

Homework 06

1. (改编自 COD_CH, P337, 5.2)

现对一个64位存储器进行访问，表 1_1 的第一列是以字地址形式给出的访问地址顺序。（假设字的大小为64位）

1) 假设存储器的 *Cache* 共有 16 个基本块，每个块大小为 1 个字。请参考表 1_1 的形式，给出这些访问地址在 *Cache* 中对应的：二进制字地址、索引、标签和访问的命中情况。

Word Address	Binary Address	index	Tag	Hit/Miss
0x03	0000 0011	3	0	Miss
0xb4	1011 0100	4	b	Miss
0x2b	0010 1011	b	2	Miss
0x02	0000 0010	2	0	Miss
0xbf	1011 1111	f	b	Miss
0x58	0101 1000	8	5	Miss
0xbe	1011 1110	e	b	Miss
0x0e	0000 1110	e	0	Miss
0xb5	1011 0101	5	b	Miss
0x2c	0010 1100	c	2	Miss
0xba	1011 1010	a	b	Miss
0xfd	1111 1101	d	f	Miss

2) 假设存储器的 *Cache* 共有 8 个基本块，每个块大小为 2 个字。请参考表 1_1 的形式，给出这些访问地址在 *Cache* 中对应的：二进制字地址、索引、标签和访问的命中情况。

(*Cache* 最初为空，替换策略采用 *LRU*，索引、标签给出二进制形式即可)

Word Address	Binary Address	index	Tag	Hit/Miss
0x03	0000 0011	1	0	Miss
0xb4	1011 0100	2	b	Miss
0x2b	0010 1011	5	2	Miss
0x02	0000 0010	1	0	Hit
0xbf	1011 1111	7	b	Miss
0x58	0101 1000	4	5	Miss
0xbe	1011 1110	7	b	Hit
0x0e	0000 1110	7	0	Miss
0xb5	1011 0101	2	b	Hit
0x2c	0010 1100	6	2	Miss
0xba	1011 1010	5	b	Miss
0xfd	1111 1101	6	f	Miss

2. (改编自 *COD_CH*, P339, 5.7)
考虑以下的程序和 *Cache* 行为:

每 1000 条指令的 数据读次数	每 1000 条指令的 数据写次数	指令 <i>Cache</i> 失效率	数据 <i>Cache</i> 失效率	块大小 (字节)
250	100	0.30%	2%	64

存储器位宽为 32-bit, 带宽单位采用: *bytes/cycle*
这里的读写带宽是指所有读写操作的期望值分别求和, 而不是瞬时最大值

- 1) 假设一个带有写直达、写分配 *Cache* 的 *CPU* 每周期执行的指令条数期望值为 0.5, 那么 *RAM* 和 *Cache* 之间的读写带宽应该为多少合适? (假设每个失效会生成一个块的请求)
2) 对于一个写回、写分配 *Cache* 来说, 假设替换出的数据 *Cache* 块中有 30% 是脏块, 那么假设 *CPU* 每周期执行的指令条数期望值依然为 0.5, *RAM* 和 *Cache* 之间的读写带宽需要达到多少?

CPI = 2 则每周期执行 0.5 条指令

读带宽: $0.5 \times 0.30\% \times 64 = 0.096 \text{ bytes/cycle}$ (指令 *Cache* 失效)

$0.5 \times \frac{250}{1000} \times 2\% \times 64 = 0.16 \text{ bytes/cycle}$ (读数据 *Cache* 失效)

$0.5 \times \frac{100}{1000} \times 2\% \times 64 = 0.064 \text{ bytes/cycle}$ (写数据 *Cache* 失效时写分配)

$0.096 + 0.16 + 0.064 = 0.32 \text{ bytes/cycle}$

写带宽: $0.5 \times \frac{100}{1000} \times 4 = 0.2 \text{ bytes/cycle}$ (写直达)

- 2) 对于一个写回、写分配 *Cache* 来说, 假设替换出的数据 *Cache* 块中有 30% 是脏块, 那么为了实现 *CPI* 为 2, 读写带宽需要达到多少?

读带宽不变, 仍为 0.32 bytes/cycle

写带宽: $0.5 \times \frac{250 + 100}{1000} \times 2\% \times 0.3 \times 64 = 0.0672 \text{ bytes/cycle}$

3.

假设 *CPU* 执行某段程序时, 共访问 *Cache* 命中 2000 次, 访问主存 50 次。已知 *Cache* 存取周期为 50ns, 主存存取周期为 200ns。求 *Cache*—主存系统的命中率和平均访问时间。

命中率: $\frac{2000}{2000 + 50} \approx 97.6\%$

平均访问时间: 认为在未命中时需考虑 *Cache* 访问时间

$\frac{2000 \times 50 + 50 \times 250}{2000 + 50} \approx 54.87 \text{ ns}$

4. (改编自 COD_CH, P341, 5.12)

多级 Cache 是一种重要的技术, 可以克服在一级 Cache 提供的有限空间的同时仍然保持速度。考虑具有一下参数的处理器:

无内存停顿的基本CPI	处理器速度	主存访问时间	每条指令的 L1 Cache 失效率	L2 直接映射 Cache 速度	L2 直接映射 Cache 全局失效率	L2 八路组 相联 Cache 速度	L2 八路组 相联 Cache 全局失效率
1.5	2GHz	100ns	7%	12 cycles	3.5%	28 cycles	1.5%

*L1 Cache 失效率是针对每条指令而言的。假设 L1 Cache 的总失效数量 (包括指令和数据) 为总指令数的 7%

1) 分别计算在下列情况下的处理器 CPI:

- 仅有 L1 Cache
- 使用 L2 直接映射 Cache
- 使用 L2 八路组相联 Cache

仅有 L1 Cache

$$1.5 + 7\% \times 2 \times 10^9 \times 100 \times 10^{-9} = 15.5$$

• 使用 L2 直接映射 Cache

$$1.5 + 7\% \times 12 + 3.5\% \times 2 \times 10^9 \times 100 \times 10^{-9} = 9.34$$

使用 L2 八路组相联 Cache

$$1.5 + 7\% \times 28 + 1.5\% \times 2 \times 10^9 \times 100 \times 10^{-9} = 6.46$$

2) 假设处理器采用 L2 直接映射 Cache, 设计人员希望添加一个 L3 Cache, 该 Cache 访问时间为 50 个时钟周期, 并且具有 13% 的失效率, 请计算此处理器的 CPI。

认为此处 13% 失效率为局部失效率

$$1.5 + 7\% \times 12 + 3.5\% \times 50 + 3.5\% \times 13\% \times 2 \times 10^9 \times 100 \times 10^{-9} = 5$$

3) 在较老的处理器中, 例如 Intel Pentium 或 Alpha 21264, L2 Cache 在主处理器和 L1 Cache 的外部 (位于不同芯片上)。虽然这种做法使得大型 L2 Cache 成为可能, 但是访问 L2 Cache 的延迟也变得很高, 并且因为 L2 Cache 以较低的频率运行, 所以带宽通常也很低。假设 512KiB 的片外 L2 Cache 的失效率为 4%, 如果每增加一个额外的 512KiB 片外 L2 Cache 能够降低 0.7% 的失效率, 并且片外 L2 Cache 的总访问时间为 50 个时钟周期, 那么片外 L2 Cache 容量必须多大才能与第一问列出的 L2 直接映射 Cache 的性能相匹配?

认为此处 4% 失效率为局部失效率

$$1.5 + 7\% \times 50 + 7\% \times 4\% \times 2 \times 10^9 \times 100 \times 10^{-9} = 5.56 < 9.34$$

则 Cache 容量为 512 KiB