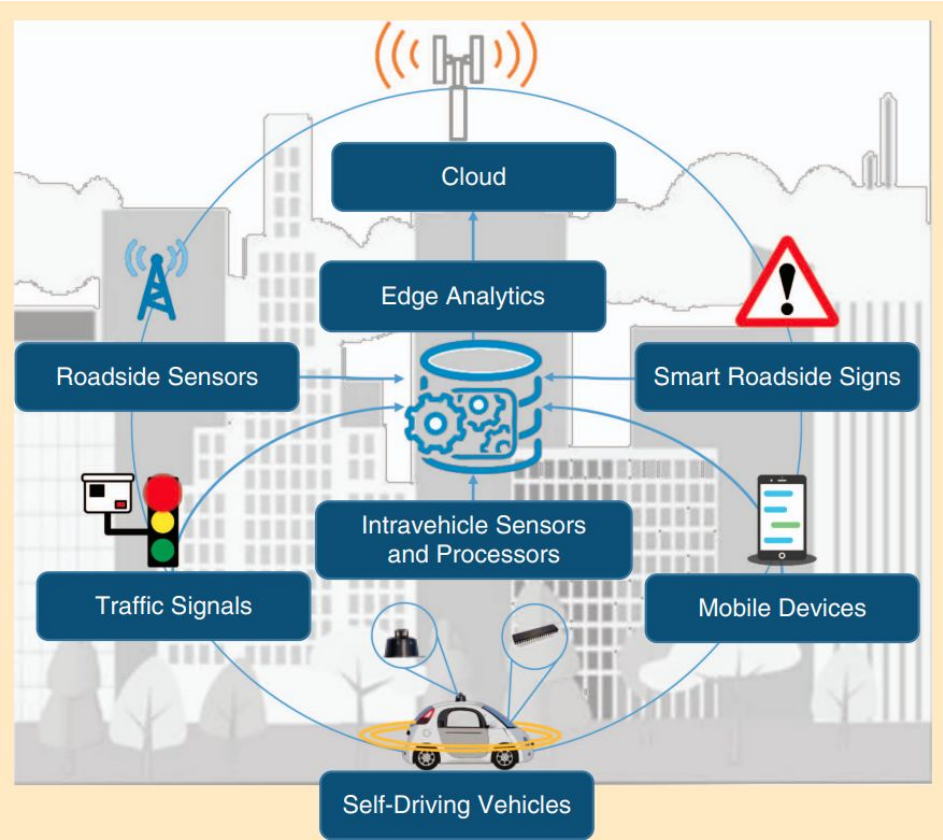
Dynamic Hardware Customisation for Mobile Users in FPGA-accelerated Edge Infrastructures

Diogo Gonçalves[1], Luiz Bittencourt[1], Edmundo Madeira[1], Ioan Petri[2] and Omer Rana[2]

[1]University of Campinas, Brazil
[2]Cardiff University, UK

Introduction - motivation



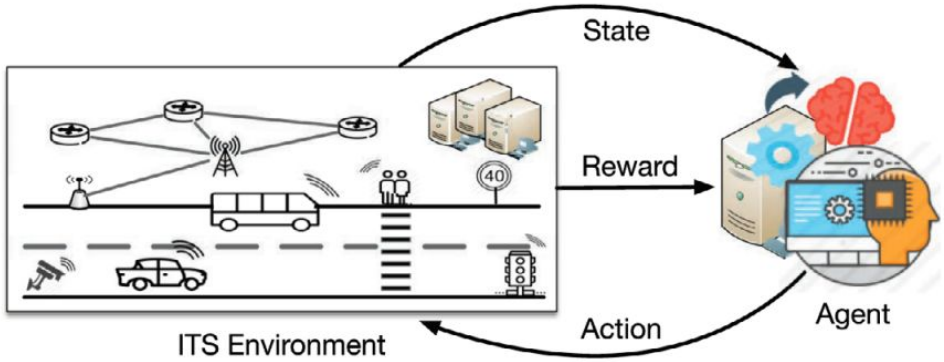
Smart city solutions for mobile devices;

Computing demand in edge is higher and higher;

Edge servers have lower computing power than cloud servers;

Mobile users add complexity in that scenario.

Introduction - motivation



Smart city solutions for mobile devices;

Machine learning algorithms and cloud/edge infrastructures have been playing a successful role as smart city supporters.

Our hypothesis

Hardware customised for a particular application avoids unnecessary resources and focuses on specific characteristics that improve the application's performance.

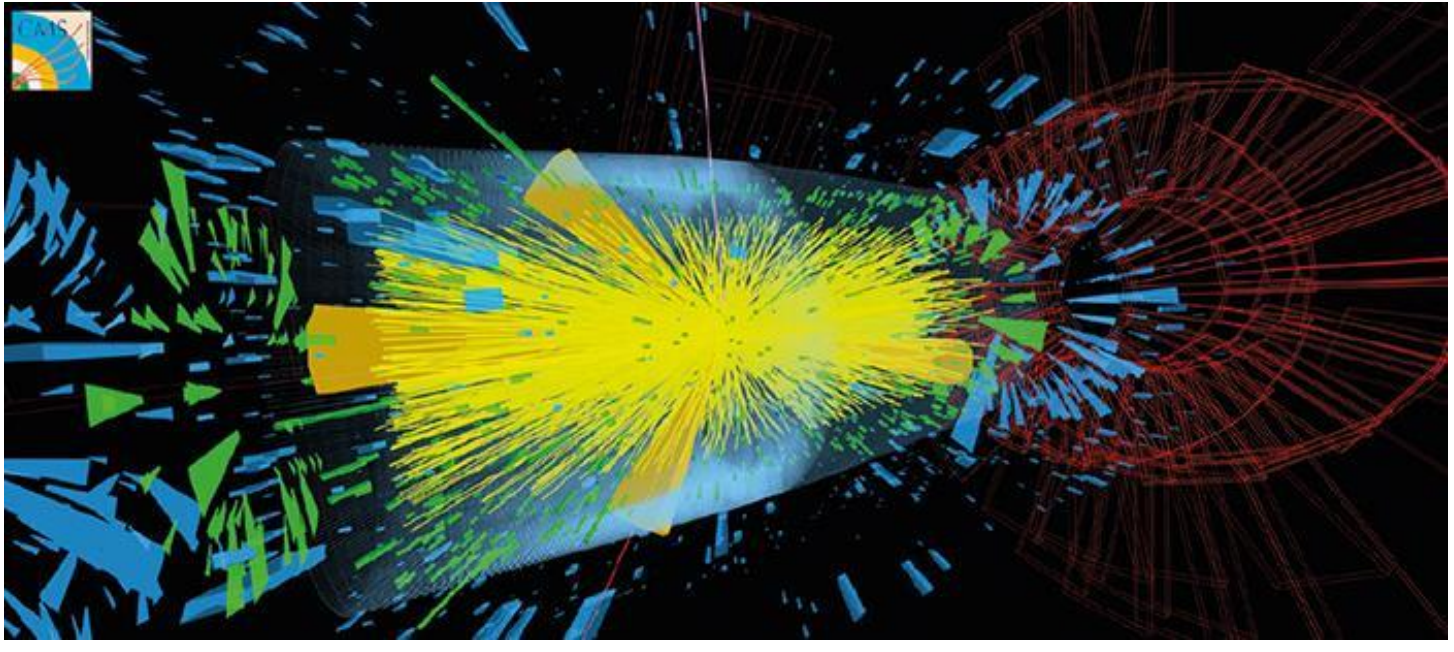
Customised hardware trades off flexibility for efficiency, enabling applications to run faster, reducing energy consumption, or both.

Hardware customisation could be key to improving application performance in edge devices.

Attaching FPGAs into edge nodes could offer tailored hardware for ML applications at the edge.

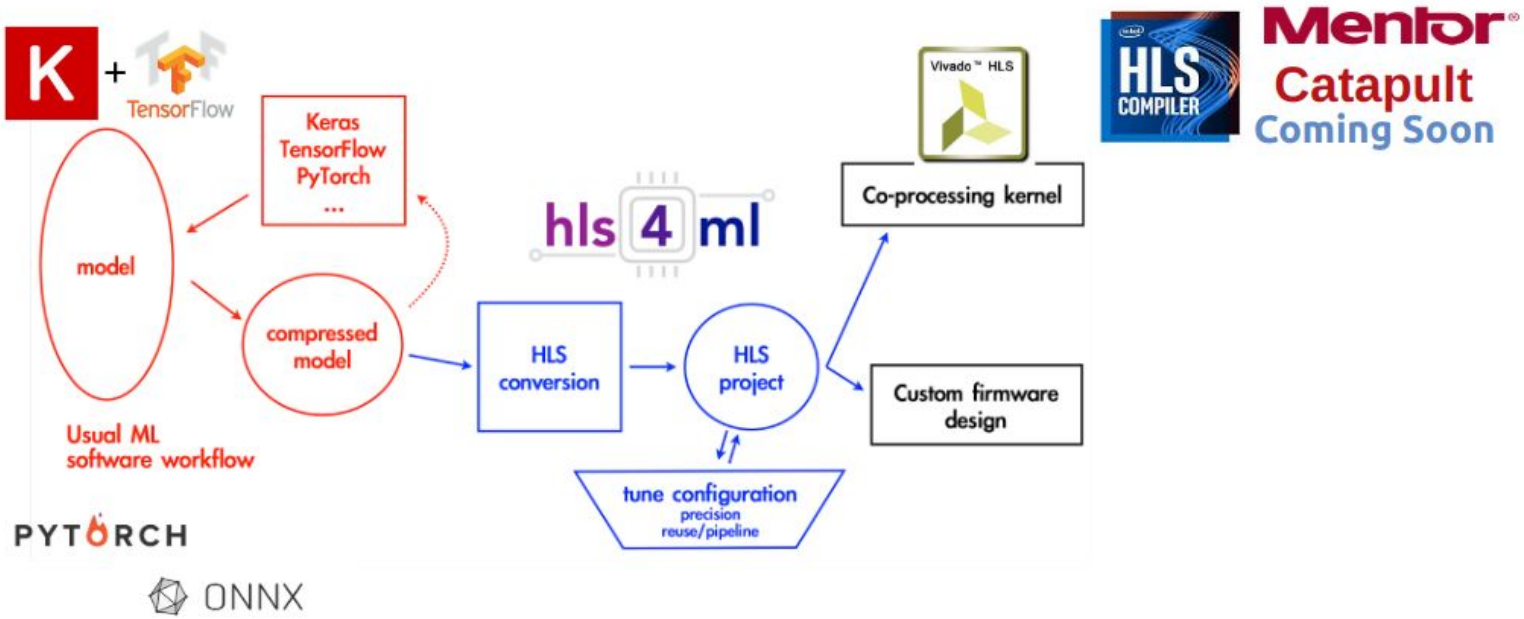
hls4ml - high level synthesis for machine learning [1]

Origins: triggering at (HL-)LHC
Extreme collision frequency of 40 MHz → extreme data rates O(100 TB/s)
Most collision “events” don’t produce interesting physics
“Triggering**”** = filter events to reduce data rates to manageable levels

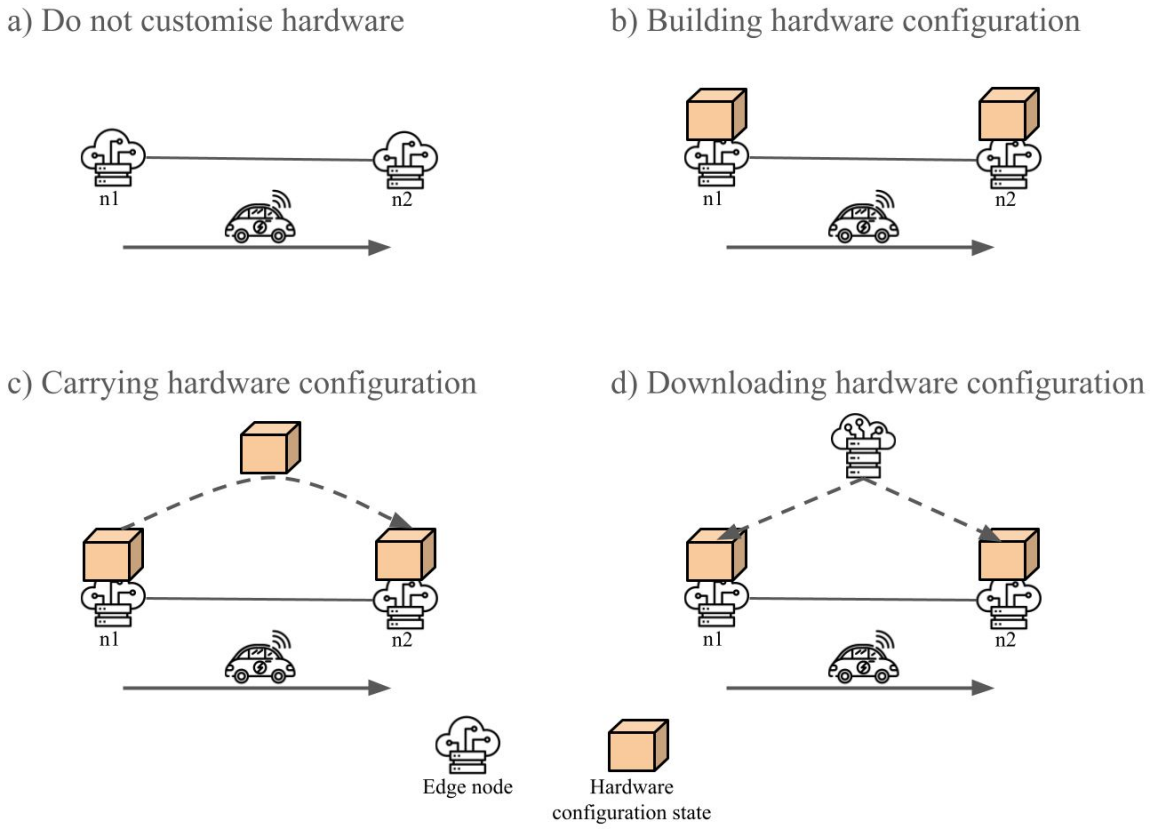


hls4ml

hls4ml is a package for translating neural networks to FPGA firmware for inference with extremely low latency on FPGAs.



Our scenarios - follow me cloud



Exploring hardware 'as a state'

Once the hardware settings are defined as a VHDL format, that 'hardware state' could be carried by the user e applied in his/her following edge servers along the path.

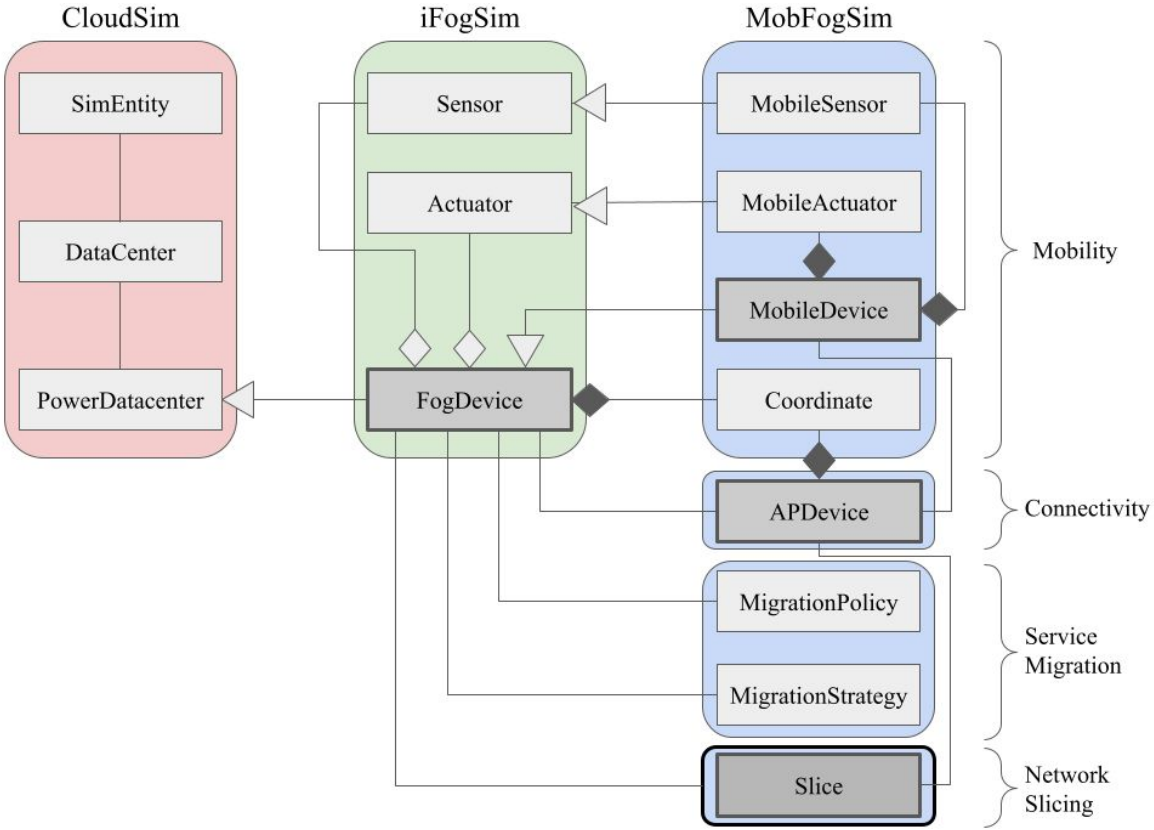
Validation Scenario

Hls4ml;

Xilinx Vivado HLS;

MobFogSim.

Validation Scenario - MobFogSim



MobFogSim - cloud/edge simulator with user mobility support.

✔User mobility support;

✔Service migration support;

✔Network Slicing support.

Next steps

Validates FPGA reconfiguration in edge as a feasible overhead;

Integrates hls4ml results into MobFogSim to validate FPGA customisation for mobile users.

Final Remarks

Hardware customisation might improves application performance in edge devices.

Attaching FPGAs into edge nodes could offer tailored hardware for ML applications at the edge.

FPGAs reconfiguration might be a solution for improving the efficiency of ML implementations at the edge.

References

[1] Fahim, Farah, et al. "hls4ml: An Open-Source Co-Design Workflow to Empower Scientific Low-Power Machine Learning Devices." Research Symposium on Tiny Machine Learning. 2020.

[2] https://www.xilinx.com/support/documentation/sw_manuals/xilinx2020_1/ug902-vivado-high-level-synthesis.pdf

[3] Puliafito, Carlo, et al. "MobFogSim: Simulation of mobility and migration for fog computing." Simulation Modelling Practice and Theory 101 (2020): 102062.

Thanks

diogomg@lrc.ic.unicamp.br

This work is part of the INCT of the Future Internet for Smart Cities (CNPq 465446/2014-0, CAPES 88887.136422/2017-00, and FAPESP 2014/50937-1).