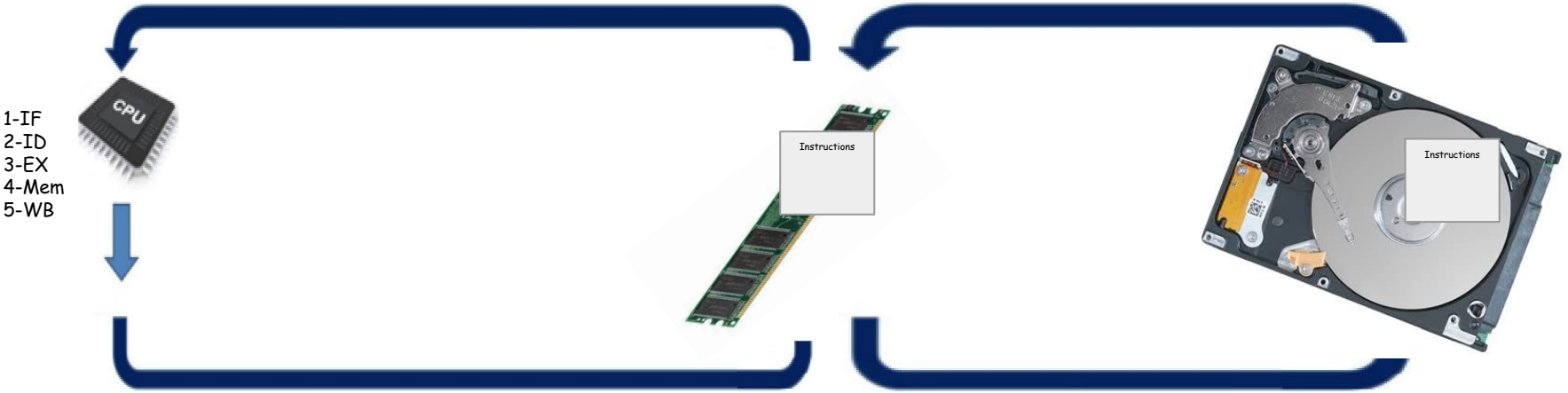
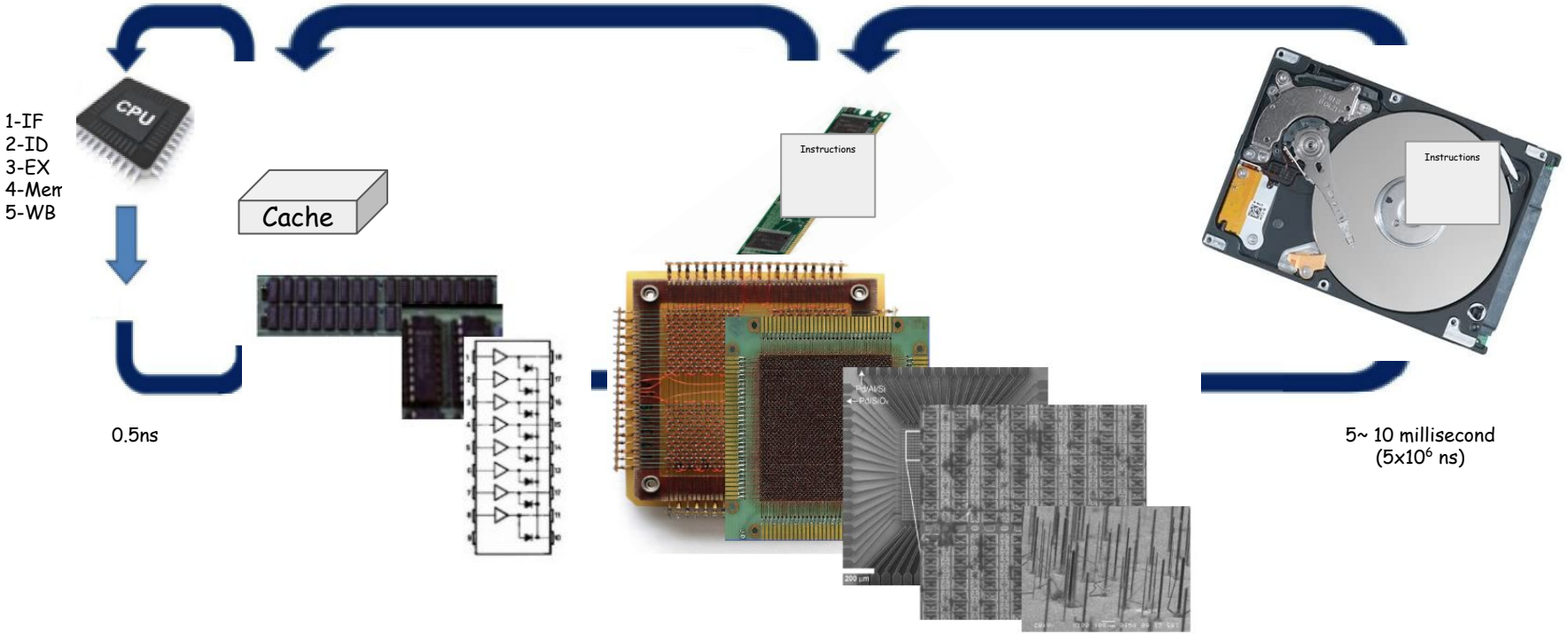
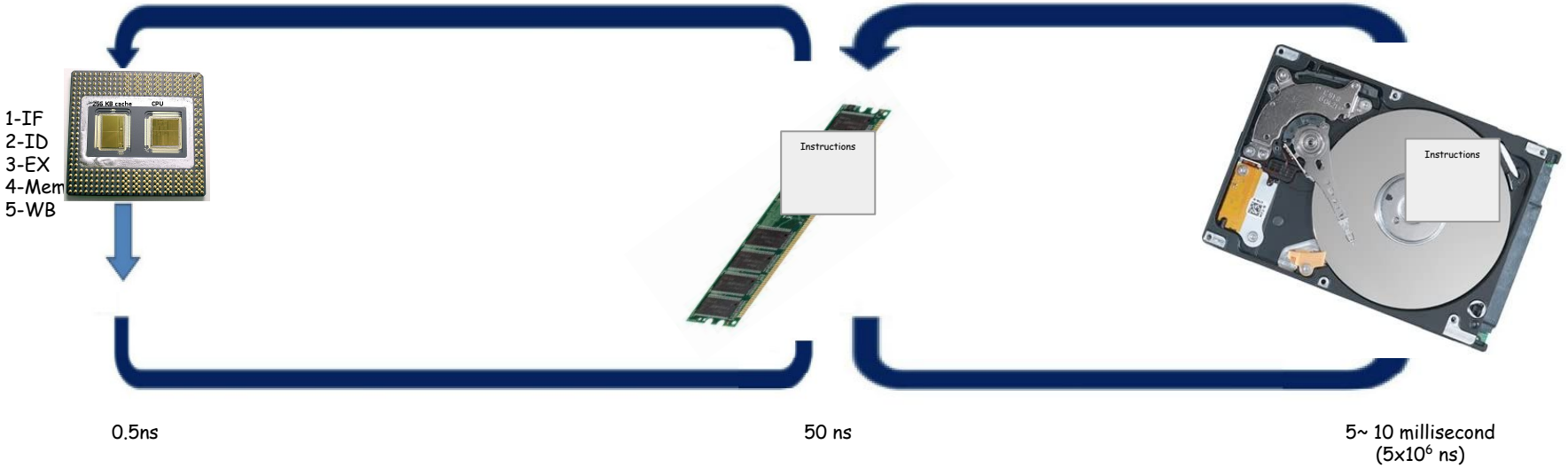
Cache Memory


1-IF
2-ID
3-EX
4-Mem
5-WB
CPU
Instructions

The Von Neumann Bottleneck


1-IF
2-ID
3-EX
4-Mem
5-WB
CPU
Instructions
Instructions
0.5ns
50 ns
5~ 10 millisecond
(5x10⁶ ns)


1-IF
2-ID
3-EX
4-Mem
5-WB
CPU
Cache
Instructions
Instructions
0.5ns
5~ 10 millisecond
(5x10⁶ ns)



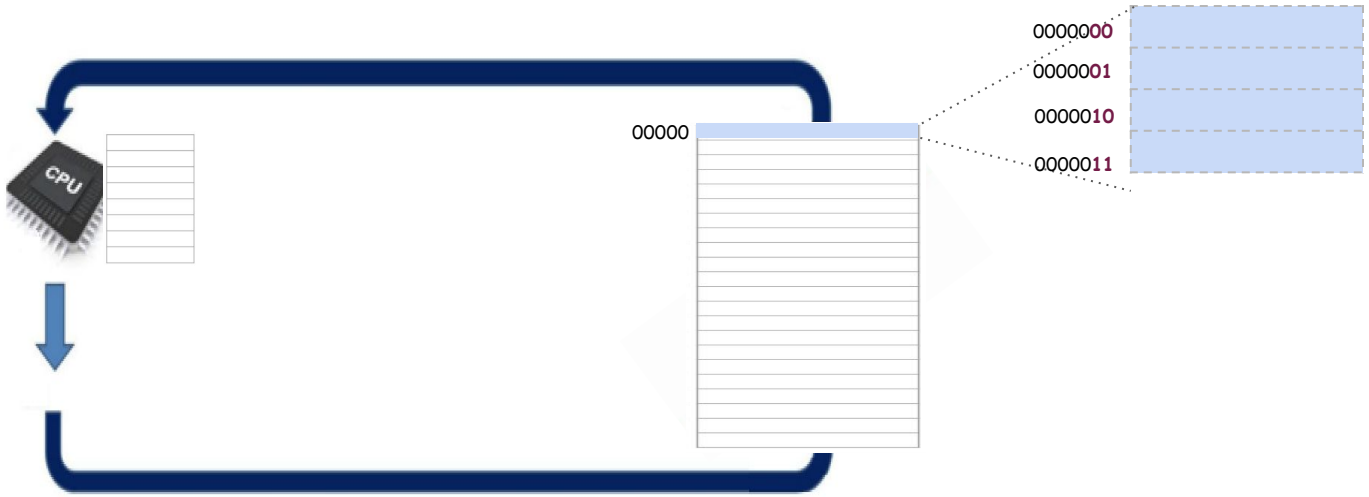
Spatial Locality: (Need for adjacent Data)

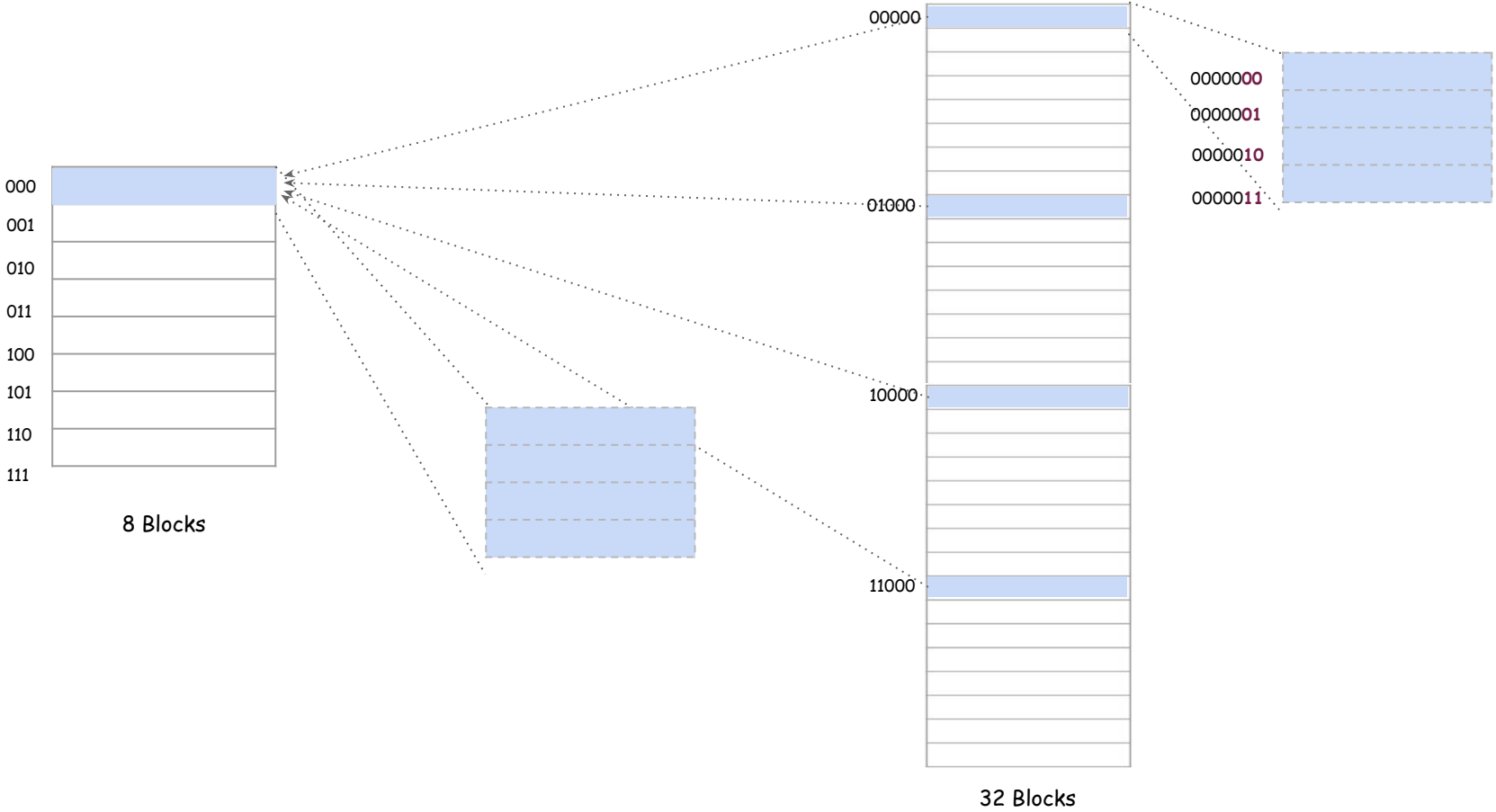
Temporal Locality: (Need to be close to data for some time)

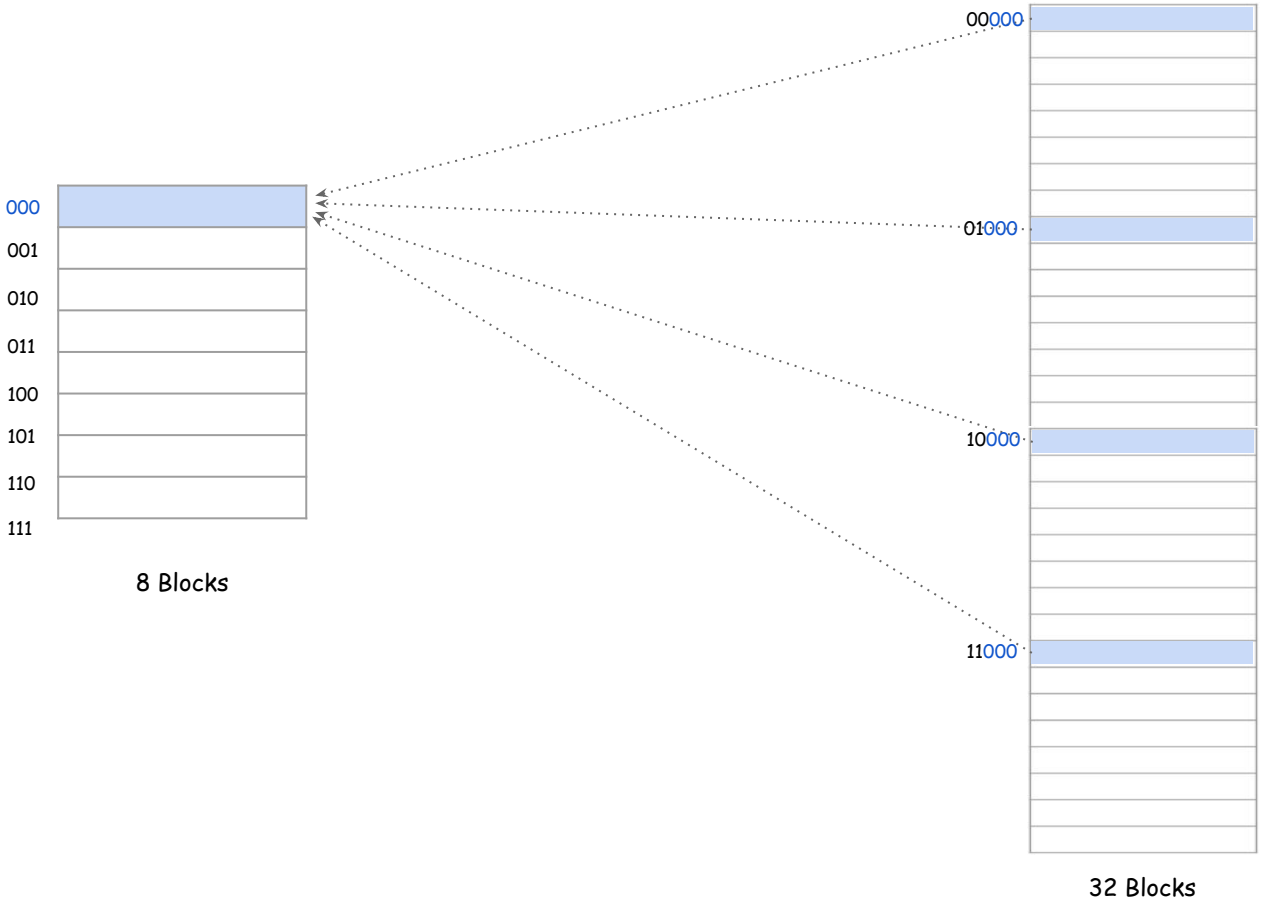
```
int a={3,4,5,6,7,1,2,3,8,3}

total=0

for(int i=0; i<10;i++){

        total+=a[i]
```

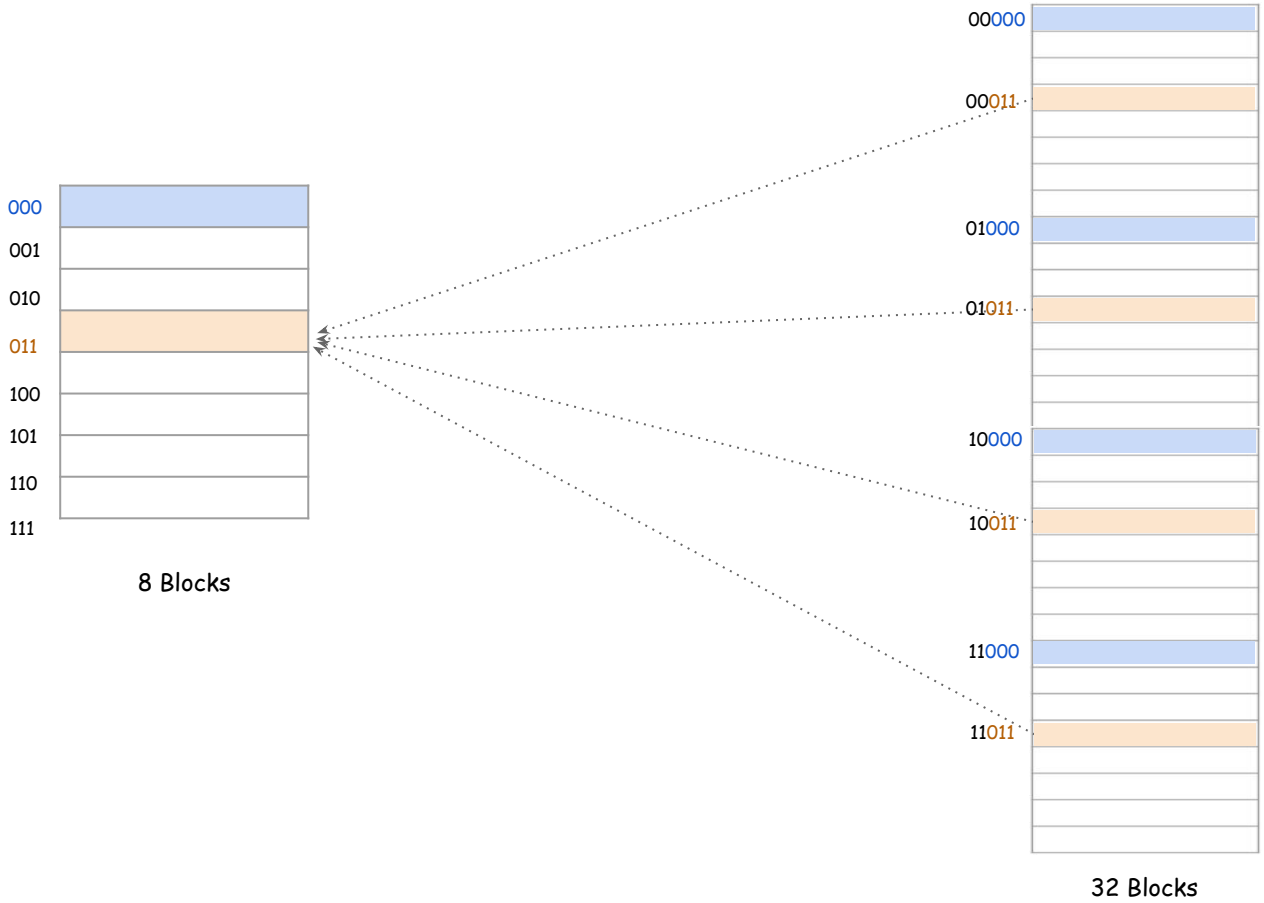
Cache is a small amount of memory that's on the CPU itself or right next to it. It can provide the cpu with same of its speed.

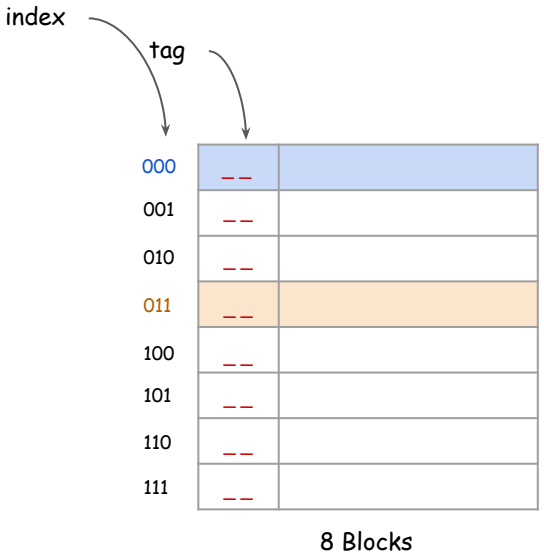
- It stores a copy of info. From the main memory.
- CPU asks cache if yes (cache hit) if not (cache miss)
- The greater the cache hits ⇒ the greater the performance
- The greater the cache misses ⇒ the lower the performance


CPU
00000
0000000
0000001
0000010
0000011

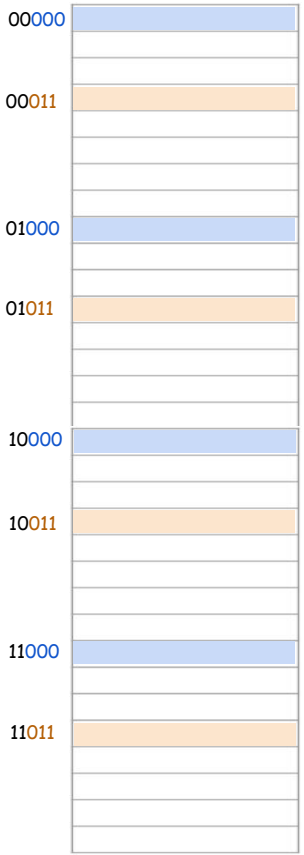

000
001
010
011
100
101
110
111
8 Blocks
00000
01000
10000
11000
32 Blocks
0000000
0000001
0000010
0000011


000
001
010
011
100
101
110
111
8 Blocks
00000
01000
10000
11000
32 Blocks

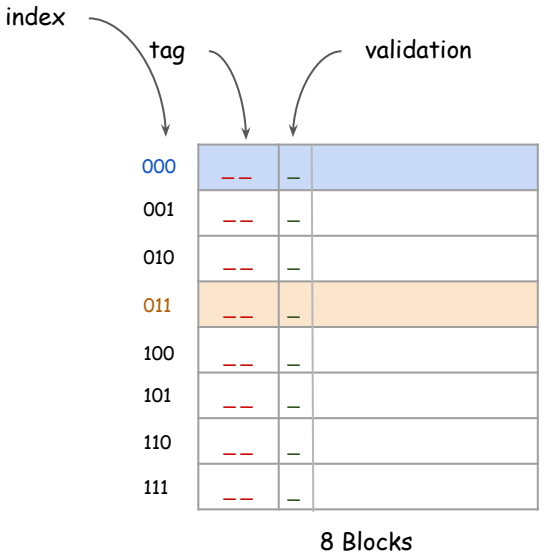

000
001
010
011
100
101
110
111
8 Blocks
00000
00011
01000
01011
10000
10011
11000
11011
32 Blocks



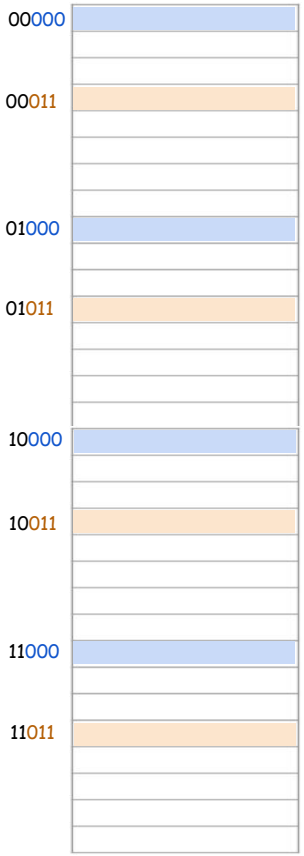
8 Blocks



32 Blocks



8 Blocks



32 Blocks