DECaxp 21264 Emulator Cache Coherency

Design Specification

Author: Jonathan D. Belanger  
Creation Date: April 7, 2018  
Modify Date: April 22, 2018

**April 18**

The information in this publication is subject to change without notice.

JONATHAN D. BELANGER SHALL NOT BE LIABLE FOR TECHNICAL OR EDITORIAL ERRORS OR OMISSIONS CONTAINED HEREIN, NOR FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES RESULTING FROM THE FURNISHING, PERFORMANCE, OR USE OF THIS MATERIAL. THIS INFORMATION IS PROVIDED "AS IS" AND JONATHAN D. BELANGER DISCLAIMS ANY WARRANTIES, EXPRESS, IMPLIED OR STATUTORY AND EXPRESSLY DISCLAIMS THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR PARTICULAR PURPOSE, GOOD TITLE AND AGAINST INFRINGEMENT.

This publication contains information protected by copyright. No part of this publication may be photocopied or reproduced in any form without prior written consent from Jonathan D. Belanger.

©22 April 2018 Jonathan D. Belanger. All rights reserved. Printed in U.S.A.

COMPAQ, DIGITAL, DIGITAL UNIX, OpenVMS, VAX, VMS, and the  logo registered in United States Patent and Trademark Office.

GRAFOIL is a registered trademark of Union Carbide Corporation.  
IEEE is a registered trademark of The Institute of Electrical and Electronics Engineers, Inc.  
Windows NT is a trademark of Microsoft Corporation.

Other product names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

Table of Contents

[1 Preface 5](#_Toc511488740)

[1.1 Audience 5](#_Toc511488741)

[1.2 Terminology and Conventions 5](#_Toc511488742)

[2 Overview 10](#_Toc511488743)

[3 Design Constraints 11](#_Toc511488744)

[3.1 Cache Block State 11](#_Toc511488745)

[3.2 Cache Block State Transitions 11](#_Toc511488746)

[3.3 CSRs Affecting Cache Coherency 12](#_Toc511488747)

[3.4 Commands sent from AXP CPU 13](#_Toc511488748)

[3.5 Commands sent to AXP CPU 14](#_Toc511488749)

[4 Uniprocessor Cache Coherency 16](#_Toc511488750)

[4.1 Instructions, Cache States, and System Messages 17](#_Toc511488751)

[4.2 CSR and IPR Settings to Support a Uniprocessor Configuration 17](#_Toc511488752)

[5 Multiprocessor Cache Coherency 19](#_Toc511488753)

[5.1 Modified-Shared-Invalid (MSI) Cache Coherency Protocol 19](#_Toc511488754)

[5.1.1 Instructions, Cache States, and System Messages for MSI Cache Coherency 21](#_Toc511488755)

[5.1.2 CSR and IPR Settings to Support a Multiprocessor Configuration for MSI 22](#_Toc511488756)

[5.2 MESI Cache Coherency Protocol 24](#_Toc511488757)

[5.3 MOSI Cache Coherency Protocol 24](#_Toc511488758)

[5.4 MOESI Cache Coherency Protocol 24](#_Toc511488759)

[5.5 MERSI Cache Coherency Protocol 24](#_Toc511488760)

[5.6 MESIF Cache Coherency Protocol 24](#_Toc511488761)

[5.7 Write-Once Cache Coherency Protocol 24](#_Toc511488762)

[5.8 Synapse Cache Coherency Protocol 24](#_Toc511488763)

[5.9 Berkeley Cache Coherency Protocol 24](#_Toc511488764)

[5.10 Firefly Cache Coherency Protocol 24](#_Toc511488765)

[5.11 Dragon Cache Coherency Protocol 24](#_Toc511488766)

Tables

[Table 3‑1 AXP CPU Supported Cache States 11](#_Toc511488619)

[Table 3‑2 Cache Block State Transitions 11](#_Toc511488620)

[Table 3‑3 Cache Coherency CSRs 12](#_Toc511488621)

[Table 3‑4 AXP CPU to System Commands 13](#_Toc511488622)

[Table 3‑5 Probe Request Data Movement Commands 14](#_Toc511488623)

[Table 3‑6 Probe Request Next Cache State Commands 15](#_Toc511488624)

[Table 3‑7 System Data Control (SysDc) Commands 15](#_Toc511488625)

[Table 4‑1 IPR and CSR Uniprocessor Settings 17](#_Toc511488626)

[Table 5‑1 MSI to AXP CPU Cache State Mapping 19](#_Toc511488627)

[Table 5‑2 IPR and CSR Uniprocessor Settings 23](#_Toc511488628)

Figures

[Figure 4‑1 Uniprocessor Cache State Transitions 16](#_Toc511488658)

[Figure 4‑2 Processing Flows 17](#_Toc511488659)

[Figure 5‑1 Cache Coherency for Multiprocessors 19](#_Toc511488660)

[Figure 5‑2 MSI State Diagram 20](#_Toc511488661)

[Figure 5‑3 Processing Flows 21](#_Toc511488662)

# Preface

## Audience

This document is for the designers and programmers who plan to code or update the DECaxp 21264 Emulator source code.

## Terminology and Conventions

This section defines the abbreviations, terminology, and other conventions used throughout this document.

**Abbreviations**

* Binary Multiples

The abbreviations K, M, and G (kilo, mega, and giga) represent binary multiples and have the following values.

For example:

* Sign extension

SEXT(*x*) means *x* is sign-extended to the required size.

* Register Access  
    
  The abbreviations used to indicate the type of access to register fields and bits have the following definitions:

| **Abbreviation** | **Meaning** |
| --- | --- |
| IGN | Ignore. Bits and fields specified are ignored on writes. |
| MBZ | Must Be Zero. Software must never place a nonzero value in bits specified as MBZ. A nonzero read produces an Illegal Operation exception. Also, MBZ fields are reserved for future use. |
| RAZ | Read As Zero. Bits and fields return a zero when read. |
| RC | Read Clears. Bits and fields are cleared when read. Unless otherwise specified, such bits cannot be written. |
| RES | Reserved. Bits and fields are reserved and should not be used. However, zeros can be written to reserved fields that cannot be masked. |
| RO | Read Only. The value may be read by software. It is written by the emulation code. Software write operations are ignored. |
| RO, *n* | Read Only, and takes the value *n* at power-on reset. The value may be read by software. It is written by the emulation code. Software write operations are ignored. |
| RW | Read Write. Bits and fields can be read and written. |
| RW, *n* | Read Write and takes the value *n* at power-on reset. Bits and fields can be read and written. |
| W1C |  |
| W1S |  |
| WO | Write Only. Bits and fields can be written but not read. |
| WO, *n* | Write Only, and takes the value *n* at power-on reset. Bits and fields can be written but not read. |

**Addresses**

Unless otherwise noted, all addresses and offsets are hexadecimal.

**Aligned and Unaligned**

The terms aligned and naturally aligned are interchangeable and refer to data objects that are powers of two in size. An aligned datum of size 2n is stored in memory at a byte address that is a multiple of 2n; that is, one that has n low-order zeros. For example, an aligned 64-byte stack frame has a memory address that is a multiple of 64.

A datum of size 2n is unaligned if it is stored in a byte address that is not a multiple of 2n.

**Bit Notation**

Multiple-bit fields can include contiguous and noncontiguous bits contained in square brackets ([]). Multiple contiguous bits are indicated by a pair of numbers separated by a colon [:]. For example, [9:7,5,2:0] specifies bits 9,8,7,5,2,1, and 0. Similarly, single bits are frequently indicated with square brackets. For example, [27] specifies bit 27. See also Field Notation.

**Data Units**

The following data unit terminology is used throughout this manual.

| **Term** | **Words** | **Bytes** | **Bits** | **Other** |
| --- | --- | --- | --- | --- |
| Byte | ½ | 1 | 8 | --- |
| Word | 1 | 2 | 16 | --- |
| Longword | 2 | 4 | 32 | Dword |
| Quadword | 4 | 8 | 64 | 2 longwords |

**Do Not Care (X)**

A capital X represents any valid value.

**External**

Unless otherwise stated, external means not contained in the chip.

**Field Notation**

The names of single-bit and multiple-bit fields can be used rather than the actual bit numbers (see Bit Notation). When the field name is used, it is contained in square brackets ([]). For example, **RegisterName[LowByte]** specifies **RegisterName[7:0]**.

**Note**

Notes emphasize particularly important information.

**Numbering**

All numbers are decimal or hexadecimal unless otherwise indicated. The prefix 0x indicates a hexadecimal number. For example, 19 is decimal, but 0x19 and 0x19a are hexadecimal (also see Addresses). Otherwise, the base is indicated by a subscript; for example, 1002 is a binary number.

**Ranges and Extents**

*Ranges* are specified by a pair of numbers separated by two periods (..) and are inclusive. For example, a range of integers 0..4 includes the integers 0, 1, 2, 3, and 4.

*Extents* are specified by a pair of numbers in square brackets ([]) separated by a colon (:) and are inclusive. Bit fields are often specified as extents. For example, bits [7:3] specifies bits 7, 6, 5, 4, and 3.

**Register Figures**

The gray areas in register figures indicate reserved or unused bits and fields.

Bit ranges that are coupled with the field name specify the bits of the named field that are included in the register. The bit range may, but need not necessarily, correspond to the bit Extent in the register. See the explanation above Table 5–1 for more information.

**Signal Names**

The following examples describe signal-name conventions used in this document.

**AlphaSignal[n:n]** Boldface, mixed-case type denotes signal names that are assigned internal and external to the EV68CB/EV68DC (that is, the signal traverses a chip interface pad).

**AlphaSignal\_x[n:n]** When a signal has high and low assertion states, a lowercase italic x represents the assertion states. For example, **SignalName\_*x*[3:0]** represents **SignalName\_H[3:0]** and **SignalName\_L[3:0]**.

**UNDEFINED**

Operations specified as UNDEFINED may vary from moment to moment, implementation to implementation, and instruction to instruction within implementations. The operation may vary in effect from nothing to stopping system operation.

UNDEFINED operations may halt the processor or cause it to lose information. However, UNDEFINED operations must not cause the processor to hang, that is, reach an unhalted state from which there is no transition to a normal state in which the machine executes instructions.

**UNPREDICTABLE**

UNPREDICTABLE results or occurrences do not disrupt the basic operation of the processor;

it continues to execute instructions in its normal manner. Further:

* Results or occurrences specified as UNPREDICTABLE may vary from moment to moment, implementation to implementation, and instruction to instruction within implementations. Software can never depend on results specified as UNPREDICTABLE.
* An UNPREDICTABLE result may acquire an arbitrary value subject to a few constraints. Such a result may be an arbitrary function of the input operands or of any state information that is accessible to the process in its current access mode. UNPREDICTABLE results may be unchanged from their previous values.  
    
  Operations that produce UNPREDICTABLE results may also produce exceptions.
* An occurrence specified as UNPREDICTABLE may happen or not based on an arbitrary choice function. The choice function is subject to the same constraints as are UNPREDICTABLE results and, in particular, must not constitute a security hole.  
    
  Specifically, UNPREDICTABLE results must not depend upon, or be a function of, the contents of memory locations or registers that are inaccessible to the current process in the current access mode.  
    
  Also, operations that may produce UNPREDICTABLE results must not:
* Write or modify the contents of memory locations or registers to which the current process in the current access mode does not have access, or
* Halt or hang the system or any of its components.

For example, a security hole would exist if some UNPREDICTABLE result depended on the value of a register in another process, on the contents of processor temporary registers left behind by some previously running process, or on a sequence of actions of different processes.

**X**

Do not care. A capital X represents any valid value.

# Overview

Cache coherency is a necessity for Symmetric Multi-Processor (SMP) Systems. This is so that if a memory location that is in more than one cache, all caches agree. Otherwise, one processor accessing the same physical memory location could be utilizing different values. There are three basic cache coherency styles. They are:

1. Where all caches are maintained with the same value, whether read from or written to.
2. Where all caches maintain the same for reading, but the one that wants to be written to will invalidate the other caches values
3. Where all caches may be out of sync with one another (called Non-conforming).

This last option we are not going to utilize within our implementation. In the EV68CB/EV68DC Hardware Reference Manual (AXP HRM) in Section 4.5.1 Cache Coherency Basics, it states that this processor provides hardware mechanisms to support several cache coherency protocols. The protocols can be separated into two classes: write invalidate cache coherency protocol and flush cache coherency protocol.

The following tasks must be performed to maintain cache coherency:

* Istream data from memory spaces will be cached in the Icache only. Icache coherency is not maintained by hardware – it must be maintained by software using the CALL\_PAL IMB instruction.
* The AXP CPU maintains the Dcache as a subset of the Bcache. The Dcache is set-associative but is kept a subset of the larger externally implemented direct-mapped Bcache.
* System logic must help the AXP CPU to keep the Bcache coherent with main memory and other caches in the system.
* The AXP CPU requires the system to allow only one change to a block at a time. This means that if the AXP CPU gains the bus to read or write a block, no other node on the bus should be allowed to access that block until the data has been moved.

# Design Constraints

In this section we will document the various specifics of the AXP CPU that will be used throughout the remainder of this design specification. This includes information maintained within the caches and Control and Status Registers (CSRs).

## Cache Block State

The following states are possible for the caches within the AXP CPU:

Table ‑ AXP CPU Supported Cache States

| **State Name** | **Description** |
| --- | --- |
| Invalid | This AXP CPU does not have a copy of the block. |
| Clean | This AXP CPU holds a read-only copy of the block, and no other agent in the system holds a copy. Upon eviction, the block is not written to memory. |
| Clean/Shared | This AXP CPU holds a read-only copy of the block, and at least one other agent is the system may hold a copy of the block. Upon eviction, the block is not written to memory. |
| Dirty | This AXP CPU holds a read-write copy of the block, and no other agent in the system holds a copy. Upon eviction, the block must be written to memory. |
| Dirty/Shared | This AXP CPU holds a read-only copy of the dirty block, which may be shared with another agent. Upon eviction, the block must be written to memory. |

## CSRs Affecting Cache Coherency

The following CSRs in the AXP CPU affect how cache coherency is performed based on their settings. Table list the CSRs, their possible values and what those values represent.

Table ‑ Cache Coherency CSRs

| **CSR** | **Description** |
| --- | --- |
| BC\_CLEAN\_VICTIM | Enable CleanVictimBlk commands to the system interface. |
| BC\_RDVICTIM | Enable RdBlkVic, RdBlkNodVic, and InvalToDirtyVic commands to the system interface. |
| ENABLE\_EVICT | Enable issue Evict command for all ECB instructions. If this field is set, then the BC\_CLEAN\_VICTIM must also be set. |
| ENABLE\_STC\_COMMAND | Enable STx\_C instructions. Systems that require an explicit indication of ChangeToDirty status changes initiated by STx\_C instructions can assert Cbox CSR ENABLE\_STC\_COMMAND [0]. When this register field = 000, CleanToDirty and SharedToDirty commands are used. The distinction between a ChangeToDirty command generated by a STx\_C instruction and one generated by a STx instruction is important to systems that want to service ChangeToDirty commands with dirty data from a source processor. In this case, the distinction between a locked exclusive instruction and a normal instruction is critical to avoid livelock for a LDx\_L/STx\_C sequence.  **NOTE**: The AXP HRM sometimes has this as STC\_ENABLE. |
| INVAL\_TO\_DIRTY\_ENABLE | Enable WH64 functionality.   | **INVAL\_TO\_DIRTY\_ENABLE**  **[1:0]** | **Cbox Action** | | --- | --- | | x0 | WH64 instructions are converted to RdModx commands at the interface. Beyond this point, no other agent sees the WH64 instruction. This mode is useful for AXP CPUs that do not want to support InvalToDirty transactions. | | 01 | WH64 instructions are enabled, but they are acknowledged within the AXP CPU. | | 11 | WH64 instructions are enabled and generate InvalToDirty transactions off chip. | |
| PRB\_TAG\_ONLY | Enable probe-tag only mode. The AXP CPU expects to hit in cache on a probe response, so it always fetches a cache block from the Bcache on system probes. This can become a performance problem for systems that do not monitor the Bcache tags, so the EV68CB/EV68DC provides Cbox CSR PRB\_TAG\_ONLY[0], which only accesses Bcache tags for system probes. For a Bcache hit, the AXP CPU retries the probe reference to get the associated data. In this mode, the AXP CPU has a cache-hit counter that maintains some history of past cache hits in order to fetch the data with the tag in the cases where streamed transactions are being performed to the host processor. |
| RDVIC\_ACK\_INHIBIT | Enable inhibition of incrementing acknowledge counter for RdBlkVic, RdBlkNodVic, and InvalToDirtyVic commands. |
| SET\_DIRTY\_ENABLE | SetDirty Acknowledge.   | **SET\_DIRTY\_ENABLE**  **[2:0]** | **Cbox Action** | | --- | --- | | 000 | Everything acknowledged internally (uniprocessor). | | 001 | Only clean blocks generate external. acknowledge (CleanToDirty commands only). | | 010 | Only clean/shared blocks generate external acknowledge (SharedToDirty command only) | | 011 | Clean and clean/shared blocks generate external acknowledge | | 100 | Only dirty/shared blocks generate external acknowledge (SharedToDirty commands only) | | 101 | Only dirty/shared and clean blocks generate external acknowledge. | | 110 | Only dirty/shared and clean/shared blocks generate external acknowledge. | | 111 | All transactions generate external acknowledge. | |
| SYSBUS\_MB\_ENABLE | Enable MB commands off chip. See AXP RTM Section 2.12.2, Memory Barrier (MB/WMB/TB Fill Flow). |

## Commands sent from AXP CPU

There are quite a few commands that can be sent by the AXP CPU to the system to request some off chip resource (memory, storage, or cache coherency information). lists all the commands that can be sent to the System from the AXP CPU.

Table ‑ AXP CPU to System Commands

| **Command** | **Function** |
| --- | --- |
| NOP | The AXP CPU drives this command on idle cycles during a reset. Once the first NZOP is generated, this command is no longer generated. |
| ProbeResponse | Returns the probe status and ID number of the VDB entry holding the requested cache block. |
| NZNOP | This nonzero NOP helps to parse the command packet. |
| VDBFlushRequest | VDB flush request. The AXP CPU sending this command to the system when an internally generated transaction Bcache index matches a Bcache victim or probe in the VDB. The system should flush all VDB entries associated with all outstanding probe and WrVictimBlk transactions that where queued up prior to this request. |
| MB | Indicates that an MB instruction was issued. |
| ReadBlk | Memory read request. Usually as the result of an LDx instruction. |
| ReadBlkMod | Memory read request with modify intent. Usually as the result of a STx instruction. |
| ReadBlkI | Memory read request for the Instruction Stream (Istream). This is internally generated when the AXP CPU attempt to parse the next instruction for execution and misses in the Icache. |
| FetchBlk | Noncached memory read request. |
| ReadBlkSpec | Speculative memory read request. |
| ReadBlkModSpec | Speculative memory read request with modify intent. |
| ReadBlkSpecI | Speculative memory read request for Istream. |
| FetchBlkSpec | Speculative noncached memory read request. |
| ReadBlkVic | Memory read request with a victim. |
| ReadBlkModVic | Memory read request with modify intent, with a victim. |
| ReadBLKVicI | Memory read request for Istream with victim. |
| WrVictimBlk | Write-back of dirty block. Sent when a dirty cache block is evicted. |
| CleanVictimBlk | Supply address of a clean victim. Sent when a clean cache block is evicted. |
| Evict | Invalidate evicted block at the given Bcache index. |
| ReadBytes | I/O read request. Mask indicates which bytes of the quadword are valid. |
| ReadLWs | I/O read request. Mask indicates which longwords of 32-byte block are valid. |
| ReadQWs | I/O read request. Mask indicates which quadwords of the 64-byte block are valid. |
| WrBytes | I/O write request. Mask indicates which bytes of the quadword are valid. |
| WrLWs | I/O write request. Mask indicates which longwords of 32-byte block are valid. |
| WrQWs | I/O write request. Mask indicates which quadwords of the 64-byte block are valid. |
| CleanToDirty | Sets a cache block to a Dirty state, but only if it is currently Clean. This is used when duplicate tags have been enabled. |
| SharedToDirty | Sets a cache block to a Dirty state, but only if it is currently in a Shared state. This is used for multiprocessor systems. |
| STCChangeToDirty | Sets a cache block to a Dirty state that was previously Clean or Shared for an STx\_C instruction. |
| InvalToDirtyVic | Invalid to Dirty state with a victim. |
| InvalToDirty | WH64 acts like a ReadBlkMod without the fill cycles. |

## Commands sent to AXP CPU

The following command are send from the System to the CPU for processing in the Cbox. The Cbox utilizes the Bcache and its Duplicate Tag (DTAG) array to respond back to the system. The commands sent by the system are broken up into two components. The first component is for a data movement request (see ). The second component is for a next cache state request.

Table ‑ Probe Request Data Movement Commands

|  |  |
| --- | --- |
| **Data Movement Commands** | **Data Movement Function** |
| NOP | No operation. |
| ReadHit | Read if hit. Return the data back to the system if block is valid. No other state matters. |
| ReadDirty | Read if dirty. Return the data back to the system if the block is valid and dirty. For both Dirty and Dirty/Shared cache blocks. |
| ReadAlways | Read anyway. Return the data at the probe index back to the system. State of the block is irrelevant. |

Table ‑ Probe Request Next Cache State Commands

| **Next Cache State Commands** | **Next Cache State** |
| --- | --- |
| NOP | No state changed. |
| Clean | State changed to Clean. |
| Clean/Shared | State changed to Clean/Shared. |
| Transition 3[[1]](#footnote-1) | Clean ⇒ Clean/Shared  Dirty ⇒ Invalid  Dirty/Shared ⇒ Clean/Shared |
| Dirty/Shared | State changed to Dirty/Shared |
| Invalid | State changed to Invalid |
| Transition 1[[2]](#footnote-2) | Clean ⇒ Clean/Shared  Dirty ⇒ Dirty/Shared |
| Reserved | Not used. |

Table ‑ System Data Control (SysDc) Commands

| **SysDc Command** | **Description** |
| --- | --- |
| NOP | NOP, SysData is ignored by the AXP CPU. |
| ReadDataError | Data is returned for read commands. The system sends SysData, I/O, or memory NXM |
| ChangeToDirtySuccess | No data. SysData is ignored by the AXP CPU. This command is also used for the InvalToDirty response |
| ChangeToDirtyFail | No data. SysData is ignored by the AXP CPU. This command is also used for the Evict response. |
| MBDone | Memory barrier operation completed. |
| ReleaseBuffer | Command to alert the AXP CPU that the RVB, RPB, and ID field are valid. |
| ReadData | Data returned for read commands. The system returns SysData. The cache status is set to Clean. The system uses the lower 2 bits of the command to control wrap order. |
| ReadDataDirty | Data is returned for Rdx and RdModx commands. The cache status is set to Dirty. The system uses the lower 2 bits of the command to control wrap order. |
| ReadDataShared | Data returned for read commands. The system returns SysData. The cache status is set to Clean/Shared. The system uses the lower 2 bits of the command to control wrap order. |
| ReadDataShareDirty | Data is returned for RdBlk commands. The cache status is set to Shared/Dirty. The system uses the lower 2 bits of the command to control wrap order. |
| WriteData | Data is sent for AXP CPU write commands or system probes. The AXP CPU sends data to the System. The AXP CPU uses the lower 2 bits of the command to control the wrote order. |

## DECchip 21272/21274

The difference between these two chipsets is that the last digit represents the number of processors that the chipset supports. The DECaxp 21264 Emulator (emulator) is implemented to the 21274 specifications. Note, though, each chipset can support just a single CPU.

## Cache Coherency in the Emulator

The emulator will maintain cache coherency a bit differently than how the actual hardware maintains it. In the hardware implementation, the chipset as part of a system board utilizes a message passing scheme to maintain and coordinate coherency among the CPUs. In the emulation, the Bcaches are going to be implemented in such a way that coherency is automatically maintained across all CPU/Cache pairs.

In order to be able to support a self-maintained coherency, the cache implementation originally implemented is going to have to be refactored. Figure 3‑1 depicts how memory and the caches are organized. One of the things we will be taking advantage of in the emulator is the fact that the Bcache is a superset of the Dcache.

Figure ‑ Cache/Memory Organization

System

Main Memory

There is going to be a single Bcache data structure that will scaled up to the number of CPUs defined. The data structure will contain CPU specific status and block but will also have associated with it the information from the other CPUs. When a block is indicated as Shared (Shared/Clean or Shared/Dirty) all blocks will contain the exact same values and be maintained as such. Where the block status is not one of the Shared ones, then the values within the blocks will be unique to that CPU.

The data structure will have the following layout:

Figure ‑ Bcache Data Structure Layout (for 2 CPUs)

Tag

State

Data Block

Tag

State

Data Block

CPU 0

CPU 1

Bcache  
Index

Dcache  
Index

Dcache  
Set

Bcache  
Index

Dcache  
Index

Dcache  
Set

Each CPU will have its own Bcache Index, Dcache Index and Set that the other CPUs will not have access to. But they will have access to the Tag/State/Data Block of the other CPUs, so if the Shared bit is set in one, it is set in all. We can do this because the Bcache is physically indexed and tagged.

One thing that will need to be maintained is the potential for more than one CPU requesting the same address at the same time and having to go to the system to get it from the memory. If the cache block in question is invalid in both, then it is likely that one CPU will fill in it’s block before the other. The end result will either be that both blocks will be Shared/Clean or one will be Shared/Dirty with the other will be Shared/Clean. They may even have started out as Invalid and one when to Clean or Dirty, when the other loads the same physical address, it will have to make sure both blocks have the same value (possibly the one in the other CPU over what was returned from memory, as memory may be stale). Also, the cache location is going to have to be locked when a store is pending but not yet retired.

The biggest difference in this design over the hardware is that it will be the CPUs that coordinate cache coherency and not the chipset. So, the following roles have been defined:

For CPUs:

* Instruction execution
* Cache Coherency with other CPUs
* Request read/write memory and I/O

For the System:

* Processing requests for memory references (addresses) reads and writes
* Processing requests for CSR reads and writes
* Processing requests for I/O reads and writes

# Uniprocessor Cache Coherency

For uniprocessors, cache coherency is the most straightforward. There is no need to utilize the Clean/Shared and Dirty/Shared states, as there is only a single cache and no coherency requirements.

Figure ‑ Uniprocessor Cache State Transitions

4

1

2

5

3

6

1. When a Rdx is performed, memory is read and stored into the cache in a Clean state (read-only).
2. When a RdModx is performed, memory is read and stored into the cache in a Dirty state (read-write).
3. When a previously Rdx block is changed to a read-write block.
4. When a previously Rdx block is evicted from the cache.
5. When a previously RdModx or read-write block is evicted and written out to memory.
6. This transition is not necessary in a uniprocessor system.

## Instructions, Cache States, and System Messages

This section describes each of the instructions that generate messages to and gets responses from the System. The processing for this functionality, is controlled by CSRs and IPRs. Please see Section 4.2 for more information about these settings.

Figure ‑ Processing Flows

| **Step** | **Instruction** | **Cache State** | **Interface** | **System** | **Memory Access** |
| --- | --- | --- | --- | --- | --- |
| L1.1 | LDx ⇒ | Invalid ⇒ | ReadBlk ⇒ | Read Memory ⇒ |  |
| L1.2 | ⇐ LDx | ⇐ Clean | ⇐ ReadData | ⇐ Return Data | ⇐ Return Data |
| L2.1 | LDx ⇒ | Clean/Hit ⇓ |  |  |  |
| L2.2 | ⇐ LDx | ⇐ Clean |  |  |  |
| L3.1 | LDx ⇒ | Clean/Miss ⇓ |  |  |  |
| L3.2 |  | Evict ⇒ | ReadBlk ⇒ | Read Memory ⇒ |  |
| L3.3 | ⇐ LDx | ⇐ Clean | ⇐ ReadData | ⇐ Return Data | ⇐ Return Data |
| L4.1 | LDx ⇒ | Dirty/Hit ⇓ |  |  |  |
| L4.2 | ⇐ LDx | ⇐ Dirty |  |  |  |
| L5.1 | LDx ⇒ | Dirty/Miss ⇓ |  |  |  |
| L5.2 |  | Evict ⇒ | WrVictimBlk ⇒ | Write Memory ⇒ |  |
| L5.3 |  |  | ⇐ WriteData | ⇐ Write Memory | ⇐ Write Memory |
| L5.4 |  |  | ReadBlk ⇒ | Read Memory ⇒ |  |
| L5.5 | ⇐ LDx | ⇐ Clean | ⇐ ReadData | ⇐ Return Data | ⇐ Return Data |
| S1.1 | STx ⇒ | Invalid ⇒ | ReadBlkMod ⇒ | Read Memory ⇒ |  |
| S1.2 | ⇐ STx | ⇐ Dirty | ⇐ ReadDataDirty | ⇐ Return Data | ⇐ Return Data |
| S2.1 | STx ⇒ | Clean/Hit ⇓ |  |  |  |
| S2.2 |  | CleanToDirty ⇓ |  |  |  |
| S2.3 | ⇐ STx | ⇐ Dirty |  |  |  |
| S3.1 | STx ⇒ | Clean/Miss ⇓ |  |  |  |
| S3.2 |  | Evict ⇒ | ReadBlkMod ⇒ | Read Memory ⇒ |  |
| S3.3 | ⇐ STx | ⇐ Dirty | ⇐ ReadDataDirty | ⇐ Return Data | ⇐ Return Data |
| S4.1 | STx ⇒ | Dirty/Hit ⇓ |  |  |  |
| S4.2 | ⇐ STx | ⇐ Dirty |  |  |  |
| S5.1 | STx ⇒ | Dirty/Miss ⇓ |  |  |  |
| S5.2 |  | Evict ⇒ | WrVictimBlk ⇒ | Write Memory ⇒ |  |
| S5.3 |  |  | ⇐ WriteData | ⇐ Write Memory | ⇐ Write Memory |
| S5.4 |  |  | ReadBlkMod ⇒ | Read Memory ⇒ |  |
| S5.5 | ⇐ STx | ⇐ Dirty | ⇐ ReadData | ⇐ Return Data | ⇐ Return Data |
| M1.1 | MB ⇒ |  |  |  |  |
| M1.2 | ⇐ MB |  |  |  |  |

## CSR and IPR Settings to Support a Uniprocessor Configuration

Table 4‑1 lists the Internal Processor Register (IPR) and Control and Status Register (CSR) settings to support uniprocessing. These settings are no necessarily required. If the multiprocessor settings are used, there will be unnecessary communication between the AXP CPU and System. Since the System knows how many processors are present, when there is just one, sending Cache Status and other requests that do not actually request data movement, are not necessary.

Table ‑ IPR and CSR Uniprocessor Settings

| **IPR or CSR** | **Setting** | **Description** |
| --- | --- | --- |
| I\_CTL[TB\_MB\_MEM] | 0 | Deasserting this field in the Ibox IPR will disable inserting an MB instruction within the TB fill flow. |
| SYSBUS\_MB\_ENABLE | 0 | Deasserting this CSR will internally acknowledge MB commands/ instructions. |
| SET\_DIRTY\_ENABLE | 000 | Deasserting this CSR will internally acknowledge SetDirty, SharedToDirty or CleanToDirty, requests (called a SetModify). |
| ENABLE\_STC\_COMMAND | 0 | Deasserting this CSR will internally acknowledge a SetDirty request for an STx\_C instruction. Deasserting means that rather than sending a STCChangeToDirty command, a SharedToDirty or CleanToDirty command would be send. Deasserting the SET\_DIRTY\_ENABLE CSR, as well, will internally acknowledge these. |
| INVAL\_TO\_DIRTY | x0 | Deasserting this CSR will internally acknowledge InvalToDirty requests. |
| ENABLE\_EVICT | 0 | Deasserting this CSR will cause the AXP CPU to not send a command to the System to indicate that an evict is being performed. |
| BC\_CLEAN\_VICTIM | 0 | This must also be deasserted when the ENABLE\_EVICT is also deasserted. |

# Multiprocessor Cache Coherency

For multiprocessors, cache coherency is the most difficult. In addition to the states in the uniprocessor case, we also introduce a Clean/Shared and a Dirty/Shared Cache State. Of these, the Clean/Shared is the simpler of the two. Depending upon how the IPRs and CSRs are set, it is possible to support more than one cache coherency protocol. The cache coherency protocols are defined in the next sections. The following figure shows the components involved in cache coherency

Figure ‑ Cache Coherency for Multiprocessors

S  
y  
s  
t  
e  
m

Memory and I/O

CPU

Cache

CPU

Cache

Coherency

## Instructions, Cache States, and System Messages

TBD

## CSR and IPR Settings to Support a Multiprocessor Configuration

TBD

# AXP CPU Cache Redesign

The initial implementation of the Icache, Dcache, and Bcache, had each cache separate from one another. See Figure 6‑1 about how the caches can be diagramed.

Figure ‑ Cache Subset Hierarchy

System

Main Memory

In reality just the Icache is separate from the Bcache and Dcache, but the Dcache is a subset of the Bcache. Another way to consider the Bcache/Dcache relationship is that the Bcache is a superset of the Dcache.

Therefore, one of the optimizations that I’d like to redesign into the CPU caches, is that the Bcache holds all the information about each cache block in the CPU and the Dcache is just an index into the Bcache. All state information is maintained in one and only one place, the Bcache. This will allow for the Cbox to more efficiently respond to system commands about the caches and we have minimized data copying within the CPU. This section of the design will document how this will be implemented to better support Cache Coherency.

1. Transition 3 is useful in non-duplicate tag systems that want to give writeable status to the reader and do not know if the block is clean or dirty. [↑](#footnote-ref-1)
2. Transition 1 is useful in non-duplicating tag systems that do not update memory on ReadBlk hits to a dirty block in another processor. [↑](#footnote-ref-2)