## Advanced and parallel architectures

## Prof. A. Massini

June 8, 2018

End-of-term test

## Student's Name

## Matricola number

| Exercise 1a (3 points) |  |
| :--- | :--- |
| Exercise 1b (3 points) |  |
| Exercise 2 (5 points) |  |
| Exercise 3 (3 points) |  |
| Exercise 4 (4 points) |  |
| Exercise 5 (4 points) |  |
| Exercise 6a (2 points) |  |
| Exercise 6b (2 points) |  |
| Exercise 7 (3 points) |  |
| Exercise 8 (4 points) |  |
| Total (34 points) |  |

## Exercise 1a (3 points) - Interconnection Networks - CLOS

Design a Clos network of size $120 \times 120$, using in the first stage modules having 15 inputs. Consider both cases, strictly non-blocking and rearrangeable network.

## Exercise 1b (3 points) - Interconnection Networks - Comparison Clos-Crossbar

Compare the cost of the crossbar $120 \times 120$ and the Clos network, strictly non-blocking and rearrangeable, designed in the previous point.

## Exercise $\mathbf{2}$ (3+2points) - GPU \& CUDA

You need to write a kernel that operates on a 2D matrix of size $\mathbf{2 0 0 0} \mathbf{2 5 0 0}$. You would like to assign one thread to each matrix element. You would like your thread blocks to use the maximum number of threads per block possible on your device, having compute capability 2.x.
a) How would you select the dimensions of a 2D grid and 2D blocks for your kernel? Consider the two cases of rectangular and square blocks.
b) What is the best choice for grid and block dimensions with respect to the number of idle threads?

| Technical specifications | Compute capability (version) |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1.0 | 1.1 | 1.2 | 1.3 | 2.x | 3.0 | 3.5 | 3.7 | 5.0 | 5.2 |
| Maximum dimensionality of grid of thread blocks | 2 |  |  |  | 3 |  |  |  |  |  |
| Maximum x-dimension of a grid of thread blocks | 65535 |  |  |  |  | $2^{31}-1$ |  |  |  |  |
| Maximum y -, or z -dimension of a grid of thread blocks | 65535 |  |  |  |  |  |  |  |  |  |
| Maximum dimensionality of thread block | 3 |  |  |  |  |  |  |  |  |  |
| Maximum x- or y-dimension of a block | 512 |  |  |  | 1024 |  |  |  |  |  |
| Maximum z-dimension of a block | 64 |  |  |  |  |  |  |  |  |  |
| Maximum number of threads per block | 512 |  |  |  | 1024 |  |  |  |  |  |
| Warp size | 32 |  |  |  |  |  |  |  |  |  |
| Maximum number of resident blocks per multiprocessor | 8 |  |  |  |  |  | 16 |  |  |  |
| Maximum number of resident warps per multiprocessor | 24 |  | 32 |  | 48 | 64 |  |  |  |  |
| Maximum number of resident threads per multiprocessor | 768 |  | 1024 |  | 1536 | 2048 |  |  |  |  |
| Technical specifications | 1.0 | 1.1 | 1.2 | 1.3 | 2.x | 3.0 | 3.5 | 3.7 | 5.0 | 5.2 |
|  | Compute capability (version) |  |  |  |  |  |  |  |  |  |

## Exercise 3 (3 points) - Interconnection networks - ( $2 \log N-1$ ) MIN

Complete the scheme of the Baseline-Baseline ${ }^{-1}$. Show the switch setting to realize permutation $P=\left(\begin{array}{llll}01 & 23 & 45 & 67 \\ 70 & 15 & 34 & 26\end{array}\right)$ according to the looping algorithm.


## Exercises 4 (4 points) Amdhal Law

The following measurements are recorded with respect to the different instruction classes for the instruction set running a given set of benchmark programs:

| Instruction Type | Instruction Count (millions) | Cycles per Instruction |
| :---: | :---: | :---: |
| Arithmetic and logic | 7 | 4 |
| Load and store | 6 | 2 |
| Branch | 3 | 5 |
| Others | 8 | 3 |

Assume that "Branch" instructions can be modified so that they take 4 cycles per instruction instead of 5, and "Arithmetic and logic" instructions can be modified so that they take 3 cycle per instruction instead of 4 as in the table. Compute the speedup obtained by introducing only one enhancement and both enhancements using the Amdhal law.

## Exercises 5 (4 points) Performance equation

Suppose we have made the following measurements, where we are considering Arithmetic and FP (Floating Point) instructions:
Frequency of Arithmetic operations $=30 \%$
Average CPI of Arithmetic operations $=3.5$
Average CPI of other instructions $=2.5$

Frequency of FP operations $=25 \%$
CPI of $F P=6.0$
Assume that the two design alternatives are to decrease the CPI of FP to 3.0 or to decrease the average CPI of Arithmetic operations to 2.5 .
Compare these two design alternatives using the processor performance equation, and compute the speedup in both cases.

## Exercise 6a (2 points) - Number representation

Given the values $A=0001110110$ and $B=0010000111$ in the signed $R B$ (Redundant Binary) representation, convert $A$ and $B$ in decimal.

## Exercise 6b (3 points) - Number representation

Show the execution of operation A-B. Verify the value of the results.

## Exercise 7 (3 points) - Number representation

- Represent the natural number range [0;399] using the residue number system, considering three different choices of the moduli set, consisting of $\mathbf{3}$ moduli.
- Compare the different choices with respect to the number of bits necessary for the representation, and consider also the number of bits needed for representing the range [0; 399] with the conventional binary system.

Represent $A=17$ and $B=21$ using the considered different choices of the moduli sets, and show how to compute the sum $A+B$ and the product $A * B$.

## Exercise 8 (4 points) - Circuit time and area

Compute the time (propagation delay) and area required by the 4-bits ripple carry array multiplier, shown here below.


