

一、假定处理器运行频率为 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。在该处理器上进行如下运算，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];
}
```

(1) 这个内核的运算密度为多少（注：运算密度指运行程序时执行的浮点运算数除以主存储器中访问的字节数）？

(2) 将此循环转换为使用条带挖掘 (Strip Mining) 的 VMIPS 汇编代码。

(3) 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）？

(4) 现在假定处理器有三条存储器流水线和链接。如果该循环的访问过程中没有组冲突，每个结果需要多少个时钟周期？

(1) 每 6 个浮点运算 (4 次乘 + 1 次加 + 1 次减，对应一个循环)，需读取 4 个单精度浮点数，最后会写回两个单精度浮点数，且每个单精度浮点数占 4 Byte

$$\therefore \text{运算密度} = \frac{6 \times 300}{4 \times 6 \times 300} = \frac{1}{4}$$

(2) 设 MVL = 64

```
li    $V1, 44
li    $r1, 0    # 下标初始化
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    # 存回 c_re
      mulvv.s $V5, $V1, $V4
      mulw.s  $V6, $V2, $V3
```

```
addvv.s $V5, $V5, $V6
sv    $V5, c_im + $r1    # 存回 c_im
bne    $r1, 0, else
addi    $r1, $r1, 4
j      loop
else:    addi    $r1, $r1, 4
skip:    blt    $r1, 1200, loop
```

见右边

(3)

mulvv.s	lv	# 计算 $a-re * b-re$, load $a-im$
lv	mulvv.s	# load $b-im$, 计算 $a-im * b-im$
subvv.s	sv	# $a-re * b-re - a-im * b-im$, 存储 $c-re$
mulvv.s	lv	# 以下类似
mulvv.s	lv	
addvv.s	sv	

共需6次钟鸣

每个数结果位需

$$\frac{6(\text{chimes}) \times 64 + 15(\text{load/store}) \times 6 + 8(\text{multiply}) \times 4 + 5(\text{add/subtract}) \times 2}{2 \times 64}$$

$$= 4.03125 \text{ 个时钟周期}$$

(4) 69

二、假定一个虚拟 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

(1) 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？

(2) 在给定存储器带宽限制下，这一吞吐量是否可持续？

(1) $1.5 \text{ GHz} \times 16 \times 16 = 384 \text{ GFLOP/s}$ \rightarrow 即每秒可完成 384 亿次运算操作

(2) 不可持续，理由如下：

每个单精度运算需 读 2 个操作数，写出 1 个数，每个单精度浮点数 为 4 Bytes，

\therefore 共操作 12 Bytes \therefore 若维持(1)中吞吐量时需有带宽 $12 \text{ Bytes} \times 384 \text{ GFLOP/s} = 4608 \text{ GB/s}$
 $> 100 \text{ GB/s}$

\therefore 不可