The Memory Hierarchy

CSC 211 – December 9, 2020
Datapath Lab Q&A

Pay attention to file names when you submit.

You can use MARS to test assembly programs
- don't set $sp to 0xFF80
- printing to terminal uses the syscall instruction
  in MARS, not $b to 0xFFFF
When Caches Fail
Cache Misses

What are the three types of cache misses? They all start with C.

Capacity Miss - the cache runs out of space, so it had to evict something we are now accessing.

Compulsory Miss - this is the first time accessing a location, so it is not yet in the cache.

Conflict Miss - the value used to be in the cache, but another cached value replaced it. The cache may not have been full at the time.
Array Accesses

`#define N ___`
`int arr[N] = {...};`

`for(int i=0; i<N; i++) {
  printf("%d\n", arr[i]);
}`

`for(int i=0; i<N; i++) {
  printf("%d\n", arr[i]);
}`

Imagine this code executing on a processor with a fully-associative cache that uses 4-byte cache lines, and has E cache entries. **How does the cache perform when...**

N is less than E

___ misses in the first loop

___ misses in the second loop

N is equal to E

___ misses in the first loop

___ misses in the second loop

N is larger than E

___ misses in the first loop

___ misses in the second loop
Array Accesses

#define N ___
int arr[N] = {...};

for(int i=0; i<N; i++) {
    printf("%d\n", arr[i]);
}

for(int i=0; i<N; i++) {
    printf("%d\n", arr[i]);
}

What if we use the same fully-associative cache, but now with 8-byte cache lines?

What values of N lead to the following outcomes? Express N in terms of E.

- N/2 cache misses in the first loop, none in the second.  
  - N ≤ E/2
  
  N/2 cache misses in the first loop, N/2 cache misses in the second loop.
  
  N > 2·E

25% miss rate

Using LRU replacement

50% miss rate

spatial locality helps, even though we over-fill the cache
Multi-Level Caching
Intel Pentium Pro 256KB Cache (source: CPU-World)
AMD Ryzen 4-Core Unit (source: PCWorld)
Virtual Memory