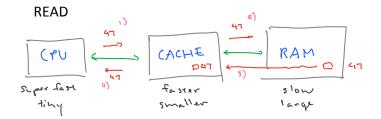
24 Cache

Friday, October 30, 2015 10:01 AM



Check if cache has a piece of data Address in RAM

TERMINOLOGY

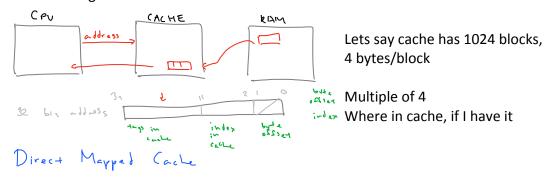
- Block/line: min amount of data between RAM and cache
- Hit Rate: fraction of memory accesses found in cache
- Miss Rate: 1 hit rate

Want a high hit rate and a low miss rate to keep the performance up

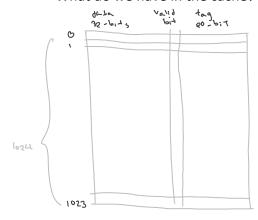
- Hit Time: Time to get the data from cache -> CPU. Includes time to check
- Miss Penalty: Time to get data RAM -> Cache -> CPU and update the cache's state

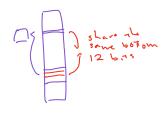
Designing the cache

- Similar to branch predictors
- What goes in and out of cache? A block



What do we have in the cache?





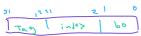
Data 32-bits: the actual data that is duplicated from RAM

<u>Valid bit</u>: whether I have the data or not. When machine is first powered on, all

valid bits are 0. But once grabbed from RAM, bit is flipped.

Tag 20-bits: the top 20-bits of the address.

Byte offset Index



Tag

Use index to look up the block in cache