

# 第三次体系结构作业

PB19051183 吴承泽

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.列出在一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)

假设所查找的块为  $Block_i$ , 回答如下:

a.

**L2中查找命中:**

从L2缓存中查找  $Block_i$ , 若命中则将L2中的块  $Block_i$  替换L1中的某个块  $Block_j$ . 若被替换的块是脏块, 则将所替换的  $Block_j$  写入L2中某个合适的位置上, 并标记为脏。

**L2中查找未命中:**

从主存中查找  $Block_i$ , 因为是包含式层次结构, L1中的块在L2中均有备份, 将  $Block_i$  替换L1与L2中的某个块  $Block_j$ 。若被替换的块为脏块, 则将所替换的  $Block_j$  写入L2中的某个位置, 标记为脏。

b.

**L2中查找命中:**

从L2缓存中查找  $Block_i$ , 若命中则将L2中  $Block_i$  与L1中某个块  $Block_j$  进行交换。若  $Block_j$  为脏块, 则将L2中替换后的该块标记为脏块。

**L2中查找未命中:**

从L2缓存中查找  $Block_i$ , 若未命中则在主存中查找  $Block_i$ , 将所查找的块替换L1缓存中的  $Block_j$ , 将原L1缓存中的  $Block_j$  替换至L2缓存上, 届时L2缓存中会有块  $Block_i$  被替换至主存中。若  $Block_j$  为脏块, 则在L2上  $Block_j$  的位置将其标记为脏, 等待写回内存。

c.

已经在a)、b)中分别进行讨论, 在此不再进行赘述。

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  $2.56\mu\text{j}$ ,读写至 DRAM 需要  $0.5\text{nJ}$ ,如果 8GB DRAM 空闲功耗为 1.6W,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 8GB

假设经过时间  $t$  后获益: (考虑在待机中启动需将数据从闪存中读回DRAM中, 开销和读至闪存相等, 假设  $1\text{GB} = 10^9\text{B}$ , 因为缓存和主存相比过小, 因此缓存写入DRAM的开销可以忽略不计, 因此仅需要比较DRAM待机和写闪存的能量开销)

$$1.6 * t = 2 * 2.56 * 10^{-6} * 8 * 10^9 / 64$$

解得:

$$t = 400\text{s}$$

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1.1GHz,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 32KB.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 512KB,统一 L2 的块大小为 64 字节,访问时间为 15ns.它由 128 位数据总线连接到 L1 缓存,运行频率为 266MHz,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 60ns,在此之后,可以在这个宽 128 位,频率为 133MHz 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.

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

一些基本信息如下:

CPU主频: **1.1GHz**      无载入存储指令时: **CPI=1**

指令分布: **75%无载入存储指令; 20%载入指令; 5%存储指令**

Cache大小: **L1I 32KB 直接映射; L1D 32KB 直接映射; L2 512KB**

块大小: **L1I 32B; L1D 16B; L2 64B**

传输总线宽度: **16B**

传输次数: **L1I 2次; L1D 1次; L2 4次**

Miss Rate; **L1I 2%; L1D 5%; L2 20%**

访问时间: **L1I 0; L1D 0; L2 15ns; 主存 60ns**

写停顿比例**5%; L2脏块比例50%**

频率: **L2 266MHz; 主存133MHz**

L1I缺失代价:  **$15ns + 2 * 1000 / 266 = 22.5ns$**

L1D缺失代价:  **$15ns + 1000 / 266 = 18.75ns$**

L2缺失代价:  **$60ns + 4 * 1000 / 133 = 90ns$**

由于L2中50%的块是脏块,则有50%的块访问需要使用额外的L2缺失代价的时间写回内存中,即

L2平均缺失代价:  **$90ns * (1 + 0.5) = 135ns$**

a.

$$T_{instruction} = 0.02 * (22.5 + 0.2 * 135) = 0.99ns$$

b.

$$T_{data} = 0.05 * (18.75 + 0.2 * 135) = 2.29ns$$

c.

若停顿消除至L1:

$$T_{write} = 0.05 * (18.75 + 0.2 * 18.75) = 2.2875ns$$

若停顿消除至L2:

$$T_{write} = 18.75 + 0.05 * 0.2 * 135 = 20.1ns$$

d.

总CPI为基础运行的CPI，取指令的CPI，读数据的CPI与写数据的CPI构成。

$$CPI = CPI_{base} + CPI_{InstructionFetch} + CPI_{DataFetch} + CPI_{Write} = 1 + 1.09 + 0.2 * 2.52 + 0.05 * 2.52 = 2.72$$