《计算机系统结构》作业 3

黄家晖 2014011330

一、“Cache-主存”层次。Cache 为 4 块，主存为 8 块。画出映像关系示意图，并计算访存地址为 5 时的 Cache 地址索引。

(1) 全相连

映像关系示意图：

（表格中第 n 行的灰色区块表示第 n 块内存区域可能被存入的 Cache 编号，例如 MEM[5]一行有四个黑色区块，就表示 Cache 的 4 个块均有可能存储内存第 5 块的内容。Cache 块编号中，逗号前代表组号，逗号后代表组中的编号）

Cache 块编号	0^{th}, 0#	0^{th}, 1#	0^{th}, 2#	0^{th}, 3#
MEM[0]				
MEM[1]				
MEM[2]				
MEM[3]				
MEM[4]				
MEM[5]				
MEM[6]				
MEM[7]				

访存地址为 5 的时候 Cache 地址索引可能为 0, 1, 2, 3。

(2) 组相连，每组两块

映像关系示意图：

（表格内容意义同上，下同）

Cache 块编号	0^{th}, 0#	0^{th}, 1#	1^{st}, 0#	1^{st}, 1#
MEM[0]				
MEM[1]				
MEM[2]				
MEM[3]				
MEM[4]				
MEM[5]				
MEM[6]				
MEM[7]				

访存地址为 5 的时候 Cache 地址索引可能为 2, 3。

(3) 直接映像

映像关系示意图：

Cache 块编号	0^{th}, 0#	1^{st}, 0#	2^{nd}, 0#	3^{rd}, 0#
MEM[0]				
MEM[1]				
MEM[2]				
MEM[3]				

MEM[4]				
MEM[5]				
MEM[6]				
MEM[7]				

访存地址为 5 的时候 Cache 地址索引可能为 1。

二、针对两个 32KB 指令-数据分离的 Cache 和一个 64KB 的混合 Cache，比较二者的失效率和平均访存时间。

对于指令-数据分离的 Cache,

失效率为：

$$\mathrm{M}_1 = 75\% \times 0.39\% + 25\% \times 4.82\% = 1.50\%$$

平均访存时间为：

$$\mathrm{t}_1 = 1 + M_1 \times 50 = 1.75$$

对于混合 Cache,

失效率为：

$$\mathrm{M}_2 = 1.35\%$$

平均访存时间为：

$$\mathrm{t}_2 = 2 + M_2 \times 50 = 2.675$$

相比之下，混合 Cache 的失效率更低；分离 Cache 的平均访存时间为 1.75s，混合 Cache 的平均访存时间为 2.675s。

三、计算和对比直接映像 Cache 和两路组相联 Cache 的平均访问时间和 CPU 性能。

设理想情况下，平均每条指令执行花费时间为$t = t_1 + t_2$，其中t_1是访存所用的时间，t_2是执行其他操作（例如运算、移位等流水线步骤）所用的时间。假设理想 Cache 不会失效，每次访存均命中，则

$$\mathrm{t}_1 = 1.2 \times 2ns = 2.4ns$$

又根据理想 CPI 为 2，可以得出：

$$\mathrm{t} = 4\mathrm{ns}$$

因此：

$$\mathrm{t}_2 = 4ns - 2.4ns = 1.6ns$$

对于 64KB 直接映像 Cache 来说，

平均访问所花费的时间为：

$$\mathrm{t}_{\mathrm{D,MA}} = 1.4\% \times 80\mathrm{ns} + 2\mathrm{ns} = 3.12\mathrm{ns}$$

平均每条指令的执行时间变为：

$$\mathrm{t}_\mathrm{D} = t_2 + 1.2 \times t_{D,MA} = 5.344ns$$

每秒执行指令条数为1.871×10^8条

对于 64KB 两路组相连 Cache 来说，

平均访问所花费的时间为：

$$\mathrm{t}_{\mathrm{L,MA}} = 1.0\% \times 80ns + 2\mathrm{ns} \times 110\% = 3\mathrm{ns}$$

平均每条指令的执行时间变为：

$$\mathrm{t}_\mathrm{L} = 1.1 \times t_2 + 1.2 \times \mathrm{t}_{\mathrm{L,MA}} = 5.36\mathrm{ns}$$

每秒执行指令条数为1.866×10^8条

根据计算结果可以看出，直接映像的 Cache 虽然在应用中失效率比组相联 Cache 高，每次访存的时间长于组相联，但是由于其设计可以简化电路，不需要多路选择器，使得整个 CPU

在执行指令的时候速度更快，最终导致其性能好于组相联 CPU。这种结果实际上是访存时间和时钟周期的 tradeoff，如果在实际的计算机系统设计过程中，还需要谨慎考虑其他因素对于最终 CPU 性能的影响，在多维空间中优化 CPU 的性能。

四、计算主存频带的平均使用比例。

在 1s 内，CPU 的读写请求数为10^9字。

(1) 使用写直达 Cache 时，

对于 75%的读访问操作，95%的请求可以直接在 Cache 中命中，无需访存；而剩余 5%的请求所需的访存操作次数为：

$$\mathrm{n_{read}} = \frac{3}{4} \times 10^9 \times 5\% \times 2 = \frac{3}{40} \times 10^9$$

上式中乘二是因为每次访存都要更新整个块 2 个字，所以需要访存两次。

对于 25%的写访问操作：其中 95%的命中请求，需要写入内存储器，访存一次；其中 5%的非命中请求，由于采用按写分配法，每次写入内存的时候，需要附加地读入一个块，总共 3 次访存，因此写操作地访存次数为：

$$\mathrm{n_{write-through}} = \frac{1}{4} \times 10^9 \times (95\% \times 1 + 5\% \times 3) = \frac{11}{40} \times 10^9$$

使用写直达 Cache 的主存频带使用比例为$\frac{\mathrm{n_{read}} + n_{write-through}}{10^9} = 35\%$

(2) 使用写回法 Cache 时，

对于 75%的读访问操作，与写直达 Cache 不同之处在于，每次读不命中之后将新块换入时，还会以 30%的几率将当前已经修改的块写入内存，所以所需访存操作次数为：

$$\mathrm{n'_{read}} = \frac{3}{4} \times 10^9 \times 5\% \times 2 \times (1 + 30\%) = \frac{39}{400} \times 10^9$$

对于 25%的写访问操作：其中 95%的命中请求，无需更新内存储器；其中 5%的非命中请求，首先要以 30%的概率将当前的块写入内存（因为任何时间总有 30%的块被修改过），再读入当前的块，期望访存次数为$2 + 30\% \times 2$次，因此写操作整体访存次数为：

$$\mathrm{n_{write-back}} = \frac{1}{4} \times 10^9 \times \big(5\% \times (2 + 30\% \times 2)\big) = \frac{13}{400} \times 10^9$$

综上所述，，写回法 Cache 的主存频带使用比例为$\frac{\mathrm{n_{read}'} + n_{write-back}}{10^9} = 13\%$。

实际上，在 PPT 中老师已经强调，使用写直达策略的 Cache 一般不采用写分配策略，因为新读入的块不会被下次写操作利用，这也是本题中写直达 Cache 的访存次数高于写回 Cache 访存次数的原因。