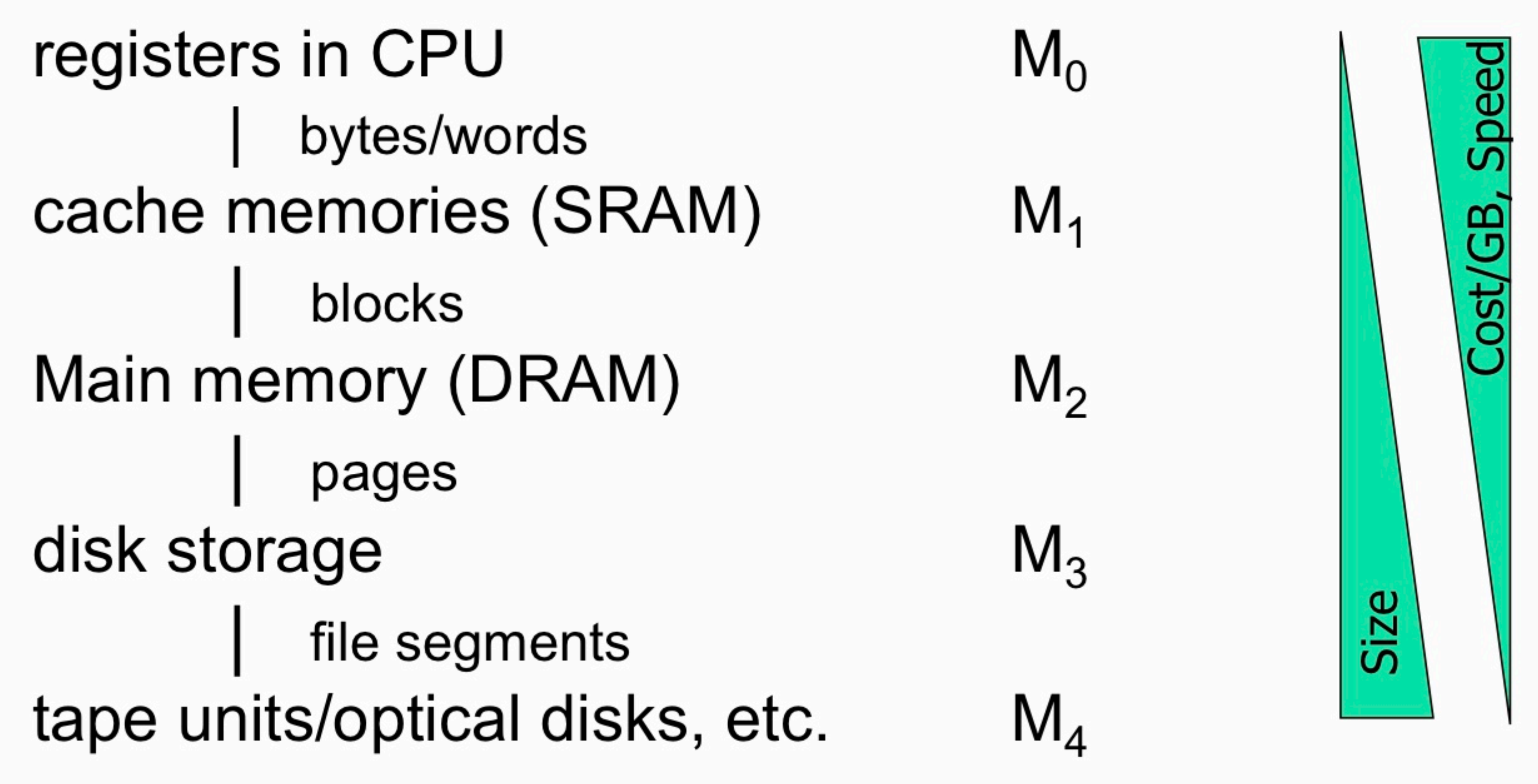
CG2028 COMPUTER ORGANIZATION

TUTORIAL 5: CACHE MEMORY PRINCIPLES

GU JING

Memory Hierarchy and Hit ratios



registers in CPU M_0

| bytes/words

cache memories (SRAM) M_1

| blocks

Main memory (DRAM) M_2

| pages

disk storage M_3

| file segments

tape units/optical disks, etc. M_4

Principle: each level holds the most frequently accessed data from the immediate higher level.

Advantage: reduces the effect of lower speed of the higher level without increasing the overall cost significantly.

CG2028 Tutorial 5: Cache Memory Principles

Q1. A computer system has 3 devices in the memory hierarchy – cache, RAM and Solid State Drive (SSD), having access times of 10 ns, 1 µs and 100 µs respectively. The access frequencies of two of the devices (not in any particular order) are 0.9 and 0.001. Compute the average access time for the memory system.

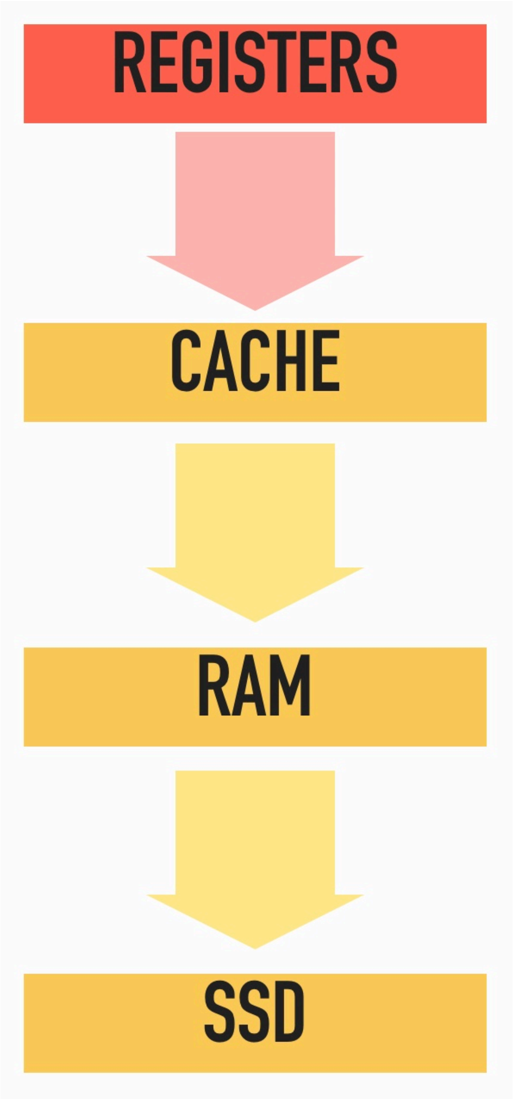
Hit Ratios and Access Frequencies

$f_i = (1-h_1)(1-h_2)(1-h_3).....(1-h_{i-1})h_i$

$f_1 + f_2 + f_3 + ... + f_i = 1$

$f_1 = h_1$

$T_{eff} = f_1t_1 + f_2t_2 + f_3t_3 + ... + f_nt_n$



Q1. A computer system has 3 devices in the memory hierarchy – cache, RAM and Solid State Drive (SSD), having access times of 10 ns, 1 µs and 100 µs respectively. The access frequencies of two of the devices (not in any particular order) are 0.9 and 0.001. Compute the average access time for the memory system.

Ans:

$f_x = 1 - 0.9 - 0.001 = 0.099$

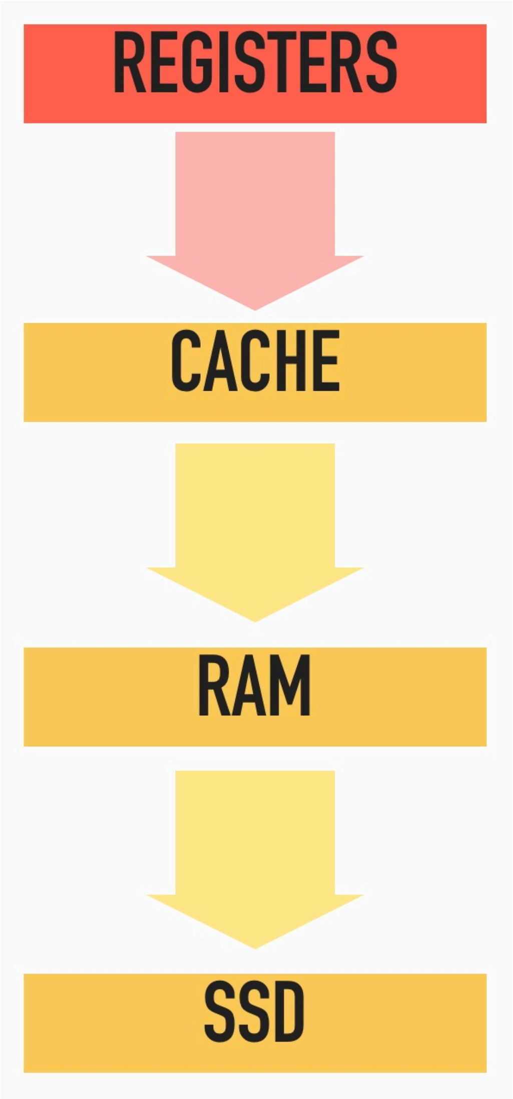
So $f_1 = 0.9$, $f_2 = 0.099$, $f_3 = 0.01$

With $t_1 = 10ns$, $t_2 = 1us$, $t_3 = 100us$, we get:

$$T_{eff} = f_1t_1 + f_2t_2 + f_3t_3$$

$$= 0.9*10 + 0.099*1000 + 0.001*100000$$

$$= 208ns$$



REGISTERS

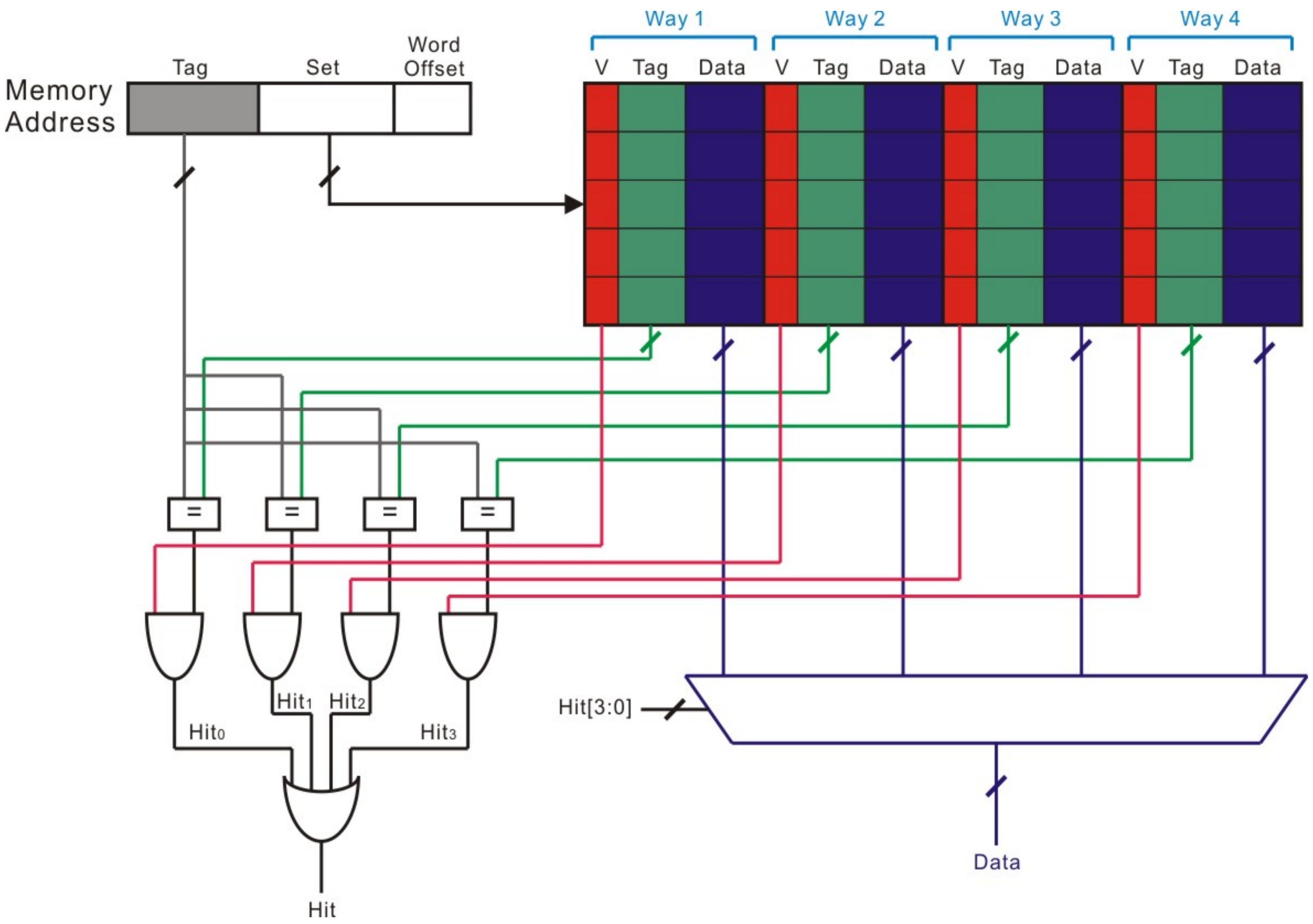
CACHE

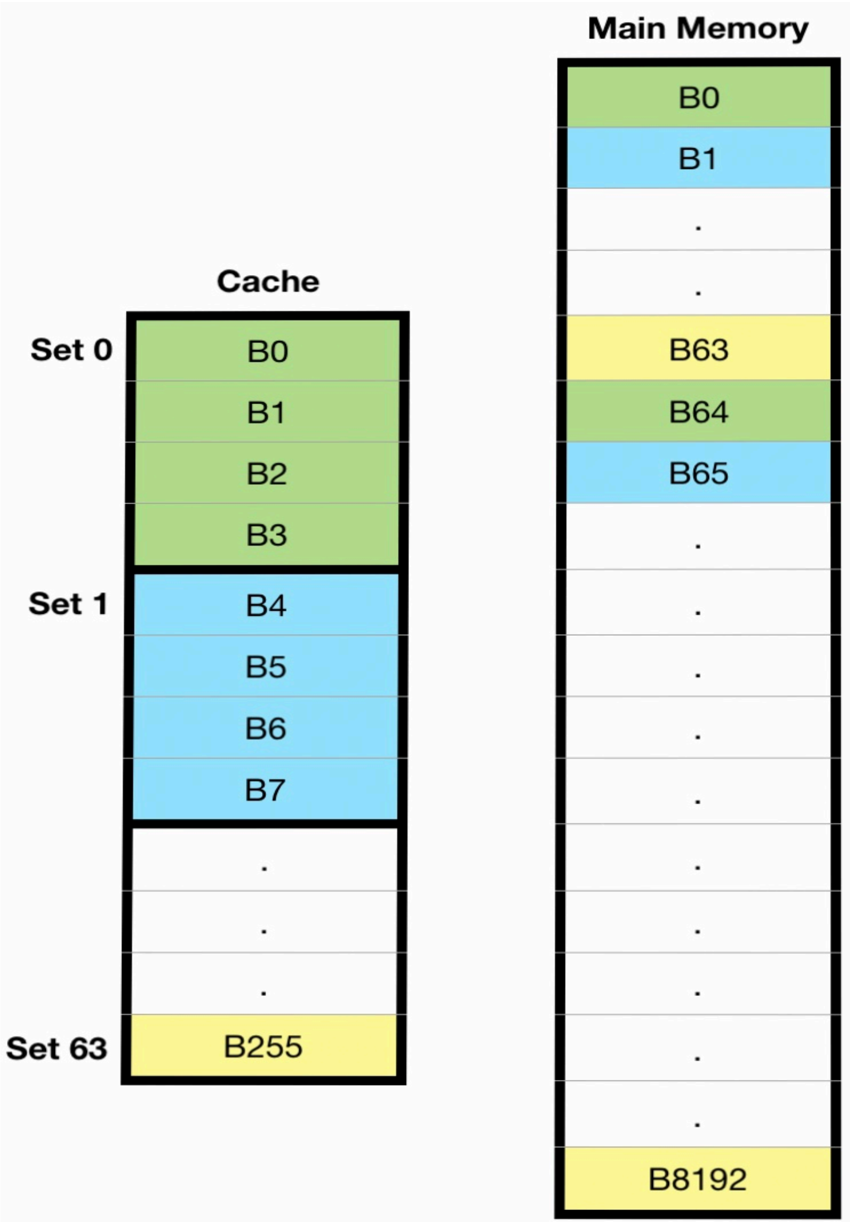
RAM

SSD

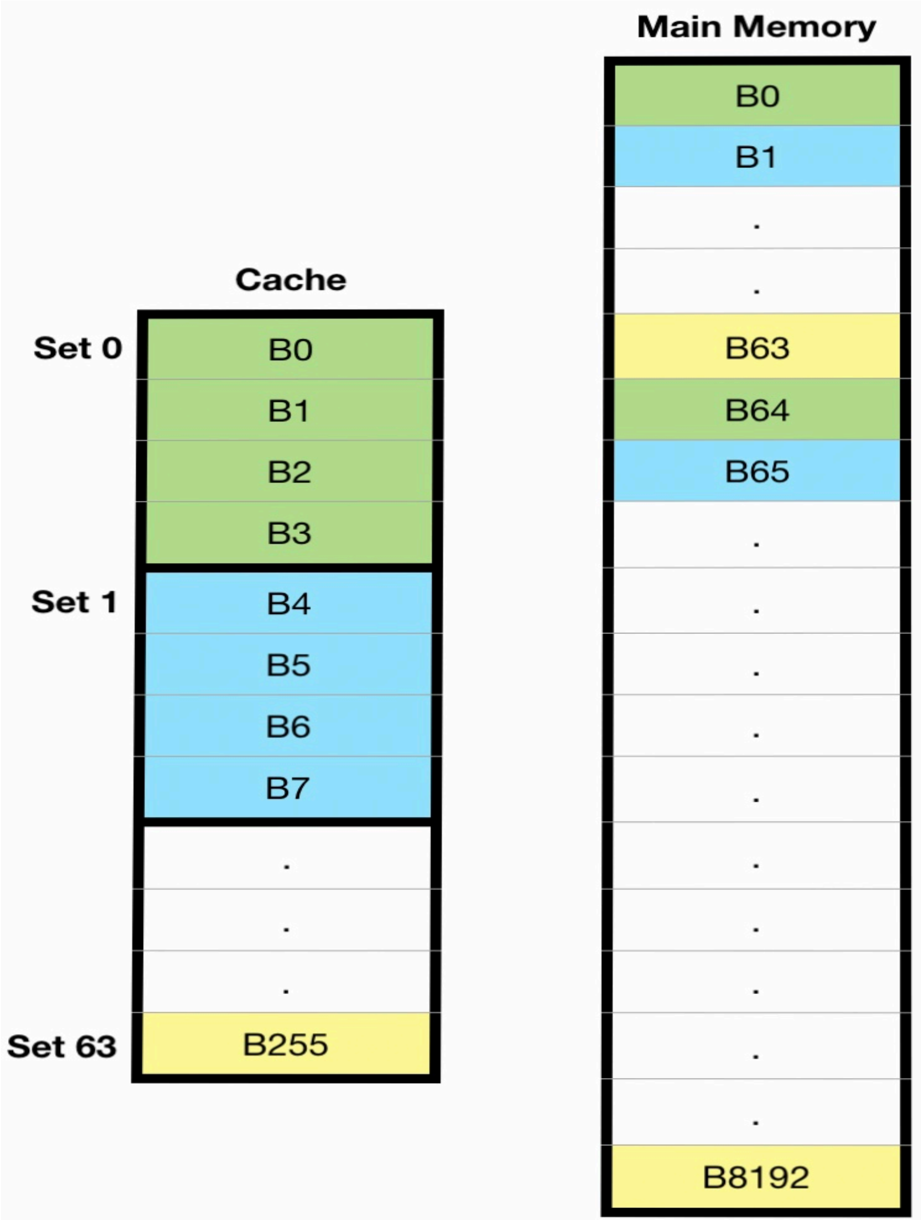
CG2028 Tutorial 5: Cache Memory Principles

Q2. A computer system has main memory capacity of 8192 blocks whereas the cache can hold 256 blocks, organized into 64 sets. What is the maximum number of comparisons required before the cache controller can determine whether the data addressed by the CPU is present in the cache? How many bits will need to be stored as overhead for the tag field in each cache block?





Q2. A computer system has main memory capacity of 8192 blocks whereas the cache can hold 256 blocks, organized into 64 sets. What is the maximum number of comparisons required before the cache controller can determine whether the data addressed by the CPU is present in the cache? How many bits will need to be stored as overhead for the tag field in each cache block?



Ans:

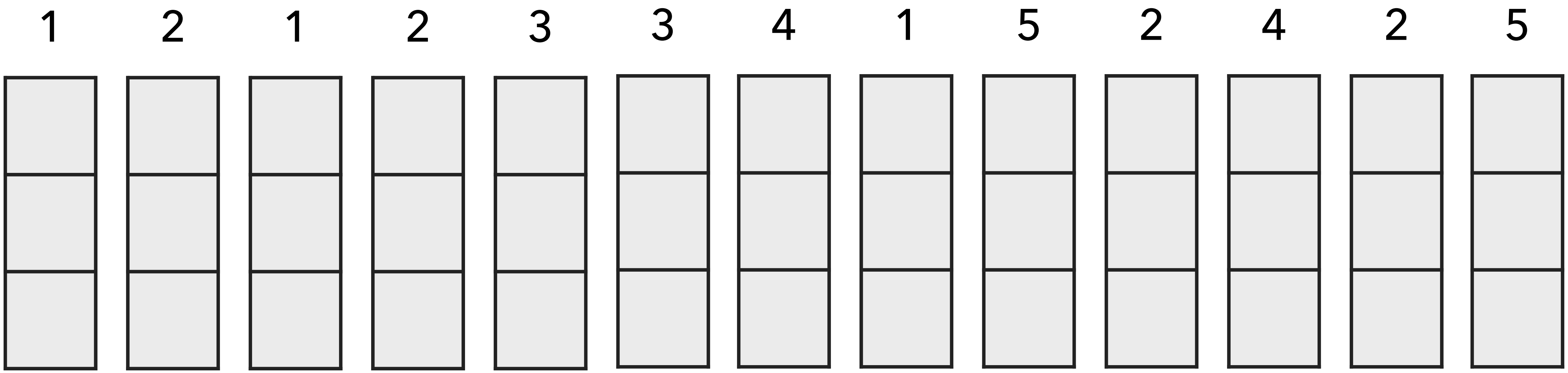
256 blocks are organized into 64 sets, so the cache is 4-way set-associative, and hence we need 4 comparisons to check if a particular block is present within the set it is mapped to.

The tag filed requires $\log_2(8192/64)$ = 7 bits.

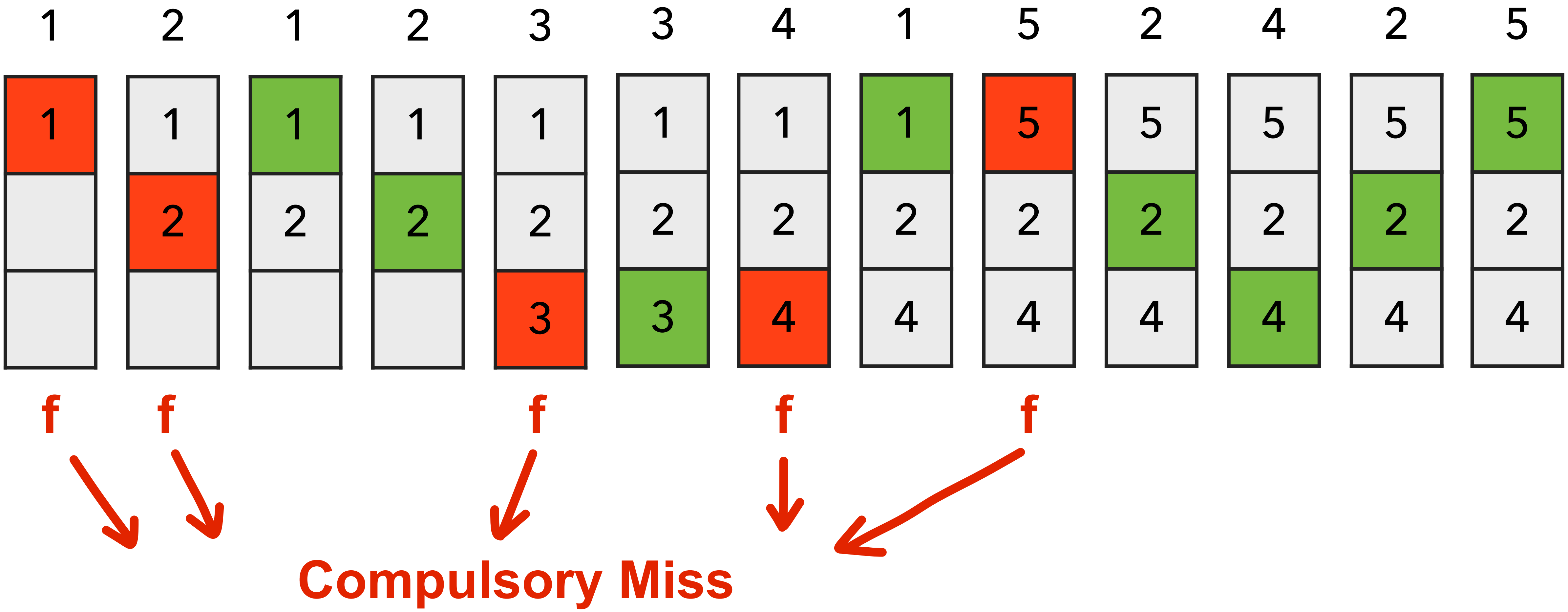
Note that additional overhead bits such as valid bit and dirty bit may be required for each block

CG2028 Tutorial 5: Cache Memory Principles

Q3. A cache can hold 3 blocks and is fully associative. Consider a request sequence as follows – 1, 2, 1, 2, 3, 3, 4, 1, 5, 2, 4, 2, 5. Draw a diagram that identifies the block to be replaced whenever a miss occurs, assuming an optimal replacement algorithm is used.



Q3. A cache can hold 3 blocks and is fully associative. Consider a request sequence as follows – 1, 2, 1, 2, 3, 3, 4, 1, 5, 2, 4, 2, 5. Draw a diagram that identifies the block to be replaced whenever a miss occurs, assuming an optimal replacement algorithm is used.

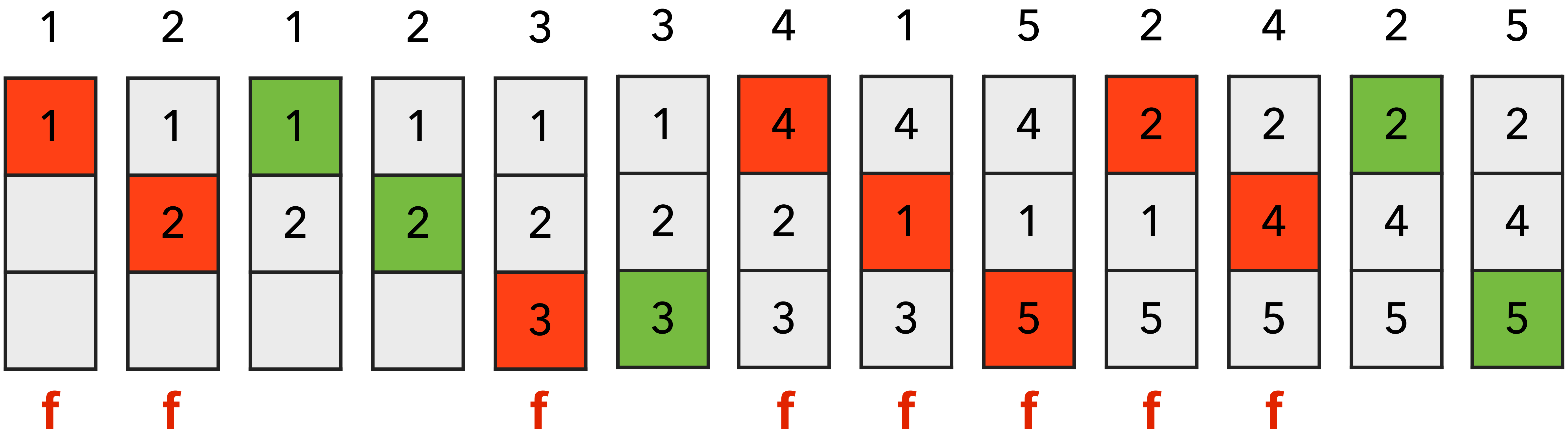


1	2	1	2	3	3	4	1	5	2	4	2	5
1	1	1	1	1	1	1	1	5	5	5	5	5
	2	2	2	2	2	2	2	2	2	2	2	2
				3	3	4	4	4	4	4	4	4
f	f			f		f		f				

Compulsory Miss

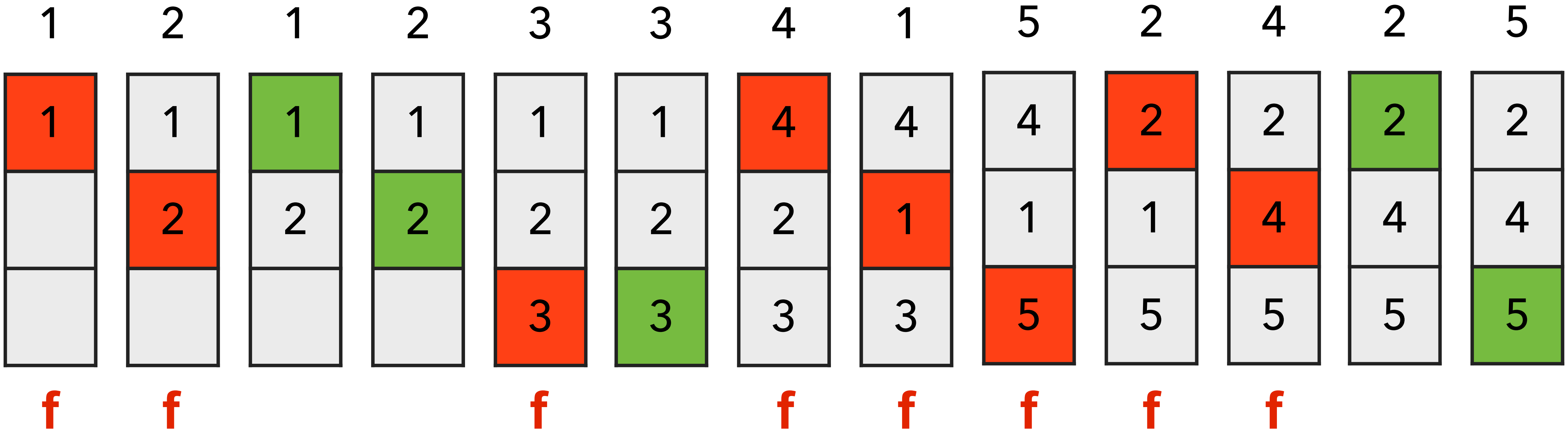
LRU or FIFO would have led to higher misses. For example, to accommodate block 4, FIFO and LRU would have replaced block 1, causing a miss at the next request.

(FIFO) First-In-First-Out:



1	2	1	2	3	3	4	1	5	2	4	2	5
🟥1	1	🟩1	1	1	1	🟥4	4	4	🟥2	2	🟩2	2
	🟥2	2	🟩2	2	2	2	🟥1	1	1	🟥4	4	4
				🟥3	🟩3	3	3	🟥5	5	5	5	🟩5
f	f			f		f	f	f	f	f		

(LRU) Least Recently Used:



1	2	1	2	3	3	4	1	5	2	4	2	5
🟥1	1	🟩1	1	1	1	🟥4	4	4	🟥2	2	🟩2	2
	🟥2	2	🟩2	2	2	2	🟥1	1	1	🟥4	4	4
				🟥3	🟩3	3	3	🟥5	5	5	5	🟩5
f	f			f		f	f	f	f	f		

CG2028 Tutorial 5: Cache Memory Principles

Q4.Which mapping strategy is used for level 1 cache by the latest processors from Intel? (not examinable)

Website https://en.wikichip.org is an excellent resource for information on various microprocessors.

THE END

For Consultation: eleguji@nus.edu.sg
Office: E4-03-10