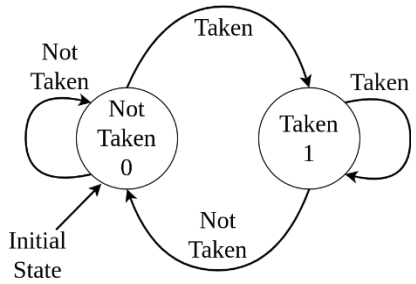
CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution

1.

Consider the following list of instructions. Assume that the initial values for `$s1`, `$s2`, `$s3`, `$s4`, and `$s5` are all 0:

```
loop:     addi $s2, $s1, -2
          bne $s5, $s2, target1
          addi $s3, $s3, 0
target1:  addi $s1, $s1, 1
          addi $s4, $s1, -3
          bne $s5, $s4, loop
```

Assume that we have a 1-bit branch predictor that stores the result of the last branch and makes the prediction based on the result. Show the results of all predictions throughout the execution. (Use T/N to represent Taken/ Not Taken)



Branch	Prediction (T/N)	Actual Result (T/N)
`1st bne->target1`	N	T
`2nd bne->loop`	T	T
`3rd bne->target`	T	T
`4th bne->loop`	T	T
`5th bne->target1`	T	N
`6th bne->loop`	N	N

The first Branch in the table is filled for you. The Prediction was Not Taken (N) and the Actual Result was Taken (T). The Prediction of current branch depends on the Prediction and Actual Result of previous branch. Here, N and T (prediction and actual results of branch 1) would produce a predicted result of T for branch 2.

Calculation of Branch 2's Actual Result:

(Note that the initial values for $s1, $s2, $s3, $s4 and $s5 are all 0)

After branch 1 execution jumps to "target1". As the instructions execute, $s1 is set to 1 and $s4 is set to -2. The next branch instruction is reach, since $s5 is not equal to $s4, the branch is taken which causes execution to jump to "loop", thus the Actual Result is Taken (T) for branch 2. T and T from branch 2 gives T as the predicted result for branch 3. The rest of the table follows the same pattern.

CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution

2. Assume the presence of the following memories in a MIPS system:

L1 cache
L2 cache
Main Memory
Hard drive

Show the memory hierarchy and order them in terms of

a. Speed
   In order of minimum to maximum
   Hard drive → Main memory → L2 cache → L1 cache
b. Memory capacity (Size)
   In order of minimum to maximum
   L1 cache → L2 cache → Main memory → Hard drive
c. Cost per byte
   In order of minimum to maximum
   Hard drive → Main memory → L2 cache → L1 cache

3.

a. If a direct mapped cache has a hit rate of 95%, a hit time of 4 ns, and a miss penalty of 100 ns, what is the AMAT?

   AMAT = Hit time + Miss rate x Miss penalty = 4 + 0.05 x 100 = 9 ns

b. If replacing the cache with a 2-way set associative increases the hit rate to 97%, but increases the hit time to 5 ns, what is the new AMAT?

   AMAT = Hit time + Miss rate x Miss penalty = 5 + 0.03 x 100 = 8 ns

c. If an L2 cache is added with a hit time of 20 ns and a hit rate of 50%, what is the new AMAT?
   AMAT = Hit TimeL1 + Miss RateL1 x (Hit TimeL2 + Miss RateL2 x Miss PenaltyL2) = 5 + 0.03 x (20 + 0.5x100) = 7.1 ns

4. Suppose that in 1000 memory references there are 40 misses in the first-level cache and 20 misses in the second-level cache.

a. What are the local and global miss rates for each level of cache?
   Assume the miss penalty from the L2 cache to memory is 200 clock cycles, the hit time of the L2 cache is 10 clock cycles, the hit time of the L1 cache is 1 clock cycle, and there are 1.5 memory references per instruction.

   Local miss rate for L1 = 40/1000 = 0.04 = 4%
   Global miss rate for L1 = 40/1000 = 0.04 = 4%
   Local miss rate for L2 = 20/40 = 0.5 = 50%
   Global miss rate for L2 = 20/1000 = 0.02 = 2%

b. What is the average memory access time?

   AMAT = Hit TimeL1 + Miss RateL1 x (Hit TimeL2 + Miss RateL2 x Miss PenaltyL2) = 1 + 4% x (10 + 50% x 200) = 5.4 clock cycles

CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution

5.

a. What is write back cache? Discuss both the advantages and disadvantages of using the write-back policy.

The information is written only to the block in the cache. The modified cache block is written to main memory only when it is replaced.

Advantages:

Low latency and high throughput for write-intensive applications.

Disadvantages:

There is data availability risk because the cache could fail (and so suffer from data loss) before the data is persisted to the backing store. This result in the data being lost.

b. What is write through cache? Discuss both the advantages and disadvantages of using the write-through policy.

The information is written to both the block in the cache and to the block in the lower-level memory.

Advantage:

Ensures fast retrieval while making sure the data is in the backing store and is not lost on case the cache is disrupted.

Disadvantage:

Writing data will experience latency as you have to write to two places every time.

6. Here is a series of address references given as word addresses: 2, 3, 11, 16, 21, 13, 64, 48, 19, 11, 3, 22, 4, 27, 6 and 11. Consider a word-addressable cache.

Assuming a Direct-mapped cache with 16 one-word blocks that is initially empty, label each reference in the list as a hit or a miss.

a. For each of these references, identify the binary address, the tag, and the index.

Given it's a word addressable cache -> no byte offset bits required

1 block holds 1 word -> No word offset bits required

There are 16 blocks in the cache, each with one word.

-> # of index bits = $\log_2$ (# blocks) = 4

tag bits = 32 - 4 = 28

b. What is the hit rate?

CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution

Index	Cache Contents
0000	~~16~~ ~~64~~ 48
0001	
0010	2
0011	~~3~~ ~~19~~ 3
0100	4
0101	21
0110	~~22~~ 6
0111	
1000	
1001	
1010	
1011	~~11~~ ~~27~~ 11
1100	
1101	13
1110	
1111	

Address	Tag	Index	Hit/Miss
2	0000 0000 0000 0000 0000 0000 0000	0010	Miss
3	0000 0000 0000 0000 0000 0000 0000	0011	Miss
11	0000 0000 0000 0000 0000 0000 0000	1011	Miss
16	0000 0000 0000 0000 0000 0000 0001	0000	Miss
21	0000 0000 0000 0000 0000 0000 0001	0101	Miss
13	0000 0000 0000 0000 0000 0000 0000	1101	Miss
64	0000 0000 0000 0000 0000 0000 0100	0000	Miss
48	0000 0000 0000 0000 0000 0000 0011	0000	Miss
19	0000 0000 0000 0000 0000 0000 0001	0011	Miss
11	0000 0000 0000 0000 0000 0000 0000	1011	Hit
3	0000 0000 0000 0000 0000 0000 0000	0011	Miss
22	0000 0000 0000 0000 0000 0000 0001	0110	Miss
4	0000 0000 0000 0000 0000 0000 0000	0100	Miss
27	0000 0000 0000 0000 0000 0000 0001	1011	Miss
6	0000 0000 0000 0000 0000 0000 0000	0110	Miss
11	0000 0000 0000 0000 0000 0000 0000	1011	Miss

Hit Rate = 1/16

7. A processor with Instruction cache miss rate of 2% and Data cache miss rate of 4% and costs 10 cycles for either of the cache miss (miss penalty). 40% of the instructions to be executed are Load and Store Instruction. The CPI with ideal cache (no misses) is 2.

CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution

a. Compute the actual CPI

- Instruction-cache miss rate = 2%
- Data-cache miss rate = 4%
- Miss penalty = 10 cycles
- Base CPI (with ideal cache performance) = 2
- Load & stores are 40% of instructions

Miss cycles per instruction
- Instruction-cache: 0.02 × 10 = 0.2
- Data-cache: 0.40 × 0.04 × 10 = 0.16

Actual CPI = 2 + 0.2 + 0.16 = 2.36

b. Consider the datapath was improved so that the CPI can be reduced from 2 to 1.5 (all the other specs remain the same). Compute the actual CPI and compare it with (a)
Miss cycles per instruction is same as (a)
Actual CPI = 1.5 + 0.2 + 0.16 = 1.86

Due to the reduced CPI, it take only 1.86 cycles for completion of 1 instruction whereas in (a) it takes 2.36 cycles to complete 1 instruction. So (b) performs better

8. Consider a direct-mapped cache of 128 blocks and each block can hold 16 words. The total memory capacity is 8 GByte with 8Byte word sizes. It is a word addressable memory.

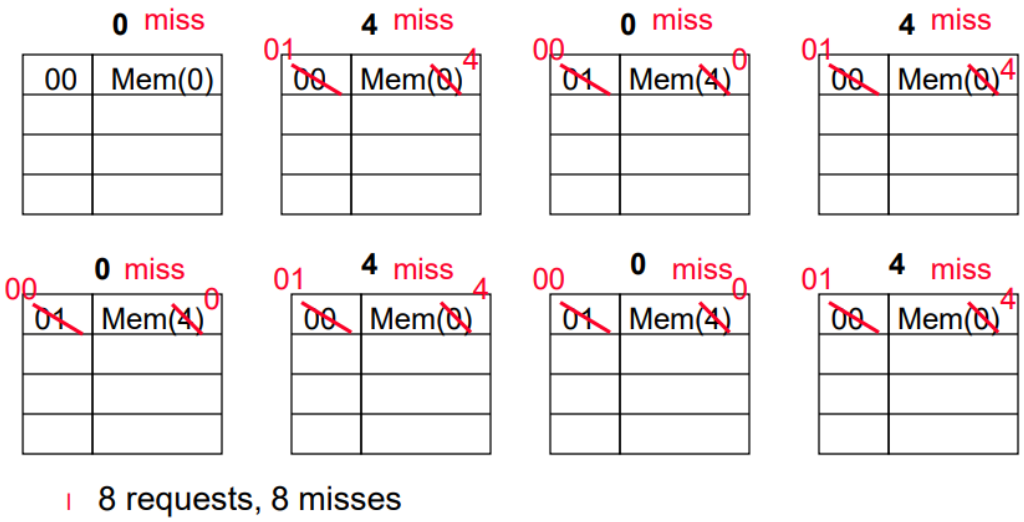
a. What is the address format?
With 8 GB memory and 8 bytes word sizes = 1 GW (i.e. 2^{30} bytes) of memory.
So we need a 30 bits for an address.
Given each block can hold 16 words (2^4) -> 4 bits for the word
Given the memory has 128 block (2^7) -> 7 bits for block number
This leaves 30 – 7 - 4 = 19 bits for the tag.
So the address format is 19 – 7 – 4.

b. If we change the cache to a 4-way set associative cache, what is the new address format?

4-way set associative cache => 128 / 4 = 32 blocks => 5 bits for block number, 4 bits for the word address
This leaves 30 – 5 - 4 = 21 bits for the tag.
So, the address format is 21 – 5 – 4.

9. a) Consider a main memory size of 4 Bytes. Each block in the cache can hold only 1 word (here 1 word = 1 Byte). Find the miss rate for the address references 0 4 0 4 0 4 0 4 in a direct mapped cache

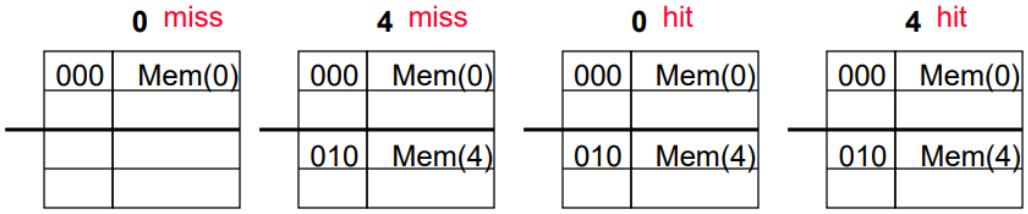
CSE 490/590 Computer Architecture Spring 2025 Homework 3 Solution



- 8 requests, 8 misses

Miss rate = 1

b) Consider a main memory size of 4 Bytes. Consider a 2 way set associative cache which are initially empty. Each block in the cache can hold only 1 word (here 1 word = 1 Byte). Find the miss rate for the same address references (0 4 0 4 0 4 0 4).



- 8 requests, 2 misses

Miss rate = 2/8

c) Compare (a) and (b) and list the advantages associative cache.

(a) Ping pong effect due to conflict misses - two memory locations that map into the same cache block

(b) Solves the ping pong effect in a direct mapped cache due to conflict misses since now two memory locations that map into the same cache set can co-exist