COE 308 – Computer Architecture

Assignment 8: Memory Hierarchy

Solution

1. **(4 pts)** Consider a processor with a 2 ns clock cycle, a miss penalty of 20 clock cycles, a miss rate of 0.05 misses per instruction, and a cache access time (hit time) of 1 clock cycle. Assume that the read and write miss penalties are the same.

a) **(1 pt)** Find the average memory access time (AMAT).

b) **(1 pt)** Suppose we can improve the miss rate to 0.03 misses per instruction by doubling the cache size. However, this causes the cache access time to increase to 1.2 cycles. Using the AMAT as a metric, determine if this is a good trade-off.

c) **(2 pts)** If the cache access time determines the processor's clock cycle time, which is often the case, AMAT may not correctly indicate whether one cache organization is better than another. If the processor's clock cycle time must be changed to match that of a cache, is this a good trade-off? Assume that the processors in part **(a)** and **(b)** are identical, except for the clock rate and the cache miss rate. Assume 1.5 references per instruction (for both I-cache and D-cache) and a CPI without cache misses of 2. The miss penalty is 20 cycles for both processors.

Solution:

a) **AMAT = Hit time + Miss rate × Miss penalty**

= 2 ns + 0.05 × (20 × 2 ns) = 4 ns

b) **AMAT = 1.2 × 2 ns + 0.03 × 20 × 2 ns = 2.4 ns + 1.2 ns = 3.6 ns**

Yes, this is a good trade-off.

c) **CPU time = Clock cycle × IC × ($CPI_{ideal\text{-}cache}$ + cache stall cycles per instruction)**

CPU time(a) = 2 ns × IC × (2 + 1.5 × 20 × 0.05) = 7 × IC

CPU time(b) = 2.4 ns × IC × (2 + 1.5 × 20 × 0.03) = 6.96 × IC

The CPU times in parts (a) and (b) are almost identical. Hence, doubling the cache size to improve the miss rate at the expense of stretching the clock cycle results in essentially no net gain.

2. (5 pts) Consider three processors with three cache configurations:

Processor 1: Direct-mapped i-cache and d-cache with one-word blocks
Instruction miss-rate = 4%, data miss-rate = 6%

Processor 2: Direct-mapped i-cache and d-cache with four-word blocks
Instruction miss-rate = 2%, data miss-rate = 4%

Processor 3: Two-way set associative i-cache and d-cache with four-word blocks
Instruction miss-rate = 2%, data miss-rate = 3%

a) (3 pts) For these processors, 50% of the instructions contain a data reference. Assume that the cache penalty is 6 + Block size in words. Determine which processor spends the most cycles on cache misses.

b) (2 pts) The cycle time is 420 ps for Processor 1 and 2, and 310 ps for the third processor. Determine which processor is the fastest and which one is the slowest.

Solution:

a) For Processor 1:
Miss penalty = 6 + 1 = 7 cycles
Stall cycles per instruction = 4% × 7 + 50% × 6% × 7 = 0.28 + 0.21 = 0.49

For Processors 2:
Miss penalty = 6 + 4 = 10 cycles
Stall cycles per instruction = 2% × 10 + 50% × 4% × 10 = 0.2 + 0.2 = 0.4

For Processor 3:
Miss penalty = 6 + 4 = 10 cycles
Stall cycles per instruction = 2% × 10 + 50% × 3% × 10 = 0.2 + 0.15 = 0.35

Therefore, Processor 1 spends the most cycles on cache misses.

b) CPU time = IC × CPI × Clock cycle

Instruction count is same for all processors

CPI = $CPI_{ideal\text{-}cache}$ + Stall cycles per instruction

$CPI_{ideal\text{-}cache}$ is the same for all processors

For processor 1: CPU time = IC × ($CPI_{ideal\text{-}cache}$ + 0.49) × 420 ps
For processor 2: CPU time = IC × ($CPI_{ideal\text{-}cache}$ + 0.4) × 420 ps
For processor 3: CPU time = IC × ($CPI_{ideal\text{-}cache}$ + 0.35) × 310 ps

Clearly, Processor 1 is the slowest and Processor 3 is the fastest.

3. **(5 pts)** A computer system has a 1 GB main memory. It also has a 4K-Byte cache organized as a 4-way set-associative, with 4 blocks per set and 64 bytes per block.

a) **(2 pts)** Calculate the number of bits in the Tag, Set Index, and Byte Offset fields of the memory address format.

b) **(3 pts)** Assume that the cache is initially empty. Suppose the processor fetches 4352 consecutive bytes from memory starting at address 0. The same fetch sequence is then repeated 9 more times for a total of 10 iterations. What is the hit rate assuming that the LRU algorithm is used for block replacement?

Solution:

a) **Numbers of Sets = (4 * 1024) / (4 * 64) = 16 sets, Set index = 4 bits, Byte offset = 6 bits**

Address = 30 bits (1 GB main memory), Tag = 30 – 4 – 6 = 20 bits

b) **4352 / 64 = 68 consecutive blocks are being fetched**

The cache can store only 64 blocks. The last 4 fetched blocks will replace the first 4 blocks.

The first iteration causes 68 cache misses to fetch the 68 blocks from memory

Then each iteration causes 8 cache misses only to fetch the first 4 and the last 4 blocks

Therefore, total cache misses = 68 + 9 * 8 = 140

If the processor is loading individual bytes from the cache, then

Total accesses to the cache = 4352 * 10 = 43520

Miss Rate = 140 / 43520 = 0.32%, Hit Rate = 100% – Miss Rate = 99.68%

If the processor is loading words from the cache and word size = 4 bytes, then

Total accesses to the cache = (4352 / 4) * 10 = 10880

Miss Rate = 140 / 10880 = 1.287%, Hit Hate = 100% – Miss Rate = 98.7%

4. **(3 pts)** Consider a main memory constructed with Synchronous DRAM chips that have the following timing requirements: 1 bus cycle to transfer the address, 10 bus cycles access latency, and 1 bus cycle to transfer a word. Assume that 32-bits of data can be transferred in parallel. If a 200-MHz clock is used for the bus and memory, and burst mode is used to transfer a block, how long does it take to access and transfer 32 bytes of data, 64 bytes of data, and 128 bytes of data?

Solution:

Clock Rate = 200 MHz, Clock cycle for bus and memory = 5 ns

Time to transfer 32 bytes (8 words) = 1 cycle (address) + 10 cycles (latency to get first word) + 7 cycles (7 remaining words) = 18 bus cycles = 18 * 5 ns = 90 ns

Time to transfer 64 bytes (16 word) = 1 + 10 + 15 = 26 bus cycles = 26 * 5 ns = 130 ns

Time to transfer 128 bytes (32 words) = 1 + 10 + 31 = 42 bus cycles = 42 * 5 ns = 210 ns

5. **(3 pts)** Assume a memory system that supports interleaving of either four reads or four writes. Given the following memory addresses in order as they appear on the memory address bus: 3, 9, 17, 2, 51, 37, 13, 4, 8, 41, 67, 10, which ones will result in a bank conflict?

Solution:

Address	Bank	Bank Conflict
3	**3**	**No**
9	**1**	**No**
17	**1**	**Yes (with 9)**
2	**2**	**No**
51	**3**	**No**
37	**1**	**Yes (with 17)**
13	**1**	**Yes (with 37)**
4	**0**	**No**
8	**0**	**Yes (with 4)**
41	**1**	**No**
67	**3**	**No**
10	**2**	**No**