THESIS SUMMARY

"**Integration of FPGA-based accelerators in distributed computing environments**"

Author: inf. Laurenţiu DUMITRU, tel. 0727.737.660, email: dumitru.laurentiu@cmsystems.ro
Ph.D Supervisor: Professor Ciprian RĂCUCIU, Ph.D.,Eng.

The thesis's main subject is the development of a platform which assures automatic FPGA accelerator management in distributed computing environments, standardized interconnection for reconfigurable acceleration modules and well-defined API interfaces which will be used by final application developers. This subject places the thesis under the "Electronic engineering, telecommunications and informational technologies" domain.

The thesis is structured on eight chapters. The first chapter, the introduction, exposes the technological context, the author's contributions and the main goal of the thesis.

The second chapter analyses the benefits of using hardware acceleration in various domains, based on current literature. From the analyzed papers, an improved execution time is observed for various algorithms implemented in FPGA. There are no common technical components between these studies; every solution implements its own transfer and resource management. By developing an access and control layer, the overall complexity of FPGA-based solutions can be reduced, thus facilitating wider adoption of such solutions in computing clusters.

In the third chapter various existing technical solutions that abstract the use of acceleration devices, from both open-source and commercial environments, are being evaluated. Advantages and disadvantages are being discussed in comparison with the proposed solution and between them. Also, some improvement ideas are being presented.

Chapter four presents the platform's global architecture, the communications channels between the accelerator card and the guest system, the communications channel between various internal elements on the accelerator and the software components that run at kernel and user-space level on the guest system.

The algorithms and procedures that are being used while the FPGA accelerator card is in use are being described in Chapter 5, along with the structures of the acceleration modules, the encapsulating components and the input/output channels. The standard signals that permit, at hardware level, reusing acceleration modules and the standardized API functions that are used, on software level, are being described. Partial and total reconfiguration flows, along with all the required elements that control the process, are being discussed in the context of integrating with the overall structure.

Chapter six focuses on the implementation of the acceleration platform in a distributed computing environment. The operation modes and the functional flows that allow transparent integration with a computing cluster in which the computing nodes are equipped with FPGA accelerator cards are being described. The complete work-flow is presented starting from the initial submission of a job to the batch system, going through all the specific steps – loading, execution, reprogramming, a.s.o. -, and finishing with the transfer of the output data from the reconfigurable accelerator module to the calling user-space application.

Chapter seven presents two study cases. The first one is done in a commercial environment while the second one is done in a research environment. Both studies demonstrate the platform's utility and the implicit benefits from both a financial/administrative and a technical point of view.

The conclusions and research directions are exposed in chapter eight. This chapter consists of a summary of the main benefits brought by FPGA acceleration cards and the centralized management of these cards in distributed computing structures. In terms of expanding the capabilities of the proposed platform, future development ideas that are not necessary related to distributed computing are being discussed.