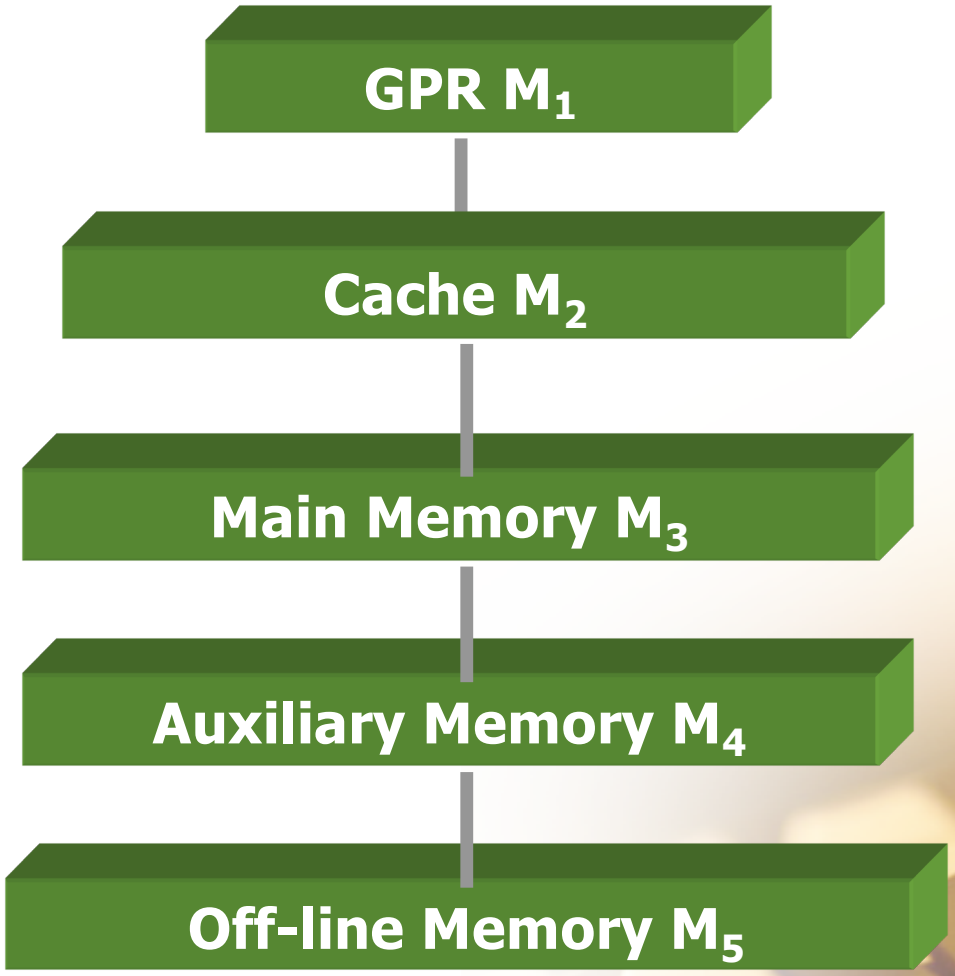
Chapter3 Memory System

Hierarchy of Memory System



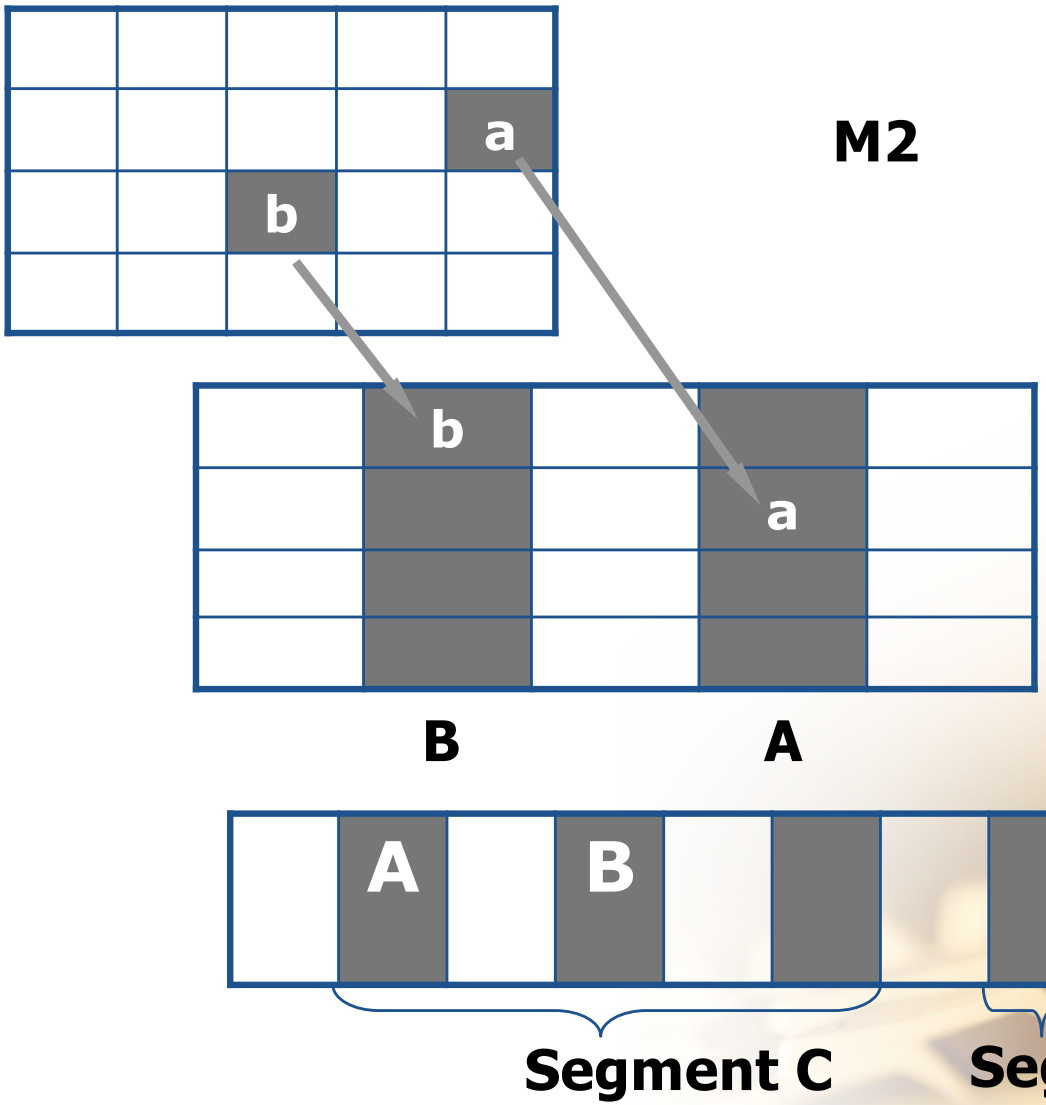
Chapter3 Memory System

Performance Parameters

- S　　B, KB, MB, GB, TB
- T　　Access Cycle
- C　　$C/bit, $C/KB

Chapter3 Memory System

Inclusiveness and Consistency



Chapter3 Memory System

Parallel Memory

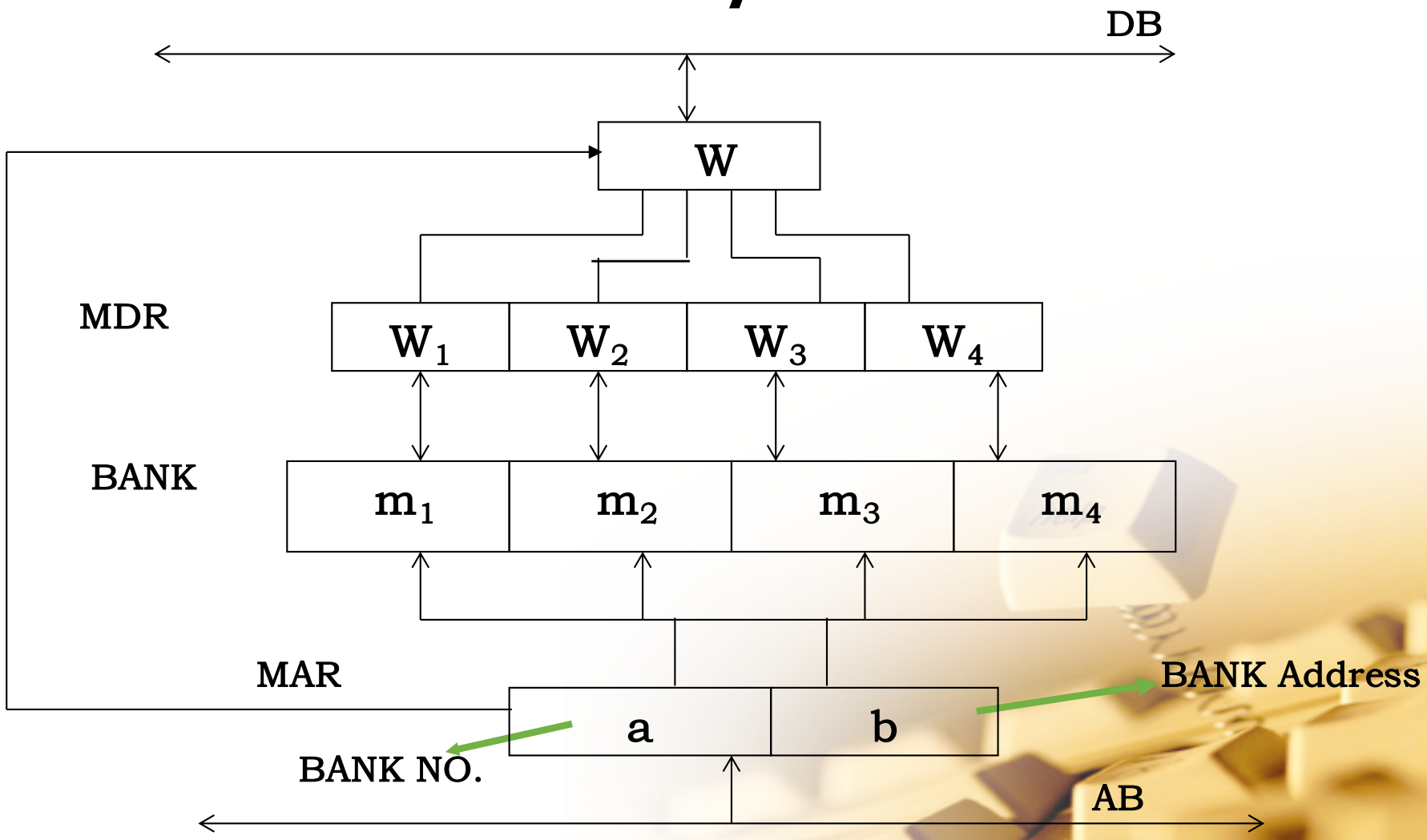
Bandwidth: Data size can be accessed per unit time

Methods for increasing bandwidth

- **Parallel Memory**
- **Buffer Memory**
- **Cache**

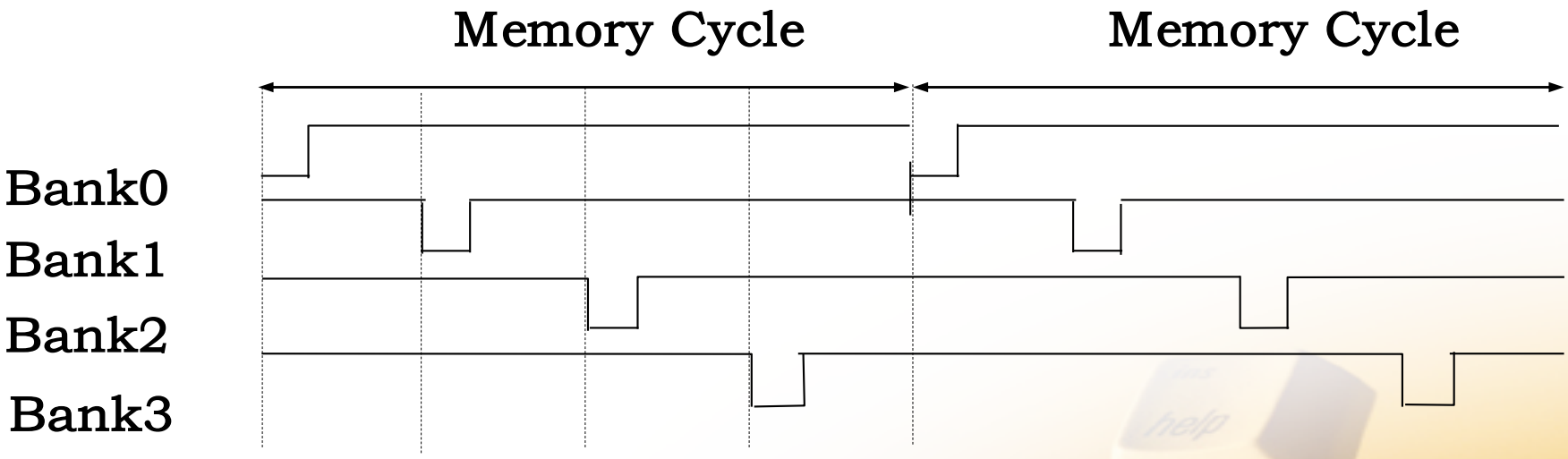
Chapter3 Memory System

Multi-bank Parallel Memory



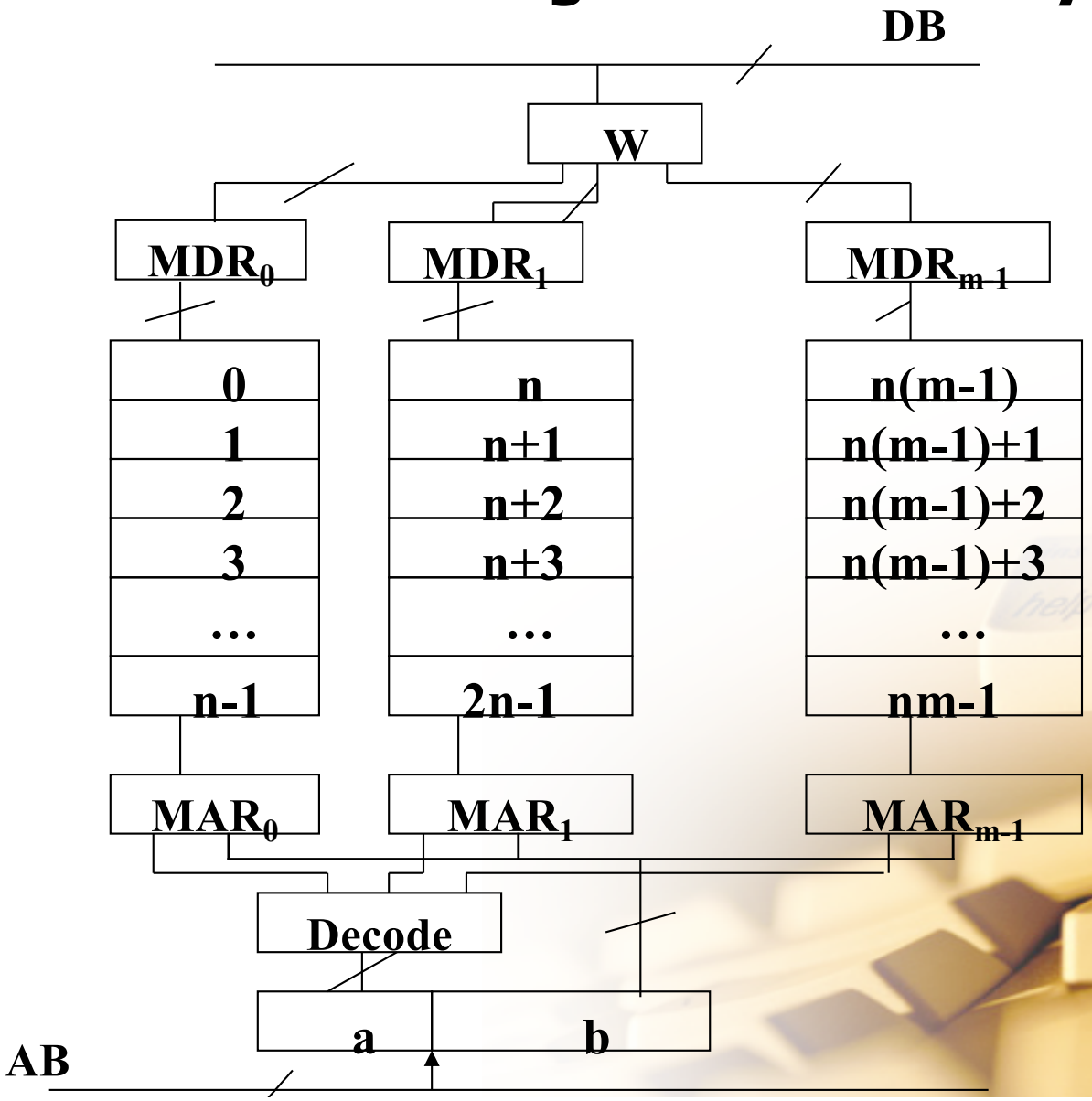
Chapter3 Memory System

Multi-bank Parallel Interleaving Access Memory



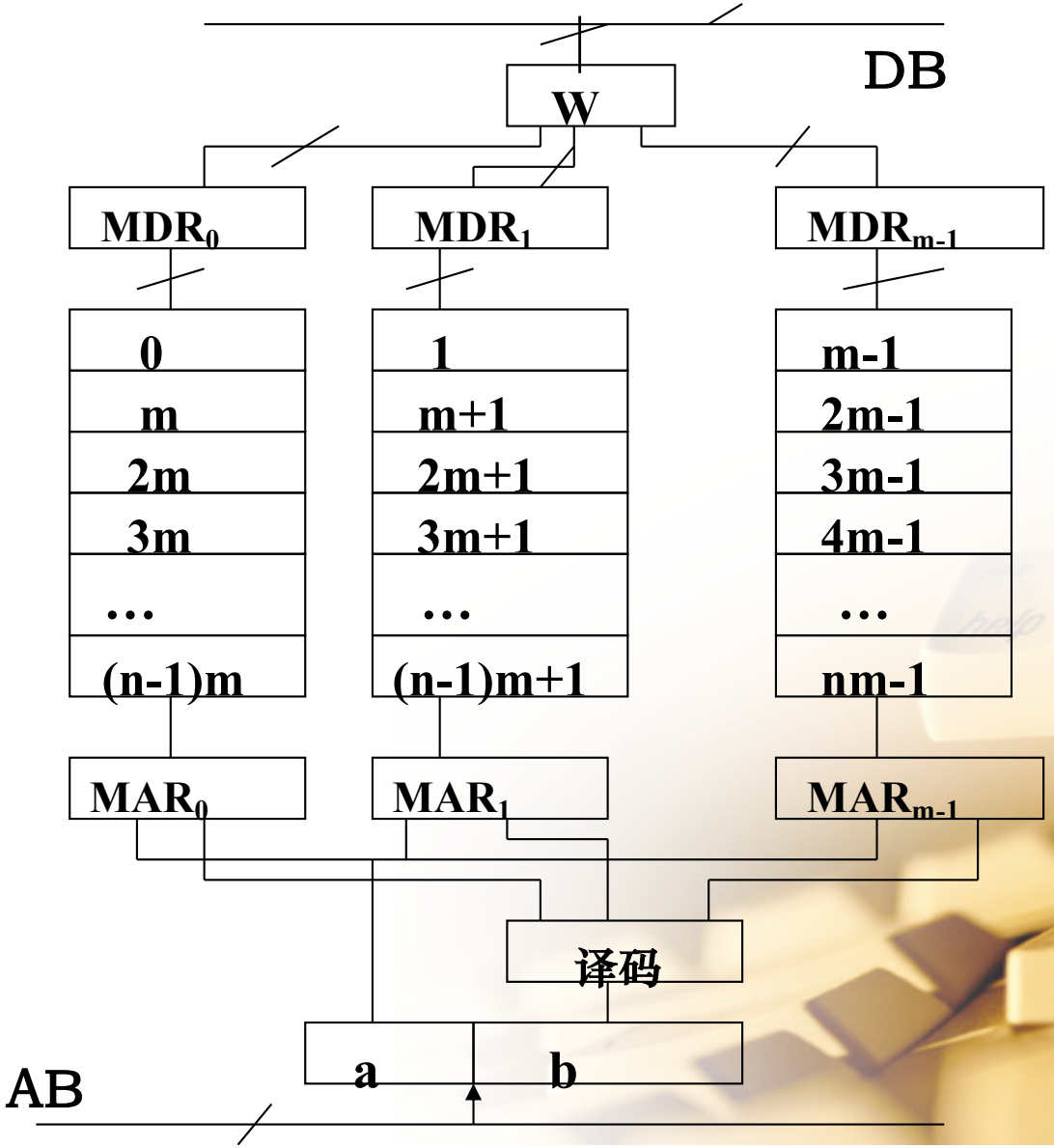
Chapter3 Memory System

High-order Interleaving Access Memory



Chapter3 Memory System

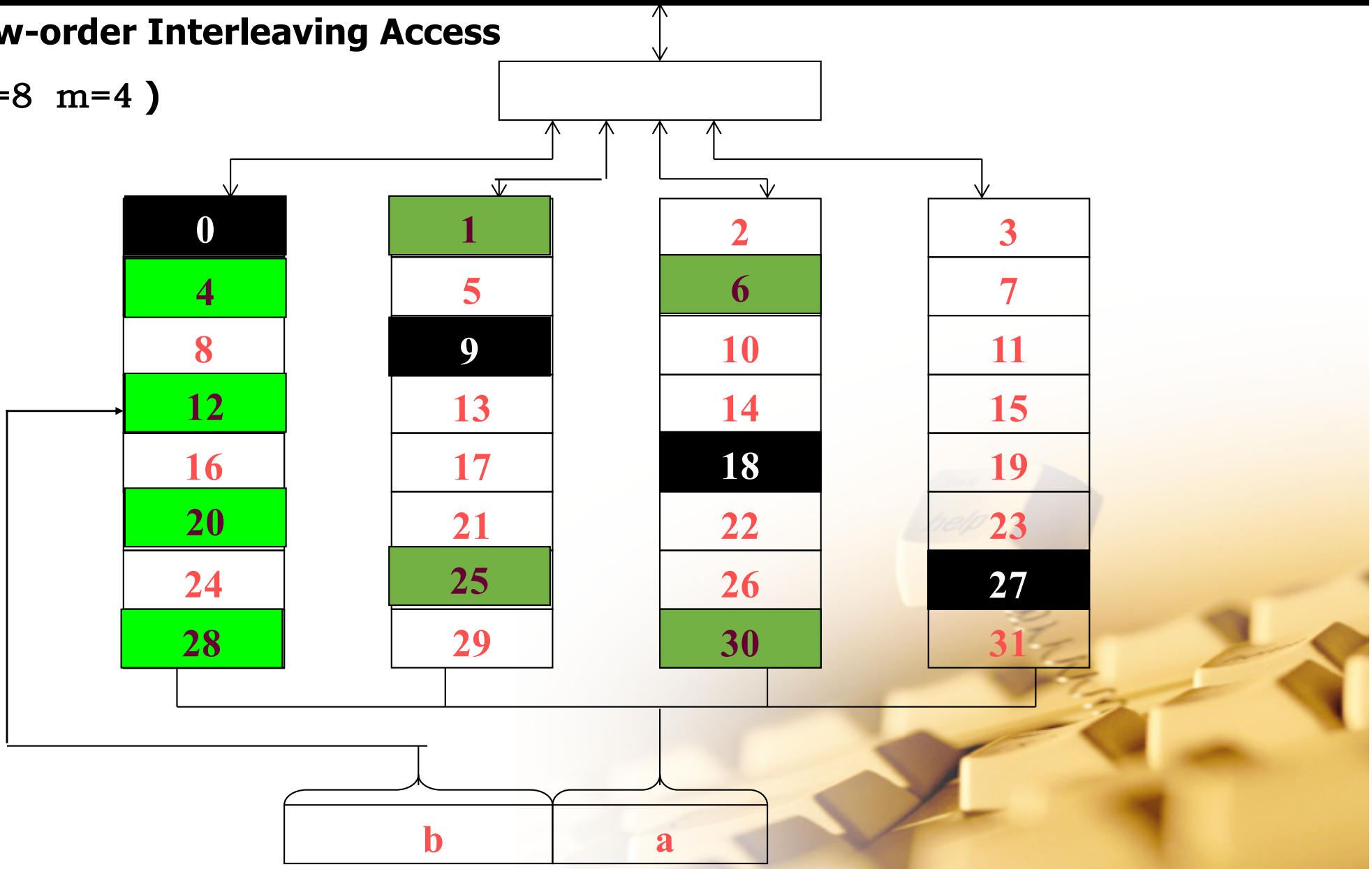
Low-order Interleaving Access Memory



Chapter3 Memory System

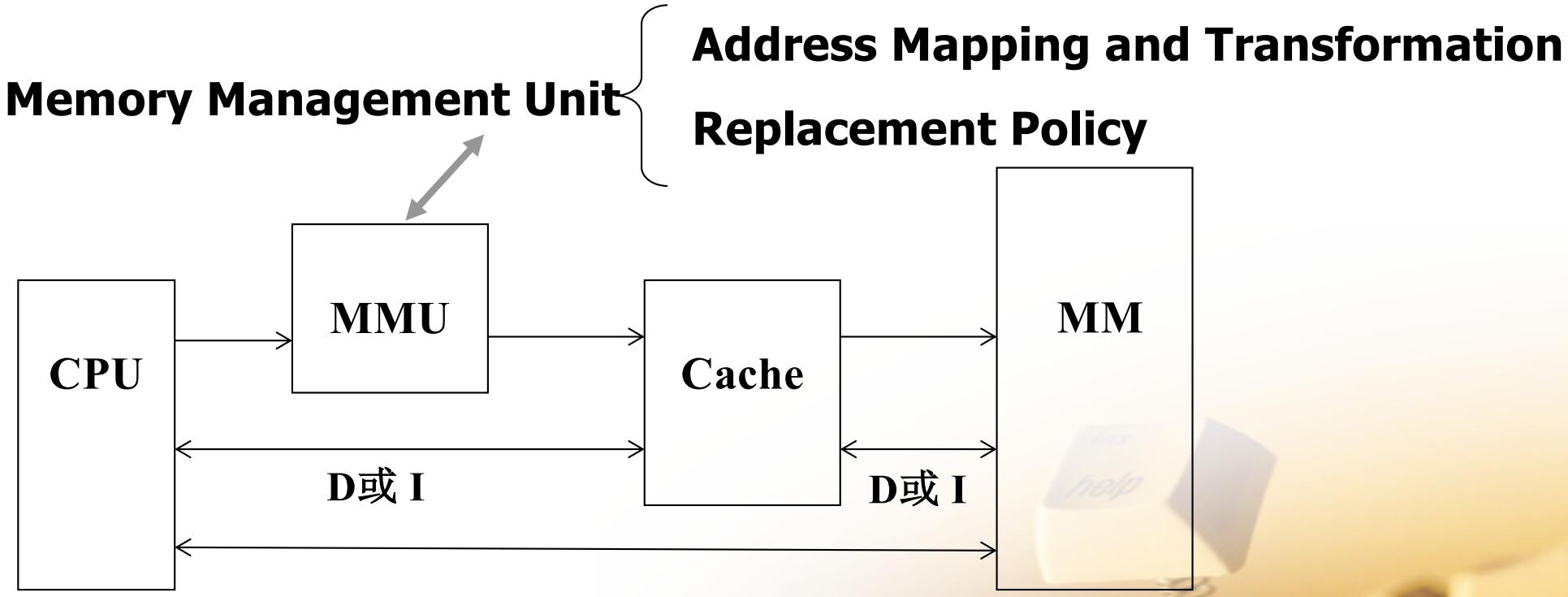
Low-order Interleaving Access

(n=8 m=4)



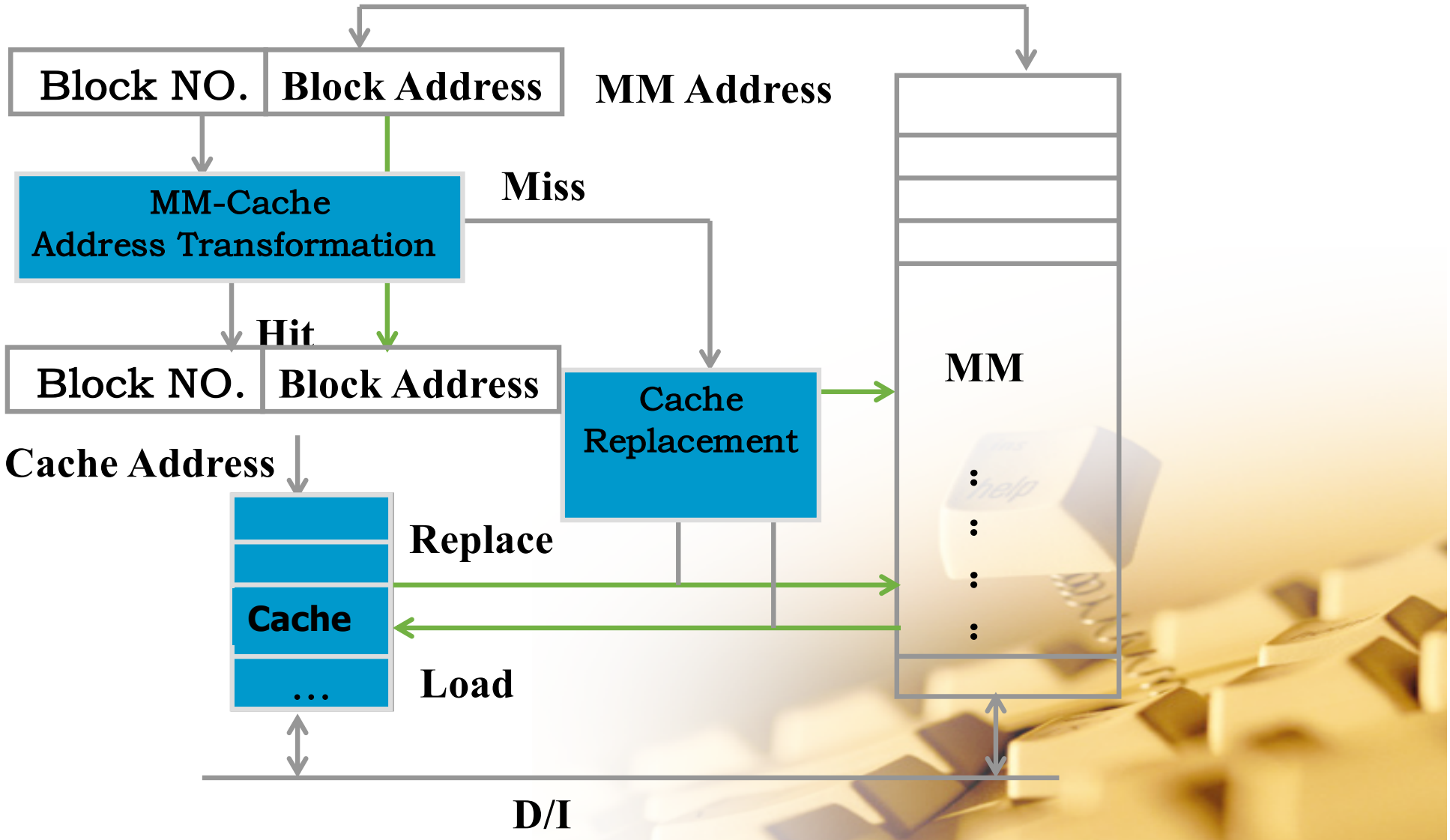
Chapter3 Memory System

Cache



Chapter3 Memory System

Cache Structure



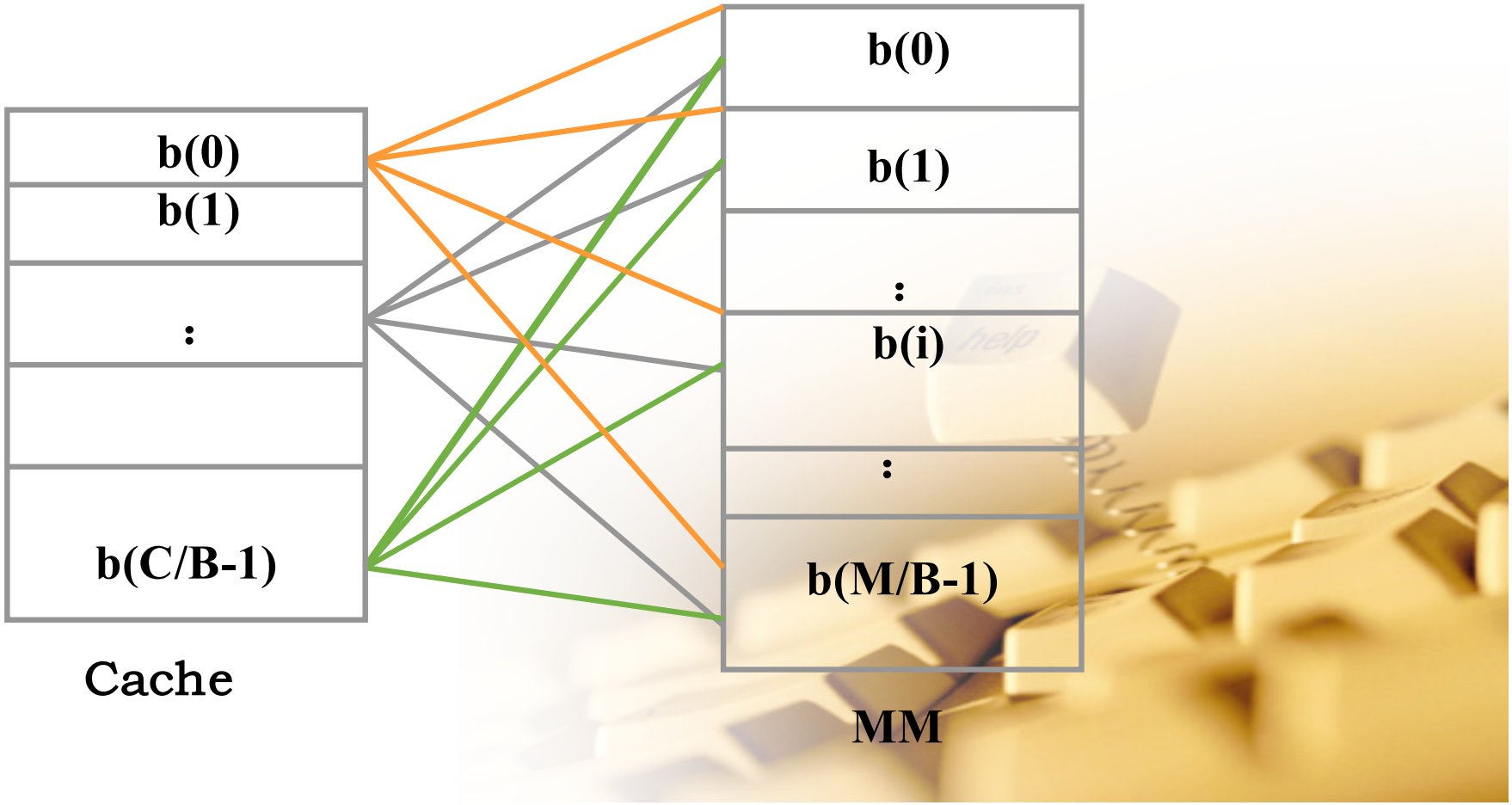
Chapter3 Memory System

Address Mapping and Transformation

- Fully Associative Mapping
- Direct Associative Mapping
- Set Associative Mapping

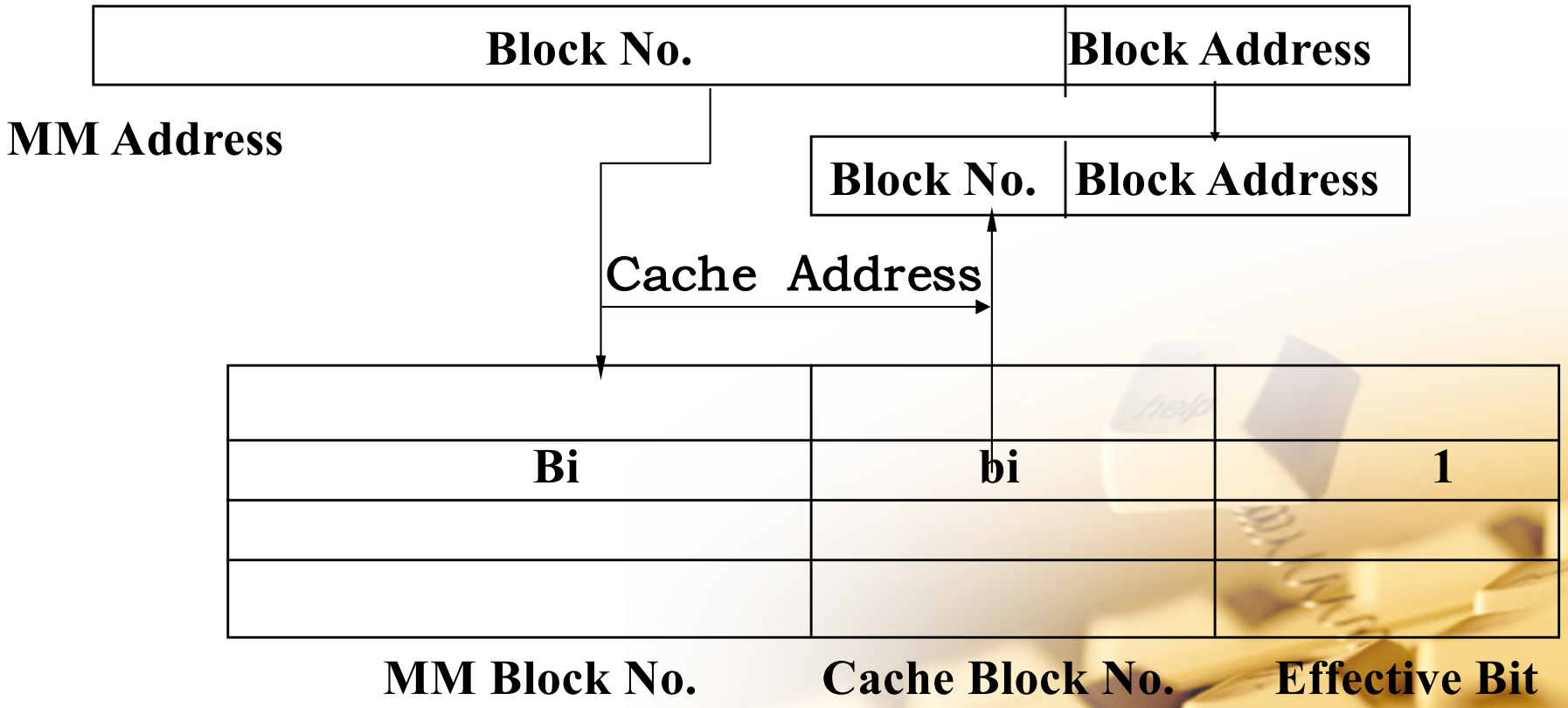
Chapter3 Memory System

- **Fully Associative Mapping**



Chapter3 Memory System

- **Fully Associative Address Transformation**



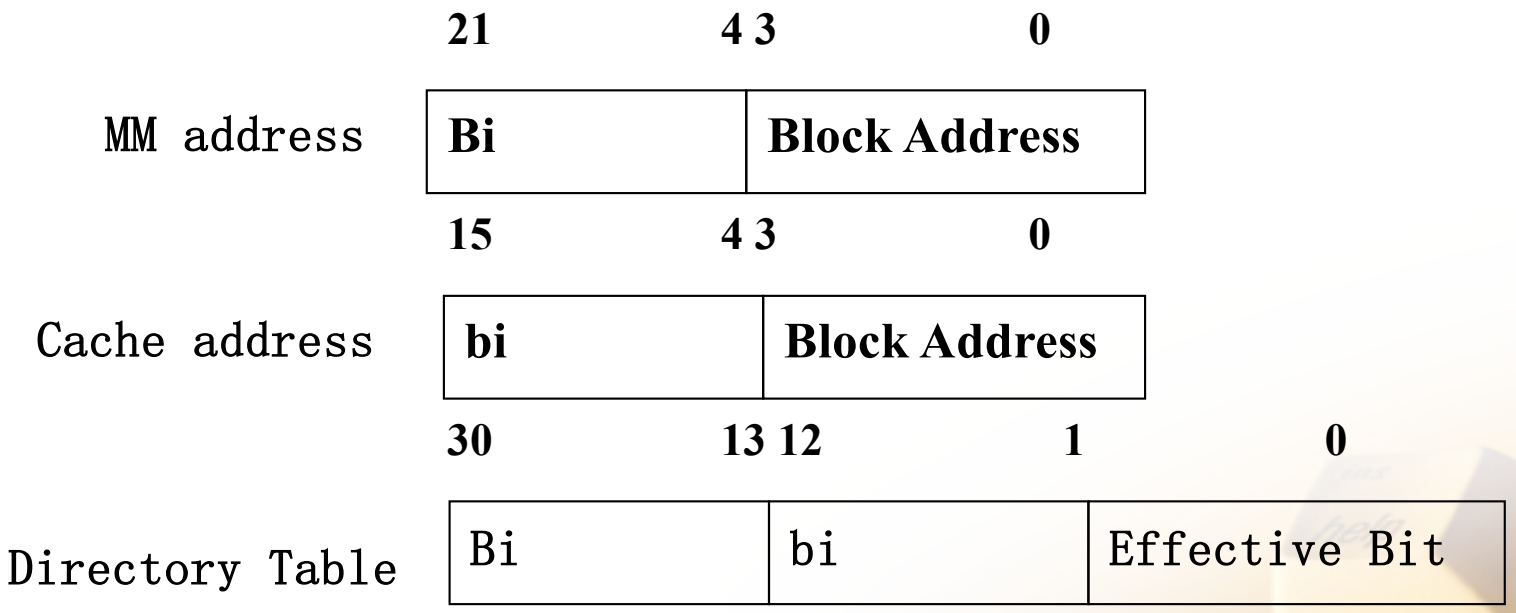
Chapter3 Memory System

Ex1:

Assume that cache capacity is 64KB in a computer system, data block size is 16 bytes, main memory capacity is 4MB, fully associative mapping is used.

Chapter3 Memory System

Solution：The address formats of MM and cache are as below.



Capacity of Directory Table：2^{12}=4096

Chapter3 Memory System

- **Direct Associative Mapping**

b=A mod C/B

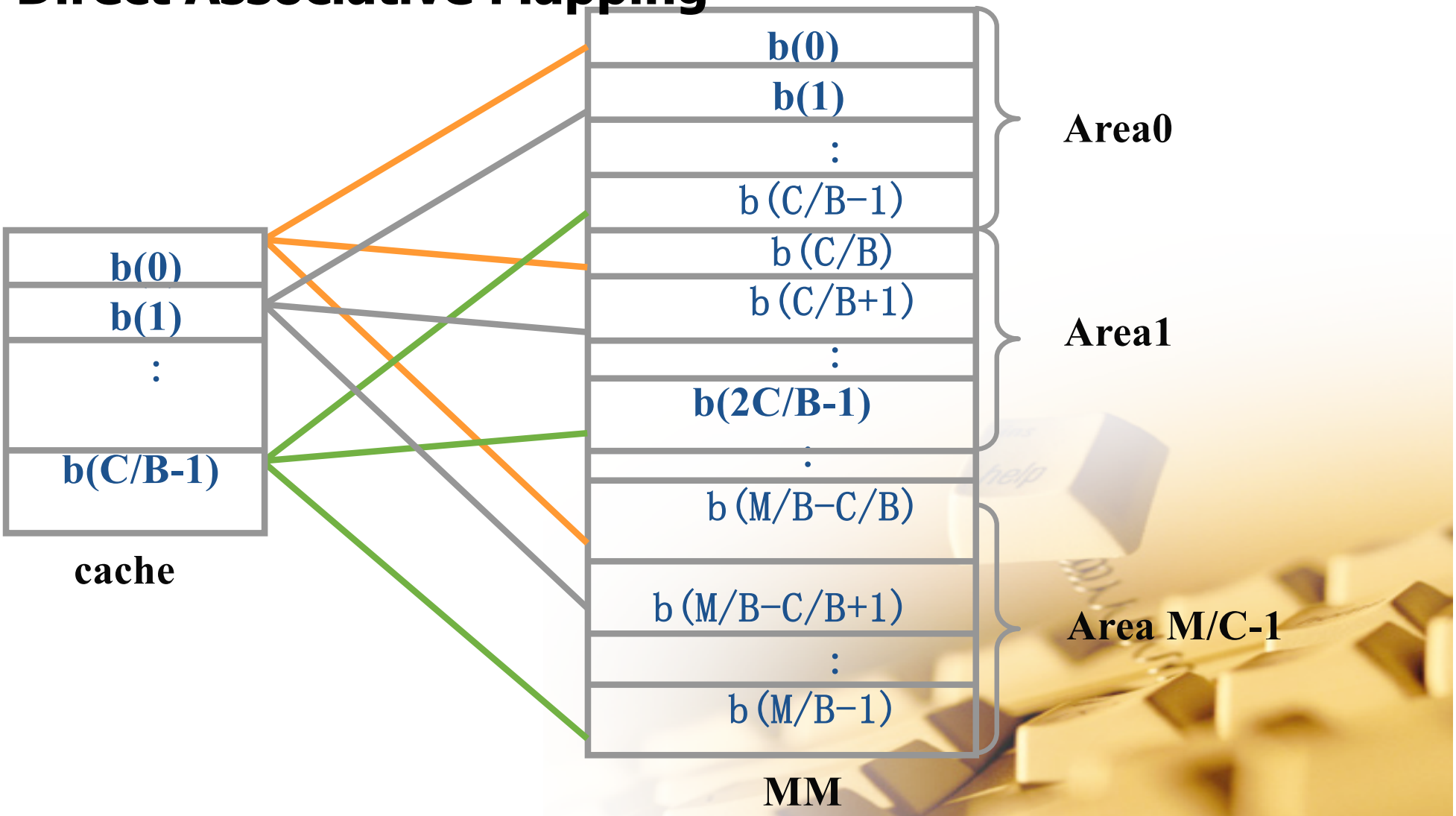
b: Cache Block No.

A: Main Memory No.

C/B: Cache Block Count

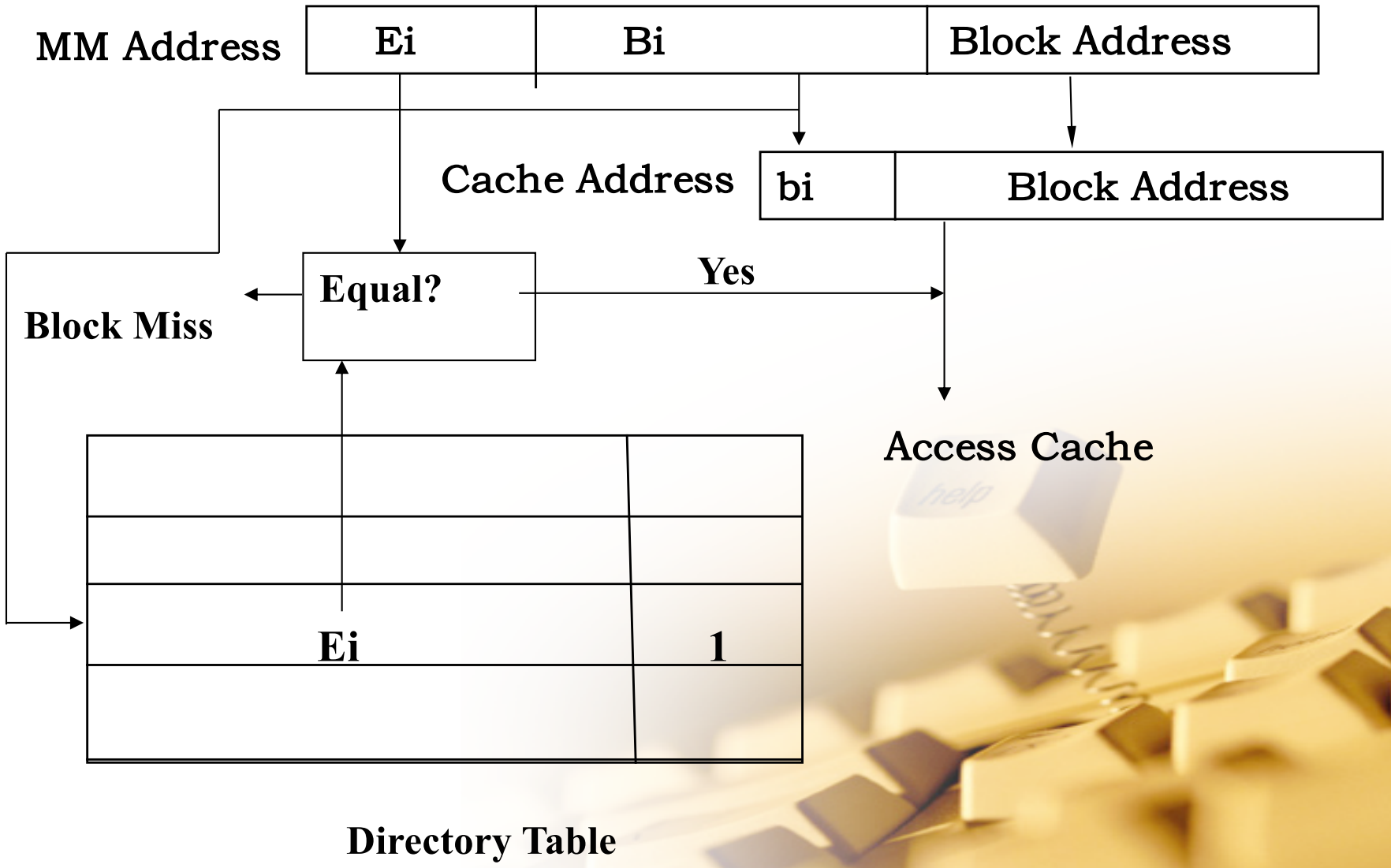
Chapter3 Memory System

- **Direct Associative Mapping**



Chapter3 Memory System

- **Direct Associative Address Transformation**



Chapter3 Memory System

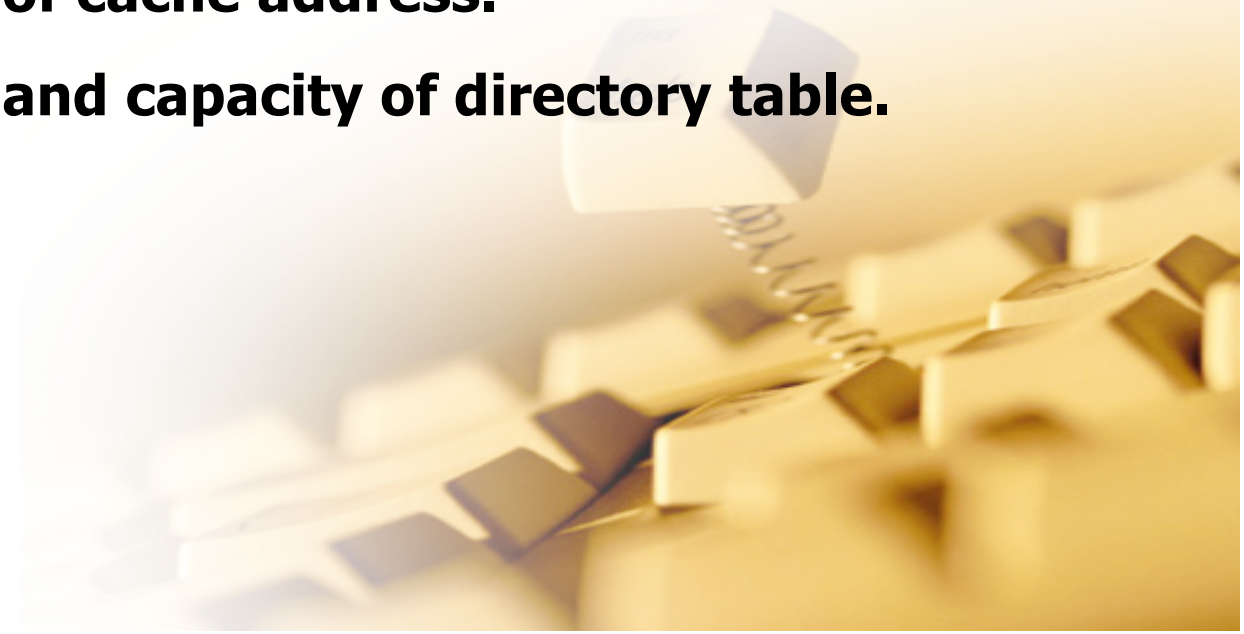
Ex2:

Assume that cache capacity is 64KB in a computer system, data block size is 16 bytes, main memory capacity is 4MB, direct associative mapping is used.

(1) Give out the format of main memory address.

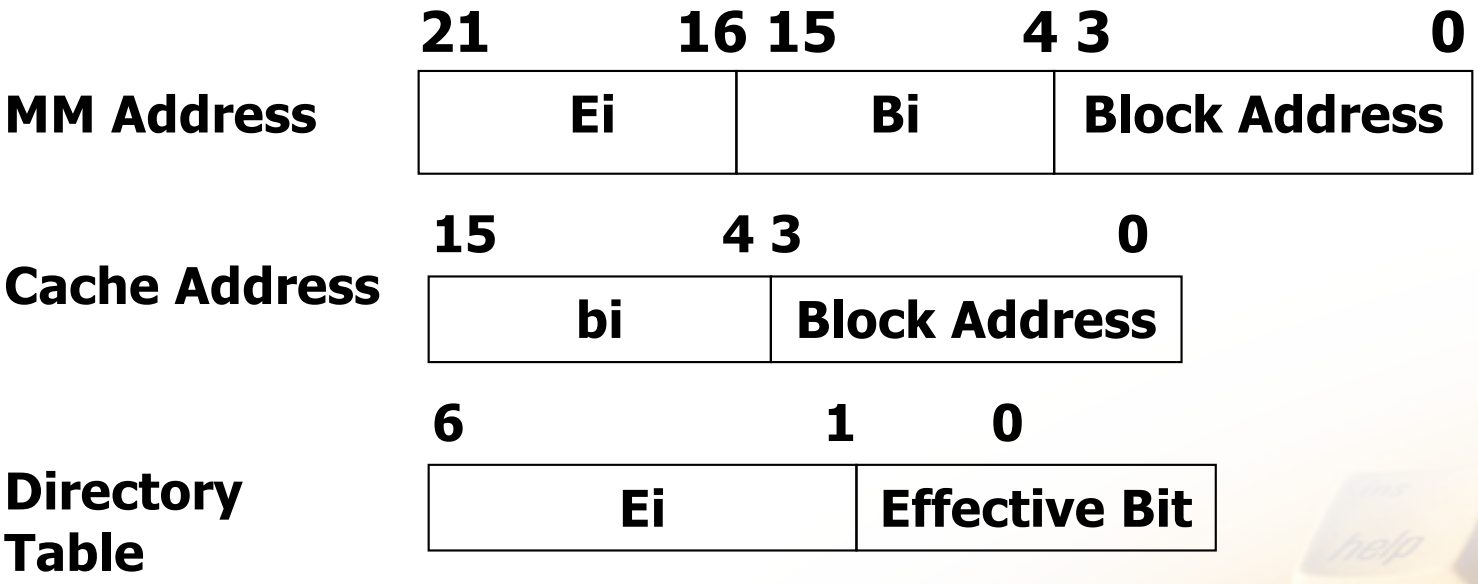
(2) Give out the format of cache address.

(3) Give out the format and capacity of directory table.



Chapter3 Memory System

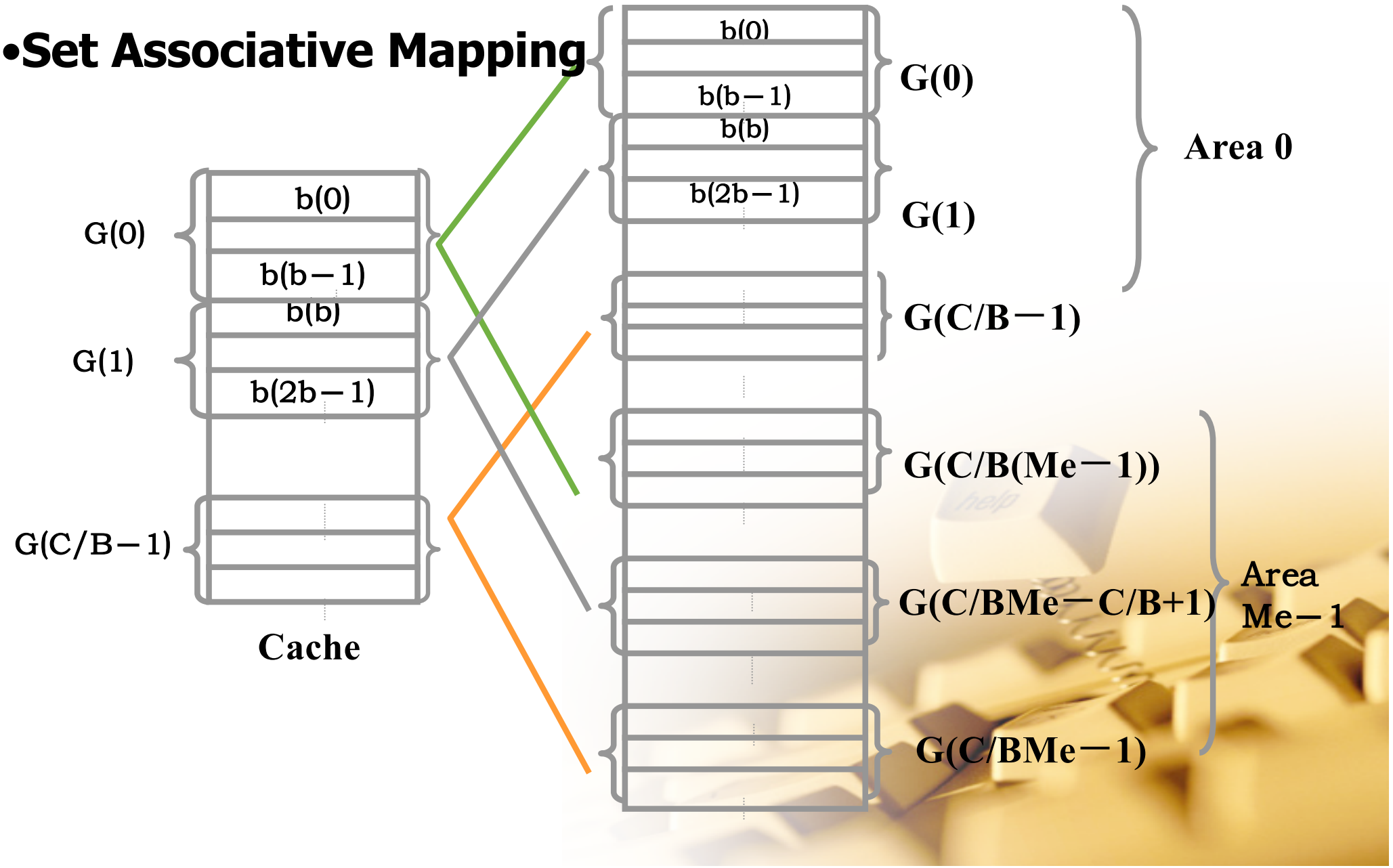
Solution:



Capacity of Directory Table: 2^{12}=4096

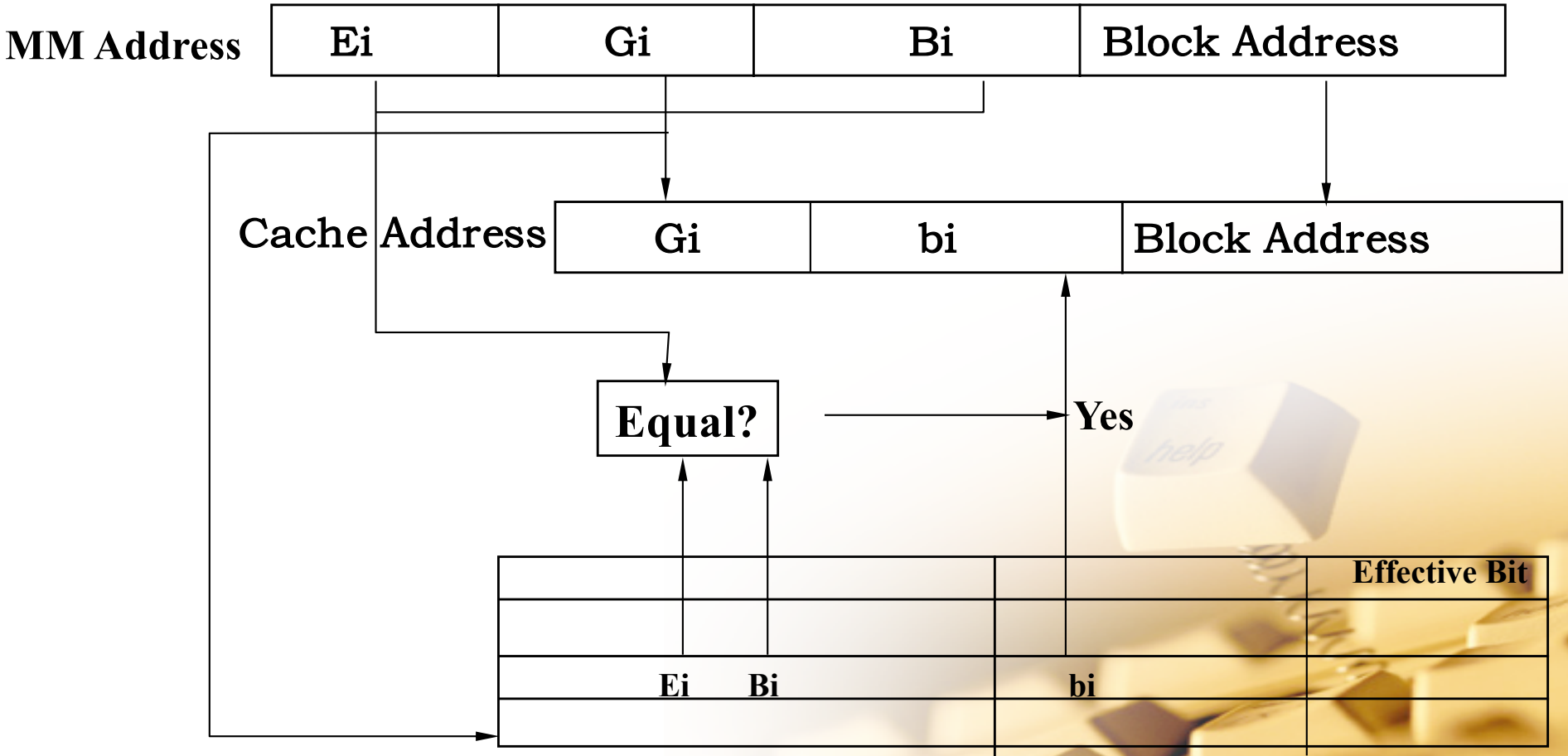
Chapter3 Memory System

- **Set Associative Mapping**



Chapter3 Memory System

- **Set Associative Address Transformation**

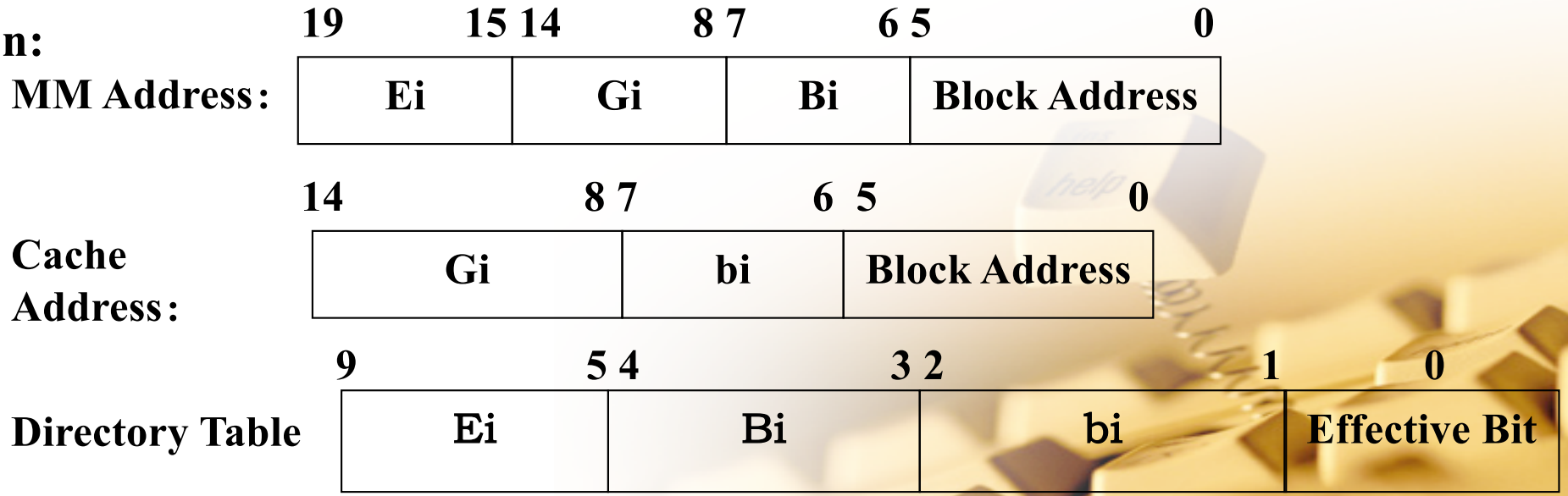


Chapter3 Memory System

Ex3:

Assume that cache capacity is 32KB in a computer system, data block size is 64 bytes, main memory capacity is 1MB, set associative mapping is used and the group count is 128.

Solution:



Capacity of Directory Table: $2^9 = 128 \times 4 = 512$

Chapter3 Memory System

Replacement Policy

- RAND (Random)
- FIFO (First-In First-Out)
- LRU (Least Recently Used)
- LFU (Least Frequently Used)

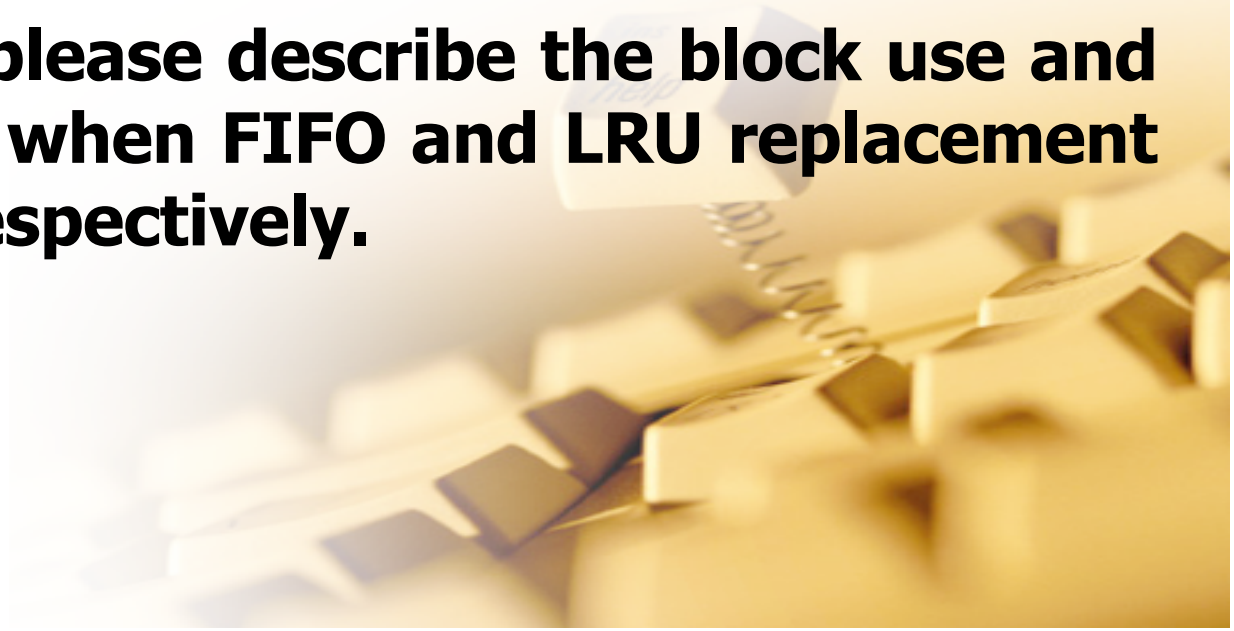
Chapter3 Memory System

Ex4:

Assume that a program uses five pieces of main memory blocks. The block address stream is

2, 3, 2, 1, 5, 2, 4, 5, 3, 2, 5, 2

If cache is 3 pieces, please describe the block use and replacement process when FIFO and LRU replacement algorithm are used respectively.



Chapter3 Memory System

Time	1	2	3	4	5	6	7	8	9	10	11	12
Block Address	2	3	2	1	5	2	4	5	3	2	5	2
FIFO	2	2	2	2*	5	5	5*	5*	3	3	3	3*
		3	3	3	3*	2	2	2	2*	5	5	5
				1	1	1*	4	4	4	4	4*	2
			Hit				Hit		Hit			

Hit

3 times

Chapter3 Memory System

Time	1	2	3	4	5	6	7	8	9	10	11	12
Block Address	2	3	2	1	5	2	4	5	3	2	5	2

LRU

2	2	2	2	2*	2	2	2*	3	3	3*	3*
	3	3	3*	5	5	5*	5	5	5*	5	5
			1	1	1*	4	4	4*	2	2	2
		Hit			Hit		Hit			Hit	Hit

Hit

5 times

Chapter3 Memory System

Implementation of LRU or LFU

✓ Counter

Block Address	4		2		3		5	
	No.	Cn	No.	Cn	No.	Cn	No.	Cn
Cache0	1	10	1	11	1	11	5	00
Cache1	3	01	3	10	3	00	3	01
Cache2	4	00	4	01	4	10	4	11
Cache3		××	2	00	2	01	2	10
Operation	Initial		Load		Hit		Replace	

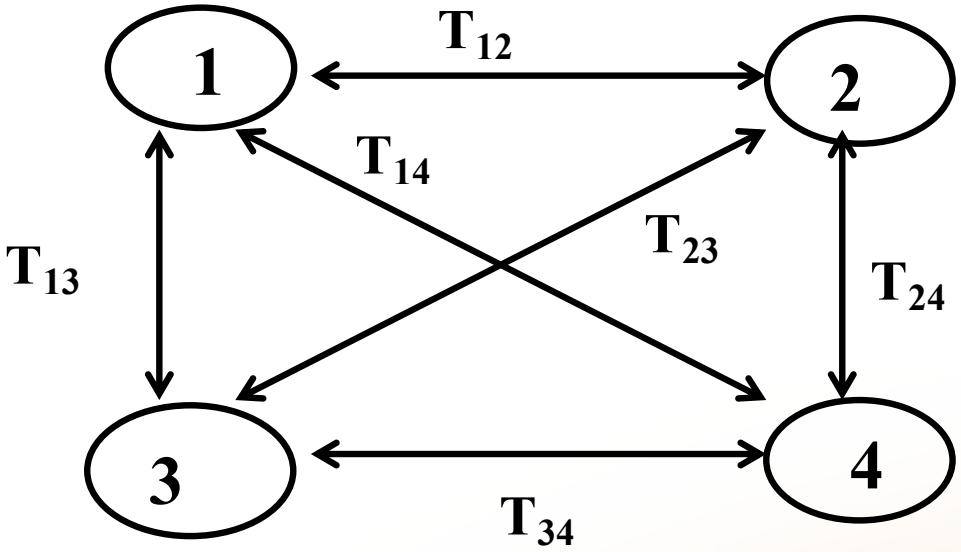
Chapter3 Memory System

- ✓**Register Stack**

Operation	Initial	Load 2	Hit 4	Replace 1
R0	3	2	4	5
R1	4	3	2	4
R2	1	4	3	2
R3		1	1	3

Chapter3 Memory System

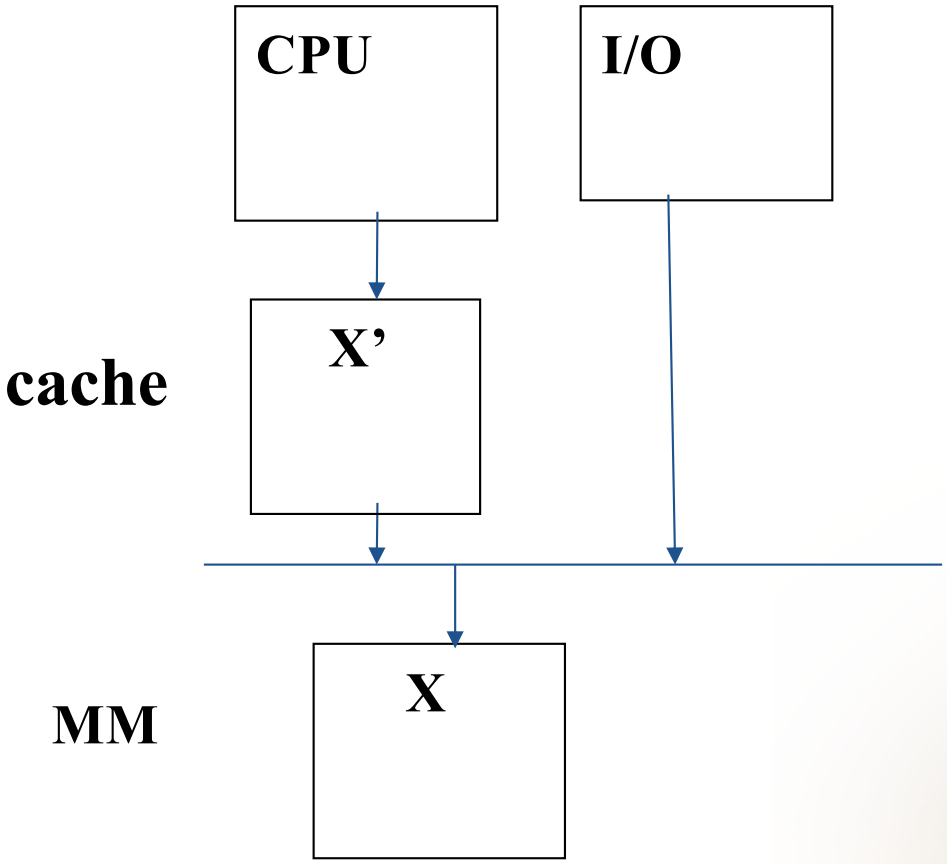
- ✓ **Comparison Pairs**

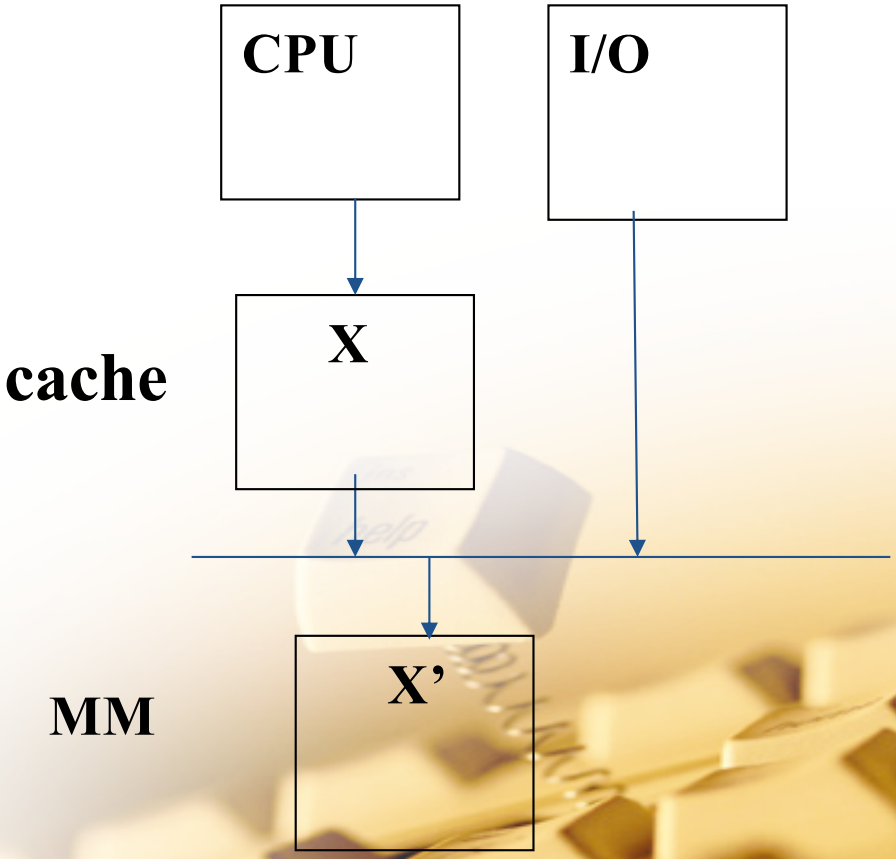


	T_{12}	T_{13}	T_{14}	T_{23}	T_{24}	T_{34}
C1	0	0	0			
C2	1			0	0	
C3		1		1		0
C4			1		1	1

Chapter3 Memory System

Write Operation of Cache





Chapter3 Memory System

No-Write Allocate

(1) WT——Write through：Not only write to cache, and also write to main memory.

Pros: Simple and Good Consistency

Cons: Low Speed

Chapter3 Memory System

Write Allocate

（2）WB——Write back：Only write to cache.

Pros：High Speed

Cons: Poor Reliability

Chapter3 Memory System

(3) Write Miss:

No-Write Allocate

Only write to main memory

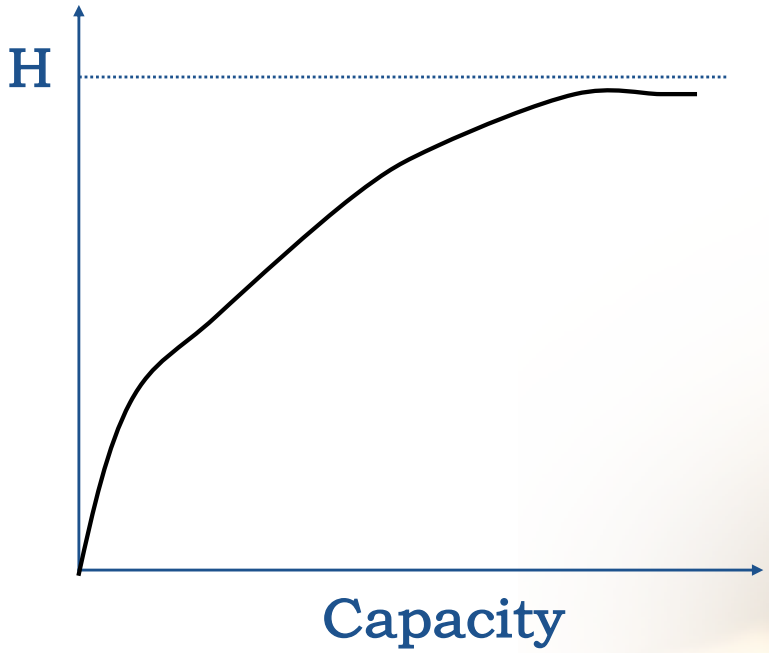
Write Allocate

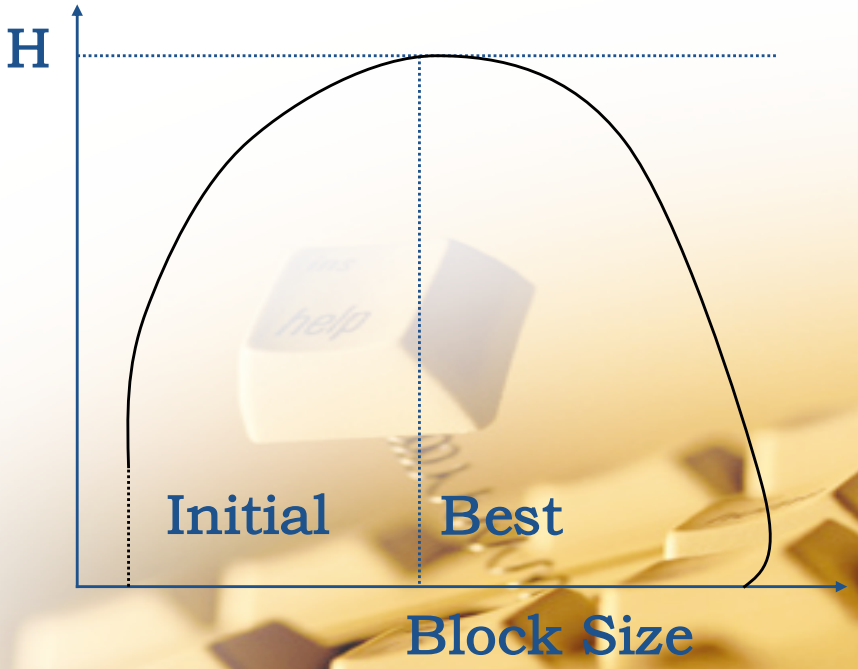
Not only write to main memory, and also write to cache

Chapter3 Memory System

Hit Rate of Cache

1. Cache Capacity
2. Cache Block Size
3. Address Mapping Mode





Chapter3 Memory System

Speedup Ratio of Cache

Average Memory Access Time: T

$$T = H_c T_c + (1 - H_c) T_m$$

Tc： Cache Access Time

Tm： Main Memory Access Time

Hc： Cache Hit Rate

Chapter3 Memory System

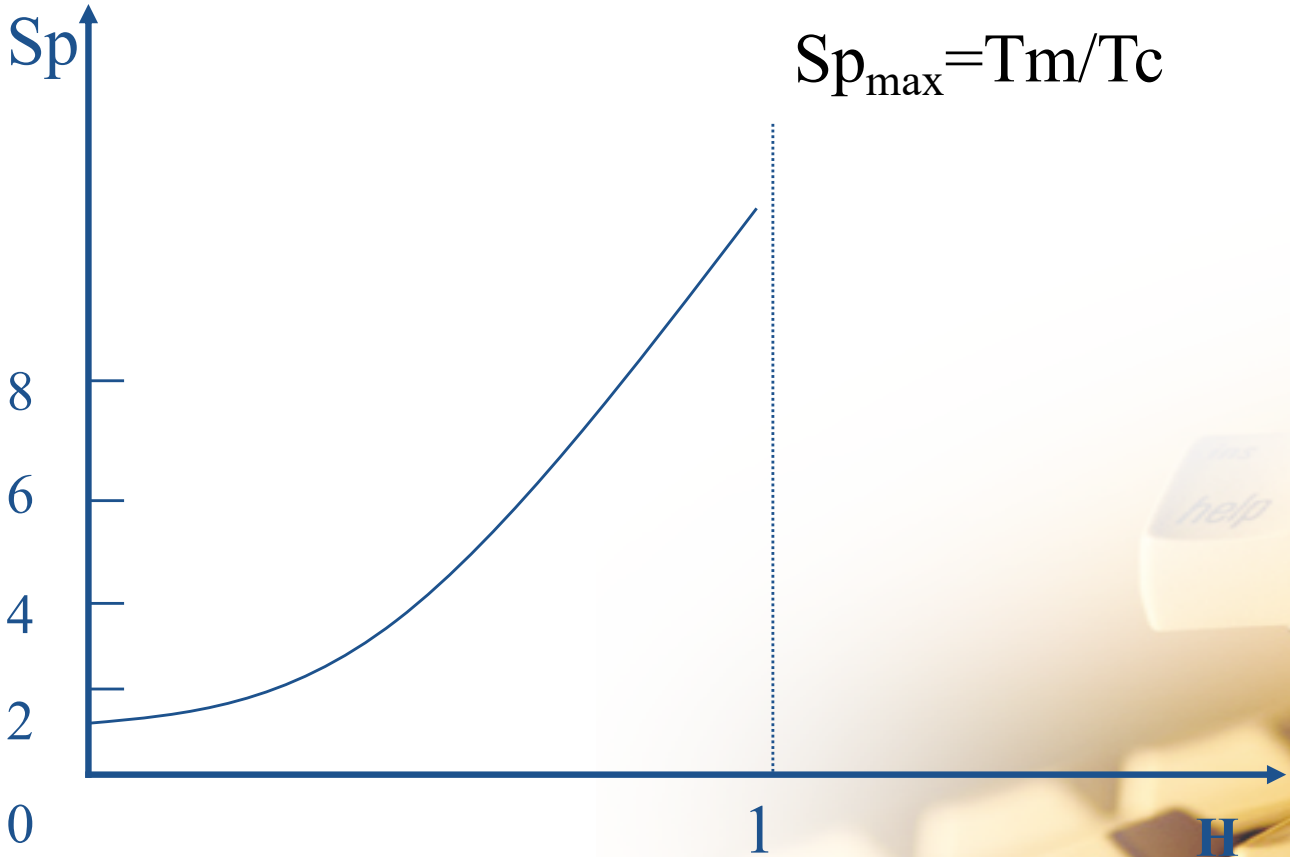
$$S_p = \frac{T_m}{T} = \frac{T_m}{H_c T_c + (1 - H_c) T_m} = \frac{1}{H_c \frac{T_c}{T_m} + (1 - H_c)}$$

When Hc→1,

$$S_p \to \frac{T_m}{T_c}$$

Chapter3 Memory System

The Relation of Cache Sp and Hc



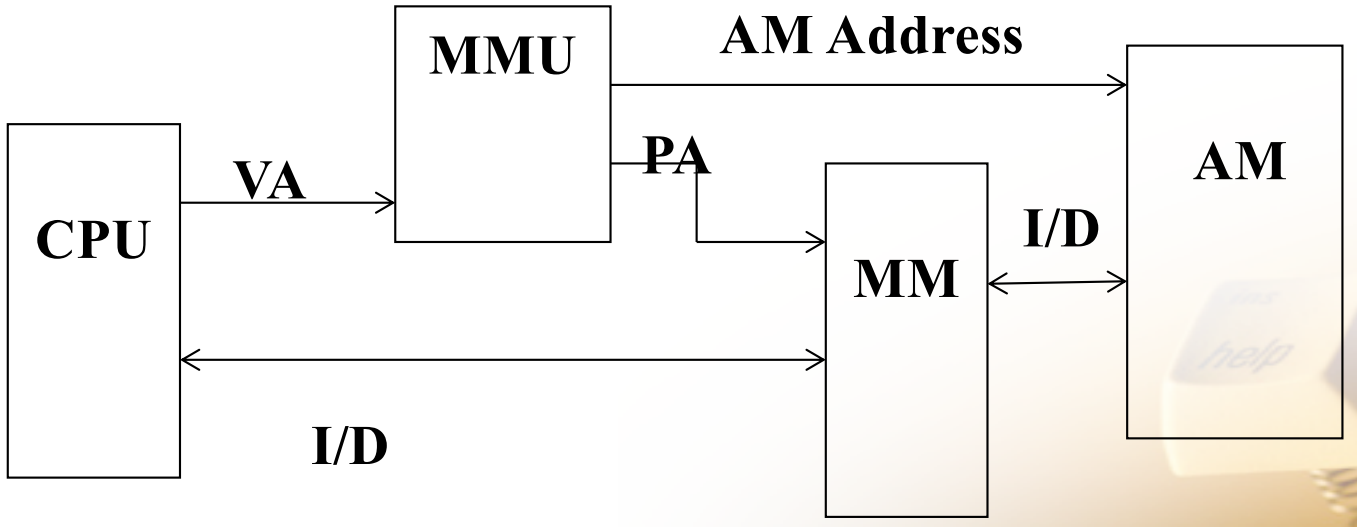
Chapter3 Memory System

Memory Access Efficiency

$$e = \frac{T_c}{T} = \frac{1}{H_c + (1 - H_c)\dfrac{T_m}{T_c}}$$

Chapter3 Memory System

Virtual Memory



Chapter3 Memory System

The Difference of VM and Cache

	Cache	VM
Function	Improve Speed	Expand Capacity
Implement	Hardware	Software
Transparency	Yes	No
Address Transformation	Simple and Fast	Complex and Slow

Chapter3 Memory System

Memory Protection

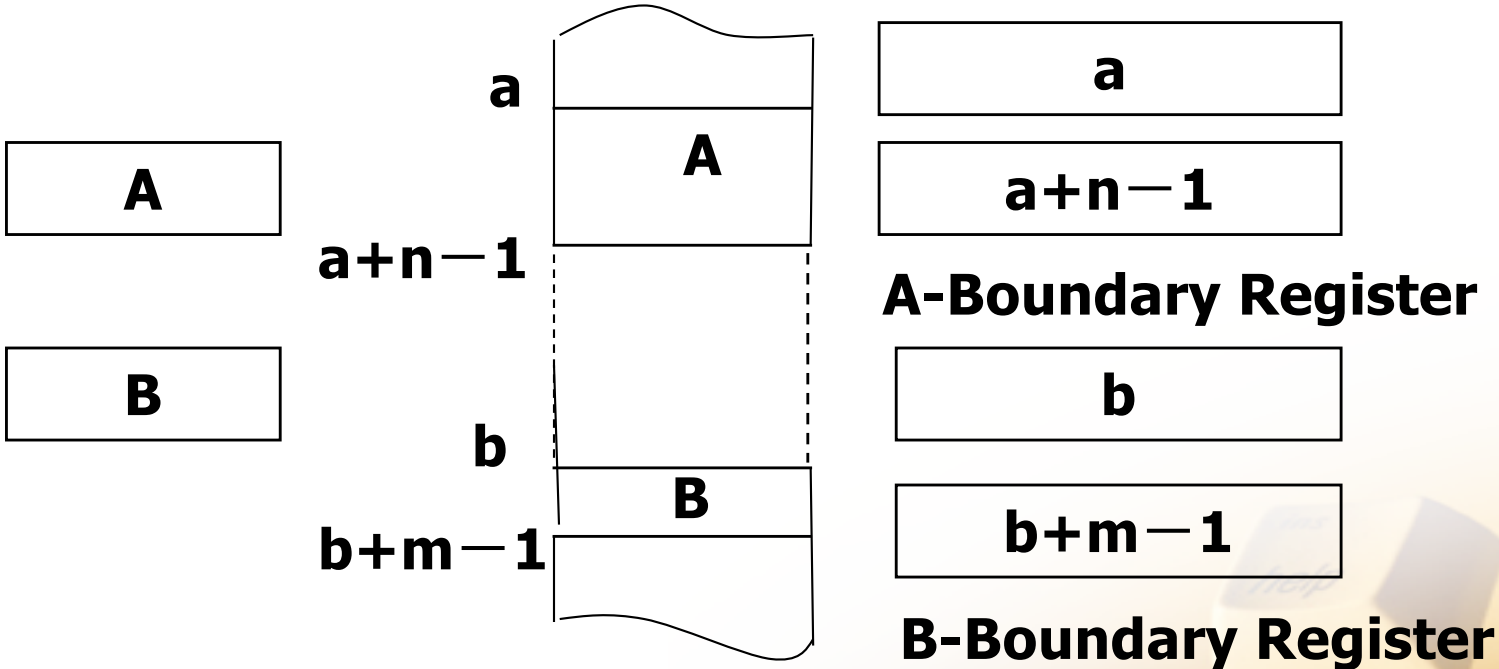
The reason:

(1) In order to prevent damage for other users' program or system software in the main memory due to a user program's error.

(2) In order to prevent that a user program not legally accesses a main memory area which is not be assigned to it, even if it does not cause damage.

Chapter3 Memory System

(1) Boundary Protection



Chapter3 Memory System

(2) Key Protection

Lock：Each memory page has a storage key with a unique key number which is stored in TLB.

Key：Access key which is assigned by OS and stored in PSW.

Process：Compare the key number of the storage key and the access key.

Chapter3 Memory System

(3) Ring Protection

Every program has an unique ring number which indicates that this program can only access the same-level or low-level program.

Chapter3 Memory System