Video explanation of solution is provided below the problem.

Cache Benefits

5/5 points (ungraded)
After his geek hit single *I Hit the Line*, renegade singer Johnny Cache has decided he'd better actually learn how a cache works. He bought three Beta processors, identical except for their cache architectures:

- **Beta1** has a 64-line direct-mapped cache
- **Beta2** has a 2-way set associative cache, LRU, with a total of 64 lines
- **Beta3** has a 4-way set associative cache, LRU, with a total of 64 lines

Note that each cache has the same total capacity: 64 lines, each holding a single 32-bit **word** of data or instruction. All three machines use the same cache for data and instructions fetched from main memory.

Johnny has written a simple test program:

```
// Try a little cache benchmark
J = 0x1000              // where program lives
A = 0x2000              // data region 1
B = 0x3000              // data region 2
N = 16                  // size of data regions (BYTES!)

.=J                     // start program here
P:    CMOVE(1000, R6)   // outer loop count
Q:    CMOVE(N, R0)      // Loop index I (array offset)
R:    SUBC(R0, 4, R0)   // I = I-1
      LD(R0, A, R1)     // read A[I]
      LD(R0, B, R2)     // read B[I]
      BNE(R0, R)
      SUBC(R6,1, R6)    // repeat many times
      BNE(R6, Q)
```

Johnny runs his program on each Beta, and finds that one Beta model outperforms the other two.

1. Which Beta gets the highest hit ratio on the above benchmark?

   Beta3 ✔ Answer: Beta3

   Explanation

Beta3 will get the highest hit ratio because there are 3 distinct regions to cache. 1000+: instructions, 2000+: data region 1, 3000+: data region 2.

2. Johnny changes the value of **B** in his program to 0×2000 (same as A), and finds a substantial improvement in the hit rate attained by one of the Beta models (approaching 100%). Which model shows this marked improvement?

Beta2 ✔ Answer: Beta2

Explanation
If data region 1 and 2 now both refer to the same data beginning at 0×2000, then there are only 2 distinct regions to cache. This means that beta2, the 2-way set associative cache will go from having a bunch of misses to practically have a 100% hit rate because now all the data fits in the 2 ways of the cache without causing any contention.

3. Finally, Johnny sets **J**, **A**, and **B** each to **0×0**, and sets **N** to **64**. What is the TOTAL number of cache misses that will occur executing this version of the program on each of the Beta models?

Total cache misses running on Beta1:

16 ✔ Answer: 16

Total cache misses running on Beta2:

16 ✔ Answer: 16

Total cache misses running on Beta3:

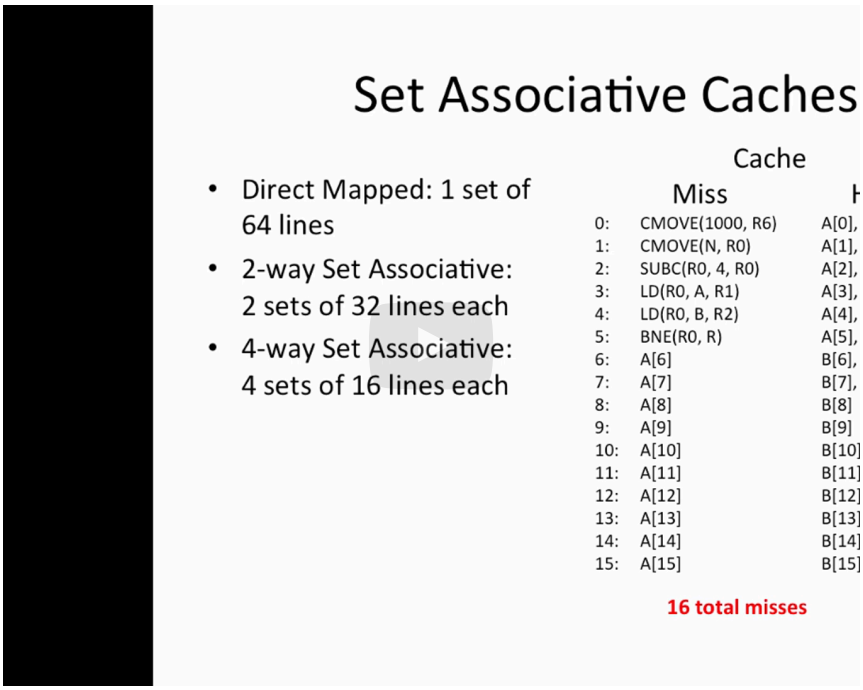
16 ✔ Answer: 16

Explanation
In this case, the instructions, data region 1, and data regions 2 are all in the same address space. This means that as long as the number of entries in the cache is large enough, there will be no cache misses after the initial load of the data into the cache. Since N = 64, that means that there are 16 total words of data to be read into the cache. Since the number of instructions is fewer than 16, that means that there will be a total of 16 compulsory misses in order to fill the cache, and after that there will be a 100% hit rate.

Submit

Answers are displayed within the problem

Cache Benefits



14:21 / 14:21 1.0x

Now, we are ready to execute our first load operation.

This operation wants to load A[15] into R1.

Because the beginning of array A is at address 0, then A[15] maps to line 15 of our cache.

Since we have not yet loaded anything into line 15 of our cache, this means that our first data access is a miss.

We continue with the second load instruction.

This instruction is not yet in the cache, so we get a cache miss and then load it into line 4 of our cache.

We then try to access B[15].

B[15] corresponds to the same piece of data as A[15], so this data access is already in the cache thus resulting in a data

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Discussion

Hide Discussion

Topic: 14. Caches and the Memory Hierarchy / WE14.1

Add a Post

Show all posts by recent activity

Question B - Why BETA 3 CACHE isn't equally good as BETA 2 CACHE? 2

Why BETA 3 (4-way) cache isn't equally good as BETA 2 (2-way) cache? The BETA 3 has 4 differe...

Question C, video explanation

I understand from the video explanation that the second run of the inner loop is executed with R0 ...

3