

Homework 3

PB17000297 罗晏宸

April 2 2020

1

考虑一个由 L1 和 L2 数据缓存组成的两级存储器层次结构，假定两个缓存在写入命中时都使用写回策略，两者的块大小相同。列出在以下事件时采取的操作。

- a** 当缓存组织方式为包含式层次结构时，发生 L1 缓存缺失 (miss)
- b** 当缓存组织方式为互斥式层次结构时，发生 L1 缓存缺失
- c** 在 a 部分和 b 部分中，考虑被替换的块是脏块的可能性（需要更新和不需要更新的可能性）

解

a 当缓存组织方式为包含式层次结构时，L1 中数据总是出现在 L2 中。L1 中块缺失，可以分为以下两种情况：

- 缺失的块在 L2 中：L2 将该块 b_0 提供给 L1，相应的 L1 中某块 b_1 会被替代。
- 缺失的块不在 L2 中：L2 需要到内存中取，取回的块 b_0 同时提供给 L1 和 L2。相应的 L1 和 L2 中有块会被替代。L2 中被替代的块 b_1 如果也在 L1 中，那么 L1 中的这一块 b_1 需要被置为失效。

b 当缓存组织方式为互斥式层次结构时, L1 中数据绝不出现在 L2 中。L1 中块 b_0 缺失, 可以分为以下两种情况:

- 缺失的块在 L2 中: L2 将该块 b_0 提供给 L1, 相应的 L1 中某块 b_1 会被替代。互斥式层次结构 L1 中的缓存缺失会导致 L1 和 L2 中的块互换, 即 b_1 和 b_0 互换。
- 缺失的块不在 L2 中: L1 需要到内存中取 b_0 , 取回的块只提供给 L1。相应的 L1 有块 b_1 会被替代。将 L1 被替代的块 b_1 写入 L2, L2 也有块 b_2 被替代。

c 考虑被替换的块是脏块的可能性。当缓存组织方式为包含式层次结构时, 对于两种情况:

- 缺失的块在 L2 中: L2 将该块 b_0 提供给 L1, 相应的 L1 中某块 b_1 会被替代。由于两个缓存都采取写回策略, L1 中被替代的块 b_1 可能是脏块, 如果是脏块, 将它写入 L2 对应的块 b_1
- 缺失的块不在 L2 中: L2 需要到内存中取, 取回的块 b_0 同时提供给 L1 和 L2。相应的 L1 和 L2 中有块会被替代。L2 中被替代的块 b_1 如果也在 L1 中, 那么 L1 中的这一块 b_1 需要被置为失效。L1 中被替代的块 b_2 如果是脏块, 将它写入 L2 对应的块 b_1 。

当缓存组织方式为互斥式层次结构时, L1 中块 b_0 缺失, 对于两种情况:

- 缺失的块在 L2 中: L2 将该块 b_0 提供给 L1, 相应的 L1 中某块 b_1 会被替代。互斥式层次结构 L1 中的缓存缺失会导致 L1 和 L2 中的块互换, 即 b_1 和 b_0 互换。
- 缺失的块不在 L2 中: L1 需要到内存中取 b_0 , 取回的块只提供给 L1。相应的 L1 有块 b_1 会被替代。将 L1 被替代的块 b_1 写入 L2, L2 也有块 b_2 被替代。 b_2 如果是脏块, 需要写回内存。

2

每当计算机空闲时, 既可以将其置于待机状态 (DRAM 仍然处于活动状态), 也可以让他休眠。为了使其进入休眠状态, 假定必须仅将 DRAM 的

内容复制到永久性介质中，比如闪存中，如果将大小为 64 字节的缓存行读写至闪存需要 2.56 μJ ，读写至 DRAM 需要 0.5 nJ，如果 8 GB DRAM 空闲功耗为 1.6 W，那么一个系统空闲多长时间后才能从休眠中获益？假定主存储器的容量为 8 GB

解

$$T = \frac{8 \text{ GB} \times 2 \times (2.56 \mu\text{J} + 0.5 \text{ nJ})}{64 \text{ B} \times 1.6 \text{ W}} \\ = 400.078 \text{ s}$$

3

你正要采用一个具有以下特征的处理器构建系统：循序执行，运行频率为 1.1 GHz，排除存储器访问在外的 CPI 为 1。只有载入和存储指令能从存储器读写数据，载入指令占全部指令的 20%，存储指令占 5%。此计算机的存储器系统包括一个分离的 L1 缓存，它在命中时不会产生任何代价。I 缓存和 D 缓存都是直接映射，分别为 32 KB。I 缓存的缺失率为 2%，块大小为 32 字节，D 缓存为直写缓存，缺失率为 5%，块大小为 16 字节。D 缓存上有一个写入缓冲区，消除了绝大多数写入操作的停顿，占总写入操作的 95%（指 CPU 的写停顿比例）。L2 为 512 kB，统一 L2 的块大小为 64 字节，访问时间为 15 ns。它由 128 位数据总线连接到 L1 缓存，运行频率为 266 MHz，每条总线每个时间周期可以传送一个 128 位字。在发往此系统 L2 缓存的所有存储器引用中，其中 80% 的引用无须进入主存储器就可以得到满足。另外，在被替换的所有块中，50% 为脏块。主存储器的宽度为 128 位，访问延迟为 60 ns，在此之后，可以在这个宽 128 位，频率为 133 MHz 的主存储器总线上以每个周期传送一个字的速率来传送任意数目的总线字。

- a 指令访问的存储器平均访问时间为多少
- b 数据读取的存储器平均访问时间为多少
- c 数据写入的存储器平均访问时间为多少
- d 包括存储器访问在内的整体 CPI 为多少

解 首先计算缓存的缺失代价

$$\text{缺失代价}_{\text{L1 I-缓存}} = 15 \text{ ns} + \frac{32 \text{ B}}{16 \text{ B}} \times \frac{1}{266 \text{ MHz}} \times 1000 = 22.5 \text{ ns}$$

$$\text{缺失代价}_{\text{L1 D-缓存}} = 15 \text{ ns} + \frac{16 \text{ B}}{16 \text{ B}} \times \frac{1}{266 \text{ MHz}} \times 1000 = 18.75 \text{ ns}$$

$$\begin{aligned} \text{缺失代价}_{\text{L2 缓存}} &= \left(60 \text{ ns} + \frac{64 \text{ B}}{16 \text{ B}} \times \frac{1}{133 \text{ MHz}} \times 1000 \right) \times (100\% + 50\%) \\ &= 135 \text{ ns} \end{aligned}$$

a

$$\begin{aligned} \text{存储器平均访问时间}_{\text{I}} &= \text{缺失率}_{\text{L1 I-缓存}} \\ &\quad \times (\text{缺失代价}_{\text{L1 I-缓存}} + \text{缺失率}_{\text{L2 缓存}} \times \text{缺失代价}_{\text{L2 缓存}}) \\ &= 2\% \times (22.5 \text{ ns} + 20\% \times 135 \text{ ns}) \\ &= 0.99 \text{ ns} \end{aligned}$$

b

$$\begin{aligned} \text{存储器平均访问时间}_{\text{read}} &= \text{缺失率}_{\text{L1 D-缓存}} \\ &\quad \times (\text{缺失代价}_{\text{L1 D-缓存}} + \text{缺失率}_{\text{L2 缓存}} \times \text{缺失代价}_{\text{L2 缓存}}) \\ &= 5\% \times (18.75 \text{ ns} + 20\% \times 135 \text{ ns}) \\ &= 2.29 \text{ ns} \end{aligned}$$

c 如果停顿消除到 L1，则有

$$\begin{aligned} \text{存储器平均访问时间}_{\text{write}} &= \text{写停顿比例} \\ &\quad \times (\text{缺失代价}_{\text{L1 D-缓存}} + \text{缺失率}_{\text{L2 缓存}} \times \text{缺失代价}_{\text{L2 缓存}}) \\ &= (1 - 95\%) \times (18.75 \text{ ns} + 20\% \times 135 \text{ ns}) \\ &= 2.29 \text{ ns} \end{aligned}$$

如果停顿消除到 L2，则有

$$\begin{aligned} \text{存储器平均访问时间}_{\text{write}} &= \text{缺失代价}_{\text{L1 D-缓存}} \\ &\quad + \text{写停顿比例} \times \text{缺失率}_{\text{L2 缓存}} \times \text{缺失代价}_{\text{L2 缓存}} \\ &= 18.75 \text{ ns} + (1 - 95\%) \times (20\% \times 135 \text{ ns}) \\ &= 20.1 \text{ ns} \end{aligned}$$

d

$$\begin{aligned}\text{CPI}_{\text{total}} &= \text{CPI}_{\text{base}} + \text{CPI}_{\text{I}} + \text{CPI}_{\text{read}} + \text{CPI}_{\text{write}} \\ &= 1 + (0.99 \text{ ns} \times +20\% \times 2.29 \text{ ns} + 5\% \times 2.29 \text{ ns}) \times 1.1 \text{ GHz} \\ &= 2.71875\end{aligned}$$