

컴퓨터구조론 (컴퓨터공학과 003/004, 인공지능공학과 001) Homework 2

1. 32비트 주소가 다음과 같이 구분되어 direct-mapped cache 접근에 사용된다. 한 word의 크기는 4 byte이다.

Tag	Index	Offset
31-10	9-5	4-0

- 1.1. cache block의 크기는 몇 워드인가? (2 points)

Offset이 4~0 (5bit) 이기 때문에

$$2^5 \text{ byte} = 32 \text{ byte} \underline{8 \text{ word}}$$

- 1.2. cache entry는 몇 개인가? (2 points)

엔트리인 index에서 결정됨

$$2^5 = 32$$

- 1.3. 캐시 구현에 필요한 총 bit 수와 데이터를 저장하는 부분의 bit 수의 비율은 어떻게 되는가? (3 points)

32개의 엔트리

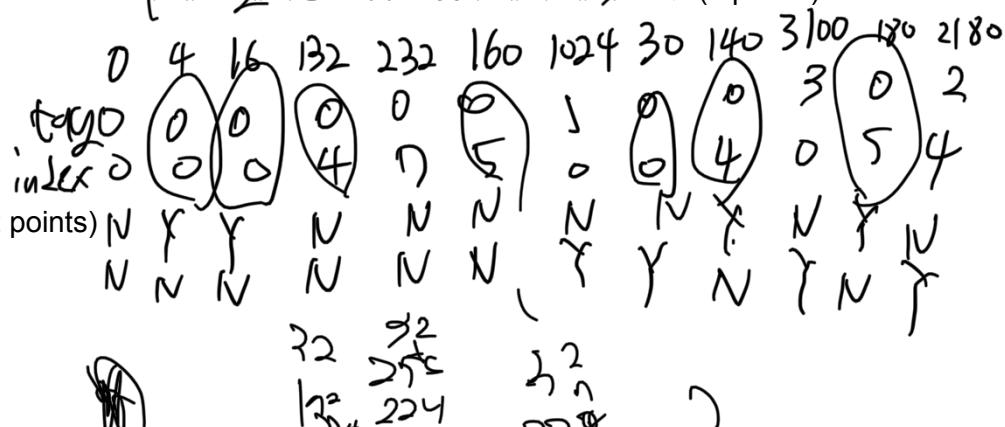
$$\begin{aligned} \text{엔트리마다 } & 1 \text{ valid bit} + 2 \text{ tag bits} + 32 \cdot 8 \text{ data} = 279 \times 32 \\ & 7 \text{ byte} \quad \text{data} \quad 32 \times 8 = 256 \times 32 \quad \frac{256}{279} = 1.09 \end{aligned}$$

전원이 켜진 후, 다음과 같은 순서로 캐시가 참조된다.

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

- 1.4. Block이 몇 번이나 교체되는가? 빈 block에 새로 대입되는 것은 횟수에서 제외한다. (2 points)

5



- 1.5. Cache hit ratio는 얼마인가? (2 points)

5
12

- 1.6. 캐시의 마지막 상태를 보이되, 유효한 엔트리는 $\langle \text{index}, \text{tag}, \text{data} \rangle$ 형태로 보여라. 주소에 대한 data는 mem[N]의 형태로 표시하라. (4 points)

0, 3, Mem(3100)				4, 2, Mem(2480)	5, 0 Mem(180)		7, 0 Mem(232)

132
232
180
4

90 304 3012 10 98 6

2. 이 연습문제에서는 용량이 전체 성능에 영향을 미치는 다양한 방법을 볼 것이다. 일반적으로 cache 접근시간은 용량에 비례한다. 메인 메모리에 접근하는데 걸리는 시간이 70 ns이고, 전체 명령어의 36%가 메모리 접근 명령이라고 가정하자. 아래 표는 두 프로세서 P1과 P2의 L1 cache에 대한 값이다.

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

2.1. L1의 hit time이 프로세서의 clock cycle time을 결정한다고 하면, P1과 P2의 clock rate은 각각 어떻게 되는가? (2 points)

$$P_1 = \frac{1}{0.66} = 1.55 \text{ GHz}$$

$$P_2 = \frac{1}{0.9} = 1.11 \text{ GHz}$$

2.2. P1과 P2 각각의 Average Memory Access Time (AMAT)는 얼마인가? (2 points)

$$0.66 + (0.08 \times 70) = 6.26 \text{ ns}$$

$$0.9 + (0.06 \times 70) = 5.1 \text{ ns}$$

2.3. 메모리로 인한 stall이 전혀 없을 때의 기본 CPI가 1이라고 할 때, 메모리로 인한 stall을 고려한다면 P1과 P2의 CPI는 어떻게 되는가? 어떤 프로세스가 더 빠른가? (3 points)

$$P_1 \rightarrow 1 + (70 \times 0.08 / 0.66) 0.36 = 4.054 \quad P_2 \rightarrow 1 + (70 \times 0.06 / 0.9) 0.36 = 2.68$$

P_2 가 2배 빠릅니다

다음 세 문제에서는 L1 cache 용량이 부족한 점을 보완하기 위해 P1에 L2 cache를 추가하는 상황을 검토한다. 문제 풀이에 앞 표의 L1 cache의 용량과 적중시간을 사용하라. L2 miss rate은 L2 cache의 local miss rate이다. (local miss rate은 L2 misses / L2 accesses를 뜻한다)

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

2.4. L2 cache가 추가된 P1의 AMAT는 얼마인가? L2 cache의 추가로 AMAT는 더 좋아졌는가 아니면 나빠졌는가? (2 points)

$$0.66 + 0.08 (5.62 + 0.95 \times 70) \rightarrow 6.4296 \text{ ns} \text{ 나빠짐}$$

2.5. 메모리 stall이 전혀 없을 때의 기본 CPI가 1이라면 L2 cache가 추가된 P1에서 메모리 stall을 고려한 CPI는 얼마인가? (2 points)

$$1 + 0.36 (0.08 (5.62 + 0.95 \times 70) / 0.66) = 4.14$$

2.6. P2와 L2 cache가 있는 P1 중 어떤 것이 더 빠른가? 만약 P1이 더 빠르다면, P2의 L1 cache miss rate이 얼마가 되어야 P1과 같은 성능을 낼 수 있는가? 만약 P2가 빠르다면, P1의 L1 cache miss rate이 얼마가 되어야 P2와 같은 성능을 낼 수 있는가? 이 경우, P1의 L2 local miss rate은 항상 일정하다고 가정한다. 또한 각 프로세서의 clock cycle time은 2.1번의 값을 사용한다. (4 points)

P_2 가 2배 빠릅니다

$$P_2 \rightarrow 2.68 \times 0.9 = 2.4 \text{ ns}$$

$$0.66 \times (0.36 \times 2.1 \times 10^6) = 4.512 \text{ ns}$$

$$\alpha = 0.06 \text{ ns} \rightarrow 69\% \text{ miss rate}$$

2.4)

컴퓨터구조론 (컴퓨터공학과 003/004, 인공지능공학과 001) Homework 1

- 1.** For a direct-mapped cache design with a 32-bit address, the following bits of the address are used to access the cache. The size of the word is 4 bytes.

Tag	Index	Offset
31-10	9-5	4-0

1.1. What is the cache block size (in words)? (2 points)

1.2. How many entries does the cache have? (2 points)

1.3. What is the ratio between total bits required for such a cache implementation over the data storage bits? (3 points)

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

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

1.4. How many blocks are replaced (excluding the insertion of a new entry into the empty block from the count)? (2 points)

1.5. What is the hit ratio? (2 points)

1.6. List the final state of the cache, with each valid entry represented as a record of <index, tag, data>. (4 points)

2. 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	2KB	8.0%	0.66ns
P2	4KB	6.0%	0.90ns

2.1. Assuming that the L1 hit time determines the cycle times for P1 and P2, what are their respective clock rates? (2 points)

2.2. What is the Average Memory Access Time (AMAT) for P1 and P2? (2 points)

2.3. Assuming a base CPI of 1.0 without any memory stalls, what is the total CPI for P1 and P2? Which processor is faster? (3 points)

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 (i.e., local miss rate = L2 misses / L2 accesses).

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

2.4. What is the AMAT for P1 with the addition of an L2 cache? Is the AMAT better or worse with the L2 cache? (2 points)

2.5. 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? (2 points)

2.6. 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? (4 points)