

作业 2 参考答案

邵艾然 <sar13@mails.tsinghua.edu.cn>

2016-06-11

作业中提到的“教材”均指：《计算机系统结构教程》，张晨曦等编著，清华大学出版社。

第 1 题（教材-习题-7.4）

设有一个“Cache-主存”层次，Cache 为 4 块，主存为 8 块；试分别对于以下三种情况，画出其映像关系示意图，并计算访存块地址为 5 时的索引（index）。（1）全相联；（2）组相联，每组两块；（3）直接映像。

参考解答

映像关系示意图请参考教材“存储系统”一章中“映像规则”小结的插图。

全相联时块地址中没有索引。

两路组相联时，Cache 中共有两个组。因此，块地址为 5 时的索引为： $5 \bmod 2 = 1$ 。因为只需 1 比特就可以区分两个组，块地址中的索引部分最终表示为 1。

直接映像时，Cache 中共有 4 个组。因此，块地址为 5 时的索引为： $5 \bmod 4 = 1$ 。因为需要 2 比特才可以区分 4 个组，块地址中的索引部分最终表示为 01。

常见问题

- 1) 请同学们熟悉 Cache 中由“组（Set）”和“路（Way）”组成的二维结构。访存块地址的两个组成部分，“索引（Index）”和“标签（Tag）”，

就分别是用来区分不同“组”和不同“路”中的 Cache 行的。本题中主存为 8 块，因此访存块地址长度为 3。

- 全相联时，组数为 1，路数为 4（Cache 为 4 块）。索引不存在，长度为 0，标签的长度为 3。认为全相联时索引可以任意取是不对的。
- 直接映像时，组数为 4（Cache 为 4 块），路数为 1。索引长度为 2，标签长度为 1。
- 两路组相连时，组数为 2（Cache 为 4 块），路数为 2。索引长度为 1，标签长度为 2。

2) 注意 Cache 块编号、内存块编号、索引，一般都是从 0 开始的。这和内存地址从 0 开始是一个道理。

第 2 题（教材-习题-7.8）

假设对指令 Cache 的访问占全部访问的 75%；而对数据 Cache 的访问占全部访问的 25%。Cache 的命中时间为 1 个时钟周期，不命中开销为 50 个时钟周期，在混合 Cache 中一次 load 或 store 操作访问 Cache 的命中时间都要增加一个时钟周期，32KB 的指令 Cache 的不命中率为 0.39%，32KB 的数据 Cache 的不命中率为 4.82%，64KB 的混合 Cache 的不命中率为 1.35%。又假设采用写直达策略，且有一个写缓冲器，并且忽略写缓冲器引起的等待。试问指令 Cache 和数据 Cache 容量均为 32KB 的分离 Cache 和容量为 64KB 的混合 Cache 相比，哪种 Cache 的不命中率更低？两种情况下平均访存时间各是多少？

参考解答

解法 1

分离 Cache 不命中率为

$$75\% \times 0.39\% + 25\% \times 4.82\% = 1.4975\%$$

又因为已知混合 Cache 不命中率为 1.35%，可见混合 Cache 的不命中率更低。

分离 Cache 的平均访存时间为

$$75\% \times (1 + 0.39\% \times 50) + 25\% \times (1 + 4.82\% \times 50) = 1.74875$$

混合 Cache 的平均访存时间为

$$75\% \times (1 + 1 + 1.35\% \times 50) + 25\% \times (1 + 1 + 1.35\% \times 50) = 2.675$$

解法 2

分离 Cache 不命中率为

$$75\% \times 0.39\% + 25\% \times 4.82\% = 1.4975\%$$

又因为已知混合 Cache 不命中率为 1.35%，可见混合 Cache 的不命中率更低。

分离 Cache 的平均访存时间为

$$1 + 1.4975\% \times 50 = 1.74875$$

混合 Cache 的平均访存时间为

$$1 + 1 + 1.35\% \times 50 = 2.675$$

常见问题

- 1) 对平均访存时间的理解主要参考教材上关于平均访存时间的公式。解法 1 把访问指令和访问数据分开考虑，比较直观。解法 2 利用已经计算出的不命中率，统一考虑二者，比较简洁。
- 2) 有些同学在计算平均访存时间时还考虑了 load 和 store 指令在所有指令中的比例。这是不需要的，因为计算的就是“访存时间”，那些不需访存的指令本来就不在考虑范围内。
- 3) Cache 缺失时间 = 命中时间 + 缺失开销。有些同学计算“缺失时间”时只考虑了“缺失开销”，这是不对的。
- 4) 有些同学忽视了“混合 Cache 命中时间增加一个周期”的题目条件。

- 5) 有的同学认为 load 和 store 操作只是用于访问数据的, 因此混合 Cache 增加的命中时间只在访问数据时有效, 因此算得混合 Cache 的平均访问时间为 1.925。这种理解也是可以的。
- 6) 本次作业中的题目提供的数据比较多, 解答过程需要稍微复杂的计算。结果有很多同学写对了算式, 却算错了最终结果。算错数的问题还请同学们想办法避免。

第 3 题 (教材-习题-7.10)

给定以下的假设, 试计算直接映像 Cache 和两路组相联 Cache 的平均访问时间以及 CPU 的性能。由计算结果能得出什么结论?

- (1) 理想 Cache 情况下的 CPI 为 2.0, 时钟周期为 2ns, 平均每条指令访存 1.2 次;
- (2) 两者 Cache 容量均为 64KB, 块大小都是 32 字节;
- (3) 组相联 Cache 中的多路选择器使 CPU 的时钟周期增加了 10%;
- (4) 这两种 Cache 的不命中开销都是 80ns;
- (5) 命中时间为 1 个时钟周期;
- (6) 64KB 直接映像 Cache 的不命中率为 1.4%, 64KB 两路组相联 Cache 的不命中率为 1.0%。

参考解答

直接映像 Cache 的平均访问时间为

$$1 \times 2ns + 1.4\% \times 80ns = 3.12ns$$

两路组相联 Cache 的平均访问时间为

$$1 \times (2ns + 0.1 \times 2ns) + 1.0\% \times 80ns = 3.0ns$$

直接映像下的 CPU 时间为

$$IC \times (2.0 \times 2 + 1.2 \times 1.4\% \times 80) = 5.344 \times IC$$

两路组相联下的 CPU 时间为

$$IC \times (2.0 \times (2 + 0.1 \times 2) + 1.2 \times 1.0\% \times 80) = 5.36 \times IC$$

由于组相联 Cache 影响了 CPU 时钟周期，直接映像 Cache 虽然平均访存时间较长，却能最终实现较低的 CPU 时间。因此本题中，直接映像 Cache 是更好的选择。

常见问题

- 1) 对 CPU 时间的理解主要参考教材上关于 CPU 时间的公式。有些同学对 CPU 时间的理解还存在问题，想要直接代入平均访存时间进行计算，结果没有算对。
- 2) 教材上 CPU 时间的公式假设“不命中开销”的单位是“时钟周期”。本题中，“不命中开销”的单位是“ns”，教材上的公式需要相应的调整才能使用。
- 3) Cache 缺失时间 = 命中时间 + 缺失开销。有些同学计算“缺失时间”时只考虑了“缺失开销”，这是不对的。
- 4) 同样的算错数问题也在本题中广泛出现了，请同学们想办法避免。

第 4 题（教材-习题-7.14）

假设一台计算机具有以下特性：

- (1) 95% 的访存在 Cache 中命中；
- (2) 块大小为两个字，且不命中时整个块被调入；
- (3) CPU 发出访存请求的速率为 10^9 字/秒；
- (4) 25% 的访存为写访问；
- (5) 存储器的最大流量为 10^9 字/秒（包括读和写）；
- (6) 主存每次只能读或写一个字；
- (7) 在任何时候，Cache 中有 30% 的块被修改过；
- (8) 写不命中时，Cache 采用按写分配法。

现欲给该计算机增添一台外设，为此首先想知道主存的带宽已用了多少。试对于以下两种情况计算主存带宽的平均使用比例。

- (1) 写直达 Cache;
- (2) 写回法 Cache。

参考解答

写直达法主存带宽平均使用比例为

$$\frac{10^9 \times (75\% \times 5\% \times 2 + 25\% \times 95\% \times 1 + 25\% \times 5\% \times 2)}{10^9} = 35\%$$

写回法主存带宽平均使用比例为

$$\frac{10^9 \times (75\% \times 5\% \times (2 + 30\% \times 2) + 25\% \times 5\% \times 2)}{10^9} = 13\%$$

常见问题

- 1) 本题的关键是搞清不同方法下读写 Cache 命中和缺失时的操作流程。
 - 写直达, 读命中: 直接读 Cache, 不访问内存。
 - 写直达, 读缺失: 直接访问内存, 读取一个块 ($75\% \times 5\% \times 2$)。
 - 写直达, 写命中: 需要写内存, 只写修改过的字即可 ($25\% \times 95\% \times 1$)。
 - 写直达, 写缺失: 因为是写分配法, 先从内存读入所需的 Cache 块 ($25\% \times 5\% \times 2$), 再写入新数据, 只写修改过的字即可 ($25\% \times 5\% \times 1$)。
 - 写回, 读命中: 直接读 Cache, 不访问内存。
 - 写回, 读缺失: 先写回被替换出 Cache 块 ($75\% \times 5\% \times 30\% \times 2$), 再从内存读入所需的 Cache 块 ($75\% \times 5\% \times 2$)。
 - 写回, 写命中: 直接写 Cache, 不访问内存。
 - 写回, 写缺失: 先写回被替换出 Cache 块 ($25\% \times 5\% \times 30\% \times 2$), 再从内存读入所需的 Cache 块 ($25\% \times 5\% \times 2$), 最后写 Cache。
- 2) 再强调一下同学们容易出错的地方。
 - 写直达法写内存时不是必须写整个 Cache 块, 可以只写修改过的字。
 - 写回法在读缺失或写缺失时, 如果被替换的 Cache 块被修改过, 则需要写回; 如果没被修改过, 则不需要写回。
- 3) 本题已知条件中给出的数据比较多也比较乱。很多同学思路是对的, 但列算式时却带入了错误的数据。另外算错数的同学也很多。请同学们尽量避免这种错误。