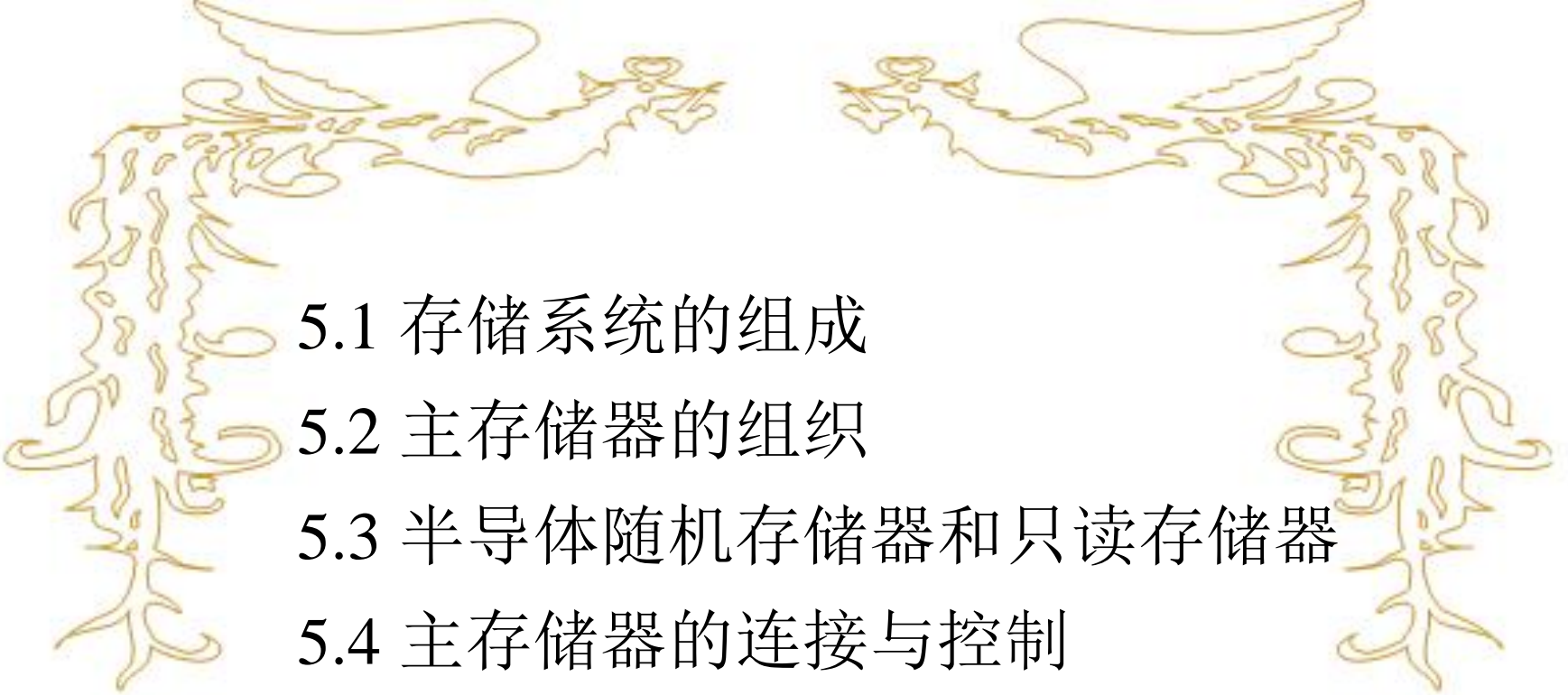





河北师范大学软件学院
Software College of Hebei Normal University

计算机组成原理

第五章 存储系统和结构

- 
- 5.1 存储系统的组成
 - 5.2 主存储器的组织
 - 5.3 半导体随机存储器和只读存储器
 - 5.4 主存储器的连接与控制
 - 5.5 提高主存读写速度的技术
 - 5.6 多体交叉存储技术
 - 5.7 高速缓冲存储器
 - 5.8 虚拟存储器
- 



5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高主存读写速度的技术

5.6 多体交叉存储技术

5.7 高速缓冲存储器

5.8 虚拟存储器



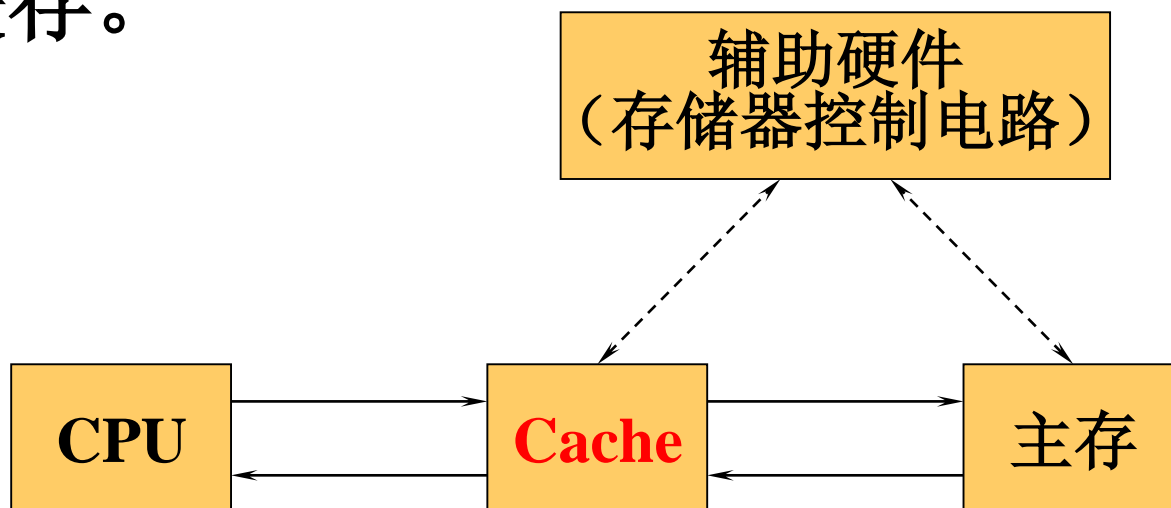
5.7.1 存储系统的层次结构

为了解决容量、速度和价格之间的矛盾，把各种不同存储容量，不同存取速度，不同价格的存储器，按一定的体系结构组织起来，使所存放的程序和数据按层次分布在各存储器中，形成一多层次的存储系统。

5.7.2 高速缓冲存储器

1.Cache—主存层次

在CPU和主存之间设置了一级容量不大，但速度很高的高速缓冲存储器（Cache），简称高速缓存。



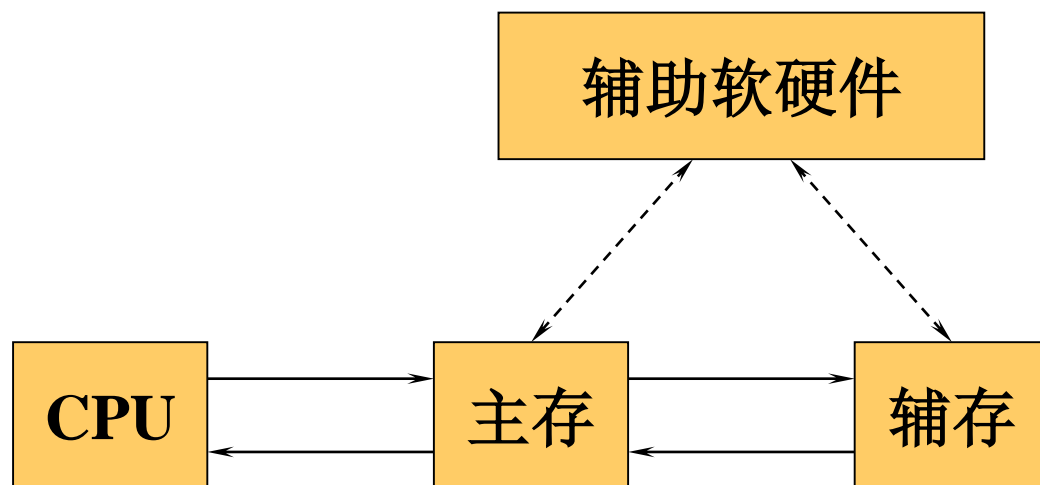
5.7.2 高速缓冲存储器

1. Cache—主存层次

- 从整体看，Cache—主存层次的存取速度接近于Cache的存取速度，但容量是主存的，每位价格接近于主存的每位平均价格。
- 解决了高速度和低成本之间的矛盾。
- 这个层次完全由硬件实现，对用户是透明的。

2.主—辅存层次

主—辅存层次通过附加的硬件及存储管理软件来控制。



高速缓冲存储器的工作原理

1.程序访问的局部性:

- 时间局部性
- 空间局部性

高速缓冲存储器的工作原理

时间局部性：

时间上的局部性指的是当前正在使用的信息很可能是后面立刻还要使用的信息，原因是程序循环（重复执行一段代码）。

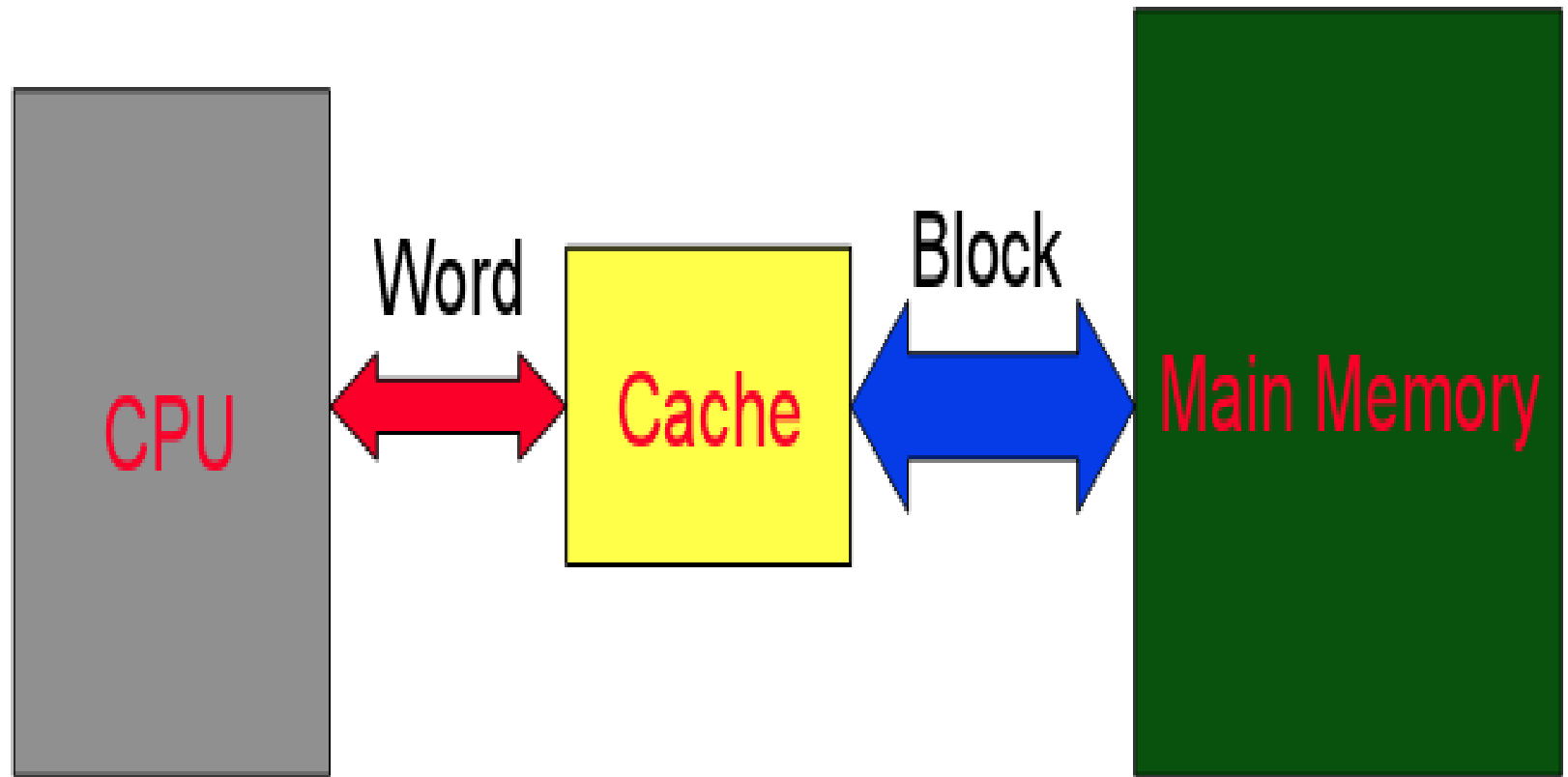
高速缓冲存储器的工作原理

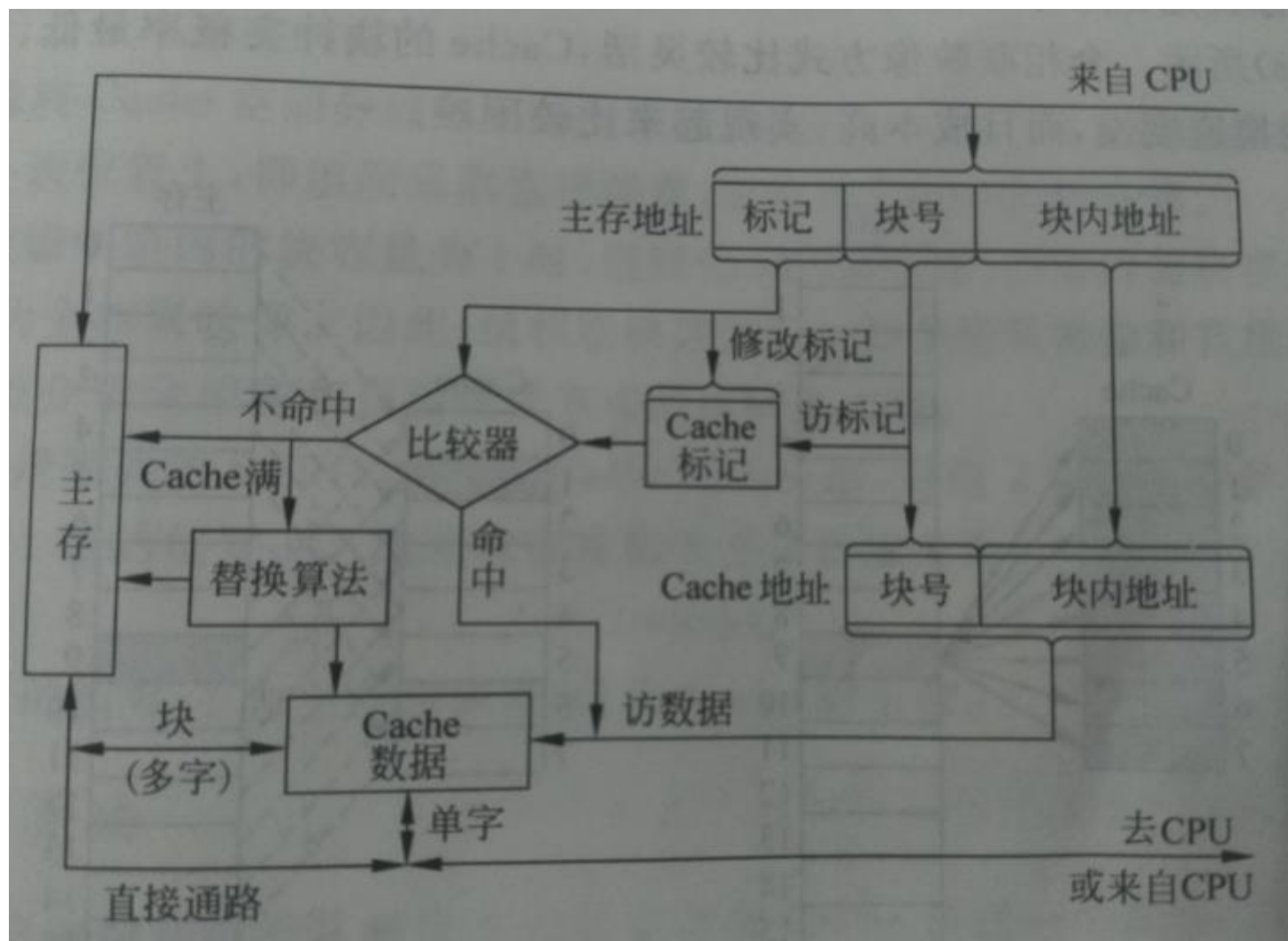
空间局部性：

空间上的局部性是指在一个较短的时间间隔内，如果一个单元被访问，则该单元邻近的单元也可能很快被访问。原因是程序的连续存放以及顺序执行，程序中的数组操作。

2.高速缓存的工作原理:

- **Cache采用SRAM器件，构成小容量高速存储器。**
- **把程序中经常使用的部分存放在Cache中。**
- **使CPU的访存操作大多数在Cache中命中，从而使程序的执行速度大大提高。**





- ▶ 命中率 H 是指对该级存储器来说，要访问的信息正好在这一级中的概率，即命中的访问次数与总访问次数之比。
- ▶ 如何保证较高的命中率？

地址映射规则和替换算法

高速缓冲存储器的地址映像方法：

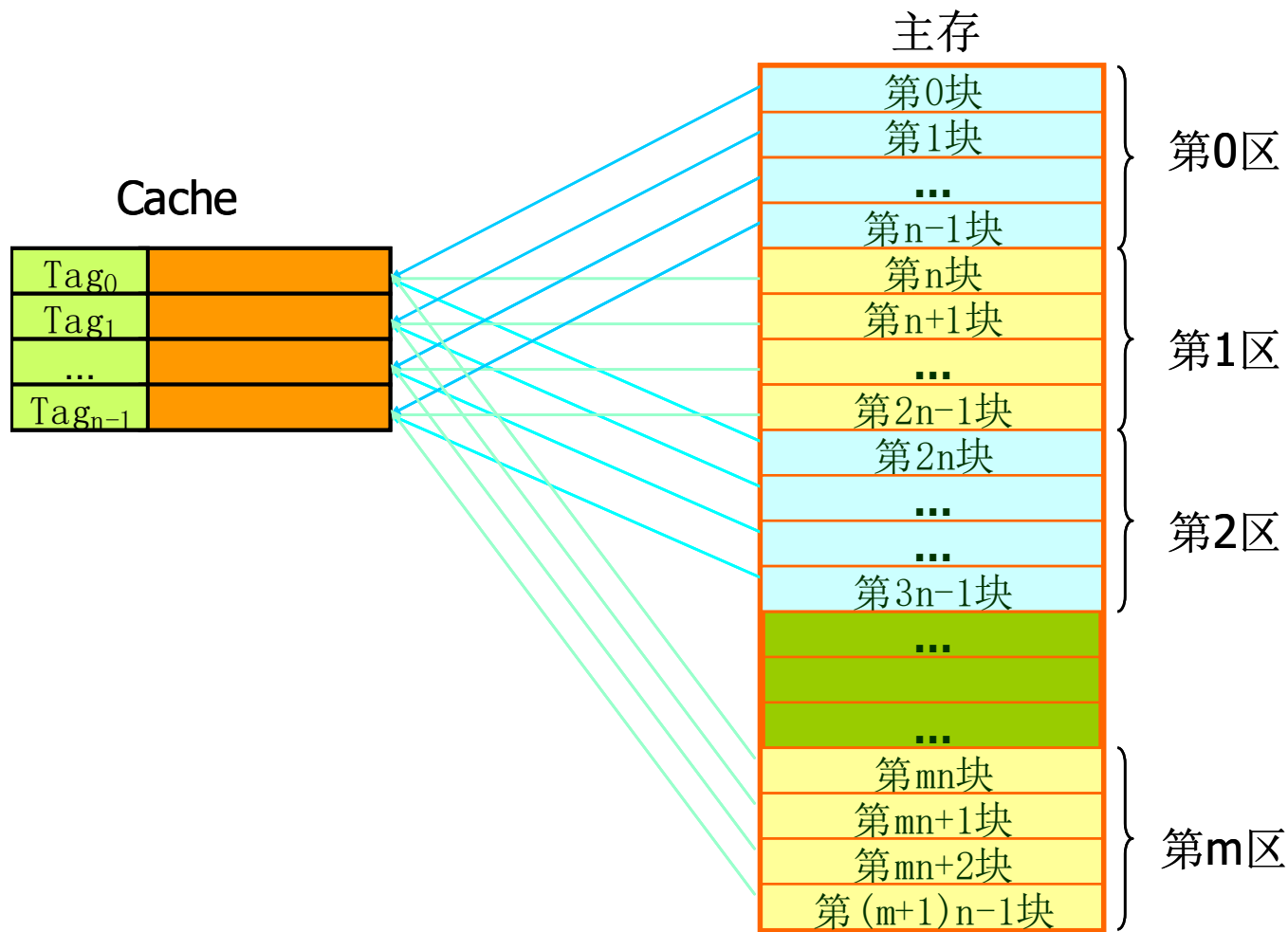
应用某种函数把主存地址映像为Cache地址，称为地址映像。

(1)直接映像

(2)全相联映像

(3)组相联映像

1.直接映射

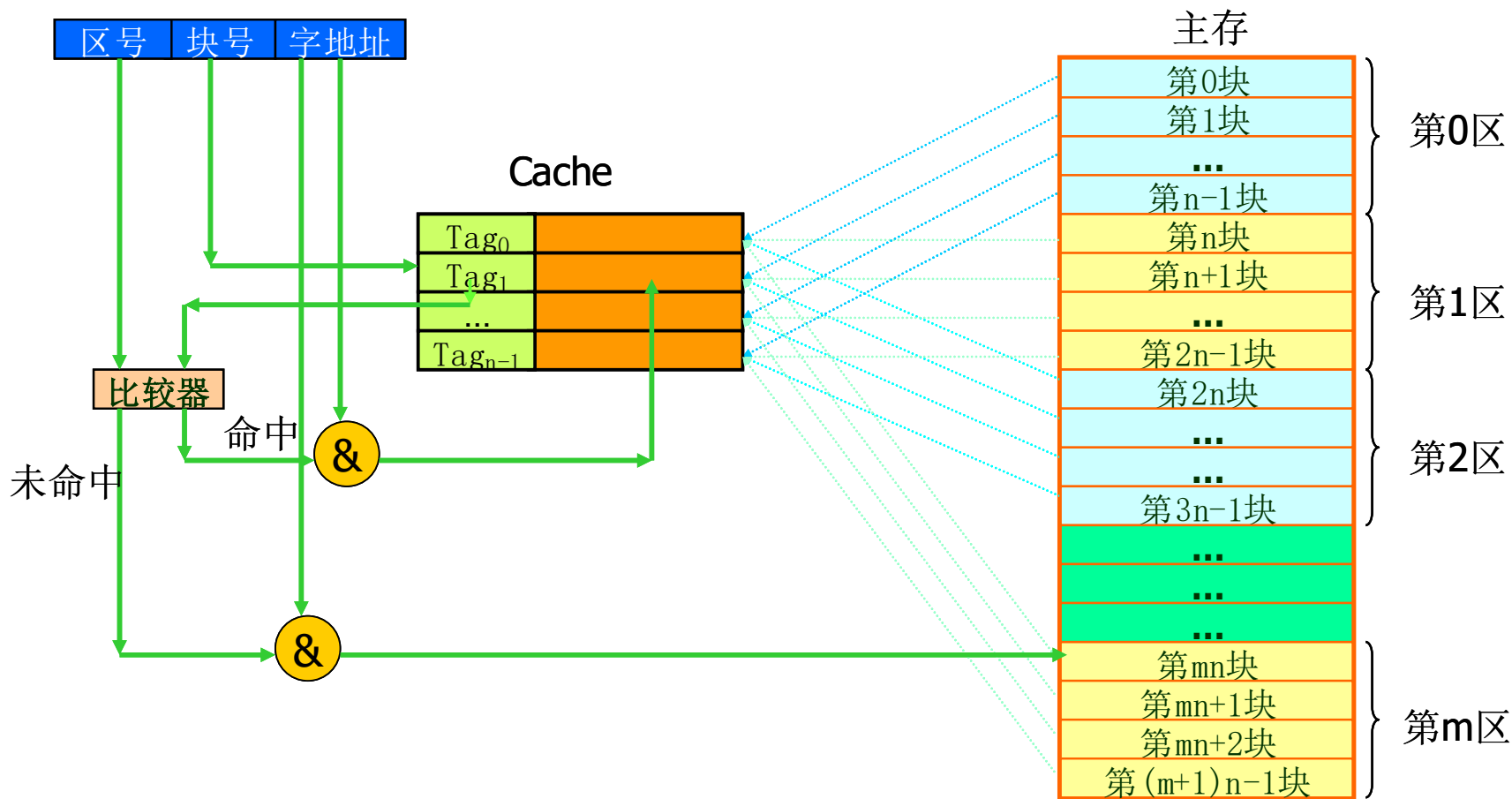


映象规则:

$$K = I \bmod 2^c$$

式中：K为Cache的块号；I为主存的块号；为Cache块数。

Tag内容是映射到该位置的主存块的主存地址中的区号，因为比较时只需确定映射的是哪一个区中的块。



Cache直接映射地址变换过程

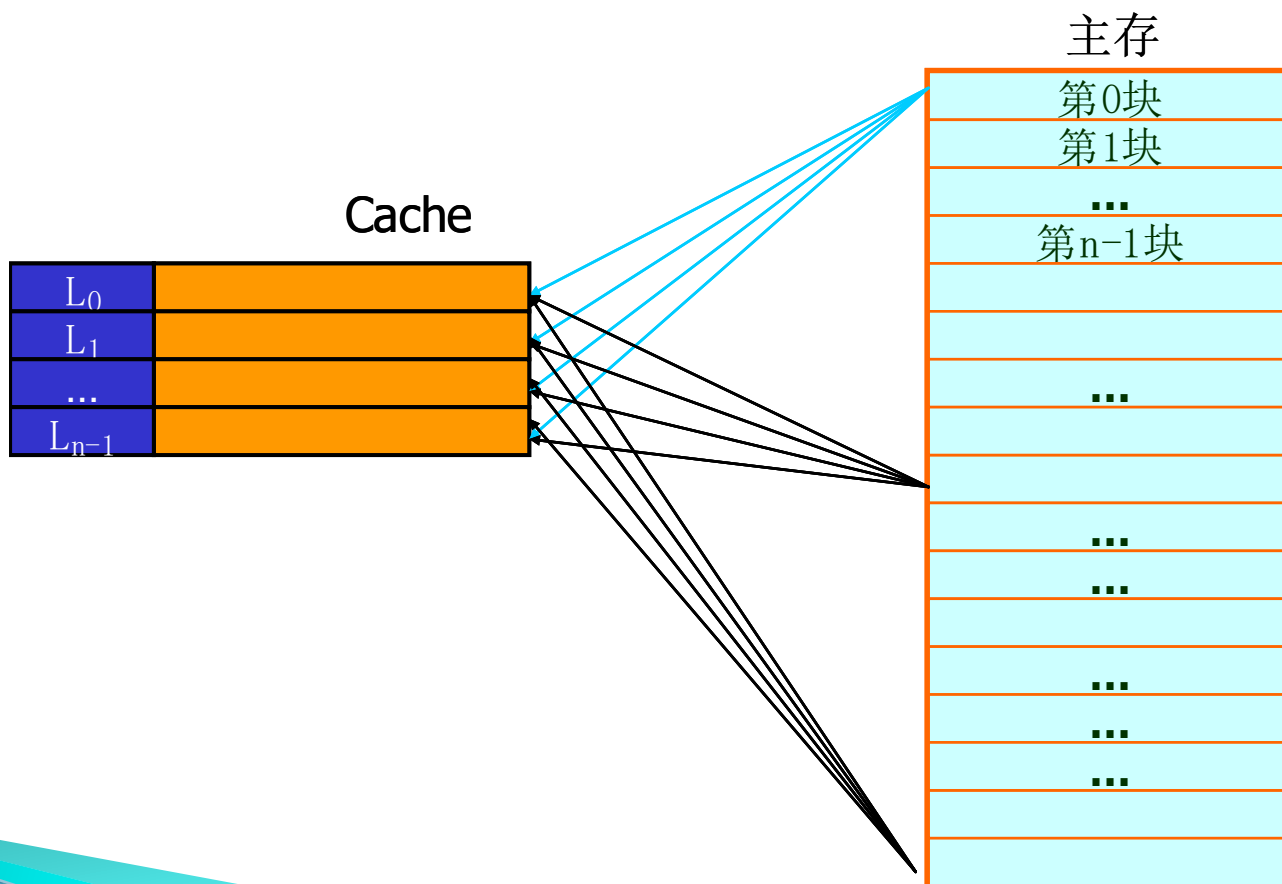
地址转换过程：

用程序中给出的主存地址中的块号找到Cache 中对应的块，读出块的Tag标记与主存地址给出的区号进行比较，按照以下几种情况进行判断：

- 如果与主存地址给出的区号相等，且有效位为1，命中。
- 如果区号相等，有效位为0，失效（作废）。
- 区号不相等，有效位为0，Cache块为空，可以直接装入
- 区号不相等，有效位为1，该块内容有用，写回后，替换

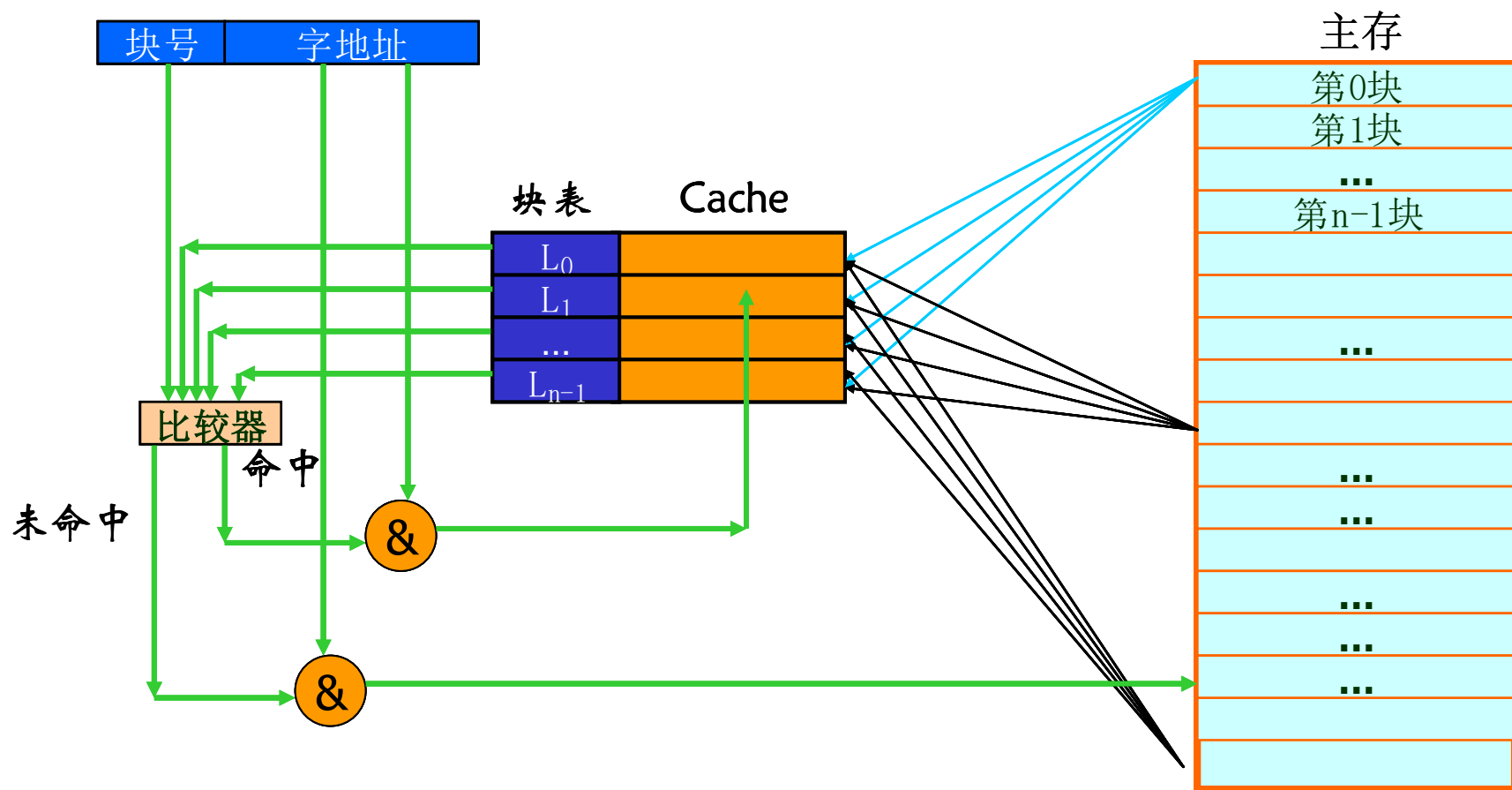
2. 全相联映射

主存中任何一块均可定位于Cache中的任意一块, 可提高命中率, 但是硬件开销增加



Cache的Tag内容:

- ▶ 主存中与该Cache数据块对应的数据块的块地址。
- ▶ 标记位数等于主存块号位数。



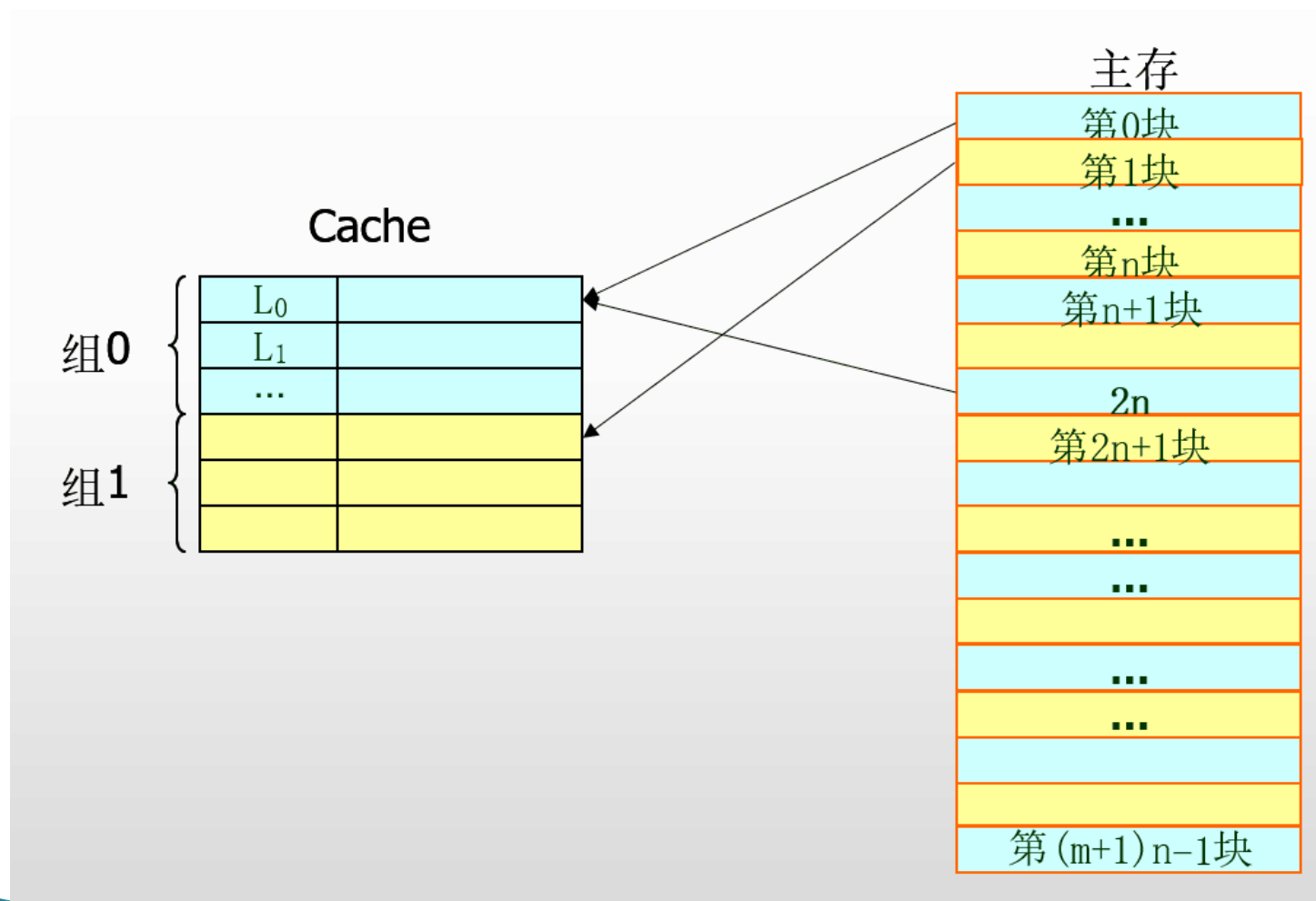
Cache全相联映射地址变换过程

地址转换过程：

因为全相联映射的主存块可能映射在任何Cache块内，所以根据主存块号，与所有Cache块的标记进行比较，有相等的，说明命中，读出Cache块号访问Cache。

需要一个**目录表**来存放映射关系，目录表容量为Cache的块数，字长等于Cache块的标记、Cache块号、有效位之和。

3. 组相联映射



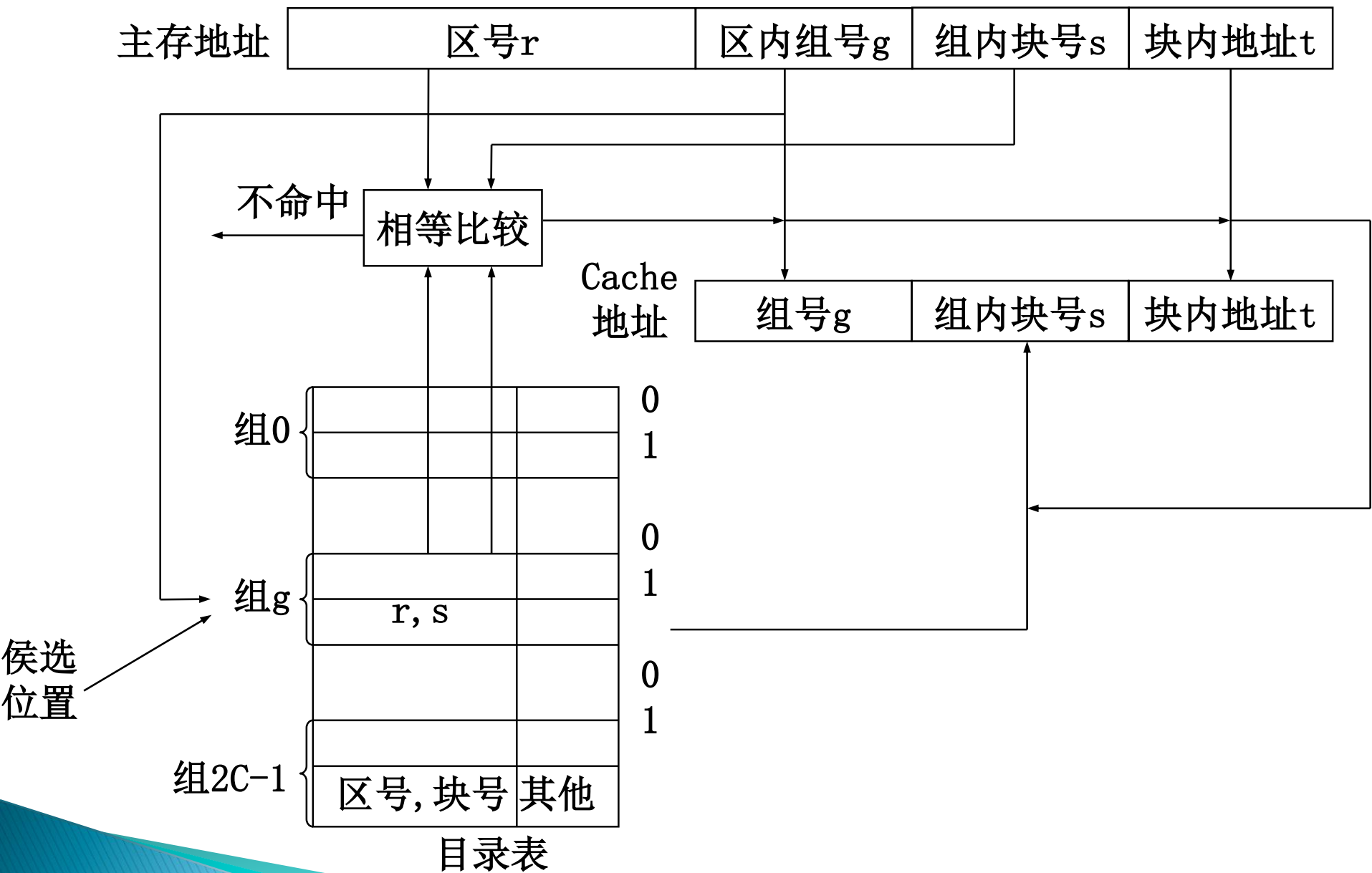
映像规则：

将Cache空间分成大小相同的组，让主存中的一块直接映像装入Cache中对应组的任何一块位置上，即主存的块对应Cache的组，采用直接映像，Cache组内采用全相联映像

$$J = I \bmod Q$$

J为Cache的组号，I为主存的块号，Q为Cache的组数

Tag内容是主存区号E和组内块号B，因为某一个Cache块中的内容可能来自任何一个区，确定了区号后，区内组号是一定的，（因为组到组是直接映像）还需要确定是组内哪一块（因为组内是全相联映像）



每组有2个块的组相联地址变换

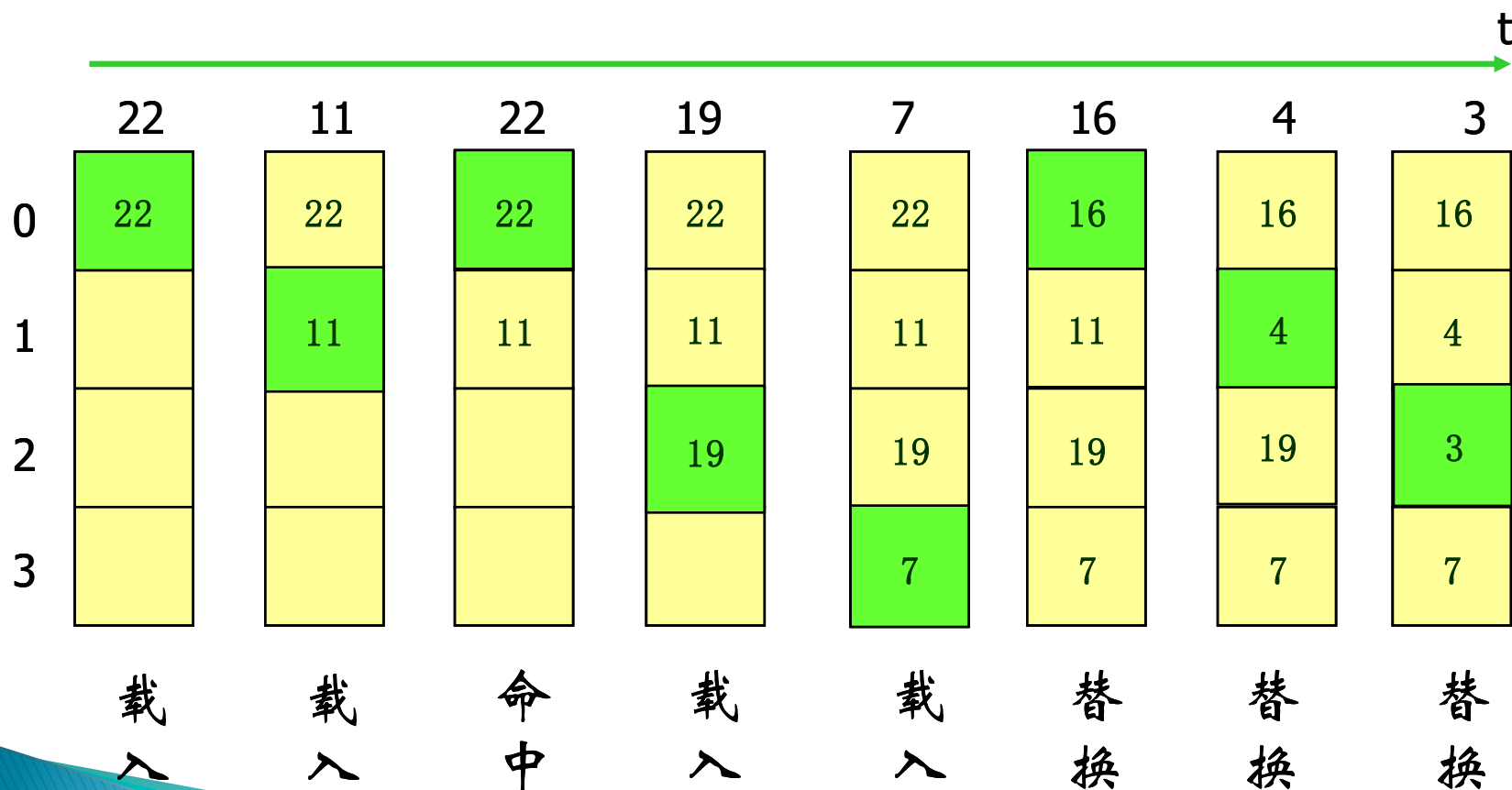
- ▶ 某计算机采用主存—Cache存储层次结构，主存容量有8个块，Cache容量有4个块，采用直接映射方式。若主存块地址流为0,1,2,5,4,6,4,7,1,2,4,1,3,7,2，一开始Cache为空，此期间Cache的命中率为（ ）。
- ▶ A.13.3% B.20% C.26.7% D.33.3%

▶ 某计算机的Cache共有16块，采用二路组相联映射方式（即每组2块）。每个主存块大小为32B，按字节编址。主存129号单元所在主存块应装入到的Cache组号是（ ）（2009年统考真题）。

▶ A.0 B.2 C.4 D.6

高速缓冲存储器的替换算法

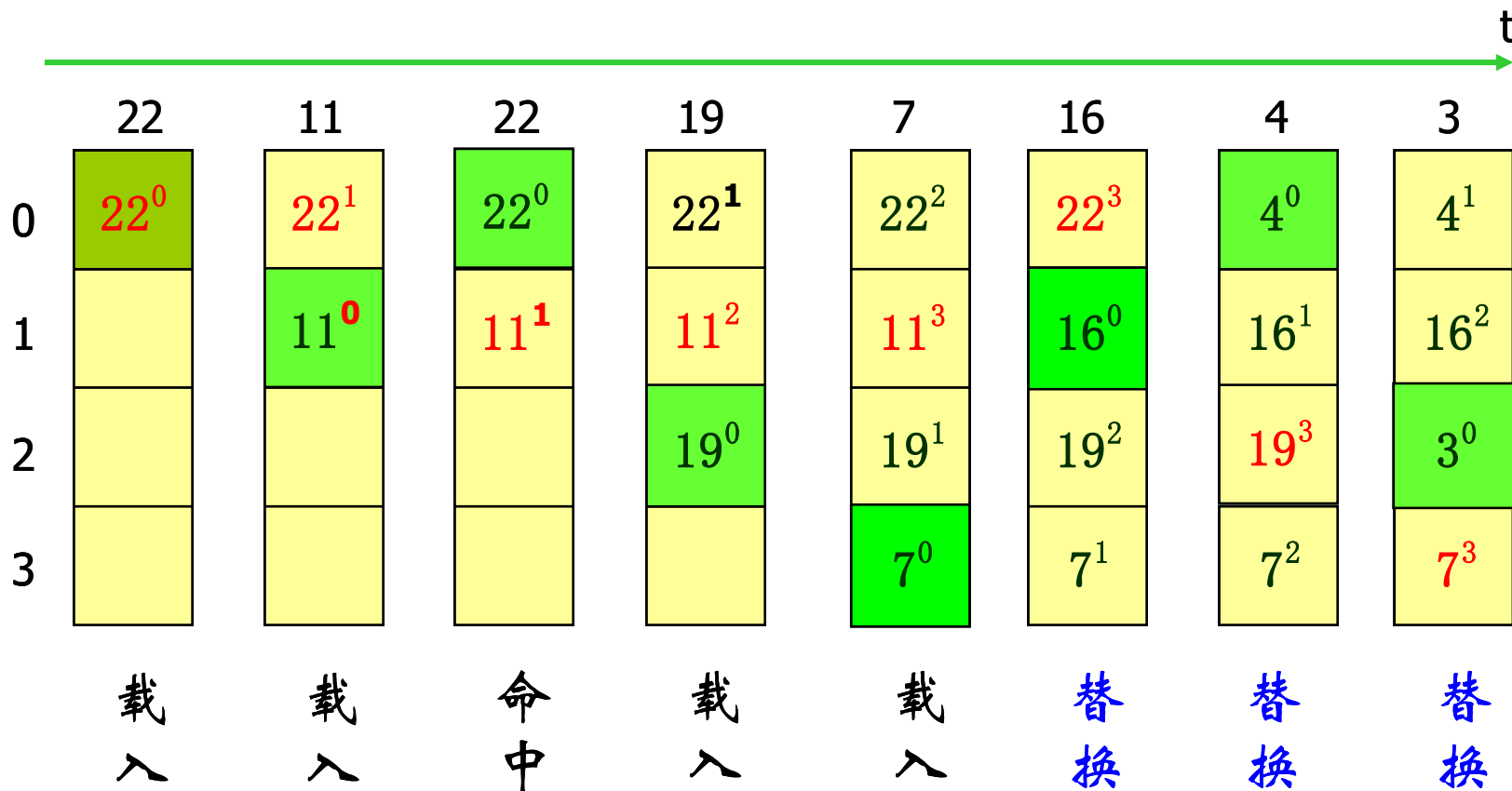
1. 先进先出（FIFO）算法



2.近期最少使用（LRU）算法：

每个Cache单元需要一个计数器跟踪最近访问的数据，每次访问的数据相应的计数器置为0；其余的计数器依次加1。替换最少使用的数据。

LRU算法：



▶ 假设某计算机按字编址，Cache有4个行，Cache和主存之间交换的块大小为1个字。若Cache的内容初始为空，采用二路组相联映射方式和LRU替换算法，当访问的主存地址依次为0,4,8,2,0,6,8,6,4,8时，命中Cache的次数是（ ）（2012年统考真题）

- ▶ A.1 B.2 C.3 D.4

Cache一致性

1.写Cache命中时，如何保持Cache与主存中的内容一致？

(1) 写直达法：每次信息从CPU写入Cache单元中时，也要写回相应的物理内存单元中。

(2) 写回法：即数据暂写入Cache，并用标志将该块注明，等需要将该块替换回到主存时，才写回主存，也称标志交换方式。

层次间应满足的原则

1. **一致性原则**：处在不同层次的同一个信息应保持相同的值。
2. **包含性原则**：处在内层的信息一定被包含在其外层的存储器中，反之则不成立，即内层存储器的全部信息，是其相邻外层信息的一部分的复制品。

2.写Cache不命中时:

(1) 按写分配法

把信息写入主存，同时将该块信息装入Cache。

(2) 不按写分配法

直接更新物理内存中的值，而不把值装入Cache

Cache 性能评价

1、命中(Cache hit)率:

在一个程序执行期间，设 N_c 表示Cache完成存取的总次数， N_m 表示主存完成存取的总次数， h 定义为命中率。则有：

$$h = \frac{N_c}{N_c + N_m}$$

命中率跟程序、Cache容量、组织方式、块的大小有关。

2、存储系统平均存取时间

Cache存取时间为 t_c ，命中率为 h ，主存的存取时间为 t_m 则系统平均存取时间为：

$$t_a = h * t_c + (1 - h) * t_m$$

3、存储系统的访问效率

$$e = \frac{t_c}{t_a}$$

例：某计算机系统的内存存储系统是由Cache和主存构成，Cache的存取周期是45ns，主存的存取周期是200ns，已知在一段给定的时间内，CPU共访问存储系统2000次，其中访问主存100次，问：

- (1) Cache的命中率是多少？**
- (2) CPU访问该内存存储系统的平均时间？**
- (3) Cache--主存的效率是多少？**

$$(1) \quad h = N_c / (N_c + N_m) = (2000 - 100) / 2000 = 0.95$$

$$(2) \quad t_a = h * t_c + (1 - h) * t_m \\ = 0.95 * 45 + 0.05 * 200 = 60\text{ns}$$

$$(3) \quad e = t_c / t_a = 45 / 60 = 75\%$$

例：Cache的存取周期是40ns，主存的存取周期是200ns，Cache /主存系统平均访问时间为50ns，求Cache的命中率？

$$50 = h * 40 + (1 - h) * 200$$

$$h = 93.8\%$$

5.7.3 虚拟存储器

虚拟存储器的基本概念

虚拟存储器是一种将大的逻辑空间映射到比它小得多的物理空间的机制。由CPU中的存储管理模块和操作系统中的相应模块共同支持

例题及习题

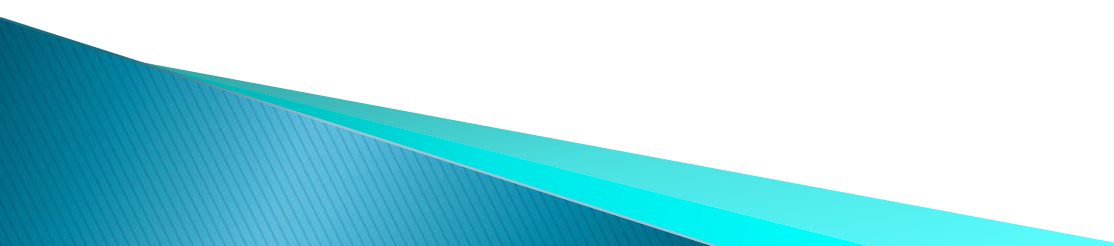
- ▶ 1、某计算机字长32位，其存储容量为64MB，若按字编址，它的存储系统的地址线至少需要（ ）条。
- ▶ 2、对存储器的要求是容量大、速度快、成本低，为了解决这三方面的矛盾，计算机采用多级存储体系结构，即（ ）、（ ）、（ ）。
- ▶ 3、主存储器的技术指标有（ ），（ ），（ ），（ ）。

例题及习题

- ▶ 4、Cache和主存构成了（ ），全由（ ）来实现。
- ▶ 5、虚拟存储器分为页式、（ ）式、（ ）式三种。

- ▶ 6、EEPROM是指（ ）。
- ▶ A 读写存储器 B 只读存储器
- ▶ C 闪速存储器 D 电擦除可编程只读存储器
- ▶ 7、某SRAM芯片，其容量为 $1\text{M} \times 8$ 位，除电源和接地端外，控制端有E和R/W#，该芯片的管脚引出线数目是（ ）。
- ▶ A 20 B 28 C 30 D 32

- ▶ 8、虚拟存储技术主要解决存储器的（ ）问题。
- ▶ A 速度 B 扩大存储容量 C 成本 D 前三者兼顾

- ▶ 9、主存储器和CPU之间增加Cache的目的是（ ）。
 - ▶ A 解决CPU和主存之间的速度匹配问题
 - ▶ B 扩大主存储器容量
 - ▶ C 扩大CPU中通用寄存器的数量
 - ▶ D 既扩大主存储器容量，又扩大CPU中通用寄存器的数量
- 

- ▶ 10、某计算机系统的存储器由Cache和主存构成。
已知在一段给定的时间内，CPU共访问内存5000次，其中400次访问主存。问Cache的命中率是多少？