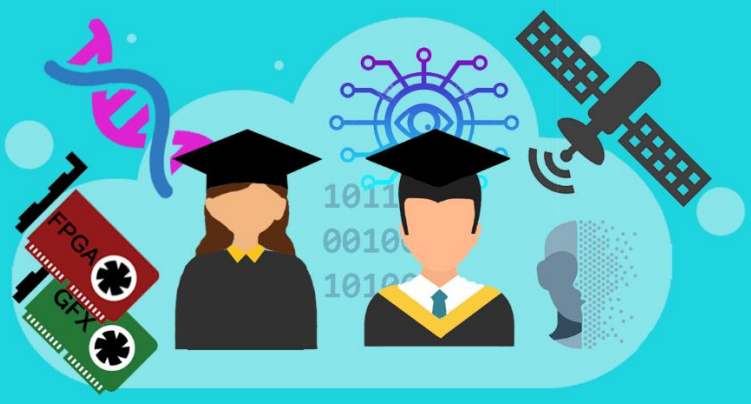
Diploma Thesis

Microproccessors and Digital Systems Laboratory



Implementation of system-aware optimizations for GPU kernels autotuning

The programmability of General Purpose Graphics Processing Units (GPGPUs) has been improved significantly over the past years. Programming languages have matured, development tools have increased capabilities, libraries were created or improved, and so on. Despite these advancements, achieving close-to-peak performance remains a task for expert programmers in many cases. Even experts might face optimization problems where the space of design decisions is too large to explore. Therefore, they need to tune their code for a wide variety of devices, each with their own sensitivities to specific parameters. Furthermore, to achieve optimal performance, code can be tailored to specific input arguments (e.g. matrix dimensions), either at compile-time (off-line) or at run-time (on-line). Hence, accurate automatic optimization heuristics are necessary for dealing with the complexity and diversity of modern hardware and software.

The scope of this thesis is to research and generate various techniques of automatic run-time tuning of GPU applications and build optimization heuristics based on system awareness (hardware device, available GPU resources, etc...). A complete hardware-independent optimization heuristic will be targeted, exploring various optimization techniques in the terms of boosting performance and limiting energy.

Prerequisites : Basic knowledge of GPUs, C, Linux, Scripting

Academic Advisors: Dimitrios Soudris (dsoudris@microlab.ntua.gr)

Member1 : Sotirios Xydis (sxydis@microlab.ntua.gr)

Member2: Ioannis Oroutzoglou (ioroutzoglou@microlab.ntua.gr)