CENG3420 Homework 3

Due: Apr. 24, 2020

Please submit PDF or WORD document directly onto blackboard.
DO NOT SUBMIT COMPRESSED ZIP or TARBALL.

Solutions

Q1 (20%) For two direct-mapped cache design with a 32-bit address and a 16-bit address, the following bits of the address are used to access the cache (as in Table 1).

Table 1: Q1

	Tag	Index	Offset
a	31-10	9-5	4-0
b	15-10	9-4	3-0

1. What is the cache line size (in word)?
2. How many entries does the cache have?
3. What is the ratio between total bits required for such a cache implementation over the data storage bits?

A1

1. a. The offset is 4-0 (5 bits). It implies 2^5bytes$=32$bytes$=8$words.
   b. The offset is 3-0 (4 bits). It implies 2^4bytes$=16$bytes$=8$words.
2. a. The range of index is 9-5 (5 bits). For the direct-mapped cache, it implies 2^5sets$=32$entries.
   b. The range of index is 9-4 (6 bits). For the direct-mapped cache, it implies 2^6sets$=64$entries.
3. a. Total bits=32 entries$*$(1 valid bit$+$22 tag bits$+32*8$ data bits$)=32*279$,
   Data bits$=32$ entries$*32*8$ data bits$=32*256$,
   Ratio$=279/256=1.09$.
   b. Total bits=64 entries$*$(1 valid bit$+$6 tag bits$+16*8$ data bits$)=64*135$,
   Data bits$=64$ entries$*16*8$ data bits$=64*128$,
   Ratio$=135/128=1.05$.

Q2 (15%) Suppose there is a byte addressable computer with main memory size 256MB and a cache with 8 lines. Each cache line size is 64B. Assume **direct mapping** in the memory hierarchy. Calculate the following items.

1. How many bits do we need at least for the main memory address? The total cache size (in bits).
2. The line number corresponding to the main memory address 2333_{10}.

A2

1. All of the following answers are correct (note: analysing one senario is enough).

(a) Assume the main memory address space is up to 256MB, then we have 28bit main memory address, thus the total cache size in bits is $8 \times (1 + 19 + 512) = 4256$bits.

(b) Assume 32-bit main memory address, the total cache size in bits is $8 \times (1 + 23 + 512) = 4288$bits.

(c) Assume 16-bit main memory address, the total cache size in bits is $8 \times (1 + 7 + 512) = 4160$bits.

2. All of the following answers are correct

(a) $\lceil 2333B/64B \rceil = 37$, line number $= 37 \mod 8 = 5$.

(b) $\lfloor 2333B/64B \rfloor = 36$, line number $= 36 \mod 8 = 4$.

Q3 (10%) Consider a cache works at 5× speed of main memory with hit rate of 75%. What is the speedup of the memory performance if such cache is used.

A3 Let cache access time to be t, then the main memory access time is $5t$. The system average access time is,

$$T_a = 0.75t + 0.25 \times 5t = 2.0t. \tag{1}$$

Then the performance speedup is $5t/2.0t = 2.5$.

Q4 (15%) There is a computer that has 64MB byte-addressable main memory. Instructions and data are stored in separated caches, each of which has eight 64B cache lines. The data cache use **direct-mapping**. Now there are two programs in the following form. `Program A:`

```
int a[64][64];
int sum_array1()
{
        int i,j,sum=0;
        for(i=0;i<64;i++)
                for(j=0;j<64;j++)
                        sum += a[i][j];
        return sum;
}
```

`Program B:`

```
int a[64][64];
int sum_array1()
{
        int i,j,sum=0;
        for(j=0;j<64;j++)
                for(i=0;i<64;i++)
                        sum += a[i][j];
        return sum;
}
```

Suppose `int` data is represented in 32-bit 2's complement and i,j,sum are stored in specific registers. Arrays are stored in row-major with the start address 320_{10} in the main memory. Answer the following questions.

1. What are the line numbers of the main memory blocks that contain `a[0][31]` and `a[2][2]` respectively? (Cache line number starts from 0)
2. What are the data cache hit rates of program A and B?

A4

1. a. $320 + 31 * 4 = 444$, $444/64 = 6$,
   b. $320 + (64 * 2 + 2) * 4 = 2376$, $2376/64 = 37$, $37 \mod 8 = 5$.
2. The size of Array a is $64 * 64 * 4 = 2^{14}$B. Since the size of a block is 64B, it takes 2^8 main memory blocks to store it. Under the condition of row-major, 2^8 times of cache miss will appear. Therefore, the hit rate of program A is $(2^{12} - 2^8)/2^{12} = 93.75\%$. For program B, the hit rate is 0.

Q5 (15%) A byte-addressable computer uses 32-bit address to access main memory. Suppose the data cache has a size of 4KiB and works in **8-way set-associative**. Each block size is 16B.

1. How many bits in Tag, Index and Offset?
2. Calculate the total cache size (in bits) if it uses write back and LRU replacement.

A5

1. The number of blocks is 4 KiB $/16$ B$= 2^8$. Therefore, $2^8/8 = 2^5$ rows need 5 bits to be indexed. Since the computer is byte-addressable, the bits of offset is $\log_2(2^4 B) = 4$. And the tag bits is calculated as: $32 - 5 - 4 = 23$ bits.
2. $(23 \text{ Tag bits } + 16 * 8 \text{ Data bits} + 1 \text{ Valid bit} + 1 \text{ Dirty bit} + 3 \text{ LRU bits}) * 2^8 = 39936$ bits.

Q6 (20%) There are several parameters that impact the overall size of the page table. Listed below are key page table parameters.

Table 2: Q6

Virtual Address Size	Page Size	Page Table Entry Size
32	16KiB	4bytes

1. Given the parameters shown above, calculate the total page table size for a system running 5 applications that utilize half of the memory available (half of the 32-bit virtual address space for each running).
2. A cache designer wants to increase the size of a virtually indexed, physically tagged cache. Given the page size shown above, is it possible to make a 64KiB direct-mapped cache, assuming 2 words per block?
3. How would the designer increase the data size of the cache?

A6

1. The tag size is $32 - \log_2(16384) = 32 - 14 = 18$ bits. All five page tables would require $5 * (2^{18} * 4)/2$ bytes $= 2621440.0$ B,
2. The page index consists of address bits 13 down to 0. So the LSB of the tag is address bit 14. A 64 KiB direct-mapped cache with 2-words per block would have 8-byte blocks and thus 64 KiB / 8 bytes = 8192 blocks, and its index field would span address bits 16 down to 3 (13 bits to index, 1 bit word offset, 2 bit byte offset). As such, the tag LSB of the cache tag is address bit 16,

3. The designer would instead need to make the cache 4-way associative to increase its size to 64 KB.

Q7 (5%) Suppose a processor supports software configuration on handling cache write hits/miss strategy.

1. What strategy should be taken if:
   (a) The processor mainly works on applications that require massive memory write operation and data access.
   (b) The processor mainly works on applications that require massive memory write operation and data access but do not allow data inconsistency.

   Explain why.
2. Exlain why virtual memory only uses **write-back** as its write strategy.

A7
1. (a) Write Back, reduce main memory access.
   (b) Write Through, keep data consistently.
2. Hard disk is much slower than main memory.