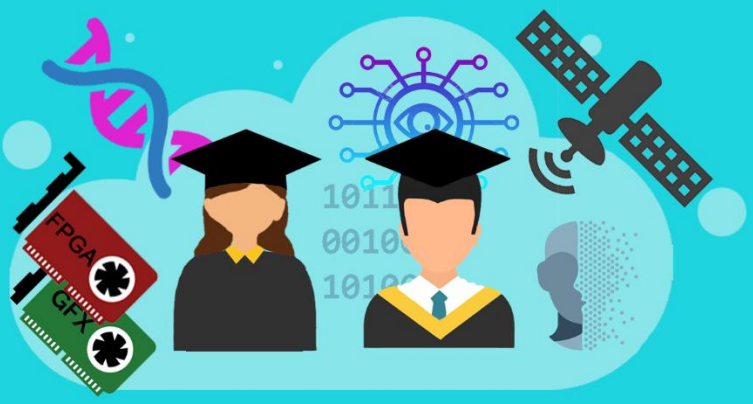
Diploma Thesis

Microproccessors and Digital Systems Laboratory

Academic Year 2019-2020



Application of Approximate Computing Techniques on the NVIDIA Deep Learning Accelerator (NVDLA)

The pervasive nature of modern computing systems has led to an increased need for high performance and energy efficiency. Towards this direction, **Approximate Computing** is considered as a promising paradigm shift for energy-efficient systems design, exploiting the inherent resilience of various applications. This relaxation in the requirements for exactness is evident in several emerging domains, e.g., machine learning, multimedia processing, etc., favored due to several factors, such as the limited human perception, the probabilistic/statistical calculations, and the user's intention to accept results of lower quality. Thus, error is considered as a commodity that can be traded for significant gains in performance, power/energy consumption.

Massive research has been reported in the field of approximate computing at various layers of software and hardware. Focusing on the hardware level, approximations can be applied at different design layers of abstraction, i.e., the application, algorithmic, gate and transistor layers. Regarding circuit designs, the main targets are the adders and the multipliers, i.e., the core units of hardware accelerators. Extensive research has also been conducted in approximate processors, using neural networks, quality programmable vectors and approximate custom instructions.

NVDLA is an industry-grade open-source Deep Neural Network (DNN) inference engine, developed by NVIDIA. It is released as Verilog source code and is configurable at the build time to meet different performance, power, and area trade-offs. NVDLA mainly targets embedded systems and IoT devices with limited power budget.

The **goal of this diploma thesis** is to enable the building of the NVDLA over an ASIC tool-flow and libraries (already available at Microlab), as well as the exploration of integrating novel approximate components to optimize the energy consumption of NVDLA.

Useful Links:

https://github.com/nvdla/

https://en.wikipedia.org/wiki/Approximate_computing

Prerequisites:

basic knowledge of Digital Design, Verilog, C, linux

Academic Advisors:

Vasileios Leon, PhD: vleon@microlab.ntua.gr

Sotirios Xydis, Dr: sxydis@microlab.ntua.gr