CENG3420 Homework 3

Due: Apr. 13, 2019

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

Solutions

Q1 (16%) 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. The total cache size (in bits).
2. The line number corresponding to the main memory address 3200_{10}.

A1

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. $3200B/64B = 50$, line number = $50 \mod 8$ =2.

Q2 (10%) Assume it takes one clock to send address to DRAM memory and one clock to send data back. DRAM has 7 cycle latency for first byte, and 4 cycles for each of subsequent bytes in the block. To transfer a 8-byte block, calculate the cycle number if we need:

1. non-interleaving;
2. 4-module interleaving.

A2

1. 37
2. 1+7+1×8=16

Q3 (15%) Consider a cache works at 5× speed of main memory with hit rate of 95%. 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,

$$T_a = 0.95t + 0.05 \times 5t = 1.2t. \quad (1)$$

Then the performance speedup is $5t/1.2t = 4.17$.

Q4 (14%) There is a computer that has 256MB 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[256][256];
int sum_array1()
{
        int i,j,sum=0;
        for(i=0;i<256;i++)
                for(j=0;j<256;j++)
                        sum += a[i][j];
        return sum;
}
```

Program B:

```
int a[256][256];
int sum_array1()
{
        int i,j,sum=0;
        for(j=0;j<256;j++)
                for(i=0;i<256;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[1][1]` respectively? (Cache line number starts from 0)
2. What are the data cache hit rates of program A and B?

A4 1. 6, 5.

2. Array a has size of $256 \times 256 \times 4$B=2^{18}B, that takes 2^{12} main memory blocks. Under the condition of row-major, there are 2^{12} times of cache miss, thus the hit rate for program A is $(2^{16} - 2^{12})/2^{16} = 93.75\%$. 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 contains 128 blocks and works in **4-way set-associative**. Each block size is 64B with one "valid" bit.

1. (9%) How many bits in Tag, Index and Offset?
2. (6%) Calculate the total cache size (in bits).

A5 1. 21-5-6.

2. $128 \times (21 + 1 + 64 \times 8) = 68352$ bits.

Q6 (15%) Describe two types of localities that are exhibited in memory hierarchy. Which locality is related to the cache replacement strategy "Least recently used"?

A6 Trivial.

Q7 (15%) There is a computer storage system that is byte addressable with virtual memory size 16MB and main memory size 1MB, page size is 4KB. How many bits are there in the virtual address and how many bits of them are for virtual page number.

A7 16MB=2^{24}B, thus the virtual address is 24bit; page size is 4KB=2^{12}B, thus the high 12bits are for virtual page number.