

体系结构第三次作业

第一题：B.5[10/10/10/10]<B.2>

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0.7。
- 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%。
- 512KB写回、统一L2缓存的块大小为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。

解: 计算各缓存缺失代价有

$$Cost_{L1,I-cache} = 15ns + 32bytes / (128bit \times 266MHz) = 22.5ns$$

$$Cost_{L1,D-cache} = 15ns + 16bytes / (128bit \times 266MHz) = 18.75ns$$

$$Cost_{L2-cache} = (60ns + 64bytes / (128bit * 133MHz)) * (1 + 50\%) = 135ns$$

a.[10]<B.2> 指令访问的存储器平均访问时间为多少?

$$\begin{aligned} AvgTime &= 0 + 2\% \times (Cost_{L1,I-cache} + (1 - 80\%) \times Cost_{L2-cache}) \\ &= 2\% \times (22.5ns + 20\% \times 135ns) \\ &= 0.99ns \end{aligned}$$

b.[10]<B.2> 数据读取的存储器平均访问时间为多少?

$$\begin{aligned} AvgTime &= 0 + 5\% \times (Cost_{L1,D-cache} + (1 - 80\%) \times Cost_{L2-cache}) \\ &= 5\% \times (18.75ns + 20\% \times 135ns) \\ &\approx 2.2875ns \end{aligned}$$

c.[10]<B.2> 数据写入的存储器平均访问时间为多少?

$$\begin{aligned} AvgTime &= 0 + (1 - 95\%) \times (Cost_{L1,D-cache} + (1 - 80\%) \times Cost_{L2-cache}) \\ &= 5\% \times (18.75ns + 20\% \times 135ns) \\ &\approx 2.2875ns \end{aligned}$$

d.[10]<B.2> 包括存储器访问在内的整体CPI为多少?

$$\begin{aligned} CPI_{total} &= CPI_0 + (0.99ns + 20\% \times 2.2875ns + 5\% \times 2.2875ns) \times 1.1ns^{-1} \\ &= 2.41875 \end{aligned}$$

第二题：B.8[20/20/15/25]<B.3>

LRU替换策略基于以下假定：如果最近访问地址A1的频率低于地址A2，那么未来再次访问A2的时机要早于A1。因此，为A2指定了高于A1的优先级。试讨论，当一个大于指令缓存的循环连续执行时，这一假定为什么不成立。例如，考虑一个全相联128字节指令缓存，其块大小为4个字节（每个块正好可以容纳一条指令）。此缓存使用LRU替换策略。

a.[20]<B.3>对于一个拥有大量迭代的64字节循环，渐近指令缺失率是多少？

解：渐近缺失率为0。设执行n次指令而n趋向于无穷，这一情景下，第一轮循环后64字节将全部放入缓存中，之后的循环指令不再缺失，所以渐近指令缺失率为

$$\lim_{n \rightarrow \infty} 16/n = 0$$

b.[20]<B.3>对于大小为192字节和320字节的循环，重复(a)部分

解：192字节的情况下，第一轮循环到128字节后，这些指令全部放入缓存中，之后从第一个块被替换开始，后面的64字节依次替换掉cache中原来从1至64字节对应的那些块，下一个循环开始后，下一个循环的第一个块紧接着替换掉65-68字节的块，以此类推，缓存以FIFO队列的形式，反复将所有块入列出列，因此缺失率为100%，320字节的情况下同理也是100%。

c.[15]<B.3>如果缓存替换策略改为最近使用最多(MRU)（替换最近访问最多的缓存行），以上三种情景（64、192、320字节的循环）中的哪一种情景将因为这一策略而受益？

解：此处考虑按照最近使用次数最多的字面含义，设“最近”的窗口大小为1.5个LOOP，那么64字节的情况不受影响。只要LOOP达到一定次数，192字节的情况下，避免了反复替换所有的块，指令缓存中31个块都将不需要被替换，所以渐近指令缺失率降低到了 $17/48 = 35.4\%$ 。320字节的情况与192字节类似，缺失率优化为 $(80 - 31)/80 = 61.25\%$ 。都受益了。

考虑其“Most Recently Used”的字面含义，每次取最近被访问的块替换，那么64字节的情况仍不受影响，当LOOP达到一定次数后，192字节的情况下，也能部分规避一个循环里反复替换所有的块，并且这种规避效果呈一定的周期性，每33个LOOP里，32个连续的LOOP缺失率均为 $(48 - 32)/48$ ，而余下的一个LOOP缺失率为 $(48 - 31)/48$ ，当LOOP数趋向于无穷时，渐近指令缺失率被优化到了 $(32 \times 16 + 17)/(33 \times 48) \approx 33.40\%$ ，对于320字节的情况则为 $(32 \times (80 - 32) + (80 - 31))/(33 \times 80) \approx 60.04\%$ ，都受益了。

综上，192和320字节的情景将因为该策略受益。

d.[25]<B.3>提出执行性能可能优于LRU的更多替换策略。

解：基于寿命周期替换，为每个块维护一个寿命周期，如果有块已经到达寿命，则优先替换该块，否则选择替换最近访问最多的。如此，只要合理配置寿命长度（比如330），可以适用于以上三种情况。

第三题：2.11[12/15]<2.2>

考虑在L2缓存缺失时使用关键字优先和提前重启动。假定L2缓存的容量为1MB、块大小为64字节、填充路径宽16字节。假定能够以每4个处理器周期16个字节的速度写入L2，从存储器控制器接收前16个字节块的时间为120个周期，每从主存储器接收另外16个字节的块需要16个周期，也可以直接将数据传送给L2缓存的读取端口。忽略向L2缓存发送缺失请求及向L1缓存传送被请求数据的周期数。

a.[12]<2.2>在使用、不使用关键字优先和提前重启动时，为L2缓存缺失提供服务分别需要多少个周期？

解：使用关键字优先和提前重启动时，需要120个周期，不使用关键字优先和提前重启动时，需要依次把整个块放进来，总共需要 $120 + 16 \times (64 - 16) / 16 = 168$ 个周期

b.[15]<2.2>你是否认为关键字优先和提前重启动对于L1缓存或L2缓存更重要一些，哪些因素影响它们的相对重要性？

解：就此题情况而言，忽略其余因素，单纯看块的大小，那么是L2更重要一些。因为只有才采用大型缓冲块的设计中，关键字优先和提前重启动的设计才能体现出较大的优化效果。

相对重要性需要具体看两个缓存的块大小、延迟和传输速度。

第四题：2.12[12/12]<2.2>

在直写L1缓存与写回L2缓存之间设计一个写缓冲区。L2缓存写数据总线的宽度为16B，可以每4个处理器周期向一个独立缓存地址执行一次写操作。

a.[12]<2.2>每个写缓冲区项目应当为多少字节？

解：应当为16字节。

b.[15]<2.2>如果所有其他指令可以与存储指令并行发射，块存在于L2缓存中，在通过执行64位存储指令将存储器置零时，使用一个合并写缓冲区来代替非合并缓冲区，在稳定状态下可以得到什么样的加速比？

解：若合并写缓冲区，稳定情况下预计消耗2个周期将两个数据填入（ $16B / 64bit = 2$ ），之后被写入存储器，整个操作包括在一个4周期的时间范围内。而若不使用，达到相同的效果则至少需要两个4周期来分别完成两个64位存储。所以加速比为2。

c.[15]<2.2>对于采用阻塞缓存与非阻塞缓存的系统，可能出现的L1缺失对于所需写缓冲区项目的个数有什么样的影响？

解：阻塞缓存的情况下没有影响，采用非阻塞缓存的情况下，写缓冲区项目个数应当不少于最多可能出现的L1缺失数。