João Guerreiro

joaofilipedg@gmail.com joaofilipedg.github.io joaofilipedg joaofdguerreiro

Education

Instituto Superior Técnico, Universidade de Lisboa — *Lisbon, Portugal*

Ph.D. in Electrical and Computer Engineering — *Feb. 2015-Jun. 2020*

- Ph.D. Thesis: "DVFS Modeling for Energy-Efficient GPU Computing".
- Jury Final Classification: Pass with Distinction and Honour
- Supervised by: Pedro Tomás, Nuno Roma and Aleksandar Ilic

Instituto Superior Técnico, Universidade de Lisboa — *Lisbon, Portugal*

M.Sc. in Electrical and Computer Engineering — *Jul. 2012-Nov. 2014*

- Major area: Computers.
- Minor area: Systems, Decision and Control.
- Final evaluation: 17 (out of 20).
- Master's Thesis: "Multi-Kernel Auto-Tuning on GPUs: Performance and Energy-Aware Optimization".

Chalmers University of Technology — *Gotenborg, Sweden*

M.Sc. in Electrical and Computer Engineering — *Jan. 2013-Jun. 2013*

- 1 semester exchange under the Erasmus programme.

Instituto Superior Técnico, Universidade de Lisboa — *Lisbon, Portugal*

B.Sc. in Electrical and Computer Engineering — *Sept. 2009-Jun. 2012*

- Final evaluation: 16 (out of 20).

Experience

INESC-ID - Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa — *Lisbon, Portugal*

Junior Researcher — *Sept. 2014 - Jun. 2020*

- Junior Researcher while doing the PhD.
- Managed the scientific group servers at INESC-ID, running CentOS.

Languages

Portuguese Native proficiency.
English Professional proficiency.

Scientific Activities

INESC-ID – Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa — *Lisbon, Portugal*

Ph.D. Scholarship — *Aug. 2019-Jun. 2020*

- Under project: "HAnDLE: Hardware Accelerated Deep Learning Framework" (PTDC/EEI-HAC/30485/2017).
- Project funded by FCT.

FCT - Fundação para Ciência e Tecnologia — *Lisbon, Portugal*

Ph.D. Scholarship — *Jul. 2015-Jun. 2019*

- Grant funded by FCT: SFRH/BD/101457/2014.

INESC-ID – Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa *Lisbon, Portugal*

Master's Scholarship *Sept. 2014-Feb. 2015*

- Under project: "THREadS: Multitask System Framework with Transparent Hardware Reconfiguration" (PTDC/EEA-ELC/117329/2010).
- Project funded by FCT.

Research

International Conference Articles

[C1] João Guerreiro, Aleksandar Ilic, Nuno Roma, and Pedro Tomás. "GPGPU Power Modeling for Multi-Domain Voltage-Frequency Scaling", in *24th IEEE International Symposium on High-Performance Computer Architecture* (**HPCA'2018**), Vienna, Austria, Feb. 2018. Available: https://ieeexplore.ieee.org/document/8327055.

[C2] João Guerreiro, Aleksandar Ilic, Nuno Roma, and Pedro Tomás. "Multi-Kernel Auto-Tuning on GPUs: Performance and Energy-Aware Optimization", in *23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing* (**PDP'2015**), Turku, Finland, Mar. 2015. Available: https://ieeexplore.ieee.org/document/7092758.

Journal Articles

[J1] João Guerreiro, Aleksandar Ilic, Nuno Roma, and Pedro Tomás. "GPU Static Modeling using PTX and Deep Structured Learning", in **IEEE Access**, Volume 7, Issue 1, pp. 159150-159161, Nov. 2019. Available: https://ieeexplore.ieee.org/document/8890640.

[J2] João Guerreiro, Aleksandar Ilic, Nuno Roma, and Pedro Tomás. "Modeling and Decoupling GPU Power Consumption for Cross-Domain DVFS", in **IEEE Transactions on Parallel and Distributed Systems**, Volume 30, Issue 11, pp. 2494-2506, Nov. 2019. Available: https://ieeexplore.ieee.org/document/8716300.

[J3] João Guerreiro, Aleksandar Ilic, Nuno Roma, and Pedro Tomás. "DVFS-aware application classification to improve GPGPUs energy efficiency", in **Parallel Computing**, Volume 83, pp. 93-117, Apr. 2019. Available: https://www.sciencedirect.com/science/article/pii/S0167819118300243.

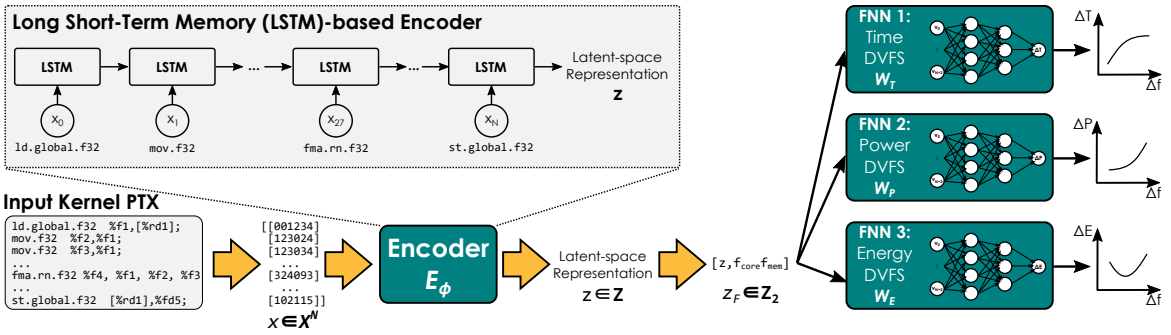
Dissertations

[D1] João Guerreiro, "Multi-Kernel Auto-Tuning on GPUs: Performance and Energy-Aware Optimization". **Thesis aproved to obtain the Master's Degree in Electrical and Computer Engineering** by Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, November 2014.

Research Highlights

GPU Static Modeling using PTX and Deep Structured Learning *IEEE Access*

Figure 1: Diagram of the proposed PTX-based characterization models.



This work presents a novel GPU modeling methodology able to provide accurate predictions on how the execution time, power and energy consumptions of applications scale when the GPU operating frequencies are scaled, without requiring the application execution. To that end, the proposed methodology uses the PTX (assembly of NVIDIA GPUs) code obtained using the compiler. Unlike previous static works that simply rely on general code statistics, such as the histogram of instructions in the PTX code, the proposed approach considers the specific sequence of kernel instructions, using a recurrent encoder architecture (based on *LSTM* blocks). This way, it models how the pattern of instructions stresses the GPU components, thus contributing to different performance, power and energy scalings. The

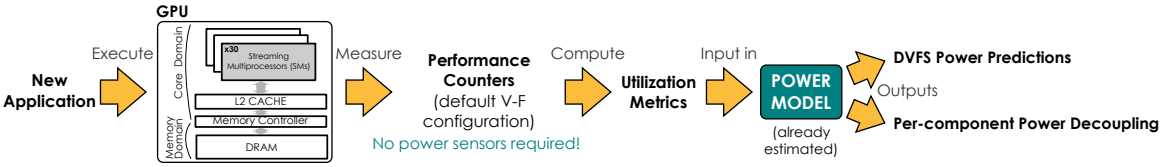
obtained experimental results show how the proposed models can be used to accurately predict the best operating frequencies for different types of applications.

Article: https://ieeexplore.ieee.org/document/8327055.

Open-source: https://github.com/hpc-ulisboa/gpupowermodel.

GPGPU Power Modeling for Multi-Domain Voltage-Frequency Scaling *HPCA, Vienna, Austria*

Figure 2: Diagram of the proposed hardware counters-based GPU power model.



A novel GPU power consumption estimation model with core and memory frequency scaling is proposed in this article. Based on the GPU components utilization during the execution of applications, the model allows the prediction of the power consumption of each component, as well as estimating how the voltage of each domain scales with its operating frequency. In order to construct an accurate model, a suite of synthetic benchmarks is proposed, specifically developed to model the unknown characteristics of the GPU device. Experimental results obtained on multiple GPU devices validate the accuracy of the proposed model.

Article: https://ieeexplore.ieee.org/document/8890640.

Open-source: https://github.com/hpc-ulisboa/gpuPTXModel.