1. One difference between a write-through cache and a write-back cache can be in the time it takes to write. During the first cycle, we detect whether a hit will occur, and during the second (assuming a hit) we actually write the data.

   Let's assume that 30% of the blocks are dirty for a write-back cache. For this question, assume that the write buffer for the write through will never stall the CPU (no penalty). Assume a cache read hit takes 1 clock cycle, the cache miss penalty is 100 clock cycles, and a block write from the cache to main memory takes 50 clock cycles. Finally, assume the instruction cache miss rate is 0.5% and the data cache miss rate is 2%. Assume that on average 25% and 10% of instructions in the workload are loads and stores, respectively.

   Estimate the performance of a write-through cache with a two-cycle write versus a write-back cache with a two-cycle write

2. Assume that we have an N-way set associative cache that has a block size of 8 bytes and the number of sets is 256. Assume that the replacement policy is least recently used (LRU).

   (a) If the capacity of the cache is 8192 bytes, what is then the associativity of the cache?
   (b) Sketch the layout of the cache content for the cache if we assume that N = 2. The figure should include the ways and columns for the valid bit, the tag, and the data. What is then the capacity of the cache?

3. Suppose a computer using fully associative cache has 2^{24} words of main memory, and a cache of 128 blocks, where each cache block contains 64 words

   (a) How many blocks of main memory are there?

   $2^{24}/2^{6} = 2^{18}$

   (b) What are the sizes of the tag and word fields?

   24 bit addresses with 18 bits in the tag field and 6 in the word field

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/

4. You purchased an old IBM System whose model number is unknown so you do not have access to any manuals or spec sheets of the computer – except those listed below by a previous user:

• 90% of all memory accesses are found in the cache.
• Each cache block is 4 words, and the whole block is read on any miss.
• The processor sends references to its cache at the rate of 10^9 words per second.
• 25% of those references are writes.
• Assume that the memory system can support 10^9 words per second, reads or writes.
• The bus reads or writes a single word at a time (the memory system cannot read or write two words at once).
• Assume at any one time, 40% of the blocks in the cache have been modified.
• The cache uses write allocate on a write miss.

You are considering adding an IBM compatible peripheral to the system, and you want to know *how much of the memory system bandwidth is already used*. Calculate the percentage of memory system bandwidth used on the average in the two cases below. Be sure to state your assumptions.

4.1 The cache is Write-Through.
4.2 The cache is Write-Back.

* Miss rate = 0.1

* Block size = 4 words (16 bytes)

* Frequency of memory operations from processor = 10^9

* Frequency of writes from processor = $0.25 * 10^9$

* Bus can only transfer one word at a time to/from processor/memory

* On average 40% of blocks in the cache have been modified (must be written back in the case of the write back cache)

* Cache is write allocate

So:

Fraction of read hits = $0.75 * 0.9 = 0.675$

Fraction of read misses = $0.75 * 0.1 = 0.075$

Fraction of write hits = $0.25 * 0.90 = 0.225$

Fraction of write misses = $0.25 * 0.1 = 0.025$

2 (a) Write through cache

- On a read hit there is no memory access
- On a read miss memory must send 4 words to the cache
- On a write hit the cache must send a word to memory
- On a write miss memory must send 4 words to the cache, and then the cache must send a word to memory

Thus:

Average words transferred = $0.675 * 0 + 0.075 * 4 + 0.225 * 1 + 0.0125 * 5 = 0.5875$

Average bandwidth used = $0.5875 * 10^9$

Fraction of bandwidth used =

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/

$[0.5875 \times 10^9] / 10^9$

$= 0.5875$ (1)

2(b) Write back cache

On a read hit there is no memory access

On a read miss:

1. If replaced line is modified (dirty) then cache must send 4 words to memory, and then

memory must send 4 words to the cache

2. If replaced line is clean then memory must send 4 words to the cache

On a write hit there is no memory access

On a write miss:

1. If replaced line is modified (dirty) then cache must send 4 words to memory, and then memory must send 4 words to the cache

2. If replaced line is clean then memory must send 4 words to the cache

Thus:

Average words transferred = $0.675 * 0 + 0.075 * (0.6 * 4 + 0.4 * 8) + 0.225 * 0 + 0.025 *$

$(0.6 * 4 + 0.4 * 8) = 0.42 + 0.14 = 0.56$

Average bandwidth used = $0.56 * 10^9$

Fraction of bandwidth used =

$0.56 \times 10^9/10^9$

$= 0.56$ (2)

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/

5. Given a 4-way set associative cache of total size 2MB that has a 26-bit cache address and blocks of size 8KB each, answer the following questions.

5.1 How many bits in the "index" field of the cache address? Explain your answer

5.2 How many bits in the "offset" field of the cache address? Explain your answer

5.3 How many bits in the "Tag" field of the cache address?

6. Consider the following program and cache behaviors.

Data Reads per 1K instructions	Data Writes per 1K instructions	Instruction Cache Miss Rate	Data Cache Miss Rate	Block Size (Bytes)
400	200	0.5%	2%	64

Suppose a CPU with a write-through, write allocate cache achieves a CPI of 2. What are the read and write bandwidths (measured by bytes per cycle) between RAM and the cache? (Assume each miss generates a request for one block.). *For a write-allocate policy, a write miss also makes a read request to RAM – please be sure to consider its impact on Read Bandwidth*

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/

Problem similar to Question 1 in this test (numbers are modified):

One difference between a write-through cache and a write-back cache can be in the time it takes to write. During the first cycle, we detect whether a hit will occur, and during the second (assuming a hit) we actually write the data.

Let's assume that 50% of the blocks are dirty for a write-back cache. For this question, assume that the write buffer for the write through will never stall the CPU (no penalty). Assume a cache read hit takes 1 clock cycle, the cache miss penalty is 50 clock cycles, and a block write from the cache to main memory takes 50 clock cycles. Finally, assume the instruction cache miss rate is 0.5% and the data cache miss rate is 1%. Assuming that on average 26% and 9% of instructions in the workload are loads and stores, respectively, **estimate the performance of a write-through cache with a two-cycle write versus a write-back cache with a two-cycle write.**

CPU performance equation: CPUTime = IC * CPI * ClockTime

CPI = CPIexecution + StallCyclesPerInstruction

We know:

Instruction miss penalty is 50 cycles

Data read hit takes 1 cycle

Data write hit takes 2 cycles

Data miss penalty is 50 cycles for write through cache

Data miss penalty is 50 cycles or 100 cycles for write back cache

Miss rate is 1% for data cache (MRD) and 0.5% for instruction cache (MRI)

50% of cache blocks are dirty in the write back cache

26% of all instructions are loads

9% of all instructions are stores

Then: CPIexecution = 0.26 * 1 + 0.09 * 2 + 0.65 * 1 = 1.09

Write through

StallCyclesPerInstruction = MRI * 50 + MRD * (0.26 * 50 + 0.09 * 50) = 0.425

so: CPI = 1.09 + 0.425 = 1.515 (1)

Write back

StallCyclesPerInstruction = MRI * 50 + MRD * (0.26 * (0.5 * 50 + 0.5 * 100) + 0.09 * (0.5 * 50 + 0.5 * 100)) = 0.5125

so: CP I = 1.09 + 0.5125 = 1.6025 (2)

Comparing 1 and 2 we notice that the system with the write back cache is 6% slower.

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/

Problem similar to Question 6 in this test (numbers are modified):

Consider the following program and cache behaviors.

Data Reads per 1K instructions	Data Writes per 1K instructions	Instruction Cache Miss Rate	Data Cache Miss Rate	Block Size (Bytes)
300	150	0.5%	5%	128

Suppose a CPU with a write-through, write allocate cache achieves a CPI of 2.

4.1 What are the read and write bandwidths (measured by bytes per cycle) between RAM and the cache? (Assume each miss generates a request for one block.). *For a write-allocate policy, a write miss also makes a read request to RAM – please be sure to consider its impact on Read Bandwidth*

Instruction Bandwidth:
When the CPI is 2, there are, on average, 0.5 instruction accesses per cycle.
0.5 instructions read from Instruction memory per cycle
0.5% of these *instruction* accesses cause a cache ***Read*** miss (and subsequent memory request).
[0.5 instr/cycle] x [0.005 misses/instruction] = missed instructions/cycle
Assuming each miss requests one block and each block is 128 bytes [16 words with 8 bytes (64 bits) per word] , instruction accesses generate an average of
[0.5 instr/cycle] x [0.005 misses/instruction] x[128 bytes/miss] =
= 0.32 bytes/cycle of *read traffic*

Read Data bandwidth:
30% of instructions generate a ***read*** request from data memory.
[0.5 instr/cycle] x [0.3 Read Data Accesses/instruction] = [0.15 Read Data Accesses / cycle]
5% of these generate a cache miss;
[0.15 Read Data Accesses / cycle] x [0.05 misses / Read Data Access] = 0.0075 Read Misses/cycle
Assuming each miss requests one block and each block is 128 bytes [16 words with 8 bytes (64 bits) per word] ,
[0.0075 Read Misses/cycle] x [128 Bytes/block] x [1 block/miss] = 0.0075 x 128 Bytes/cycle
= 0.96 Bytes/cycle
Write Data bandwidth:
15% of instructions generate a ***write*** request into data memory.
[0.5 instr/cycle] x [0.15 Write Data Accesses/instruction] = [0.075 Write Data Accesses / cycle]
All of the words written to the cache must be written into Memory:
[0.075 Write Data Accesses / cycle] x [8 bytes/word] x [1 word/write-through] = 0.6 Bytes/cycle

For a *Write-allocate policy*, a Write miss also makes *a **read** request* to RAM
[0.5 inst/cycle] x [0.15 Write Data Accesses/instruction] x [0.05 misses/Write Data Access] x [128 Bytes/miss]
= 0.48 Bytes/cycle
Assuming each miss requests one Word (8 bytes) since this is a write-through cache *with only 1 word written per miss into memory*,
[0.00375 Write Misses/cycle] x [8 Bytes/word] x [1 word/miss] = 0.03 Bytes/cycle

Total Read Bandwidth
0.32 (Instruction memory) + **0.96** (data memory) + **0.48** (Write-miss in Write-through cache with Write Allocate)
Bytes/cyle = 1.76 Bytes/cycle
Total Write Bandwidth:
0.6 Bytes/cycle + 0.03 Bytes/cycle = 0.63 Bytes/cycle

This study source was downloaded by 100000877991777 from CourseHero.com on 12-19-2023 13:51:39 GMT -06:00

https://www.coursehero.com/file/202807397/ECE-6913-Final-Section-A-solpdf/