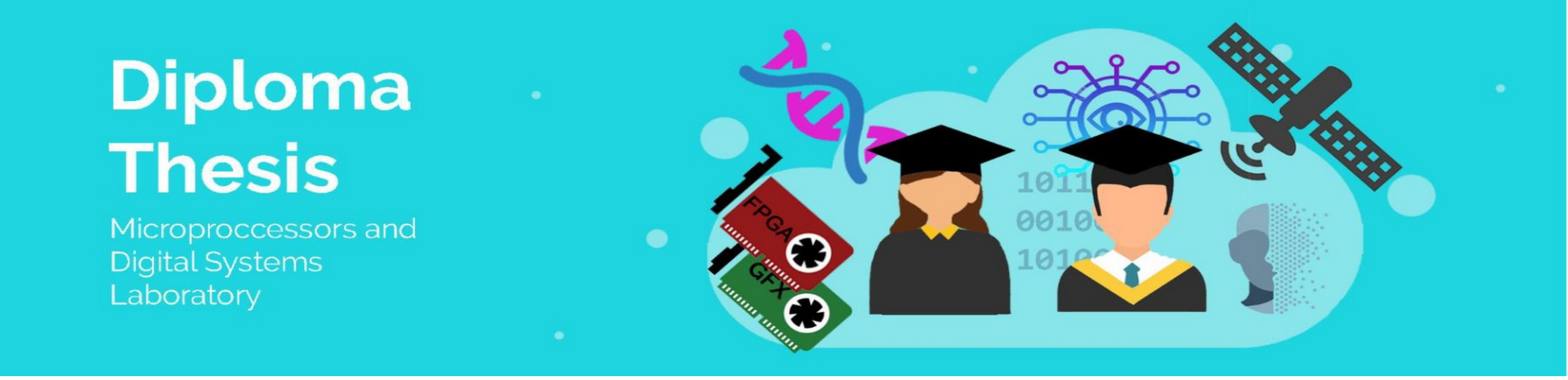
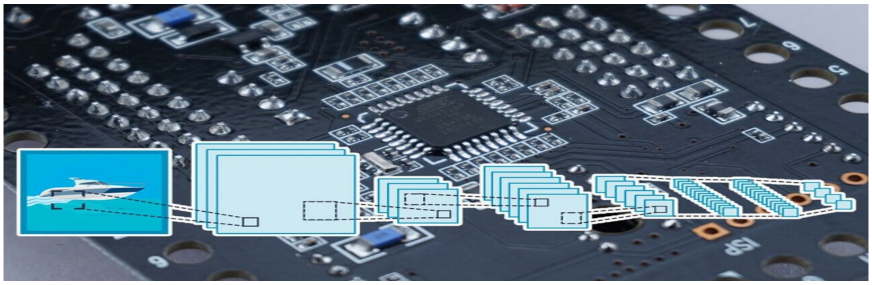


Approximate TinyML systems for extreme resource constrained devices

The next wave of growth in AI will be driven in good part by embedding intelligence into ubiquitous devices such as wearables, implanted devices, smart home devices, etc. Most of these devices present extremely tight size, cost, weight and power constraints. Tiny ML is broadly defined as a fast growing sub-field of Machine Learning (ML) that addresses the design of *algorithms*, *software frameworks*, *hardware* and *systems* for such applications.

The nature of such constrained devices raise the need for new design methodologies that will increase the efficiency (e.g., in area and power) of ML implementations. Towards this direction, Approximate Computing (AC) is considered a promising design paradigm for energy-efficient systems, exploting the inherent error resilience of various applications, including ML.



Traditionally, we are used to design methodologies, that given one specification in input can generate hundreds of different implementation versions, and these may vary in terms of energy efficiency, delay, memory usage, etc. But playing with the accuracy of the output was not ever something that was on the table. What if it is. What if we can automatically tell our design tools: “and this is what I am prepared to pay in terms of accuracy loss”. Then we can say: “give me a less accurate implementation, as long as it’s both smaller and faster”. That sounds great; but what are the challenges that AC poses, and how far are we from seeing Approximate Computing everywhere? Are there enough and significant applications that can tolerate a loss in accuracy? Should we consider AC just in software, before we commit into generating inaccurate hardware?

These are the key questions that this thesis targets to answer. The goal is to explore and develop frameworks and methodologies to build customized hardware accelerators for Tiny ML inference with respect to power/throughput/resource minimization.

Prerequisites:
Basic knowledge of Digital Design, FPGAs, VHDL, Python, Bash

Related Material:

- https://www.youtube.com/watch?v=lcX77I9bLJo
- https://thenextweb.com/news/tinyml-deep-learning-microcontrollers-syndication

Available Diploma Thesis: 2

Contact Information:

- Giorgos Armeniakos PhD candidate: (armeniakos@microlab.ntua.gr)
- Prof. Dimitrios Soudris: (dsoudris@microlab.ntua.gr)