Cache Block Size

Cache Specifications

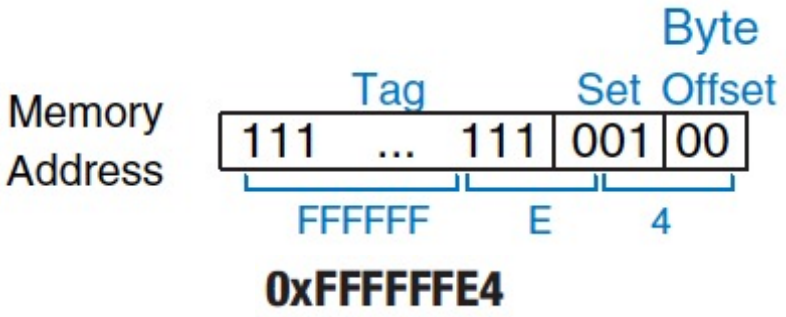
1. Capacity
   How much data can we store?
2. Number of Sets
   How many mappings between cache and main memory addresses?
3. Block Size
   Grab related words around a piece of data added to the cache.
4. Number of Blocks
   How many blocks can we store?
5. Degree of Associativity
   Different types of cache handle organization differently.

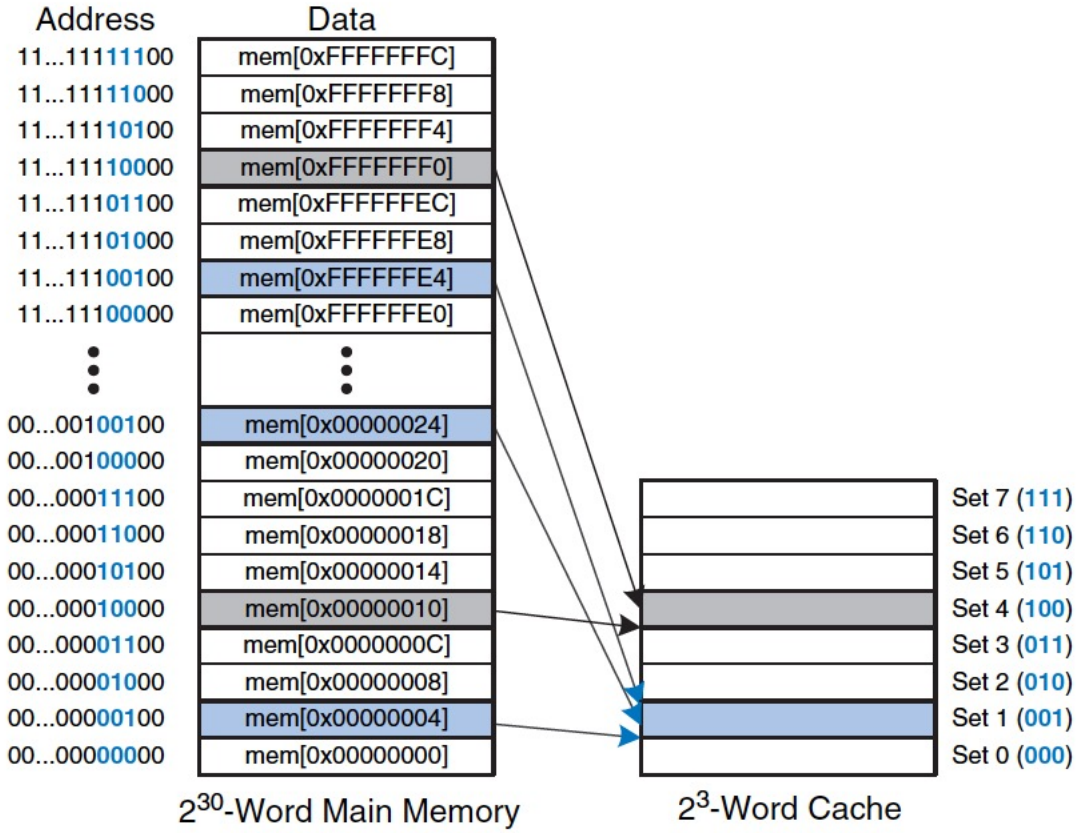
Direct Mapped Cache

One block in each set.

$S = B$ sets

One-to-one mapping, but there are more blocks in memory than there are in cache.



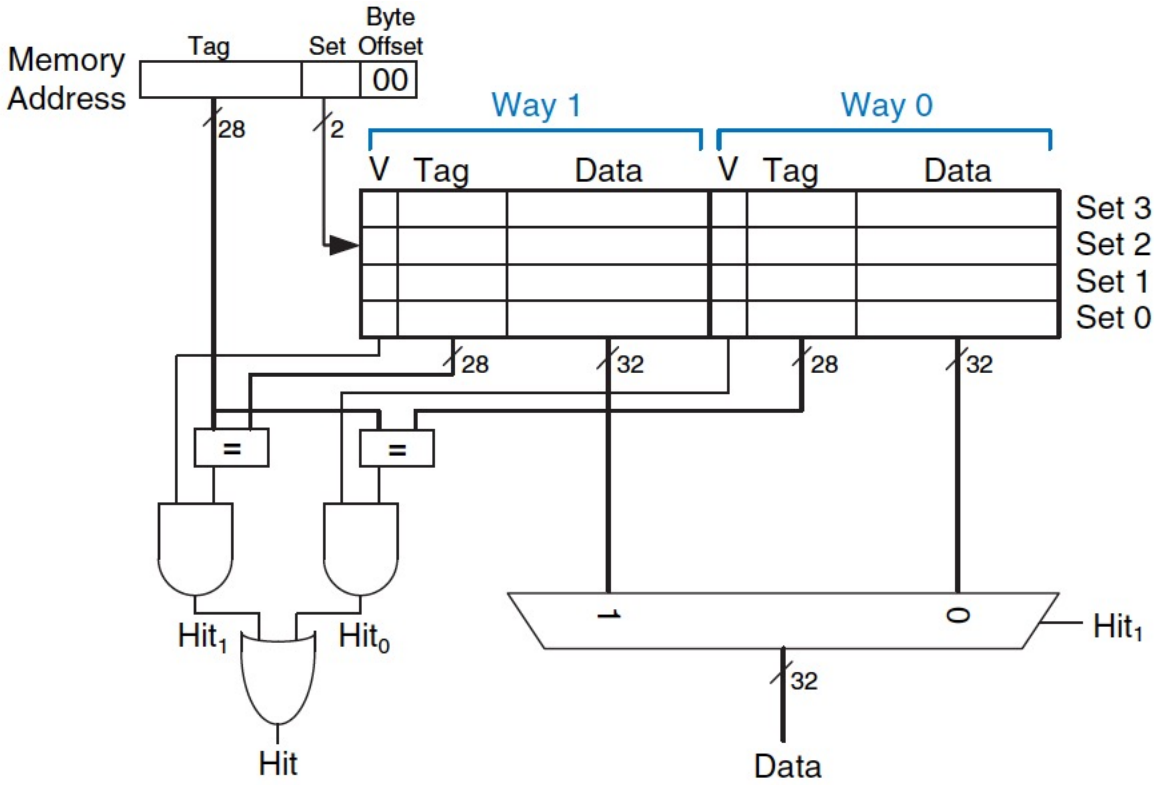


2-Way Set Associative Cache

Each set has two ways of associativity.

Cache reads blocks from both ways and checks tags and valid bits for a hit.

- If hit, multiplexer selects data from that way.



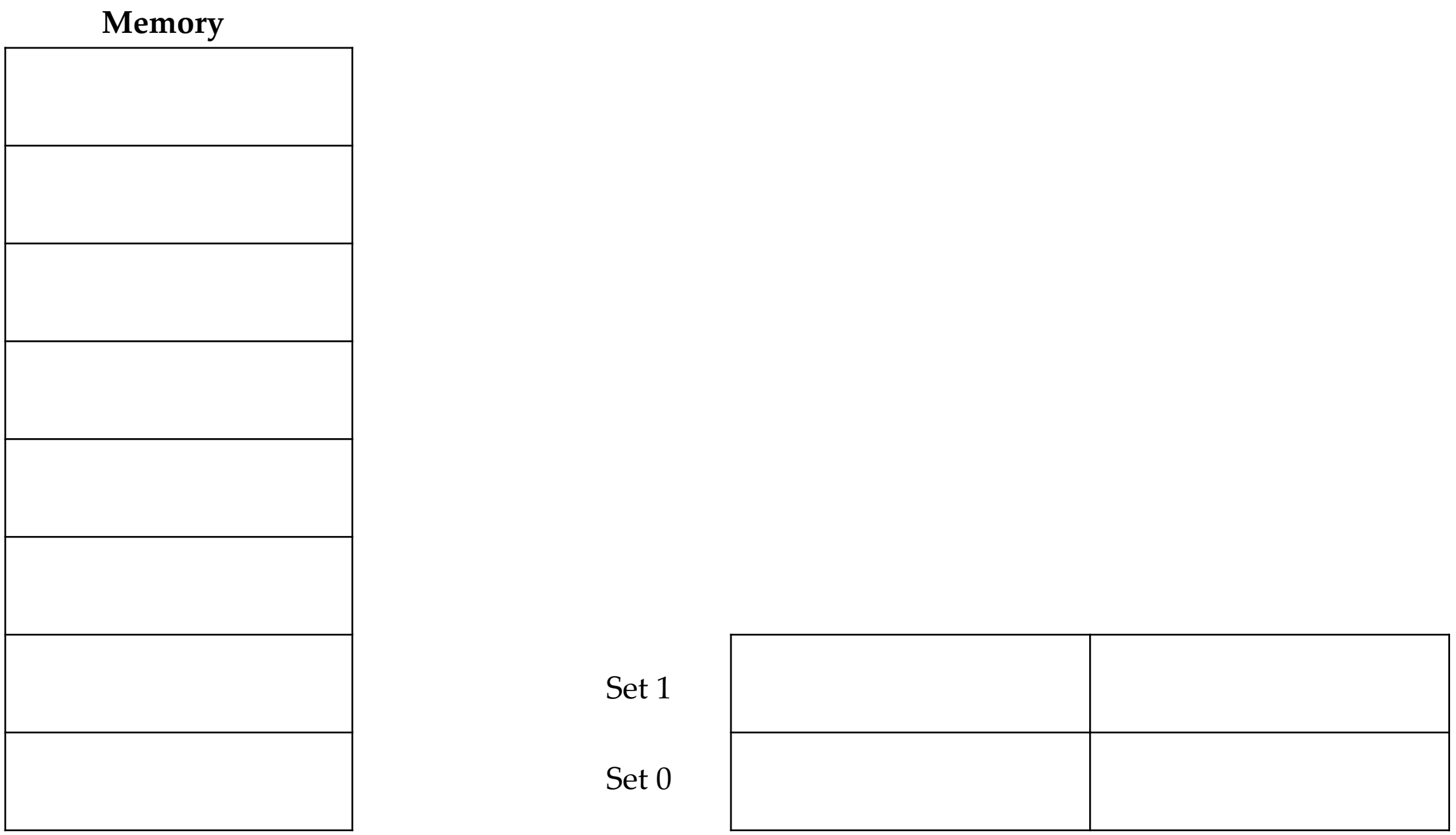
Block Size

Previous examples take advantage of temporal locality.

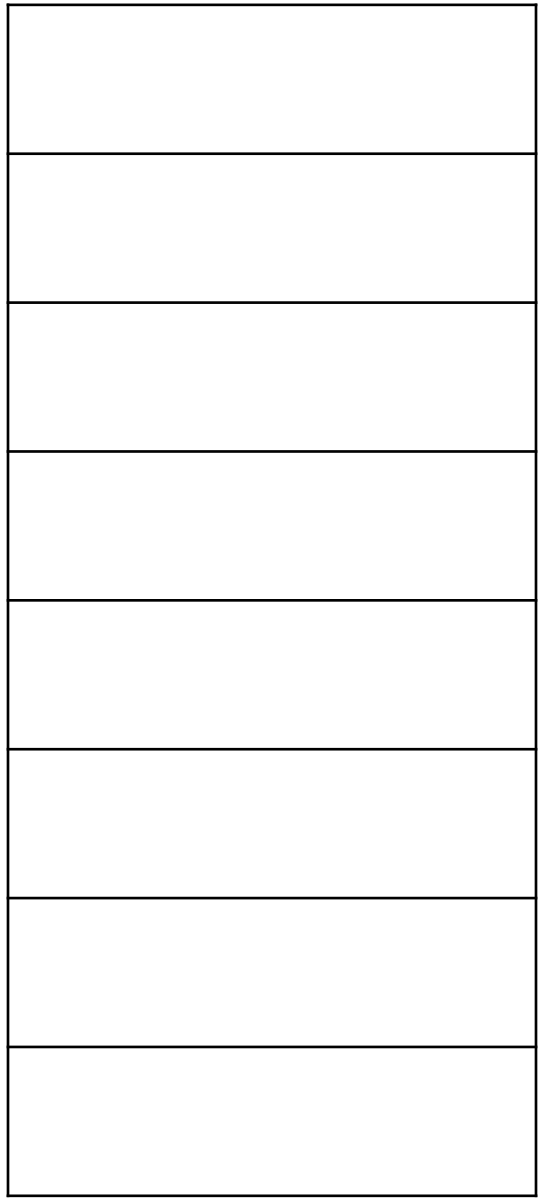
Why?

Must use larger blocks to hold several consecutive words for spatial locality.

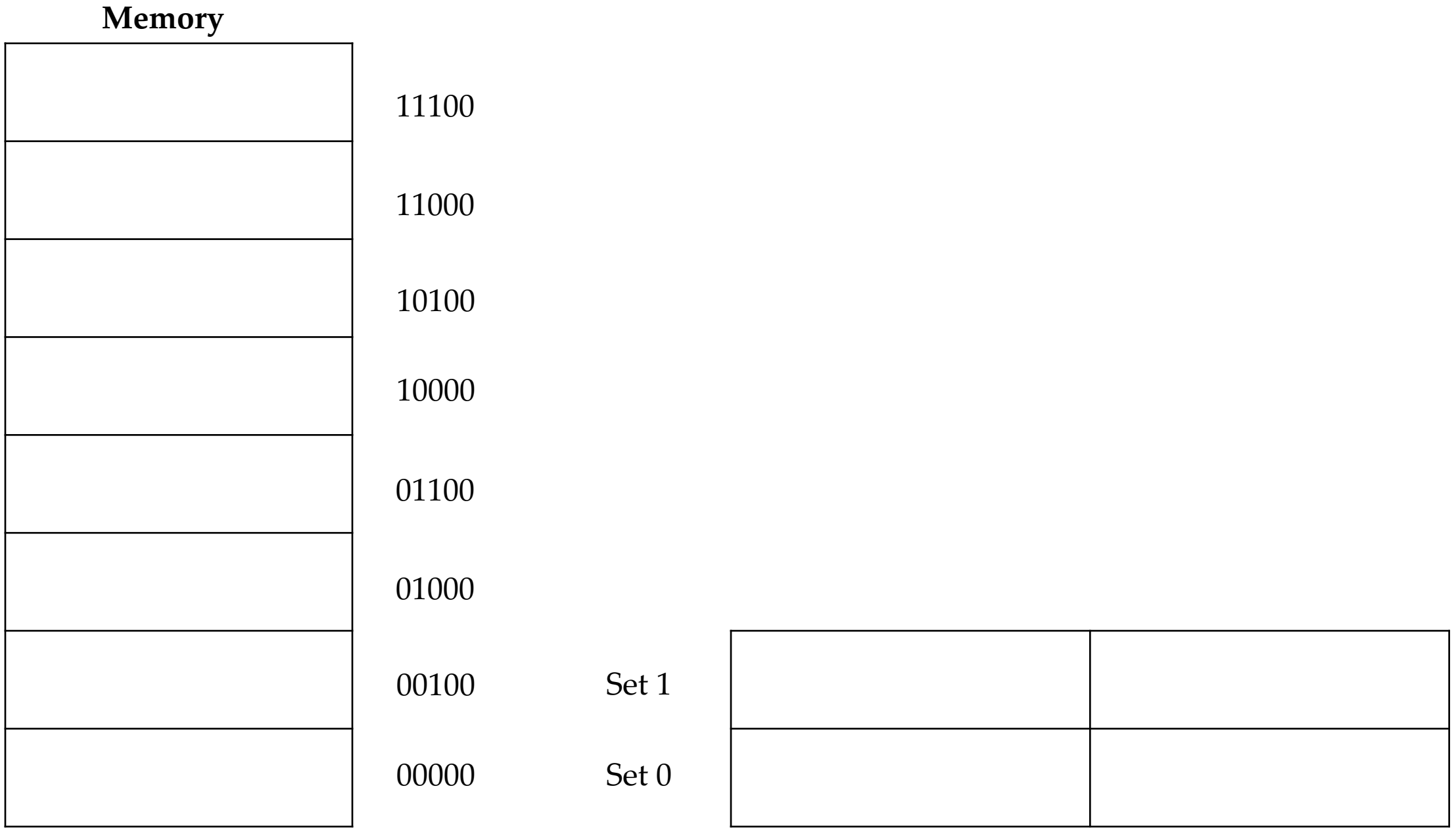
- Larger block sizes allow us to fetch multiple words into the cache
- Exploit the adjacency of data
- Subsequent accesses are more likely to hit because of spatial locality


Memory
Set 1
Set 0

Memory



Blocks are fixed! The addresses belonging to a block do not change based on on which address is requested


Memory
11100
11000
10100
10000
01100
01000
00100
00000
Set 1
Set 0