体系结构第五次作业

第一题

在以下的循环中，找出所有真相关、输出相关和反相关。通过重命名来消除输出相关和反相关。

```
for(i = 0;i < 100;i++){
    A[i] = A[i] * B[i];     /*S1*/
    B[i] = A[i] + c;        /*S2*/
    A[i] = C[i] * c;        /*S3*/
    C[i] = D[i] * A[i];     /*S4*/
}
```

解答：

- **真相关：**
  - 由于A[i]，S1和S2、S3和S4之间存在真相关
- **输出相关：**
  - 由于A[i]，S1和S3之间存在输出相关
- **反相关：**
  - 由于B[i]，S1和S2之间存在反相关
  - 由于A[i]，S2和S3之间存在反相关
  - 由于C[i]，S3和S4之间存在反相关

消除相关后代码：

```
//假设At、Bt、Ct复制了A、B、C的内容
for (i=0;i<100;i++) {
    X[i] = At[i] * Bt[i];   /*S1*/
    B[i] = X[i] + c;        /*S2*/
    A[i] = Ct[i] * c;       /*S3*/
    C[i] = D[i] * A[i];     /*S4*/
}
```

第二题 [10/20/20/15/15]<4.2>

考虑以下代码，它将两个包含单精度复数值的向量相乘:

```
for(i = 0;i < 300;i++){
    c_re[i] = a_re[i] * b_re[i] - a_im[i] * b_im[i];
    c_im[i] = a_re[i] * b_im[i] + a_im[i] * b_re[i];
}
```

假定处理器的运行频率为 $700\mathrm{MHz}$，最大向量长度为 64 。载入 / 存储单元的启动开销为 15 个时钟周期，乘法单元为 8 个时钟周期，加法 / 减法单元为 5 个时钟周期。

a. [10] < 4.2 >

这个内核的运算密度为多少？给出理由。

解：指令序列每六次FLOP中执行4次浮点读和2次浮点写，所以其运算密度为 $(4+2)/6=1$ 。

b. [20] < 4.2 >

将此循环转换为使用条带挖掘的 $VMIPS$ 汇编代码。

解： 依题目假设有

```
        li          $VL     ,44
        li          $r1     ,0
        lv          $v1     ,a_re + $r1
        lv          $v3     ,b_re + $r1
        mulvv.s     $v5     ,$v1    ,$v3
        lv          $v2     ,a_im + $r1
        lv          $v4     ,b_im + $r1
        mulvv.s     $v6     ,$v2    ,$v4
        subvv.s     $v5     ,$v5    ,$v6
        sv          $v5     ,c_re + $r1
        mulvv.s     $v5     ,$v1    ,$v4
        mulvv.s     $v6     ,$v2    ,$v3
        addvv.s     $v5     ,$v5    ,$v6
        sv          $v5     ,c_im + $r1
        addi        $r1     ,$r1    ,#44
LOOP:   lv          $v1     ,a_re + $r1
        lv          $v3     ,b_re + $r1
        mulvv.s     $v5     ,$v1    ,$v3
        lv          $v2     ,a_im + $r1
        lv          $v4     ,b_im + $r1
        mulvv.s     $v6     ,$v2    ,$v4
        subvv.s     $v5     ,$v5    ,$v6
        sv          $v5     ,c_re + $r1
        mulvv.s     $v5     ,$v1    ,$v4
        mulvv.s     $v6     ,$v2    ,$v3
        addvv.s     $v5     ,$v5    ,$v6
        sv          $v5     ,c_im + $r1
        addi        $r1     ,$r1    ,#256
        blt         $r1     ,1200   ,LOOP
```

c. [20] < 4.2 >

假定采用链接和单一存储器流水线，需要多少次 Chime？每个复数结果值需要多少个时钟周期?

解答： 6次。每次结果值需要$(6\times 64+15\times 6+8\times 4+5\times 2)/128=516/128=4$ 个周期。

第三题 [10/15] <4.4>

假定有一种包含 10 个 SIMD 处理器的 GPU 体系结构。每条 SIMD 指令的宽度为 32，每个 SIMD 处理器包含 8 个车道，用于执行单精度运算和载入 / 存储指令，也就是说，每个非分岔 SIMD 指令每 4 个时钟周期可以生成 32 个结果。假定内核的分岔分支将导致平均 80% 的线程为活动的。假定在所执行的全部 SIMD 指令中， 70% 为单精度运算、20\%为载入 / 存储。由于并不包含所有存储器延迟, 所以假定 SIMD 指令平均发射率为 0.85 。假定 GPU 的时钟速度为 1.5 GHz 。

a. [10] <4.4>

计算这个内核在这个 GPU 上的吞吐量，单位为 GFLOP/s 。

解答：罗列如下

$$1.5\ \mathrm{GHz} \times 0.8 \times 0.85 \times 0.7 \times 10 \times 32/4 = 57.12\ \mathrm{GFLOPs/s}$$

b. [15] <4.4>

解答： 假定我们有以下选项：

- **(1)** 将单精度车道数增大至 16 。

$$1.5\ \mathrm{GHz} \times 0.8 \times 0.85 \times 0.7 \times 10 \times 32/2 = 114.24\ \mathrm{GFLOPs/s}$$

- **(2)** 将 SIMD 处理器数增大至 15 (假定这一改变不会影响所有其他性能度量，代码会扩展到增加的处理器上) 。

$$1.5\ \mathrm{GHz} \times 0.8 \times 0.85 \times 0.7 \times 15 \times 32/4 = 85.68\ \mathrm{GFLOPs/s}$$

- **(3)** 添加缓存可以有效地将存储器延迟缩减 40% ， 这样会将指令发射率增加到 0.95 ， 对于这些改进中的每一项，吞吐量的加速比为多少?

$$1.5\ \mathrm{GHz} \times 0.8 \times 0.95 \times 0.7 \times 10 \times 32/4 = 63.84\ \mathrm{GFLOPs/s}$$

第三种是最好的

第四题 [10] <4.4>

假定一个虚设 GPU 具有以下特性

- 时钟频率为 $1.5\mathrm{GHz}$；
- 包含 16 个 SIMD 处理器，每个处理器包含 16 个单精度浮点单元；
- 片外存储器带宽为 100 GB/s 。

不考虑存储器带宽，假定所有存储器延迟可以隐藏，则这一 GPU 的峰值单精度浮点吞吐量为多少 $\mathrm{GFLOP/s}$ ？ 在给定存储器带宽限制下，这一吞吐量是否可持续？

解答： 峰值单精度浮点吞吐量为 $1.5 \times 16 \times 16 = 384\ \mathrm{GFLOP/s}$ ，若想始终保持该吞吐量，存储器带宽应当达到 $384\ \mathrm{GFLOP/s} \times 3 \times 4 bytes/\mathrm{FLOP} = 4.6\ \mathrm{TB/s}$ ，远超本题的存储器带宽，所以吞吐量不可持续。