

17. 负大小 64bit, 则页内偏移 6 bit
 TLB 16 位, 四路组相联, 则有 4 组, 索引 2 位
 块大小 4 byte, 则偏移 2 bit
 L1 有 16 位, 且直接映射: 索引 4 位

- 1) 对于 0x0504, 二进制为 00 0101 1010 0100
 对应第二组, 有效 0x05, 且双击为 1.
 : 发生命中, 物理页号为 0x1C, 物理地址为 01110010 0100, 即 0x724.
- 2) 虚拟地址长度 14bit, 块大小 64byte = 2^9 bit
 : 需要 $\frac{2^{14}}{2^9} = 2^5 = 32$ 位
 : 有 32 个条目
- 3) 物理地址 01110000100
 索引 1001 为 9, 读偏移为 0xD
 补全: 011100 为 0x1C, 相同, 双击为 1!
 : 发命中, 有效后果为 0x63.

18. 一段程序循环往复地按顺序访问 A、B、C、D 四个地址上的数据。考虑一个拥有 2 条目的全相联缓存, 回答以下问题。

- 1) 使用 LRU 替换策略时, 填写下表。当程序长时间运行时, 缓存的命中率为多少?

访存地址	A	B	C	D	A	B	C	D
way 0	—	A	A	C	C	A	A	C
way 1	—	—	B	B	D	D	B	B
命中?	N	N	N	N	N	N	N	N

- 2) 提出一种缓存替换策略, 使得上述程序可以在该缓存中拥有最大的命中率, 并计算该命中率。

1) 命中率为 0%.

2) 若每次替换掉最近访问的。

A	B	C	D	A	B	C	D	A	B	C	D
—	A	A	A	A	B	C	C	C	D	A	A
—	—	B	C	D	D	D	A	B	B	B	C
N	N	N	N	Y	N	N	Y	N	N	Y	N

命中率 33.3%.

19. 一些处理器引入了“微标签”(microtag)的技术来降低组相联缓存标签匹配过程的时序压力。该技术将地址的标签部分进一步拆分为高位标签 (HTag) 和低位标签 (LTag), 在判断缓存命中与否时, 控制器仅取出低位标签进行比较, 将匹配的缓存块预测为一次命中并把数据前馈给处理器。在随后的剩余周期内, 高位标签被取出并进一步用于判断该预测最终是否构成真正的命中。回答以下问题:

- 1) 低位标签在同一缓存组内通常被要求是唯一的, 试说明原因。
- 2) 基于对 1) 的讨论, 简要说明该技术的引入对于通常的缓存替换策略有什么影响。
- 3) 考虑到虚拟页偏移和物理页偏移是一致的, 为了提高访存性能, 系统可以进一步要求地址的低位标签和组索引位完全位于页偏移段内, 这样低位标签的匹配过程就完全不需要经过地址翻译而可以直接进行, 后续的高位标签则使用页表翻译后的结果判断是否构成真实命中。基于上述过程, 对于 16KB 页大小的内存系统, 一个 8KB 大小的四路组相联缓存至多可以拥有几比特的低位标签?

- 1) 组相联缓存中多个缓存块共用同一个组索引。如果多个缓存块的地址差相同，会导致地址冲突造成命中率下降。
- 2) 缓存替换策略通常是以整个标签进行匹配，而分为局部标签使替换策略只能在局部标签范围内进行，可能影响替换策略的公平性和准确性。
- 3) 16KB的页，页向移14位

假设大小为 x byte。

$$\text{高位位数} = \log_2(x) \text{ bit. 新: } \log_2\left(\frac{2^{11}}{x}\right)$$

$$\text{低位位数} = 14 - \log_2(x) - \log_2\left(\frac{2^{11}}{x}\right) = 3 \text{ 位}$$

20. 监听一致性和目录一致性各有什么优缺点？简述缓存一致性的实现代价体现在哪些方面？

1) 监听一致性：

优点：及时更新缓存，可立即对同一个缓存一致性域内的所有缓存进行更新
低延迟，可以更快地响应缓存命中和缺失
实现相对简单

缺点：产生更高的网络负载。
容易产生局部风暴

2) 目录一致性：

优点：更加灵活，可支持更广泛的系统
可扩展性，可更容易扩展到大规模系统中
更低的网络负载

缺点：更为昂贵的空间开销，占用更多内存空间
更高的延迟，尤其是在多处理器系统中。

3) 网络负载，导致网络拥堵与延迟

可能导致缓存失效而访问延迟增加。
实现较复杂，带来一定的设计和实现成本
内存开销，比如目录一致性中需要维护目录表。