Assignment5

5.3

For a direct-mapped cache design with a 32-bit address, the following bits of the address are used to access the cache.

Tag	Index	Offset
31–10	9–5	4–0

5.3.1

What is the cache block size (in words)?

The offset contains 5 bits, thus the block size is 32 bytes/block

Since 1 word = 4 byte, thus the cache block size is

$$\frac{2^5}{4} = 2^3 = 8(word/block)$$

5.3.2

How many entries does the cache have?

The index contains 5 bits, thus the the cache have $2^5 = 32$ entries.

5.3.3

What is the ratio between total bits required for such a cache implementation over the data storage bits?

The data storage bits = $8 \times 32 \times 32 = 8192(bit)$

The total bits requirse = $32 \times (8 \times 32 + (32 - 5 - 3 - 2) + 1) = 8928(bit)$

The ratio is:

$$\frac{8928}{8192} = 1.089$$

5.3.4

Starting from power on, the following byte-addressed cache references are recorded.

Address											
0	4	16	132	232	160	1024	30	140	3100	180	2180

How many blocks are replaced?

ADDRESS	BINARY	INDEX	TAG	SITUATION
0	0000 0000 0000	00000	00	miss
4	0000 0000 0100	00000	00	hit
16	0000 0001 0000	00000	00	hit
132	0000 1000 0100	00100	00	miss
232	0000 1110 1000	00111	00	miss
160	0000 1010 0000	00101	00	miss
1024	0100 0000 0000	00000	01	miss + replace
30	0000 0001 1110	00000	00	miss + replace
140	0000 1000 1100	00100	00	hit
3100	1100 0001 1100	00000	11	miss + replace
180	0000 1011 0111	00101	00	hit
2180	1000 1000 0100	00100	10	miss + replace

5.3.5

What is the hit ratio?

The hit ratio is:

$$\frac{4}{12} = \frac{1}{3}$$

5.3.6

List the final state of the cache, with each valid entry represented as a record of <index, tag, data>

The final state will be

INDEX	TAG	DATA ADDRESS IN THE MEMORY
00000	11	11 00000 00000 ~ 11 00000 11111
00100	00	10 00100 00000 ~ 10 00100 11111
00101	00	00 00101 00000 ~ 00 00101 11111
00111	00	00 00111 00000 ~ 00 00111 11111

5.6

In this exercise, we will look at the different ways capacity affects overall performance. In general, cache access time is proportional to capacity. Assume that main memory accesses take `70 ns` and that memory accesses are `36%` of all instructions. The following table shows data for L1 caches attached to each of two processors, P1 and P2.

	L1 Size	L1 Miss Rate	L1 Hit Time
P1	2 KiB	8.0%	0.66 ns
P2	4 KiB	6.0%	0.90 ns

5.6.1

Assuming that the L1 hit time determines the cycle times for P1 and P2, what are their respective clock rates?

- $Clock\ \ rate_{P1} = \frac{1}{0.66 \times 10^9} = 1.52 \times 10^9 Hz = 1.52 GHz$
- $Clock\ \ rate_{P2} = \frac{1}{0.90 \times 10^9} = 1.11 \times 10^9 Hz = 1.11 GHz$

5.6.2

What is the Average Memory Access Time for P1 and P2?

The cycle time caused due to the memory access is

- $Memory\ \ Access\ \ Time_{P1} = \frac{70ns}{0.66ns} = 107(cycles)$
- $Memory\ \ Access\ \ Time_{P1} = \frac{70ns}{0.9ns} = 78(cycles)$

Thus the AMAT are

- $AMAT_{P1} = 1 + 8\% \times 107 = 9.56(cycle)$
- $AMAT_{P2} = 1 + 6\% \times 78 = 5.68(cycle)$

5.6.3

Assuming a base CPI of 1.0 without any memory stalls, what is the total CPI for P1 and P2? Which processor is faster?

Suppose we have I instructions

Since

- $Clock\ \ Cycle\ \ Time_{P1} = 0.66ns$
- $Clock\ \ Cycle\ \ Time_{P2} = 0.90ns$

The cycle time caused due to the memory access is

- $Memory\ \ Access\ \ Time_{P1} = \frac{70ns}{0.66ns} = 107(cycles)$
- $Memory\ \ Access\ \ Time_{P1} = \frac{70ns}{0.9ns} = 78(cycles)$

Thus the cycle time we get due to the loss is

- $Loss\ \ Cycle\ \ Time_{P1} = I \times (1 + 36\%) \times 8\% \times 107 = 11.6416$
- $Loss\ \ Cycle\ \ Time_{P2} = I \times (1 + 36\%) \times 6\% \times 78 = 6.3648$

And the overall CPI of P1 and P2 is

- $CPI_{P1} = 11.6416$
- $CPI_{P2} = 7.3648$

Thus, P2 is faster

5.6.4

For the next three problems, we will consider the addition of an L2 cache to P1 to presumably make up for its limited L1 cache capacity. Use the L1 cache capacities and hit times from the previous table when solving these problems. The L2 miss rate indicated is its local miss rate.

L2 Size	L2 Miss Rate	L2 Hit Time
1 MiB	95%	5.62 ns

What is the AMAT for P1 with the addition of an L2 cache? Is the AMAT better or worse with the L2 cache?

- if the data is in L2, then P1 needs $\frac{5.62}{0.66} = 8.52 \approx 9(cycle)$ to get this data into L1
- if the data is not in L2, then P1 needs 107 cycles to get this data

Thus, the total miss penalty is $9 + 0.95 \times 107 = 110.65(cycle)$

$AMAT_{P1'} = 1 + 8\% \times 110.65 = 9.852(cycle)$

It is even worse with L2.

5.6.5

Assuming a base CPI of 1.0 without any memory stalls, what is the total CPI for P1 with the addition of an L2 cache?

$$CPI_{P1'} = 1 + (1 + 36\%) \times 8\% \times 110.65 = 13.03872$$

5.6.6

Which processor is faster, now that P1 has an L2 cache? If P1 is faster, what miss rate would P2 need in its L1 cache to match P1's performance? If P2 is faster, what miss rate would P1 need in its L1 cache to match P2's performance?

- The average time need to execution an instruction on P1' is $13.03872 \times 0.66ns = 8.61(ns)$
- The average time need to execution an instruction on P1' is $7.3648 \times 0.9ns = 6.62832(ns)$

Since P2 is faster, if P1 has a new CPI $CPI_{P1''}$, which satisfy

$$CPI_{P1''} \times 0.66ns = 6.62832ns$$

We can find $CPI_{P1''} = 10.04$

Suppose the new miss rate is R, we have

$$CPI_{P1''} = 1 + (1 + 36\%) \times R \times 110.65 = 10.04$$

Solve the equation, we get

$$R = 0.06 = 6\%$$