資訊工程學系 計算機結構 期末考(105 下)

姓名： 學號：

[Instruction Level Parallelism and Its Exploitation] (34 pnts)

1. (5 pnts) 請說明何謂 hardware speculation。

當還不確定 branch 指令的 outcome 時，以預測的方式猜測可能的 outcome，之後繼續執行這個路徑上的指令。

Overcome control dependence by hardware speculating on outcome of branches and executing program as if guesses were correct

2. (10 pnt) 為了將原本的 Tomasulo 機制擴展成可以達到 hardware-based speculation 運作，需要額外在原 Tomasulo 機制增加(a)甚麼階段與增加(b)甚麼硬體? (c)新的階段所要做的任務是甚麼?

(a) Commit

(b) Reorder buffer

(c) 當指令進到 commit 階段時，經判定該指令是必須執行的，則將記錄在 ROB entry 中的結果存入記憶體或是暫存器，否則需要清除此 entry 之後的所有 entry

3. (5 pnts) Reorder Buffer 是如何確保 precise interrupt model?

因為 ROB 可以讓指令 In order commit，當指令執行階段發生 interrupt 或是 exception 時，相關訊息將會先被記錄但不處理，直到指令到達 commint 階段時，才處理。

If an instruction caused an interrupt, we could simply wait until it reached the head of the ROB and take the interrupt, flushing any other pending instructions from the ROB. Because instruction commit happens in order, this yields a precise exception.

4. (4 pnt) 給予可以使得 CPI<1 的兩個機制。

Superscalar processors

VLIW (very long instruction word) processors

5. (10 pnts) (a)請說明 Branch Target Buffer (BTB)機制 (b)與 Branch Prediction Buffer 相較，BTB 的好處與額外增加的成本是甚麼?

(a)每個 entry 中紀錄了目前指令的 PC 與所預測之接下來指令位置

Next PC prediction buffer, indexed by current PC

(b)減少計算下個指令的時間，需要多額外空間做 PC 紀錄

[Multiprocessors and Thread-Level Parallelism] (56 pnts)

6. (5 pnts) 請說明何謂 Cache Coherency problem。

在不同處理器的 cache 中，所看到的同一變數之值不同

7. (5 pnts) 比較在 Snoopy protocol 中，如果使用 Write invalidate protocol 或是 Write update protocol 有甚麼優缺點?

由於前者只傳送 invalidation 訊息，並不會將新的資料廣播給所有處理器知道，因此較節省頻寬。而後者需要傳送新資料至所有處理器，且要確保收到，時間成本大。

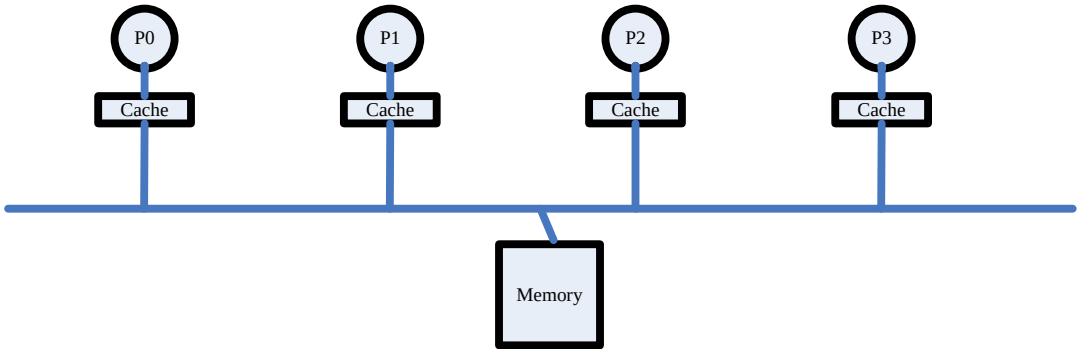
8. (6 pnts) 請解釋 directory-based protocol 中的三種 node：local、home、remote nodes。.

A. **Local node** where a request originates

B. **Home node** where the memory location and the directory entry of an address reside

C. **Remote node** has a copy of a cache block, whether exclusive or shared

9. (20 pnts) A write-back cache use the write-invalidate cache coherence protocol. In the snoopy protocol, the cache controller must maintain the state transitions for each cache block. We assume that the memory addresses X1 and X2 are in different memory blocks and map to the same cache block in different processors. Their initial values are 0. The initial stats of cache block and each memory block are Invalid and Uncached, respectively. (Cache state: Invalid, Shared, or Exclusive)



Please describe

甲、the stat change of each corresponding cache block

乙、the values of X1 and X2 in the memory and cache

丙、what message is placed in the bus

When the following requests are issued under the snoopy protocol.

(1) P1 writes 1 to X1 (3 pnts)

In P1, cache block: Invalid→ Exclusive X1=1 write miss

Memory block: X1=0

(2) P2 writes 2 to X1 (4 pnts)

In P2, cache block: Invalid→Exclusive X1=2 write miss

In P0, cache block: Exclusive→Invalid write back the dirty block, abort the memory access

Memory block: X1=1

(3) P0 reads X1 (4 pnts)

In P0, cache block: Invalid→Shared X1=2 read miss

In P2, cache block: Exclusive→Shared X1=2 write back the dirty block

Memory block: X1=2

(4) P2 writes 1 to X2 (4 pnts)

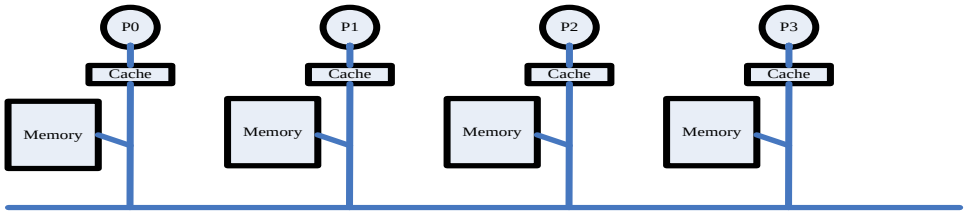
In P2, cache block: Shared→ Exclusive X2=1 write miss

Memory block: X1=2 X2=0

(5) P2 writes 3 to X1 (5 pnts)

In P2, cache block: Exclusive→Exclusive X1=3 write miss, write back the dirty block
In P0, cache block: Shared →Invalid
Memory block: X1=2 X2=1

10. (20 pnts) A write-back cache use the write-invalidate cache coherence protocol. In the directory protocol, the cache controller must maintain the state transitions for each cache block and each memory block. Besides, we assume that the memory addresses X1 and X2 are in different memory blocks on the processor P2 and map to the same cache block. Their initial values are 0. The initial stats of each cache block and each memory block are Invalid and Uncached, respectively. States of a memory block: Uncached, Exclusive, and Shared. States of a cache block: Invalid, Exclusive, and Shared.



Please describe

丁、the stat change of each corresponding cache block and memory bock

戊、the sharer set for the memory block

己、the values of X1 and X2 in the memory and cache

庚、what message is placed in the bus

When the following requests are issued under the directory protocol. (Hint: home node, local node, and remote node)

(1) P1 writes 1 to X1 (3 pnts)

In P1, cache block: Invalid→ Exclusive X1=1 write miss

In P2, X1 memory block: Uncached→ Exclusive Sharer={P1} X1=0, Data Value Reply

(1) P2 reads X1 (3 pnts)

In P2, cache block: Invalid→ Shared X1=1 read miss

In P1, cache block: Exclusive→Shared

In P2, X1 memory block: Exclusive→Shared Sharer={P1,P2} X1=1, Fetch, Data Value Reply

(2) P1 writes 2 to X1 (3 pnts)

In P1, cache block: Shared →Exclusive X1=2 Invalidate

In P2, cache block: Shared →Invalid

In P2, X1 memory block: Shared→Exclusive, Sharer={P1} X1=1, Invalidate

(3) P3 reads X1 (4 pnts)

In P3, cache block: Invalid→Shared X1=2 read miss

In P1, cache block: Exclusive→Shared X1=2

In P2, X1 memory block: Exclusive→Shared Sharer={P1,P3} X1=2, Fetch, Data Value Reply

(4) P1 writes 1 to X2 (3 pnts)

In P1, cache block: Shared→ Exclusive X2=1 write miss

In P2, X2 memory block: Uncached→Exclusive Sharer={P1} X2=0, Data Value Reply

X1 memory block: Shared→Shared Sharer={P3} X1=2

(5) P3 writes 2 to X2 (4 pnts)

In P3, cache block: Shared→Exclusive X2=2 write miss

In P1, cache block: Exclusive→Invalid

In P2, X1 memory block: Shared→Uncached Sharer={} X1=2

X2 memory block: Exclusive→Exclusive Sharer={P3} X2=1, Fetch/Invalidate, Data Value

Reply

[Data-Level Parallelism in Vector, SIMD, and GPU Architectures] (10 pnts)

11. (5 pnts) 如果驗一個 vector processor 系統中有 4 個 memory bank，當欲從記憶體載入資料到一個 64 個 element 的暫存器時，資料在 bank 中如何放置時，會有最大的讀取時間。

如果欲載入的資料皆是從同一個 bank 中讀取，將會有此情況

12. (5 pnts) 如果在可以放置相同電晶體個數的晶片中，分別用來設計一般的 CPU 與 GPU，為何在 GPU 上可以同時執行上百個執行緒，而在 CPU 上能夠執行的執行數只有個數個。

GPU 中的執行緒所使用的硬體資源都很單純簡單，然而 CPU 的設計則相對複雜。