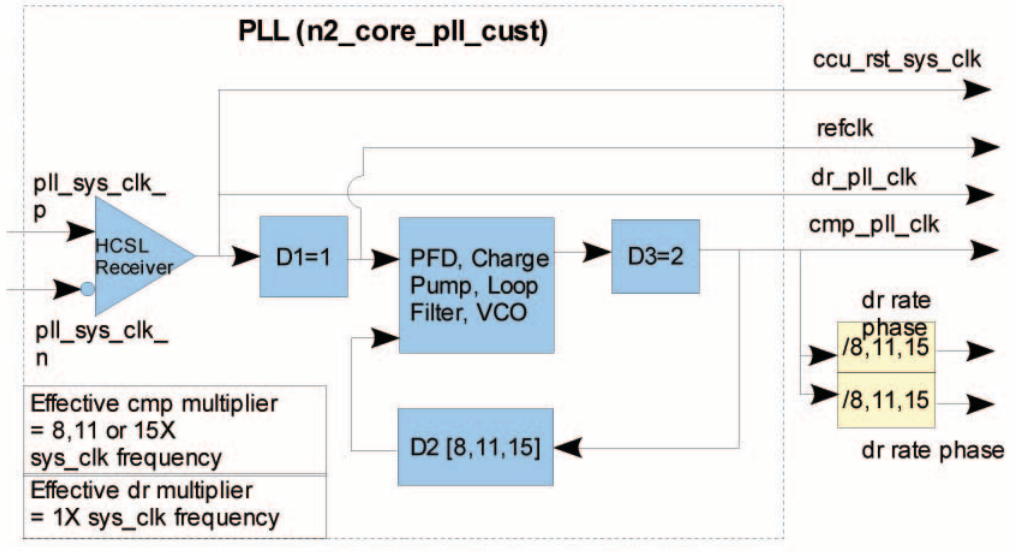


FIGURE 5-21 shows how the CCU PLL would be configured for DTM. Note that the muxes are not shown. They will configured such that the mux for the cmp clock output would be in functional mode whereas the mux for the dr clock would be in bypass mode.

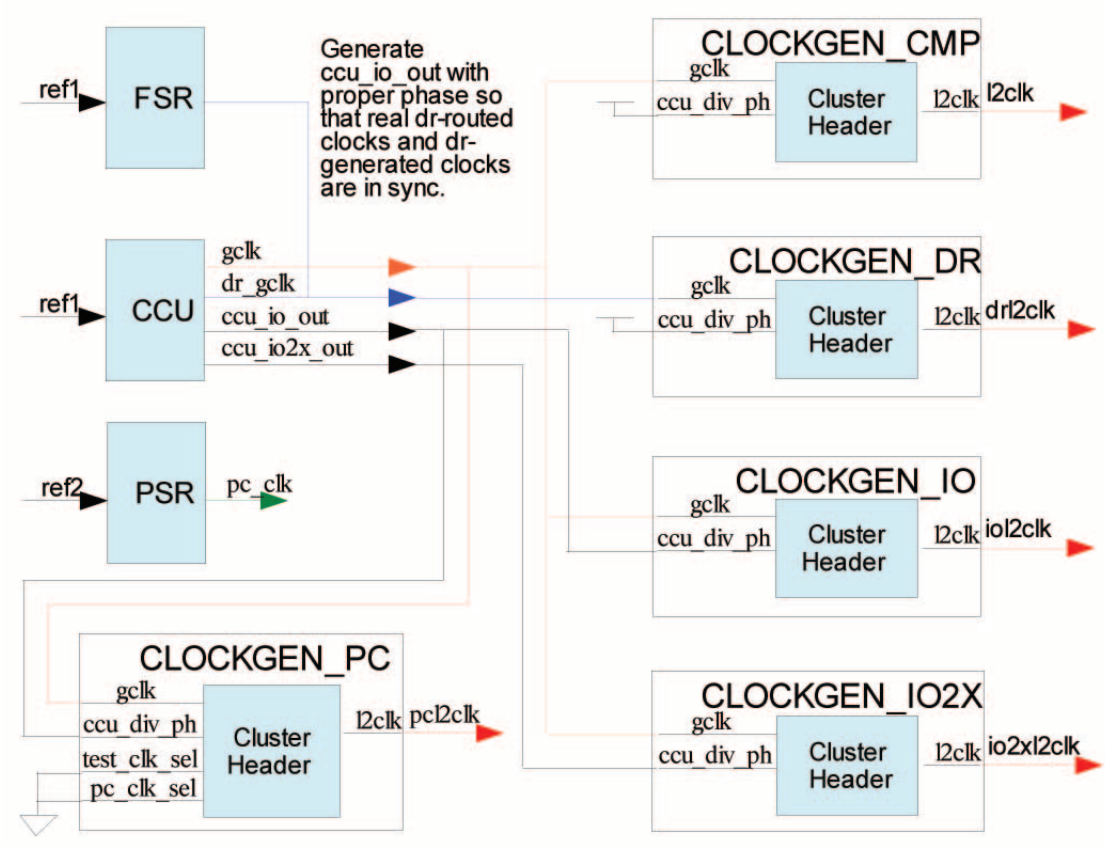
As shown in FIGURE 5-21 and FIGURE 5-22, clusters that will get IO, IO2X, and PC clocks will operate at DR rate by dividing down from the cmp clock. This is different from the actual DR clock from the distribution tree. The former is dubbed virtual dr clock as opposed to the real dr. The ramification of this approach is simplification of cluster header muxing controls, easier gclk distribution, and more robust timing since there are no new timing paths for analysis. However, the real dr clock will have a 50-50 duty cycle, while the virtual dr clock will have a duty cycle of 50/50, 55/45 and 53/47 respectively for D2=8, 11, and 15. This is not expected to be an issue.

Even though PC, IO, IO2X and DR clocks operate at the same frequency albeit with perhaps different duty cycles during DTM, direct data crossing between these clocks needs to be handled with care due to high possibility of hold-time violations. In normal mode, there is no direct communication between the four domains, or is handled via asynchronous fifos, so these paths are false. The min-time issues

encountered in DR<->IO crossings are addressed by using lock-up latches in the MCU, while for PC<->IO they are preemptively addressed by inverting the phase of PC clock (inside the cluster header) with respect to the IO clock in the PEU.

Sync pulse positions for domain transfers between CMP<->IO, CMP<->IO2X, and CMP<->DR are shown in FIGURE 5-23. They depend on the divider value D2. The sync pulse pairs, sys_cmp and cmp_sys are unchanged. However, all other sync pulses appear in different positions depending on the value of D2.

FIGURE 5-22 Chip Level DTM Clocking Scheme



5.8.4.4 Programmation and Sequencing

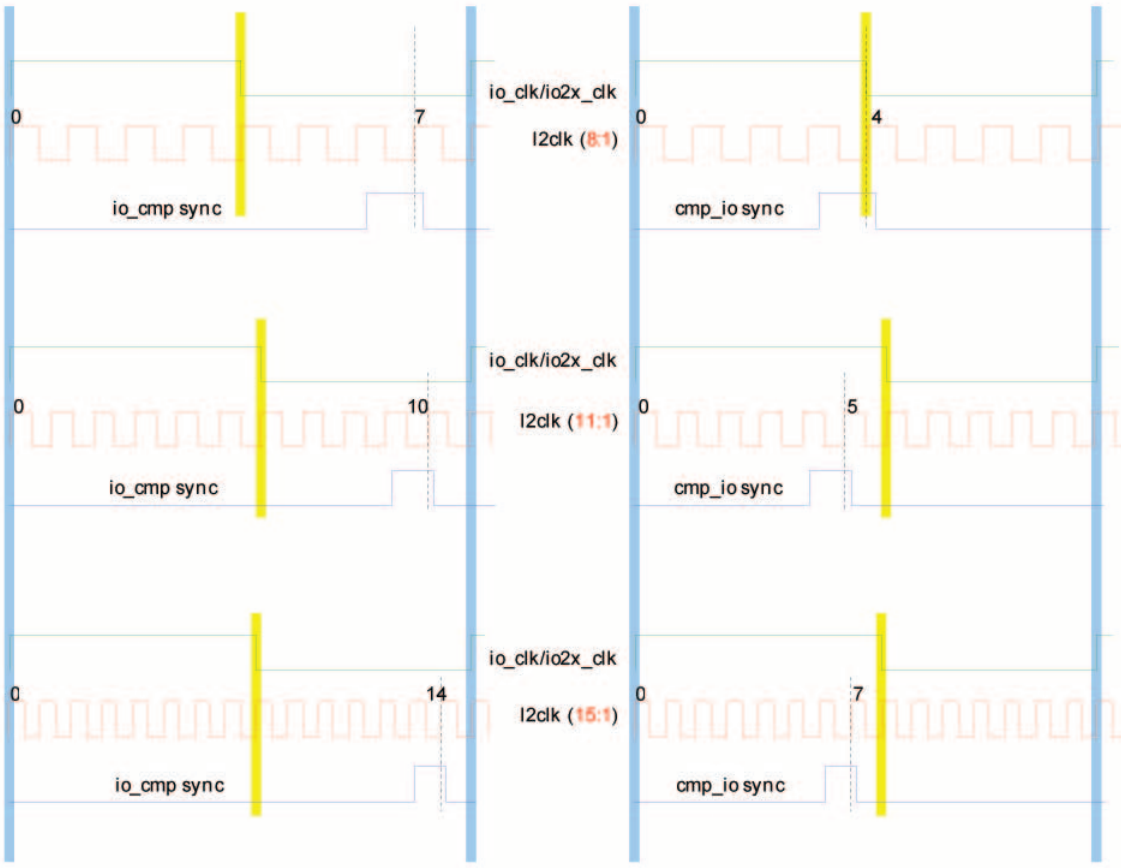
After power on reset, the pll_reg has to be programmed to set either of the DTM bits to '1', and set the PLL divider values to match one of the three acceptable gear ratios. The "CHANGE" field in PLL_REG should also be set to indicate frequency will change. (Depending on whether DTM1 or DTM2 is selected, DBG and MIO mux controls will be affected. However, from the CCU's perspective, it is only one mode. The only check CCU will perform that if both DTM1 and 2 will are asserted, mode 1 will be considered active).

The RST block will see the "ccu_rst_change" signal asserted and issue a warm reset.

The PLL will load the divider values upon de-assertion of reset and begin to lock for DTM to provide new clock frequencies (cmp and dr).

During reset active, the cluster header will see a change on the ccu_serdes_dtm signal which will be used to disable PC clock select. All other clockgen modules will require no connectivity change as a result of DTM; the process of generating a different clock frequency will be handled transparently by the CCU and cluster header. Everything will function normally here on.

FIGURE 5-23 New Sync Pulse Positions for DTM



All locations are at final destination after being flopped once in cluster

5.9 Appendix A.1 – Sync Pulse Design Procedure

This appendix focuses on the design methodology for sync pulses between CMP and DR domains which are non-integer multiples of one another. Consider the near ideal scenario for synchronizing between two such domains. We make the following assumptions:

- There is no jitter, skew, or PVT mismatches.
- At some point in time, both positive clock edges are perfectly aligned.
- There is zero phase offset between the PLL input reference and output.
- The setup and hold requirements on flip-flops are small.
- Propagation delay is experienced only by data (through wires, and clock-Q).

No matter which direction data is crossing domains, it makes sense to maximize the amount of time available between data launch and capture. This is illustrated in FIGURE 5-24 and FIGURE 5-25.

For cmp_clk to dr_clk transfers, the launch edge should be the first positive edge of cmp_clk after a dr_clk sampling point. The enable control generation would then occur a cmp cycle before launch.

On the other hand, for dr_clk to cmp_clk synchronization, the capture edge would be the last cmp_clk rising edge prior to a dr_clk sampling event. This time, the enable control would assert a cmp cycle before capture.

The cycle repeats when both the launch and capture clocks are perfectly aligned on the rising edge, or every *M* cycles of slow clock (equivalently *N* cycles of fast clock).

TABLE 5-12 Waveform Parameters for Ideal Case

Parameter	Description
N	Multiplication factor of fast clock
M	Multiplication factor of slow clock
T	Fast clock period (cmp_clk)
Tref	Reference clock period = *N.T*
Tslow	Slow clock period (dr_clk) = *M.T*
k	Cycle count of slow clock starting with 0

FIGURE 5-24 Synchronization from Fast to Slow Clock

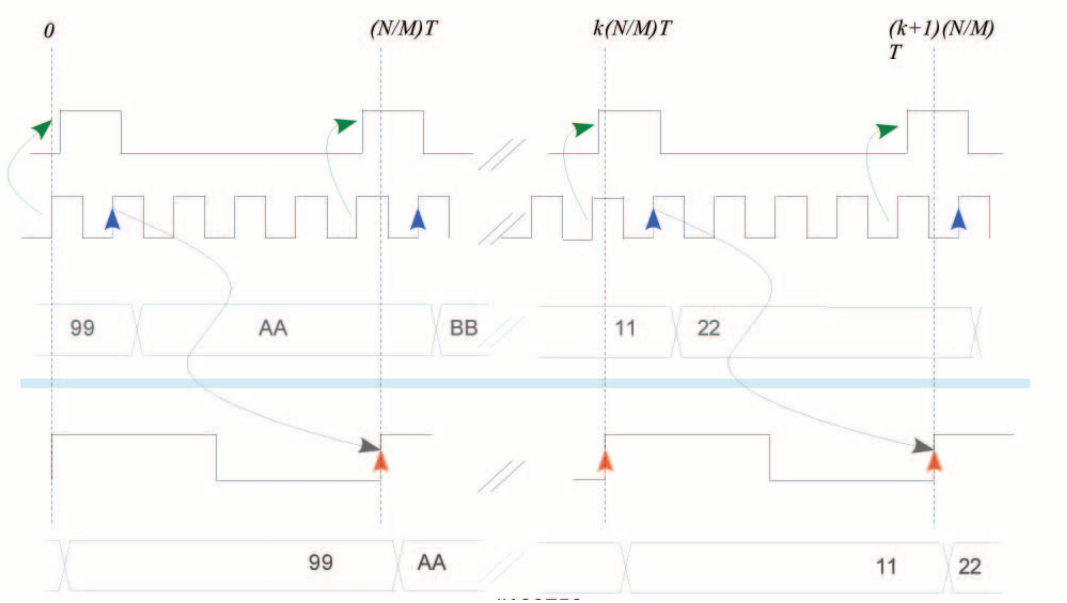
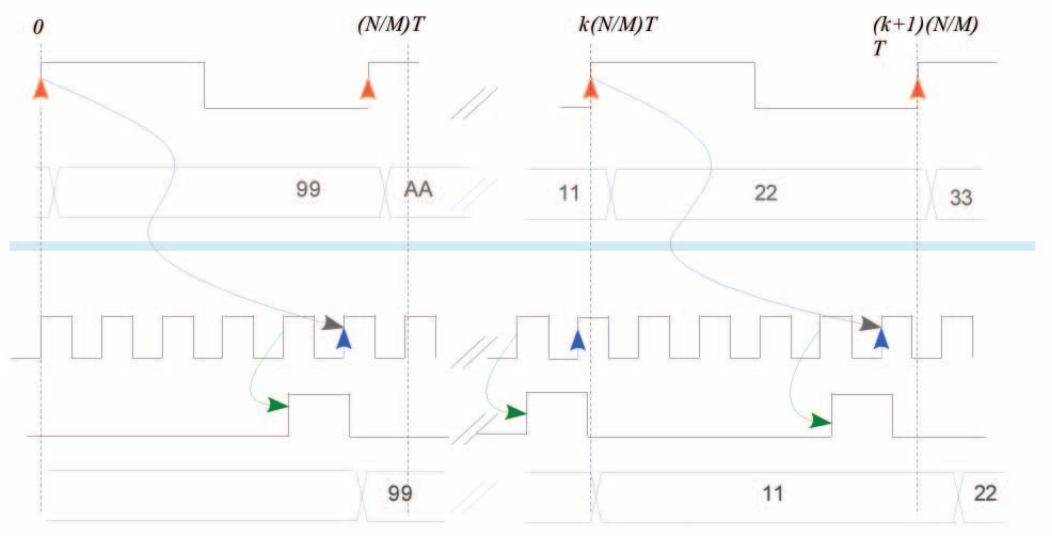


FIGURE 5-25 Synchronization from Slow to Fast Clock



5.10 Appendix A.2 – Sync Pulse Timing Analysis

Looking at the idealized timing diagrams and parameters in Appendix A.1, we can analyze the conditions one at a time for *min* and *max* conditions.

5.10.1 Fast to Slow Clock Synchronization

As per the proposal, the capture edge for cycle *k* is at *(k+1)(N/M)T.* Working our way backwards, the corresponding launch edge is the first edge on cmp_clk after *k(N/M)T*. This works out to be *FLOOR(k(N/M)+1)T.* The pulse would have to be generated a cycle before at *FLOOR(k(N/M)T.*

Amount of time available for setup,

tmax = (k+1)(N/M)T – FLOOR(k(N/M)+1)T

Similarly, the launch edge lags the last capture edge by

tmin = FLOOR(k(N/M)+1)T – k(N/M)T

There is a subtle difference between *FLOOR(k(N/M)+1)T* and *CEILING(k(N/M)T)* which shows up when k= 0.

5.10.2 Slow to Fast Clock Synchronization

The situation is reversed, where we need to launch data on the dr_clk as early as possible, ie, at *k(N/M)T*, and capture on cmp_clk as late as possible, at *FLOOR((k+1)(N/M)T).* Enable pulse generation thus occurs at *FLOOR((k+1)(N/M)-1)T.*

Therefore, setup margin is given by:

tmax = FLOOR((k+1)(N/M)T) – k(N/M)T

And the lag from last capture is

tmin = k(N/M)T – FLOOR(k(N/M))T

5.10.3 Modifications for Non-Ideal Scenario

This time, we revisit the approach while factoring in real conditions. Some other parameters that need to be considered are:

TABLE 5-13 Additional Parameters for Non-ideal Scenario

Parameter	Description
tsu	Setup time of capture flip-flop.
th	Hold time of launch flop.
tcq	Clock-to-Q time in flip flop.
tdata	Data delay from launch flop's Q output to capture flop D-pin
tskew	Skew and static phase offsets between slow and fast clocks.
tjitter	Jitter (cycle to cycle and long-term) between the clocks.

The constraints for max and min timing under non-ideal conditions for data launch and capture to work correctly are:

```
tmax >  tcq + tsu + tdata + tskew + tjitter

tmin > th + tskew + tjitter - tcq
```

Corresponding timing margins are given by:

```
tmargin,max = tmax - tdata - tcq - tsu - tskew - tjitter

tmargin,min = tmin - th - tskew - tjitter + tcq
```

Both sets of equations hold true, regardless of synchronization direction. Only the parameters *tmax* and *tmin* are derived differently as in the past section.

5.10.4 Computation and Selection of Sync Pulses

Now that a scheme has been proposed, and sync pulse generation formalized, here is the algorithm for the complete solution:

1. Compute which phase of slow clock the pulses should be generated (under ideal conditions) for fast to slow clock.
2. Find the corresponding timing margins available *tmax* and *tmin* (also ideal).
3. Estimate the amount of *skew, jitter, tdata, tcq, tsu* and *th*, and calculate tmargin,*max* and *tmargin,min*.

4. If *tmargin,max < 0 OR tmargin,min < 0* for any sync pulse, adjust phase for that pulse and repeat steps two through 4.
5. Repeat steps one through four for all ratios of *N/M.*
6. Repeat steps one through five for slow to fast clock.
7. Repeat steps one through six for all 3 refclk frequencies.

We are done when *tmargin,max > 0 AND tmargin,min > 0* for every ratio, otherwise for any particular ratio if either tmargin < 0, this scheme will not work.

CHAPTER 6

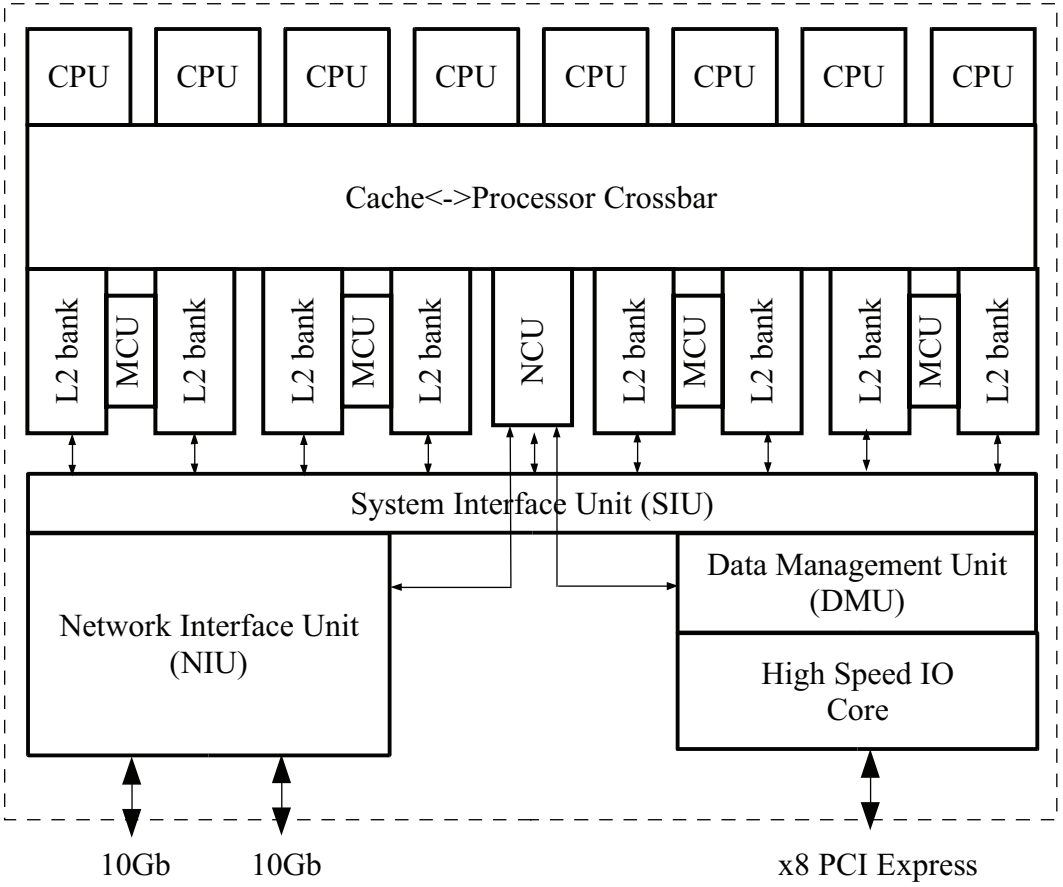
System Interface Unit (SIU)

This chapter contains the following sections:

6.1 Overview

OpenSPARC T2 has on chip multiple system I/O subsystems. OpenSPARC T2 integrates Fire's high speed IO core and connect directly to a x8 PCI Express channel (2GB/s/direction). OpenSPARC T2's integrated network I/O unit includes two 10Gb Ethernet MACs (2.5GB/s/direction). The System Interface Unit (SIU) provides 12GB/s of raw bandwidth per direction and has flexible interfaces for the Network Interface Unit (NIU) and Data Management Unit (DMU) to access memory via eight secondary level cache (L2) banks. SIU supports Fire's PCI Express. For the NIU, SIU was designed with the ability to allow write traffic to bypass other posted write traffic. SIU does not support coherency.

FIGURE 6-1 SIU Top Level Block Diagram



The SIU also provides a data return path for reads to the Peripheral I/O subsystems. The data for these PIO Reads and interrupt messages generated by the PCI Express subsystem are ordered in the SIU prior to delivery to the NonCacheable Unit (NCU).

6.2 Terminology

Cacheable: Can be stored in the L2 cache.

Cacheline: 64Byte

CSR: Configuration Status Register. A storage element for holding status or configuration information. They can exist either on OpenSPARC T2 or off OpenSPARC T2.

Core clock domain: reference to the speed of the SPARC processor core and L2 cache. Target is 1.4GHz

DMA: Direct Memory Access. A load or store originating from the IO subsystem that targets the memory subsystem.

Inbound: Logically toward the CPU and memory subsystem and NCU, away from the IO subsystems (NIU or DMU).

IO clock domain: references to the speed of the internal I/O interfaces units. Either 1/3 or core clock frequency. Also referred to as System clock domain

JBUS: Jalapeno bus. A coherency bus used in OpenSPARC T1 and other Sun processors.

Nonposted: A transaction in which the sender does want and require an acknowledge of delivery. A read is always nonposted.

Outbound: Logically toward the IO subsystems (NIU or DMU) and away from the CPU, L2 caches and the NCU

Packet: A structure for transferring information between interfaces. A Packet can consists of just a header or a header followed by a payload.

PIO: Peripheral Input Output. A load or store originating from the cpu that does not target the caches and memory. The target of a PIO can either be onchip or offchip.

Posted: A transaction in which the sender does not want nor require an acknowledgement of delivery

RDD: Read and Discard – All DMA accesses are noncoherent. Thus the data for a DMA read should be treated as use once and discard. All DMA reads are converted into RDDs by L2 caches and do not allocate in the L2 cache.

WR8: Write8 bytes – A WRM is decomposed into up to eight WR8 when the WRM is forwarded to an L2 bank. The L2 does a read modify write operation for each eight byte store. The eight byte enables for a WR8 can be randomly on or off.

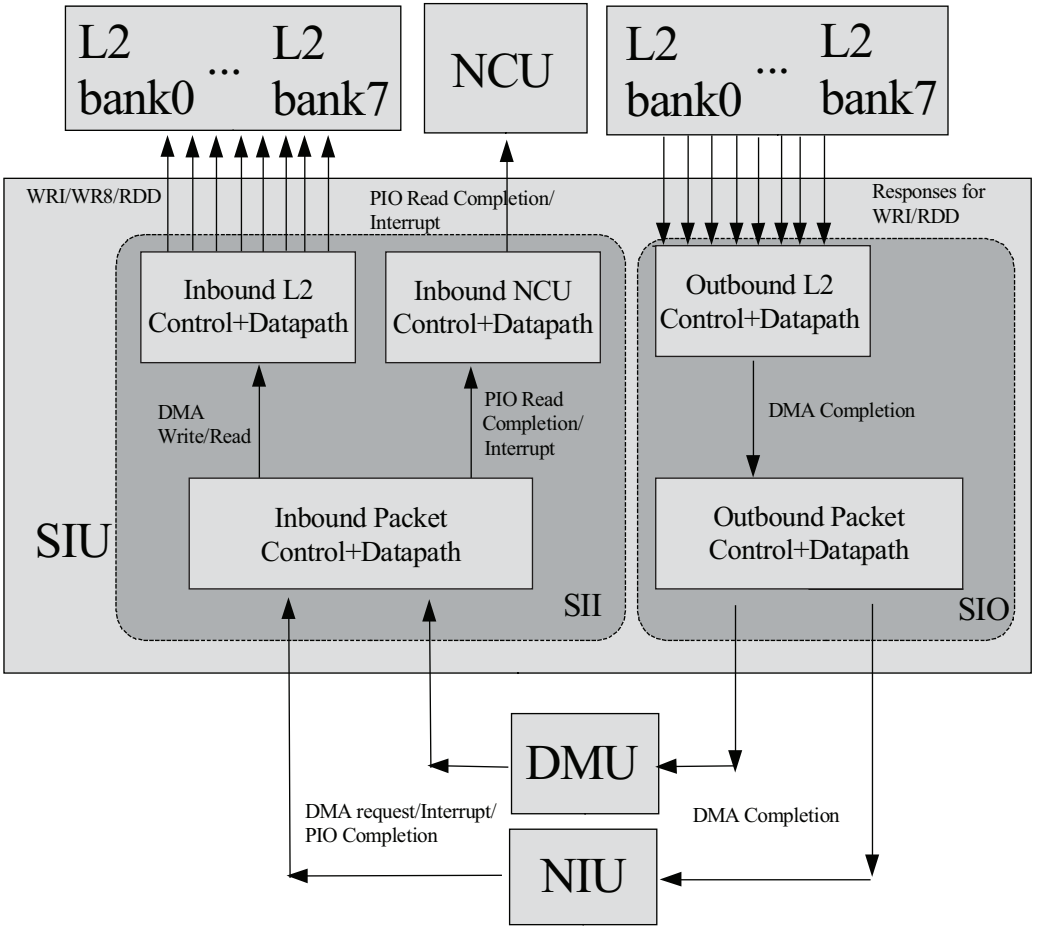
WRI: Write Invalidate – all DMA Writes that are aligned to a cacheline address boundary and writes 64 Byte of data will invalidate any matching address in the L2 cache tag array prior to data being forwarded to memory. (Terminology comes from the JBUS architecture)

WRM: Write Merge – a DMA Write with one or more bytes of the 64 byte payload not enabled. (Terminology comes from the JBUS architecture). A WRM may not cross cacheline boundary.

6.3 SIU Top Level Logical Block Diagram

The SIU is partitioned physically and logically into two parts based on flow direction – SIU Inbound (SII) for inbound traffic and SIU Outbound (SIO) for outbound traffic.

FIGURE 6-2 SIU Logical Block Diagram



All inbound traffic continues inbound through SIU until it reaches NCU or an L2 bank. All outbound traffic from NCU or L2 must leave SIU in the outbound direction. NCU and L2 banks cannot send traffic to each other through the SIU.

DMU and NIU cannot send traffic toward each other through the SIU. Because the L2 banks have their own paths through the memory controllers to memory, the SIU sees each L2 bank as a slave device. SIU assumes L2 never initiates requests to SIU. Likewise, Network blocks are always seen as master devices pulling from and pushing data to L2 only.

SIU does not support coherency.

All traffic uses a packet transfer interface. Each packet is one or two consecutive address/header cycles immediately followed by 0 or more consecutive data/payload cycles. SIU follows L2's addressing convention: big endian where the databytes for the lowest address are transferred first. Where applicable, byte enables are positional where byte_enable[0] always refer to databits[7:0] for all interfaces.

The interfaces between SIU and L2 are in the core clock domain - 1.5GHz. The interfaces between SIU and DMU, NIU, NCU are in the IO clock domain – 350 MHz or 1/4 core clock frequency.

TABLE 6-1 shows the packet types from DMU and NIU that are supported by SIU:

TABLE 6-1 Supported Packet Types from NIU and DMU

Source	Packet type	Posted	Queue in which packet may enter
NIU	RDD	Nonposted	Ordered
			Bypass
	WRI	Posted	Ordered
			Bypass
	WRI	Nonposted	Ordered
			Bypass

TABLE 6-1 Supported Packet Types from NIU and DMU *(Continued)*

Source	Packet type	Posted	Queue in which packet may enter
DMU	RDD	Nonposted	DMA/INT (Ordered)
	WRI	Posted	DMA/INT (Ordered)
	WRM	Posted	DMA/INT (Ordered)
	INT (Mondo)	Nonposted	DMA/INT (Ordered)
	PIO Rd Completions	Posted	
			PIO (Bypass)

TABLE 6-2 shows how the mapping between inbound DMA addresses from DMU and NIU to the L2 bank number. This is to support partial L2 banks when entire set(s) of L2 banks are disabled.

Five bits (PM, BA01, BA23, BA45, BA67) are used to indicate partial mode active, and which of the four pairs of banks are available. X is a don't care. When PM is on, it is illegal for only three of the BAs to be asserted and illegal if all four BAs are deasserted.

TABLE 6-2 Partial L2 Bank Mapping

PM	BA67	BA45	BA23	BA01		L2Bank[2]	L2Bank[1]	L2Bank[0]
0	X	X	X	X		PA[8]	PA[7]	PA[6]
1	0	0	0	0	Illegal	0	0	PA[6]
1	0	0	0	1		0	0	PA[6]
1	0	0	1	0		0	1	PA[6]
1	0	0	1	1		0	PA[7]	PA[6]
1	0	1	0	0		1	0	PA[6]
1	0	1	0	1		PA[7]	0	PA[6]
1	0	1	1	0		PA[7]	~PA[7]	PA[6]
1	0	1	1	1	Illegal	0	PA[7]	PA[6]
1	1	0	0	0		1	1	PA[6]

TABLE 6-2 Partial L2 Bank Mapping *(Continued)*

PM	BA67	BA45	BA23	BA01		L2Bank[2]	L2Bank[1]	L2Bank[0]
1	1	0	0	1		PA[7]	PA[7]	PA[6]
1	1	0	1	0		PA[7]	1	PA[6]
1	1	0	1	1	Illegal	0	PA[7]	PA[6]
1	1	1	0	0		1	PA[7]	PA[6]
1	1	1	0	1	Illegal	1	PA[7]	PA[6]
1	1	1	1	0	Illegal	1	PA[7]	PA[6]
1	1	1	1	1		PA[8]	PA[7]	PA[6]

SIU along with the gasket block inside each of the SPARC cores implement the same index hashing algorithm for the L2 cache. The purpose is to improve performance for certain software application – to reduce the thrashing of certain L2 indexes. Software can enable this feature by writing to a CSR in NCU.

If hashing is enabled and PA[39]==0, SIU converts the PA[39:0] to a different PA for L2:

L2_PA[39:18] = PA[39:18];

L2_PA[17:13] = PA[32:28] ^ PA[17:13];

L2_PA[12:11] = PA[19:18] ^ PA[12:11];

L2_PA[10:0] = PA[10:0];

6.4 Logical Subblocks

The SIU is partitioned physically and logically into two parts based on flow direction – SIU Inbound (SII) for inbound traffic and SIU Outbound (SIO) for outbound traffic. SIU is also partitioned into two clock domains.

A SIU subunit's name consists of three - five characters. All subunits are listed below:

- SII subunits:

  ILC0, ILC1, ILC2, ILC3, ILC4, ILC5, ILC6, ILC7,

  ILD0, ILD1, ILD2, ILD3, ILD4, ILD5, ILD6, ILD7,

  IPCC, IPCS0, IPCS1, IPD, INC, IND

- SIO subunits:

  OLC0, OLC1, OLC2, OLC3, OLC4, OLC5, OLC6, OLC7,

  OLD0, OLD1, OLD2, OLD3, OLD4, OLD5, OLD6, OLD7,

  OPCC, OPCS0, OPCS1, OPDS, OPDC

The first letter of a logical subunit's name indicates direction: **I**nbound or **O**utbound

The second letter of a logical subunit's name indicates either the destination for inbound data or the source object for outbound data:

- Second letter:

  L = L2 cache

  N = NCU

  P = Packets (DMU & NIU)

The third letter of a logical subunit's name indicates **C**ontrol path or **D**atapath

The optional last one or two character represents either an instance number (i.e. L2 bank number) or the subunit's clock domain (***C***ore or ***S***ystem).

6.4.1 Clocks

Target operating frequencies, CPU:IO = 1500MHz:350MHz

- Supported operating frequencies:

  4:1 CPU:IO synchronous clock ratio

  L2 and NCU @ 1500MHz,1400MHz,1300MHz, 1200MHz

  NIU and DMU @ 350MHz, 325MHz, 300MHz

6.4.2 Interface Datapath Access Mechanism

TABLE 6-3 Interface Datapath Access Mechanism

Datapath	Mechanism	Comment
SII to L2	Credits	SIU initially given: Two request credits, Four 64B-write-invalidate-data credits
From L2 to SIO	None	SII must not send request to L2 if SIO will not have space to receive response from L2
SII to NCU	Request/Grant Arbitration	Receiver of packets schedules resources and asserts Grant for one cycle to winning Requestor. Requestor must send packet the cycle after its Grant is asserted. Winning requestor may reassert Request for subsequent packet while delivering current packet. Requestor's Request must stay asserted until its Grant is received. SII can buffer 16 PIO Read data returns SII can buffer four interrupt mondo data or 4 x 16B
From DMU to SII, From NIU to SII	Credits	SII provides each IO subsystem two dedicated inbound packet queues. Each inbound queue can hold a maximum of 16 requests + 64 Byte data payload per request. SII notifies each IO subsystem when an request has been dequeued from either of the two dedicated inbound packet queue.
From SIO to DMU, From SIO to NIU	None	DMU or NIU must not send request to SII if it will not have space to receive response.
Internal SIU Datapaths	Request/Grant Arbitration	Receiver of packets schedules resources and asserts Grant for one cycle to winning Requestor. Requestor must send packet the cycle after its Grant is asserted. Winning requestor may reassert Request for subsequent packet while delivering current packet. Requestor's Request must stay asserted until its Grant is received.

6.4.3 Inbound

The parity protected interfaces between SIU and the DMU and NIU are 128 bit wide with side band signals for packet control. Having a 128 bit for header allows SIU to provide a rich set of transaction types and allows SIU to provide a uniform and

generic but flexible enough for most IO architectures.

The inbound packet interface protocol works as follows: (replace 'ext' with 'niu' or 'dmu')

Cycle 1: Header Cycle

- ext_sii_hdr_vld asserts for one cycle to indicate ext is sending the packet header.
- ext_sii_reqbypass indicates that ext is sending this packet to the SIU's 'bypass' inbound queue.
- ext_sii_datareq indicates this packet has a payload following the header cycle.
- ext_sii_datareq16 indicates this packet has only one cycle (16 Byte maximum) of payload. If datareq is asserted and datareq16 is deasserted, this packet has four cycles of payload for a maximum transfer size of 64Bytes.
- ext_sii_data[127:0] contains a valid header.

Cycle 2-5: Payload Cycle(s)

- ext_sii_hdr_vld is deasserted.
- ext_sii_data[127:0] contains the payload data
- ext_sii_parity[3:0] contains the parity for each 32 bit of data. Parity[N]= xor(data[32N+31: 32N])
- ext_sii_be[15:0] contains the byteenables for each byte of data if applicable. BE[N]==1 implies write data[8N + 7: 8N]

Each IO subsystem must keep track of number of available entries in the SIU Inbound queues and not overflow the SIU. Each time SIU Inbound forwards a request from its packet queue to its inbound L2 or NCU queue, SIU Inbound returns a credit back to the appropriate IO subsystem via sideband signals.

Each of the SIU Inbound queue allows for a maximum of 16 packets.

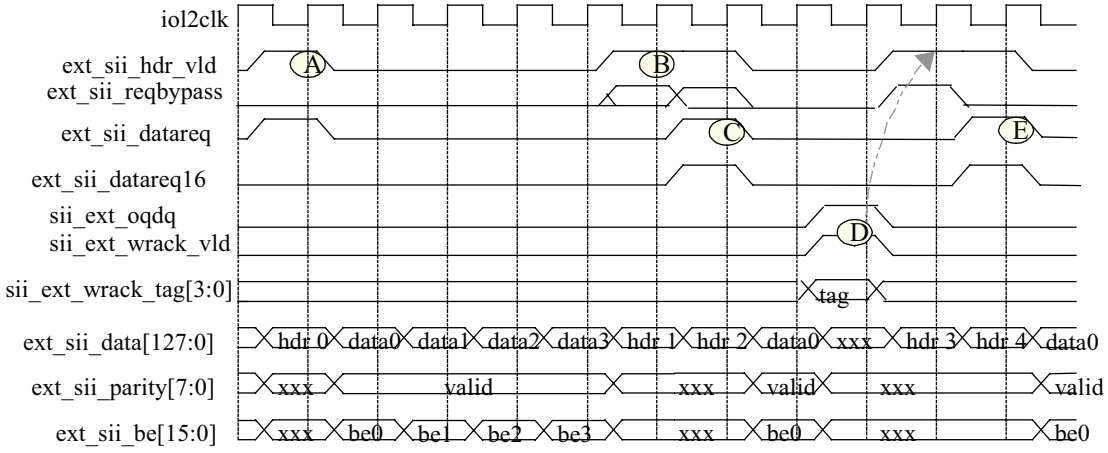
SIU supports back-to-back packet transfers with no dead cycle in between packets.

SIU architecturally supports PIO Read returns, DMA read requests, Interrupt Writes, DMA Write full cacheline (posted and nonposted), DMA write merge 64 bytes (posted only). The same logic is instantiated twice. One for each IO subsystem interface.

Another type of DMA access come from TCU/JTAG interface. The Read/Write access from the JTAG interface is eight bytes. The address need to be eight-bytes aligned (addr[2:0] = 3'b000). There is only one outstanding request allowed from the JTAG interface. The signals sending from JTAG interface should be running on cmp clock domain. When SII sending TCU read request, the adddr[2] need to be zero, so that L2 will return most critical eight-bytes first.

6.4.4 Interface Timing Diagrams and Protocols

FIGURE 6-3 Inbound Packet Interface Timing Diagram



A: *A 64-byte DMA write request with 4 cycles of data payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from NIU, ext may choose which queue in SIU is appropriate for the behavior wanted.
This example shows the DMA write request is for the Ordered Queue.

B: *A read request, no payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from NIU, ext may choose which queue in SIU is appropriate for the behavior wanted.
If this was the 16th outstanding credit, ext must stop issuing transactions to this Queue.

C: *A 16-byte PIO read data return; 1 cycle of data payload following the header.*
If from NIU, this must be destined for the Ordered Queue in SIU.
If from Fire-DMU, this must be destined for the Bypass Queue in SIU.
If this was the 16th outstanding credit, ext must stop issuing transactions to this Queue.

D: *SIU returns a credit after forwarding a DMA write data from the Ordered Queue to the Inbound L2$ Queue.*
If write request was from Fire-DMU, sii_dmu_wrack_vld asserts with tag information and the Fire-DMU can add this credit back to the credit list,
If write request was from NIU and the write request was in the Ordered Queue, sii_ext_oqdq asserts.
Ext may now resume sending transaction to the SIU – this example has a DMA read request following credit.

E: *(Only applicable for Fire-DMU) INT header plus 1 data beat of data payload, the SIU checks the header to distinquish PIO read completion from INT payload;*

6.4.4.1 From NIU to SIU

Single and back-to-back DMA read request from NIU to SIU

For each DMA read request, NIU must always guarantee it has buffer space to receive the DMA read response that will return from SIU outbound.

A DMA read does not allocate in the cache.

This describes the protocol for a single DMA read request from NIU to SIU.

1. NIU first checks that it has a credit available for the packet transfer.

   If NIU does not have credit for SIU's Ordered Queue and wishes to send a DMA read request to SIU's Ordered Queue, NIU must wait for sii_niu_oqdq to assert.

   If NIU does not have credit for SIU's Bypass Queue and wishes to send a DMA read request to SIU's Bypass Queue, NIU must wait for sii_niu_bqdq to assert.

   Once NIU has guaranteed that it will not overflow SIU, NIU can send the DMA read request packet on the interface.

2. Send packet. This transfer takes one IO clock cycle.

   a. NIU asserts header valid signal (niu_sii_hdr_vld) high,

   b. NIU drives all the header bits appropriately on niu_sii_data[127:0]

      Note that SIU does not require the byte address field in the header to be aligned to a cacheline boundary. Memory will always return the critical 32 bit word first and wrapped back to the beginning cacheline address boundary. Because the current software ethernet driver model has the ethernet transmit/control information structures in memory aligned to 64B address boundary, NIU sets DMA Read address[5:0] set to 0.

   c. NIU drives data request signals (niu_sii_datareq and niu_sii_datareq16) low.

   d. NIU drives the destination queue signal (niu_sii_reqbypass) to high for the bypass queue or low for the ordered queue.

   e. The parity lines (niu_sii_parity) are a don't care for the header cycle.

   f. SIU does not support byte enables for reads. If byteenable wires exist at the top level interface (niu_sii_be), then they are a don't care for the header cycle.

3. NIU must reduce the appropriate credit counter.

SIU supports back-to-back transfers from NIU. A second packet may be sent immediately the cycle after the first DMA Read packet if there is credit available for the second transfer.

Single and back-to-back DMA write request from NIU to SIU

For each DMA write request that NIU requires an acknowledgement of completion, NIU must always guarantee it has buffer space to receive the DMA write completion response that will return from SIU outbound. For N2's NIU, this is always true. A DMA write packet that needs an completion ack returned must be marked nonposted in the packet header.

A DMA Write does not allocate in the cache.

This describes the protocol for a single DMA write request from NIU to SIU.

1. NIU first checks that it has a credit available for the packet transfer.

   If NIU does not have credit for SIU's Ordered Queue and wishes to send a DMA write request to SIU's Ordered Queue, NIU must wait for sii_niu_oqdq to assert.

   If NIU does not have credit for SIU's Bypass Queue and wishes to send a DMA write request to SIU's Bypass Queue, NIU must wait for sii_niu_bqdq to assert.

   Once NIU has guaranteed that it will not overflow SIU, NIU can send the DMA write request packet on the interface.

2. Send packet. This transfer takes five IO clock cycle.

   a. On the first cycle,

      i. NIU asserts header valid signal (niu_sii_hdr_vld) high.

      ii. NIU drives all the header bits appropriately on niu_sii_data[127:0].

      Note that SIU does not require the byte address field in the header to be aligned to a cacheline boundary and allows for byte mask field in the header to be set for a WRM, because the current software ethernet driver model has the ethernet receive/control information structures in memory aligned to 64B address boundary, NIU sets DMA Write address[5:0] to 0, and command field to be write full 64 bytes, byte mask active to 0.

      iii. NIU drives data request signals (niu_sii_datareq high and niu_sii_datareq16 low).

      iv. NIU drives the destination queue signal (niu_sii_reqbypass) to high for the bypass queue or low for the ordered queue.

      Writes to the ordered queue will always be issued by the SIU to L2 after the youngest write in the bypass queue and after all prior writes to L2 has been sent from L2 to MCU. The writes in the bypass queues are not ordered with respect to other writes.

      v. The parity lines (niu_sii_parity) are don't cares for the header cycle.

      vi. If byte enables wires exist at the top level interface, the byteenable lines (niu_sii_be) are don't cares for the header cycle.

b. On the second, third, fourth, and fifth cycle,

i. NIU drives header valid signal (niu_sii_hdr_vld) low.

ii. NIU drives data payload on niu_sii_data[127:0] in big endian format.

iii. The data request signals (niu_sii_datareq and niu_sii_datareq16) are don't cares for nonheader cycles.

iv. The destination queue signal (niu_sii_reqbypass) is a don't care for nonheader cycle.

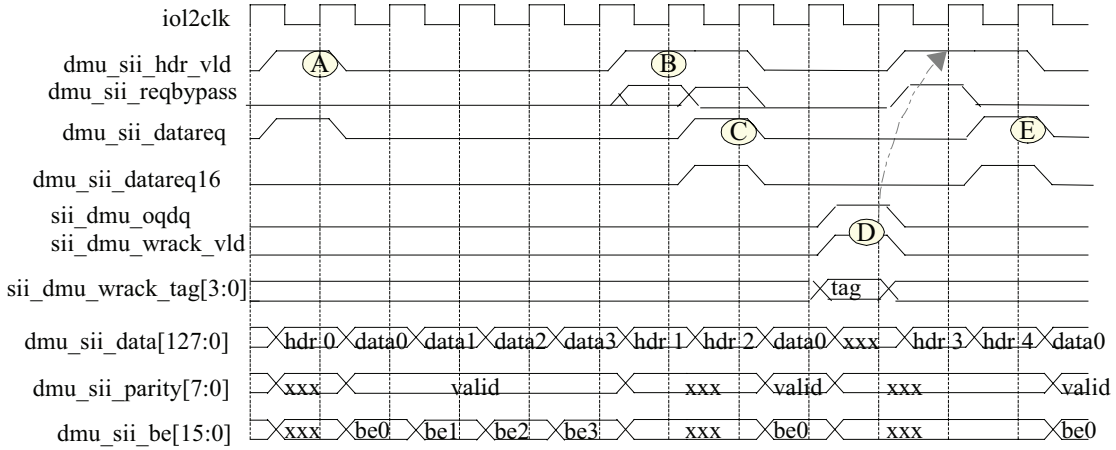
v. The parity lines (niu_sii_parity[3:0]) are driven.

vi. If byte enables wires exist at the top level interface, the byteenable lines (niu_sii_be[15:0]) should be all 1's to be safe but are treated as don't cares if during the header cycle, the byte mask active field was 0.

3. NIU must reduce the appropriate credit counter.

SIU supports back-to-back transfers from NIU. A second packet may be sent immediately the cycle after the first DMA Write packet if there is credit available for the second transfer.

FIGURE 6-4 Timing Diagram for SIU Inbound Packet from DMU



***A**: A 64-byte DMA write request with 4 cycles of data payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from HT-DMU, dmu may choose which queue in SIU is appropriate for the behavior wanted.
This example shows the DMA write request is for the Ordered Queue.

***B**: A read request, no payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from HT-DMU, dmu may choose which queue in SIU is appropriate for the behavior wanted.
If this was the 16th outstanding credit, dmu must stop issuing transactions to this Queue.

***C**: A 16-byte PIO read data return; 1 cycle of data payload following the header.*
If from HT-DMU, this must be destined for the Ordered Queue in SIU.
If from Fire-DMU, this must be destined for the Bypass Queue in SIU.
If this was the 16th outstanding credit, dmu must stop issuing transactions to this Queue.

***D**: SIU returns a credit after forwarding a DMA write data from the Ordered Queue to the Inbound L2$ Queue.*
If write request was from Fire-DMU, sii_dmu_wrack_vld asserts with tag information and the Fire-DMU can add this credit back to the credit list,
If write request was from HT-DMU and the write request was in the Ordered Queue, sii_ext_oqdq asserts.
DMU may now resume sending transaction to the SIU – this example has a DMA read request following credit.

***E:** INT header plus 1 data beat of data payload, the SIU checks the header to distinquish PIO read completion from INT payload;*

6.4.4.2 From a Fire-PCI Express-DMU to SIU

For SIU to support Fire's version of a DMU that connects to PCI Express, SIU Inbound must adapt to a stricter form of credit management where not only is a credit returned when SII dequeues a packet, but the corresponding id for the packet that was dequeued.

With respect to packet types, Fire does not support nonposted DMA writes and does not allow DMA Writes to pass other writes so all DMAs would effectively go into the SIU's Ordered Queue. The packet format differences between a prior DMU<->SIU and Fire-DMU<->JBC is handled in a thin layer within the new DMU called the DSN.

Fire's PCI-Express DMU requires SIU to be able to accept without flow control all completions for all PIO reads that had originated from NCU.

SIU's Inbound architecture has a 16 deep FIFO for its Ordered Queue and a 16 deep FIFO for its Bypass Queue. Fire's PCI-Express DMU supports 16 'credits' of DMAs+Interrupts and 16 credits of PIOs.

A DMA write credit id may be reused once the write has been posted (dequeued) from SIU's Inbound Packet Ordered Queue. A DMA read credit id may only be reused after the DMA read data response has returned from SIU's Outbound to DMU. Interrupt (Mondo type only) credit id may only be reused after NCU has acked or nacked the Interrupt. NCU must adapt to Fire's PCI-Express DMU and manage 16 PIO credits.

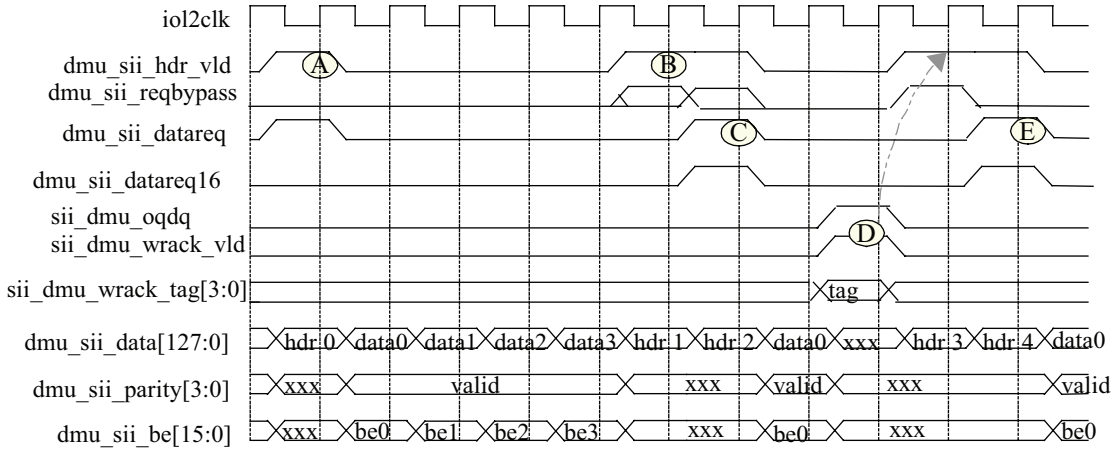
Without a design change, SIU's ordered Queue can support 16 outstanding DMA+Interrupt. But expanding that to an 32 deep ordered queue and gutting the design of the bypass queue to accommodate the 16 PIO completions would significantly impact schedule and not allow for code reuse. The estimated net area savings from gutting the design and extending the queue depth was determined to be small (at most couple hundred square microns in Epic9).

The solution proposed and implemented is conditioned on the fact that SIU already has dependency pointers for each of the 16 entries in the Inbound Packet Bypass Queue and dependency pointers for each of the 16 entries in the Inbound Packet Ordered Queue. An internal mode wire is added to indicates the interface is connected to PCI-Express DMU instead of NIU. Effectively, the 'Bypass Queue' becomes an ordered PIO completion queue. When in PCIExpress mode, rather than using the results of address cams to set up dependencies for a new packet entering the Bypass Queue, SIU forces that new packet to wait for the youngest packet existing in the Ordered Queue. Note that new packets entering the Ordered Queue by default (even for NIU) depends on the youngest packet existing in the Bypass Queue. Forcing this dependency when in PCIExpress mode converts the Bypass Queue into another ordered queue IF ALL the packets in the Bypass Queues drains to the same place. If there are multiple DMA Writes in the bypass queue, the existing ordering mechanism used for NIU would not guarantee a younger DMA Write from the bypass queue entry would complete later than an older DMA Write from the bypass queue that targets a different L2 bank.

Therefore, ALL DMA Writes (and interrupts) from Fire's PCI-Express DMU must be steered into the Ordered Queue. Likewise, because there is only one drain from SIU for all PIO completions (one FIFO path from SIU to NCU), when ALL PIO Completions from Fire's PCI-Express DMU are steered in the Bypass Queue, the

ordering requirement that PIO completions must pull all prior DMAs writes and all prior PIO completions is satisfied and SIU now achieved a 16 entry PIO completion queue at the cost of a few muxes added the original design. If physical areas become more critical, the cam logic remaining may be optimized out during physical implementation.

FIGURE 6-5 Timing Diagram for SIU Inbound Packet from DMU



***A**: A 64-byte DMA write request with 4 cycles of data payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from HT-DMU, dmu may choose which queue in SIU is appropriate for the behavior wanted.
This example shows the DMA write request is for the Ordered Queue.

***B**: A read request, no payload.*
If from Fire-DMU, this must be destined for the Ordered Queue in SIU.
If from HT-DMU, dmu may choose which queue in SIU is appropriate for the behavior wanted.
If this was the 16th outstanding credit, dmu must stop issuing transactions to this Queue.

***C**: A 16-byte PIO read data return; 1 cycle of data payload following the header.*
If from HT-DMU, this must be destined for the Ordered Queue in SIU.
If from Fire-DMU, this must be destined for the Bypass Queue in SIU.
If this was the 16th outstanding credit, dmu must stop issuing transactions to this Queue.

***D**: SIU returns a credit after forwarding a DMA write data from the Ordered Queue to the Inbound L2$ Queue.*
If write request was from Fire-DMU, sii_dmu_wrack_vld asserts with tag information and the Fire-DMU can add this credit back to the credit list,
If write request was from HT-DMU and the write request was in the Ordered Queue, sii_ext_oqdq asserts.
DMU may now resume sending transaction to the SIU – this example has a DMA read request following credit.

***E:** INT header plus 1 data beat of data payload, the SIU checks the header to distinquish PIO read completion from INT payload;*

Single and Back-to-Back DMA Read Request from Fire-DMU to SIU

For each DMA read request, DMU must always guarantee it has buffer space to receive the DMA read response that will return from SIU outbound. For Fire-DMU, this is always true.

A DMA read does not allocate in the cache.

This describes the protocol for a single DMA read request from Fire-DMU to SIU.

1. Fire-DMU first checks that it has a credit available for the packet transfer.
   a. If Fire-DMU does not have credit for SIU's Ordered Queue and wishes to send a DMA read request to SIU, Fire-DMU must wait for sii_dmu_wrack_vld to assert, a prior DMA Read to complete from SIU Outbound or an interrupt credit to return from NCU.
   b. Once Fire-DMU has guaranteed that it has a DMA credit, Fire-DMU can send the DMA read request packet on the interface.
2. Send packet. This transfer takes one IO clock cycle.
   a. Fire-DMU asserts header valid signal (dmu_sii_hdr_vld) high.
   b. Fire-DMU drives all the header bits appropriately on dmu_sii_data[127:0]
      i. Fire-DMU always generate a 64 Byte aligned address and never cross a cacheline boundary. Fire-DMU sets DMA Read address[5:0] set to 0.
   c. Fire-DMU drives data request signals (dmu_sii_datareq and dmu_sii_datareq16) low.
   d. Fire-DMU drives the destination queue signal (dmu_sii_reqbypass) to low for the ordered queue.
   e. The parity lines (dmu_sii_parity) are a don't care for the header cycle.
   f. The byteenable lines (dmu_sii_be) are a don't care for the header cycle.
3. Fire-DMU must reduce the DMA credit counter.

SIU supports back-to-back transfers from Fire-DMU. A second packet may be sent immediately the cycle after the first DMA Read packet if there is credit available for the second transfer.

Single and Back-to-Back DMA Write Request from Fire-DMU to SIU

For Fire-DMU, all DMA Writes are posted and address aligned to cacheline boundary although byte(s) may be deasserted at the beginning or the end. this is always true.

A DMA Write does not allocate in the cache.

This describes the protocol for a single DMA write request from Fire-DMU to SIU.

1. Fire-DMU first checks that it has a credit available for the packet transfer.

a. If Fire-DMU does not have credit for SIU's Ordered Queue and wishes to send a DMA write request to SIU, Fire-DMU must wait for sii_dmu_wrack_vld to assert, a prior DMA Read to complete from SIU Outbound or an interrupt credit to return from NCU.

b. Once Fire-DMU has guaranteed that it has a write credit, Fire-DMU can send the DMA write request packet on the interface.

2. Send packet. This transfer takes five IO clock cycle.

   a. On the first cycle,

      i. Fire-DMU asserts header valid signal (dmu_sii_hdr_vld) high.

      ii. Fire-DMU drives all the header bits appropriately on dmu_sii_data[127:0].

      For Fire-DMU, DMA Writes with or without bytemask active has address aligned to 64-Byte boundary. Fire-DMU set DMA Write address[5:0] to 0. Fire-DMU does not support nonposted writes.

      iii. Fire-DMU drives data request signals (dmu_sii_datareq high and dmu_sii_datareq16 low).

      iv. Fire-DMU drives the destination queue signal (dmu_sii_reqbypass) to low for the ordered queue.

      Writes to the ordered queue will always be issued by the SIU to L2 after the youngest PIO completion in the bypass queue and after all prior writes to L2 has been sent from L2 to MCU.

      v. The parity lines (dmu_sii_parity) are don't cares for the header cycle.

      vi. The byteenable lines (dmu_sii_be) are don't cares for the header cycle.

   b. On the second, third, fourth, and fifth cycle,

      i. Fire-DMU drives header valid signal (dmu_sii_hdr_vld) low.

      ii. Fire-DMU drives data payload on dmu_sii_data[127:0] in big endian format.

      iii. The data request signals (dmu_sii_datareq and dmu_sii_datareq16) are don't cares for nonheader cycles.

      iv. The destination queue signal (dmu_sii_reqbypass) is a don't care for nonheader cycle.

      v. The parity lines (dmu_sii_parity[3:0]) are driven.

      vi. The byteenable lines (dmu_sii_be[15:0]) should be driven. They are treated as don't cares if during the header cycle, the byte mask active field was 0.

3. Fire-DMU must reduce the DMA credit counter.

SIU supports back-to-back transfers from Fire-DMU. A second packet may be sent immediately the cycle after the first DMA Write packet if there is credit available for the second transfer.

Single and Back-to-Back Interrupt Request from Fire-DMU to SIU

PCI-Express requires that an interrupt it send to the CPU via whatever path is delivered after the youngest corresponding DMA write has been sent to memory (interrupt from Fire-DMU must guarantee that prior writes sent to SIU has left L2). Fire-DMU take advantage of the ordering maintained by SIU by simply send an interrupt to NCU via SIU's ordered queue.

Because these interrupt types are mondo, Fire-DMU must be capable of retrying NAcked mondo and Fire-DMU must have an interrupt response path from NCU.

This describes the protocol for a single interrupt from Fire-DMU to SIU.

1. Fire-DMU first checks that it has a credit available for the packet transfer.
   a. If Fire-DMU does not have credit for SIU's Ordered Queue and wishes to send a DMA write request to SIU, Fire-DMU must wait for sii_dmu_wrack_vld to assert, a prior DMA Read to complete from SIU Outbound or an interrupt credit to return from NCU.
   b. Once Fire-DMU has guaranteed that it has a DMA-INT credit, Fire-DMU can send the interrupt request packet on the interface.
2. Send packet. This transfer takes two IO clock cycle.
   a. On the first cycle,
      i. Fire-DMU asserts header valid signal (dmu_sii_hdr_vld) high.
      ii. Fire-DMU drives all the header bits appropriately on dmu_sii_data[127:0].
      iii. Fire-DMU drives data request signals (dmu_sii_datareq high and dmu_sii_datareq16 high).
      iv. Fire-DMU drives the destination queue signal (dmu_sii_reqbypass) to low for the ordered queue.

          Interrupt in the ordered queue will always be issued by the SIU to NCU after the youngest PIO completion in the bypass queue and after all prior writes to L2 has been sent from L2 to MCU.
      v. The parity lines (dmu_sii_parity) are don't cares for the header cycle.
      vi. The byteenable lines (dmu_sii_be) are don't cares for the header cycle.
   b. On the second cycle,

i. Fire-DMU drives header valid signal (dmu_sii_hdr_vld) low.

ii. Fire-DMU drives mondo data payload on dmu_sii_data[127:0].

This is the first 16Bytes of the mondo data.

iii. The data request signals (dmu_sii_datareq and dmu_sii_datareq16) are don't cares for nonheader cycles.

iv. The destination queue signal (dmu_sii_reqbypass) is a don't care for nonheader cycle.

v. The parity lines (dmu_sii_parity[3:0]) are driven.

vi. The byteenable lines (dmu_sii_be[15:0]) should be all 1's to be safe.

3. Fire-DMU must reduce the DMA-INT credit counter.

SIU supports back-to-back transfers from Fire-DMU. A second packet may be sent immediately the cycle after the first Interrupt packet if there is credit available for the second transfer.

Single and Back-to-Back PIO Read Data Return from Fire-DMU to SIU

PCI-Express requires that an interrupt it send to the CPU via whatever path is delivered after the youngest corresponding DMA write has been sent to memory (interrupt from Fire-DMU must guarantee that prior writes sent to SIU has left L2). Fire-DMU take advantage of the ordering maintained by SIU by when in PCI-Express mode by simply sending the PIO Completion to NCU via SIU's Bypass Queue.

This describes the protocol for a single PIO Read data return from Fire-DMU to SIU.

1. Send packet. This transfer takes two IO clock cycle. Because NCU guarantees that NCU will stop sending PIO request if all 16 PIO credits are used, Fire-DMU does not need to check for credit available prior to PIO read completion transfers.

   a. On the first cycle,

      i. Fire-DMU asserts header valid signal (dmu_sii_hdr_vld) high.

      ii. Fire-DMU drives all the header bits appropriately on dmu_sii_data[127:0].

      iii. Fire-DMU drives data request signals (dmu_sii_datareq high and dmu_sii_datareq16 high).

      iv. Fire-DMU drives the destination queue signal (dmu_sii_reqbypass) to high for the bypass queue.

A PIO completion in the bypass queue will always be issued by the SIU to NCU after the youngest PIO completion in the bypass queue and after all prior DMA writes to L2 has been sent from L2 to MCU and all prior interrupts are on its way to NCU.

v. The parity lines (dmu_sii_parity) are don't cares for the header cycle.

vi. The byteenable lines (dmu_sii_be) are don't cares for the header cycle.

2. On the second cycle,

a. Fire-DMU drives header valid signal (dmu_sii_hdr_vld) low.

b. Fire-DMU drives data payload on dmu_sii_data[127:0] in big endian format.

i. The CPU knows how many bytes it wanted.

ii. Note that if this PIO read is for a CSR, Fire-DMU must replicate the CSR's 64 bit value on the 128 bit payload. The CPU expects the lower 64 data bits to be the same as the upper 64 bits.

iii. The data request signals (dmu_sii_datareq and dmu_sii_datareq16) are don't cares for nonheader cycles.

iv. The destination queue signal (dmu_sii_reqbypass) is a don't care for nonheader cycle.

v. The parity lines (dmu_sii_parity[3:0]) are driven.

vi. The byteenable lines (dmu_sii_be[15:0]) should be all 1's.

SIU supports back-to-back transfers from Fire-DMU. A second packet may be sent immediately the cycle after the first PIO read completion packet if there is credit available for the second transfer.

6.4.4.3 From SIU to L2

Back to Back Read Requests to L2

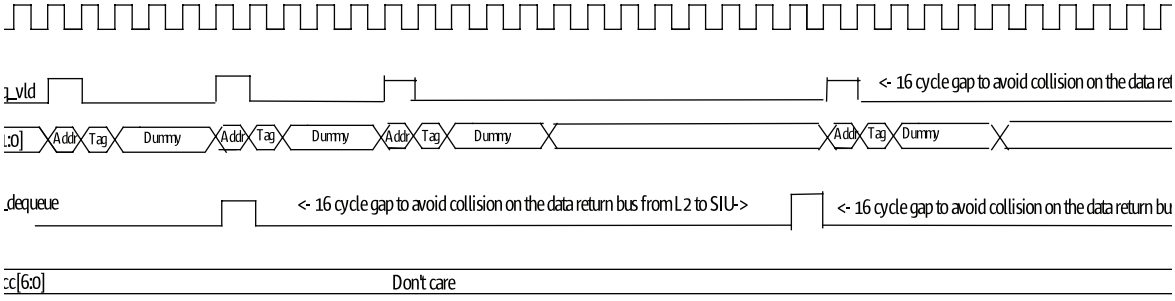
A read request packet (on sii_l2t_req[31:0]) consists of two header cycles followed by three dummy cycles. The first two dummy cycles are required by L2 for pipeline alignment - so the initial pipeline stages of a read request looks like the stages of a write 8 byte request. The last dummy cycle is for turnaround required by L2. The signal sii_l2t_req_vld asserts for one cycle to indicate the first cycle of the packet transfer. SIU only has two request tokens so SIU can burst only up to two back to back read requests to the L2 Tag. After two outstanding requests without an indication of a dequeue from L2 (l2t_sii_iq_dequeue asserting high for one cycle), SIU must wait for an entry in the input queue in L2Tag to drain. FIGURE 6-6 shows

the best case of back to back read requests. SIU bursts two reads, then sees l2t_sii_iq_dequeue asserting during the second transfer and proceeds to send out a 3rd read request. After which point, resource constraints prevent further back to back requests. According to the L2 pipeline, the earliest l2t_sii_iq_dequeue asserts is two cycles after L2Tag receives the second dummy data cycle from the 1st read request. The next possible assertion of l2t_sii_iq_dequeue must be a minimum of 16 clock cycles after the 1st assertion. This minimum pulse period of once every 17 cycles avoids the bus contention on the data return path from L2 Bank to SIU (1 header cycle + 16 payload cycles).

A read request inbound to L2 will generate a response packet from L2 on the outbound path. An inbound read request from SIU should not overflow SIU's receive header and data buffers in the outbound direction. SIU's outbound L2 subunit sends dequeue signals to the inbound L2 subunit to communicate buffer resource availability and SIU's inbound L2 subunit increments and decrements its credit counters.

For a read request, there is no data payload to protect, so the ECC lines (sii_l2b_ecc) are a don't care.

FIGURE 6-6 SIU to L2: Back to Back Reads



Because the L2 Tags are physically way across the chip from SIU or a one way distance of 9 to 10mm, two cycles of delay staging flops per direction will most likely be required to accommodate the paths between SIU and L2. The timing diagrams shown in this specification do not push out the signals to account for the delay stages.

Back to Back WR8 Request follow by WRI Request

A Write8 byte is a partial store in L2 cache and the packet transfer consists of two two cycles of header followed by two cycles of payload and one dummy turnaround cycle. The bytemask is encoded in the header. In the current L2 implementation, a WR8 request does not consume any data I/O write buffer entry. Instead, L2 pumps

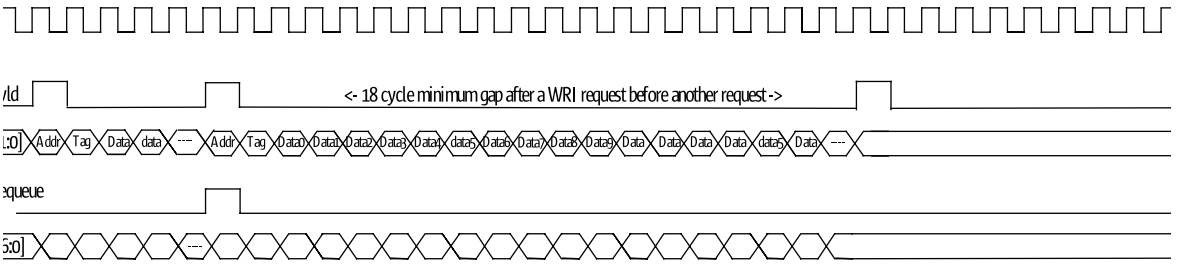
the 64 bit data into pipeline stages of 64 flops. Like the read request to L2, SIU must have a request token available before it can send a write request. L2 asserts l2t_sii_iq_dequeue for one cycle when it sends the WR8 down its pipeline during the first pass. Note that for a WR8, SIU does not need a data token and should not decrement its data credit for a WR8 transfer. For performance purpose, SIU does not issue any WR8 with all bytes off.

A Write Invalidate (WRI) invalidates L2 if there's a tag match and moves all 64 bytes to memory. A WRI packet transfer consists of two cycles of header followed by 16 cycles of payload and one dummy turnaround cycle. L2 does not move the write data to the data cache array (for either WR8 or WRI) until L2 has accumulated the entire data payload. Like the read request to L2, SIU must have a request token available before it can send a write invalidate request. SIU will also need a data token (initially set to four to match the four I/O write buffer entries in L2 Tag). SIU decrement its data credit for a WRI transfer. Although not shown in FIGURE 6-7, L2 asserts l2t_sii_iq_dequeue for one cycle when it sends the WRI request down its pipeline and L2 also asserts l2t_sii_wib_dequeue for one cycle when L2 moves the 64 byte write data out of the I/O write buffer.

ECC (sii_l2b_ecc[6:0]) is generated to protect the content of each data cycle. The ECC algorithm used by SIU is the same as used by L2 for its data array and produces seven check bits for a set of 32 data bit. Note that because ECC algorithm used by L2 is different from memory, L2 will check the ECC from SIU and will regenerate new ECC for memory for a WRI request.

Any nonposted write request inbound to L2 will generate an ack packet from L2 on the outbound path. An inbound write request from SIU should not overflow SIU's receive header buffers in the outbound direction. SIU's outbound L2 subunit sends dequeue signals to the inbound L2 subunit to communicate buffer resource availability and SIU's inbound L2 subunit increments and decrements its credit counters.

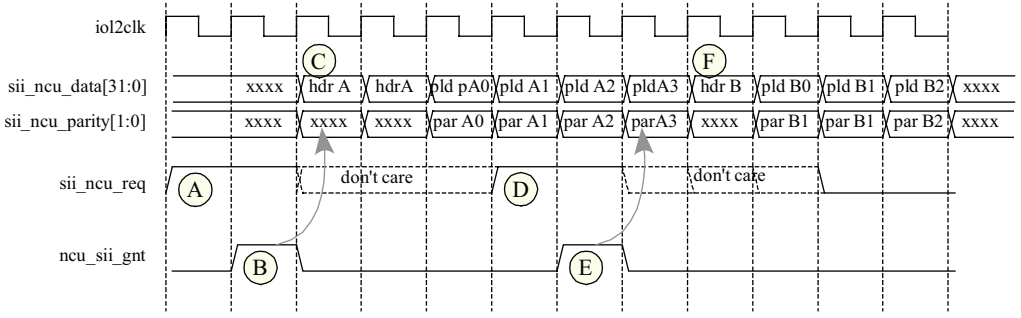
FIGURE 6-7 SIU to L2: Back to Back Writes (WR8 followed by WRI)



6.4.4.4 From SIU to NCU

FIGURE 6-8 shows fastest possible back to back transfer from the SIU to NCU. This could be for Interrupt or PIO completion. SIU signals NCU it wants to transfer a packet. The request signal is held high until SIU sees ncu_sii_gnt asserts (it only asserts for one clock cycle). The cycle after grant, SIU drives for five cycles always, beginning with the header and four cycles of payload/parity. While SIU is driving the data lines, SIU may reassert the request line if it has more work to do. NCU is expected to ignore the request line until two cycles before the current packet to check if another transfer is requested. If NCU has space, it will reassert grant for one cycle again for the next transfer. The earliest the second grant can assert is on the fourth payload cycle of the first transfer. This would allow for back to back transfer with no bubbles on the sii_ncu_data bus.

FIGURE 6-8 Timing Diagram for Packet from SIU to NCU (Back to Back Transfer)



Ⓐ : *SIU signals NCU for transaction request;*
Ⓑ : *NCU sends grant signal to indicate ready to accept packet in next cycle;*
Ⓒ : *SIU starts new packet;*
Ⓓ : *2 cycles before finishing the current packet, NCU checks if there is another request;*
Ⓔ : *NCU sends grant signal to indicate ready to accept new packet in next cycle;*
Ⓕ : *SIU starts new packet..*

6.4.4.5 From TCU to SIU

There will be two bits interface (tcu_sii_vld, tcu_sii_data) from TCU to SII for DMA read/write access.

tcu_sii_data is a 128/64 bit data stream with 64-bit header, 64-bits of data in case of write. tcu_sii_vld asserted at the first, 64nd cycle on valid tcu_sii_data. There will be 128 bits (header+data) for DMA write and 64-bits for DMA read.

Header format:

bit[63:56] = 0x81 for read, 0x82 for write

bit[55:40] = 0x00 reserved

bit[39:0] = eight byte aligned physical address (bit[5:0])

6.4.5 SIU's Inbound Pipeline

6.4.5.1 Major Pipeline Stages

There are four major pipeline stages in the inbound transfer. The best case total latency is eight cmp clock cycles + four IO clock cycles. The worst case total latency to NCU is 15 cmp clock cycles + 15 IO clock cycles. The worst case total latency to L2 Tag is 15 cmp clock cycles +32 cmp clock cycles. The following subsections will discuss the latencies of each stages in details. Refer to FIGURE 6-9.

Stage1: Interface (3-7 IO clock cycles)

The interface latency between DMU/NIU to SIU is between from three to seven cycles. Request and grant arbitration costs two cycles. Although that latency can be hidden when there are back to back requests and the first request transfers at least one payload cycle, it must be taken into account. The best case is one cycle of read request to either L2 or NCU. The worst case is one cache line write request from DMU/NIU to L2. It is one cycle of header plus four cycles of payload. The transaction on the bus will be registered and written to the fifo (register file) in this stage.

Stage2: Write to Fifo (1 to 4 cmp clock + 1 IO clock cycles)

This stage includes the header decoding and address lookup to set dependency for DMU packets followed by write to the fifo (register file). Once the last cycle of packet has been written into the FIFO and to disallow flow through FIFO, read pointer synchronization across the clock domain takes a minimum of one to four cmp clock cycles. one IO cycle (3 or four cmp clock cycles) for header decoding and register file write, and one to four cmp clocks for the read ptr synchronization.

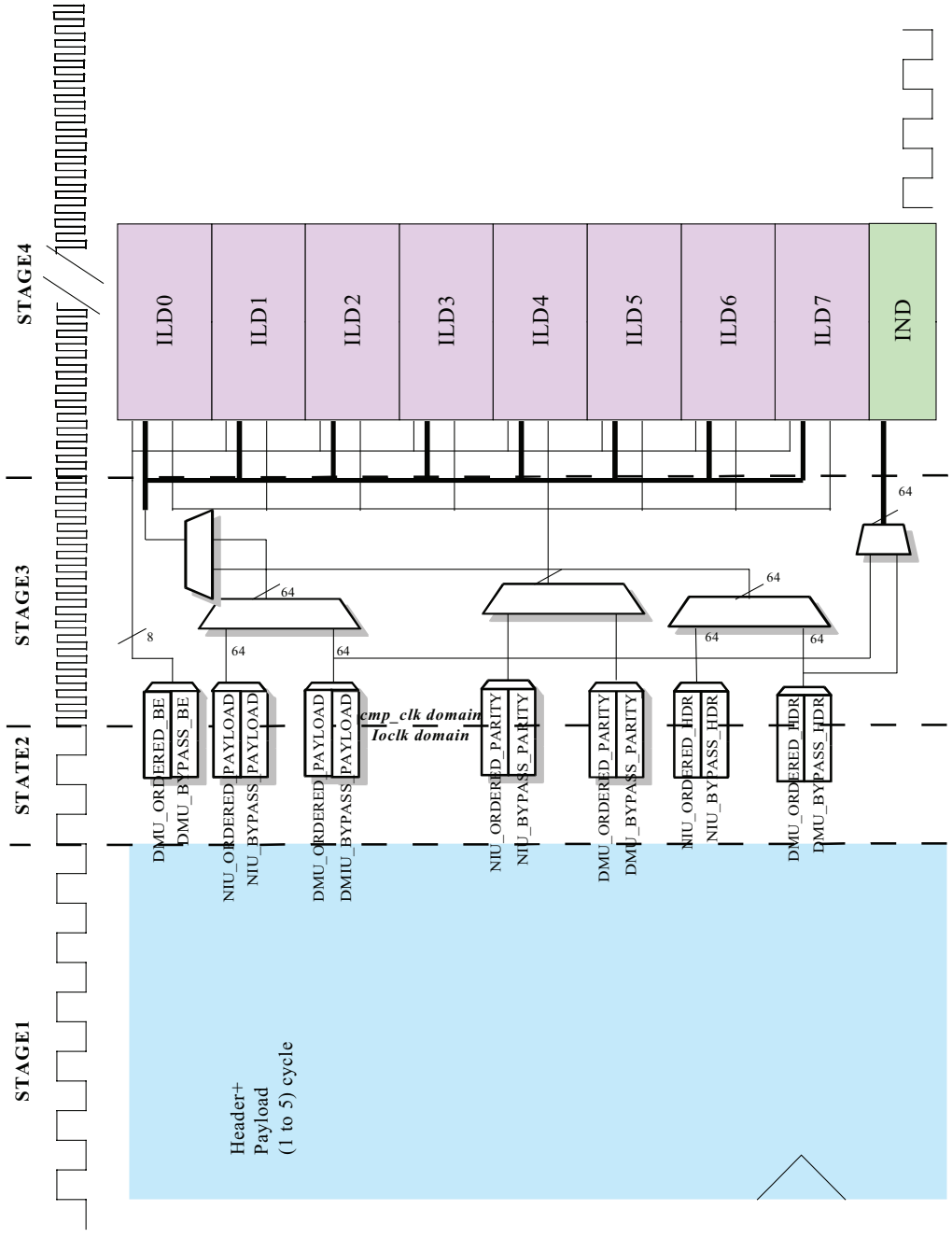
Stage3: Read from Fifo and arbitration (3-11 cmp clock cycles)

There are two cycles for arbitration between different fifo queues (DMU ordered, DMU bypass, NIU). In this stage, the arbitor check for the resources availability and priority of each queue to grant the transfer. Depending on the type of transaction, the transfer going to the queues in the inbound L2 subunits may take one or 9 cmp cycles. The transfer going to the queues in the inbound NCU subunit takes 1, 2 or 3 cmp cycles. Inbound toward L2 contributes to both the best case and worse case latency.

Stage4: L2 interface (4-32 cycles), NCU interface (4 to 7 IO clock cycles)

This stage includes latency of either the Ilks or IND. In the IND, SIU crosses back from the cmp clock domain into the IO domain. That pointer synchronization plus writing and reading from the width conversion FIFO take two IO clock cycles. The subsequent transfer to NCU takes two to five IO clock cycles. In the ILDs, there are two cycles of header (Addr Tag) and (0 to 16 cycles) for regular RDD, WRI, WR8 transaction. However, when there is a WRM request, since L2 only merges eight bytes, a WRM will be broken down to a maximum of eight WR8 transfers. With each WR8 being four cmp cycles (two cycles of header and two cycles of payload), and assuming L2 can stream the merge pipeline, the worse case is 32 cmp cycles for eight transfers. The best case is RDD (two cycles of header + three dummy data cycles) and the worse case is WRM (32 cycles).

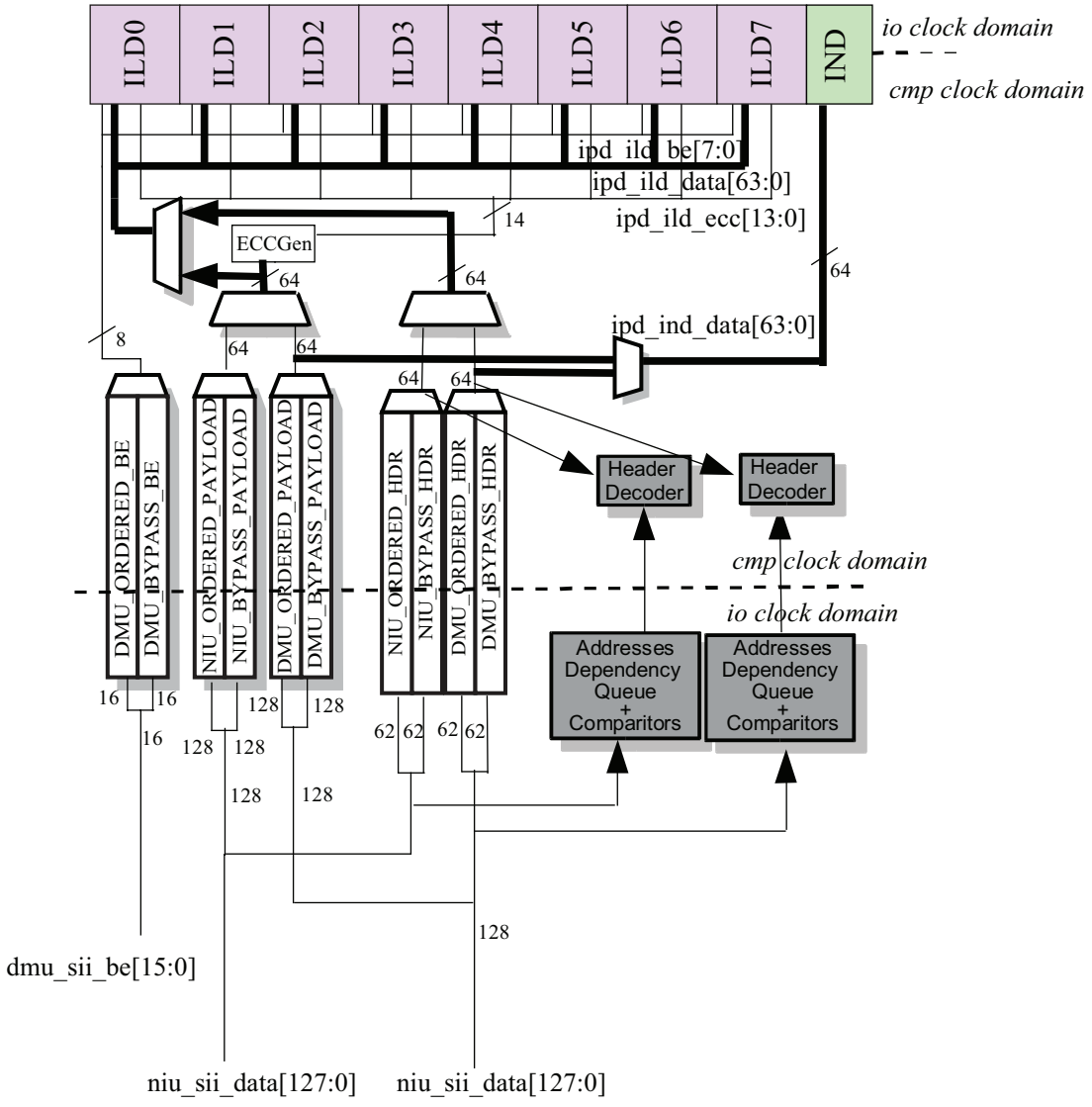
FIGURE 6-9 Inbound Pipeline Diagram



6.4.6 Block Diagrams of SIU Inbound

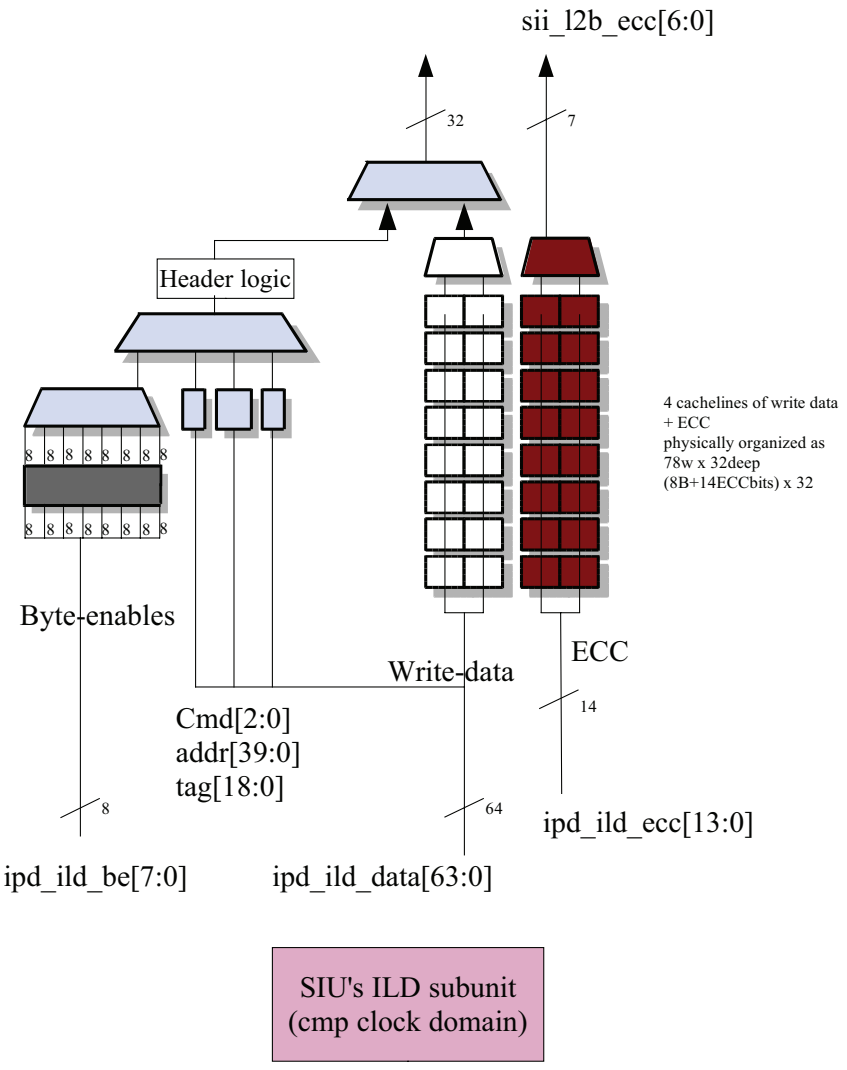
6.4.6.1 Top Level Block Diagrams

FIGURE 6-10 SIU Inbound Top Level



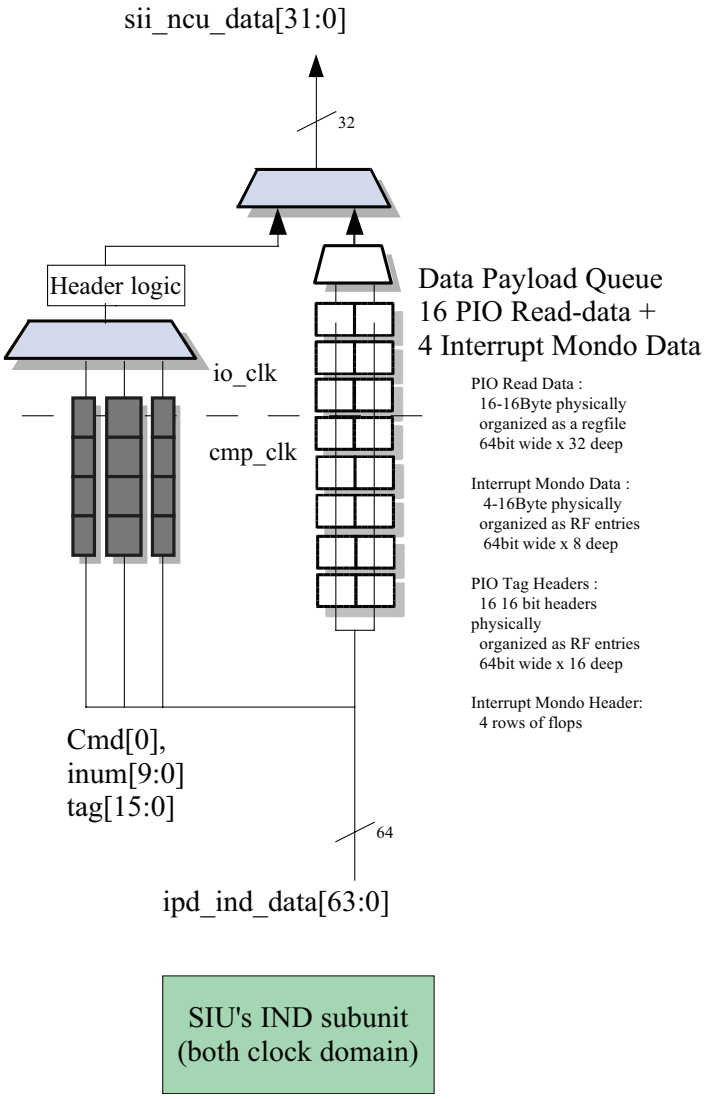
6.4.6.2 Sub-Blocks - ILD

FIGURE 6-11 SIU Inbound L2 Datapath (ILD) Subunit



6.4.6.3 Sub-Block - IND

FIGURE 6-12 SIU Inbound NCU Datapath (IND) Subunit



6.4.6.4 Sub-Block Descriptions

Top

The logical top level of the SIU inbound logic consist of control sub-blocks and data path sub-blocks. The datapath include the following sub-blocks:

ILDs: There are eight identical copies of ILD instantiated in the inbound logic. Each ILD can hold four requests and four cachelines of write data and associated byte-enables, ECC and header information. ILDs handle transaction going to L2 cache, it has one unified data queue 78bits wide x 32 deep to hold the four cachelines - writing in 64 bit data + 14 bit ECC per cycle. It acts as buffer for assembling of SIU_L2 packets.

IND: IND handles transfer going to NCU. It acts as a buffer for assembling SIU_NCU packets. It can buffer 16 PIO Read 16Byte responses and four interrupt 16Byte mondo write requests (2 32-bit x 40 data queues, 4x26-bit register for Interrupt header storage, 16 16-bit register for PIO Read Return header storage). IND runs in both clock domains. It down converts the cmp clock frequency of IPD into the IO clock frequency because NCU's interface to SIU runs in the IO clock domain.

DMU_ORDERED header queue: it is a 62-bit wide x 16 depth register file with input and output registered. The write operation is running at Io_clock domain, and the read operation is running at L2_clock domain.

DMU_BYPASS header queue: it is a 62-bit wide x 16 depth register file with input and output registered. The write operation is running at Io_clock domain, and the read operation is running at L2_clock domain.

NIU_ORDERED header queue: it is a 62-bit wide x 16 depth register file with input and output registered. The write operation is running at Io_clock domain, and the read operation is running at L2_clock domain.

NIU_BYPASS header queue: it is a 62-bit wide x 16 depth register file with input and output registered. The write operation is running at Io_clock domain, and the read operation is running at L2_clock domain.

DMU_ORDERED parity queue: it is a 1-bit x 16 register. It hold the parity bit of the whole packet including header and payload.

DMU_BYPASS parity queue: it is a 1-bit x 16 register. It hold the parity bit of the whole packet including header and payload.

NIU_ORDERED parity queue: it is a 1-bit x 16 register. It hold the parity bit of the whole packet including header and payload.

NIU_BYPASS parity queue: it is a 1-bit x 16 register. It hold the parity bit of the whole packet including header and payload.

DMU_ORDERED_BE queue: it is a 16-bit x 16 register file with input and output registered. The write operation running at Io_clock domain, and the read operation is running at L2 clock domain.

DMU_BYPASS_BE queue: it is a 16-bit x 16 register file with input and output registered. The write operation running at Io_clock domain, and the read operation is running at L2 clock domain.

DMU_ORDERED payload queue: it is a 128-bit x 64 depth queue logically. It is implemented with one 128-bit x 32 depth register file with input and output registered. The write operation runs at Io_clock domain and the read operation runs at L2_clock domain.

DMU_BYPASS payload queue: it is a 128-bit x 64 depth queue logically. It is implemented with one 128-bit x 64 depth register file with input and output registered. The write operation runs at Io_clock domain and the read operation runs at L2_clock domain.

NIU_ORDERED payload queue: it is a 128-bit x 64 depth queue logically. It is implemented with one 128-bit x 32 depth register file with input and output registered. The write operation runs at Io_clock domain and the read operation runs at L2_clock domain.

NIU_BYPASS payload queue: it is a 128-bit x 64 depth queue logically. It is implemented with one 128-bit x 64 depth register file with input and output registered. The write operation runs at Io_clock domain and the read operation runs at L2_clock domain.

The control path include the following sub-blocks: IPC, ILC, Refer to the next subsections for the details of IPC and ILC sub-blocks.

ILC Sub-block

There are eight identical copies of ILCs instantiated in SIU. ILCs are running at core clock domain. There are two major functions of ILC sub-block:

1. It checks the L2 bank's availability. There are two counters in each ILC to keep track of outstanding L2 transactions. One is the transaction counter, which keep track of outstanding L2 requests which are issued to L2 and no acknowledgments come back yet. The second counter keeps track of the WRI requests. The requirements from L2 stated that L2 allow two outstanding requests (WR8, RDD, WRM) and four outstanding WRI requests. As long as the counters' value satisfy the L2 requirements, the particular L2 bank is considered as available. The availability information will be passed to IPC for arbitration purposes.

2. ILC will drive the SIU-L2 interface bus according to the protocol defined in the previous sections. Also it will assemble the packets with header and payload formats according to the SIU-L2 packet format defined in the previous sections.

Note – For the WR8 transfer with no bit set in the byte-mask, SIU will just discard the transaction.

For WRM of 64 byte merge, ILC will break it down to eight WR8 requests and assemble eight packets for it.

Should a WRM of 64 bytes have no bytes on, SIU will not send out any WR8 packets to L2 but will wait for all outstanding acks from L2 to return and then inject a single fake write response into the outbound L2 subblock for return to DMU if needed.

INC Sub-block

The INC sub-block is running at core clock domain, however part of the logic is running at IO clock domain for the purpose of driving the SIU-NCU interface signals. There is no flow control between SIU and NCU for read returns, NCU is considered as always available with respect to read returns. However INC need to keep track of outstanding interrupts, there are four outstanding interrupts allowed in SIU. Once it reached that number, it will signal the IPC to block further interrupt requests to NCU. Another function of INC is to assemble the packet following the SIU-NCU packet format and drive the SIU_NCU interface accordingly.

IPC Sub-block

There are two identical copies of IPCs instantiated in SIU, one for DMU and one for NIU. IPCs are mainly running at IO clock domain. However, part of the logic is running at core clock domain to handle the cross clock domain situation. The logic will be partitioned into two parts (IPCC and IPCS) according their clock domain.

IPCS sub-block implement two major functions:

1. IPCS will drive the DMU/NIU – SIU interface bus according to the protocol defined in previous section. It check for the availability of different queues to maintain flow control for the interface.

2. IPCS maintains the ordering rules for the input side from the DMU and NIU. It uses the sideband signals and header information to dispatch requests from DMU to DMU_ORDERED queue and DMU_BYPASS queue and from NIU to NIU_ORDERED queue and NIU_BYPASS queue.

IPCS has access to two FIFOs containing addresses and write/read bit duplicating the addresses and command type in the ordered and bypass FIFOs. Assuming the register-file FIFOs are not cam-able, the two FIFOs must be made of flops for address comparison. IPCS maintains 5 pointers – the location of the youngest write entry in the bypass queue, the youngest entry in the bypass queue, the youngest entry in the

ordered queue, the oldest entry in the ordered queue, the oldest entry in the bypass queue. IPCS and IPCC communicates with each other to maintain these pointers. All packets from an interface updates the duplicate address FIFOs for that interface.

For the newest write entering the ordered queue, IPCS tags it as dependent on the younger of the two entries: youngest write in the bypass queue or the youngest matching cacheline address in the bypass queue. This is done by storing a pointer to the bypass queue and setting a dependency pointer valid bit.

For the newest read entering the ordered queue, IPCS tags it as dependent on the youngest entry with the same address in the bypass queue. This is done by storing a pointer to the bypass queue and setting a dependency pointer valid bit.

For the newest write or read entering the bypass queue, IPCS tags it as dependent on the youngest entry with the same address in the ordered queue. This is done by storing a pointer to the ordered queue and setting a dependency pointer valid bit.

In PCIExpress mode, IPCS tags each newest entry as dependent on the youngest entry in the opposite queue.

IPCC sub-block implements two major functions:

1. IPCC will drive the output buses of the inbound packet fifos according to the protocol defined in previous section. It check for the availability of different queues in IND and ILDs to maintain flow control.

2. IPCC maintain the ordering rules for the output side for DMU packet and does a two level arbitration between NIU and DMU. The top level arbitration is between NIU and DMU packets on a deficit round robin basis. NIU packets have no ordering requirement with respect to other packets from DMU. DMU packets have no ordering requirement with respect to packets from NIU. The second level arbitration is between the two DMU FIFOs.

IPCC maintains two counters – a bypass write counter and an ordered write counter. The bypass write counter counts the number of writes sent from the bypass queue but have not received their acknowledgements. The ordered write counter counts the number of writes or interrupts that were sent from the ordered queue but have not received their acknowledgements. An ordered target ID tracks which of the nine targets (eight L2, one NCU) was the last issued write/interrupt from the ordered queue.

When a write/interrupt/PIO read return reaches the top of the ordered queue and has its dependency pointer valid bit set, IPCC first checks if that entry in the bypass queue has been dequeued. If it has and the bypass write counter reaches zero, then a second check is made. After the bypass write counter has reached zero, if it's a read return then it may continue, else the following is decided. A comparison between the ordered target ID against the destination of the write/interrupt is made. If they are the same, then it may continue. If they are not the same, then the write/interrupt

must wait until the ordered write counter has reached zero before it can continue. When the write/interrupt packet dequeues, the ordered write counter is incremented and the ordered target ID is updated.

When a read reaches the top of the ordered queue and has its dependency pointer valid bit set, IPCC first checks if that entry in the bypass queue has been dequeued. If it has and the bypass write counter reaches zero, then a second check is made. A comparison between the ordered target ID against the destination of read is made. If they are the same, then it may continue. If they are not the same, then the read must wait until the ordered write counter has reached zero before it can continue. No counter is incremented when the read dequeues.

When a write or read reaches the top of the bypass queue and has its dependency pointer valid bit set, then IPCC first checks if that entry in the ordered queue has been dequeued. If it has, then the write may continue. Once dequeued and if its a write, then the bypass write counter is incremented. No counter is incremented when a read dequeues.

When a flush reaches the top of the ordered queue and has its dependency pointer valid bit set, IPCC checks if that entry in the bypass queue has been dequeued. If it has and the bypass write counter reaches zero, then a second check is made. After the bypass write counter has reached zero, must wait until the ordered write counter has reached zero before it is dequeued. When a flush dequeues and it's nonposted, a response packet is injected into the outbound path to return to DMU.

6.4.6.5 Reliability, Availability, and Serviceability (RAS)

Syndrome format of sii_ncu_syn_data[63:0]

sii_ncu_syn_vld will cover the 16 io cycles of syndrome.

Cycle0: send sii_ncu_syn_data[3:0]

Cycle1: send sii_ncu_syn_data[7:4]

.

.

Cycle15: send sii_ncu_syn_data[63:60]

```
==================================

>> bit[63:62]     = 2'b00         (in case of future changes!)

>> bit [61]      = niud_pe

>> bit [60]      = niua_pe

>> bit [59]      = niuctag_ue
```

```
>> bit [58]      = dmud_pe
>> bit [57]      = dmua_pe
>> bit [56]      = dmuctag_ue
>> bit [55:40]   = ctag[15:0]
>> bit [39:0]    = Physical Address[39:0]
>> ================================
```

6.5 Outbound

6.5.1 Interface Timing Diagrams

6.5.1.1 From L2 to SIU

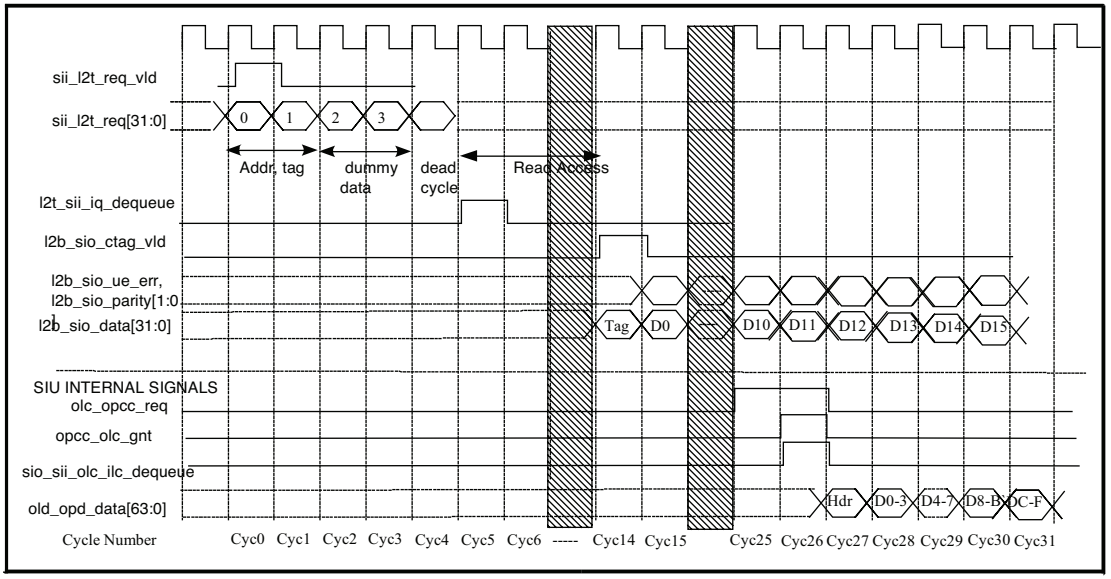
Single Read Response from L2

FIGURE 6-13 shows the quickest a read request inbound to an L2 bank will return outbound to SIU. The greyed bars in the diagram mean some clock cycles are not shown.

After L2Tag has accumulated the read request packet with the dummy cycles for pipeline alignment (shown as sii_l2t_req[31:0]), the read request can be dispatched down L2's pipeline. The l2t_sii_iq_dequeue signal asserts high for one cycle when L2 dispatches the read. That signal is used by SIU's Inbound L2 Control subunit for credit based flow control. In the current L2 pipeline, the earliest L2 can assert l2t_sii_iq_dequeue is two cycles after the last dummy data cycle of the request packet.

L2 Bank asserts the l2b_sio_ctag_vld signal high for one cycle to indicate the completion of the read. On the cycle that l2b_sio_ctag_vld asserts, l2b_sio_data has a read response header. The subsequent 16 cycles of l2b_sio_data[31:0] contain the 64B read data. Parity is generated on l2b_sio_parity[1:0] and corresponds to the parity of each data word of l2b_sio_data. The l2b_sio_ue_err signal is also active during the data cycles to indicate that L2 had detected an uncorrectable error for that data word.

FIGURE 6-13 L2 Read Data Return Timing Diagram (Fastest case is shown)



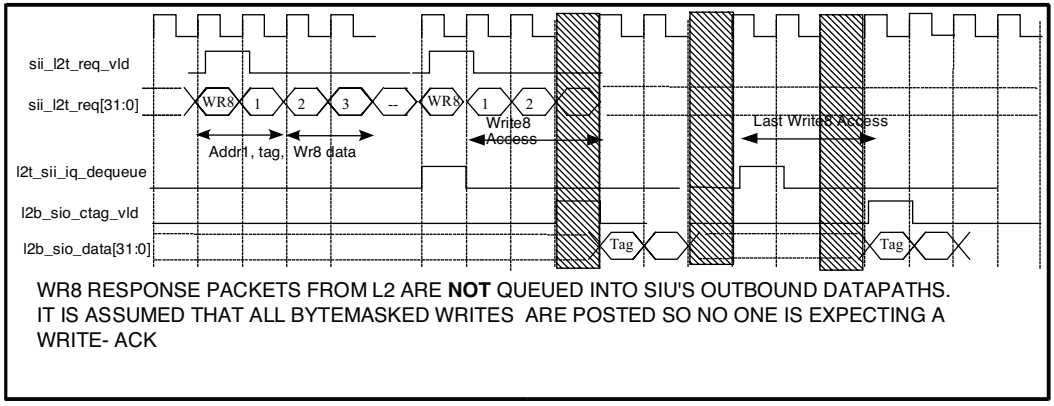
The bottom part of the timing diagram represents the read response continuing outbound internally through SIU. Read requests can be pipelined to L2 and the responses pipelined back to SIU. However because there is no direct flow control to stop a response from L2 from overflowing SIU's outbound response buffers, SIU's Outbound L2 Control subunit asserts internal signal sio_sii_olc_ilc_dequeue high for a cycle when a credit is returned to SIU's Inbound L2 Control subunit. The diagram shows SIU forwarding the read response packet to the outbound packet data subunit on the next data bus (old_opd_data[63:0]) as early as possible. The earliest this can happen is after SIU has received 3/4 the response payload from L2. The assumption made here is that SIU can forward the accumulated result of all the parity checks and uncorrectable errors later. Otherwise, SIU pays the latency of accumulating the full 64B response payload. Working backward from the databus old_opd_data are SIU's outbound internal arbitration request-grant signals olc_opc_req and opc_olc_gnt. The dequeue signal sio_sii_olc_ilc_dequeue for the inbound path can simply be a buffered version of the grant signal opc_olc_gnt as shown, if the timing can be made. Otherwise it will assert a cycle later.

Due to stall conditions such as a cache miss, the latency between the assertion of l2t_sii_iq_dequeue and l2b_sio_ctag_vld can exceed the 9 cycles shown in the above diagram. Likewise, due to back stall conditions further downstream in the outbound paths, the latency between the assertion of sii_l2t_req_vld and sio_sii_olc_ilc_dequeue can exceed the 26 cycles shown in the above diagram.

Write 8 Responses from L2 to SIU

FIGURE 6-14 shows two WR8 requests inbound to an L2 bank followed by three signals to related to the WR8 responses returning outbound to SIU from L2. The greyed bars in the diagram mean some clock cycles are not shown.

FIGURE 6-14 L2 Write8 Acknowledgement Timing Diagram



After L2Tag has accumulated the WR8 request packet (shown as sii_l2t_req[31:0]), the request can be dispatched down L2's pipeline. The l2t_sii_iq_dequeue signal asserts high for one cycle when L2 dispatches the write. That signal is used by SIU's Inbound L2 Control subunit for credit based flow control. In the current L2 pipeline, the earliest L2 can assert l2t_sii_iq_dequeue is two cycles after the last data cycle of the request packet. The l2t_sii_iq_dequeue signal guarantees write ordering has occurred in L2. However, it does not mean the write has completed.

For the WR8 acknowledgement, the latency between l2t_sii_iq_dequeue and l2b_sio_ctag_vld depends on whether the store missed the L2 or not. L2 handles the merge like an atomic read-modified-write operation. A subsequent request from SIU would not overtake the WR8 should the WR8 miss L2 on the first pass and must wait to be reissued after a fill. Note that there is no new assertion of l2t_sii_iq_dequeue on second pass of a WR8.

The timing diagram shows that when l2b_sio_ctag_vld asserts, L2 Bank responds with a tag on signal l2b_sio_data[31:0]. This tag does not contain enough information for SIU to know if the WR8 response is the last of a sequence of up to eight WR8 packets that were decomposed from a single write transaction. The assumption made in the SIU is that DMU will never need a ack response for writes with bytemasks (all memory writes with bytemask from DMU are posted) and hence SIU Inbound and Outbound paths do not need to coordinate to scoreboard the tag.

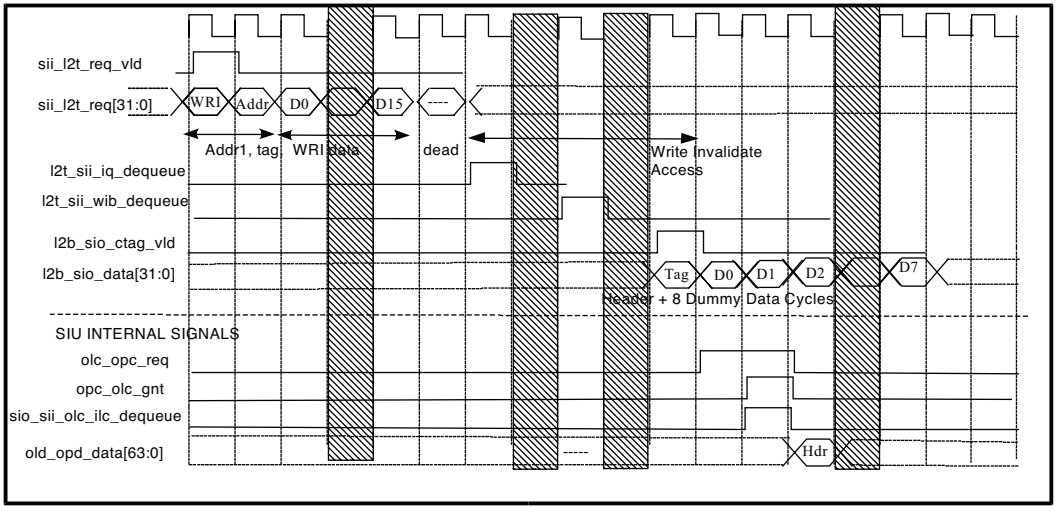
Similarly, the assumption is that NIU will never generate bytemask writes requests to SIU and thus SIU will never generate WR8 for NIU and again SIU would not need to scoreboard the tag after SIU issue to L2.

Because the outbound path never needs to respond to a WR8 response from L2, those responses never enters any queue in the SIU outbound L2 data or control paths.

Write Invalidate Response from L2 to SIU

FIGURE 6-15 shows a Write Invalidate request inbound to an L2 bank and the outbound return of the write response. The greyed bars in the diagram mean some clock cycles are not shown.

FIGURE 6-15 L2 Write Invalidate Acknowledgement Timing Diagram



After L2Tag has accumulated the 18-cycle WRI request packet (shown as sii_l2t_req[31:0]), the request can be dispatched down L2's pipeline. The l2t_sii_iq_dequeue signal asserts high for one cycle when L2 dispatches the write. That signal is used by SIU's Inbound L2 Control subunit for credit based flow control. In the current L2 pipeline, the earliest L2 can assert l2t_sii_iq_dequeue is two cycles after the last data cycle of the request packet. The l2t_sii_iq_dequeue signal guarantees write ordering has occurred in L2. However, it does not mean the write has completed. When L2 Tag drains the 64 bytes of write data from its I/O Write Buffer and the data are enroute to the memory controller, the l2t_sii_wib_dequeue signal asserts for one cycle. This signal is used only for a Write Invalidate request as a Write8 request does not consume any data buffer in L2. The

SIU Inbound L2 subunit monitors the signal l2t_sii_wib_dequeue for credit based flow control of L2's four I/O Write Buffers. Many cycles later, the l2b_sio_ctag_vld signal asserts for one cycle to indicate the completion of the write. When l2b_sio_ctag_vld asserts, l2b_sio_data contains the header of the write response packet. eight cycles of dummy data follows to align L2's store pipeline. The bottom part of the timing diagram represents the write response continuing outbound through SIU. For flow control, the outbound L2 control subunit signals to the inbound L2 control subunit that an entry has been dequeued.

6.5.1.2 From SIU to NIU

The assumption is that the NIU has buffers to receive all DMA responses from SIU. There is no flow control from NIU to throttle SIU Outbound and thus SIU is allowed to send responses back to back without any bubble to NIU.

Writes from NIU can be either posted or nonposted. For nonposted writes, SIU must return an acknowledgement response back to NIU when the L2 has acknowledged completion of the write. There is no requirement of the SIU to return the ack response packet in the same order that NIU delivered the write request nor in the order that the write completed in memory.

The parity protected interfaces between SIU and the NIU are 128 bit wide with side band signals for packet control. Having a 128 bit for header allows SIU to provide a rich set of transaction types and allows SIU to provide a uniform and generic but flexible enough for most IO architectures. See SIU-DMU Interface List and SIU-NIU Interface List.

The outbound packet interface protocol works as follows:

Cycle: Header Cycle

- sio_niu_hdr_vld asserts for one cycle to indicate SIU is sending the packet header to niu.
- sio_niu_datareq is set to 1 to indicate this packet has a four cycle payload following the header cycle to transfer 64 Bytes of data. It is set to 0 to indicate this packet is an write acknowledge and has no data payload.
- sio_niu_data[127:0] contains a valid header.

Cycle 2-5: Payload Cycles if sio_niu_datareq was asserted during Header Cycle.

- sio_niu_hdr_vld is deasserted.
- sio_niu_data[127:0] contains the payload data. Data is returned big endian and critical four Byte first and wraps back to the beginning of the cacheline when it reaches the cacheline boundary.
- sio_niu_parity[3:0] contains the parity for each 32 bit of data. Parity[N]= xor(data[32N+31: 32N])

6.5.1.3 From SIU to DMU

The assumption is that the DMU has buffers to receive all DMA responses from SIU. There is no flow control from DMU to throttle SIU Outbound and thus SIU is allowed to send responses back to back without bubble to DMU.

Fire-DMU never issues nonposted DMA writes so all responses from SIU to Fire-DMU has a four cycle payload.

The parity protected interfaces between SIU and the DMU are 128 bit wide with side band signals for packet control. Having a 128 bit for header allows SIU to provide a rich set of transaction types and allows SIU to provide a uniform and generic but flexible enough for most IO architectures.

The outbound packet interface protocol works as follows:

Cycle 1: Header Cycle

- sio_dmu_hdr_vld asserts for one cycle to indicate SIU is sending the packet header to dmu.
- sio_dmu_datareq is set to 1 to indicate this packet has a four cycle payload following the header cycle to transfer 64 Bytes of data.
- sio_dmu_data[127:0] contains a valid header.

Cycle 2-5: Payload Cycles if sio_dmu_datareq was asserted during Header Cycle.

- sio_dmu_hdr_vld is deasserted.
- sio_dmu_data[127:0] contains the payload data. Data is returned big endian and critical four Byte first and wraps back to the beginning of the cacheline when it reaches the cacheline boundary.
- sio_dmu_parity[3:0] contains the parity for each 32-bit of data. Parity[N]= xor(data[32N+31: 32N])

6.5.1.4 From SIO to TCU

There will be two bits interface (sio_tcu_vld, sio_tcu_data) from SIO to TCU for DMA read/write response.

sio_tcu_data is a 64-bit data stream for read request, and 1bit of data=0 for write request.

Header format:

bit[63:0] = eight bytes of read return data for read request

bit[0] = 0x0

6.5.2 Outbound Pipeline

6.5.2.1 From L2

L2 Bus Cycle Packets (Write Acknowledge)

1. L2B-OLD Header Enqueue (one L2 cycle: 32 bit bus)
2. OLD-OPD Request (one+ L2 cycle)
3. OLD-OPD Grant (one L2 cycle)
4. OLD-OPD Transmit/Muxing Wire Delay (one to two L2 cycles)
5. OLD-OPD Header Enqueue (one1 L2 cycle: 64 bit bus)
6. OPD-OPD Domain Crossing (one-three L2 cycles)
7. OPD-DMU/NIU Header Enqueue (one IO cycle: 128 bit bus)

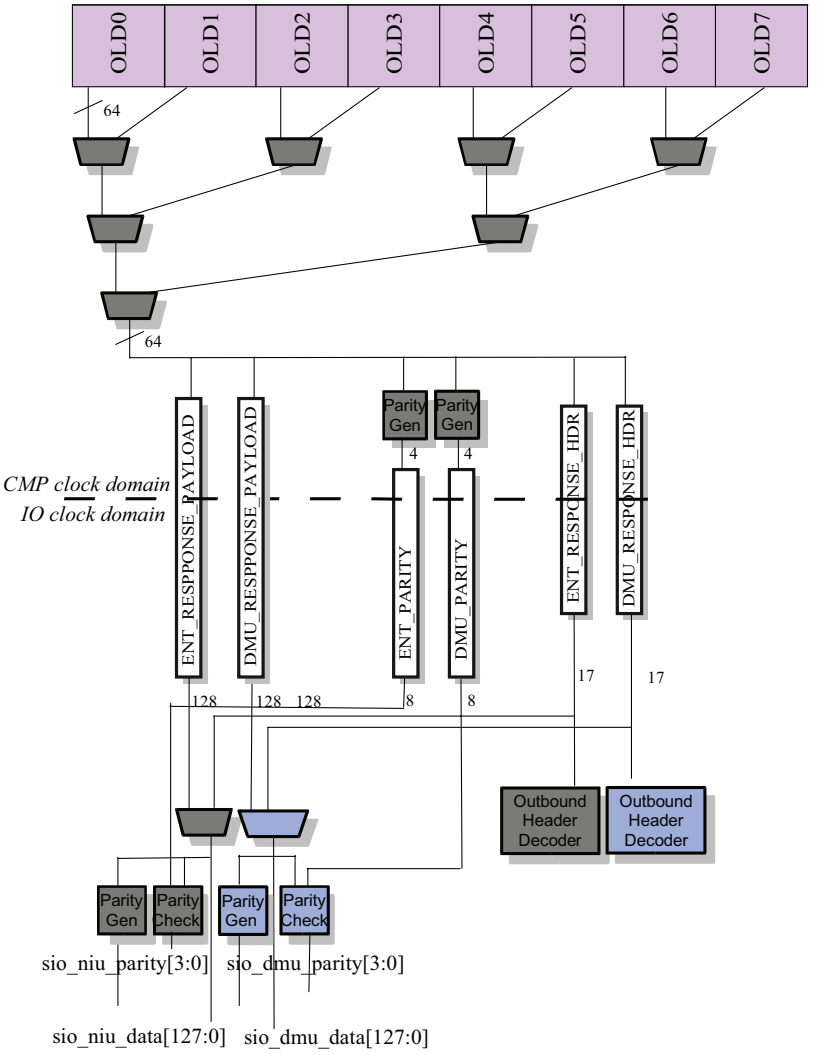
L2 Bus Cycle Packets (Read Response)

1. L2B-OLD Header Enqueue (one L2 cycle: 32 bit bus)
2. L2B-OLD Data Payload Enqueue (16 L2 cycles: 32 bit bus)
3. OLD-OPD Request (one+ L2 cycles)
4. OLD-OPD Grant (one L2 cycle)
5. OLD-OPD Transmit/Muxing Wire Delay (one to two L2 cycles)
6. OLD-OPD Header Enqueue (one L2 cycle: 64 bit bus)
7. OLD-OPD Data Payload Enqueue (8 L2 cycles: 64 bit bus)
8. OPD-OPD Domain Crossing (one-three L2 cycles)
9. OPD-DMU/NIU Header Enqueue (one IO cycle: 128 bit bus)
10. OPD-DMU/NIU Data Payload Enqueue (four IO cycles: 128 bit bus)

6.5.3 SIU Outbound Block Diagram

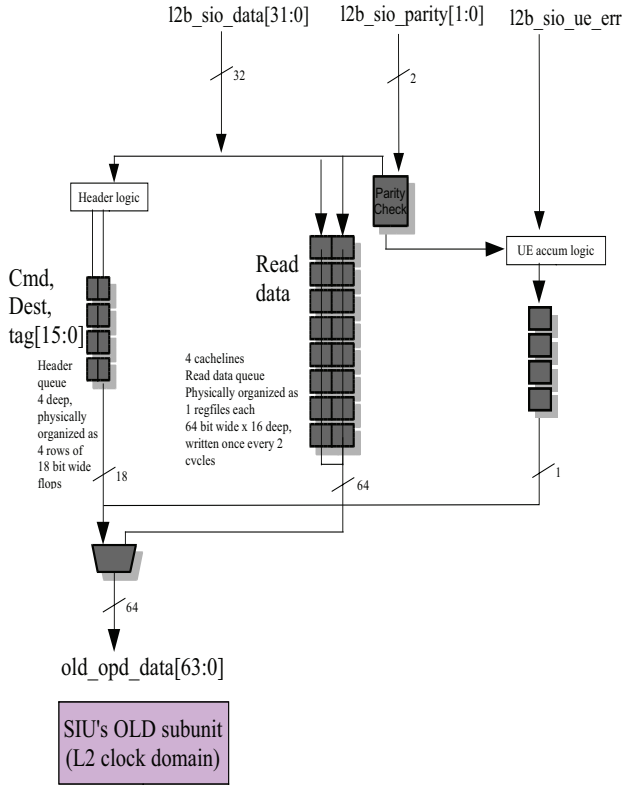
6.5.3.1 OPD: Outbound Packet Datapath

FIGURE 6-16 SIU Outbound Packet Datapath (OPD) Subunit



6.5.3.2 OLD: Outbound L2 Datapath

FIGURE 6-17 SIU Outbound L2 Datapath (OLD) Subunit



6.5.4 SIU Outbound Subunit Descriptions

6.5.4.1 Datapath

OLD0 to OLD7 receive and store response packets from L2 buffer. Each interface has a 64Byte four-deep payload buffers physically organized as two regfiles. Each regfile is 32bit wide x 32 deep – written in at cmp clock @ 32bit/cycle, read out at cmp clock @ 64bit/cycle. There is also a header queue in each interface to hold the tag information. That queue is 18 bits wide x four-deep.

The physical placement and organization of these OLDx queues are critical to the critical path in the outbound direction. The distance between the farthest two regfile's IO flops determines how far the mux-select lines would need to travel on the first level of 8:1 muxing of these 64 bit wire bundles.

OPD: The outbound packet datapath subunit holds packets from the OLDx subunits and stream them out to DMU and NIU. NIU and DMU each has a separate outbound packet queue 16 entry deep for responses from L2. The packet queues are physically separated into header queues, payload queues, and parity. The packets crosses clock domain from core clock to IO clock in OPD.

6.5.4.2 Control Path

OLC0 to OLC7 are identical instantiations of the control logic to enqueue and dequeue from the OLD FIFOs. Each makes a request to OPCC and waits for a grant before it ships the DMA response over to the Outbound packet datapath and FIFOs.

OPCC: The outbound packet control logic in the core clock domain monitors all the eight request lines from OLC0 to OLC7, checks that the destination FIFO is available and then drives grant and controls the mux selects for the 8:1 muxes. It also transfer FIFO write pointers to the IO clock domain so the corresponding outbound packet control logic in the IO clock domain can handle pushing the data out to DMU or NIU. OPCC also monitors the L2 response headers to signal to the inbound side how many and type of responses received so IPCC can do bookkeeping.

OPCS: The outbound packet control logic in the IO clock domain monitors the pessimistic FIFO write pointers from OPCC and communicates with DMU or NIU.

6.6 Packet Formats

6.6.1 Inbound To L2

6.6.1.1 WRI Packet

Write Invalidate (WRI) request must be 64 byte aligned and a full 64 bytes is written to memory. The bottom six bits of address for a WRI request are set to zero.

FIGURE 6-18 Write Invalidate Request

V	*ECC[6:0]*	*Packet Data [31:0]*			
	6 5 4 3 2 1 0	31 30 29 28 27 26 25 24 21 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8	7 6 5 4 3 2 1 0		
1	CtagECC[5:0]	J O P E S 1 0 0 Tag[15:0]	Address[39:32]		
	Invalid	Address[31:6], 6'b000000			
	Valid ECC	DataByte0	...	...	DataByte3
	Valid ECC	DataByte4	...	...	DataByte7
	Valid ECC	DataByte8	...	...	DataByte11
	Valid ECC	DataByte12	...	...	DataByte15
	Valid ECC	DataByte16	...	...	DataByte19
	Valid ECC	DataByte20	...	...	DataByte23
	Valid ECC	DataByte24	...	...	DataByte27
	Valid ECC	DataByte28	...	...	DataByte31
	Valid ECC	DataByte32	...	...	DataByte35
	Valid ECC	DataByte36	...	...	DataByte39
	Valid ECC	DataByte40	...	...	DataByte43
	Valid ECC	DataByte44	...	...	DataByte47
	Valid ECC	DataByte48	...	...	DataByte51
	Valid ECC	DataByte52	...	...	DataByte55
	Valid ECC	DataByte56	...	...	DataByte59
0	Valid ECC	DataByte60	...	...	DataByte63

J :Jtag access : 1= Jtag access from tcu, 0= regular dma packet from IO

O:Ordered bit : 1=From SIU Inbound Ordered Queue. Needed by SIU.

P:Posted bit : 1=Posted => Completion Ack by SIU to the source NOT needed,

0=Nonposted => Completion Ack by SIU to the source NEEDED

E: Error bit: (parity or uncorrectable error in header)

S:Source: 1=DMU, 0=NIU

T:Tag[15:0] generated by the source to track the transaction.

If the P bit is zero, the response packet will contain this 16-bit tag.

For this implementation, DMU and NIU will both guarantee packets with address errors (like unmapped address or illegal access) will not be sent to SIU. This was different than in OpenSPARC T1 where OpenSPARC T1's JBI set the error bit and changed the physical address to all zeros.

Parity error will be logged in the header if SIU or a prior unit had detected an uncorrectable error in the data payload.

Legal WRI Packet Encodings:

L2 does not look at the O, P, S and T fields from SIU but simply pipe them along back to SIU Outbound when L2 generates the response packet.

For this implementation, NIU never sends bytemasked writes which means all DMA writes from NIU will become WRI's not WR8. In NIU's case, DMA writes can be issued to the SIU's ordered queue or the SIU's bypass queue. However, a high percentage of NIU's DMAs will go into the bypass queue. NIU's writes can be posted or nonposted. Fire-DMU never issue nonposted memory writes. In the Fire-DMU implementation, all their DMA writes are restricted to the ordered queue.

So if the NIU and DMU interfaces are behaving correctly, the following header restrictions apply for WRI:

P must be 1 if from Fire-DMU.

O must be 1 if Fire-DMU.

PA[5:0] must be all 0s.

Header cycle bits 26:24 == 3'b100

6.6.1.2 WR8 Packet

For the Write8 bytes (WR8) request, random byte writes are supported provided at least one byte gets written. A byte mask field [7:0] is supported for the random byte writes with at least one byte mask = 1. Bytemask field is positional. The address for a WR8 request must be eight-byte aligned (the lower 3 bits of the address must be 0).

FIGURE 6-19 Write 8 Bytes Request

V	*ECC[6:0]*	*Packet Data [31:0]*			
	6 5 4 3 2 1 0	31 30 29 28 27 26 25 24 21 22 21 20 19 18 17 16	15 14 13 12 11 10 9 8	7 6 5 4 3 2 1 0	
1	CtagEcc[5:0]	J O P E S 0 1 0	ByteMask[7:0]	Address[39:32]	
0	Invalid	Address[31:3], 3'b000			
0	Valid ECC	DataByte0	...	...	DataByte3
0	Valid ECC	DataByte4	...	...	DataByte7

J :Jtag access: 1= Jtag access from tcu, 0= regular dma packet from IO

O:Ordered bit: 1=From SIU Inbound Ordered Queue. Needed by SIU.

P:Posted bit: 1=posted. Needed by SIU

E: Error bit: (parity or uncorrectable error)

S:Source: 1=DMU, 0=NIU;

Legal WR8 Packet Encodings:

L2 does not look at the O, P, S fields from SIU but simply pipe them along back to SIU Outbound when L2 generates the response packet.

For this implementation, NIU never sends bytemasked writes which means all DMA writes from NIU will become WRI's not WR8. Fire-DMU can send DMA writes that will decompose into WR8. Fire-DMU never issue nonposted memory writes. In the Fire-DMU implementation, all their DMA writes are restricted to the ordered queue.

So if the NIU and DMU interfaces are behaving correctly, the following header restrictions apply for WR8:

- S must be 1.
- P must be 1.
- O must be 1 if Fire-DMU.
- PA[2:0] must be 0s.
- Header cycle bits 26:24 == 3'b010

6.6.1.3 RDD Packet

FIGURE 6-20 RDD Requests

V	*ECC[6:0]*							*Packet Data [31:0]*																									
	6	5	4	3	2	1	0	31	30	29	28	27	26	25	24	21	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	CtagEcc[5:0]							J	O	P	E	S	0	0	1	Tag[15:8]								Tag[7:0]									Address[39:32]						
0	Invalid							Address[31:0]																															
0	Invalid							Dummy Cycle0																															
0	Invalid							Dummy Cycle1																															

J :Jtag access : 1= Jtag access from tcu, 0= regular dma packet from IO

O:Ordered bit : 1=From SIU Inbound Ordered Queue. Needed by SIU.

P:Posted bit : reads are nonposted so this bit should always be 0

E: Error bit : (parity or uncorrectable error)

S:Source : 1=DMU, 0=NIU

T:Tag[15:0] generated by the source to track the transaction.

The response packet will contain this 16-bit tag.

Legal RDD Packet Encodings:

RDD requests will always read a full 64 byte cache line, although L2 does not require the address to be any alignment. NIU and Fire-DMU will always align to a 64 Byte address boundary. Note there must be two dummy data cycles in the read request from SIU to L2 to match the pipeline format for WR8.

L2 does not look at the O, P, S and T fields from SIU but simply pipe them along back to SIU Outbound when L2 generates the response packet.

For this implementation, NIU can issue reads to either the SIU's ordered or bypass queue. However, a high percentage of NIU's DMAs will go into the bypass queue. For Fire-DMU implementation, all their DMA reads are restricted to the ordered queue.

So if the NIU and DMU interfaces are behaving correctly,

the following header restrictions apply for RDD:

- P must be 0
- O must be 1 if Fire-DMU
- PA[5:0] must be all 0s if from Fire-DMU or NIU
- Header cycle bits 26:24 == 3'b001

6.6.2 Outbound from L2

6.6.2.1 RDD Response Packet

L2 drives back a packet with a 64 Byte cacheline payload, critical 32-bits first. Should the address sent on the SIU-L2 interface be not aligned to a 64 Byte boundary, L2 will align the responses to a 4-word (32 bit) boundary. The initial 32-bit doubleword within the 64 Byte line is indicated by Address[5:2]. Data responses will start at the specified address, continuing sequentially to the end of the cache line and then wrap. The Read bit (bit 16) is set to 1. The same tag is returned to SIU. For this example, the requested address had addr[5:0]=101101 (or decimal 45). Note that byte 44 is returned first by L2.

Note that L2 returns the CBA (Critical Byte Address – address[2:0]) from the original RDD request. NIU and Fire-DMU always read on a 64 Byte boundary so this is not an issue and data is therefore always returned at the cacheline boundary and CBA[2:0] should always be 0 when the SIU->L2 read request is legal.

L2 also pipes back to SIU Outbound the fields O, P, S, Tag as sent by SIU Inbound. The field E indicates there was an error. SIU Outbound will pass the error condition downstream to DMU or NIU.

Example FIGURE 6-21 shows PA[5:0]=101101 (decimal 45).

FIGURE 6-21 RDD Response Packet when PA[5:0] is not all zeros

V	*UE, Paritys*	*Packet Data [31:0]*			
	UE 1 0	31 30 29 28 27 26 25 24	21 22 21 20 19 18 17 16	15 14 13 12 11 10 9 8	7 6 5 4 3 2 1 0
1	Invalid	J CtagEcc[5:0]	UE O P E S CBA 1	Tag[15:0]	
	Valid	DataByte0	DataByte1	...	...
	Valid	DataByte4	...	...	...
	Valid	DataByte8	...	...	...
	Valid	DataByte12	...	...	...
	Valid	DataByte16	...	...	...
	Valid	DataByte20	...	...	...
	Valid	DataByte24	...	...	...
	Valid	DataByte28	...	...	...
	Valid	DataByte32	...	...	...
	Valid	DataByte36	...	...	...
	Valid	DataByte40	...	...	...
	Valid	DataByte44	...	...	...
	Valid	DataByte48	...	...	...
	Valid	DataByte52	...	...	...
	Valid	DataByte56	...	...	...
0	Valid	DataByte60	...	...	DataByte63

Legal RDD Response Packet Encodings:

Refer to RDD Packet.

Given those restrictions, if the NIU and DMU interfaces and SIU Inbound are behaving correctly, the following header are expected for RDD responses:

J Jtag access: 1= Jtag access from tcu, 0= regular dma packet from IO

UE: Uncorrectable error generated by L2, internal L2 error

Ctag Ecc: check bit for Tag[15:0]

P must be 1

O must be 1 if Fire-DMU

CBA[2:0] must be all 0s if from Fire-DMU or NIU

S must be 1 if from Fire-DMU. Must be 0 if from NIU

Header cycle bit 16 must be 1

Tag[15:0] matches Tag sent by NIU/Fire-DMU

6.6.2.2 Write Invalidate Response Packet

FIGURE 6-22 WRI Response Packet

V	*UE, Paritys*	*Packet Data [31:0]*
	UE 1 0	31 30 29 28 27 26 25 24 21 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
1	Invalid	J CtagEcc[5:0] UE O P E S 0 0 0 0 Tag[15:0]
0	Invalid	Dummy Cycle0
0	Invalid	Dummy Cycle1
0	Invalid	Dummy Cycle2
0	Invalid	Dummy Cycle3
0	Invalid	Dummy Cycle4
0	Invalid	Dummy Cycle5
0	Invalid	Dummy Cycle6
0	Invalid	Dummy Cycle7

A write invalidate response packet looks like a read response packet filled with eight cycles of dummy data. Because of the L2 pipeline, after a write response, the next response (either read or write) cannot come until after eight pad cycles. SIU Outbound will drop all write invalidate responses with the 'P'osted bit set.

L2 pipes back to SIU Outbound the fields O, P, S tag as sent by SIU Inbound. The field E indicates there was an error. SIU Outbound will pass the error condition downstream to DMU or NIU if P bit was a 0.

Legal WRI Response Packet Encodings:

Refer to WRI Packet.

Given those restrictions, if the NIU and DMU interfaces and SIU Inbound are behaving correctly, the following header bits are expected for WRI responses:

UE: Uncorrectable error generated by L2, internal L2 error

Ctag Ecc: check bit for Tag[15:0]

P must be 1 if from Fire-DMU.

O must be 1 if Fire-DMU

S must be 1 if from Fire-DMU. Must be 0 if from NIU

Header cycle bit 19:16 must be 0s

Tag[15:0] matches Tag sent by NIU/Fire-DMU

6.6.2.3 Write8 Response Packet

FIGURE 6-23 WR8 Response Packet

V	UE, Paritys	Packet Data [31:0]																															
						UE	1	0	31	30	29	28	27	26	25	24	21	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	Invalid	J	Ctagecc[5:0]						UE	O	P	E	S	0	0	0	0	Tag[15:0]																							
0	Invalid	Dummy Cycle0																																					
0	Invalid	Dummy Cycle1																																					
0	Invalid	Dummy Cycle2																																					
0	Invalid	Dummy Cycle3																																					
0	Invalid	Dummy Cycle4																																					
0	Invalid	Dummy Cycle5																																					
0	Invalid	Dummy Cycle6																																					
0	Invalid	Dummy Cycle7																																					

A write8 response packet looks like a read response packet filled with eight cycles of dummy data. Because of the L2 pipeline, after a write response, the next response (either read or write) cannot come until after eight pad cycles. SIU Outbound will drop all write8 responses with the 'P'osted bit set. Because the SIU to L2 WR8 request packet header does not provide enough bits for a full 16 bit Tag field and because it is guaranteed that all writes with bytemasks from Fire-DMU are posted and because NIU does not do any writes with bytemasks, SIU outbound's implementation drops all WR8 responses. However, SIU Outbound does decode the header fields so SIU Inbound can do buffer management and therefore expects L2 to pipe back to SIU Outbound fields O, P, S as sent by SIU Inbound. The field E indicates there was an error. Currently, there is no mechanism to log an WR8 response packet with an error.

Legal WR8 Response Packet Encodings:

Refer to WR8 Packet.

Given those restrictions, if the NIU and DMU interfaces and SIU Inbound are behaving correctly, the following header bits are expected for WR8 responses:

J Jtag access: 1= Jtag access from tcu, 0= regular dma packet from IO

UE: Uncorrectable error generated by L2, internal L2 error

Ctag Ecc: check bit for Tag[15:0]

P must be 1.

O must be 1 if Fire-DMU

S must be 1.

Header cycle bit 19:16 must be 0s

6.6.2.4 DMA Read Request Packet from NIU to SIU

All read requests by NIU will return 64 Bytes aligned to 64B boundary. The one cycle packet on niu_sii_data[127:0] contains only the header. The header format is shown below with the SUPPORTED settings needed for Read Requests by NIU. The 16 bit ID is the captured into the tag field sent up to L2 by SIU and later returned back to NIU. There are no payload cycles for the request packet. During the packet transfer,

NIU indicates whether the DMA Read packet will go into an ordered queue or a bypass queue. The reads do not fill in L2. A high percentage of DMA Reads go to the bypass queue. Rads are by default non-posted.

TABLE 6-4 NIU to SIU: DMA Read Request Header Format

NIU's Header Cycle niu_sii_data[msb:lsb] for a Read	Name	Usage
127:122	Command	
	127=Response bit	Must be 0
	126=Posted request bit	Must be 0
	125=Read bit	Must be 1
	124=Write ByteMask Active	Must be 0
	123=L2 bit	Must be 1
	122=NCU bit	Must be 0
121:85	Reserved	Must Be Zero
84:83	Address parity	Bit 84 for odd address bit Bit 83 for even address it
82:82	TimeOutError	1=This packet had Timed Out
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address
80:80	UncorrectableError	Must be 0 because read request does not carry data payload
79:64	ID[15:0]	NIU supplies an ID that it can use to track the responses later
63	Reserved	Must be Zero
62	Command Parity	Parity bit for bit127-122
61:56	Ctag Ecc	6bit sec-dec check bit for ID
55-40	Reserve	Must be Zero
39:0	PA[39:0]	Must be 64B aligned - PA[5:0] must be zeros

6.6.2.5 DMA Write Request Packet from NIU to SIU

NIU always sends a full cacheline of data for writes. The header format is shown below with the SUPPORTED settings needed for Write Requests by NIU. The 16 bit ID is the captured into the tag field sent up to L2 by SIU. The address is always 64 Byte aligned. All DMA Write Request from NIU can be either nonposted (requires a write response from SIU to indicate completion) or posted (no response from SIU).

Prior to the packet transfer, NIU indicates whether the packet will go into an ordered queue or a 'bypass' queue. The writes do not allocate in L2. Writes to the ordered queue will always be issued by the SIU to L2 after the youngest write in the bypass queue. The writes in the bypass queues are not ordered with respect to other writes in the bypass queue. NIU will never use byte-masks so if niu_sii_be[15:0] exists at the interface, NIU always drives them to all 1's.

TABLE 6-5 NIU to SIU: Write Request Packet Format

Packet Cycle Number	Packet Content of niu_sii_data[127:0]
1	NIU's Write Header
2	Byte0, Byte1,..., Byte15
3	Byte16, Byte17, ..., Byte31
4	Byte32, Byte33, ..., Byte47
5	Byte48, Byte49, ..., Byte63

The header encoding for a DMA Write from NIU is shown in TABLE 6-6.

TABLE 6-6 NIU to SIU: DMA Write Request Header Format

NIU's Header Cycle niu_sii_data[msb:lsb] for a Write	Name	Usage
127:122	Command	
	127=Response bit	Must be 0
	126=Posted request bit	0=NIU needs an ack
	125=Read bit	Must be 0
	124=Write ByteMask Active	Must be 0 (all bytes on)
	123=L2 bit	Must be 1
	122=NCU bit	Must be 0
121:85	Reserved	Must Be Zero
84:83	Address parity	Odd, even address parity
82:82	TimeOutError	1=This packet had Timed Out
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address
80:80	UncorrectableError	1=data payload has uncorrectable error
79:64	ID[15:0]	NIU supplies an ID that it can use to track the responses later
62	Command parity	Command parity[127:122]

TABLE 6-6 NIU to SIU: DMA Write Request Header Format *(Continued)*

NIU's Header Cycle niu_sii_data[msb:lsb] for a Write	Name	Usage
61:56	CtagEcc[5:0]	6-bit ECC check bit for ID
55:40	Reserved	Must be Zero
39:0	PA[39:0]	Must be 64B aligned - PA[5:0] must be zeros

6.6.3 Outbound to NIU

6.6.3.1 DMA Write Response Packet from SIU to NIU

If NIU had requested a response (write-ack), SIU will return a packet to indicate the completion of the write. The write responses can return out of order.

The write response is a one cycle packet containing only a header. Immediately following a write response can be another write or read response.

TABLE 6-7 SIU to NIU: DMA Write Response Header Format

SIO's Header Cycle sio_niu_data[msb:lsb] for Write Response	Name	Usage
127:122	Command	
	127=Response bit	Must be 1
	126=Posted request bit	Must be Ignored (driven to 0)
	125=Read bit	Must be 0
	124=Write ByteMask Active	Must be Ignored (driven to 0)
	123=L2 bit	Must be Ignored (driven to 1)
	122=NCU bit	Must be Ignored (driven to 0)
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	Reserved	Must Be Zero
81:81	UncorrectableError for prior address, Ctag Ecc	1=This packet's address mapped to an nonexistent, reserved, or erroneous address

TABLE 6-7 SIU to NIU: DMA Write Response Header Format *(Continued)*

SIO's Header Cycle sio_niu_data[msb:lsb] for Write Response	Name	Usage
80:80	Data Error	1=data payload has a detected uncorrectable error. This could be: 1. timeout errors 2. unmapped errors 3. data ue error from L2$ or dram
79:64	ID[15:0]	ID supplied originally by NIU
63:40	Reserved	Must be Zero
39:0	PA[39:0]	Must be Ignored

6.6.3.2 DMA Read Response Packet from SIU to NIU

TABLE 6-8 SIU to NIU: DMA Read Response Packet Format

Packet Cycle Number	Packet Content of sio_niu_data[127:0]
1	SIU to NIU DMA Read Response Header
2	Byte0, Byte1, ..., Byte15
3	Byte16, Byte17, ..., Byte31
4	Byte32, Byte33, ..., Byte47
5	Byte48, Byte49, ..., Byte63

The read response packet from SIU to NIU is one cycle of header followed by four cycles of data payload. The above data order assumes that NIU had set the lower six bits of the PA to zero. Otherwise L2 and SIU will return data critical 32-bit doubleword first and wrap around the 64 Byte boundary.

The read response header encoding is defined in TABLE 6-9:

TABLE 6-9 SIU to NIU Read Response Header Format

SIO's Header Cycle sio_niu_data[msb:lsb] for Read Response	Name	Usage
127:122	Command	
	127=Response bit	Must be 1
	126=Posted request bit	Must be Ignored (driven to 0)
	125=Read bit	Must be 1
	124=Write ByteMask Active	Must be Ignored (driven to 0)
	123=L2 bit	Must be Ignored (driven to 1)
	122=NCU bit	Must be Ignored (driven to 0)
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	Reserved	1=This packet had Timed Out
81:81	UncorrectableError for prior address, Ctag Ecc	1=This packet's address and Ctag Ecc err
80:80	DE data error	1=data payload has a detected uncorrectable error. This could be: 1. timeout errors 2. unmapped errors 3. data ue error from L2$ or dram
79:64	ID[15:0]	ID supplied originally by NIU
63:40	Reserved	Must be Zero
39:0	PA[39:0]	Must be Ignored

6.6.4 Inbound from DMU.

6.6.4.1 Packet from =Fire-DMU to SIU.

There are four expected/supported types of packet transfers from Fire-DMU to SIU

DMA Read Request: A packet with only a header cycle and no payload cycles. Addresses must be 64-Byte aligned. Must be steered into SIU's Inbound Ordered Queue.

Interrupt Write Request: A packet with a header cycle and fixed size of one payload cycle with all 16 bytes in the payload valid. Must be steered into SIU's Inbound Ordered Queue.

DMA Write Request: A packet with a header cycle and fixed size of four payload cycles. Addresses must be 64-Byte aligned and always posted. Must be steered into SIU's Inbound Ordered Queue.

PIO Read Data Return: A packet with a header cycle and fixed size of one payload cycle. SIU and NCU will transport the full 16 bytes. Only the cpu cares which byte(s) within the 16 bytes are enabled. Must be steered into SIU's Inbound Bypass Queue.

TABLE 6-10 for dmc_tag[15:0] is referred to by all packets from Fire-DMU. See *OpenSPARC T2 SoC Microarchitecture Specification, Part 2 of 2* for DMU information.

TABLE 6-10 Fire-DMC Tag

Field	Bits	Description
	DMA transactions	
dmc_tag[15]	type	0b-indicates DMA/Int transactions
dmc_tag[14:11]	cl_tag[3:0]	Dmc transaction number for tracking credits
dmc_tag[10:6]	d_ptr[4:0]	Used for DMA Rds only-dou dma rd buffer address
dmc_tag[5:1]	pkt_tag[4:0]	Used for DMA Rds only-PSB index for building packet records
dmc_tag[0]	cl_sts	Used for DMA Rds only-indicates 1st cacheline in packet sequence
Int Transactions		
dmc_tag[15]	type	0b-indicates DMA/Int transactions
dmc_tag[14:11]	cl_tag[3:0]	Dmc transaction number for tracking credits
dmc_tag[10:3]	Rsv[7:0]	reserved
dmc_tag[2:1]	mdo_tag[1:0]	mondo_tag for mondo-reply to IMU
dmc_tag[0]	rsv	Must be 0
MMU Tablewalk Transactions		
dmc_tag[15]	type	1b-indicates MMU Tablewalk transactions
dmc_tag[14:11]	cl_tag[3:0]	Dmc transaction number for tracking credits
dmc_tag[10:6]	Rsv[4:0]	reserved
dmc_tag[5:0]	Mtag[5:0]	Used for MMU tablewalks only-MMU tag for tracking tablewalks

TABLE 6-10 Fire-DMC Tag *(Continued)*

Field	Bits	Description
PIO Cpl Transactions		
dmc_tag[15:13]	Rsv[2:0]	Must be 3'b100
dmc_tag[12:9]	jbc_trans_#[3:0]	Pio transaction credit id
dmc_tag[8:0]	thread_id[8:0]	Thread id of PIO read request

Note – The NCU will distinguish interrupts from PIO cpl's by using dmc_tag[15].

DMA Read Request packet from Fire-DMU to SIU

All read requests by Fire-DMU will return 64 Bytes aligned to 64B boundary. The one cycle packet on dmu_sii_data[127:0] contains only the header. The header format is shown below with the SUPPORTED settings needed for Read Requests by Fire-DMU. The 16 bit ID is the captured into the tag field sent up to L2 by SIU and later returned back to Fire-DMU. There are no payload cycles for the request packet. During the packet transfer, Fire-DMU must always steer DMA Reads into the ordered queue. The reads do not fill in L2. Reads are by default non-posted.

TABLE 6-11 Fire-DMU to SIU: DMA Read Request Header Format

Fire-DMU's Header Cycle dmu_sii_data[msb:lsb] for a Read	Name	Usage
127:122	Command	
	127=Response bit	Must be 0
	126=Posted request bit	Must be 0
	125=Read bit	Must be 1
	124=Write ByteMask Active	Must be 0 Must be 1
	123=L2 bit	Must be 0
	122=NCU bit	
121:85	Reserved	Must Be Zero
84:83	Address parity	Bit 84 for odd address bit Bit 83 for even address it
82:82	TimeOutError	1=This packet had Timed Out, used only by Fire-DMU for PIO Rd Completions

TABLE 6-11 Fire-DMU to SIU: DMA Read Request Header Format *(Continued)*

Fire-DMU's Header Cycle dmu_sii_data[msb:lsb] for a Read	Name	Usage
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address used only by Fire-DMU for PIO Rd Completions
80:80	UncorrectableError	used only by Fire-DMU for PIO Rd Completions
79:64	ID[15:0] (dmc_tag[15:0])	DMU supplies an ID that it can use to track the responses later. See dmc_tag TABLE 6-10
63	Reserved	Must be Zero
62	Command Parity	Parity bit for bit127-122
61:56	Ctag Ecc	6bit sec-dec check bit for ID
55-40	Reserve	Must be Zero
39:0	PA[39:0]	Must be 64B aligned - PA[5:0] must be zeros

DMA Write Request Packet from Fire-DMU to SIU

Fire-DMU does not always send a full cacheline of data for writes. The header format is shown below with the SUPPORTED settings needed for Write Requests by Fire-DMU. The 16 bit ID is the captured into the tag field sent up to L2 by SIU. The address is always 64 Byte aligned. All DMA Write Request from Fire-DMU are posted (no response from SIU). Fire-DMU must always steer DMA Writes into the SIU's ordered queue. The writes do not allocate in L2. Fire-DMU can send DMA writes with all bytes enabled and with one or more bytes at the beginning and/or the end of the 64Bytes disabled.

TABLE 6-12 Fire-DMU to SIU Write Request Packet Format

Packet Cycle Number	Packet Content of dmu_sii_data[127:0]
1	Fire-DMU's Write Header
2	Byte0, Byte1, ..., Byte15
3	Byte16, Byte17, ..., Byte31
4	Byte32, Byte33, ..., Byte47
5	Byte48, Byte49, ..., Byte63

The header encoding for a DMA Write from Fire-DMU is shown in TABLE 6-13.

TABLE 6-13 Fire-DMU to SIU: DMA Write Request Header Format

Fire-DMU's Header Cycle dmu_sii_data[msb:lsb] for a Write	Name	Usage
127:122	Command	
	127=Response bit	Must be 0
	126=Posted request bit	Must be 1
	125=Read bit	Must be 0
	124=Write ByteMask Active	0=Write full cacheline. 1=WRM Must be 1
	123=L2 bit	Must be 0
	122=NCU bit	
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	TimeOutError	1=This packet had Timed Out, used only by Fire-DMU for PIO Rd Completions
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address used only by Fire-DMU for PIO Rd Completions
80:80	UncorrectableError	1=data payload has uncorrectable error used only by Fire-DMU for PIO Rd Completions
79:64	ID[15:0] (dmc_tag[15:0])	Fire-DMU supplies an ID that it can use to track the responses later. See dmc_tag TABLE 6-10.
63	Reserved	Must be Zero
62	Command Parity	Parity bit for bit127-122
61:56	Ctag Ecc	6bit sec-dec check bit for ID
55-40	Reserve	Must be Zero
39:0	PA[39:0]	Must be 64B aligned - PA[5:0] must be zeros

Interrupt Write Request packet from Fire-DMU to SIU

Fire-DMU can send an interrupt to NCU via SIU. The interrupt is always a mondo type with 16 bytes of payload. NCU decodes the ID field to determine how to process the mondo and to differentiate it from a PIO Completion. SIU transports the full 16 bit ID to NCU. Interrupts must be steered toward the ordered queue.

TABLE 6-14 Fire-DMU to SIU: Interrupt Write Request Packet Format

Packet Cycle Number	Packet Content of dmu_sii_data[127:0]
1	Fire-DMU's Interrupt Write Header
2	Mondo Byte0, Byte1, ..., Byte15

The header encoding for an Interrupt Write from Fire-DMU is shown in TABLE 6-15.

TABLE 6-15 Fire-DMU to SIU: Interrupt Write Request Header Format

Fire-DMU's Header Cycle dmu_sii_data[msb:lsb] for an Interrupt	Name	Usage
127:122	Command	
	127=Response bit	Must be 0
	126=Posted request bit	Must be 0
	125=Read bit	Must be 0
	124=Write ByteMask Active	Must be 0
	123=L2 bit	Must be 0
	122=NCU bit	Must be 1
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	TimeOutError	1=This packet had Timed Out used only by Fire-DMU for PIO Rd Completions
81:81	UnmappedAddressError	Must be 0 used only by Fire-DMU for PIO Rd Completions
80:80	UncorrectableError	Must be 0 used only by Fire-DMU for PIO Rd Completions
79:64	ID[15:0]	See NCU spec or see dmc_tag TABLE 6-10 for Interrupt ID encoding
63:40	Reserved	Must be Zero
39:0	PA[39:0]	Must be Ignored

PIO Read Completion Packet from Fire-DMU to SIU

Fire-DMU can send a PIO Read completion packet to NCU via SIU. The PIO Read completion has a one cycle payload. SIU transports the full 16 bit ID and 16 byte payload to NCU. PIO Read completions from Fire-DMU must be steered toward the bypass queue. Only cpu cares which byte(s) within the 16 bytes are enabled.

TABLE 6-16 Fire-DMU to SIU: PIO Read Completion Response Packet Format

Packet Cycle Number	Packet Content of dmu_sii_data[127:0]
1	Fire-DMU's PIO Read Completion Header
2	Byte0, Byte1, Byte2, Byte3, Byte4, Byte5, Byte6, Byte7, Byte8, Byte9, Byte10, Byte11, Byte12, Byte13, Byte14, Byte15

The header encoding for a PIO Read Completion from Fire-DMU is shown in TABLE 6-17.

TABLE 6-17 Fire-DMU to SIU: PIO Read Completion Packet Header Format

Fire-DMU's Header Cycle dmu_siu_data[msb:lsb] for PIO completions	Name	Usage
127:122	127=Response bit	Must be 1
	126=Posted request bit	Must be 0 (Ignored by SIU if Response bit is set)
	125=Read bit	Must be 1
	124=Write ByteMask Active	Must be 0 (Ignored by SIU if Response bit is set or if Read bit is set)
	123=L2 bit	Must be 0
	122=NCU bit	Must be 1
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	TimeOutError	1=This packet had Timed Out
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address
80:80	UncorrectableError	1=data payload has a detected uncorrectable error

TABLE 6-17 Fire-DMU to SIU: PIO Read Completion Packet Header Format *(Continued)*

Fire-DMU's Header Cycle dmu_siu_data[msb:lsb] for PIO completions	Name	Usage
79:64	ID[15:0]	for PIO read completions this is PIOID [15:13]=3'b100 [12:9] will be the credit id returned on PIO rd completions, [8:0] will be the {3'b000, cpu-thread ID[5:0]}. See NCU or DSN specification.
62	Command parity	Command parity for [127:122]
63:40	Reserved	Must be Zero
39:0	PA[39:0]	Must be Ignored.

6.6.5 Outbound to DMU.

6.6.5.1 Packet from SIU to Fire-DMU.

There is one type of packet transfers from SIU to Fire-DMU.

DMA Read Response: A packet with a header cycle followed by a fixed size of four payload cycles with all bytes valid. Because Fire-DMU's DMA Read Request are always cacheline aligned, the data returned always starts at the cacheline boundary. Data format is big endian.

TABLE 6-18 SIU to Fire-DMU: DMA Read Response Packet Format

Packet Cycle Number	Packet Content of sio_dmu_data[127:0]
1	SIU to Fire-DMU's DMA Read Response Header
2	Byte0, Byte1, ..., Byte15
3	Byte16, Byte17, ..., Byte31
4	Byte32, Byte33, ..., Byte47
5	Byte48, Byte49, ..., Byte63

The header format is shown in TABLE 6-19.

TABLE 6-19 SIU to Fire-DMU: Outbound Packet Header Format

SIO to Fire-DMU's Header Cycle sio_dmu_data[msb:lsb]	Name	Usage
127:122	Command Legal combinations - DMA Read Response	 1010_10
	127=Response bit	Must be 1
	126=Posted request bit	Must be Ignored (driven to 0)
	125=Read bit	Must be 1
	124=Write ByteMask Active	Must be Ignored (driven to 0)
	123=L2 bit	Must be 1
	122=NCU bit	Must be 0
121:84	Reserved	Must Be Zero
83:83	Reserved	Must Be Zero
82:82	TimeOutError	1=This packet had Timed Out
81:81	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address
80:80	UncorrectableError	1=response has a detected uncorrectable error. This could be: 1. timeout errors 2. unmapped errors 3. data ue error from L2$ or dram
79:64	ID[15:0] (dmc_tag[15:0])	For Response, this is DMU's ID
63:62	Reserved	Must be Zero
61:56	CtagEcc[5:0]	6-bit ECC check bit for ID
55:40	Reserved	Must be Zero
39:0	Bus Address[39:0]	For Responses, SIU does not return the Address.

6.6.6 Inbound to NCU

6.6.6.1 Packet from SIU to NCU

There is one cycle of header followed by four cycles of payload data. The format of the header is shown in TABLE 6-20.

TABLE 6-20 SIU to NCU: Inbound Packet Header Format

Header Cycle sii_ncu_data[msb:lsb]	Name	Usage
31:31	TimeOutError	1=This packet had Timed Out
30:30	UnmappedAddressError	1=This packet's address mapped to an nonexistent, reserved, or erroneous address
29:29	UncorrectableError	1=packet has an uncorrectable error
28	SIU Ctag Uncorrectable Error	1= ctag uncorrectable error
27-22	Reserve	Must be Zero
21:16	ctag	ECC check bit for ID [15:0]
15:0	ID[15:0] as originally sent by DMU. dmc_tag[15:0]	If Interrupt is from DMU, NCU returns the entire tag back to DMU with mondo_ack or mondo_nack signal asserted. If PIO Completion is from DMU, dmc_tag[12:9] = NCU credit id will be returned to the credit pool dmc_tag[8:0] = {3'b000, cpu-thread id[5:0]}

```
DMU sending to SIU as the following [127:0]
Byte0     Byte1...          Byte15
Byte16    Byte17...     Byte31
Byte 32   Byte33...     Byte47
Byte 48   Byte49...     Byte63

SIU send to L2 is  [31:0]
Byte0 Byte1 Byte2 Byte3                   cycle1
Byte 4 Byte5 Byte6 Byte7                  cycle2
:
:
Byte 60 Byte61 Byte62  Byte63             cycle16
```

```
SIU send to NCU will  be [31:0]
Byte0 Byte1 Byte2 Byte3                          cycle1
Byte4 byte5 Byte6  Byte7                         cycle2
:
Byte12 Byte13 Byte14 Byte15                      cycle4
```

6.7 CSR

SIU has no CSRs. Data parity errors and uncorrectable errors detected are signaled in the packet header to the receiving unit. It is assumed that the end unit will log the error.

The following is a summary of what the CSR tool looks like and may be a candidate for connecting/generating CSRs within OpenSPARC T2. If debug control status registers are defined in the future, then SIU will participate with whatever methodology is defined for accessing CSRs.

Control Status Register (CSR)

All the control and status registers inside the OpenSPARC T2 are generated by the CSR Tool, so that we can reduce some of the coding effort and standardize the register access among different modules within OpenSPARC T2. All the modules with CSR registers will be connected with CSR specific interface in a ring fashion.

CSR interface signals:

```
Inputs:

clk           - clock signal for the design

rst           - reset signal for the flops

csrbus_data_in - data to writeto CSR

csrbus_addr   - address to send to select one CSR

csrbus_src_bus - source bus identification

csrbus_valid  - specifies that address and data lines are valid
```

```
Outputs:

csrbus_mapped    - asserted when CSR within module is selected

csrbus_acc_vio   - improper access is attempted

csrbus_done      - transaction is completed

csrbus_data_out  - data read from the CSR
```

CSR Read/Write Access:

For read from CSR, csrbus_valid is asserted along with address and source bus information, data come back on csrbus_data_out when csrbus_done is asserted.

For write operation, csrbus_valid is asserted along with address and data info, then the carbus_done is asserted shortly after within the same cycle.

6.8 Unit Level Signals

6.8.1 SIU-L2 Interface List

TABLE 6-21 SIU-L2 Interface List

Signal Name	I/O	Size	From/To	Description
SII to L2Tag signals				
sii_l2t0_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank0
sii_l2t1_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank1
sii_l2t2_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank2
sii_l2t3_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank3
sii_l2t4_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank4
sii_l2t5_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank5
sii_l2t6_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank6
sii_l2t7_req_vld	O	1	SIU->L2T	Packet request valid (first cycle) to L2Tag for bank7
sii_l2t0_req	O	32	SIU->L2T	Packet header/data for bank0
sii_l2t1_req	O	32	SIU->L2T	Packet header/data for bank1
sii_l2t2_req	O	32	SIU->L2T	Packet header/data for bank2
sii_l2t3_req	O	32	SIU->L2T	Packet header/data for bank3
sii_l2t4_req	O	32	SIU->L2T	Packet header/data for bank4
sii_l2t5_req	O	32	SIU->L2T	Packet header/data for bank5
sii_l2t6_req	O	32	SIU->L2T	Packet header/data for bank6
sii_l2t7_req	O	32	SIU->L2T	Packet header/data for bank7
SII to L2Buffer signals				
sii_l2b0_ecc	O	7	SIU->L2T	Packet ECC for bank0
sii_l2b1_ecc	O	7	SIU->L2B	Packet ECC for bank1
sii_l2b2_ecc	O	7	SIU->L2B	Packet ECC for bank2
sii_l2b3_ecc	O	7	SIU->L2B	Packet ECC for bank3
sii_l2b4_ecc	O	7	SIU->L2B	Packet ECC for bank4
sii_l2b5_ecc	O	7	SIU->L2B	Packet ECC for bank5

TABLE 6-21 SIU-L2 Interface List *(Continued)*

Signal Name	I/O	Size	From/To	Description
sii_l2b6_ecc	O	7	SIU->L2B	Packet ECC for bank6
sii_l2b7_ecc	O	7	SIU->L2B	Packet ECC for bank7
L2Tag to SII signals				
l2t0_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank0 request
l2t1_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank1 request
l2t2_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank2 request
l2t3_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank3 request
l2t4_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank4 request
l2t5_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank5 request
l2t6_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank6 request
l2t7_sii_iq_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank7 request
l2t0_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank0 write invalidate data buffer
l2t1_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank1 write invalidate data buffer
l2t2_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank2 write invalidate data buffer
l2t3_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank3 write invalidate data buffer
l2t4_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank4 write invalidate data buffer
l2t5_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank5 write invalidate data buffer
l2t6_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank6 write invalidate data buffer
l2t7_sii_wib_dequeue	I	1	L2T->SIU	L2Tag is unloading a bank7 write invalidate data buffer
L2Buffer to SIO signals				
l2b0_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 0
l2b1_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 1
l2b2_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 2
l2b3_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 3
l2b4_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 4
l2b5_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 5
l2b6_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 6
l2b7_sio_ctag_vld	I	1	L2B->SIU	Response packet Valid (First Cycle) from L2 bank 7
l2b0_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank0
l2b1_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank1

TABLE 6-21 SIU-L2 Interface List *(Continued)*

Signal Name	I/O	Size	From/To	Description
l2b2_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank2
l2b3_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank3
l2b4_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank4
l2b5_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank5
l2b6_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank6
l2b7_sio_data	I	32	L2B->SIU	Read data/write response packet from L2 bank7
l2b0_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank0
l2b1_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank1
l2b2_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank2
l2b3_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank3
l2b4_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank4
l2b5_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank5
l2b6_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank6
l2b7_sio_ue_err	I	1	L2B->SIU	UE on Read data from L2 bank7
l2b0_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank0
l2b1_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank1
l2b2_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank2
l2b3_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank3
l2b4_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank4
l2b5_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank5
l2b6_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank6
l2b7_sio_parity[1:0]	I	2	L2B->SIU	Parity on Read data from L2 bank7

6.8.2 SIU-NCU Interface List

TABLE 6-22 SIU-NCU Interface List

Signal Name	I/O	Size	From/To	Description
NCU to SII				
ncu_sii_gnt	I	1	NCU->SIU	Grant – xfr packet to NCU starting next cycle
ncu_sii_dmuctag_uei	I	1	NCU->SIU	Inject uncorrectable error for ctag
ncu_sii_dmuctag_cei	I	1	NCU->SIU	Inject correctable error for ctag
ncu_sii_dmua_pei	I	1	NCU->SIU	Inject address parity error
ncu_sii_dmud_pei	I	1	NCU->SIU	Inject Data parity error
ncu_sii_niuctag_uei	I	1	NCU->SIU	Inject uncorrectable error for ctag
ncu_sii_niuctag_cei	I	1	NCU->SIU	Inject correctable error for ctag
ncu_sii_niua_pei	I	1	NCU->SIU	Inject address parity error
ncu_sii_niud_pei	I	1	NCU->SIU	Inject data parity error
SII to NCU				
sii_ncu_req	O	1	SIU->NCU	Packet request from SIU to NCU
sii_ncu_data	O	32	SIU->NCU	Packet header/data from SIU to NCU
sii_ncu_parity[1:0]	O	2	SIU->NCU	Parity on data from SIU to NCU
sii_ncu_dmuctag_ue	O	1	SIU->NCU	Uncorrectable error for ctag
sii_ncu_dmuctag_ce	O	1	SIU->NCU	Correctable error for ctag
sii_ncu_dmua_pe	O	1	SIU->NCU	Address parity error
sii_ncu_dmud_pe	O	1	SIU->NCU	Data parity error
sii_ncu_niuctag_ue	O	1	SIU->NCU	Uncorrectable error for ctag
sii_ncu_niuctag_ce	O	1	SIU->NCU	Correctable error for ctag
sii_ncu_niua_pe	O	1	SIU->NCU	Address parity error
sii_ncu_niud_pe	O	1	SIU->NCU	Data parity error
sii_ncu_syn_vld	O	1	SIU->NCU	Syndrome valid signal
sii_ncu_syn_data[3:0]	O	4	SIU->NCU	Syndrome bus total 16 cycle xfr syndrome
SIO to NCU				
sio_ncu_ctag_ue	O	1	SIU->NCU	Uncorrectable error for ctag
sio_ncu_ctag_ce	O	1	SIU->NCU	Correctable error for ctag
sio_ncu_data_parity	O	1	SIU->NCU	Data parity error

TABLE 6-22 SIU-NCU Interface List *(Continued)*

Signal Name	I/O	Size	From/To	Description
Partial L2 Bank Mode bits				
ncu_sii_pm	I	1	NCU->SIU	0=all 8 banks available 1=partial mode and need to look at each ncu_sii_ba* signals
ncu_sii_ba01	I		NCU->SIU	0=bank0 and bank1 unavailable 1=both banks available
ncu_sii_ba23	I		NCU->SIU	0=bank2 and bank3 unavailable 1=both banks available
ncu_sii_ba45	I		NCU->SIU	0=bank4 and bank5 unavailable 1=both banks available
ncu_sii_ba67	I		NCU->SIU	0=bank6 and bank7 unavailable 1=both banks available
L2 Index Hashing Enable bit				
ncu_sii_l2_idx_hash_en	I		NCU->SIU	1=enable hashing of PA for L2 index.

6.8.3 SIU-NIU Interface List

TABLE 6-23 SIU-NIU Interface List

Signal Name	I/O	Size	From/To	Description
NIU to SII signals				
niu_sii_hdr_vld	I	1	NIU->SIU	Asserted during the header phase of any requests from NIU to SIU. Not asserted during the data transfer phase.
niu_sii_reqbypass	I	1	NIU->SIU	Valid during the header phase only. 0: Current request is for the bypass queue 1: Current request is for the ordered queue
niu_sii_datareq	I	1	NIU->SIU	Valid during the header phase only. 0: Current request is a read, with no payload; 1: Current request is a write, with one or four cycles of data payload
niu_sii_datareq16	I	1	NIU->SIU	Valid during the header phase only. Don't care if niu_sii_datareq is 0. Otherwise should always be 0 for the supported modes expected from NIU: Current write request has 64B data payload;
niu_sii_data[127:0]	I	128	NIU->SIU	Packet header/data for L2. (Big-endian)
niu_sii_parity[7:0]	I	4	NIU->SIU	Parity of data payload cycles (127:0)
SII to NIU signals				
sii_niu_oqdq	O	1	SIU->NIU	Transaction credit for the ordered queue
sii_niu_bqdq	O	1	SIU->NIU	Transaction credit for the ordered queue
SIO to NIU signals				
sio_niu_hdr_vld	O	1	SIU->NIU	Envelops the header of any requests from SIU to NIU. Not asserted during the data transfer phase. NIU determines from the header if and how much data will follow.
sio_niu_data[127:0]	O	128	SIU->NIU	Packet header/data for NIU
sio_niu_parity[7:0]	O	4	SIU->NIU	Parity of payload cycles (127:0)

6.8.4 SIU-DMU Interface List

TABLE 6-24 SIU-DMU Interface List

Signal Name	I/O	Size	From/To	Description
DMU to SII signals				
dmu_sii_hdr_vld	I	1	DMU->SIU	Asserted during the header phase of any requests from DMU to SIU. Not asserted during the data transfer phase.
dmu_sii_reqbypass	I	1	DMU->SIU	Valid during the header phase only. Asserted for PIO rd cpl's
dmu_sii_datareq	I	1	DMU->SIU	Valid during the header phase only. 0: Current request is a read, with no payload; 1: Current request is a write, with one or four cycles of data payload
dmu_sii_datareq16	I	1	DMU->SIU	Valid during the header phase only. Don't care if dmu_sii_datareq is 0. 0: Current write request has 64B data payload; 1: Current write request has 16B data payload. (meant for NCU – int/PIO read data)
dmu_sii_data[127:0]	I	128	DMU->SIU	Packet header/data for L2/NCU. (Big-endian) For PIO read completions, the 64 bit PIO payload will be duplicated on both halves of the 128 bit data bus. Which 64 bits to replicate will be determined by dmu_sii_be[15:0]
dmu_sii_be[15:0]	I	16	DMU->SIU	Packet data byte enables/errors. Only valid during data transfer phase. (Bit position matches Byte position on the data bus.)
dmu_sii_parity[7:0]	I	4	DMU->SIU	Parity of data payload cycles (127:0)
dmu_sii_be_parity	I	1	DMU->SIU	Parity of dmu_sii_be[15:0]
SII to DMU signals				
sii_dmu_wrack_tag[3:0]	O	4	SIU->DMU	j2d_d_wrack_tag[3:0] DSN/DMU name Transaction credit id for dma wrack
sii_dmu_wrack_parity	O	1	SIU->DMU	Parity bit for the sii_dmu_wrack_tag
sii_dmu_wrack_vld	O	1	SIU->DMU	j2d_d_wrack_vld DSN/DMU name Valid signal for j2d_d_wrack_tag

TABLE 6-24 SIU-DMU Interface List *(Continued)*

Signal Name	I/O	Size	From/To	Description
SIO to DMU signals				
sio_dmu_hdr_vld	O	1	SIU->DMU	Envelops the header of any requests from SIU to DMU. Not asserted during the data transfer phase. DMU determines from the header if and how much data will follow.
sio_dmu_data[127:0]	O	128	SIU->DMU	Packet header/data for DMU
sio_dmu_parity[7:0]	O	4	SIU->DMU	Parity of payload cycles (128:0)

6.8.5 SIU-TCU Interface List

TABLE 6-25 SIU-TCU Interface List

Signal Name	I/O	Size	From/To	Description
TCU to SII				
tcu_sii_vld	I	1	TCU->SII	Valid signal to qualify the tcu_sii_data
tcu_sii_data	I	1	TCU->SII	Serial data bus for dma rd/wr request
SIO to TCU				
sio_tcu_vld	O	1	SIO->TCU	Valid signal to qualify the sio_tcu_data
sio_tcu_data	O	1	SIO->TCU	Serial bus for DMA return data/hdr
SII to DBG				
sii_dbg_l2t0_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t1_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t2_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t3_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t4_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t5_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t6_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8
sii_dbg_l2t7_req[1:0]	O	2	SII->DBG	00=no req, 01=rd,10=wr,11=wr8

CHAPTER 7

Non-Cacheable Unit (NCU)

This chapter contains the following sections:

7.1 Overview

The main function of NCU is to communicate between the CMP cores (64 threads total) and the various blocks in the IO subsystem. FIGURE 7-1 shows the connectivity of NCU with various IO subsystem blocks as well as the XBAR, which connects to CMP core on the other side. Traffic on XBAR side runs at CPU clock frequency whereas traffic on IO subsystem side is at IO clock frequency. In general, traffic goes to NCU does not require high performance and can tolerate high latency.

FIGURE 7-1 NCU Connectivity

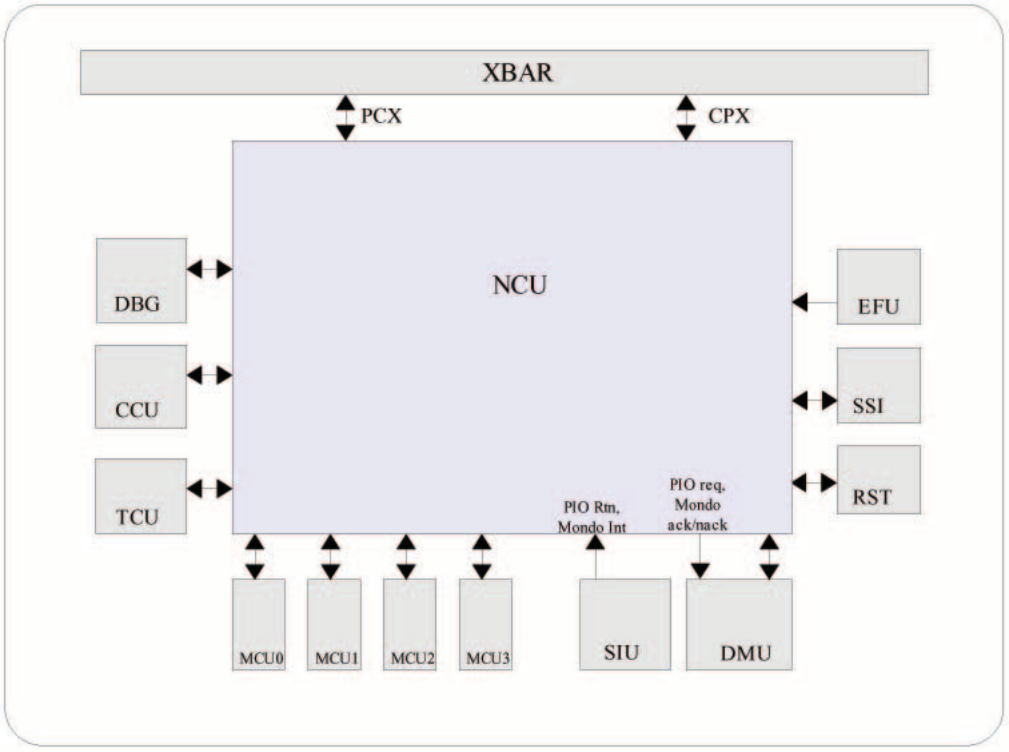


TABLE 7-1 shows a summary of types of traffic and bus size between each unit control block and NCU.

TABLE 7-1 NCU/UCB Communication Type and Bus Size

Unit Control Block (UCB)	Bus Size (downstream/ upstream (protocol))	External req/ack/intr	CSR	On Chip Interrupts	Boot Up
SII	--/32 (SII/NCU Table4)	Mondo intr., PIO rtns	---	---	---
DMU	Mondo resp.: 4/-- PIO: 64/-- (NCU/DMU 5) CSR: 32/32 (UCB Table2,3)	Mondo Interrupt. Resp. PIO rd_req/wr requests ---	 CSR rw	---	---
MCUs	4/4 (UCB 2,3)	---	CSR rw	OCI	---
CCU	4/4 (UCB 2,3)	---	CSR rw	---	---
TCU	8/8 (UCB 2,3)	---	CSR rw	---	---

TABLE 7-1 NCU/UCB Communication Type and Bus Size *(Continued)*

Unit Control Block (UCB)	Bus Size (downstream/ upstream (protocol))		External req/ack/intr	CSR	On Chip Interrupts	Boot Up
SSI (integrated into NCU)	4/4	(UCB 2,3)	---	CSR rw	OCI	Instructions
RST	4/4	(UCB 2,3)	---	CSR rw	---	---
DBG	4/4	(UCB 2,3)	---	CSR rw	---	---

7.1.1 Changes from OpenSPARC T1 IOB

Changes from two MCUs to four MCUs.

CTU changes to CCU + TCU.

JBUS changes to SIU + DMU with different interface format

Adds DMU CSR support.

Adds DMU PIO token ID engine to limit numbers of outstanding PIO to DMU

Adds DMU PIO memory due to OpenSPARC T2 IOMMU changes

Adds support for Mondo Interrupt ID return for DMU

Adds ASI register to comply with SUN's CMP specification

Adds L2 partial bank mode support

Internal memories upsizing to accommodate 64 threads and memory pipelines adjustment

XBAR packet format changes

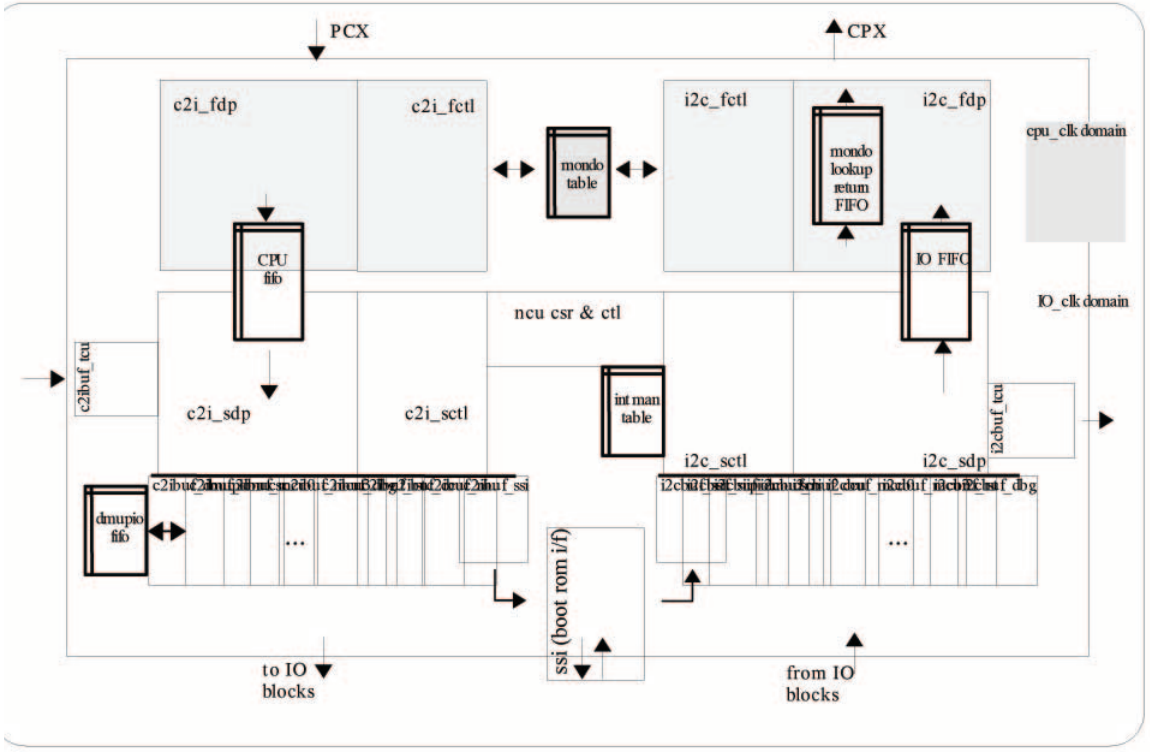
Modifies reset handling to comply with OpenSPARC T2's reset scheme

Integrates SSI (boot ROM i/f logics)

RAS logics

OpenSPARC T2 naming rule compliance.

FIGURE 7-2 NCU Internal Logical Block Diagram



NCU retains most of the internal block name from OpenSPARC T1 IOB since it does not violate the OpenSPARC T2 naming rule:

c2i – cpu to io

i2c – io to cpu

sdp – slow clock (io_clk domain) data path

sctl – slow clock control logic

fdp – fast clock (cpu_clk domain) data path

fctl – fast clock control logic

7.2 Clock Domains

There are two clock domains in NCU: CPU clock domain and IO clock domain.

- CPU Clock Domain: XBAR side communications at CPU clock frequency (targeted for 1.4GHz)
- IO Clock Domain: IO subsystem side communications at IO clock frequency (targeted for 1.4GHz /4 = 350MHz or 1.4GHz/3 = 467MHz)

7.3 Data Flow

Data Flow can be subdivided into the following categories:

Downstream -- packets/information going from XBAR to IO subsystem.

1. CPU non-cacheable external PIO store requests (8B max) /load requests (16B max).
2. CPU external instruction fetch (IFILL) requests.
3. On chip CSR read/write requests (8B only) from CPU to IO subsystem.
4. Upstream – packets/information going from IO subsystem to XBAR.
5. PIO read returns (16B max).
6. External Instruction fetch returns (4B only).
7. CSR read returns (8B only).
8. Mondo Interrupts (with 16B mondo payload).
9. On-chip interrupts (MCU, SSI).

Loopback – packet/information going from XBAR and being sent back to XBAR

1. CPU Mondo Interrupt Table lookup.
2. NCU's internal CSR/ ASI register access.

Undeliverable – NCU adopted the OpenSPARC T1 IOB's packet delivery policy. All writes/STORE_REQs to NCU from core is non-posted which means core requires STORE_ACK to confirm the packets delivery. NCU generates STORE_ACK back to core automatically whenever a STORE_REQ is dequeued successfully from the main downstream FIFO or the DMUPIO fifo. Therefore, core still gets a STORE_ACK for

an undeliverable STORE_REQ, and the actual STORE_REQ packet is discarded. For undeliverable LOAD_REQ/read, NCU generates an CPX Load Return packet with uncorrectable error bit set.

7.3.1 Downstream Path Block Diagrams

FIGURE 7-3 and FIGURE 7-4 show the logical block diagram for downstream communication path, PCX interface is a 130 bit-wide data bus running at 1.5GHz., sourcing packets into NCU. The packet is then decoded by c2i_fdp and c2i_fctl blocks to extract the CPU Mondo interrupt table access from other traffic, which includes on-chip CSR read/write access, CPU non-cacheable external load/store request (PIO read/write), and CPU instruction fetch request. All requests other than Mondo interrupt table access are sent to the CPU command FIFO, which is a 32 deep domain crossing FIFO. The write control is managed by c2i_fctl block in CPU clock domain and the read control is managed by c2i_fctl block in IO clock domain. When a CPU packet is dequeued from the FIFO, c2i_sdp and c2i_sctl blocks will determine which of UCB output buffers to send the packet to. Each of the UCB output buffer contains a double buffer and a working buffer as shown in FIGURE 7-4.

FIGURE 7-3 Downstream Path Logic Block Diagram

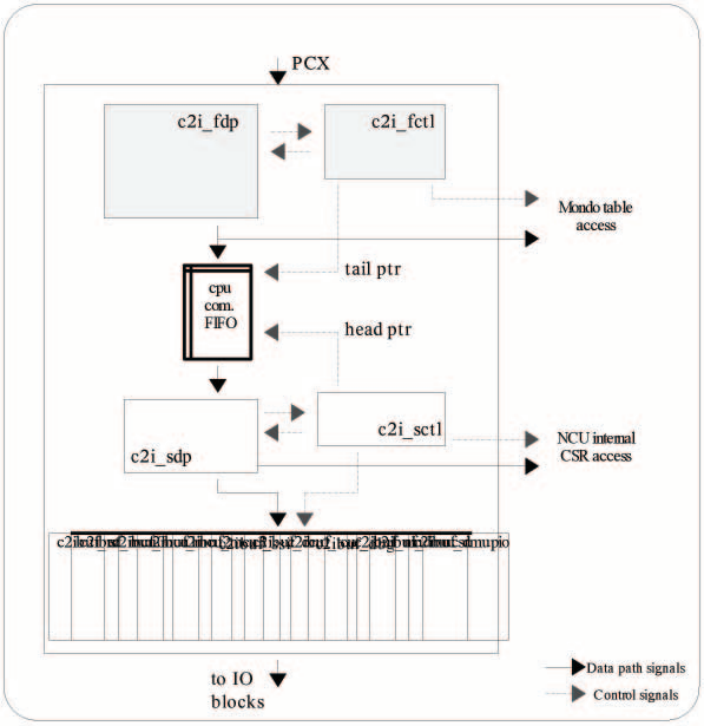
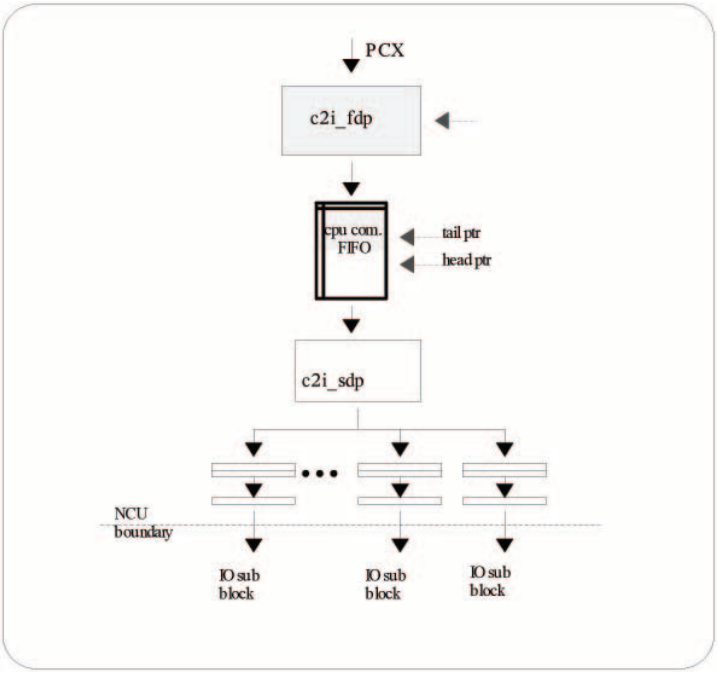


FIGURE 7-4 Downstream Data Path Block Diagram



Note – There is no downstream communications to SIU because all non cacheable external CPU load/store requests (PIO requests) are sent directly to DMU, and there is no CSR in SIU neither. However, SIU talks to NCU on the upstream data return path, and this will be covered in the upstream path.

7.3.2 Upstream Path Block Diagrams

FIGURE 7-5 and FIGURE 7-6 show the logical block diagram for the upstream communication path. NCU collects packets from each IO block and push them into the upstream main FIFO in IO clock domain. For Mondo Interrupt case (originated from DMU and sent via SIU to NCU), NCU checks the internal status table for the target CPU thread's availability and responses with an “ack” or “nack,” An “ack” means the Mondo Interrupt is accepted, and a “nack” means it is rejected. The upstream main FIFO is a 32 deep domain crossing FIFO, shared by all IO blocks. Data is written in IO clock domain and read out in CPU clock domain. The head pointer and tail pointer of the FIFO are controlled by the i2c_sctl block in IO clock

domain and i2c_fctl block in CPU clock domain, respectively. C2i_fdp and c2i_fctl blocks also arbitrate and mux between the upstream main FIFO output and CPU Mondo lookup data output as shown in FIGURE 7-5 and FIGURE 7-6.

FIGURE 7-5 Upstream Path Logic Block Diagram

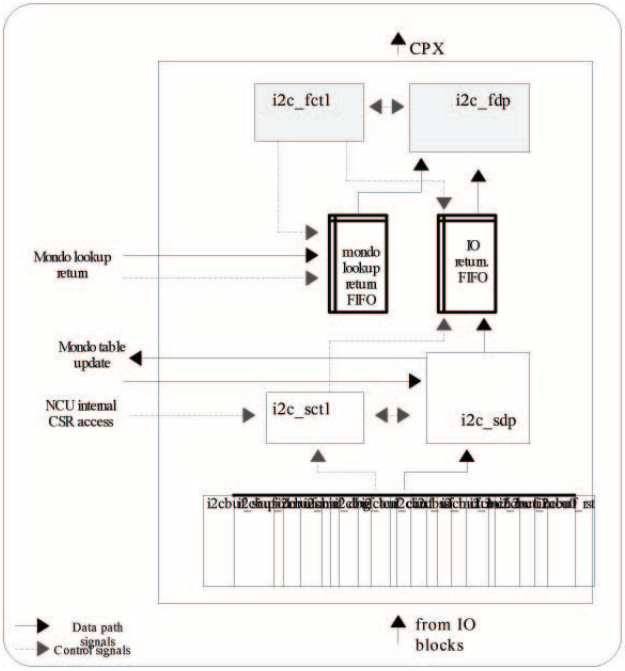
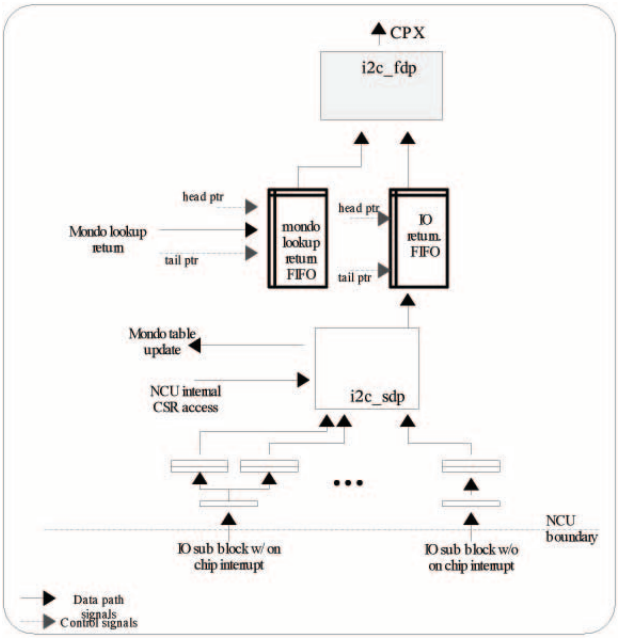


FIGURE 7-6 Upstream Data Path Block Diagram



7.4 Interface Signals, Protocols, and Timing Diagrams

TABLE 7-2 NCU/XBAR (CCX) Interface Signals

NCU/XBAR(CCX) Interface Signals	Direction	Comment
ncu_pcx_stall_pq	Output	NCU back pressure control signal to CCX
pcx_ncu_data_rdy_px1	Input	PCX to NCU data valid (px1 version)
pcx_ncu_data_px2[129:0]	Input	PCX to NCU data bus (px2 version)
cpx_ncu_grant_cx[7:0]	Input	CPX grant indicates the corresponding packet has reached its CPU destination and there is room for more packet for the same corresponding CPU destination.
ncu_cpx_req_cq[7:0]	Output	NCU to CPX request signals
ncu_cpx_data_ca[144:0]	Output	NCU to CPX data bus

TABLE 7-3 NCU/MCU0 Interface Signals

NCU/MCU0 Interface Signals	Direction	Comment
mcu0_ncu_stall	Input	MCU0 back pressure control signal to NCU
ncu_mcu0_vld	Output	NCU to MCU0 data valid
ncu_mcu0_data[3:0]	Output	NCU to MCU0 data bus
ncu_mcu0_stall	Output	NCU back pressure control signal to MCU0
mcu0_ncu_vld	Input	MCU0 to NCU data valid
mcu0_ncu_data[3:0]	Input	MCU0 to NCU data bus
ncu_mcu0_e0i	Output	Error injection0 to MCU0
mcu0_ncu_e1	Input	Error strobe1 from MCU0
ncu_mcu0_e1i	Output	Error injection1 to MCU0
mcu0_ncu_e2	Input	Error strobe2 from MCU0
ncu_mcu0_e2i	Output	Error injection2 to MCU0

TABLE 7-4 NCU/MCU1 Interface Signals

NCU/MCU1 Interface Signals	Direction	Comment
mcu1_ncu_stall	Input	MCU1 back pressure control signal to NCU
ncu_mcu1_vld	Output	NCU to MCU1 data valid
ncu_mcu1_data[3:0]	Output	NCU to MCU1 data bus
ncu_mcu1_stall	Output	NCU back pressure control signal to MCU1
mcu1_ncu_vld	Input	MCU1 to NCU data valid
mcu1_ncu_data[3:0]	Input	MCU1 to NCU data bus
mcu1_ncu_e0	Input	Error strobe0 from MCU1
ncu_mcu1_e0i	Output	Error injection0 to MCU1
mcu1_ncu_e1	Input	Error strobe1 from MCU1
ncu_mcu1_e1i	Output	Error injection1 to MCU1
mcu1_ncu_e2	Input	Error strobe2 from MCU1
ncu_mcu1_e2i	Output	Error injection2 to MCU1

TABLE 7-5 NCU/MCU2 Interface Signals

NCU/MCU2 Interface Signals	Direction	Comment
mcu2_ncu_stall	Input	MCU2 back pressure control signal to NCU
ncu_mcu2_vld	Output	NCU to MCU2 data valid
ncu_mcu2_data[3:0]	Output	NCU to MCU2 data bus
ncu_mcu2_stall	Output	NCU back pressure control signal to MCU2
mcu2_ncu_vld	Input	MCU2 to NCU data valid
mcu2_ncu_data[3:0]	Input	MCU2 to NCU data bus
mcu2_ncu_e0	Input	Error strobe0 from MCU2
ncu_mcu2_e0i	Output	Error injection0 to MCU2
mcu2_ncu_e1	Input	Error strobe1 from MCU2

TABLE 7-5 NCU/MCU2 Interface Signals *(Continued)*

NCU/MCU2 Interface Signals	Direction	Comment
ncu_mcu2_e1i	Output	Error injection1 to MCU2
mcu2_ncu_e2	Input	Error strobe2 from MCU2
ncu_mcu2_e2i	Output	Error injection2 to MCU2

TABLE 7-6 NCU/MCU3 Interface Signals

NCU/MCU3 Interface Signals	Direction	Comment
mcu3_ncu_stall	Input	MCU3 back pressure control signal to NCU
ncu_mcu3_vld	Output	NCU to MCU3 data valid
ncu_mcu3_data[3:0]	Output	NCU to MCU3 data bus
ncu_mcu3_stall	Output	NCU back pressure control signal to MCU3
mcu3_ncu_vld	Input	MCU3 to NCU data valid
mcu3_ncu_data[3:0]	Input	MCU3 to NCU data bus
mcu3_ncu_e0	Input	Error strobe0 from MCU3
ncu_mcu3_e0i	Output	Error injection0 to MCU3
mcu3_ncu_e1	Input	Error strobe1 from MCU3
ncu_mcu3_e1i	Output	Error injection1 to MCU3
mcu3_ncu_e2	Input	Error strobe2 from MCU3
ncu_mcu3_e2i	Output	Error injection2 to MCU3

TABLE 7-7 NCU/SSI Interface Signals

NCU/SSI Interface Signals	Direction	Comment
ncu_mio_ssi_sck	Output	Boot ROM interface clk (iol2clk/8)
ncu_mio_ssi_mosi	Output	Boot ROM interface data ssi Output to ROM
mio_ncu_ssi_miso	Input	Boot ROM interface data ROM to ssi
mio_ncu_ssi_ext_int_l	Input	Low active external trigger interrupt

TABLE 7-8 NCU/DBG1 Interface Signals

NCU/DBG1 Interface Signals	Direction	Comment
dbg1_ncu_stall	Input	DBG1 back pressure control signal to NCU
dbg1_ncu_vld	Input	DBG1 to NCU data valid
dbg1_ncu_data[3:0]	Input	DBG1 to NCU data bus
ncu_dbg1_vld	Output	NCU to DBG1 data valid
ncu_dbg1_data[3:0]	Output	NCU to DBG1 data bus
ncu_dbg1_stall	Output	NCU back pressure control signal to DBG1
ncu_dbg1_error_event	Output	NCU error happens, enabled with wmr_vec_mask

TABLE 7-9 NCU/CCU Interface Signals

NCU/CCU Interface Signals	Direction	Comment
ccu_ncu_stall	Input	CCU back pressure control signal to NCU
ncu_ccu_vld	Output	NCU to CCU data valid
ncu_ccu_data[3:0]	Output	NCU to CCU data bus
ncu_ccu_stall	Output	NCU back pressure control signal to CCU
ccu_ncu_vld	Input	CCU to NCU data valid
ccu_ncu_data[3:0]	Input	CCU to NCU data bus

TABLE 7-10 NCU/TCU Interface Signals

NCU/TCU Interface Signals	Direction	Comment
tcu_ncu_stall	Input	TCU back pressure control signal to NCU
tcu_ncu_vld	Input	TCU to NCU data valid
tcu_ncu_data[7:0]	Input	TCU to NCU data bus
ncu_tcu_stall	Output	NCU back pressure control signal to TCU
ncu_tcu_vld	Output	NCU to TCU data valid.
ncu_tcu_data[7:0]	Output	NCU to TCU data bus.

TABLE 7-10 NCU/TCU Interface Signals *(Continued)*

NCU/TCU Interface Signals	Direction	Comment
ncu_tcu_soc_error	Output	One pulse signal to TCU each time when an soc error packet is generated from NCU to the core
ncu_tcu_bank_avail[7:0]	Output	Copy from bankavail[7:0].
tcu_ncu_mbist_start[1:0]	Input	Mbist start (1'b0 for normal function mode)
ncu_tcu_mbist_done[1:0]	Output	Mbist done
ncu_tcu_mbist_fail[1:0]	Output	Mbist fail
tcu_dbr_gateoff	Input	Turn off all the vld and stall when it is 1'b1.

TABLE 7-11 NCU/RST Interface Signals

NCU/RST Interface Signals	Direction	Comment
rst_ncu_stall	Input	RST back pressure control signal to NCU
ncu_rst_vld	Output	NCU to RST data valid
ncu_rst_data[3:0]	Output	NCU to RST data bus
ncu_rst_stall	Output	NCU back pressure control signal to RST
rst_ncu_vld	Input	RST to NCU data valid
rst_ncu_data[3:0]	Input	RST to NCU data bus
rst_ncu_unpark_thread	Input	After each “warm reset” is de-asserted and all BISX activities are completed, RST send in a 1-clock wide pulse to tell NCU the system is ready to wake up the master thread, which is the lowest available thread basing on core_enable_status ASI register.
rst_ncu_xir_	Input	External Initiated Interrupt (multi-clock wide pulse signal) This signal triggers interrupts to cpu_thr that based on XIR_steering register. It will be deasserted when ncu_rst_xir_done is high.
ncu_rst_xir_done	Output	NCU asserts this signal back to RST block to indicate all XIR interrupts have been generated (multi-clock wide pulse). It will be deasserted when rst_ncu_xir_ is back to high.

TABLE 7-12 NCU/DMU CSR Interface Signals

NCU/DMU CSR Interface Signals	Direction	Comment
dmu_ncu_stall	Input	DMU CSR bus back pressure control signal to NCU.
ncu_dmu_vld	Output	NCU to DMU CSR data valid.
ncu_dmu_data[31:0]	Output	NCU to DMU CSR data bus.
ncu_dmu_stall	Output	NCU CSR bus back pressure control signal to DMU.
dmu_ncu_vld	Input	DMU to NCU CSR data valid.
dmu_ncu_data[31:0]	Input	DMU to NCU CSR data bus.

TABLE 7-13 NCU/DMU PIO and Mondo Interface

NCU/DMU PIO and Mondo Interface	Direction	Comment
ncu_dmu_pio_hdr_vld	Output	Indicates ncu_dmupio_data is valid for PIO header transaction.
ncu_dmu_mmu_addr_vld	Output	Indicates ncu_dmupio_data is valid for one cycle for "mmu invalidate vector." The vector is coming a write operation into CSR register 0x80_0000_2030
ncu_dmu_pio_data[63:0]	Output	NCU to DMU data bus
dmu_ncu_wrack_par	Input	Odd parity check for dmu_ncu_wrack_tag[3:0].
dmu_ncu_wrack_vld	Input	Indicates dmu_ncu_wrack_tag[3:0] is valid
dmu_ncu_wrack_tag[3:0]	Input	Credit ID back to NCU for PIO write completion.
ncu_dmu_mondo_ack	Output	Mondo Interrupt ack (ncu_dmu_mondo_id[5:0] is valid when this signal is asserted to indicate the mondo_id it is acking.)
ncu_dmu_mondo_nack	Output	Mondo Interrupt nack (ncu_dmu_mondo_id[5:0] is valid when this signal is asserted to indicate the mondo_id it is nacking.)
ncu_dmu_mondo_id[5:0]	Output	Mondo Interrupt ID, valid when ncu_dmu_mondo_ack or ncu_dmu_mondo_nack is asserted.
dmu_ncu_ctag_ue	Input	Ctag double bit ue from SIO DMA read return.
ncu_dmu_ctag_uei	Output	Ctag double bit ue injected by RASEJR.
dmu_ncu_ctag_ce	Input	Error strobe from dmu (ctag ue)

TABLE 7-13 NCU/DMU PIO and Mondo Interface *(Continued)*

NCU/DMU PIO and Mondo Interface	Direction	Comment
ncu_dmu_ctag_cei	Output	Error injection signal
dmu_ncu_d_pe	Input	Error strobe from dmu (data parity error)
ncu_dmu_d_pei	Output	Error injection signal
dmu_ncu_siicr_pe	Input	Error strobe from dmu (siicr parity error)
ncu_dmu_siicr_pei	Output	Error injection signal
dmu_ncu_ncucr_pe	Input	Error strobe from dmu (ncucr parity error)
ncu_dmu_ncucr_pei	Output	Error injection signal
dmu_ncu_ie	Input	Error strobe from dmu (internal error)
ncu_dmu_iei	Output	Error injection signal

TABLE 7-14 NCU/SII Interface Signals

NCU/SII Interface Signals	Direction	Comment
ncu_sii_gnt	Output	NCU to SII grant signal to indicate there is room to receive one more packet. Transaction should starts in the next cycle
sii_ncu_req	Input	SII to NCU packet available request
sii_ncu_data[31:0]	Input	SII to NCU data bus
sii_ncu_dparity[1:0]	Input	SII to NCU data parity (covers sii_ncu_data[31:0] bus for data cycle only). Odd parity: bit[0] covers even bits, and bit[1] covers odd bits.
sii_ncu_dmua_ue	Input	Error strobe from sii (dmu pkt address ue)
ncu_sii_dmua_uei	Output	Error injection signal
sii_ncu_dmuctag_ue	Input	Error strobe from sii (dmu pkt ctag ue)
ncu_sii_dmuctag_uei	Output	Error injection signal
sii_ncu_dmuctag_ce	Input	Error strobe from sii (dmu pkt ctag ce)
ncu_sii_dmuctag_cei	Output	Error injection signal
sii_ncu_dmud_pe	Input	Error strobe from sii (dmu pkt data parity error)
ncu_sii_dmud_pei	Output	Error injection signal
sii_ncu_niua_ue	Input	Error strobe from sii (niu pkt address ue)
ncu_sii_niua_uei	Output	Error injection signal

TABLE 7-14 NCU/SII Interface Signals *(Continued)*

NCU/SII Interface Signals	Direction	Comment
sii_ncu_niuctag_ue	Input	Error strobe from sii (niu pkt ctag ue)
ncu_sii_niuctag_uei	Output	Error injection signal
sii_ncu_niuctag_ce	Input	Error strobe from sii (niu pkt ctag ue)
ncu_sii_niuctag_cei	Output	Error injection signal
sii_ncu_niud_pe	Input	Error strobe from sii (niu pkt data parity ue)
ncu_sii_niud_pei	Output	Error injection signal
sii_ncu_syn_vld	Input	Error syndrome vld for sii_ncu_syn_data[3:0]
sii_ncu_syn_data[3:0]	Output	Error syndrome data bus for
ncu_sii_pm	Output	L2 bank partial mode. (Value is from BANK_ENABLE_STATUS register)
ncu_sii_ba01	Output	L2 bank0,1 available (Value is from BANK_ENABLE_STATUS register)
ncu_sii_ba23	Output	L2 bank2,3 available (Value is from BANK_ENABLE_STATUS register)
ncu_sii_ba45	Output	L2 bank4,5 available (Value is from BANK_ENABLE_STATUS register)
ncu_sii_ba67	Output	L2 bank6,7 available (Value is from BANK_ENABLE_STATUS register)
ncu_sii_l2_idx_hash_en	Output	L2 index hash enable. (Value is from L2_IDX_HASH_EN_STATUS register)

TABLE 7-15 SIO/NCU Interface Signals

SIO/NCU Interface Signals	Direction	Comment
sio_ncu_ctag_ue	Input	Error strobe from sio (ctag ue)
ncu_sio_ctag_uei	Output	Error injection signal
sio_ncu_ctag_ce	Input	Error strobe from sio (ctag ce)
ncu_sio_ctag_cei	Output	Error injection signal
sio_ncu_d_pe	Input	Error strobe from sio (data parity error)
ncu_sio_d_pei	Output	Error injection signal

TABLE 7-16 eFuse/NCU Interface Signals

eFuse/NCU Interface Signals	Direction	Comment
efu_ncu_fuse_data	Input	Fuse unit serial data signal
efu_ncu_coreavl_xfer_en	Input	Indicates data bit is valid for core available register.
efu_ncu_bankavl_xfer_en	Input	Indicates data bit is valid for bank available register.
efu_ncu_fusestat_xfer_en	Input	Indicates data bit is valid for fuse status reg.
efu_ncu_sernum0_xfer_en	Input	Indicates data bit is valid for sernum0 reg.
efu_ncu_sernum1_xfer_en	Input	Indicates data bit is valid for sernum1 reg.
efu_ncu_sernum2_xfer_en	Input	Indicates data bit is valid for sernum2 reg.

TABLE 7-17 CCU/NCU Interface Signals

CCU/NCU Interface Signals	Direction	Comment
ccu_cmp_io_sync_en	Input	Sync. pulse for cmp domain to io domain
ccu_io_cmp_sync_en	Input	Sync. pulse for io domain to cmp domain
tcu_pce_ov	Input	TEST control. 1'b0 for normal functional mode
tcu_ncu_clk_stop	Input	TEST control. 1'b0 for normal functional mode
tcu_ncu_io_stop	Input	TEST control. 1'b0 for normal functional mode
ccu_io_out	Input	CCU output goes to NCU io clkgen.
tcu_aclk	Input	SCAN clock
tcu_bclk	Input	SCAN clock

TABLE 7-18 Global Signals

Global Signals	Direction	Comment
scan_in	Input	SCAN IN (1'b0 or 1'b1 for normal function mode simulation)
scan_out	Output	SCAN OUT
tcu_ncu_mbist_scan_in	Input	Mbist scan in (1'b0 or 1'b1 for normal function mode simulation)
ncu_tcu_mbist_scan_out	Output	Mbist scan out

TABLE 7-18 Global Signals *(Continued)*

Global Signals	Direction	Comment
tcu_mbist_bisi_en	Input	Bist engine enable (1'b0 for normal function mode)
tcu_scan_en	Input	SCAN enable. 1'b0 for normal functional mode
tcu_se_scancollar_in	Input	TEST control. 1'b0 for normal functional mode
tcu_se_scancollar_out	Input	TEST control. 1'b0 for normal functional mode
tcu_array_wr_inhibit	Input	TEST control. 1'b0 for normal functional mode

TABLE 7-19 Signals to L2T

Signals to L2T	Direction	Comment
ncu_l2t_pm	Output	L2 bank partial mode. (Value is from BANK_ENABLE_STATUS register)
ncu_l2t_ba01	Output	L2 bank0,1 available (Value is from BANK_ENABLE_STATUS register)
Ncu_l2t_ba23	Output	L2 bank2,3 available (Value is from BANK_ENABLE_STATUS register)
ncu_l2t_ba45	Output	L2 bank4,5 available (Value is from BANK_ENABLE_STATUS register)
ncu_l2t_ba67	Output	L2 bank6,7 available (Value is from BANK_ENABLE_STATUS register)

TABLE 7-20 Signals to all SPC

Signals to all SPC	Direction	Comment
ncu_spc_pm	Output	L2 bank partial mode. (Value is from BANK_ENABLE_STATUS register)
ncu_spc_ba01	Output	L2 bank0,1 available (Value is from BANK_ENABLE_STATUS register)
ncu_spc_ba23	Output	L2 bank2,3 available (Value is from BANK_ENABLE_STATUS register)
ncu_spc_ba45	Output	L2 bank4,5 available (Value is from BANK_ENABLE_STATUS register)
ncu_spc_ba67	Output	L2 bank6,7 available (Value is from BANK_ENABLE_STATUS register)

TABLE 7-20 Signals to all SPC *(Continued)*

Signals to all SPC	Direction	Comment
ncu_spc_l2_idx_hash_en	Output	L2 index hash enable. (Value is from L2_IDX_HASH_EN_STATUS register)
cmp_tick_enable	Output	ASI register cmp_tick_enable signal.
tcu_wmr_vec_mask	Output	ASI register wmr_vec_mask signal.

TABLE 7-21 SPC 0

SPC 0	Direction	Comment
ncu_spc0_core_enable_status	Output	For gating off clock to SPC0
ncu_spc0_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC0 for each thread
spc0_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC0 status is active or parking.

TABLE 7-22 SPC1

SPC1	Direction	Comment
ncu_spc1_core_enable_status	Output	For gating off clock to SPC1
ncu_spc1_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC1 for each thread
spc0_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC1 status is active or parking.

TABLE 7-23 SPC2

SPC2	Direction	Comment
ncu_spc2_core_enable_status	Output	For gating off clock to SPC2
ncu_spc2_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC2 for each thread
spc2_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC2 status is active or parking.

TABLE 7-24 SPC3

SPC3	Direction	Comment
ncu_spc3_core_enable_status	Output	For gating off clock to SPC3
ncu_spc3_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC3 for each thread
spc3_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC3 status is active or parking.

TABLE 7-25 SPC4

SPC4	Direction	Comment
ncu_spc4_core_enable_status	Output	For gating off clock to SPC4
ncu_spc4_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC4 for each thread
spc4_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC4 status is active or parking.

TABLE 7-26 SPC5

SPC5	Direction	Comment
ncu_spc5_core_enable_status	Output	For gating off clock to SPC5
ncu_spc5_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC5 for each thread
spc5_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC5 status is active or parking.

TABLE 7-27 SPC6

SPC6	Direction	Comment
ncu_spc6_core_enable_status	Output	For gating off clock to SPC6
ncu_spc6_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC6 for each thread
spc6_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC6 status is active or parking.

TABLE 7-28 SPC7

SPC7	Direction	Comment
ncu_spc7_core_enable_status	Output	For gating off clock to SPC7
ncu_spc7_core_running[7:0]	Output	8-bit signals to indicate parking or unparking request to SPC7 for each thread
spc7_ncu_core_running_status[7:0]	Input	8-bit signals to indicate the current SPC7 status is active or parking.

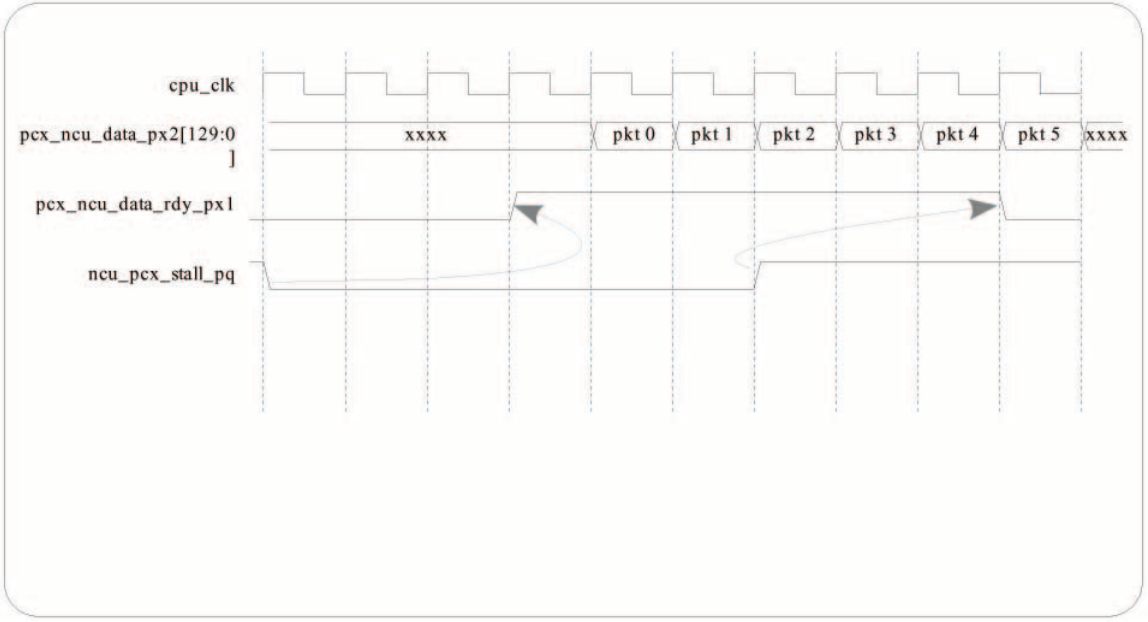
7.4.1 XBAR Interface

7.4.1.1 NCU/XBAR PCX Interface (Downstream)

The PCX interface is a 130 bit-wide bus with two bits flow control signals. The signal “ncu_pcx_stall_pq” is to back pressure XBAR when the downstream CPU shared buffer becomes full. An asserted “pcx_ncu_data_rdy_px1” indicates the data bus “pcx_ncu_data_px2[129:0]” to NCU is valid in next cycle. FIGURE 7-7 shows the case that CPU shared buffer has available entry and is filled by PCX packets again. The de-assertion of “ncu_pcx_stall_pq” indicates there is room for at least six more PCX packets. Due to the nature of pipelining design, there may be three more packets in flight after the assertion of “ncu_pcx_stall_pq” signal. Therefore, when “ncu_pcx_stall_pq” signal is asserted, NCU is guaranteed to be able to accept as least three more packets from XBAR PCX interface as shown in FIGURE 7-7.

Note – NCU is using px1 version of “data ready” signal and px2 version for “data” bus.

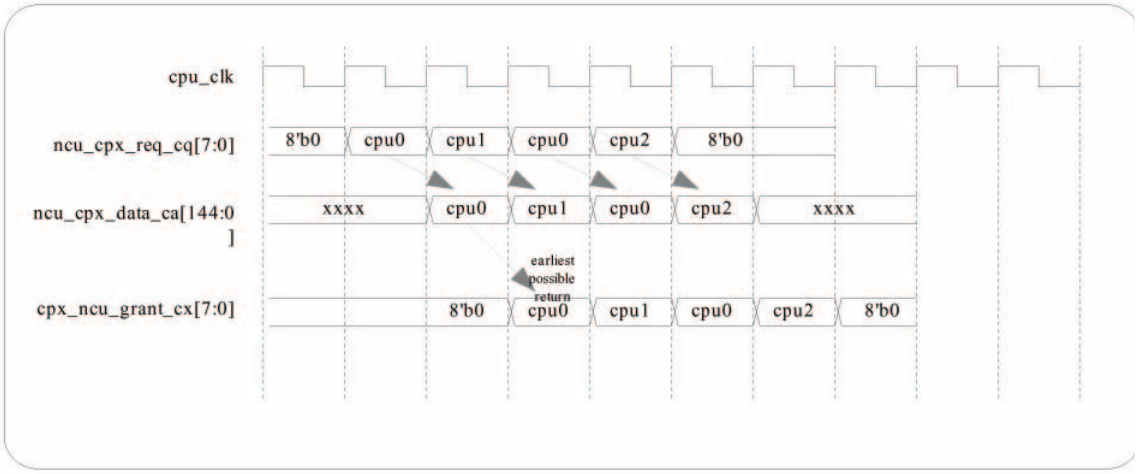
FIGURE 7-7 Downstream PCX Interface Timing



7.4.1.2 NCU/XBAR CPX Interface (Upstream)

The CPX interface is a 146 bit-wide data bus plus two sets of eight bit-wide flow control. NCU keeps track of the number outstanding requests without grant for each of the eight CPU. When the number of outstanding requests without grant reaches two for a particular CPU, NCU will stop sending the third request to the same CPU until the first grant has returned. This is mainly due to CCX has two levels of buffering for each CPU destination, and sending a third outstanding packet to the same destination will result in packet being lost. The timing diagram for the CPX bus is shown in FIGURE 7-8.

FIGURE 7-8 Upstream PCX Interface Timing



7.4.2 NCU/MCU Interface

There are four MCUs on OpenSPARC T2, and they are all connected to NCU in the same manner. The downstream and upstream paths are both four bit-wide data bus with two control signals. The interface protocol is a 128 bit packet being spread into 32 cycles of transactions. NCU only sends type “READ_REQ,” and “WRITE_REQ.” with 8B request size to MCU for CSR access.

The packet types that MCU sends upstream to NCU are:

“READ_ACK,” with 8B payload in response to a successful “READ_REQ,” (128-bit UCB packet)

“READ_NACK,” without payload in response to an unsuccessful “READ_REQ,” (64-bit UCB packet without payload)

“INT,” for on chip interrupt, resulting from some error conditions in MCU (64-bit UCB Int. packet with dev_id = 1)

FIGURE 7-9 and FIGURE 7-10 show the downstream and upstream timing diagram for NCU/MCU interface.

FIGURE 7-9 NCU to MCU/SSI/RNG/CCU/RST Downstream Timing Diagram (back-to-back case)

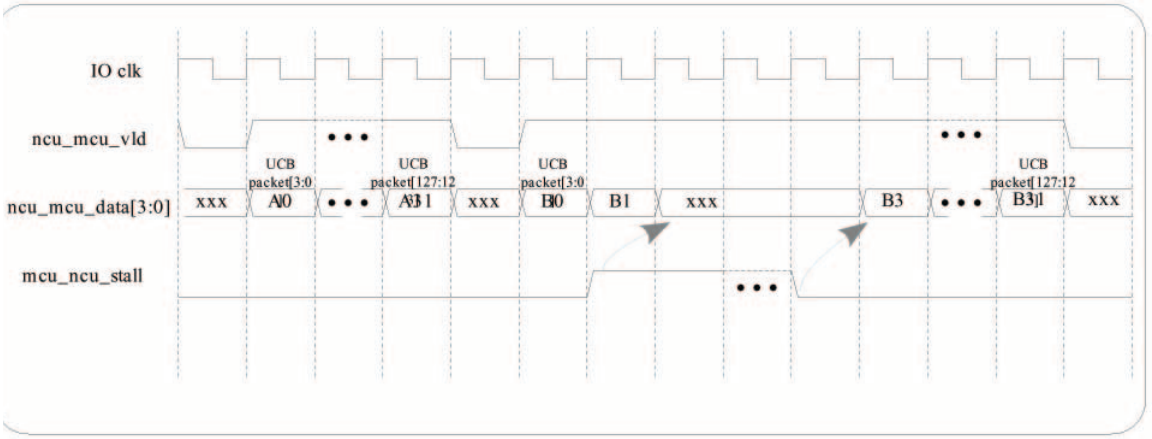
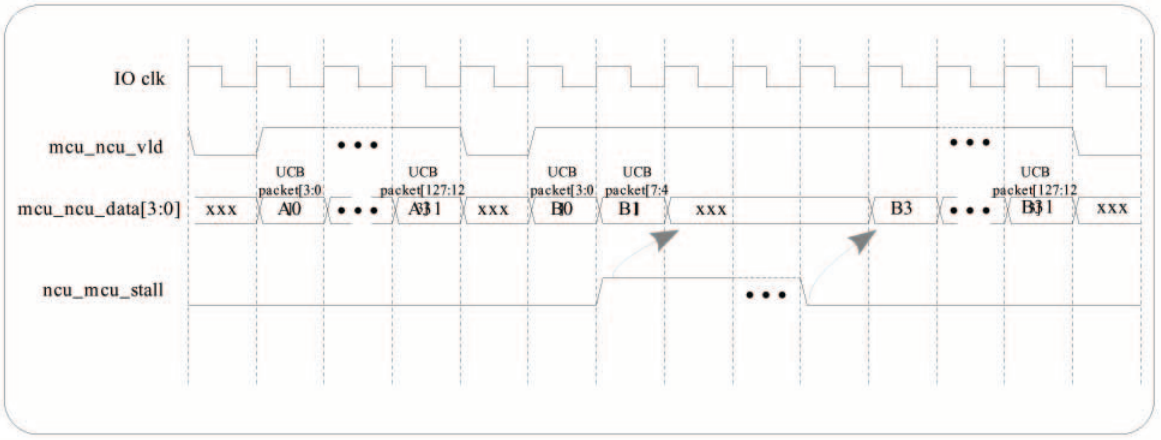


FIGURE 7-10 MCU/SSI/RNG/CCU/RST to MCU Upstream Timing Diagram



Back-to-back read ack return of IFIKK return with 8B payload.

7.4.3 Boot ROM Interface (NCU/SSI))

NCU has integrated the SSI interface logics which originated from OpenSPARC T1. With modification, the ncu_mio_ssi_sck frequency is now programmable. It could be iol2clk/8 (default) or iol2clk/4. There are four i/o pins directly connecting to the

external. Reference Appendix A regarding the boot ROM interface. The original SSI UCB interface has become NCU's internal signals and is no longer visible from outside of NCU cluster. If a request from CPU is an IFILL request, but the 40-bit PA is not addressed to SSI (0xFF_Fxxxx_xxxx), NCU/SSI classifies this as an un-deliverable packet, and will reply with IFILL return to CPU with uncorrectable error set. When CPU initiates an IFILL request to NCU, CPU expects only four bytes IFILL load return. For IFILL request, CPU should only request 4Byte. The F4B field in CPX packet should always be set to "1". However, for SSI's CSR request, CPU should always do 8-byte access similar to all other CSR access. SSI also supports non-IFILL and non CSR type access, which read or write to external. For this type of requests, SSI supports 1,2,4,8 bytes access.

ncu_mio_ssi_sck could be programmed as iol2clk/8 or iol2clk/4, depends on the CSR register NCU_SCKSEL. This register is warm_reset protected. The new value programmed into NCU_SCKSEL register, can't effects current ncu_mio_ssi_sck until next warm reset.

After warm reset, NCU holds up to five ms before sending first request. This is the time FPGA needs to lock sck clock. However, during test and debug mode TCU can drive tcu_sck_bypass signal to "1", and this will cause NCU/SSI to skip the 5ms wait. System developer should make sure the external boot ROM interface logic is stabilized and ready before SSI sends out the first request.

7.4.4 NCU/CCU Interface

The CCU interface is same as the NCU/MCU interface in UCB packet format. The request size is always 8B and the request types that NCU sends downstream to CCU are "READ_REQ," and "WRITE_REQ."

CCU returns the following to NCU:

- "READ_ACK, "with 64 bit payload in response to a successful CSR "READ_REQ;"
- "READ_NACK," without payload in response to an unsuccessful CSR "READ_REQ."

The interface timing diagram for NCU/CCU is same as NCU/MCU, which can be found in FIGURE 7-9 and FIGURE 7-10.

7.4.5 NCU/RST Interface

The RST interface is same as the NCU/MCU interface in UCB packet format. The request size is always 8B and the request types that NCU sends downstream to RST are "READ_REQ," and "WRITE_REQ." RST returns the following to NCU:

- "READ_ACK,"with 64 bit payload in response to a successful CSR "READ_REQ;"
- "READ_NACK," without payload in response to an unsuccessful CSR "READ_REQ."

The interface timing diagram for NCU/RST is same as NCU/MCU, which can be found in FIGURE 7-9 and FIGURE 7-10.

7.4.6 NCU/DMU CSR Interface

The DMUCSR Interface is same as the NCU interface in UCB packet format. The request size is always 8B, and the request types that NCU sends downstream to DMUCSR interface are "READ_REQ," and "WRITE_REQ." DMUCSR interface returns the following to NCU:

- "READ_ACK,"with 64 bit payload in response to a successful CSR "READ_REQ;"
- "READ_NACK," without payload in response to an unsuccessful CSR "READ_REQ."

The interface timing diagram for NCU/DMUCSR is same as NCU/NIU, which can be found in FIGURE 7-11 and FIGURE 7-12.

7.4.7 NCU/DBG Interface

NCU/DBG interface is similar to NCU/MCU interface. The downstream and upstream paths are both four bit-wide data bus with two control signals. The interface protocol is a 128 bit packet being spread into 32 cycles of transactions. NCU only sends type "READ_REQ," and "WRITE_REQ." with 8B request size to DBG for CSR access. The packet types that DBG sends upstream to NCU are:

- "READ_ACK," with 8B payload in response to a successful "READ_REQ," (128-bit UCB packet)
- "READ_NACK," without payload in response to an unsuccessful "READ_REQ," (64-bit UCB packet without payload)

Refer to NCU/MCU Interface for NCU/DBG interface.

7.4.8 NCU/TCU Interface

The TCU interface is similar to NCU/MCU or NCU interfaces using UCB packet format except it is an eight-bit data bus plus two control signals each way. For write_req, read_req, and read_ack type, the packet size is 128-bit and requires 16 cycles to complete the transaction. For read_nack type, the packet size is 64-bit, and

requires eight cycles to complete the transaction. TCU is intended to connect to the external service processor (JTAG/TAP controller,) and, therefore, is capable to initiate request type UCB packet into NCU. TCU can be in both master or slave mode.

NCU does not support JTAG/TAP access across the Crossbar to L2s nor CPUs. Therefore, TCU is limited to access the following via NCU: NCU's CSR, MCUs' CSR, NIU's CSR/PIO, SSI's CSR, DMU's CSR+PIO, RST's CSR, CCU's CSR.

The request type UCB packet from TCU to NCU should contain the following fields: Buffer ID (always 2'b01), a valid 40-bit PA field, a valid Packet Type and a Request Size field. On the return path, when an UCB packet returned to NCU with the Buffer ID field marked as TAP packet, NCU routes the packet back to TCU accordingly. All write requests from TCU are “posted,” which means no “ack” is generated back to TCU after a write request, and the packet will be dropped silently if address is illegal. This implies TCU can generated multiple consecutive write requests (possibly back-to-back) in a short period of time because it does not require “ack” for a write request. However, once TCU generates a read requests to NCU, there should not be any more requests until a “READ_ACK” or “READ_NACK” UCB packet has returned back to TCU. NCU does not support Interrupt type packet nor IFILL type packet on NCU/TCU interface.

TCU/NCU interface is designed as a low performance, infrequent access interface. Unlike other interface, NCU provides only minimum buffering for TCU accessing. Excessive traffic from TCU can possibly slow down performance of NCU due to lack of buffering issue.

TABLE 7-29 UCB Packet Types supported on TCU/NCU interface

WRITE_REQ	TCU->NCU	128-bit UCB packet (8B header + 8B payload)
READ_REQ	TCU->NCU	128-bit UCB packet (8B header + 8B meaningless payload)
READ_ACK	NCU->TCU	128-bit UCB packet (8B header + 8B payload). Note: PA and Size fields are invalid in a return packet.
READ_NACK	NCU->TCU	64-bit UCB packet (8B header only). Note: PA and Size fields are invalid in a return packet.

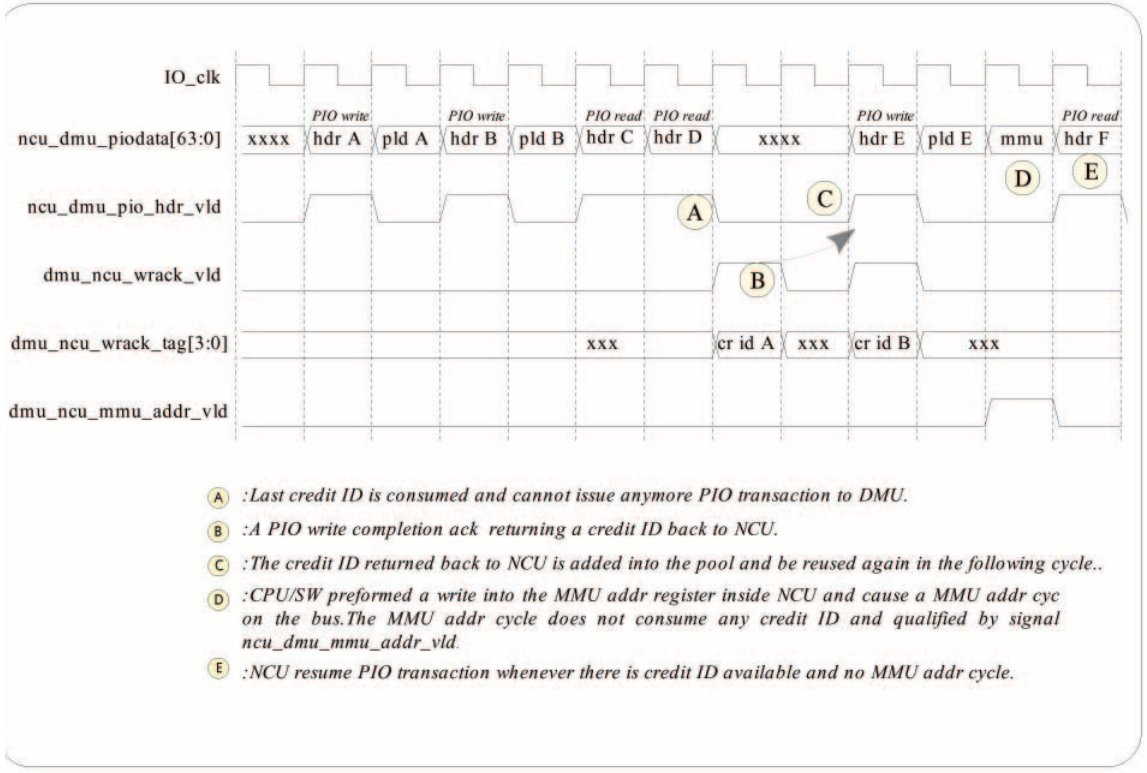
The interface timing diagram for NCU/TCU is same as NCU/MCU, which can be found in FIGURE 7-9and FIGURE 7-10 with data bus set to [7:0] and number of cycles set to 16.

7.4.9 NCU/DMU PIO Interface

NCU sends PIO read/write request (non cacheable LOAD_REQ/STORE_REQ) directly through the NCU/DMUPIO interface. NCU keeps a total of 16 “credit Ids. Each PIO request sends to DMU will consume a “credit ID,” which will be returned from the signal dmu_ncu_wrack_tag[3:0] after a PIO write is completed. The PIO

read returned packet from SIU also has a returning “credit ID” embedded in its header. These returned “credit Ids are put back to the pool and will be reused again. Therefore, there can be a maximum of 16 outstanding PIO read and PIO write requests. Note that DMU has a limit of processing up to 16 PIO requests FIGURE 7-11 show the timing diagram the DMUPIO interface. Signal dmu_ncu_mmu_addr_vld will be asserted when CPU perform a write to NCU's MMU_ID_ADDR register, and the value of the register is put on to the ncu_dmu_piodata[63:0] bus.

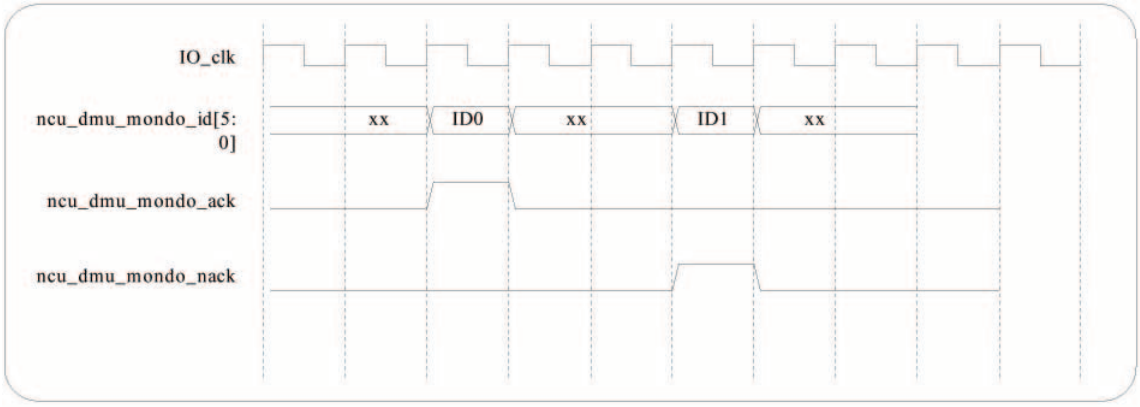
FIGURE 7-11 NCU/DMUPIO Interface Timing Diagram



7.4.10 NCU/DMU Mondo Response Interface

After receiving Mondo Interrupt packet from SIU, NCU directly response to DMU with the Mondo ID which is qualified by an “ack” or a “nack” at the same cycle. The six-bit ID bus is valid when “ncu_dmu_mondo_ack” or “ncu_dmu_mondo_nack” signal is asserted. FIGURE 7-12 shows the timing diagram for NCU/DMU Mondo Response interface.

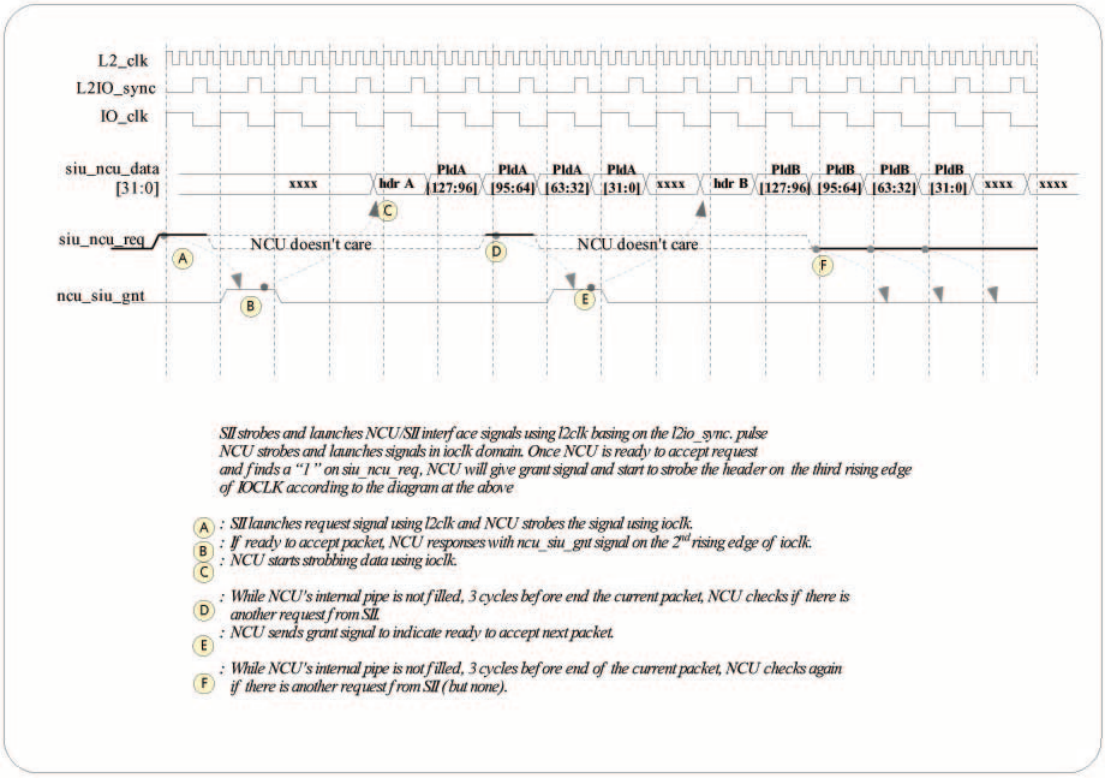
FIGURE 7-12 NCU/DMU Mondo Response Interface Timing Diagram (from NCU to DMU.)



7.4.11 NCU/SII Interface

SII only has upstream path to NCU. Packets received from SII are either PIO read returns or Mondo Interrupts. When signal “ncu_siu_gnt” is asserted, in response to an asserted “siu_ncu_req”, a new packet transaction should start in the next cycle. Once a packet transaction is in progress, NCU ignores the signal “siu_ncu_req” until two cycle before ending of the current packet. Details of the interface timing is shown in FIGURE 7-13. packet from SII to NCU always have five cycles (one header cycle with four payload cycles.)

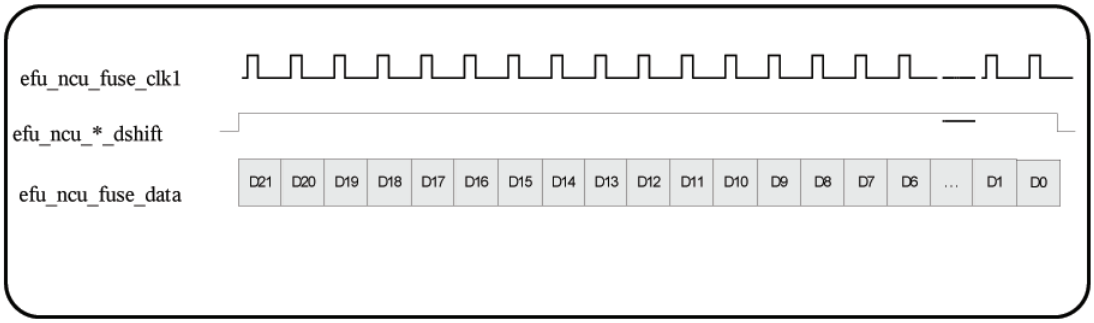
FIGURE 7-13 NCU/SIU Interface Timing Diagram (from SIU to NCU)



7.4.12 eFuse Interface

NCU only receive signals from eFuse. This interface has a serial data signal shared by different register. eFuse will guarantee there are 22 consecutive bits/data with MSB first per "*dshift" assertion as shown if the following diagram. The only exception is efu_ncu_fusestat_dshift signal, which is 64-bit status information inputting into NCU's eFuse Status CSR register.

FIGURE 7-14 EFU/NCU Interface Timing Diagram.



7.4.13 Packet Format

7.4.13.1 UCB (Unit Control Block) Data Packet Format

TABLE 7-30 UCB Data Packet Format

Bit	Name	Definition
[127:64]	Payload Data	8B Payload Data. If packet is from PCX, this field is from PCX's data field. Payload is valid only for "WRITE_REQ," "READ_ACK," and "IFILL_ACK" types packets For "READ_REQ," and "IFILL_REQ" packet, this 8B payload is meaningless. For "READ_NACK," or "IFILL_NACK" types, there is no payload.
[63:55]	ByteMask	(use only for TCU to DMUPIO packets) This field is being ignored in general. Only exception is when TCU initiates a DMUPIO request. In such case, this field is being treated similar to PCX packet's size field.
[54:15]	40 bit PA	This field is valid only for request type packet, and should be ignored for return type packet. CSR Address: Upper eight bits indicates the block that it should go to. Individual block should only look at the lower 32 bits If packet is from PCX, this field is from PCX's address field.
[14:12]	Request Size	(This field is being ignored in TCU to DMUPIO packets) This field is valid only for request type packets and should be ignored for return type packets. If packet is from PCX, this field is from PCX's size field. Supported size for request type: 3'b000 : 1 Byte (valid only for SSI non-ifill type) 3'b001 : 2 Byte (valid only for SSI non-ifill type) 3'b010 : 4 Byte (valid only for SSI, can be ifill or non-ifill type) 3'b011 : 8 Byte (for all other access) NOTE: UCB protocol only support up to 8Byte payload max. DMUPIO is not in UCB protocol, and, therefore,is not limited by this restriction.
[11:10]	Buffer ID	NCU sends 2'b00 if a request is originated from Core and sends 2'b01 if request is originated from TCU. All UCB clients returning packets to NCU must return the same value in this field as in the original request packet.

TABLE 7-30 UCB Data Packet Format *(Continued)*

Bit	Name	Definition
[9:7]	CPU ID [2:0]	This field indicates the source CPUID this packet is from, or the target CPUID this packet should be send back to.
[6:4]	Thread ID [2:0]	This field indicates the target thread this packet is from or targeting to.
[3:0]	Packet Type	4'b0000: READ NACK (generates CPX NCU Load Return with U.E. if packet is to CPX) 4'b0001: READ ACK (generates CPX NCU Load Return if packet is to CPX) 4'b0011: IFILL ACK (generates CPX NCU Ifill Return if packet is to CPX) 4'b0111: IFILL NACK (generates CPX NCU Ifill Return with err if packet is to CPX) 4'b0100: READ REQ (from PCX LOAD if packet is from PCX) 4;b0101: WRITE REQ (from PCX STORE if packet is from PCX) 4'b0110: IFILL REQ (from PCX Inst. FILL if packet is from PCX)

7.4.13.2 UCB (Unit Control Block) Interrupt Packet Format

TABLE 7-31 UCB Interrupt Packet Format

Bit	Name	Definition
[63:57]	reserved	Reserved (may not be 0)
[56:51]	Interrupt vector	interrupt vector (valid only when Packet Type=INT_VEC, MCU, SSI should always use Packet Type = INT, which cause s NCU to ignore this field)
[50:19]	Reserved	Reserved (may not be 0)
[18:10]	Device ID	This field identify the entry of the int_man mem. lookup table.
[9:7]	CPU ID	This field indicates the target CPU this interrupt packet should be sent to when Packet Type = INT (should not be use by MCU, or SSI)
[6:4]	Thread ID	This field indicates the target thread this interrupt packet should be sent to when Packet Type = INT
[3:0]	Packet Type	4'b1000: INT (interrupt) 4'b1100: INT_VEC (interrupt w/ vector field,cpu_id,and thread_id valid)

7.4.13.3 SII to NCU Header Format

Each header is followed by four cycles of payload data.

TABLE 7-32 SIU to NCU Header Format

Header Cycle "siu_ncu_data[31:0]"	Name	Definition
[31]	TimeOut	Packet timed out (this will cause a CPX NCU Load Return packet with 'err' field set to uncorrectable error)
[30]	DmuAE	Unmapped Error (this will cause a CPX NCU Load Return packet with 'err' field set to uncorrectable error)
[29]	DmuUe	Uncorrected error from DMU (this will normally cause a CPX NCU Load Return packet with 'err' field set to uncorrectable error)
[28]	Ebit	Packet error bit (indicates packet has error and already reported by SII, and NCU will terminate this packet silently with any other action)

TABLE 7-32 SIU to NCU Header Format *(Continued)*

Header Cycle "siu_ncu_data[31:0]"	Name	Definition
[27:22]	Reserved	Reserved (ignore by NCU)
[21:16]	dmc_tag_ecc[5:0]	dmc_ctagecc check bits
[15:0]	dmc_tag[15:0]	dmc_tag[15] = 0 --> Mondo Interrupt packet For Mondo Interrupt, NCU returns {dmc_tag[14:11],dmc_tag[2:1]} back to DMU with Mondo_ack or Mondo_nack signals asserted. Targeted cpu_thread ID = payload[75:70] mondo_data0=payload[127:64] mondo_data1=payload[63:0] . dmc_tag[15] = 1 --> PIO read return packet dmc_tag[14:12]: reserved (must be 0) dmc_tag[11:8]: NCU credit ID return, dmc_tag[7:6]: buf_id[1:0], 2'b00=normal, 2'b01=JTAG access dmc_tag[5:0]: cpu_thread[5:0]

7.4.13.4 NCU to DMUPIO Header Format

TABLE 7-33 NCU to DMUPIO Header Format

Header Cycle "ncu_dmupio_data[63:0]"	Name	Definition
[63:61]	Reserved	Reserved (may not be 0)
[60]	PIO read	1'b1 for PIO read 1'b0 for PIO write
[59:56]	NCU Credit ID	4-bit Credit from that will eventually return back to NCU for reuse. This is to guarantee that there can only be 16 outstanding PIO transactions as DMU cannot take more than 16.
[55:48]	Byte Count/Byte Mask	This field is directly come from 'size' field of a PCX packet For PIO read: (cannot count on upper 5bit to be zero for read case from PCX packet) 8'bxxxx_x000: 1 Byte 8'bxxxx_x001: 2 Bytes 8'bxxxx_x010: 4 Bytes 8'bxxxx_x011: 8 Bytes 8'bxxxx_x100: 16Bytes For PIO write: 8bit byte mask indicates which of the 8B of store data should be updated
[47:40]	NCU PIO ID	{buf_id[1:0],CPU_thrID[5:0]}
[39:38]	reserved	Must be 0
[37:36]	Command Mapping	See PCIE base/mask CSR registers for mapping details. 2'b11: Mem64 space 2'b10: Mem32 space 2'b01: IO space (if PA[28]=1'b1) 2'b00: Config space (if PA[28]=1'b0)
[35:0]	Bus Address	PA address [35:0]. This address is PCX packet address, masked with mask registers.

7.4.13.5 DMUPIO Read Request Address and Data Format

When CPU sends a non-cacheable external LOAD_REQ request (PIO read request) to NCU via the PCX interface, the packet contains a 40 bit address, request type, request size, etc. Followings are rules for read access to the external PCI space.

1. The 40-bit address is a byte address pointing to the 1st byte CPU is interested in.
2. The most CPU asks for are 16 bytes which is indicated by the PCX packet's size field.
3. DMU/NCU returns 16B to CPU via XBAR (with 64 msb is replicated to 64 lsb for non 16B returns)
4. DMU/NCU do not align the return data.
5. For 2 Byte LOAD_REQ, the fetch should not cross the two Byte boundary

   (i.e. ByteAddress should be 0,2,4,6,8..., case 3 below)
6. For 4 Byte LOAD_REQ, the fetch should not cross the four Byte boundary

   (i.e. ByteAddress should be 0,4,8,..., case four below)
7. For 8 Byte LOAD_REQ, the fetch should not cross the eight Byte boundary

   (i.e. ByteAddress should be 0,8,..., case 5 below)
8. For 16 Byte LOAD_REQ, the fetch should not cross the 16 Byte boundary

   (i.e. ByteAddress should be 0,x10,x20,x30,..., case 6 below)

Example: Focus on the lower 16 bits of the 36-bit address, and assume the upper 24 bits point to the correct PIO region in the following cases.

TABLE 7-34 PIO Read Address and Data Format

Case (in plain English)	Note	Lower 16 bits of PA field in PCX packet	Lower 2 bit of size field in PCX packet	16 Byte return from SIU some time later
Case 1: LOAD_REQ Byte Address 1 for 1 Byte	(Byte Address can be any value for 1 Byte)	16'h0001	3'b000 (means 1 Byte)	{2{64'hxxAB_xxxx_xxxx_xxxx}} (x's means unknown, cpu 'don't care,' and could be anything depends on what DMU is reading)
Case 2: LOAD_REQ Byte Address 3 for 1 Byte	(Byte Address can be any value for 1 Byte)	16'h0003	3'b000	{2{64'hxxxx_xxAB_xxxx_xxxx}}
Case 3: LOAD_REQ Byte Address 4 for 2 Bytes	(Byte Address 1,3,5,7 are illegal)	16'h0004	3'b001 (means 2 Bytes)	{2{64'hxxxx_xxxx_ABCD_xxxx}}
Case 4: LOAD_REQ Byte Address 4 for 4 Bytes	(Byte Address 1~3 or 5~7 are illegal)	16'h0004	3'b010 (means 4 Bytes)	{2{64'hxxxx_xxxx_ABCD_EF01}}
Case 5: LOAD_REQ Byte Address 8 for 8 Bytes	(Byte Address 1~7 or 9~f are illegal)	16'h0008	3'b011 (means 8 Bytes)	{2{64'hABCD_EF01_2345_6789}}
Case 6: LOAD_REQ Byte Address 0x10 for 16 Bytes	(Byte Address 1~f are illegal)	16'h0010	3'b100 (means 16 Bytes)	{64'hABCD_EF01_2345_6789_ 0123_4567_89AB_CDEF} Note: this does not imply any data replication, just all bits are valid and no replication for a 16B load return.

7.4.13.6 DMUPIO Write Request Address and Data Format

When CPU sends a non-cacheable external STORE_REQ (PIO write request) to NCU via the PCX interface, the packet contains a 40 bit byte address, request type, request size, etc. Following are rules for write access to the external PCI space.

1. The 40-bit address from PCX is a byte address pointing to the 1st byte CPU is interested in.
2. The most CPU generated store is eight Bytes which is indicated by the PCX packet's byte mask field.

3. NCU support “partial store” feature by sending the 8bit 'byte mask field', which is position mask, directly to DMU. The 8bit byte mask field can be at any combinations. DMU in OpenSPARC T2 also supports “partial store” feature.
4. NCU does not align any payload data. The 8B payload data is sent to DMU unmodified. The masked 36-bit byte address is sent to DMU with the lower 3 bits PA[2:0] being turn off (always 0) for write only.

Focus on the lower 16 bits of the 40-bit address, and assume the upper 24 bits point to the correct PCI PIO region in the following cases.

TABLE 7-35 PIO Write Address and Data Format

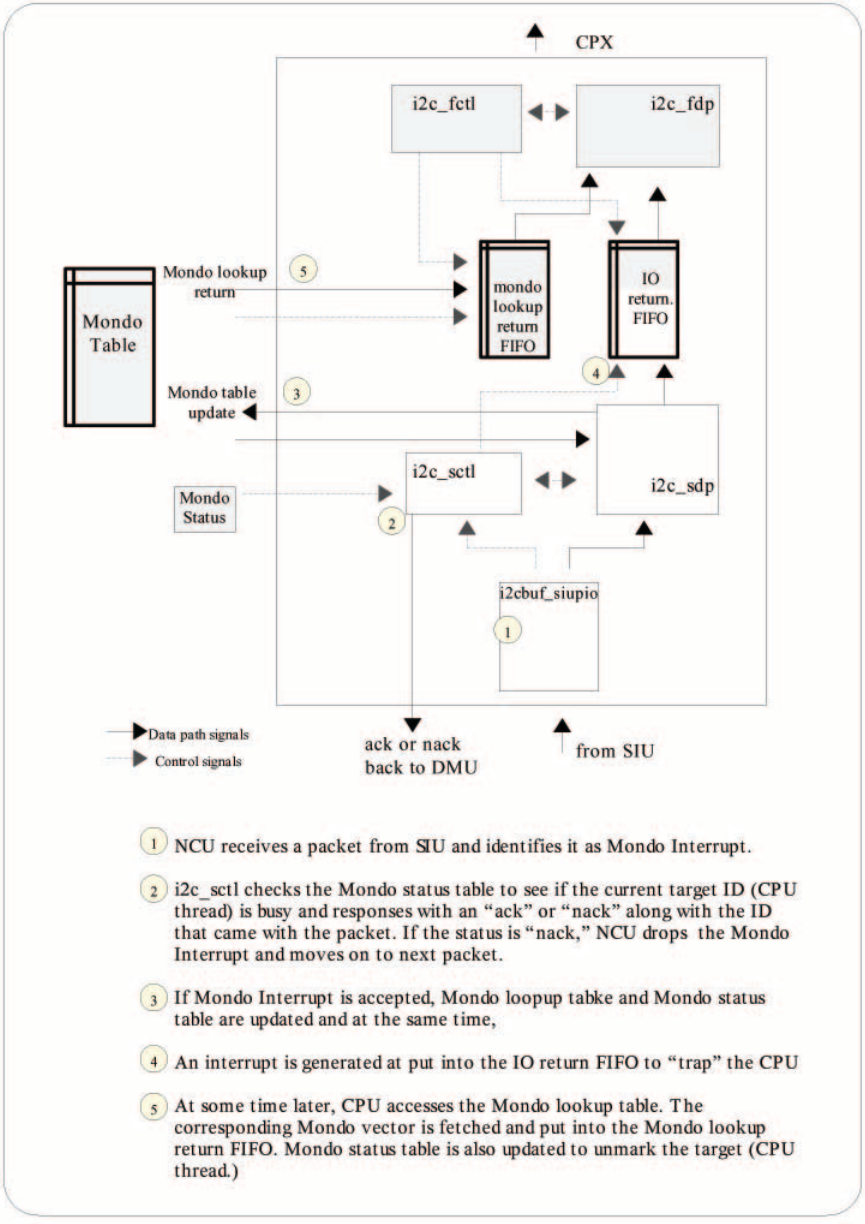
Case (in plain English)	Note	Lower 16 bits of address field from PCX packet	8 bit size field from PCX packet	8 Byte payload from PCX and to DMU
Ex. 1: STORE_REQ Byte Address 1 for 1 Byte	(Byte Address from CPU can be any value pointing to 1st critical Byte)	16'h0001 (sending to DMU as 16'h0000)	8'b0100_0000	64'hxxAB_xxxx_xxxx_xxx (x's means unknown, cpu 'don't care,' and could be anything depends on what DMU is reading)
Ex. 2: STORE_REQ Byte Address 5 for 1 Byte	(Byte Address from CPU can be any value pointing to 1st critical Byte)	16'h0005 (sending to DMU as 16'h0000)	8'b0000_0100	64'hxxxx_xxxx_xxAB_xxxx
Ex. 3: STORE_REQ Byte Address 3 for 2 Bytes (contiguous)	(Byte Address from CPU can be any value pointing to 1st critical Byte)	16'h0003 (sending to DMU as 16'h0000)	8'b0001_1000	64'hxxxx_xxAB_CDxx_xxxx
Ex. 4: STORE_REQ Byte Address 3 for 2 Bytes (non-contiguous)	(Byte Address from CPU can be any value pointing to 1st critical Byte)	16'h0003 (sending to DMU as 16'h0000)	8'b0001_0010	64'hxxxx_xxAB_xxxx_CDxx
Ex. 5: STORE_REQ Byte Address 1 for 4 Bytes (non-contiguous)	(Byte Address from CPU can be any value pointing to 1st critical Byte)	16'h0001 (sending to DMU as 16'h0000)	2'b0101_0011	64'hxxAB_xxCD_xxxx_EF01

7.5 Interrupts

7.5.1 Mondo Interrupt Path (External Interrupts)

All interrupts come from DMU/SIU are treated as Mondo Interrupts. It is originated from DMU but send to NCU via SIU. Mondo Interrupts in NCU, actually, including external devices interrupts and MSI Interrupts. However, NCU does not distinguish between them. FIGURE 7-15 shows detail for the Mondo Interrupt path and control. Since NCU serves Mondo Interrupts in the order of receiving, DMU is guarantee to have Mondo “ack” or “nack” back in the order of sending

FIGURE 7-15 Mondo Interrupt Path



1. NCU receives a packet from SIU and identifies it as Mondo Interrupt.
2. i2c_sctl checks the Mondo status table to see if the current target ID (CPU thread) is busy and responses with an "ack" or "nack" along with the ID that came with the packet. If the status is "nack," NCU drops the Mondo Interrupt and moves on to next packet.
3. If Mondo Interrupt is accepted, Mondo loopup tabke and Mondo status table are updated and at the same time,
4. An interrupt is generated at put into the IO return FIFO to "trap" the CPU
5. At some time later, CPU accesses the Mondo lookup table. The corresponding Mondo vector is fetched and put into the Mondo lookup return FIFO. Mondo status table is also updated to unmark the target (CPU thread.)

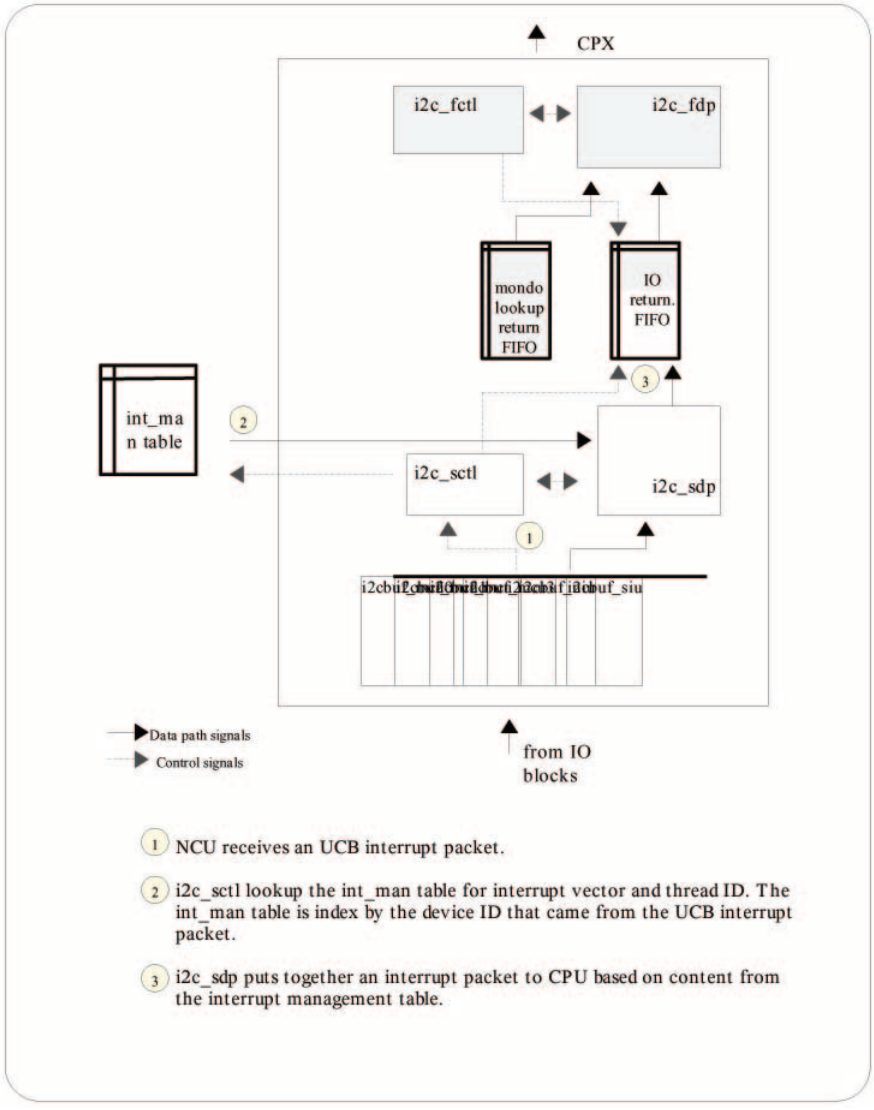
7.5.2 Non Mondo Interrupt (On Chip Interrupt)

The on chip interrupt or Non Mondo Interrupt is identified by its device ID. Each device ID associates with an interrupt source and also index to an entry in Interrupt Management Table (int_man table). This non-mondo interrupt is a “fire-and-forget” type interrupt because once NCU fires the interrupt to the processor, no further information is retained inside NCU. FIGURE 7-15 shows details of the non-mondo type interrupt path, and a list of device ID can be found in TABLE 7-36.

TABLE 7-36 Device ID Assignments

Device ID	Definition
0	Reserved
1	MCU ECC errors, counter rollover, SSI errors
2	SSI Interrupt from EXT_INT_L pin
3 ~ 63	Reserved

FIGURE 7-16 Non Mondo Interrupt Path



7.6 NCU Global Physical Address (PA) Assignments

7.6.1 Global Physical Address Assignments

NCU also serves as the Physical Address parser. Each packet dequeued from the downstream main FIFO carries a 40-bit address, also known as physical address (PA). NCU determines the destination of the packet by examining the eight MSB (bit[39:32]) of the physical address. The address range of each IO subsystem block can be found in TABLE 7-37.

In OpenSPARC T2, CCX (Crossbar) automatically filters out all L2 related packets (L2 non-cacheable CSR and all cacheable packet,) and send them directly to L2. Therefore, packets come from PCX interface to NCU are guaranteed to have bit[39] set to “1'b1.”

TABLE 7-37 Global Physical Address Assignments

MSB Address Range[39:32]	Assignment
0x80	NCU
0x82	Reserved
0x83	CCU
0x84	MCUs [13:12] = 2'b00 for MCU0 [13:12] = 2'b01 for MCU1 [13:12] = 2'b10 for MCU2 [13:12] = 2'b11 for MCU3
0x85	TCU
0x86	DBG
0x87	Reserved
0x88	DMUCSR
0x89	RST
0x90	ASI CPU shared registers
0x91 ~ 0x9F	Reserved

TABLE 7-37 Global Physical Address Assignments *(Continued)*

MSB Address Range[39:32]	Assignment
0xA0 ~ 0xBF	L2 CSR (handles by CCX directly and does not come to NCU)
0xD0 ~ 0xFE	Reserved
0xFF	SSI (boot ROM)

7.6.2 NCU Local CSR Assignments

7.6.2.1 NCU Management

Each device sends its device ID to NCU along with the UCB interrupt packet. The device ID is used to index into the Interrupt Management table.

TABLE 7-38 Interrupt Management – INT_MAN (0x80_0000_0000) (count 128 step 8)

Bit	Name	Initial Value	R/W	Description
[63:14]	Reserved	0	RO	Reserved
[13:8]	CPU	X	RW	CPUID to manage the device
[7:6]	Reserved	0	RO	Reserved
[5:0]	Vector	X	RW	Interrupt Vector

MONDO_INT_VEC performs the identical function for Mondo Interrupts that INT_MAN performs for other IO interrupts, except that the CPU_ID (thread ID) is specified in the Mondo interrupt transaction.

TABLE 7-39 Mondo Interrupt Vector Register – MONDO_INT_VEC (0x80_0000_0a00)

Bit	Name	Initial Value	R/W	Description
[63:6]	Reserved	0	RO	Reserved
[5:0]	Vector	0	RW	Interrupt Vector for Mondo interrupts (encodes bit set in ASI_SWVR_INTR_RECEVIE)

TABLE 7-40 Processor Serial Number – SER_NUM (0x80_0000_1000)

Bit	Name	Initial Value	R/W	Description
[63:44]	sernum2	0	RO	Chip's serial number programmed by eFuse
[43:22]	sernum1	0	RO	Chip's serial number programmed by eFuse
[21:0]	sernum0	0	RO	Chip's serial number programmed by eFuse

This register is warm reset protected.

TABLE 7-41 eFuse Status – EFU_STAT (0x80_0000_1008)

Bit	Name	Initial Value	R/W	Description
[63:0]	efu_status	0xFFFFFFFFFF FFFFFF	RO	eFuse status programmed by eFuse block

This register is warm reset protected.

TABLE 7-42 Core Available – CORE_AVAIL (0x80_0000_1010)

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_avail	0xFFFFFFFFFF FFFFFF	RO	Core available programmed by eFuse This register is same as core_available in ASI

This register is warm reset protected.

The following Bank Available, Bank Enable and Bank Enable Status works the same fashion as the ASI's Core Available, Core Enable and Core Enable status. Bank Available is only programmable by eFuse after POR event, and will not change. This default value is propagated in to Bank Enable register which is programmable at any time after POR event. Finally, the Bank Enable Status register is the one that being used by different clusters/clock disabling. The Bank Enable Status is only updated from Bank enable register at the de-assertion of WMR event.

TABLE 7-43 Bank Available – BANK_AVAIL (0x80_0000_1018)

Bit	Name	Initial Value	R/W	Description
[63:8]	Reserved	**0**	**RO**	Reserved
[7:0]	Bank_avail	**0xFF**	**RO**	Bank available programmed by eFuse This register indicates the availability of each L2 bank.

This register is warm reset protected.

TABLE 7-44 Bank Enable – BANK_ENABLE (0x80_0000_1020)

Bit	Name	Initial Value	R/W	Description
[63:8]	Reserved	0	RO	Reserved
[7:0]	Bank_enable	0xFF	RW	Received initial from Bank Available after POR event. This programed value is reflected onto Bank Enable Status at the de-assertion of WMR event. (note: hardware forces a non-available bank indicated by Bank Available register to 0, so that SW cannot enable a non-available bank)

This register is warm reset protected.

There are certain rules for L2 partial bank mode, refer to Level 2 Cache, System Interface Unit (SIU), Clock Control Unit (CCU) specifications for details. The Bank Enable Status changes only after the de-assertion of WMR event. NCU provides a signal for each bank pair. For example, BA01 means bank 0 and bank 1. BA01 is “0” if bank 0 or bank 1 is disable or unavailable.

Since there are encoding involves between Bank Enable and Bank Enable Status. A preview version of partial bank signals is provided whenever there is a change in Bank Enable register. However, the final/usable copy that provided to different clusters will not be changed until after WMR event.

The PM (partial mode) signal is 1 if any of the bank is not enable. Each of the BA* signal is a result of two bank enable being “ANDED” together. However, there are some illegal combinations of BA* signals, and NCU is the BA* rule enforcer. Refer to Level 2 Cache, System Interface Unit (SIU), Clock Control Unit (CCU) chapters for illegal case details. The following is the illegal case mapping by NCU.

TABLE 7-45 Illegal Case Mapping

Illegal PM,BA67,BA45,BA23,BA01 combinations		Resulting BA67,BA45,BA23,BA01
1,0,0,0,0	-->	1,0,0,0,0 (Bad Chip, no mapping)
1,0,1,1,1	-->	1,0,0,1,1
1,1,0,1,1	-->	1,0,0,1,1
1,1,1,0,1	-->	1,1,1,0,0
1,1,1,1,0	-->	1,1,1,0,0

TABLE 7-46 Bank Enable Status – BANK_ENABLE_STATUS (0x80_0000_1028)

Bit	Name	Initial Value	R/W	Description
[63:13]	Reserved	0	RO	Reserved
[12]	PM_preview	0	RO	L2 partial mode preview value
[11]	BA67_preview	1	RO	BA67 preview value
[10]	BA45_preview	1	RO	BA45 preview value
[9]	BA23_preview	1	RO	BA23 preview value
[8]	BA01_preview	1	RO	BA01 preview value
[7:5]	Reserved	0	RO	Reserved
[4]	PM	0	RO	L2 partial mode (final copy to different clusters)
[3]	BA67	1	RO	Availability of bank 6 and bank 7 (final copy to different clusters)
[2]	BA45	1	RO	Availability of bank 4 and bank 5 (final copy to different clusters)
[1]	BA23	1	RO	Availability of bank 2 and bank 3 (final copy to different clusters)
[0]	BA01	1	RO	Availability of bank 0 and bank 1 (final copy to different clusters)

TABLE 7-47 L2 Index Hash Enable – L2_IDX_HASH_EN (0x80_0000_1030)

Bit	Name	Initial Value	R/W	Description
[63:1]	Reserved	0	RO	Reserved
[0]	L2_Idx_Hash_en	0	RW	L2 indexing enable. New value will not propagate to L2_Idx_hash_en_status until the next wrm_reset.

This register is warm reset protected.

TABLE 7-48 L2 Index Hash Enable Status – L2_IDX_HASH_EN_STATUS (0x80_0000_1038)

Bit	Name	Initial Value	R/W	Description
[63:1]	Reserved	0	RO	Reserved
[0]	L2_Idx_hash_en_status	0	RO	Final/usable copy of l2_index_hash_en to SII and SPC

TABLE 7-49 NCU/SSI SCK clock select – NCU_SCKSEL (0x80_0000_3040)

Bit	Name	Initial Value	R/W	Description
[63:2]	Reserved	0	RO	Reserved
[1:0]	ncu_scksel	0	RW	when “01”, ssi_sck = iol2clk/4 all other cases, ssi_sck = iol2clk/8

This register is warm reset protected.

7.6.2.2 RAS Related Registers

Logics sets ESR bit on the error indication (thus recording the error) if the corresponding bit in the ELE is set. Errors will continue to be recorded until a logged error also has its respective EIE bit set. This causes the NCU to dispatch an “SocError” message using a CPX Error Indication Packet. A “snapshot” of the ESR be taken/stored in the PER register and the ESR cleared.

All CPUTHR ID fields are protected by ECC (SecDed) in NCU's memories. As a general policy, if an uncorrectable error happens at CPUTHR ID, NCU terminates the corrupted packet silent without replying to any of the CPUTHR ID since the CPITHR ID is unknown. TABLE 7-50 shows expected NCU behavior when NCU detected an error.

TABLE 7-50 NCU Response to Error

Error type	Cause of source	Transaction (return packet)	Syndrome reg
NcuDmuCredit	DMUPIO store from PCX interface	When NCU received wack with parity error, NCU drop the wack_tag.	No
NcuCtagCe [23]	1 DMUPIO read return for sii interface	Complete. send load return cpx packet without error	No
	2. MONDO interrupt from sii interface	Complete. send INT cpx packet without error	
NcuCtagUe [22]	1. DMUPIO read return from sii interface	Terminated. Do not send load return CPX packet	Format 2 data (ctag is corrupted)
	2. MONDO interrupt from sii interface	Continue. send INT CPX packet with error but not ack back mondo id	
NcuDataParity [14]	1. DMUPIO read return from SII interface	Continue. Send load return CPX packet with error	Format 2 data
	2. MONDO interrupt from sii interface	terminate, do not send INT CPX packet and not ack back mondo id	
NcuDmuUe[21]	1. DMUPIO store/load (read) from PCX interface	Terminate. Not forward packet to DMUPIO. If cputhr id is corrupted: no response is generated If cputhr id is not corrupted: return without error bit set for store, return with error bit set for load.	Format 1 RCTP=4'hf RCTP data match with PCX packet

TABLE 7-50 NCU Response to Error *(Continued)*

Error type	Cause of source	Transaction (return packet)	Syndrome reg
NcuCpxUe [20]	1. PIO/CSR store from PCX interface.	Continue transfer packet to target. Don't send store ack return CPX packet. (EJR write can't be affected) \	No
	2. Load return/IO interrupt from IOs	Terminate. Don't send load return/Interrupt CPX packet.	
NcuPcxUe [19]	PIO/CSR load/store form PCX interface	Terminate. Don't pass done packet to target. Don't send store ack return CPX packet	Format 1 RCTP=4'h1011 RCTP data match with PCX packet (data is corrupted data)
NcuPcxData[18]	PIO/CSR Load from PCX interface	Continue (read)	Format 1 RCTP = 4'h0110 RCTP data match with PCX packet
	PIO/CSR store from PCX interface	Terminated. (write) Ack back to CPU without error.	
NcuIntTable [17]	Load INT table from PCX interface	Continue, send load ack return CPX packet with error	Format 1 RCTP = 4'hf RCTP data match with PCX packet
	IO interrupts from IO (NIU/MCU/SSI) interface	terminate. Don't' send INT CPX packet	format 1 RCTP = 4'h6 RCTP data match with interrupt table (corrupted)
NcuMondoFifo[16]	load MONDO table from PCX interface	terminate. Don't sent ack return CPX packet	
NcuMondoTable[15]	Load the MOND table	Continue, send load ack return CPX packet with error	

TABLE 7-51 Error Status Register - ESR (0x80_0000_3000)

Bit	Name	Initial Value	R/W	Description
[63]	valid	0	RW	Valid: indicates that any error or multiple error has been recorded.
[62:41]	Reserved	0	RO	Reserved
[42]	NcuDmuCredit	0	RW	Credit token to NCU for DMU pio write credits

TABLE 7-51 Error Status Register - ESR (0x80_0000_3000) *(Continued)*

Bit	Name	Initial Value	R/W	Description
[41]	Mcu3Ecc	0	RW	MCU3 ECC Correctable (exceeded data CE threshold)
[40]	Mcu3Fbr	0	RW	MCU3 Fbdimm Recoverable
[39]	SpareBit[4]	0	RW	This bit is always set to 0 and does not capture anything regardless of EJR settings (see note below table)
[38]	Mcu2Ecc	0	RW	MCU2 ECC Correctable (exceeded data CE threshold)
[37]	Mcu2Fbr	0	RW	MCU2 Fbdimm Recoverable
[36]	SpareBit[3]	0	RW	This bit is always set to 0 and does not capture anything regardless of EJR settings (see note below table)
[35]	Mcu1Ecc	0	RW	MCU1 ECC Correctable (exceeded data CE threshold)
[34]	Mcu1Fbr	0	RW	MCU1 Fbdimm Recoverable
[33]	SpareBit[2]	0	RW	This bit is always set to 0 and does not capture anything regardless of EJR settings (see note below table)
[32]	Mcu0Ecc	0	RW	MCU0 ECC Correctable (exceeded data CE threshold)
[31]	Mcu0Fbr	0	RW	MCU0 Fbdimm Recoverable
[30]	SpareBit[1]	0	RW	This bit is always set to 0 and does not capture anything regardless of EJR settings (see note below table)
[29]	NiuDataParity	0	RW	Data Parity error in the DMA read return from the SIO
[28]	NiuCtagUe	0	RW	Ctag double bit Uncorrected error from the SIODMA read return
[27]	NiuCtagCe	0	RW	Ctag single bit Corrected Error from the SIO DMA read return
[26]	SioCtagCe	0	RW	Ctag single bit Corrected Error after the OLD Fifo.
[25]	SioCtagUe	0	RW	Ctag double bit Uncorrected Error from the OLD Fifo. Recommended Fatal Error
[24]	SpareBit[0]	0	RW	(Does not capture anything)
[23]	NcuCtagCe	0	RW	Ctag single bit Corrected error on Interrupt write or PIO read return.
[22]	NcuCtagUe	0	RW	Cag double bit error on Interrupt write or PIO read return. Recommended Fatal Error. (NCUSYN)
[21]	NcuDmuUe	0	RW	For IOMMU conflicts. (NCUSYN)
[20]	NcuCpxUe	0	RW	Error in Output Fifo to CPX.

TABLE 7-51 Error Status Register - ESR (0x80_0000_3000) *(Continued)*

Bit	Name	Initial Value	R/W	Description
[19]	NcuPcxUe	0	RW	CPU PIO/CSR commands, may be Fatal. (NCUSYN)
[18]	NcuPcxData	0	RW	Error in CPU PCX Fifo. (NCUSYN)
[17]	NcuIntTable	0	RW	Error in NCU read of Interrupt table. (NCUSYN)
[16]	NcuMondoFifo	0	RW	Parity/ECC error in read of Mondo Fifo
[15]	NcuMondoTable	0	RW	Parity/ECC error in CPU read Mondo table
[14]	NcuDataParity	0	RW	Parity for Interrupt write or PIO read return from the SIO. (NCUSYN)
[13]	DmuDataParity	0	RW	Data Parity error in the DMA read return from the SIO
[12]	DmuSiiCredit	0	RW	Parity error in the DMA write acknowledge Credit from the SII. Recommended Fatal Error
[11]	DmuCtagUe	0	RW	Ctag double bit Uncorrected error from the SIO DMA read return. Recommended Fatal Error
[10]	DmuCtagCe	0	RW	Ctag single bit Corrected error from the SIO DMA read return
[9]	DmuNcuCredit	0	RW	Parity error in the PIO read/Mondo acknowledge Credit from the NCU Recommended Fatal Error
[8]	DmuInternal	0	RW	Recommended Fata Error
[7]	SiiDmuAparity	0	RW	Parity error for Address field for DMA transactions from DMU Fifo. (SIISYN)
[6]	SiiNiuDParity	0	RW	Data parity error for DMA writes from DMU Fifo. (SIISYN)
[5]	SiiDmuDParity	0	RW	Data parity error for DMA writes from DMU Fifo. (SIISYN)
[4]	SiiNiuAParity	0	RW	Parity error fro Address field for DMA transactions from NIU Fifo. (SIISYN)
[3]	SiiDmuCtagCe	0	RW	Ctag single bit Corrected error, in transaction from NIU Fifo
[2]	SiiNiuCtagCe	0	RW	Ctag single bit Corrected error in transaction from NIU Fifo
[1]	SiiDmuCtagUe	0	RW	Ctag double bit Uncorrected ECC or Command Parity Error in transaction from DMU Fifo. (SIISYN)
[0]	SiiNiuCtagUe	0	RW	Ctag double bit Uncorrected ECC or Command Parity Error in transaction from NIU Fifo. (SIISYN)

This register is warm reset protected.

Note – Bit[30], [33], [36], and [39] in the ESR register does not capture anything from hardware point of view. Even though the corresponding EJE is set, these bit still capture 0. However, software can set these bit to 1 and cause SOC interrupt or fatal error for testing purposes.

ELE provides the capability to select individual error events to be logged in the ESR. If a 'Log Enable' bit is set, and the corresponding error type signal is asserted, then the respective bit position in the ESR's “Recorded Error Type” field is set.

TABLE 7-52 Error Log Enable - ELE (0x80_0000_3008)

Bit	Name	Initial Value	R/W	Description
[63:43]	Reserved	0	RO	Reserved
[42:0]	Error Log Enable	0x7FFFFFFFFF	RW	1-to-1 corresponding to ESR register

This register selects individually logged errors to dispatch an SocError message. Each bit enables interrupting (dispatching SOCError) for the respective bit position in the ESR. Interrupts may be sent if no other SOCError is still pending as indicated by the PER valid bit=1. Thus, if no pending SOCError (i.e. PER valid=0) and the respective “Interrupt Enable” bit is set, the SOCError indication will be dispatched for logged errors at this bit location.

TABLE 7-53 Error Interrupt Enable - EIE (0x80_0000_3010)

Bit	Name	Initial Value	R/W	Description
[63:43]	Reserved	0	RO	Reserved
[42:0]	Error Interrupt Enable	0	RW	1-to-1 corresponding to ESR register

This register provides the capability to select individual error checking nodes to have their parity/ECC bits flipped. When the respective bit is set, the parity/ECC will be flipped, causing on error, for this particular parity/ECC checking location.

TABLE 7-54 Error Injection Register - EJR (0x80_0000_3018)

Bit	Name	Initial Value	R/W	Description
[63:43]	Reserved	0	RO	Reserved
[42:0]	Error Injection Enable	0	RW	1-to-1 corresponding to ESR register with exception of bit[30],[33],[36], and [39], which the corresponding ESR bit always capture 0 regardless of EJR is set.

Each error type may be programmed to cause a Fatal Error This register enables an error to cause the signal “ncu_rst_fatal_error” to be asserted to the Reset Unit. If the respective “Fatal Error Enable” bit is set, and the corresponding error type is asserted, a fatal error will be dispatched to the Reset Unit. This functionality is not dependent on the settings of the ESR, PER, ELE or EIE.

TABLE 7-55 Fatal Error Enable - FEE (0x80_0000_3020)

Bit	Name	Initial Value	R/W	Description
[63:43]	Reserved	0	RO	Reserved
[42:0]	Fatal Error Enable	0	RW	1-to-1 corresponding to ESR register

This register is a snapshot copy of the entire 64-bit of the ESR. This “snapshot” is taken when NCU initiates an SOC Error packet dispatch. This is caused when an error type occurs that has both the respective “log” enable and respective “interrupt” enable bit positions set. After an SOC Error message, the thread's trap handler may read this register to determine the error “cause”. When this register's Valid bit is set further SOCError message dispatches are disabled.

TABLE 7-56 Pending Error Register - PER (0x80_0000_3028)

Bit	Name	Initial Value	R/W	Description
[63:0]	(same as ESR)	(same as ESR)	(same as ESR)	(same as ESR)

This register is warm reset protected.

The SII Error Syndrome Register stores the syndrome (header) information from an SII caused error event. This register is located in NCU. Data is sent to NCU from SII on a special 4-bit serial bus (refer to Appendix B). When the logging is disable for this error type, NCU will simply ignore the data syndrome data transfer from SII. In

this case, it will retain the prior data already stored in the SIISYN register. If bit[63],"Valid"-bit, is already set, NCU will ignore further SIISYN coming from SII until software clears this bit.

TABLE 7-57 SII Error Syndrome - SIISYN (0x80_0000_3030)

Bit	Name	Initial Value	R/W	Description
[63]	valid	0	RW	valid
[62:59]	Reserved	0	RO	Reserved
[58:56]	Etag	0	RW	Indicates which type of error is associated with this syndrome. This is the lower 3-bit of the error type index in ESR (SII error types only limited to bit7-bit0). For example, the "Etag" of a "SiiNiuAParity" error = 4.
[55:40]	Ctag	0	RW	16-bit CTAG or ID field from SII or DMU/SII header
[39:0]	PA	0	RW	40 physical address

This register is warm reset protected.

TABLE 7-58 NCU Error Syndrome - NCUSYN (0x80_0000_3038)
If bit[62] is 0: format 1

Bit	Name	Initial Value	R/W	Description
[63]	Valid	0	RW	Valid
[62]	Format=0	0	RW	Format 0
[61:58]	RCTP	0	RW	Rqtyp, Cpu, Thr ,PA valid
[57:56]	Reserved	0	RO	Reserved
[55:51]	etag	0	RW	Which bit in ncuesr causes loading of syndrome
[50:46]	Rqtyp	0	RW	Packet request type
[45:43]	Cpu_id	0	RW	CPU ID
[42:40]	Thr_id	0	RW	Thread ID
[39:0]	PA	0	RW	40bit PA

This register is warm reset protected.

TABLE 7-59 NCU Error Syndrome - NCUSYN (0x80_0000_3038)
If bit[62] is 1

Bit	Name	Initial Value	R/W	Description
[63]	Valid	0	RW	Valid
[62]	Format=1	0	RW	Format 1
[61:58]	Reserved	0	RW	Rqtyp,Cpu,Thr,PA valid
[57:56]	Reserved	0	RO	Reserved
[55:51]	etag	0	RW	Which bit in ncuesr causes loading of syndrome
[50:46]	Reserved	0	RW	Packet request type
[45:43]	Reserved	0	RW	CPU ID
[42:40]	Reserved	0	RW	Thread ID
[39:0]	CTAG	0	RW	{24'b0,ctag[5:0]}

This register is warm reset protected.

TABLE 7-60 DBG1 Error Event Trigger Enable - NCU_CREG_DBGTRIG_EN (0x80_0000_4000)

Bit	Name	Initial Value	R/W	Description
[63:1]	Reserved	0	R/O	reserved
[0]	dbgtrigen	0	R/W	Enable dbg1 error event trigger.

This register is warm reset protected.

7.6.2.3 Mondo Table Access

The following register are used to manage the Mondo Interrupts.

When NCU receives a Mondo interrupt, it sets the Busy bit and ack DMU. When a Busy bit is set, it means an interrupt is waiting to be serviced or is being serviced. Software needs to reset the Busy bit after it completes servicing the interrupt. If the Busy bit is already set when an interrupt arrives at NCU, a NACK will be sent back to DMU. The Busy bit is set after a reset and software has to clear it to begin receiving interrupts.

There are two Mondo Interrupt Mondo Tables. The tables are read-only by software and the entries are updated by DMU Mondo interrupts, provided that corresponding Busy bit is not currently set. NCU will ack the interrupt if it is not busy, otherwise the NCU will NACK it.

TABLE 7-61 Mondo Interrupt Data0 – MONDO_INT_DATA0 (0x80_0004_0000) (Count 64 Step 8)

Bit	Name	Initial Value	R/W	Description
[63:0]	Data0	X	RO	First 64 bits of Mondo interrupt data

TABLE 7-62 Mondo Interrupt Data1 – MONDO_INT_DATA1 (0x80_0004_0200) (Count 64 Step 8)

Bit	Name	Initial Value	R/W	Description
[63:0]	Data1	X	RO	Second 64 bits of Mondo interrupt data

When a thread reads the following alias register, it is reading its own entry in the Mondo Data0 table (i.e. The PA will from PCX bus will be ignored, and the cputhr[5:0] will be used for accessing the table entry.) This is designed for a CPU thread accessing its own entry without doing address calculation or knowing its own cpu thread I.D. If access if from JTAG the cputhr[5:0] in UCB packet will be used for table indexing.

TABLE 7-63 Alias Mondo Interrupt Data0 – MONDO_INT_ADATA0 (0x80_0004_0400)

Bit	Name	Initial Value	R/W	Description
[63:0]	Data0	X	RO	First 64 bits of Mondo interrupt data

When a thread reads the following alias register, it is reading its own entry in the Mondo Data1 table (i.e. The PA will from PCX bus will be ignored, and the cputhr[5:0] will be used for accessing the table entry.) This is designed for a CPU thread accessing its own entry without doing address calculation or knowing its own cpu thread I.D. If access if from JTAG the cputhr[5:0] in UCB packet will be used for table indexing.

TABLE 7-64 Alias Mondo Interrupt Data1 – MONDO_INT_ADATA1 (0x80_0004_0600)

Bit	Name	Initial Value	R/W	Description
[63:0]	Data1	X	RO	Second 64 bits of Mondo interrupt data

TABLE 7-65 Mondo Interrupt Busy – MONDO_INT_BUSY(0x80_0004_0800) (Count 64 Step 8)

Bit	Name	Initial Value	R/W	Description
[63:7]	Reserved	0	RO	Reserved
[6]	Busy	1	RW	Hardware set Busy to "1" when an interrupt is received. Hardware nacks an incoming Mondo interrupt if Busy bit is already set.
[5:0]	Reserved	0	RO	Reserved

When a thread reads the following alias register, it is reading its own entry in the Mondo Busy table (i.e. The PA will from PCX bus will be ignored, and the cputhr[5:0] will be used for accessing the table entry.) This is designed for a CPU thread accessing its own entry without doing address calculation or knowing its own cpu thread I.D. If access if from JTAG the cputhr[5:0] in UCB packet will be used for table indexing.

TABLE 7-66 Alias Mondo Interrupt Busy – MONDO_INT_ABUSY(0x80_0004_0a00)

Bit	Name	Initial Value	R/W	Description
[63:7]	Reserved	0	RO	Reserved
[6]	Busy	1	RW	Hardware set Busy to "1" when an interrupt is received. Hardware nacks an incoming Mondo interrupt if Busy bit is already set.
[5:0]	Reserved	0	RO	Reserved

7.6.3 ASI Registers

The ASI registers could be accessible by both JTAG and core. The algorithm for mapping from ASI address to IO address is as follows:

PA[39:32] = 0x90

PA[31:29] = core_id[2:0] (physical core id)

PA[28:26] = tid[2:0] (thread id)

PA[25:18] = asi[7:0]

PA[17:3] = VA[17:3]

PA[2:0] = 000

If it's a register that is shared by all virtual cores, then the core_id, PA[31:29] and thread_id, PA[28:26] are ignored. NCU always decode only PA [25:0] if PA[39:32]= 0x90.

7.6.3.1 Core Available Register – ASI_CORE_AVAILABLE (0x90_0104_0000)

(ASI:41 VA:00)

This register is programmed by eFuse controller after POR is deasserted. NCU will detect the de-assertion of efu_ncu_coreavail_dshift signal which triggers update to Core Enable, Core Enable Status and XIR Steering registers. The granularity of the fuses is at each physical core level, and there are eight core in OpenSPARC T2. Therefore, physically there are only eight bits for this register. Hardware automatically expands each bit (representing a core) to eight bits and becomes a total of 64 bit representing 64 threads.

- JTAG accessible (RO)

TABLE 7-67 Core Available Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_available	0xFFFFFFFFFFFFFFFF(by POR)	RO	A one means the thread is available

This register is warm reset protected.

7.6.3.2 Core Enable Status Register – ASI_CORE_ENABLE STATUS (0x90_0104_0010)

(ASI:41 VA:10)

The Core Enable Status Register is updated from Core Enable register at the deassertion of “warm reset”, or from Core Available register at de-assertion of efu_ncu_coreavail_dshift signal (after POR deasserted). JTAG could program the Core Enable register after POR and before the “warm reset,” so that Core Enable Status register takes the value of Core enable at the next “warm reset” deassertion.

Hardware implements only 8-bit for this register. When SW reads, NCU automatically expands each bit to 8-bit wide and becomes 64 bits total to represent 64 threads. In OpenSPARC T2, CPU uses the value of this register to gate off the clock to the appropriate physical core.

- JTAG accessible (RO)

- A thread that is not available in the Core Available register must have its corresponding status bit set to 0 by hardware.

TABLE 7-68 Core Enable Status Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_enable_status	0xFFFFFFFFFFFFFFF	RO	A one means the thread is currently enabled

7.6.3.3 Core Enable Register – ASI_CORE_ENABLE (0x90_0104_0020)

ASI:41 VA:20

This register is first update after POR (actually at the assertion of efu_ncu_coreavail_dshift) based on Core Available register. When SW uses this register to enable/disable a core or thread, the effect of programming this register will take place only after the following "warm reset."

Hardware implements only eight bits, representing eight cores for this register. When reading, NCU expands each bit into eight bits and becomes a total of 64 bits, representing 64 threads. When writing, NCU ANDed eight corresponding bits to a physical core to reduce the 64 bits signals down to eight bits which representing eight cores.

- JTAG accessible: RW
- Bits corresponding to the same core is ANDed together by NCU before writing into the register. So, if one thread is being disabled, all threads within the same physical core are also being disabled.
- Hardware forces all threads in an unavailable core's thread (based on Core Available register) to be disabled.
- Hardware enforces "no all-core-disabled" rule to protect the situation that all cores are disabled by SW or by JTAG. If JTAG writes all 0s to this register, NCU will set the lowest available core (based on Core Available register) to 1. If CPU writes all 0 to this register, NCU will keep the bit corresponding to CPU that initiates the command to 1. A disabled/unavailable thread (basing on core available and core enable status registers) should never access this register. Unpredictable hardware behavior will be resulted in such case.

TABLE 7-69 Core Enable Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_enable	0xFFFFFFFFFFFFFFF (by POR)	R/W	A one means the thread will be enable following the next "warm reset"

This register is warm reset protected.

7.6.3.4 XIR Steering Register – ASI_XIR_STEERING (0x90_0104_0030)

(ASI:41 VA30)

SW can program which thread gets XIR when XIR pin is asserted. SW can program this register such that all threads, a subset of threads, a thread, or none of the threads will get XIR.

XIR Steering register first receives a default value based on Core available register after POR (actually at deassertion of efu_ncu_coreavail_dshift). At each deassertion of "warm reset," XIR Steering register gets new default value basing on Core Enable register which could be programmed by SW or JTAG.

- JTAG accessible RW
- If a core is not enable, all corresponding bits in XIR Steering register are force to 0 by hardware.

TABLE 7-70 XIR Steering Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Xir_steering	0xFFFFFFFFFFFFFF	R/W	A one means the thread will receive a "reset" interrupt when XIR external pin is asserted

7.6.3.5 Core Running RW Register –ASI_CORE_RUNNING_RW(0x90_0104_0050)

(ASI:41 VA:50)

SW uses this register to park or unpark a thread. Each bit position corresponds to a thread. If the bit is set to 1, the thread is running. When set to 0, the thread is parked. A parked thread stops execute new instructions and will not initiate transaction except in response to a coherency transaction initiated by other threads. It could take arbitrarily long from the time this register is programmed to the thread is actually parked or unparked.

Upon "warm reset," this register is set to all 0. When NCU receives the rst_ncu_wake_thread signal from RST cluster, NCU will set the lowest available thread bit to 1 based on Core Enable Status register. This thread becomes the master thread. Privileged software, running on the master thread, will subsequently write to

this register to unpark other threads. It is up to software to perform the initialization that are required by thread upon unparking. There are 3 ways to program this register:

1. Writing directly to Core Running RW register
2. Alternatively, SW can write a 1 to the corresponding thread bit in Core Running W1S register. This results in setting the corresponding bit in Core Running RW register to 1.
3. Alternatively, SW can write a 1 to the corresponding thread bit in Core Running W1C register. This results in clearing the corresponding bit the Core Running RW register to 0.

- JTAG accessible RW
- Hardware forces all unavailable or disabled threads to be parked (base on Core Enable Status register) Writing 1 into the disabled thread bits will have no effect.

Only JTAG is able to park all threads during debug by writing all '0' to core_running register. Other than JTAG, hardware enforce “no all-thread-parked” rule. When core write to core_running register to park all threads, the hardware will keep the requesting thread unparked. A disabled/unavailable thread (basing on core available and core enable status registers) or a parked thread should never access this register. Unpredictable hardware behavior will be resulted in such case.

TABLE 7-71 Core Running RW Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_running RW	0x1 (by POR and WMR)	R/W	A one means the thread is being unparked. A zero means the thread is current park or disabled. The status is reported in Core_running_status register

7.6.3.6 Core Running Status Register – ASI_CORE_RUNNING_STATUS (0x90_0104_0058)

(ASI:41 VA:58)

Each SPC thread will send spc_core_running_status to indicate its status. The SPC thread determines the status of each thread by the following criteria. The SPC thread receives a request to park or unpark the based upon a '1' to '0' or '0' to '1' transition on the ncu_spc_core_running signal from NCU. An indeterminate time later, once all activity for that thread has been processed (the store buffer is empty, any pending cache and TLB misses have bee processed, and all instructions have completed execution), the SPC will drive the spc_cmp_core_running_status signal to a '0' (to signal the thread is parked) or to a '1' (to signal the thread is running). Upon “warm reset,” Core Running Status should be all 0s.

TABLE 7-72 Core Running Status Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_running_status	0x1 (by POR and WMR)	RO	A one means the thread is currently running. A zero means the thread is currently parked or disabled.

7.6.3.7 Core Running W1S Register – ASI_CORE_RUNNING_W1S (0x90_0104_0060)

(ASI:41 VA:60)

TABLE 7-73 Core Running W1S Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_running_W1S	N/A	WO	Write one to a bit will cause the corresponding bit in core_running_rw register to be set to a one.

7.6.3.8 Core Running W1C Register – ASI_CORE_RUNNING_W1C (0x90_0104_0068)

(ASI:41 VA:68)

TABLE 7-74 Core Running W1C Register

Bit	Name	Initial Value	R/W	Description
[63:0]	Core_running_W1C	N/A	WO	Write one in a bit will cause the corresponding bit int core_running_rw register to be cleared to a zero. Write zero means no change on the bit.

7.6.3.9 Interrupt Vector Dispatch Register – INT_VEC_DISP (0x90_01CC_0000)

(ASI:73 VA:00)

A thread may write to the following register to trigger an interrupt to another thread. NCU will generate an interrupt packet and send to a targeted CPU Thread specified in CPU_TH[5:0]. TCU may also write into this register to generate an interrupt to a specific CPU Thread.

TABLE 7-75 Interrupt Vector Dispatch Register

Bit	Name	Initial Value	R/W	Description
[63:14]	Reserved	0	RO	Reserved (NCU ignores write to these bits)
[13:8]	Thread	0	WO	CPU_TH[5:0]
[7:6]	Reserved	0	RO	Reserved (NCU ignores write to these bits)
[5:0]	Vector	0	WO	Interrupt Vector (encodes bit set in ASI_SWVR_INTR_RECEVIE)

7.6.3.10 RAS Error Steering Register – RAS_ERR_STEERING (0x90_0104_1000)

(ASI:41 VA:1000)

This register stores the virtual core ID (VCID), which is used by NCU to determine the error thread target of socerror messages. This six-bit cpuid + threadid will be included in cpx packet. Refer to RAS Related Registers

TABLE 7-76 RAS Error Steering Register

Bit	Name	Initial Value	R/W	Description
[63:6]	reserved	0	RO	Reserved.
[5:0]	VCID	0	RW	cpuID+threadID for target error thread location.

This register is warm reset protected.

7.6.3.11 ASI CMP Tick Enable Register – ASI_CMP_TICK_ENABLE(0x90_0140_0038)

(ASI:41 VA:38)

This register is used to synchronize the TICK register of all physical cores. Refer to Op*enSPARC T2 Programmer's Reference Manual*.

TABLE 7-77 ASI CMP Tick Enable Register

Bit	Name	Initial Value	R/W	Description
[63:1]	reserved	0	RO	Reserved.
[0]	tick_enable	0	RW	Set to '1' to enable incrementing of TICK registers in all physical cores.

Its value is preserved across warm reset.

7.6.3.12 ASI Warm Reset Vector Mask Register – ASI_WMR_VEC_MASK(0x90_0114_0018)

(ASI:45 VA:18)

When this register is set to '1' by software, POR, WMR or DBR will be able to be directed to RAM, at location (0x000000020).

TABLE 7-78 ASI Warm Reset Vector Mask Register

Bit	Name	Initial Value	R/W	Description
[63:1]	reserved	0	RO	Reserved.
[0]	Wmr_vec_mask	0	RW	Send to TCU for wmr protect.

Its value will be preserved during warm reset.

Note that POR and WMR are events, not signals.

7.7 Appendix A

SSI Software Interface

Addresses within the SSI address range (0xFF_F000_0000 to 0xFF_FFFF_FFFF) are issued to the off-chip SSI interface bus. The only transactions that are supported directly to the SSI interface 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 is explicitly always big-endian, all accesses to the SSI interface bus are treated as big-endian.

1. SSI Register Interface The SSI registers all deal with error handling, so are described in the *OpenSPARC T2 Programmer's Reference Manual*.

2. SSI Error Handling TABLE 7-79 describes the SSI's handling of errors. The error indication on read returns is delivered regardless of the ERREN bit, where it is up to the processor to ignore the error or receive it. Logging the error and sending an error interrupt are controlled by the ERREN bit. Note that returning zeros on an I-fetch timeout will tend to cause an illegal instruction trap.

TABLE 7-79 SSI Error Handling

Error	TType	Severity	Logs	Returns	ERREN
SSI Parity Error	Read	Uncorrectable	Just the bit	Data with error indication	Async Intr
SSI Parity Error	Write	Uncorrectable	Just the bit	N/A	Async Intr
SSI Timeout	Read	Uncorrectable	Just the bit	All Zeros with error indication	Async Intr
SSI Timeout	Write	Uncorrectable	Just the bit	N/A	Async Intr

3. SSI Interrupts

   SSI generates interrupts for two reasons: either the EXT_INT_L pin was asserted, or an error was detected.

   The external interrupt pin is intended to be used by the FPGA, and has NO ordering protection, meaning when EXT_INT_L is asserted, an interrupt is issued to the IOB, without checking any transactions in flight. The interrupt is delivered to the IOB using the SSI device ID, i.e. (device ID == 2).

   EXT_INT_L is treated as an asynchronous input, meaning the JBI must synchronize it to its internal clock before using it. Also, EXT_INT_L is treated as an edge-triggered interrupt, meaning that JBI will detect a rising edge on the synchronized signal, and issue an interrupt to the IOB on those rising edges. If the

actual use is level-sensitive, software is responsible for querying the FPGA device (or whatever is driving EXT_INT_L), to see if the interrupt is still asserted, at the end of the interrupt handler.

To guarantee being seen, EXT_INT_L must be asserted for at least 4.5 JBUS cycles.

Error interrupts, when enabled, are delivered to the IOB using the error device ID, (device ID == 1).

4. SSI Interface The Serial System Interface (SSI) is defined for to allow microprocessors to access peripherals in a low pin count fashion. The OpenSPARC T2 chip will not directly interface to peripherals but instead will provide a interface that can be easily converted to peripheral protocols by an external Programmable Logic Device (PLD). Isolating the OpenSPARC T2 chip from these peripherals allows the devices to use higher voltage signalling and provides a mechanism for protocol conversion.

   For the purposes of this discussion, some assumptions of the environment will be made. The JBUS will be assumed to run at 200 MHz nominally, although the actual frequency could be somewhat less than 200. In addition the OpenSPARC T2 chip is assumed to interface to either a CPLD or a more complex FPGA. In the former case, the CPLD may just interface to a Flash PROM. In the latter case, the FPGA may include peripherals of its own (e.g. RS232 UART or system management microprocessor) and have a dedicated parallel (8-bit or wider) interface to Flash ROMs and potentially SRAMs. All of these peripherals would be memory mapped into the 256 Megabyte SSI addressable location area (FF_F000_0000 FF_FFFF_FFFF). All devices accessible off the SSI interface will be only targets OpenSPARC T2 will always be the master of the bus.

5. Functional Interface

   The SSI interface includes three pins: SSI_SCK (clock), SSI_MOSI (master out/slave in), and SSI_MISO (master in/slave out). SSI_CLK and SSI_MOSI are outputs of OpenSPARC T2, and SSI_MISO is an input. The SSI_SCK is a free running clock, toggling whenever the on chip JBUS clock is toggling. It is assumed to be nominally 50 Mhz, but is always a divide by four or eight of the JBUS clock.

6. SSI Request

   An SSI request is transmitted on the SSI_MOSI line. It can be either a read command or a write command. The format of all these requests is one start bit, 3 bit command (CMD[2:0]), a 28 bit address, 0-64 bits of data, and a parity bit. The high order (most significant) bit within the command, address and data are always transmitted first, with the low order bit transferred last. Zeros are transmitted as a low voltage value and ones are transmitted as a high value. A start bit is a high value.

   CMD[2] is 0 for write, 1 for read

   CMD[1:0] encodes the transaction size as follows:

- 2'b00 - 1 byte
- 2'b01 - 2 byte
- 2'b10 - 4 byte
- 2'b11 - 8 byte

For every SSI request, a SSI response is expected. A succeeding request can not be sent until the preceding request has had a response. (No command pipelining is supported.)

When OpenSPARC T2 has no request to transfer or is waiting for a response, the SSI_MOSI line is held in the low voltage state.

The parity bit is set such that the number of 1s in the start bit, the command, the address, any data bits, and the parity bit is an even number.

7. SSI Response

An SSI response is received on the SSI_MISO line. It can be either a read response which must contain data or a write response which must contain no data. The format of a read response is one start bit, 8-64 data bits, and one parity bit. The format of a write response is one start bit and one parity bit. The high order (most significant) bit within the data are always transmitted first, with the low order bit transferred last. Zeros are transmitted as a low voltage value and ones are transmitted as a high value. A start bit is a high value.

The parity bit is set such that the number of 1s in the start bit, any data bits, and the parity bit is an even number. This means a write response is two 1's in consecutive cycles.

When the target has no response to transfer or is processing a request, the SSI_MISO line is held in the low voltage state.

Electrical Interface

The SSI_SCK, SSI_MOSI, SSI_MISO, and EXT_INT_L signals will be HSTL signals at 1.5V. Care must be taken on the input so that overshoot doesn't exceed the 1.5V VDD for long enough to induce gate oxide breakdown for the CO27.C process. (See the signal ERS for voltage levels and currents.)

When driving OpenSPARC T2 will drive SSI_MOSI for 3 JBUS cycles prior to a SSI_SCK rising edge and hold SSI_MOSI for one JBUS cycle after the SSI_SCK rising edge. When receiving, OpenSPARC T2 will wait 3 JBUS cycles after a rising SSI_SCK edge to sample the input line.

7.8 Appendix B

The following is the SII/NCU interface data format which results in the SIISYN syndrome register.

siisyn_data[63:0] comes from SII to NCU, 4-bit at a time (see following timing diagram),

starting 1st transfer in bit[3:0], then bit[7:4], and so on

```
Siisyn_data[39:0] = PA,

siisyn_data[55:40] = ctag,

siisyn_data[61] = niud_pe,

Siisyn_data[60] = niua_pe,

siisyn_data[59] = niuctag_ue,

siisyn_data[58] = dmud_pe,

Siisyn_data[57] = dmua_pe,

siisyn_data[56] = dmuctag_ue,
```

NCU will encode siisyn_data[61:56] to 3-bit Etag, siisyn[58:56] as in TABLE 7-80

TABLE 7-80 SII/NCU Interface Data Format

siisyn_data[61]	"000001"	"000001"	"000001"	"000001"	"000001"	"000001"
Etag[2:0]	"001"	"111"	"101"	"000"	"100"	"110"

FIGURE 7-17 SII to NCU Error Strobe

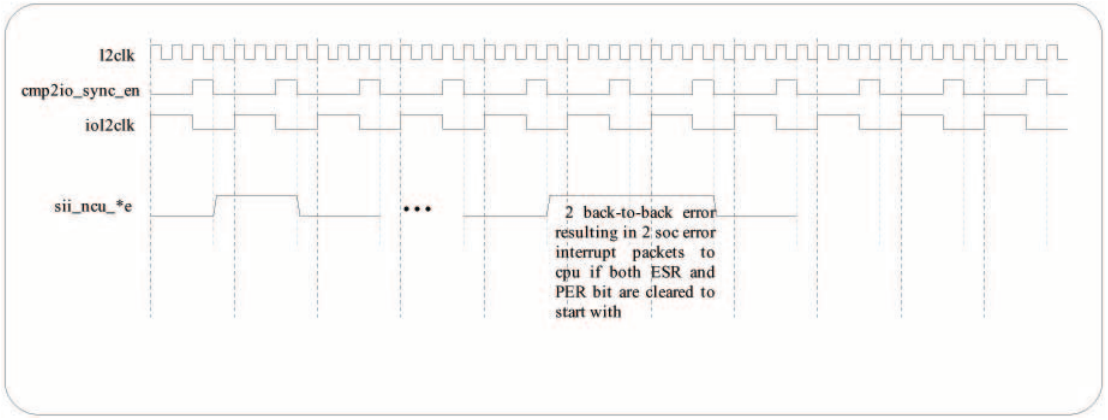


FIGURE 7-18 SII to NCU Error Syndrome

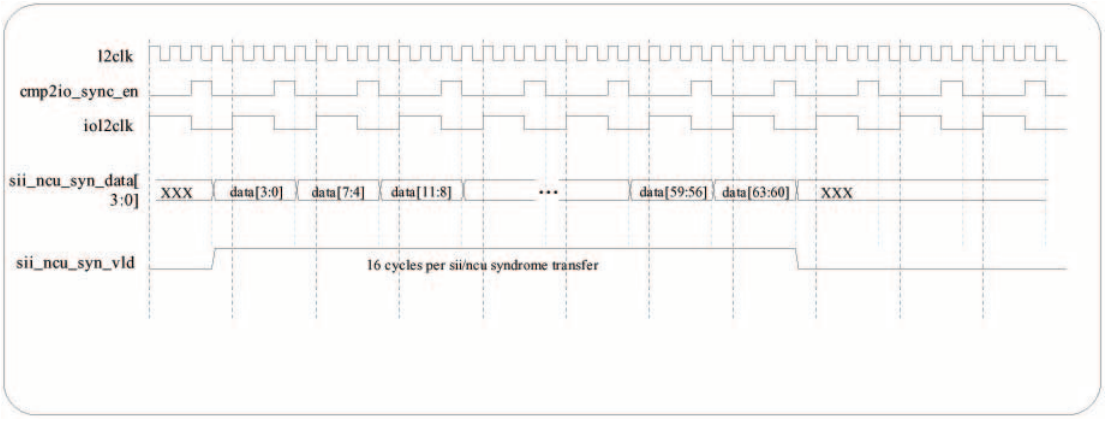


FIGURE 7-19 SII to NCU Error Strobe and Syndrome Transfer Example