

Cache design / implementation details

→ Why you should care?

→ there are caches every where in H/W

→ even software has caches

→ understanding caches gives huge perf.

key design point: as fast as possible

simplest / fastest: direct-mapped cache

→ exactly 1 place to look

4-byte
entries

000
001
010
011
100
101
110
111

cache
blocks

16-bits / addr.

0x29ac

1111
1110
1101

0010 1001 1010 1100

tag

Need to find where in cache:

indexing

3 64KB memory 15

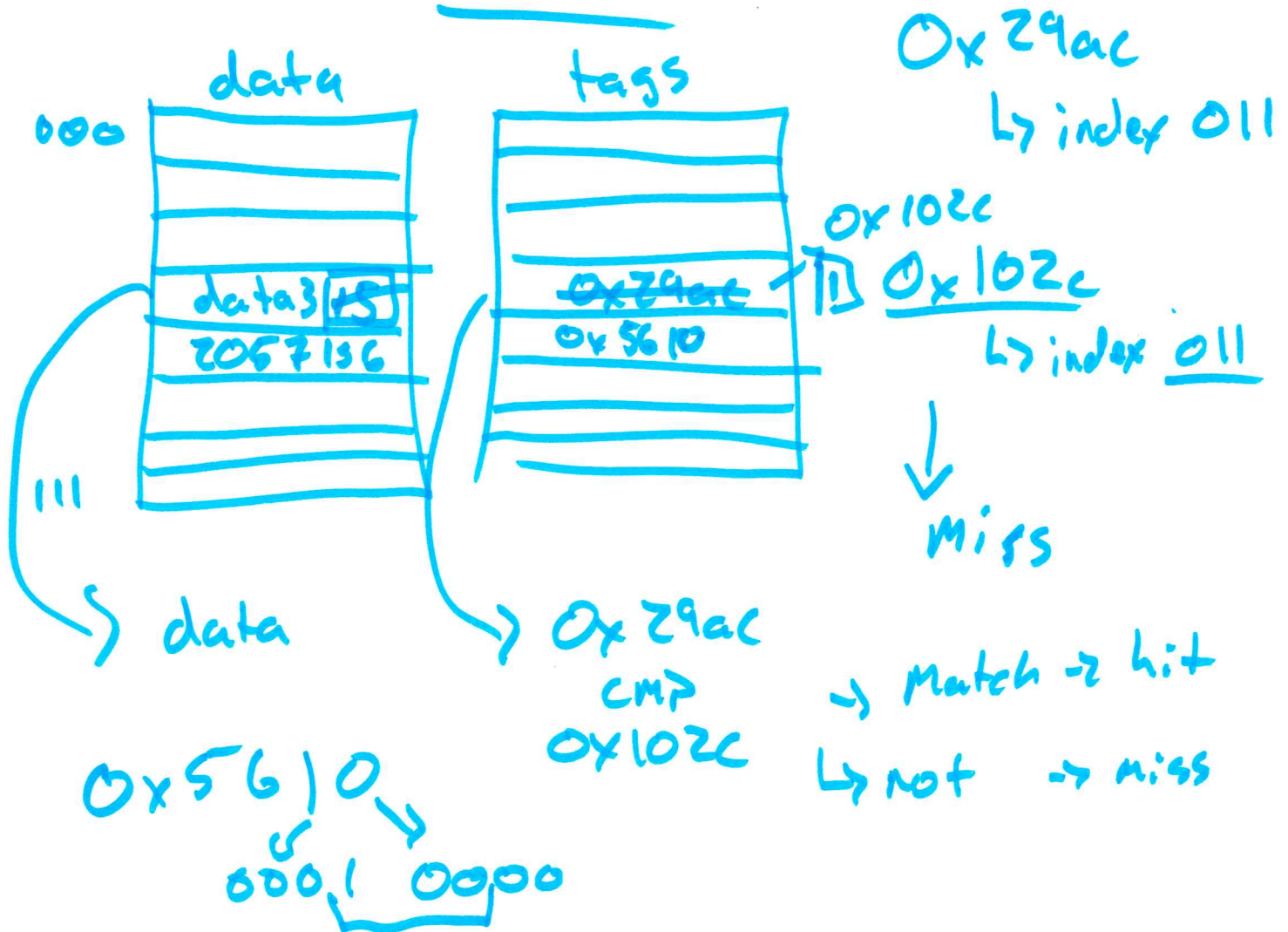
0x102c

0001 0000 0010 1100

↳ index

- How to tell which address the data in the cache actually corresponds to

→ add a TAG



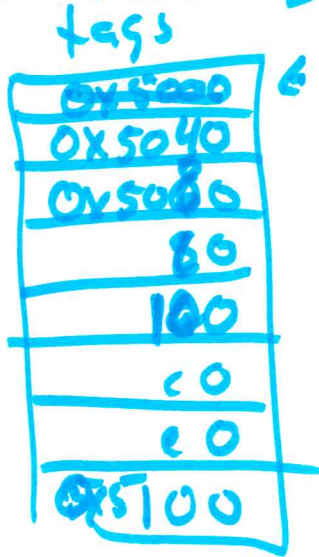
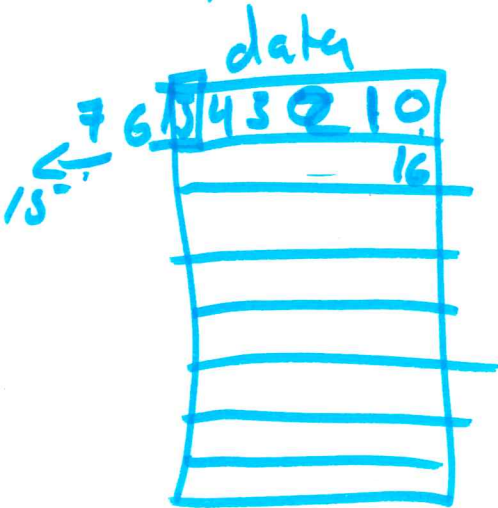
- Want to take advantage of spatial locality

→ increase size of entries

→ Block Size line size

↳ commonly this is

64-bytes $a[i] = i$



int arr[1000];
↳ 0x5000

0x5004

~~0x5008~~ 0x500c

0x5010

0x5014

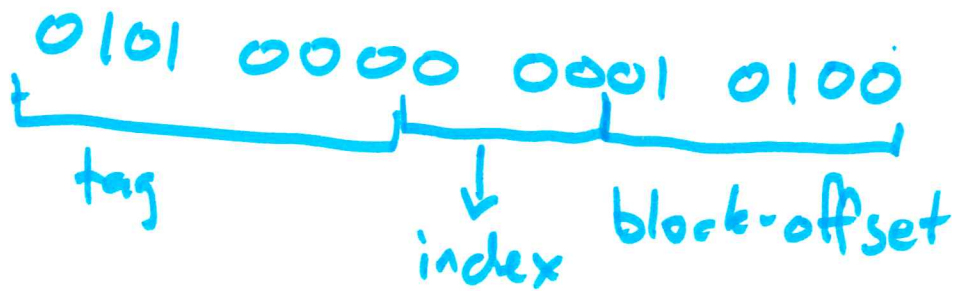
8

↳

64-bytes

2⁶

0x5014



0x5040

0101 0000 0100 0000

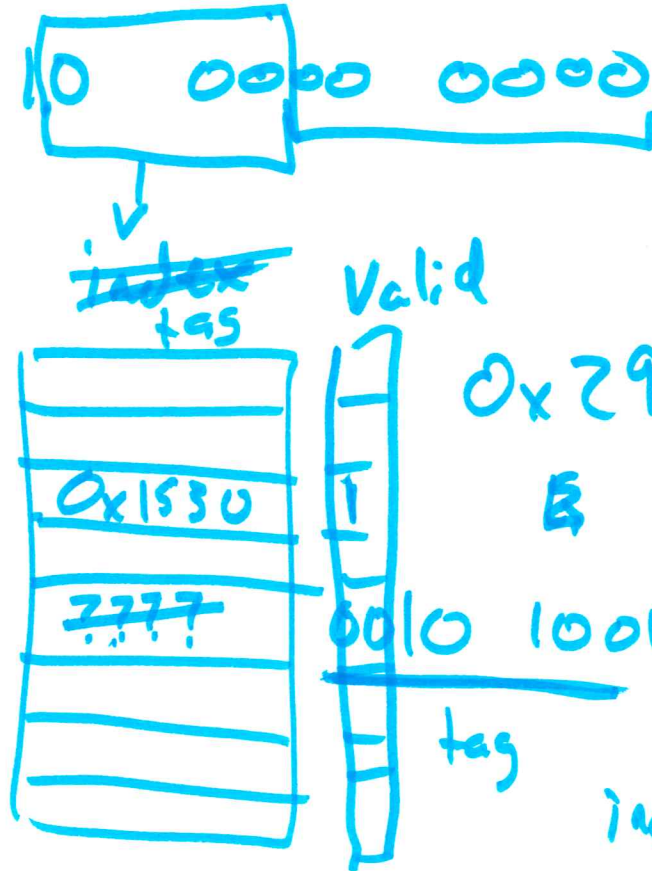
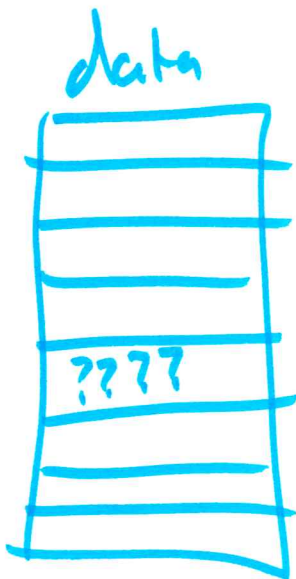
index

~~0x5200~~

~~0101 0001 0100 0000~~

0x5200

0101 0010 0000 0000



32
-byte blocks