

CS61C Fall 2013 – 5 – Caches

Caches

Conceptual Questions: Why do we cache? What is the end result of our caching, in terms of capability?

What are temporal and spatial locality? Give high level examples in software of when these occur.

Break up an address:

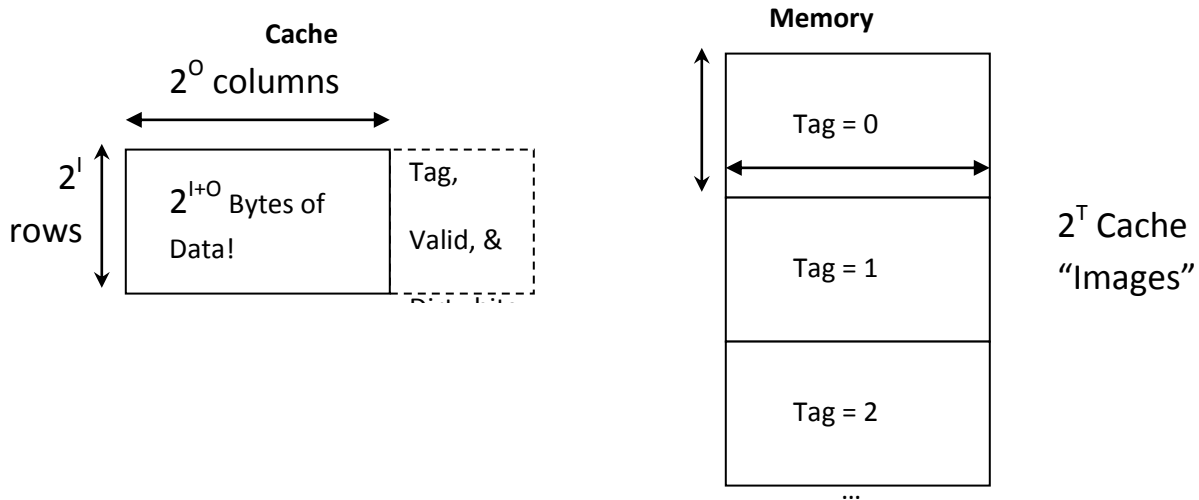


Offset: “column index”, Indexes into a block. (O bits)

Index: “row index,” Indexes blocks in the cache. (I bits)

Tag: Where from memory did the block come from? (T bits)

Segmenting the address into TIO implies a geometrical structure (and size) on our cache. Draw memory with that same geometry!



Cache Vocab:

Cache hit – Correct item is found and we write to the cache directly.

Cache miss – Nothing in checked cache block, so read from memory and write to cache.

Cache miss, block replacement – The right block was found, but it had the wrong tag. Do above.

CS61C Fall 2013 – 5 – Caches

Assuming a write-through policy, fill out the table:

Address Bits	Cache Size	Block Size	Tag Bits	Index Bits	Offset Bits	Bits per Row
16	16KB	1B				
16	16KB	16KB				
16	16KB				3	
32	32KB			12		
32	64KB		16			145
32		32B				270
			42	14	8	

Assume 16 B of memory and an 8B direct-mapped cache with 2-byte blocks. Classify each of the following byte-addr. memory accesses as hit (H), miss (M), or miss with replacement (R).

- | | |
|------|-------|
| a. 0 | e. 10 |
| b. 4 | f. 12 |
| c. 1 | g. 0 |
| d. 1 | h. 4 |

You want your AMAT to be ≤ 2 cycles. You have two levels of cache.

- | | |
|-------------------------|-------------------------------|
| L1 hit time is 1 cycle. | L1 miss rate is 20% |
| L2 hit time is 4 cycles | L2 miss penalty is 150 cycles |

What does your L2 miss rate need to be?

You know you have 1 MiB of memory (maxed out for processor address size) and a 16 KiB cache (data size only, not counting extra bits) with 1 KiB blocks.

```
#define NUM_INTS 8192
int A[NUM_INTS]; // lives at 0x100000
int i, total = 0;
for (i = 0; i < NUM_INTS; i += 128) A[i] = i; // Line 1
for (i = 0; i < NUM_INTS; i += 128) total += A[i]; // Line 2
```

- a) What is the T:I:O breakup for the cache (assuming byte addressing)?
- b) Calculate the hit percentage for the cache for the line marked "Line 1".
- c) Calculate the hit percentage for the cache for the line marked "Line 2".
- d) How could you optimize the computation?