

HW02

PB19071405 王昊元

2022 年 04 月 18 日

1. a. 包含式层次结构中, $L1$ 中数据总是出现在 $L2$ 中。则 $L1$ 中块缺失, 可以分为缺失的块在 $L2$ 中与不在 $L2$ 中两种情况考虑。
 - i. 缺失的块 $b0$ 在 $L2$ 中: $L2$ 将 $b0$ 提供给 $L1$, 替换掉 $L1$ 中的块 $b1$, 如果 $b1$ 是脏块 (即数据被修改过), 将它写入 $L2$ 对应的块 $b1$ 即可。
 - ii. 缺失的块 $b0$ 不在 $L2$ 中: $L2$ 到内存中取, 取回的块 $b0$ 同时提供给 $L1$ 和 $L2$, 替换掉 $L1$ 和 $L2$ 中的块 $b1$ 和块 $b2$, 如果 $b1$ 是脏块, 则将它写入 $L2$ 中对应的块 $b1$ 中, 如果 $b2$ 也在 $L1$ 中, 则 $L1$ 中 $b2$ 需要被置失效,
 - b. 互斥式层次结构中, $L1$ 中数据绝不出现在 $L2$ 中。则 $L1$ 中块缺失, 也可以分为缺失的块在 $L2$ 中与不在 $L2$ 中两种情况。
 - i. 缺失的块 $b0$ 在 $L2$ 中: $L2$ 将 $b0$ 提供给 $L1$, 替换掉 $L1$ 中的块 $b1$, 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 $b0$ 和 $b1$ 互换。
 - ii. 缺失的块 $b0$ 不在 $L2$ 中: $L1$ 到内存中取, 取回的块 $b0$ 提供给 $L1$, 替换掉 $L1$ 中的块 $b1$, 再用 $b1$ 替换掉 $L2$ 中的块 $b2$, 如果 $b2$ 是脏块, 则将 $b2$ 写回内存。
 - c. 如 i. 和 ii. 所述。
2. 要使系统从休眠中获益, 则要使待机期间能量消耗高于进入和退出休眠状态所需的能量。进入和退出休眠状态所需要的能量为

$$\eta = \frac{8 \times 2^{30}}{64} \times (2.56\mu J + 0.5nJ) \times 2 \approx 687.328985088J$$

则可以使系统收益的空闲时间至少

$$T = \frac{\eta}{P} = \frac{687.328985088J}{1.6W} \approx 429.58061567999994s$$

3. 由题可知一下信息:

- CPU 运行频率为 1.1GHz, 1 个 cycle 为 $\frac{10}{11}ns$
- 非访存指令 CPI 为 1
- 指令构成为 75% 非访存指令 + 20% 载入指令 + 5% 存储指令
- Block Size: L1 I-cache: 32B, L1 D-cache: 16B, L2 cache: 64B
- Miss rate: L1 I-cache: 2%, L1 D-cache: 5%, L2 cache: $1 - 80\% = 20\%$
- L1/L2 cache 传输位宽: $128b = 16B$, L2/主存储器传输位宽: $128b = 16B$
- block 缺失传输次数: L1 I-cache: 2, L1 D-cache: 1, L2 cache: 4
- 写停顿比例: $1 - 95\% = 5\%$
- 缺失延时: L1 cache: 15ns, L2 cache: 60ns
- 传输周期: L1/L2: $1/266MHz = 3.75ns$, L2/主存储器: $1/133MHz = 7.5ns$
- 脏块比例: 50%

则我们可以计算出各个 cache 的缺失代价如下:

- L1 I-cache: $15ns + 3.75ns \times \frac{32B}{16B} = 22.5ns$
- L1 D-cache: $15ns + 3.75ns \times \frac{16B}{16B} = 18.75ns$
- 不考虑脏块的情况下, L2 cache: $60ns + 7.5ns \times \frac{64B}{16B} = 90ns$
- 考虑脏块的情况下, L2 cache: $90ns \times (1 + 50\%) = 135ns$

a. $2\% \times (22.5ns + 20\% \times (135ns)) = 0.99ns$

b. $5\% \times (18.75ns + 20\% \times (135ns)) = 2.2875ns$

c. 分两种情况讨论:

i. 写入考虑了读取, 消除的停顿是 L1 写回 L2, 直写法一定写回, 由于考虑了读取, 则 L2 一定命中, 而 L2 不需要写回主存储器 (写回法命中时不直接写回), 此时平均访问时间为 $2.29ns + 5\% \times 18.75ns = 3.2275ns$

ii. 写入没有考虑读取,

- 如果消除的停顿为到 L1, 则平均访存时间为 $5\% \times (18.75 + 20\% \times 135ns) = 2.2875ns$
- 如果消除的停顿为到 L2, 则平均访存时间为 $18.75ns + 5\% \times 20\% \times 135ns = 20.1ns$

d. $CPI = \text{base CPI} + \text{inst CPI} + \text{load CPI} + \text{store CPI} = 1 + 0.99ns / \frac{10}{11}ns + 20\% \times 2.2875ns / \frac{10}{11}ns + 5\% \times 2.2875ns / \frac{10}{11}ns = 2.718$