Politecnico di Milano

Dipartimento di Elettronica, Informazione e Bioingegneria

prof. Fabrizio Ferrandi

Parallel Computing– II part–Thursday, January 12th, 2023

Polimi ID _____ _____

Surname _____ _____ **Name** _____

- This is a closed-book examination. You cannot use computers, phones, or laptops during the exam.
- Paper will be provided, but you should bring and use writing instruments that yield marks dark enough to be read easily. Erasable pens can be used.
- Total available time: 1h:30m.

Exercise 1 (4 points) _____

Exercise 2 (4 points) _____

Exercise 3 (4 points) _____

Exercise 3 (4 points) _____

Exercise n. 1

Answer the following questions about parallel patterns and briefly explain (without an explanation, the answer will be considered invalid)

A. Please define what a parallel pattern reduction is: assumptions, complexity, and example. (1)

B. Please describe how a parallel pattern Bin works. (1)

C. Please elaborate on issues coming from data structure based on arrays of structure vs structure of arrays. (1)

D. Please describe how a Pack parallel pattern could be implemented (1)

Exercise n. 2

Answer the following questions about parallel programming models (without an explanation, the answer will be considered invalid).

A. Describe three scenarios in which a combination of different parallel programming languages is suitable. (1)

B. What is the difference between joining and waiting on a barrier in Pthreads? (1)

C. What error will arise compiling the following code snippet? (1)

```
#pragma omp parallel for private(x) reduction(+:sum) default(none)
for (i=1;i<= num_steps; i++)
{
  x = (i-0.5)*step;
  sum = sum + 4.0/(1.0+x*x);
}
```

D. Describe the main advantages of OpenMP over Pthreads. (1)

Exercise n. 3

Answer the following questions about GPGPU computing with CUDA (without an explanation, the answer will be considered invalid).

A. Describe the types of memory available on NVIDIA GPUs and who can access each of them. (1)

B. The expression to map thread index to data index for a kernel that needs to process an element of a 1D vector is i=blockIdx.x + threadIdx.x. Is this statement true or false? Explain. (1)

C. Describe the impact of CUDA atomic operations on the throughput of a program. (2)

Exercise n. 4

Answer the following questions about Domain-Specific Languages and Halide (without an explanation, the answer will be considered invalid).

A. What is a DSL? (1)

B. Describe what is the effect of the following Halide schedule: (3)

```
gradient.tile(x, y, x_outer, y_outer, x_inner, y_inner, 4, 4);
gradient.fuse(x_outer, y_outer, tile_index);
gradient.parallel(tile_index);
```