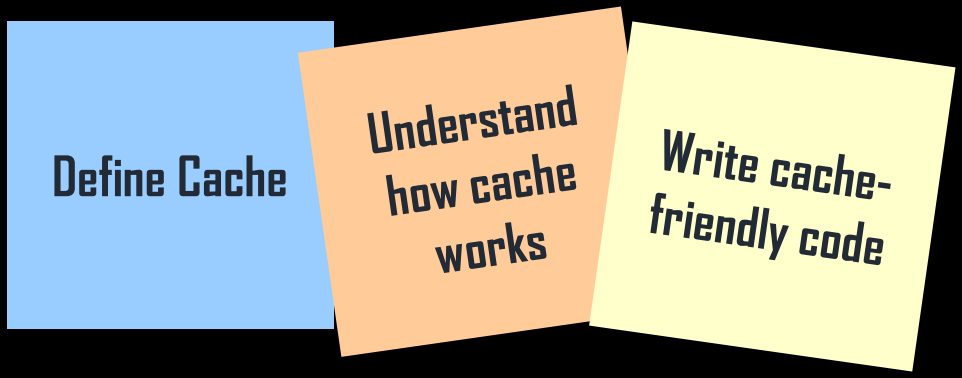


TRUONG DAI HOC KHOA HOC TU NHIEN
KHTN
TP. HO CHI MINH

Cache Memory

① Cache Concepts

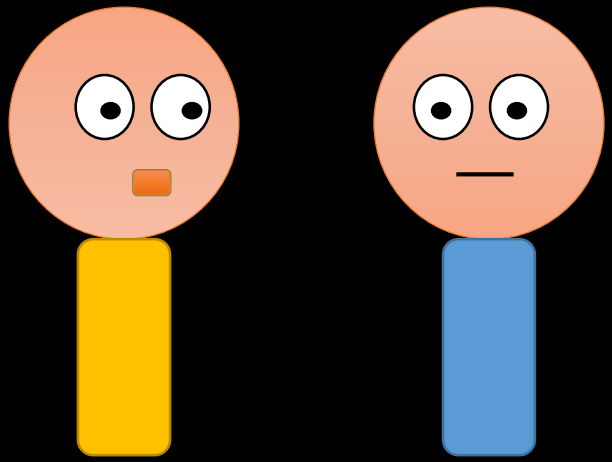
② Cache Organization

③ Writing Cache-friendly Code

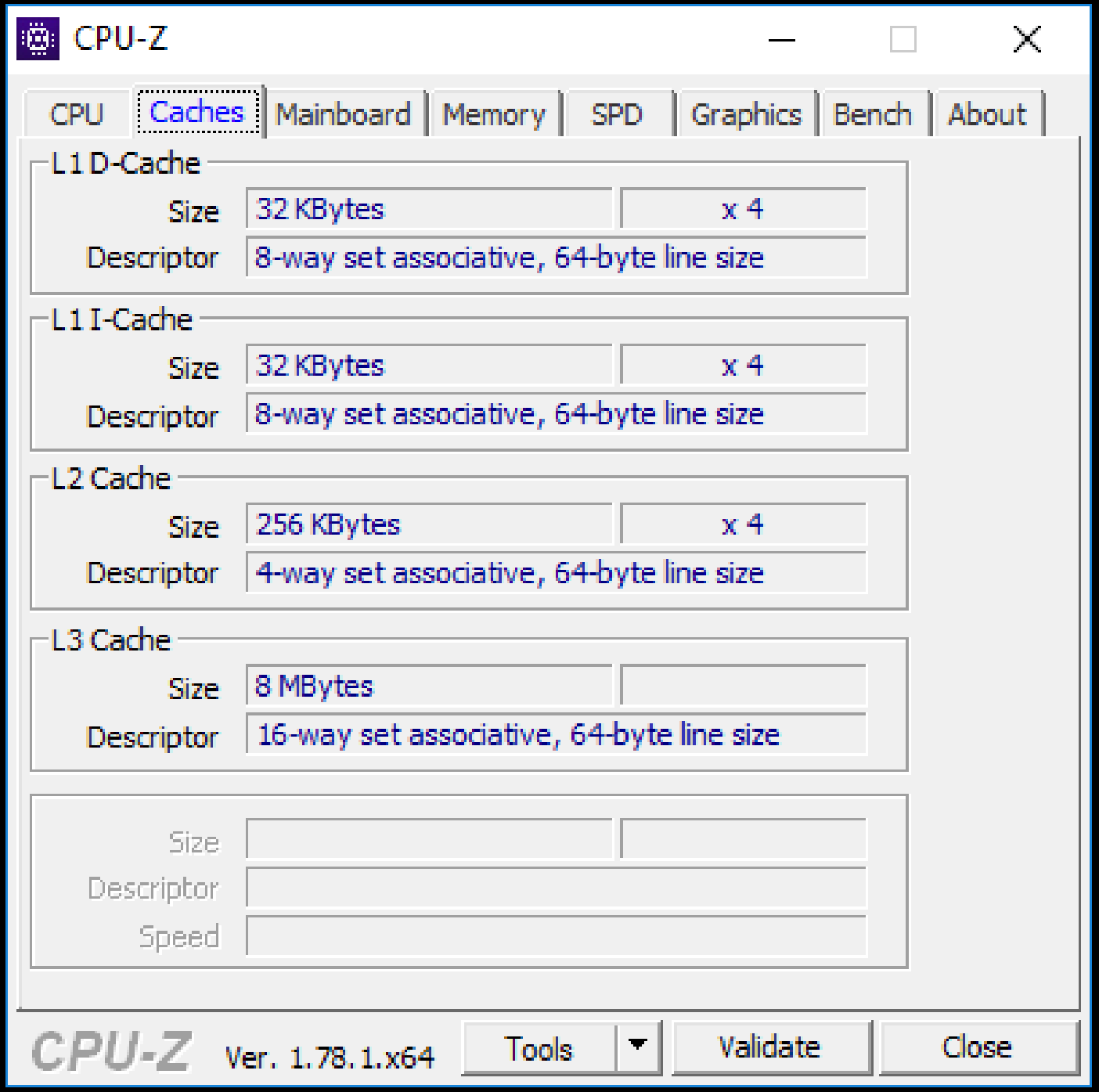
Cache Concepts

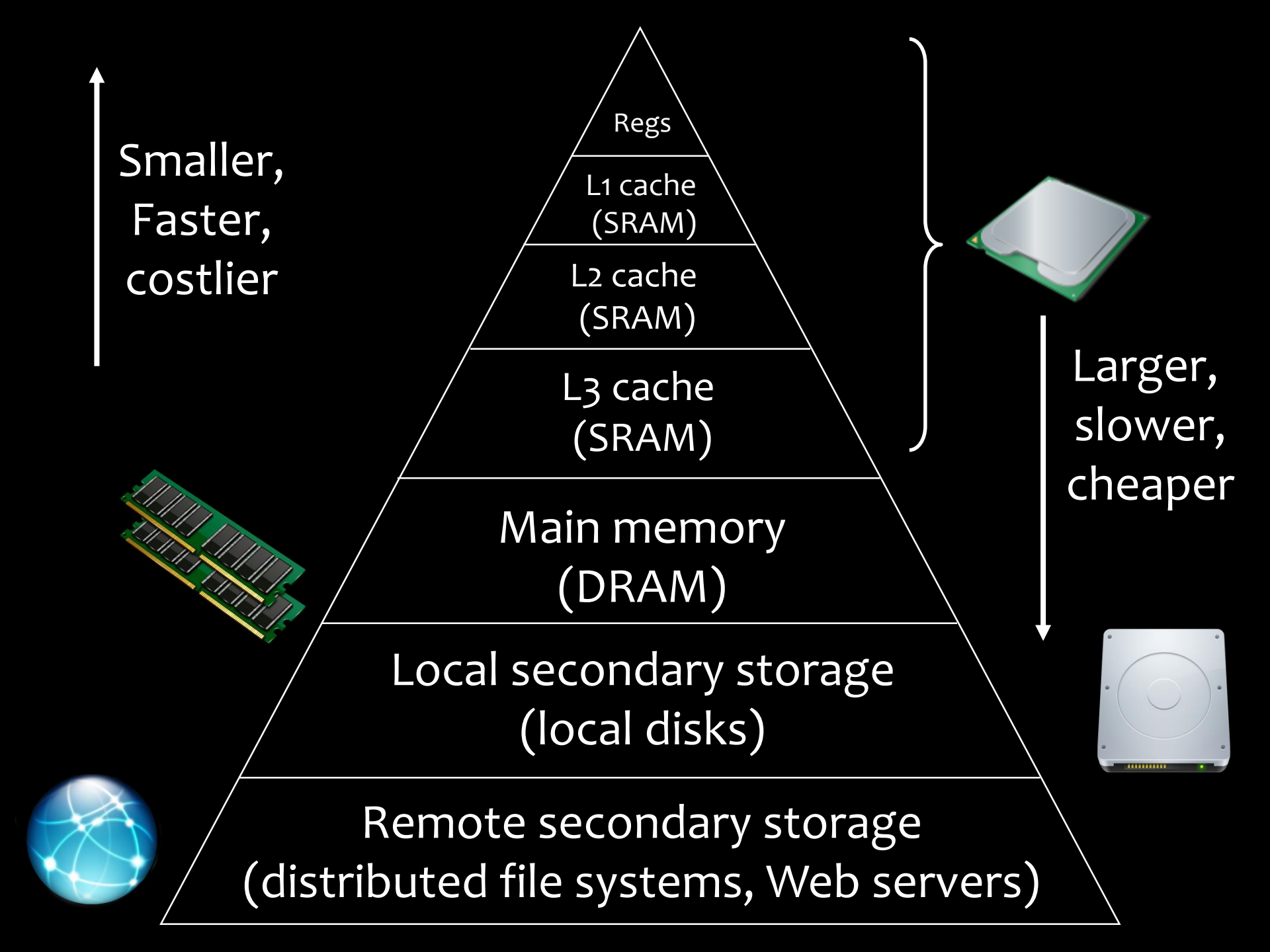
Memory

CPU

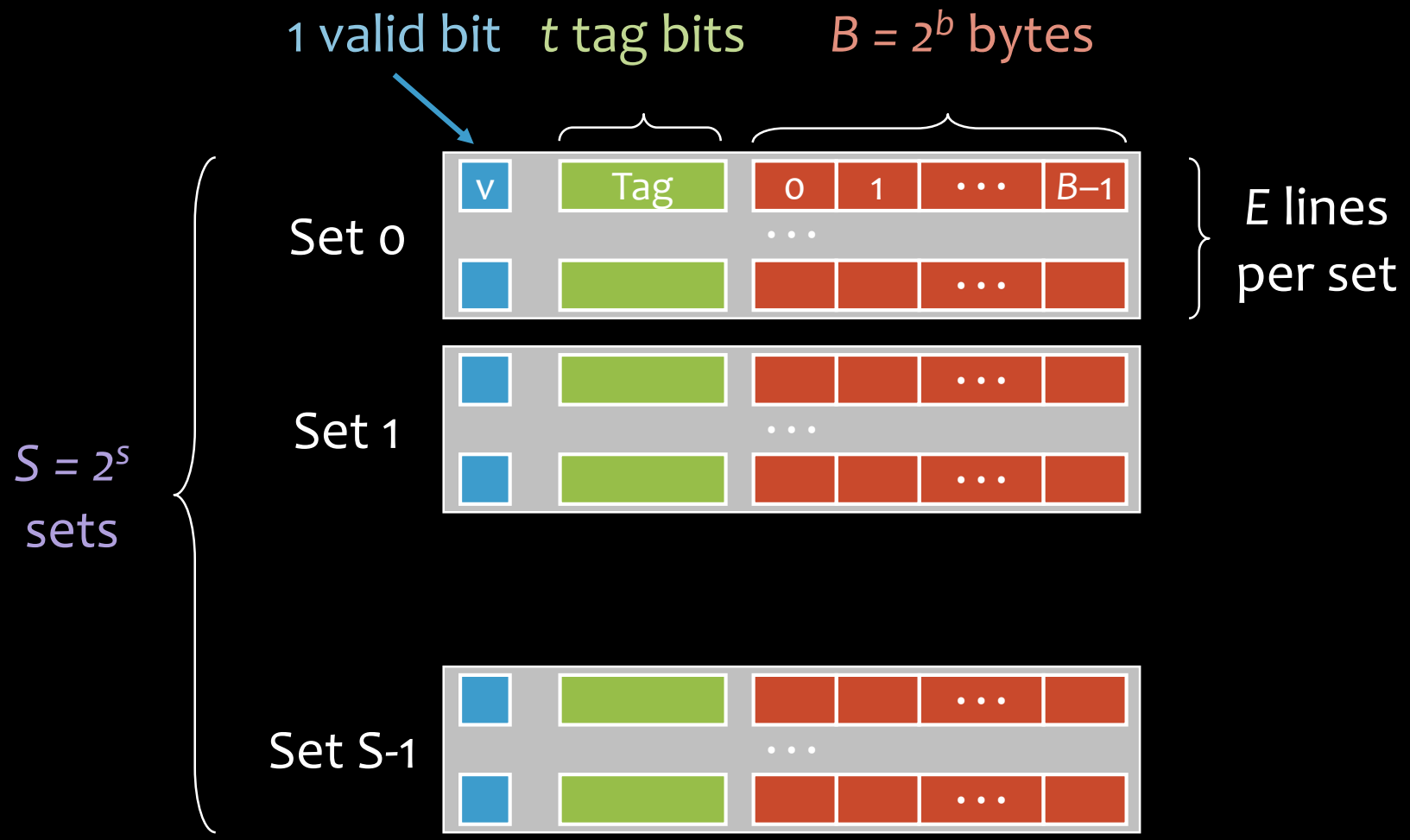


Cache

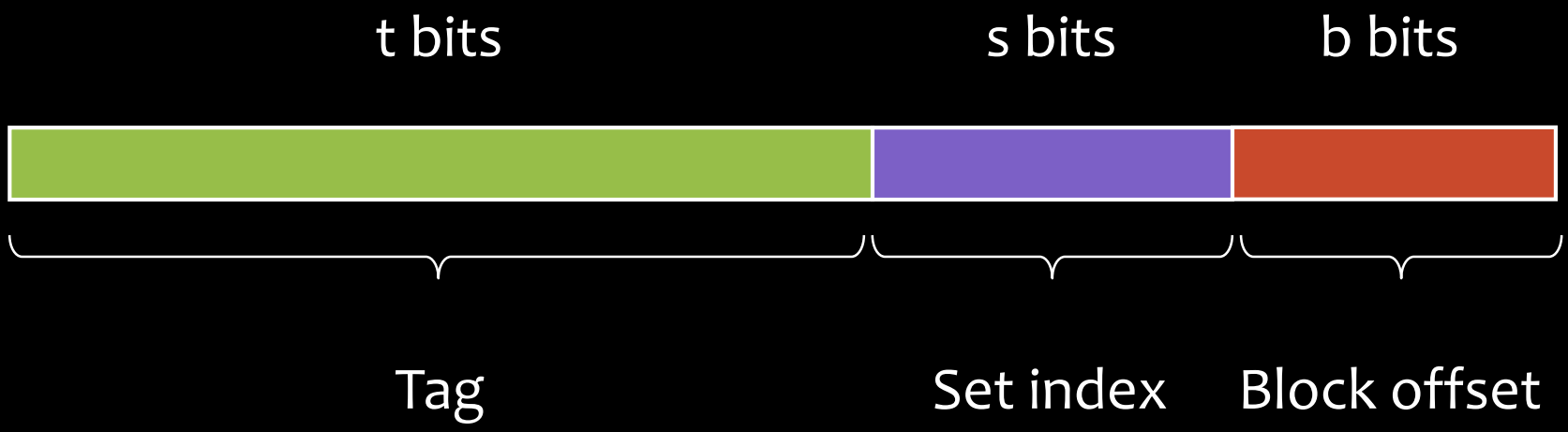

CPU-Z
CPU
Caches
Mainboard
Memory
SPD
Graphics
Bench
About
L1 D-Cache
Size
32 KBytes
x 4
Descriptor
8-way set associative, 64-byte line size
L1 I-Cache
Size
32 KBytes
x 4
Descriptor
8-way set associative, 64-byte line size
L2 Cache
Size
256 KBytes
x 4
Descriptor
4-way set associative, 64-byte line size
L3 Cache
Size
8 MBytes
Descriptor
16-way set associative, 64-byte line size
Size
Descriptor
Speed
CPU-Z
Ver. 1.78.1.x64
Tools
Validate
Close


Smaller,
Faster,
costlier
Regs
L1 cache
(SRAM)
L2 cache
(SRAM)
L3 cache
(SRAM)
Main memory
(DRAM)
Local secondary storage
(local disks)
Remote secondary storage
(distributed file systems, Web servers)
Larger,
slower,
cheaper

Cache organization



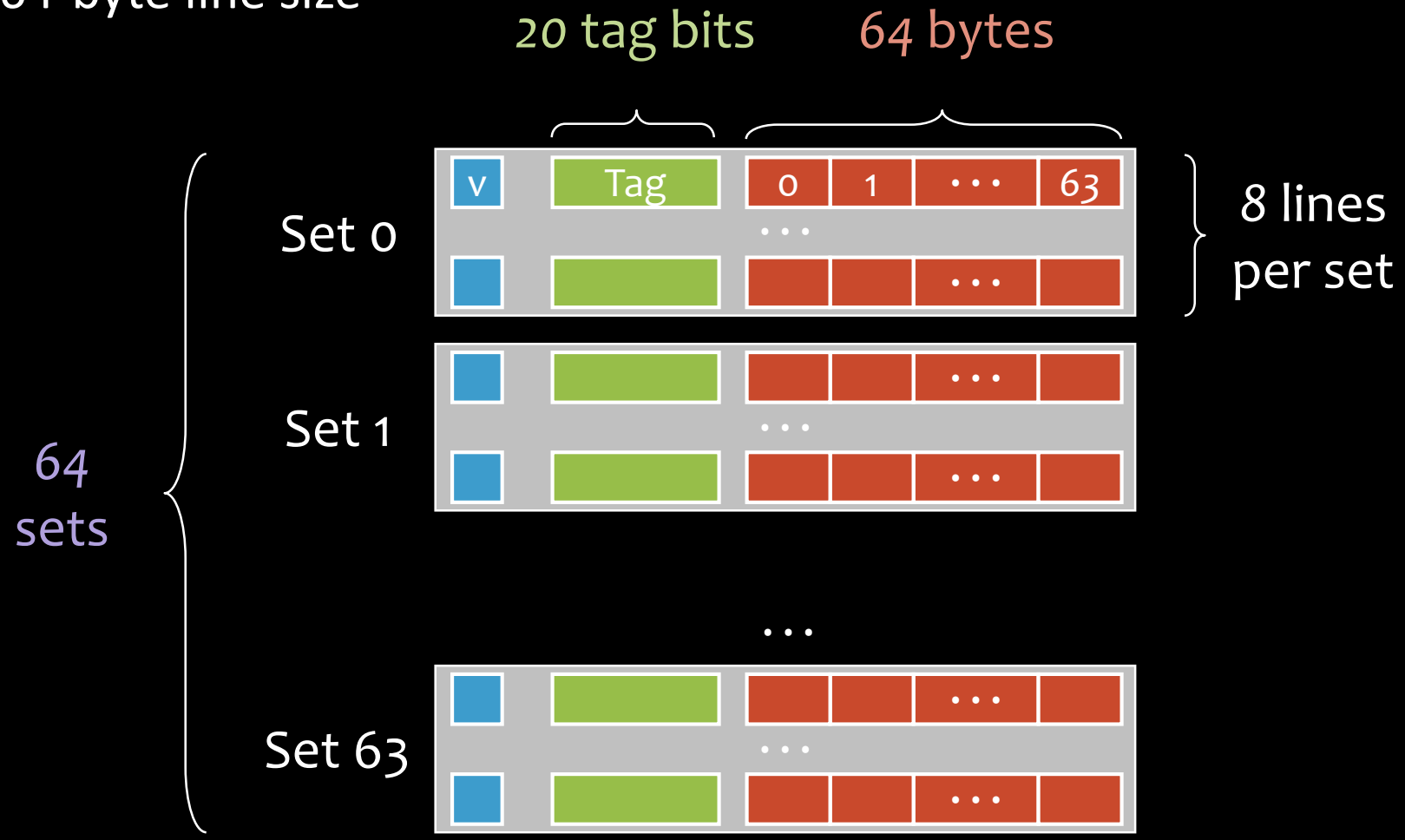
Partition of Address

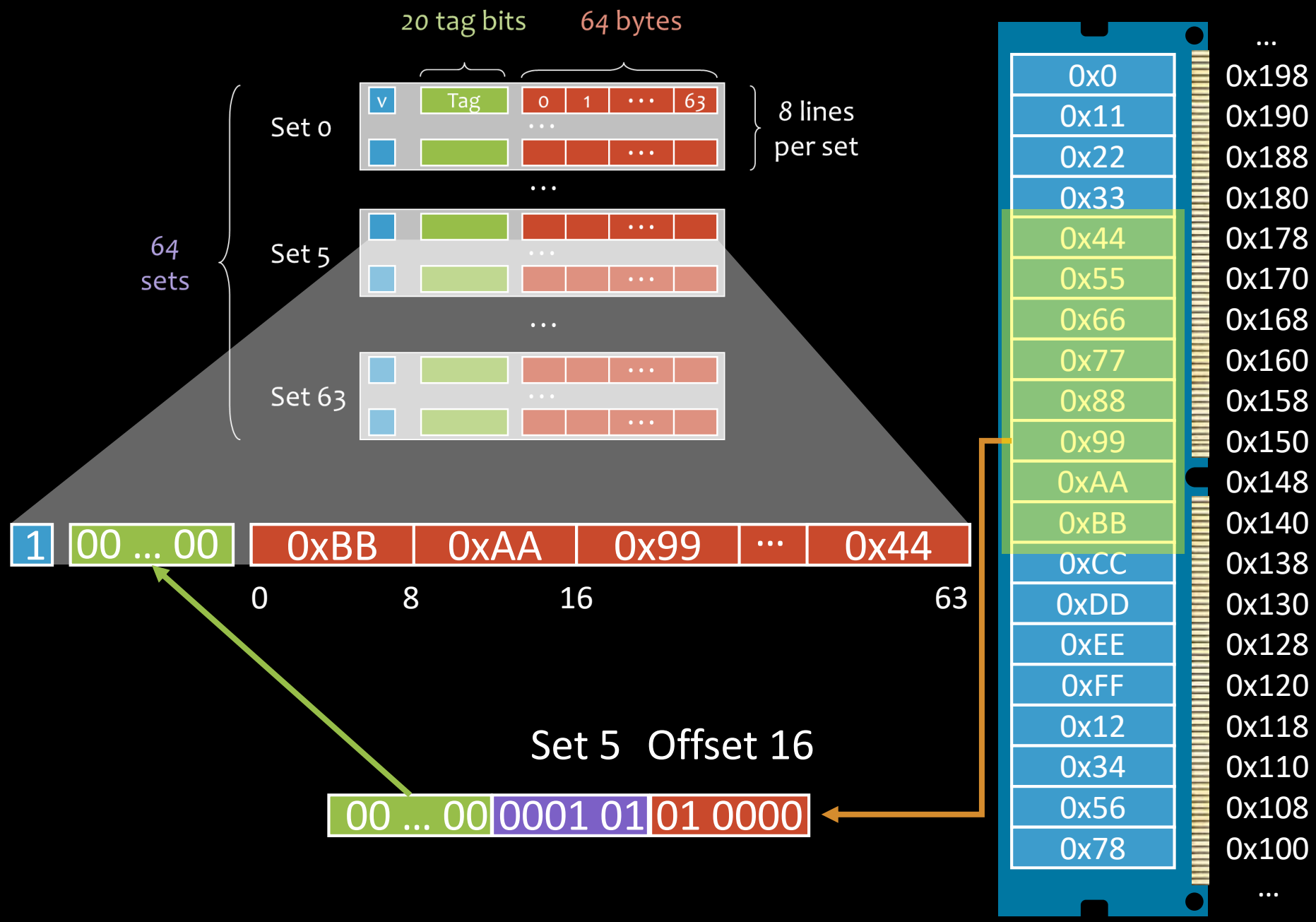


Memory address has 32 bits, determine S, t, s, b?

C	B	E	S	t	s	b
1,024	4	1				
1,024	8	4				
1,024	32	32				
32KB	64	8				
246KB	64	4				
8MB	64	16				

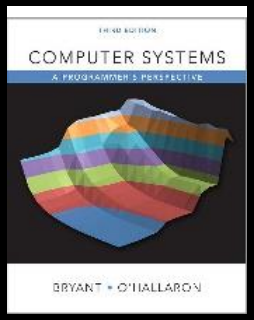
32 KBytes
8-way set associative
64-byte line size

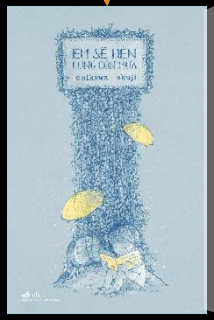


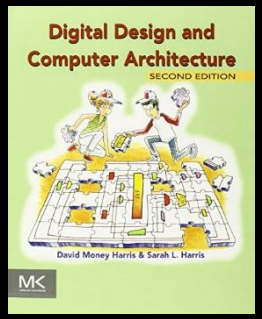

20 tag bits
64 bytes
v
Tag
0
1
63
Set 0
Set 5
Set 63
8 lines per set
64 sets
1
00 … 00
0xBB
0xAA
0x99
0x44
0
8
16
63
Set 5 Offset 16
00 … 00
0001 01
01 0000
0x0
0x11
0x22
0x33
0x44
0x55
0x66
0x77
0x88
0x99
0xAA
0xBB
0xCC
0xDD
0xEE
0xFF
0x12
0x34
0x56
0x78
0x198
0x190
0x188
0x180
0x178
0x170
0x168
0x160
0x158
0x150
0x148
0x140
0x138
0x130
0x128
0x120
0x118
0x110
0x108
0x100

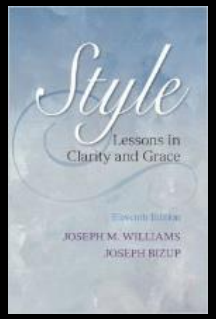
Line (Book)

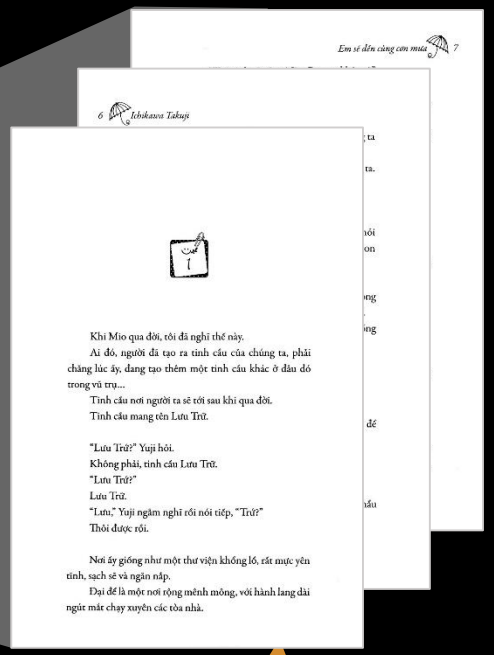
Set











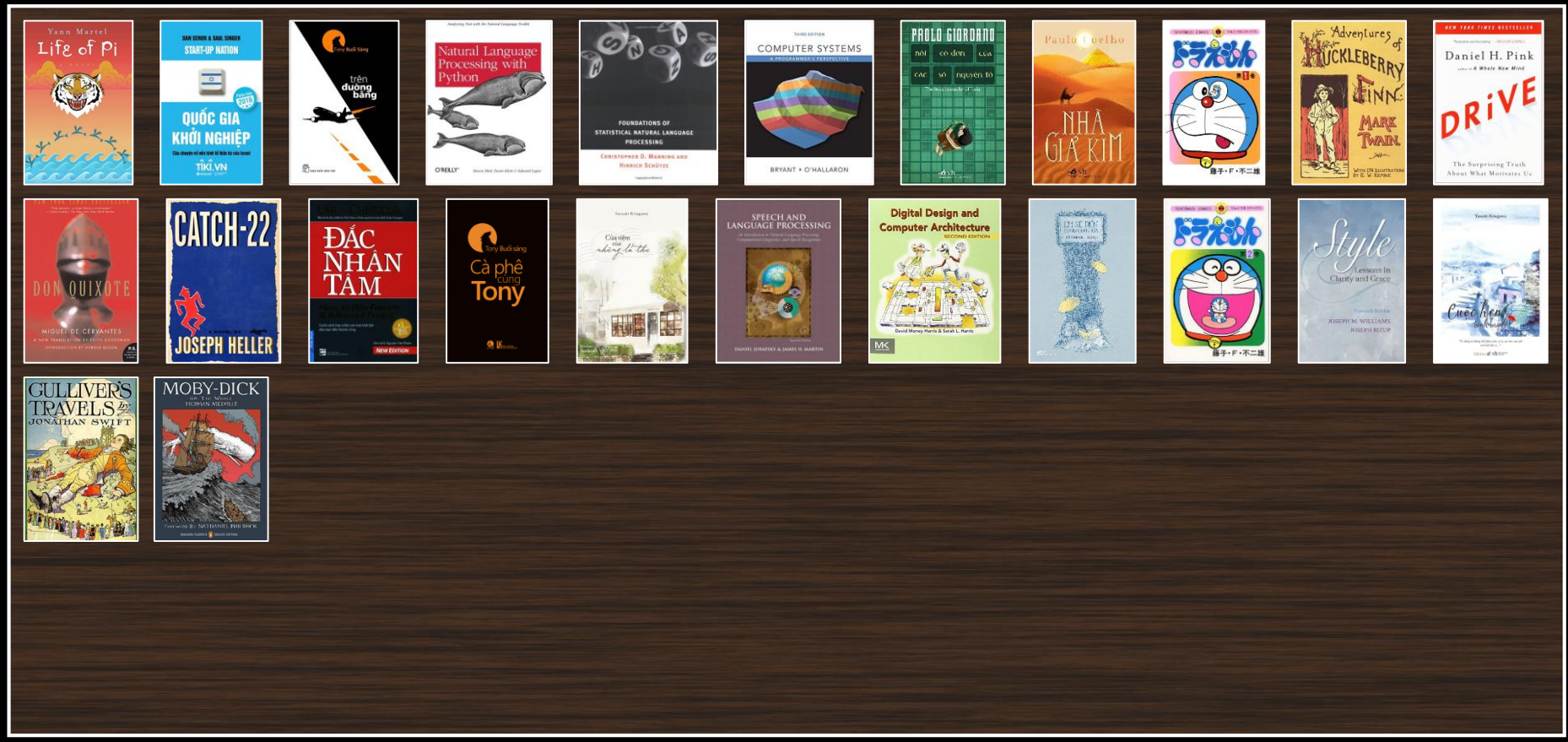
Block (Pages)

Cache Concepts

Memory

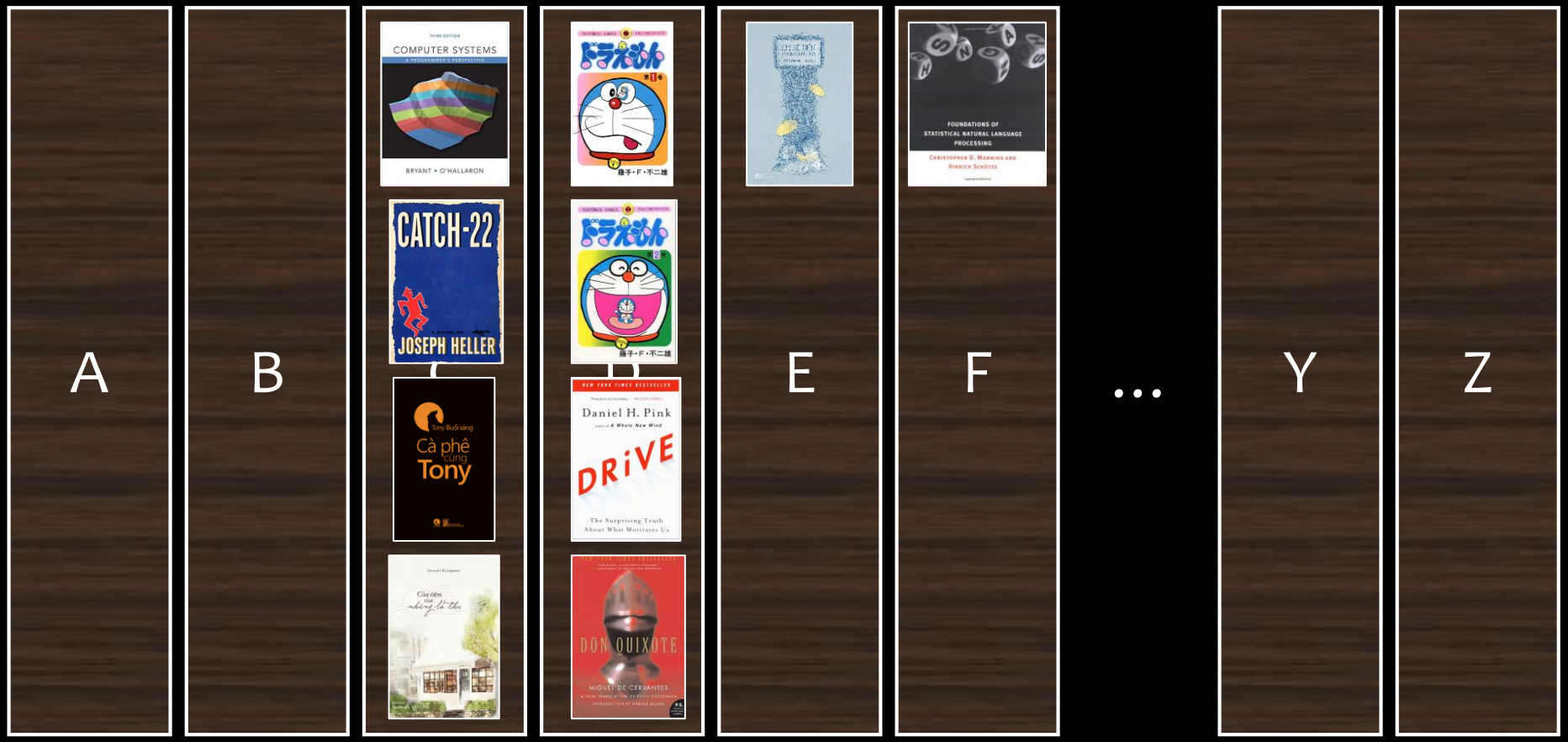
CPU

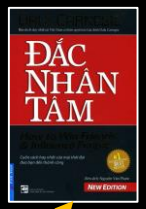
Cache



Put any book anywhere

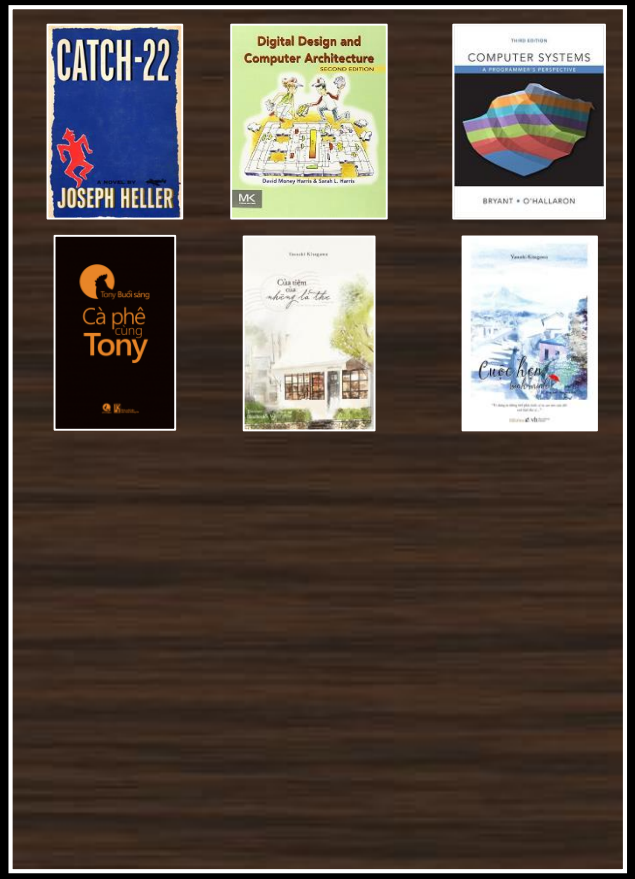
→ hard to find





conflict

→ easy to find



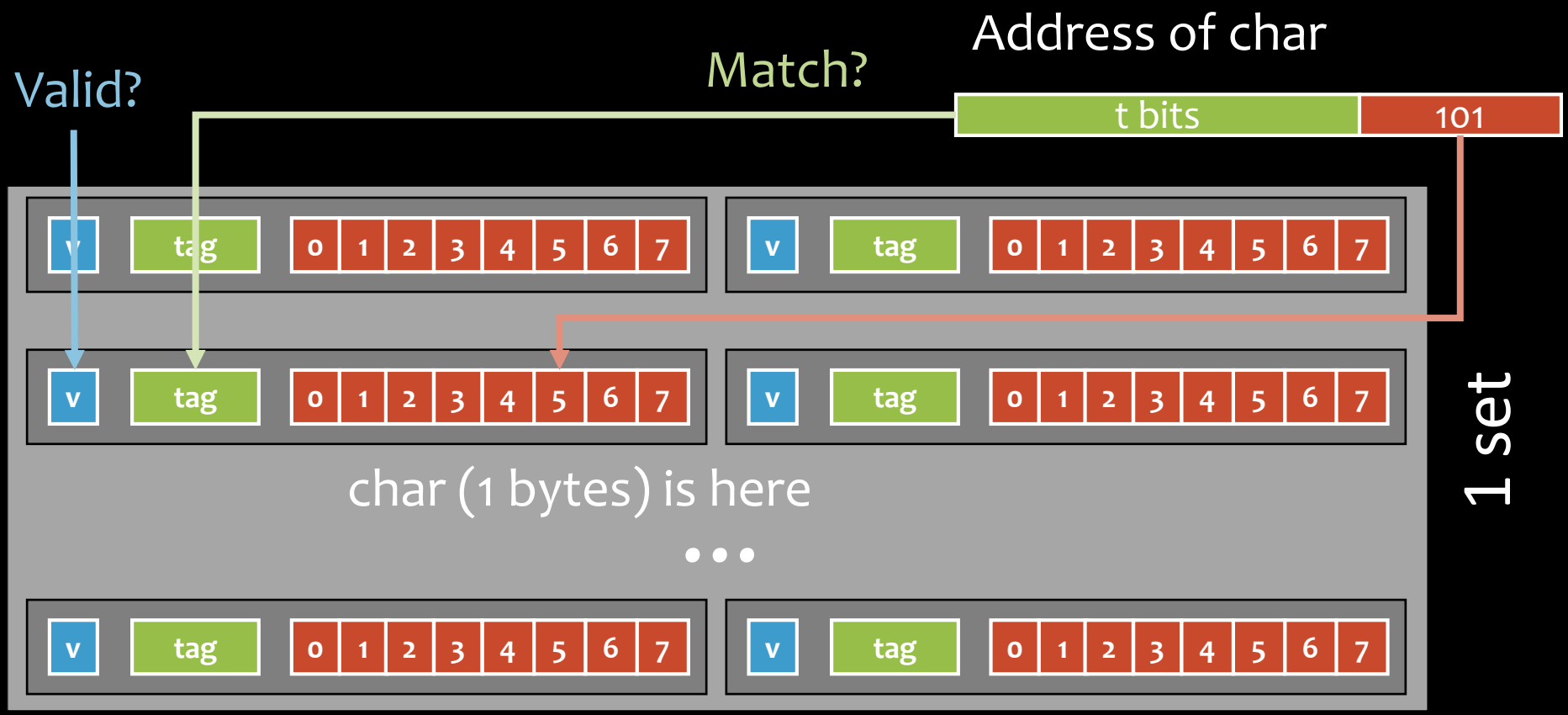
A or B or C

D or E or F

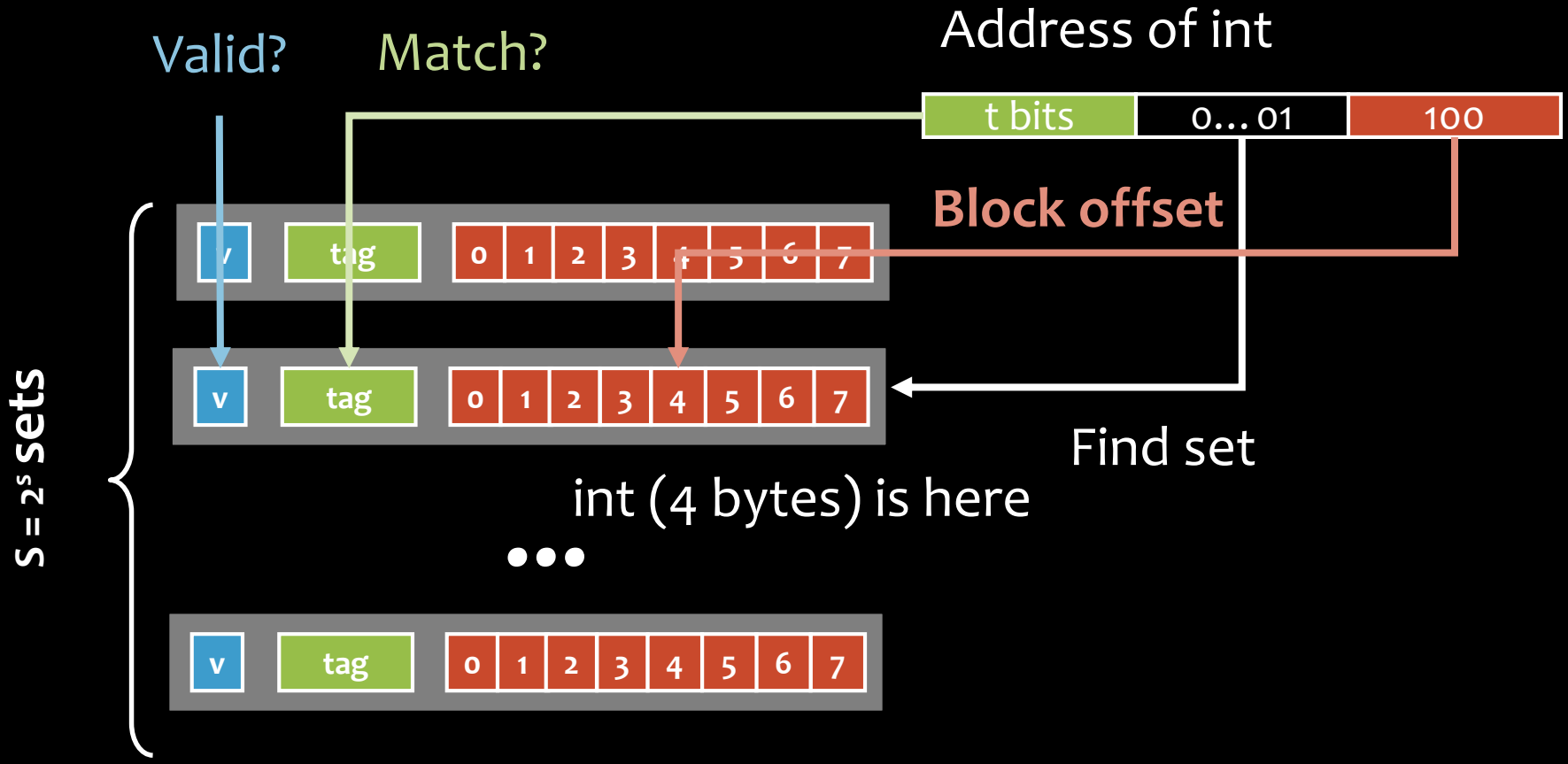
...

Y or Z

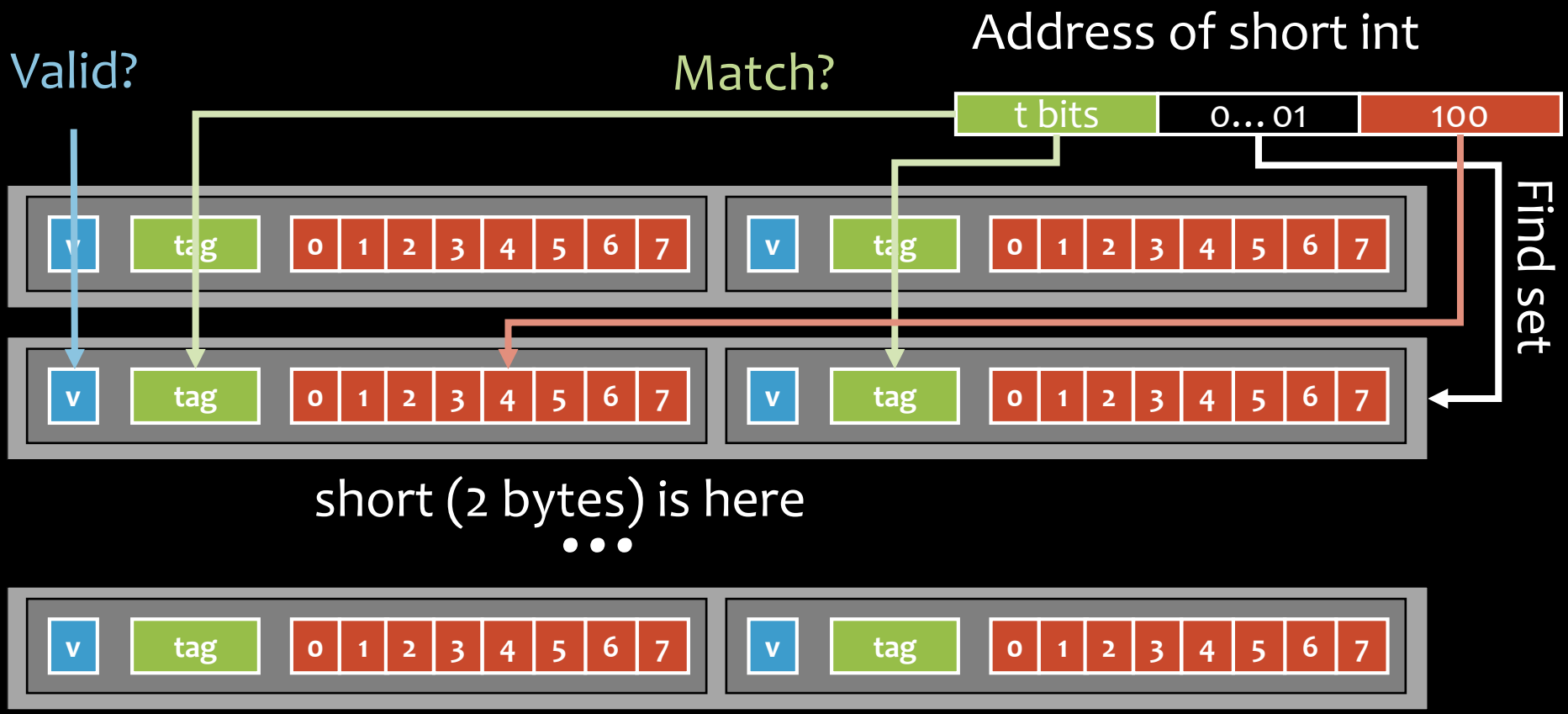
Fully-Associative Cache



Direct-Mapped Cache

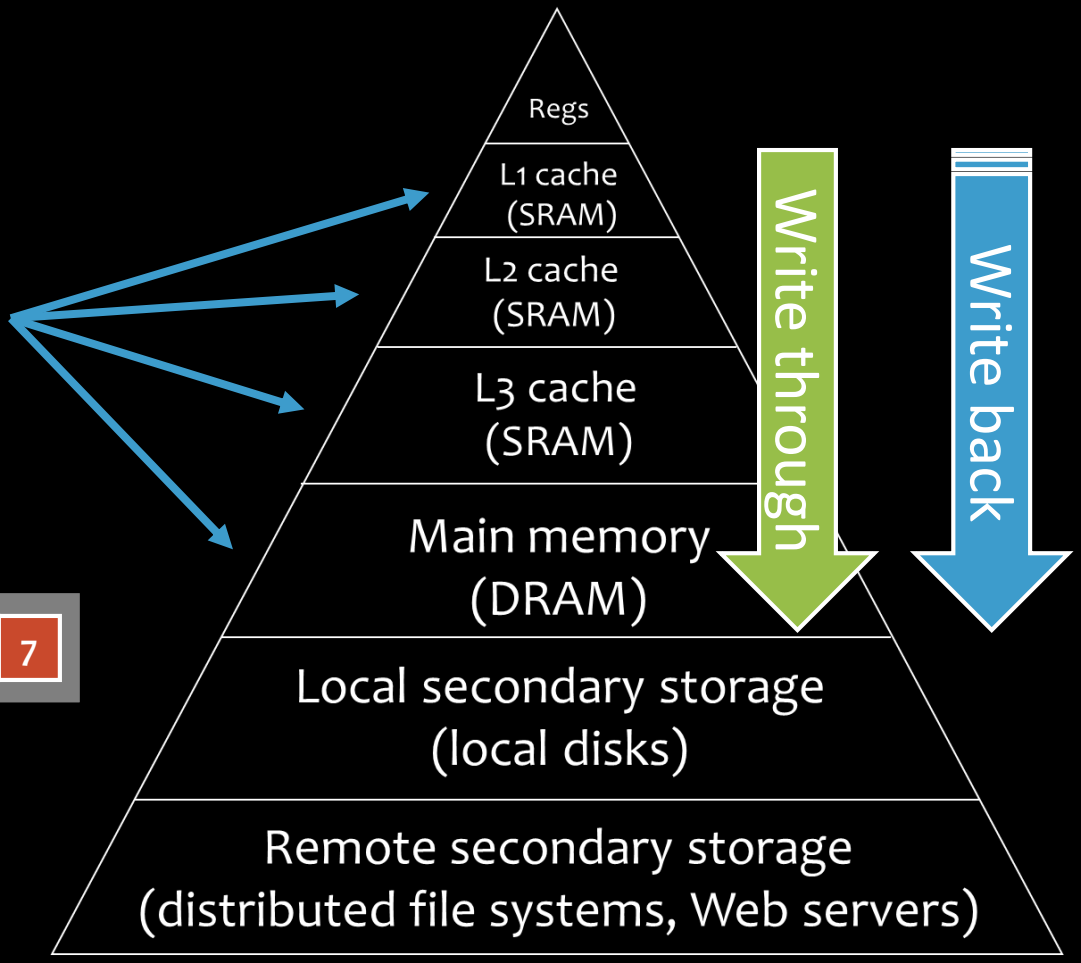


2-Way Set-Associative Cache



Valid bit

multiple copies



Summary

- Cache Concepts
  - Cache Hit
  - Cache Miss
- Cache Organization
  - Direct-mapped
  - E-way Set Associative
  - Fully Associative

Gene Myron Amdahl

formulating Amdahl's law

“

$$S_{latency}(s) = \frac{1}{(1-p) + \frac{p}{s}}$$

”