

中国科学技术大学计算机学院

《计算机体系结构》作业

2021.06.07



作业题目：第六章作业

学生姓名：胡毅翔

学生学号：PB18000290

计算机实验教学中心制

2019 年 9 月

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。

```

1  for(i = 0; i < 100; i++){
2      A[i] = A[i] * B[i];    /* S1 */
3      B[i] = A[i] + c;      /* S2 */
4      A[i] = C[i] * c;      /* S3 */
5      C[i] = D[i] * A[i];   /* S4 */
6  }

```

解：

真相关：S1 到 S2 基于 A[i]，S3 到 S4 基于 A[i]。

输出相关：S1 到 S3 基于 A[i]。

反相关：S3 到 S4 基于 C[i]。

修改代码如下：

```

1  for(i = 0; i < 100; i++){
2      T[i] = A[i] * B[i];
3      B1[i] = T[i] + c;
4      A1[i] = C[i] * c;
5      C1[i] = D[i] * A1[i];
6  }

```

第二题 考虑以下代码，它将两个包含单精度复数值的向量相乘：

```

1  for ( i = 0; i < 300; i++){
2      c_re[i] = a_re[i] * b_re[i] - a_im[i] * b_im[i];
3      c_im[i] = a_re[i] * b_im[i] + a_im[i] * b_re[i];
4  }

```

假定处理器的运行频率为 700MHz，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) 这个内核的运算密度为多少？给出理由。

解：

这一代码每 6 个 FLOP，读 4 个浮点数，写 2 个浮点数。所以，运算密度 = $6/6 = 1$ 。

(b) 将此循环转换为使用条带挖掘的 VMIPS 汇编代码。

解：

假设 $MVL = 64$ ：

```

1      li      $VL,44          # perform the first 44 ops
2      li      $r1,0          # initialize index

```

```

3  loop:
4      lv      $v1,a_re+$r1    # load a_re
5      lv      $v3,b_re+$r1    # load b_re
6      mulvv.s $v5,$v1,$v3     # a_re * b_re
7      lv      $v2,a_im+$r1    # load a_im
8      lv      $v4,b_im+$r1    # load b_im
9      mulvv.s $v6,$v2,$v4     # a_im * b_im
10     subvv.s $v5,$v5,$v6     # a_re * b_re - a_im * b_im
11     sv      $v5,c_re+$r1    # store c_re
12     mulvv.s $v5,$v1,$v4     # a_re * b_im
13     mulvv.s $v6,$v2,$v3     # a_im * b_re
14     addvv.s $v5,$v5,$v6     # a_re * b_im + a_im * b_re
15     sv      $v5,c_im+$r1    # store c_im
16     bne     $r1,0,else      # check if first iteration
17     addi    $r1,$r1,#44     # first iteration, increment by 44
18     j loop      # guaranteed next iteration
19 else:
20     addi    $r1,$r1,#256    # not first iteration, increment by 256
21 skip:
22     blt     $r1,1200,loop   # next iteration?

```

- (c) 假定采用链接和单一存储器流水线，需要多少次钟鸣？每个复数结果值需要多少个时钟周期（包括启动开销在内）？

解：

```

mulvv.s  lv      # a_re * b_re, load a_im
lv      mulvv.s  # load b_im, a_im * b_im
subvv.s  sv      # subtract and store c_re
mulvv.s  lv      # a_re * b_im, load next a_re vector
mulvv.s  lv      # a_im * b_re, load next b_re vector
addvv.s  sv      # add and store c_im

```

需要 6 次钟鸣。

每次循环的周期数 = $6\text{chimes} \times 64\text{elements} + 15\text{cycles}(\text{load/store}) \times 6 + 8\text{cycles}(\text{multiply}) \times 4 + 5\text{cycles}(\text{add/subtract}) \times 2 = 516$

每个复数结果值需要的周期数 = $516/128 = 4$

- (d) 如果向量序列被链接在一起，每个复数结果值需要多少个时钟周期（包含开销）？

解：

每次循环的周期数 = $6\text{chimes} \times 64\text{elements} + 15\text{cycles}(\text{load/store}) \times 6 + 8\text{cycles}(\text{multiply}) \times 4 + 5\text{cycles}(\text{add/subtract}) \times 2 = 516$

每个复数结果值需要的周期数 $= 516/128 = 4$

第三题 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 1.5GHz。

(a) 计算这个内核在这个 GPU 上的吞吐量，单位为 GFLOP/s。

解：

吞吐量为 $1.5\text{GHz} \times 0.80 \times 0.85 \times 0.70 \times 10\text{cores} \times 32/4 = 57.12\text{GFLOP/s}$ 。

(b) 假定我们有以下选项：

- i. 将单精度车道数增大至 16。
- ii. 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)。
- iii. 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至 0.95，对于这些改进中的每一项。

吞吐量的加速比为多少？

解：

- i. $1.5\text{GHz} \times 0.80 \times 0.85 \times 0.70 \times 10\text{cores} \times 32/2 = 114.24\text{GFLOP/s}$
 $\text{Speedup} = 114.24/57.12 = 2$
- ii. $1.5\text{GHz} \times 0.80 \times 0.85 \times 0.70 \times 15\text{cores} \times 32/4 = 85.68\text{GFLOP/s}$
 $\text{Speedup} = 85.68/57.12 = 1.5$
- iii. $1.5\text{GHz} \times 0.80 \times 0.95 \times 0.70 \times 10\text{cores} \times 32/4 = 63.84\text{GFLOP/s}$
 $\text{Speedup} = 63.84/57.12 = 1.11$

第四题 假定一个虚设 GPU 具有以下特性：

- 时钟频率为 1.5GHz;
- 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 100GB/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 GFLOP/s? 在给定存储器带宽限制下，这一吞吐量是否可持续？

解：

该 GPU 的峰值吞吐量为 $1.5 \times 16 \times 16 = 384\text{GFLOP/s}$ 的单精度吞吐量。但是假设每个单精度操作都需要两个 4 字节操作数并输出一个 4 字节结果。维持这一吞吐量（假设没有时间局限性）将需

要 $12\text{bytes}/\text{FLOP} \times 384\text{GFLOP}/\text{s} = 4.6\text{TB}/\text{s}$ 的片外存储器带宽。因此，这种吞吐量是不可持续的，但当使用片上存储器时，仍然可以在短时间内实现。