

计算机体系结构 HW3

PB18111697 王章瀚

第一题

总结题目信息如下:

存储器级	块大小	总大小	访问时间	缺失率	特性
L1(I)	32B	32KB	命中无代价	2%	
L1(D)	16B	32KB	命中无代价	5%	直写缓存
L2	64B	512KB	15ns	1-80%=20%	写回, 统一

L1(D) 有写缓冲区, 消除 95% 写操作的停顿

主存: 位宽 128 bit, 访问延迟 60 ns

总线连接情况:

- L2 到 L1: 以频率 266 MHZ, 每个周期传 128 bit, 即传输一个 16B 的块需要

$$2 \times \frac{1}{266MHZ} \approx 3.75ns$$

- 主存到 L2: 以频率 133 MHZ, 每周期传 128 bit, 即传输一个 64B 的块需要

$$4 \times \frac{1}{133MHZ} \approx 30ns$$

缺失代价:

- L1(I): $15ns + 2 \times 3.75ns = 22.5ns$
- L1(D): $15ns + 3.75ns = 18.75ns$
- L2: $(1 + 50\%) \times (60ns + 30ns) = 135ns$

a.

对于指令访问, $AMAT_I = 2\% \times [22.5ns + 0.2 \times 135ns] = 0.99ns$

b.

对于数据读取, $AMAT_{DR} = 5\% \times [18.75ns + 0.2 \times 135ns] = 2.2875ns$

c.

写失效时, 写回法通常使用写分配策略, 写直达通常使用非写分配策略.

只有当 L1(D) 写失效, 且用不了写缓存的时候, 才会有写入 L2 的额外开销, 比例是 $(1 - 95\%)$, 开销至少是 $18.75ns$

而 L2 采用写回+写分配策略的话, 如果 miss, 额外的写入开销来自于从 memory 调块(写分配)的开销 $135ns$, 比例是 20%

对于数据写入, $AMAT_{DW} = (1 - 95\%) \times (18.75ns + 0.2 \times 135ns) = 2.2875ns$

d.

整体 $CPI = \text{基本}CPI + \text{取指}CPI + \text{载入}CPI\text{代价} + \text{存储}CPI\text{代价}$

其中

- 基本 $CPI = 0.7$
- 取指 $CPI = 1.1GHz \times AMAT_I = 1.1GHz \times 2\% \times [22.5ns + 0.2 \times 135ns] \approx 1.09$
- 载入 $CPI\text{代价} = 20\% \times 1.1GHz \times 2.2875ns \approx 0.5$
- 存储 $CPI\text{代价} = 5\% \times 1.1GHz \times 2.2875ns \approx 0.125$

故 整体 $CPI = 0.7 + 1.09 + 0.5 + 0.125 = 2.415$

第二题

a.

只有首次循环的时候载入指令出现了 $\frac{64}{4} = 16$ 次缺失, 由于大量循环, 故可忽略不计, 渐进指令缺失率是 0%

b.

对于 192B 的循环, 有 48 条指令. 刚开始可以读入 $\frac{128}{4} = 32$ 条指令, 每次都缺失, 此后 16 条需要替换最早的 16 条;

进入下一轮循环后, 却需要载入最早的 16 条, 导致中间的 16 条被替换掉, 以此类推.

因此, 可以说, 每次访问都发生了缺失, 缺失率为 100%

对于 320B 的循环, 共有 80 条指令, 刚开始读入 0-31 条, 然后需要替换成 32-63 条, 再替换成 (64-79 和 0-15) \Rightarrow (16-47) \Rightarrow (48-79). 依然都是没有命中
所以缺失率也是 100%

c.

如果使用 MRU, 那么

对于 64B 的循环, 显然缺失率还是 0%

对于 192B 的循环,

访问前 32 条指令后, 之后 32-47 条需要把逐次把第 31 条换掉, 循环结束时 cache 里是 (0-30 和 47).

第二轮循环则前 31 条都能命中, 然后又不断替换最后一个.

可见对于大量该循环, 每次循环(48条指令)会发生 $48 - 31 = 17$ 次缺失, 这意味着渐进缺失率为 $\frac{17}{48} = 35.4\%$

对于 320B 的循环, 也是类似. 每次循环(80条指令)会发生 $80 - 31 = 49$ 次缺失, 这意味着渐进缺失率为 $\frac{49}{80} = 61.25\%$

综上, 192B 和 320B 的场景都能受益.

d.

考虑硬件层面针对循环模式和非循环模式优化. 对于循环模式采用 MRU, 否则采用其他算法 如 LRU.

第三题

a.

- **不使用**: 需要完整取完整块, $120 + \frac{64-16}{16} \times 16 = 168$ 个时钟周期
- **关键字优先**: 120 个时钟周期(取到需要的 16 个字节即可)

b.

根本决定因素是看哪种策略 AMAT 更小; 其中影响因素主要是块大小. 如果块比较大, 则关键字优先和提前重启动带来的收益比较大.

第四题

a.

可以匹配 L2 缓存写数据总线大小, 取 16B

b.

题述指令需要对 8B 数据操作, 若采用合并写缓冲, 则平均值需要 2 个时钟周期, 若采用非合并写缓冲, 则要处理 16 B 数据, 要 4 个时钟周期, 因此**加速比是 2**

c.

采用阻塞缓存, 则 L1 是否缺失并不影响所需写缓冲区大小, 因为一旦缺失必然阻塞;

若采用非阻塞缓存, 则 L1 更多的缺失将导致更多的数据要写在缓冲区, 所以缓冲区应该更大一些.