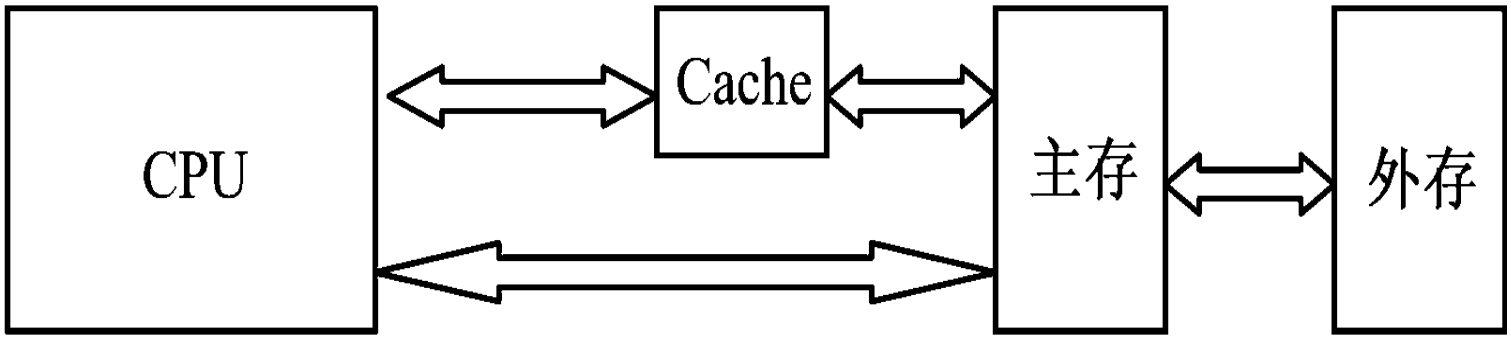
3.6 cache存储器

3.6.1 Cache的基本原理

- 原理：基于程序和数据访问的局部性
- 目标：减少访存次数，加快运行速度

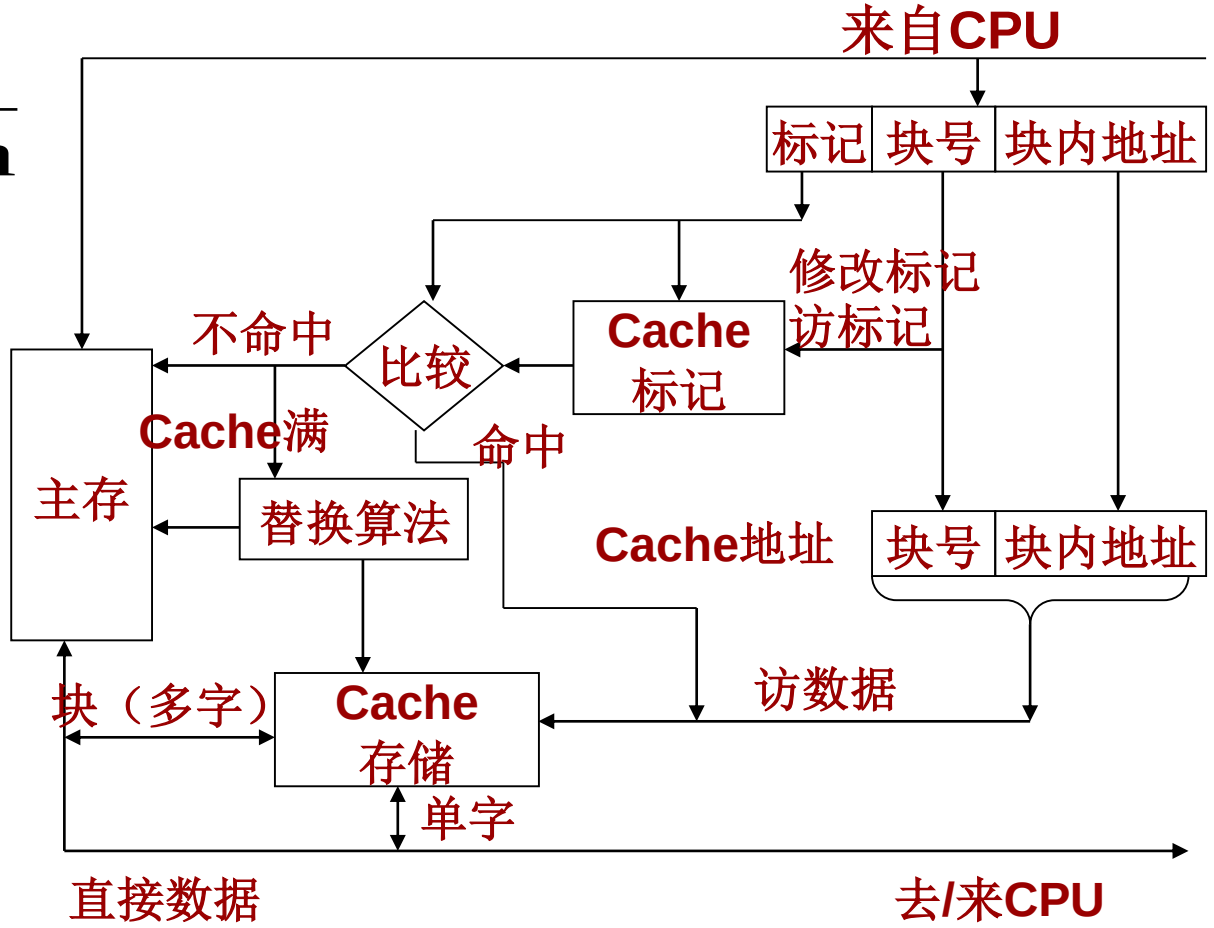


Cache与CPU及主存的关系

- 组成：由存储体、Cache-主存地址映射和Cache替换机构组成。
- 命中率：设一个程序执行期间，Nc表示cache完成存取的总次数，Nm表示主存完成存取的总次数，则命中率

$$h = \frac{Nc}{Nc + Nm}$$

高速缓存的基本结构



- t_c为命中的cache访问时间，t_m为未命中的主存访问时间，对存储系统的平均访问时间：

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

- 效率为：

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

- 设 $r = \frac{t_m}{t_c}$

$$e = \frac{t_c}{h t_c + (1-h) t_m} = \frac{1}{h + (1-h) r} = \frac{1}{r + (1-r) h}$$

例 CPU执行一段程序，cache完成存取的次数为1900次，主存完成存取次数为100次，已知cache存取周期为50ns，主存存取周期为250ns，求cache/主存系统的效率和平均访问时间。

- 方法一：

h=1900/(1900+100)=0.95

t_a=0.95*50+(1-0.95)*250=60ns

e=50/60=83.3%

- 方法二：

h=0.95；　　r=250/50=5

e=1/[5+(1-5)*0.95]=83.3%

由于 $e = \frac{t_c}{t_a}$ 得出$t_a=t_c/e$=60ns

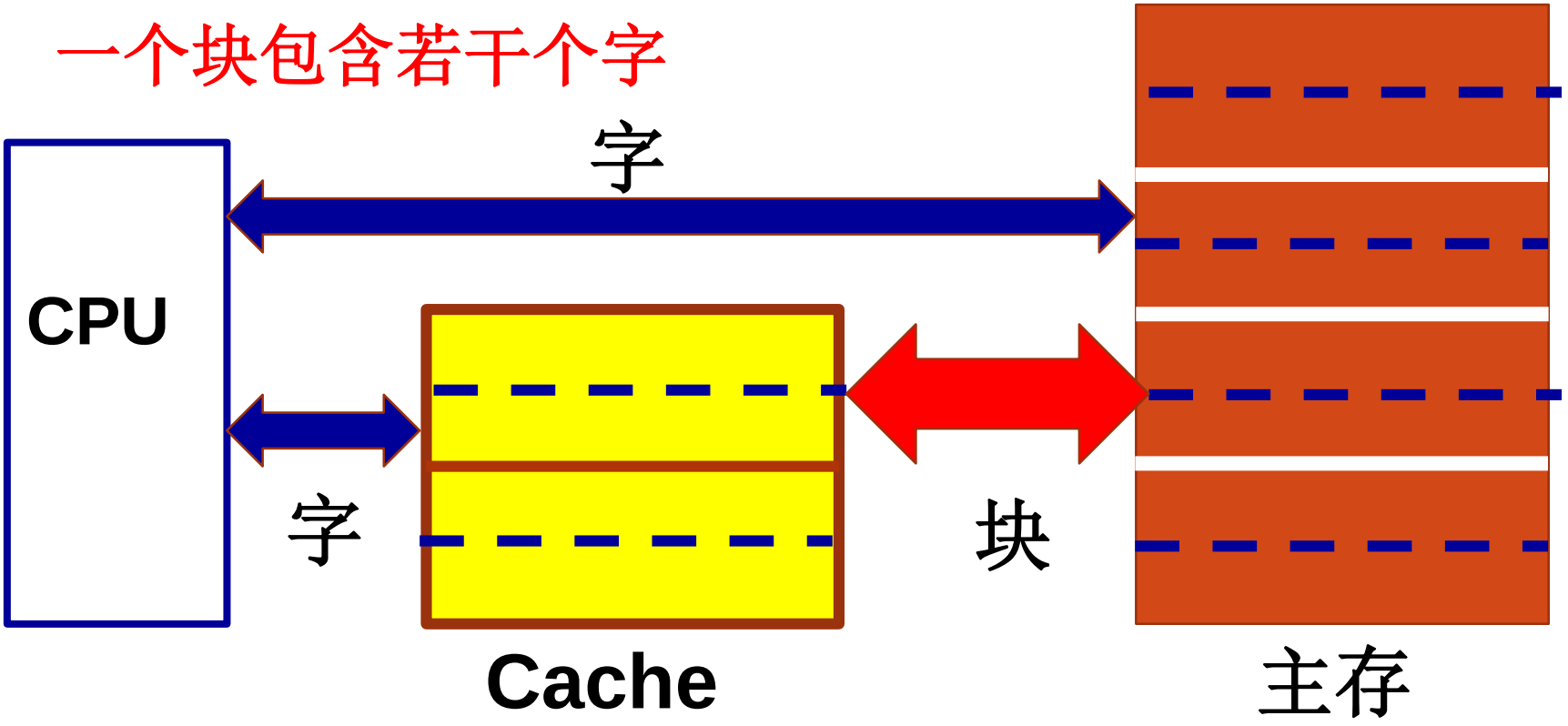
3.6.2 CPU、主存与Cache的信息交换方式

1、CPU与主存之间，以字为单位交换信息

CPU与Cache之间，以字为单位交换信息

2、Cache与主存之间，以块为单位交换信息

一个块包含若干个字



Q：主存按何种规则将内容复制到Cache中？

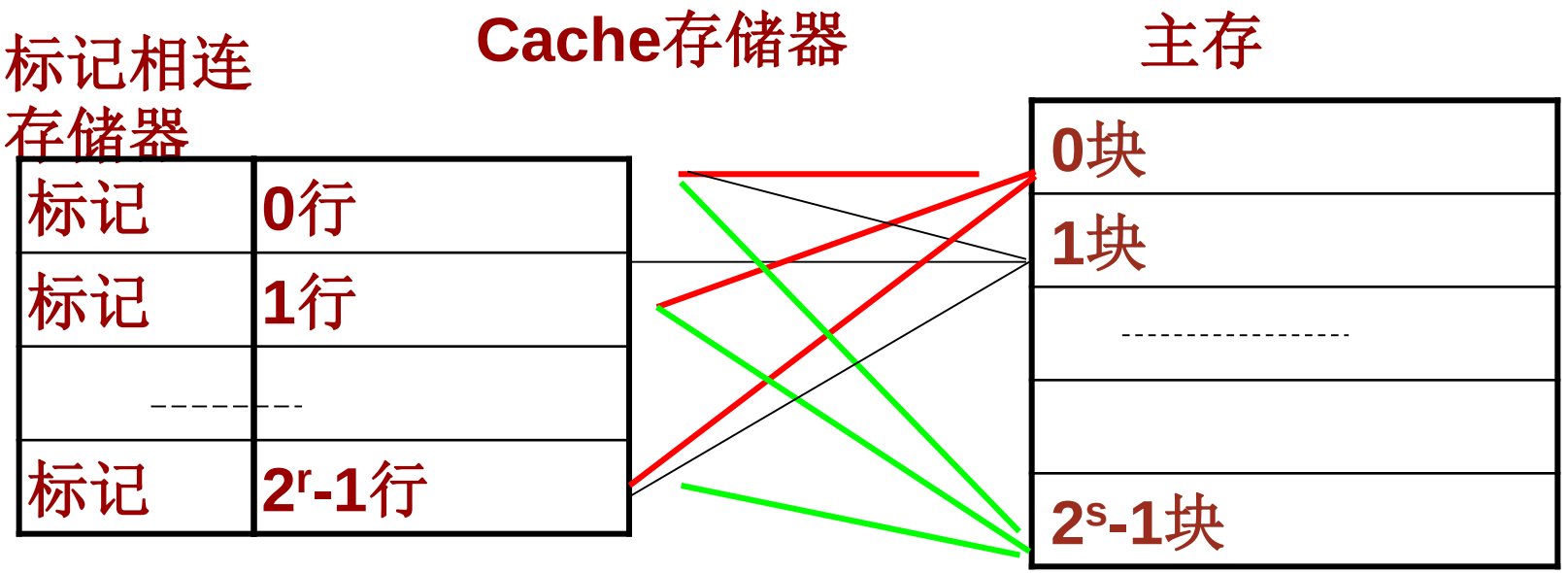
主存与Cache的地址映射方式：

- 全相联映射方式
- 直接映射方式
- 组相连映射方式

□ 主存与cache地址映射方式

1、全相连映射方式

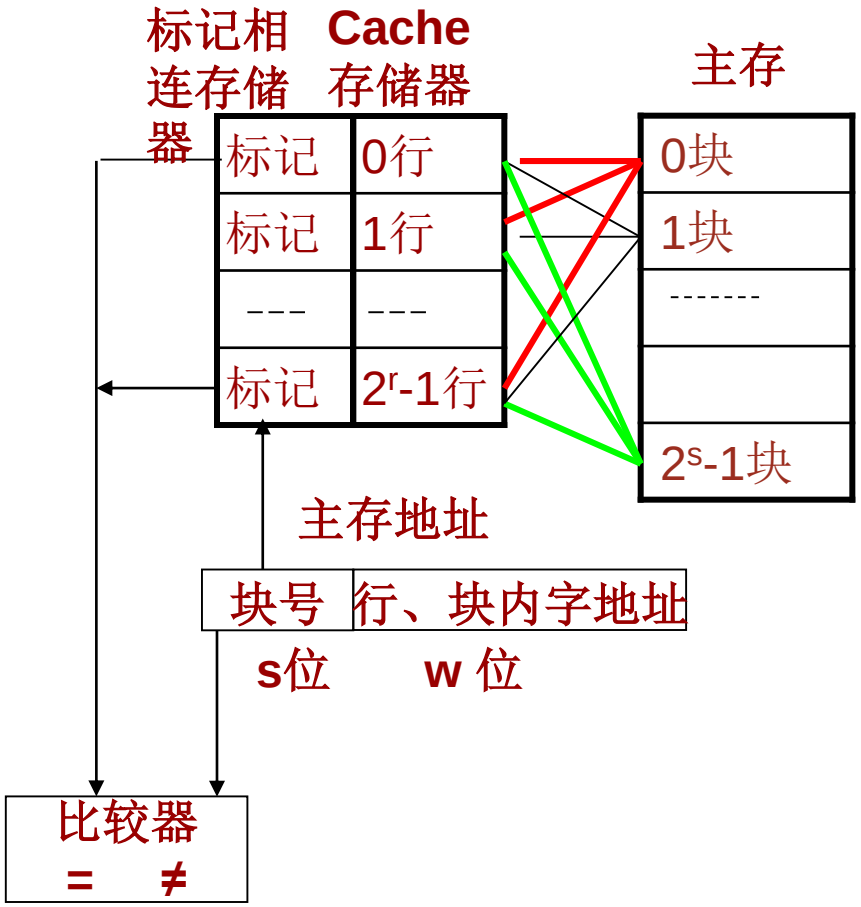
- 设:主存划分2^s个块,每块2^w个字.Cache分为2^r个行,每行大小同主存的块。
- 主存中每个块可复制到任一行的Cache行中，块号地址存于标记。



- **CPU读Cache检索步骤**

① 将块号与所有标记依次进行比较

② 命中，则根据块（行）内地址从Cache中读取该字；否则，CPU从主存中读取。

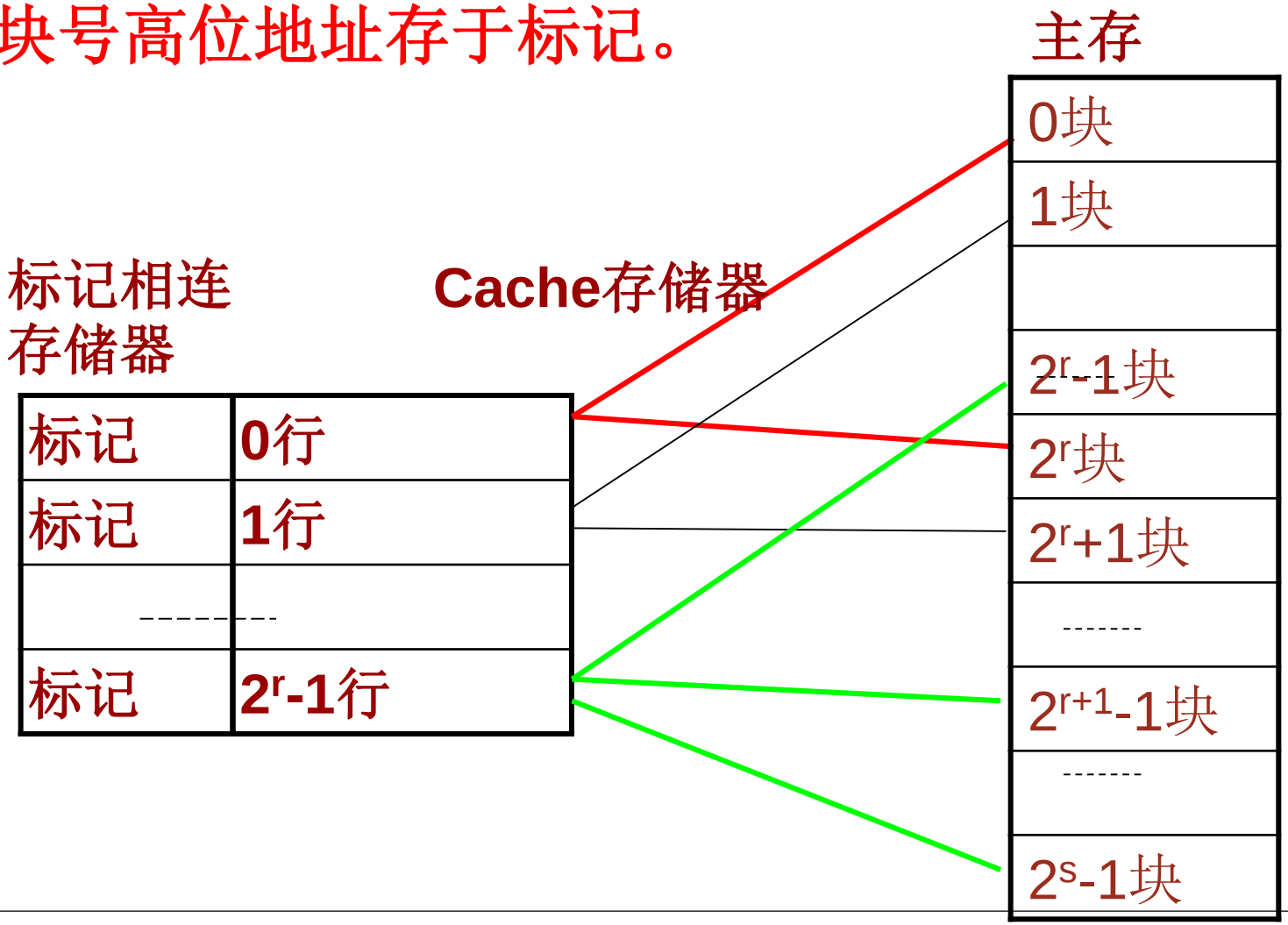


全相连映射Cache结构

- 优点：cache的利用率高。
- 缺点:比较多，速度慢。

2、直接映射方式

- 主存与cache的划分方式同全相连。
- 主存中每个块只能复制到某一固定行的Cache中，块号高位地址存于标记。



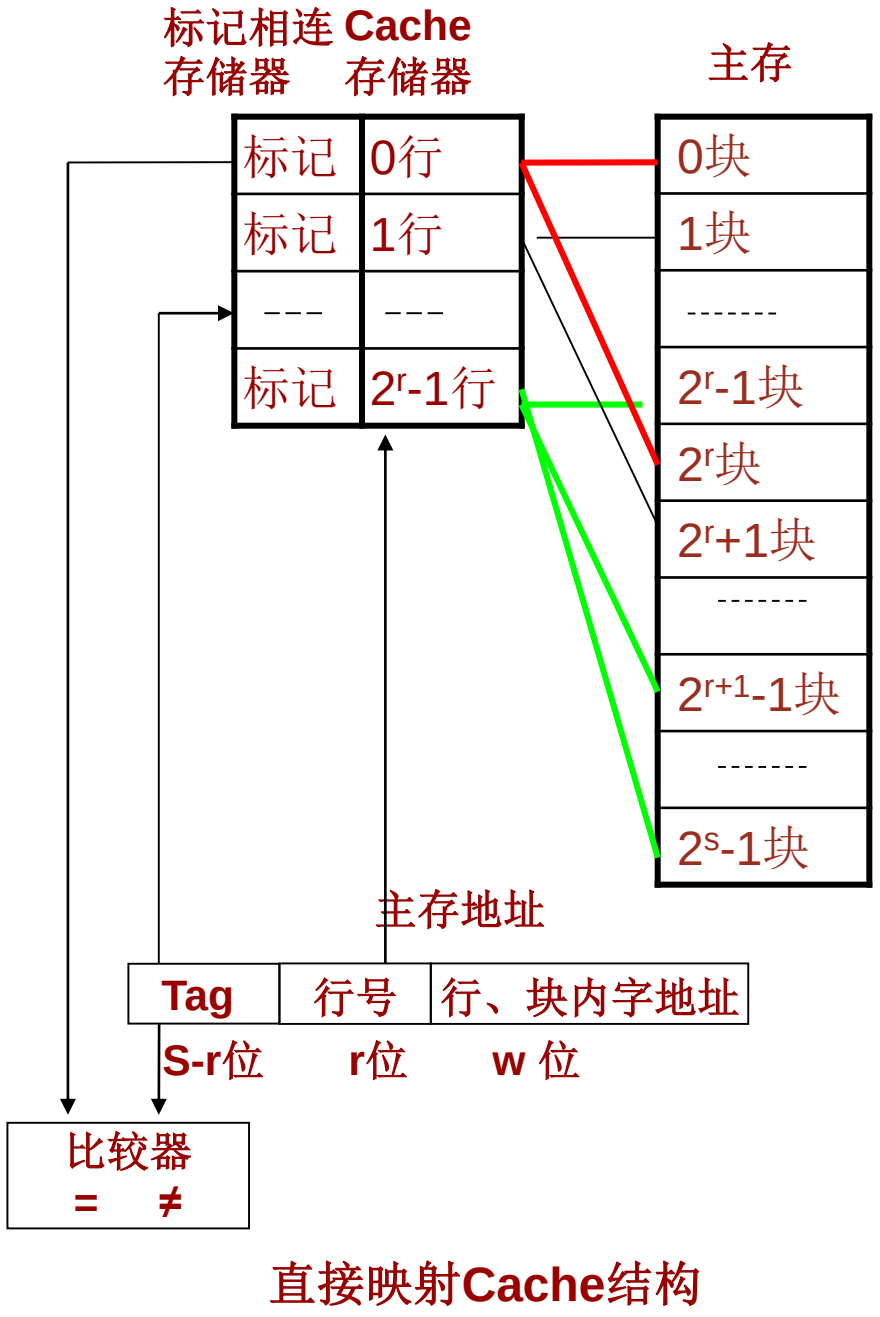
- **CPU读Cache检索步骤**

① 利用行号定位Cache中的某一行

② 将Tag与该行标记器中的值进行比较

③ 命中，则根据块内地址从Cache中读取该字

优点：比较简单，速度高。
缺点:块的冲突高，利用率低。



直接映射Cache结构

例11　设主存容量1MB，高缓容量16KB，行的大小为512B，采用直接映射：

（1）写出主存每部分地址位长；　（2）CACHE地址格式；

（3）行标记的容量为多大；（4）画出直接地址映像关系。

（1）主存地址

19　~ 1413 ~　9 8　~

0 Tag	行号	行内地址

（2）CACHE地址

13　9 8　0

行号	行内地址

（3）行标记容量

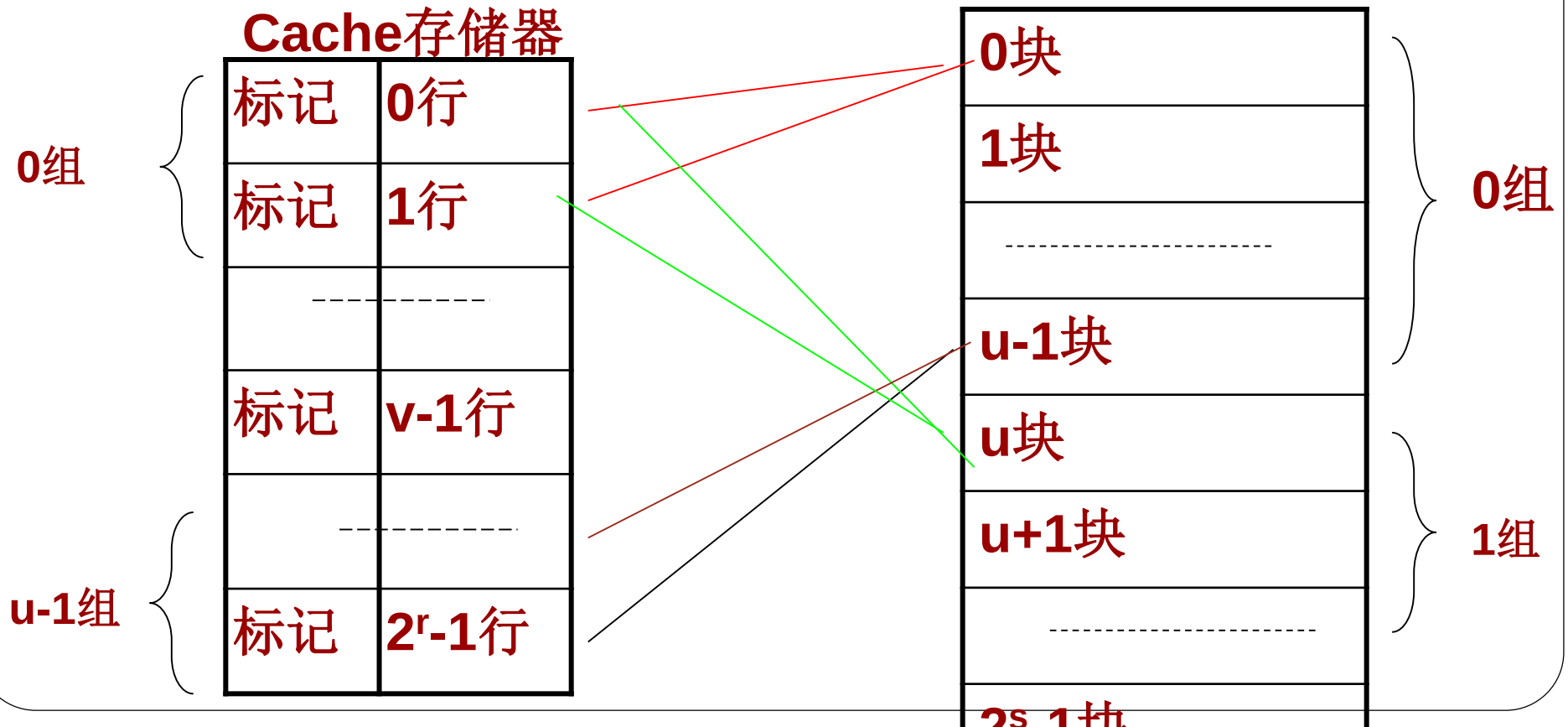
32行x6位=192b

（4）映像关系

CACHE行	主存块
0	0,32,64,……,2016
1	1,33,65,……,2017
2	2,34,66,……,2018
……	
30	30,62,94,……,2046
31	31,63,95,……,2047

3、组相联映射方式

- 将Cache与主存分组，设Cache中分成u(2^d)个组，每组v($2^{v'}$)个行,即r=d+v'。主存中一个组的块数与Cache中的分组数相同。
- 主存中的各块与Cache的组号有固定的映射关系,但可自由映射到对应的Cache组中的任何一行。



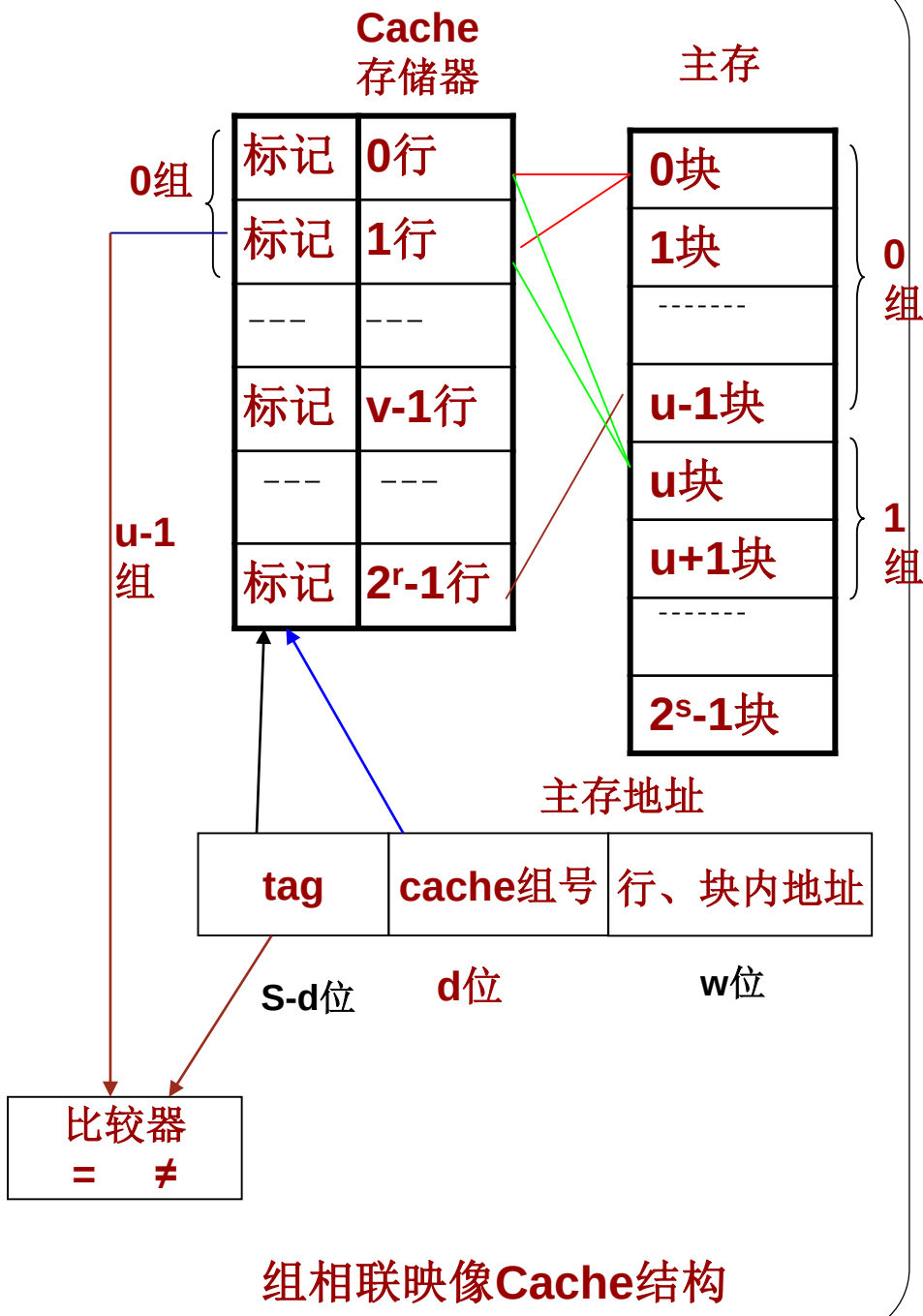
主存高位地址存入标记。

- **CPU读Cache检索步骤**

① 利用组号定位Cache中的某一组

② 将Tag与组内所有行的标记器中的值依次进行比较

③ 命中，则根据块内地址从Cache中读取该字



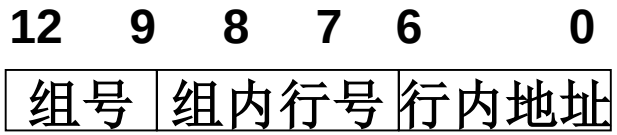
组相联映像Cache结构

- 组间直接映射，组内全相联映射
- **v=1**，即是直接映像；**u=1**，则是全相联映像，即组间为直接映射,组内为全相联映射。
- 组相联兼有全相连及直接相连的优点。

例12：一个组相连cache由64个存储行构成，每组4个存储行。主存包含4096个存储块，每块由128字组成，按字访存：（1）写出CACHE地址位数和地址格式；（2）写出主存地址位数和地址格式；（3）页标记容量（4）画出组相连映像关系。

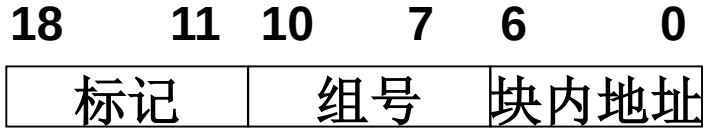
（1）CACHE地址

Cache容量：64x128=2^{13}



（2）主存地址

主存容量：4096x128=2^{19}



（3）页标记容量

64X8=512b

（4）映射表

CACHE行(16组)	主存块
0组（行0,1,2,3）	0,16,……,4080
1组（行4,5,6,7）	1,17,,……,4081
2组（行8,9,10,11）	2,18,,……,4082
……	
14组（行56,57,58,59）	14,30,,……,4094
15组（行60,61,62,63）	15,31,,……,4095

□ 替换策略

（1）最不经常使用算法LFU

（2）最近最少使用算法LRU

（3）先进先出算法FIFO

•Cache的读/写过程

读 将主存地址同时送往主存和Cache

- Cache命中 从cache读
- Cache失败 从主存读

写

- 写回法 Cache行被替换时，才写入主存
- 全写法 同时写Cache和主存
- 写一次法 第一次写时，同时写Cache和主存