1. 一标准流水线 CPU 的理想 CPI 为 1，其执行的程序片段中 load 和 store 类指令占 30， 其余指令均为 R 型指令。主存延迟为 50 个时钟周期。L1 级 cache 性能为：命中时间为1 个时钟周期，缺失率为 2。L2 级 cache 性能为：命中时间为 10 个时钟周期，缺失率为 5。针对如下情况分别计算流水线的实际 CPI：没有 cache；只有 L1 级 cache；L2 级cache。
2. 由于没有 cache，因此指令和数据的缺失率均为 100%；

暂停周期数 = 指令缺失率× 指令缺失代价+ 数据缺失率× 数据缺失代价

= 50 + 30% × 50 = 65

CPI无 = CPI基准 + 暂停周期数 = 1 + 65 = 66

1. 只有 L1 级 cache

暂停周期数 = 指令缺失率× 指令缺失代价+ 数据缺失率× 数据缺失代价

= 2% × 50 + 30% × 2% × 50

= 1 + 0.3

= 1.3

CPIL1 = CPI基准 + 暂停周期数 = 1 + 1.3 = 2.3

1. 有 L2 级 cache：

由于不分指令 cache 和数据 cache，因此 L1 指令缺失代价与 L1 数据缺失代价是相同的。

暂停周期数 = 指令缺失率× L1 指令缺失代价+ 数据缺失率× L1 数据缺失代价

= 2% × L1 缺失代价+ 30% × 2% × L1 缺失代价

= (2% + 0.6%) × (10 + 5% × 50)

= 0.026 × 12.5

= 0.325

CPIL2 = CPI基准 + 暂停周期数 = 1 + 0.325 = 1.325

1. 直接映射 cache 参数如下：cache 的数据容量为 16KB，cache 块为 16B。现该 cache 接入一个 32 位 cpu。请给出 cache 的 TIO 结构以及包含标记后的 cache 块容量。

1) Offset：log2(16)=4

2) Index：16KB/16=1024 块，log2(1024)=10

3) Tag：32-4-10=18

4) cache 块总容量：数据+Tag+Valid=16B+18b+1b=16B+19b

1. 组相联映射 cache 参数如下：cache 的数据容量为 512KB，16 路相联，cache 块为 32B。现该 cache 接入一个 32 位 cpu。请给出 cache 的 TIO 结构。
   1. offset：log2(32)=5
   2. Index：512KB/32B=16K 块，16K/16=1K 组，log2(1K)=10
   3. Tag：32-5-10=17
2. 对于第 3 问的系统，cache 采用写回策略，cache 命中时间为 1 个时钟周期，从主存调入一块或回写一块的代价均为 100 个时钟周期。某程序片段将起始地址为 0000\_0000h 的 1MB 内存单元全部初始化为 0，代码如下所示。

int \*p=0x0 ;

for ( int i=0; i<1024\*1024/4; i++ )

\*p++ = 0 ;

* 1. 计算该程序片段中的数据访问缺失率（忽略指令缺失等因素）。
  2. 计算该程序片段的数据访问实际需要多少个时钟周期。
  3. 计算程序片段的存储访问次数与实际时钟周期数之比。

答案：

1)

* 1. 1MB 主存对应的块数：1MB/32B=32K（块）
  2. 由于是线性初始化，因此这 32K 块均会访问 1 次。可以看成是如下循环：
  3. 第 1 次循环：
     1. 当初始化主存块 0 字 0 时，缺失。于是从将主存块 0 写入 cache 的组 0 块

0。之后的 7 个字（8 字-1 字）均命中。

* + 1. 当初始化主存块 1 字 0 时，缺失。于是从将主存块 1 写入 cache 的组 1 块

0。之后的 7 个字（8 字-1 字）均命中。

* + 1. 类似的，主存块 15，映射到 cache 组 15 块 0。
  1. 第 2 次循环：
     1. 主存块 16 字 0 时，缺失，会被映射到 cache 组 0 块 1；之后 7 个字均命中。
     2. 依次类推，主存块 31 字 0 时，缺失，会被映射到 cache 组 15 块 1。之后

7 个字均命中。

* 1. 重复上述过程，直至第 1K 次循环。在第 1K 次循环中，主存块（16K-16）至主存块（16K-1）被分别装入 cache 组 0 块 15 和 cache 组 15 块 15。
  2. 至此，主存的前 16K 块已经访问结束了，cache 所有块均被写满了，再无空闲块了。
  3. 而主存的后 16K 块的行为与前 16K 块完全一样，区别在于需要把前 16K 块写回主存。
  4. 主存访问时，其对应的每个主存块的访问情况完全相同：字 0 缺失，字 1~字

7 均命中。因此，数据访问缺失率=1/8=12.5%

2)

1. 对于每个块来说，有 1 次缺失，其代价是 100 个时钟周期；7 次命中，7 个周期。故每个块是 100+7=107 个时钟周期。
2. 总共有 32K 块，因此访问时间=107\*32K（时钟周期）。
3. 但是后 16K 块还会导致 16K 次替换，每次替换都会有 100 时钟周期，因此替换时间=100\*16K（时钟周期）

iv. 总时间=107\*32K+100\*16K=5024K（时钟周期）

3)

* 1. 程序总存储访问次数=1MB/4=256K
  2. 程序总存储访问次数：实际时钟周期数=256K：5024K≈1:20
  3. 意味着每 20 个时钟周期才能完成一次存储访问。

1. 假设某系统的虚页和物理页尺寸均为 8KB，40 位虚地址，物理主存容量为 32GB。请问虚页号与物理页号的位数分别是多少。
   1. 8KB 页面，则页内地址为 13 位。

2) 虚页号（VPN）位数=40-13=27 位

3) 物理页号（PPN）位数=log2(32G)-13=35-13=22 位

1. 设计师给第 5 问的系统设计了 TLB。TLB 采用 2 路组相联结构，共有 256 个页表项。TLB 中的每个页表项结构如下图所示。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Valid | Dirty | 访问权限 | TLB Tag | PPN |
| 1 位 | 1 位 | 2 位 |  |  |

* 1. 请问每个页表项的位数是多少。
  2. 页表的总存储容量是多少位？
  3. 操作系统设计团队希望将页面尺寸从 8KB 降低至 4KB，但硬件设计团队认为会增加硬件开销，因此不同意改变页面尺寸。你作为硬件设计师团队负责人，请陈述理由。

答案：

1) TLB 的组数=256/2=128 组

1. TLB 的 index 位数=log2(128)=7
2. TLB 的 tag 位=虚地址位数-页内偏移位数-index 位数=40-13-7=20 位

4) 页表项位数=1+1+2+20+22=46 位

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Valid | Dirty | 访问权限 | TLB Tag | PPN |
| 1 位 | 1 位 | 2 位 | 20 | 22 |

1. 现有 4K × 8 位容量的 DRAM 存储芯片。
   1. DRAM 芯片内置译码器输出的行/列选择线各是多少条。
   2. DRAM 芯片内置的刷新地址计数器位数为多少位。
   3. DRAM 每刷新周期刷新一行存储单元是多少位。
   4. 主存容量为 8K × 8 位，需使用 DRAM 芯片的数量。
   5. 给出每个 DRAM 芯片的片选控制信号的逻辑表达式。CPIL2 = CPI基准 + 暂停周期数 = 1 + 0.325 = 1.325答案：
2. DRAM 芯片共计 4K 个单元，因此内部共需地址线的位数=log2(4K)=12，即内部地址为 A[11:0]。由于 DRAM 采用行地址和列地址共享，因此行/列地址选择线均为

12/2=6。

1. 由于 DRAM 每次刷新一行，因此刷新地址计数器位数与行地址的位数相同，即刷新地址计数器位数为 6 位。
2. 一行的单元数由列地址数决定，即一行单元数=26=64 个。故总的存储单元位数为

64x8=512 位。

1. 需要的 DRAM 芯片数量=(8K x 8）/ (4K x 8)=2 片。
2. 主存地址为 A[12:0]。每片 DRAM 芯片的位数为 A[11:0]，因此 A12 用来选择 2 个芯片。假设芯片 1 对应 0~4K，芯片 2 对应 4K-8K，则：

芯片 0 的片选=!A12 芯片 1 的片选= A12