Politecnico di Milano

Dipartimento di Elettronica, Informazione e Bioingegneria

prof. Fabrizio Ferrandi

Parallel Computing– II part– Wednesday, January 10th, 2024

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:00m.

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 parallel pattern gather: size of the collection, complexity, and example. (1)

B. Please describe how DOALL loops could be parallelized using a known parallel pattern. (1)

C. Please describe how the shift parallel pattern could be implemented in a parallel way. (1)

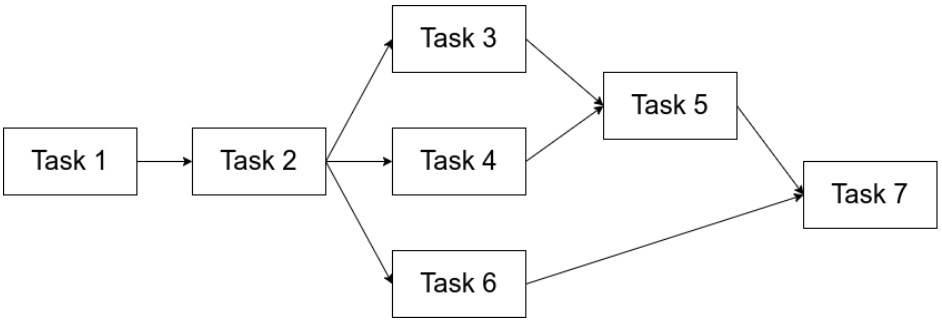
D. Which loop can be fully parallelized? Are there partially parallelizable loops? (1)

```
for (i=0; i<99; i++)
        for (j=5; j<99; j++)
            a[i][j-5] = f(a[i][j]);
```

Exercise n. 2

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

A. Consider the following task graph:



W(T1) = 100
W(T2) = 10
W(T3) = 250
W(T4) = 250
W(T5) = 10
W(T6) = 250
W(T7) = 10

Calculate work and parallelism. If 3 threads are used and Tasks 2, 5, and 7 represent the thread management overhead, how much does performance improve over a sequential version? (1)
Write a corresponding OpenMP implementation. (2)

B. List the factors that determine the number of threads in an OpenMP parallel region. (1)

Exercise n. 3

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

A. Describe two situations where the use of CUDA Unified Memory provides enough benefits to compensate the latency overhead of on-demand migration. (2)

B. What are the advantages of using asynchronous memory prefetching when transferring data from the host to a device? (1)

C. Is an execution configuration of <<<150, 512>>> suitable for a device with 48 Streaming Multiprocessors and 32-thread warps? (1) **This question can be skipped in case of a passed challenge.**

Exercise n. 4

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

A. What are the characteristics of a program that make it amenable to parallelization with multiple threads? (1)

B. What are the characteristics of a program that make it amenable to parallelization on a GPU? (1)

C. Why is synchronization necessary when programming with the shared memory model? (1)

D. Why would a programmer intentionally apply transformations that introduce redundant computation? (1)