Cache performance and associativity

Write through vs write back

Write through: every time data is changed in cache, change is done to lower level in hierarchy

Pro:

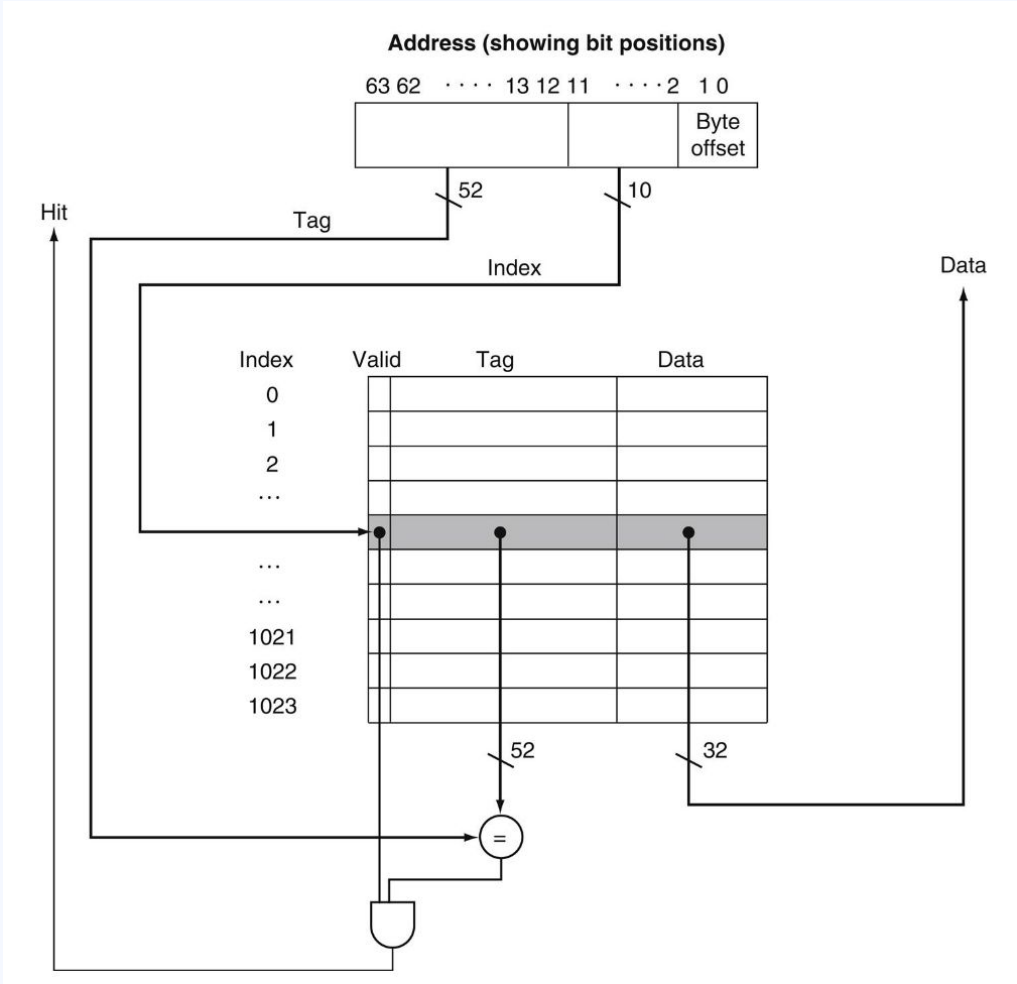
Con:

Write back: changes to lower level in hierarchy are only done when data is evicted from cache

Pro:

Con:

Cache controller



P&H fig. 5.10

What does the memory controller do when a cache miss occurs? What does the CPU do?

Write buffers

A way to hide the cost of writing to lower level of hierarchy

Example: instruction `sw t1, 4(a0)` in write-through cache

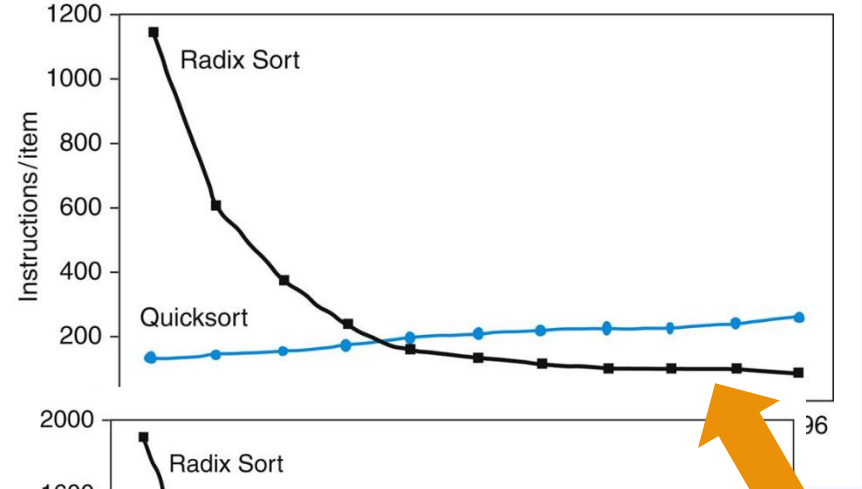
1. Write to cache and write to write buffer happen immediately (simultaneously, 1 cycle)
2. Rest of execution can happen at the same time that write to main memory is happening from buffer

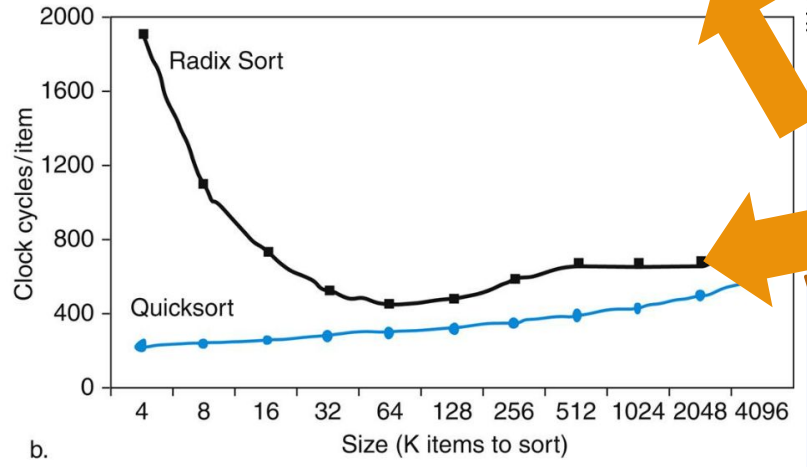
What happens if data that has been evicted from the cache is waiting in the write buffer and a read instruction for that address executes?

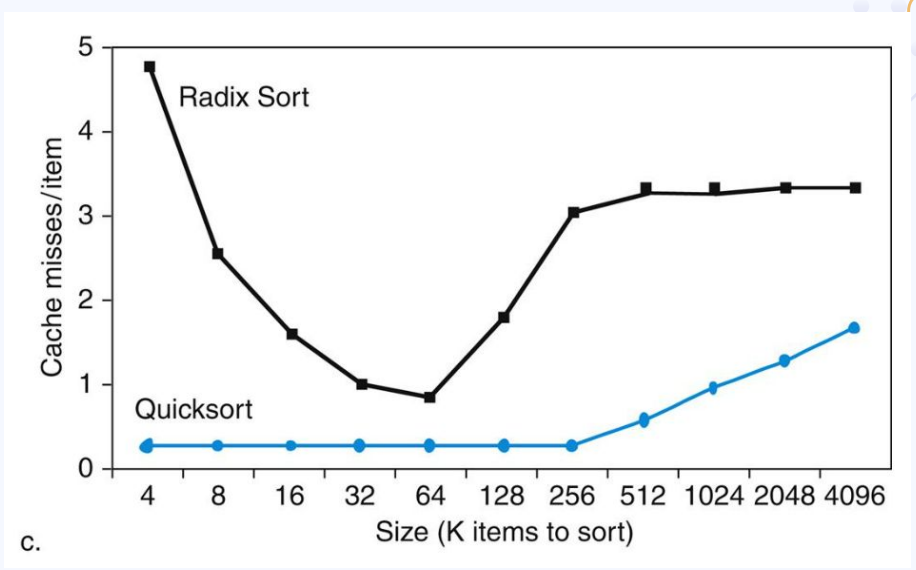
How many physical bits of space do we need to store our 1KB cache?

How do we measure the performance of a processor that uses caching?

Effect of algorithm on CPU time







why??

P&H fig. 5.19

Formulas from P&H 4.3

$$\text{CPU time} = (\text{CPU execution clock cycles} + \text{Memory-stall clock cycles}) \times \text{Clock cycle time}$$

$$\text{Memory-stall clock cycles} = (\text{Read-stall cycles} + \text{Write-stall cycles})$$

$$\text{Read-stall cycles} = \frac{\text{Reads}}{\text{Program}} \times \text{Read miss rate} \times \text{Read miss penalty}$$

$$\text{Write-stall cycles} = \left(\frac{\text{Writes}}{\text{Program}} \times \text{Write miss rate} \times \text{Write miss penalty}\right) + \text{Write buffer stalls}$$

$$\text{Memory-stall clock cycles} = \frac{\text{Memory accesses}}{\text{Program}} \times \text{Miss rate} \times \text{Miss penalty}$$

What causes cache misses? 3 Cs

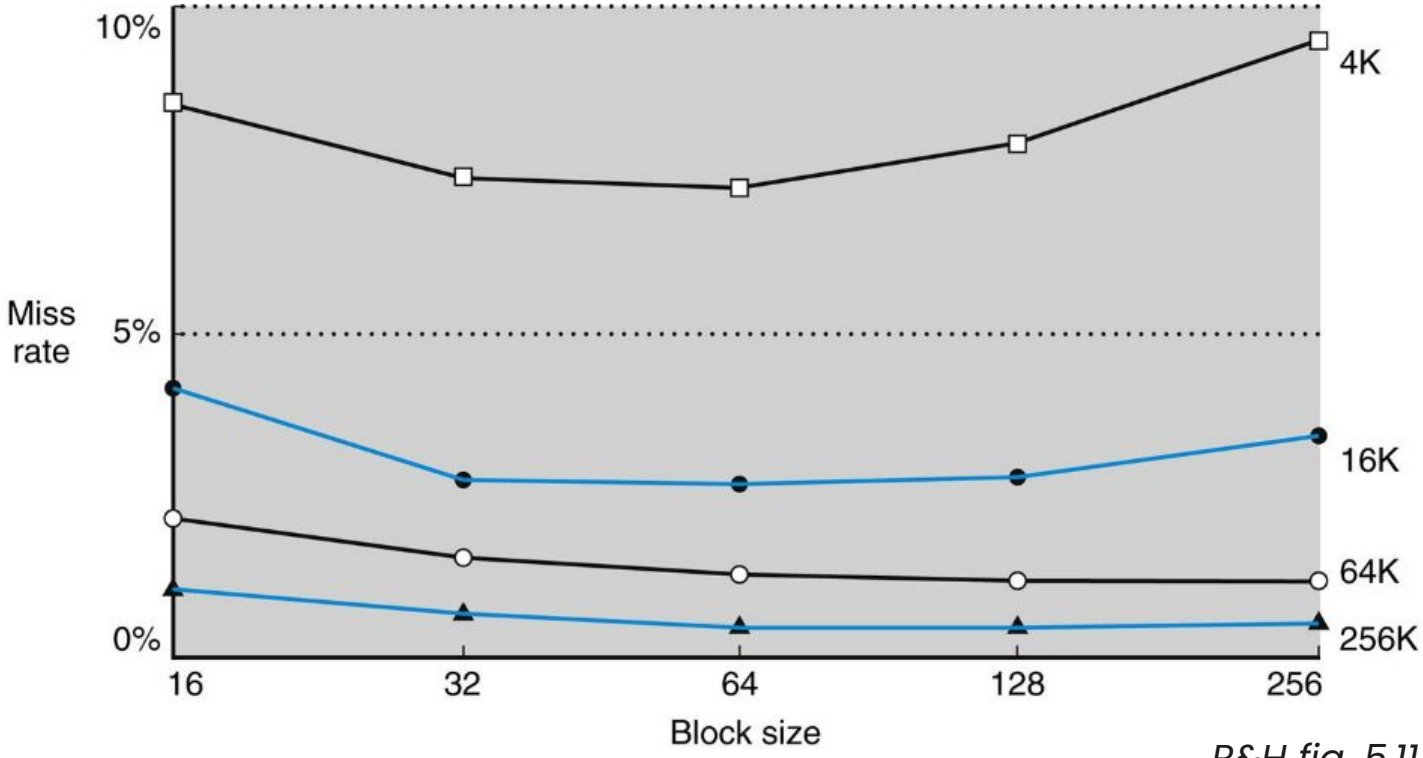
Compulsory – bringing the first blocks into a cache ("warming up" the cache)

Capacity – cache not big enough to contain all of the blocks it needs

Conflict – blocks constantly evicted due to cache collisions

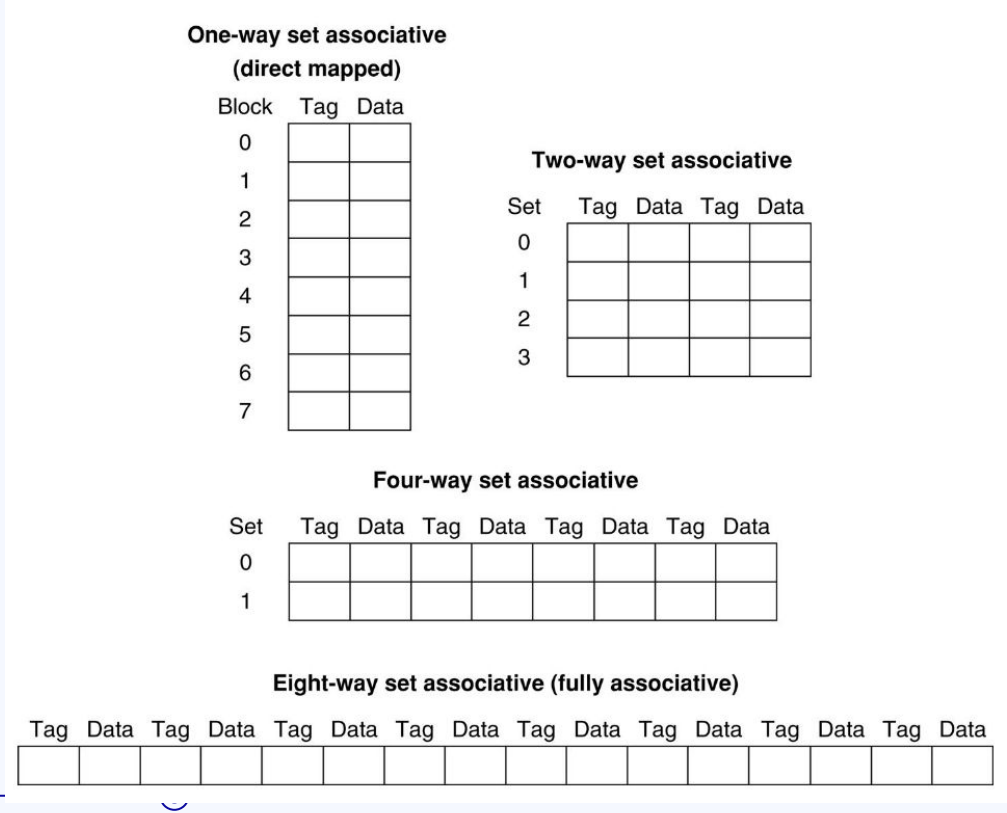
Can we decrease compulsory misses?

Increasing block size has limited effects



P&H fig. 5.11

Set-associative caches



P&H fig. 5.15