Homework Assignment 4

For chapter 4

College of Computer Science, Zhejiang University

推荐时间：1-3h

补充题也需要写，占分

1. Consider the following code, which multiplies two vectors that contain single-precision complex values:

```
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];
}
```

Assume that the processor runs at 700 MHz and has a maximum vector length of 64. The load/store unit has a start-up overhead of 15 cycles; the multiply unit, 8 cycles; and the add/subtract unit, 5 cycles.

a. [10] What is the arithmetic intensity of this kernel? Justify your answer.

注： Arithmetic intensity is the ratio of floating-point operations per byte of memory accessed. 单位是 FLOP/byte。

b. [20] Convert this loop into RV64V assembly code using strip mining.

假设 a0=300, a1=a_re, a2=a_im, a3=b_re, a4=b_im, a5=c_re, a6=c_im

注：RISCV-V 指令集手册：https://github.com/riscv/riscv-v-spec/releases/download/v1.0/riscv-v-spec-1.0.pdf

有关 RISCV-V 指令集，详见 12.26 的讲解 PPT。

c. [20] Assuming chaining and a single memory pipeline, how many chimes are required? How many clock cycles are required per complex result value, including start-up overhead?

d. [15] If the vector sequence is chained, how many clock cycles are required per complex result value, including overhead?

e. [15] Now assume that the processor has three memory pipelines and chaining. If there are no bank conflicts in the loop's accesses, how many clock cycles are required per result?

2. [30] In this problem, we will compare the performance of a vector processor with a hybrid system that contains a scalar processor and a GPU-based coprocessor. In the hybrid system, the host processor has superior scalar performance to the GPU, so in this case all scalar code is executed on the host processor while all vector code is executed on the GPU. We will refer to the first system as the vector computer and the second system as the hybrid computer. Assume that your target application contains a vector kernel with an arithmetic intensity of 0.5 FLOPs per DRAM byte accessed; however, the application also has a scalar component that must be performed before and after the kernel in order to prepare the input vectors and output vectors, respectively. For a sample dataset, the scalar portion of the code requires 400 ms of execution time on both the vector processor and the host processor in the hybrid system. The kernel reads input vectors consisting of 200 MB of data and has output data consisting of 100 MB of data. The vector processor has a peak memory bandwidth of 30 GB/s and the GPU has a peak memory bandwidth of 150 GB/s. The hybrid system has an additional overhead that requires all input vectors to be transferred between the host memory and GPU local memory before and after the kernel is invoked. The hybrid system has a direct memory access (DMA) bandwidth of 10 GB/s and an average latency of 10 ms. Assume that both the vector processor and GPU are performance bound by memory bandwidth. Compute the execution time required by both computers for this application.

3. In this exercise, we will examine several loops and analyze their potential for parallelization.

a. [10] Does the following loop have a loop-carried dependency?

```
for (i = 0; i < 100; i++) {
    A[i] = B[2 * i + 4];
    B[4 * i + 5] = A[i];
```

```
}
```

b. [15] In the following loop, find all the true dependences, output dependences, and antidependences. Eliminate the output dependences and antidependences by renaming.

```
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 */
}
```

c. [15] Consider the following loop:

```
for (i = 0; i < 100; i++) {
    A[i] = A[i] + B[i];     /* S1 */
    B[i + 1] = C[i] + D[i]; /* S2 */
}
```

Are there dependences between S1 and S2? Is this loop parallel? If not, show how to make it parallel.