COEN 122: Computer Architecture
Homework #5

Tamir Enkhjargal

March 11, 2020

Question #1

Cache block size (B) can affect both miss rate and miss latency. Assuming a machine with a base CPI of 1, and an average of 1.35 references (both instruction and data) per instruction, find the block size that minimizes the total miss latency given the following miss rates for various block sizes.

8: 4%, 16: 3%, 32: 2%, 64: 1.5%, 128: 1%

1. What is the optimal block size for a miss latency of 20 x B cycles?

$$AMAT = Hit + (miss\ rate * latency)$$
$$AMAT = 1 + 0.04 * 160 = 7.4$$
$$AMAT = 1 + 0.03 * 320 = 10.6$$
$$AMAT = 1 + 0.02 * 640 = 13.8$$
$$AMAT = 1 + 0.015 * 1280 = 20.2$$
$$AMAT = 1 + 0.01 * 2560 = 26.6$$

   The best average memory access time is the 8 B block.

2. What is the optimal block size for a miss latency of 24 + B cycles?

$$AMAT = Hit + (miss\ rate * latency)$$
$$AMAT = 1 + 0.04 * 32 = 2.28$$
$$AMAT = 1 + 0.03 * 40 = 2.2$$
$$AMAT = 1 + 0.02 * 56 = 2.12$$
$$AMAT = 1 + 0.015 * 88 = 2.32$$
$$AMAT = 1 + 0.01 * 152 = 2.52$$

   The best average memory access time is the 32 B block.

3. For constant miss latency, what is the optimal block size?

   With constant miss latency, the optimal block size is the one with the lowest miss rate, which is a 128 B block size.

Question #2

An 8 KB direct-mapped cache with 2^8 cache lines.

What is the line size? How many sets does it have?

8 KB = 2^{13} total size. If we split into 2^8 cache lines, then there are 256 cache lines, with each holding $2^{13-8} = 2^5 = 32 = 8$ words.

Question #3

A 4-way associative cache has 256 lines, each is 32 bytes long.

1. What is the cache size?

   $Cache\ Size = Lines * Line\ Size$
   $= 256 * 32 = 8192 = 2^{13} = 8$ KB

2. How many sets does it have?

   Each set has 4 lines, so total 256 lines means number of sets = 256/4 = 64 sets.

3. What is the total number of valid bits it has? $Valid\ Bits = Sets + Offset$ $= 6 + 5 = 11$ bits.

Question #4

Consider a memory system with a cache access time of 100 ns. If the hit ratio is 90% and effective access time is 10% greater than the cache access time, what is memory access time? Show your work.

Cache access time = 100ns
Effective access time = 110ns
110 = 0.9*100+0.1*x
x = 200

A hit would mean the data was in cache, and a miss would need to access the memory, which is 200ns.

Question #5

Assume a 32 bit address, 20 bits tag, and 12 bits page size. How many page table levels do we need, if every page table fits in 2 pages, and every page table entry is 4 bytes? Show your work.

We have 32 bits total. We have 12 bits for page size, 1 bit for page number per table, and 5 bits for each page entry.

32 - 12 - 1 - 5 = 14 bits. 2^{14} = 16 KB page tables. We would access a single page through the entry bit (1) and page number (1). Therefore we would need a minimum of 2 page table levels to account for each page. We could increase the number of levels to account for each entry, and depending on how far into each page we want. However, we need minimum two page table levels, and can increase more for better organization.