GUSTAVO LEITE

Av. Albert Einstein, 1251, Cidade Universitária, Campinas-SP, Brazil, ZIP 13083-852

+55 (19) 99721-4443 ⋄ contact@gustavoleite.me

EDUCATION

BSc in Computer Science   *2013–2016*
São Paulo State University (UNESP)   *Rio Claro-SP, Brazil*
I finished the course with grade 8.58 out of 10, placed as the first student of the class. I developed a
study on parallel computing with application to the n-body problem as my final paper. This work was
supervised by Prof. Alexandro Baldassin.

MSc in Computer Science   *2017–2019*
São Paulo State University (UNESP)   *Rio Claro-SP, Brazil*
I obtained grade A in 5 out of 6 courses attended. I defended the thesis titled "Performance Evaluation
of Code Optimizations in FPGA Accelerators" in August, 2019 under supervision of Prof. Alexandro
Baldassin.

Complementary Training   *2018*
University of Alberta   *Edmonton-AB, Canada*
During a research internship, I audited two courses on compilers in the University of Alberta. The
courses were "Compiler Design" and "Machine Learning in Optimizing Compilers", both lectured by
Prof. José Nelson Amaral during the Fall of 2018.

PhD in Computer Science   *2020–present*
University of Campinas (Unicamp)   *Campinas-SP, Brazil*
I am currently a PhD candidate at the University of Campinas, working on the Computer Systems
Laboratory (LSC) under the supervision of Prof. Guido Araújo and Prof. Marcio Machado Pereira.

AWARDS AND DISTINCTIONS

Best Academic Performance ⋄ UNESP   *2017*
In March 2017, I was awarded a distinction from the São Paulo State University for the best academic
performance in the Bachelors in Computer Science 2016's class.

Winner of the Parallel Programming Challenge ⋄ ERAD-SP, ICMC-USP   *2017*
In April 2017, my team won the Parallel Programming Challenge carried during the 8th Regional School
of High-Performance (ERAD-SP), ICMC-USP, São Carlos-SP, Brazil.

INTERESTS

Following, a non-exhaustive list of my professional and research interests:

- high-performance computing;
- compilers;
- machine learning systems;
- computer architecture.

PROFESSIONAL EXPERIENCE

I do not have professional experience.

RESEARCH EXPERIENCE

Self-Healing Software   *2014–2015*
São Paulo State University (UNESP)   *Rio Claro-SP, Brazil*
During the bachelors course, I have participated in two research projects aimed at developing better

recommendation systems to assist decision-making in self-healing systems. I was supervised by Prof. Frank José Affonso and received funding from CNPq and FAPESP. During this period I coauthored two papers published in international conferences [1, 2].

Parallel Programming   *2016*
São Paulo State University (UNESP)   *Rio Claro-SP, Brazil*

I have developed a term paper where I studied the parallel programming paradigm using OpenMP and OpenCL. With the knowledge obtained, I compared the performance of implementations in CPU and GPU of the n-body problem. This work was supervised by Prof. Alexandro Baldassin.

High-Performance Computing   *2017–2019*
São Paulo State University (UNESP)   *Rio Claro-SP, Brazil*

During the masters course, I have investigated workload balancing techniques for NUMA systems and, after that, I have analyzed the performance of code optimizations aimed at FPGA accelerators present in the literature. I defended the thesis titled "Performance Evaluation of Compiler Optimizations in FPGA Accelerators" [4] in August 2019 under the supervision of Prof. Alexandro Baldassin. This work was carried with funding from CAPES and FAPESP. During this period I have also been a teacher assistant in the undergrad course "Microprocessors II".

High-Performance Computing   *2018*
University of Alberta   *Edmonton-AB, Canada*

I visited the University of Alberta between September 2018 and November 2018 for a research internship. The objetive of this project was to conduct a bibliographic survey on existing compiler optimizations for FPGA accelerators. I was co-supervised by Profs. José Nelson Amaral (UAlberta) and Guido Araújo (IC-Unicamp). This internship was funded by the BEPE/FAPESP program. From this collaboration, we published a paper in a brazilian conference [5].

OmpCluster Programming Model   *2020–2021*
University of Campinas (Unicamp)   *Campinas-SP, Brazil*

During the first two years of my PhD, I was part of a research project that consisted in extending the OpenMP programming model for working in cluster environments. I was responsible for implementing the first version of the task scheduler in the OmpCluster runtime [6, 3]. I also created a benchmarking tool called OmpcBench that standardized the process of collecting benchmark results for the runtime. This tool was adopted by the entire team and is continued to be used to this day. During this period I also helped Prof. Guido by preparing and lecturing classes about CUDA programming in the "Parallel Programming" course offered at the Institute of Computing at Unicamp.

Machine Learning Systems   *2022–present*
University of Campinas (Unicamp)   *Campinas-SP, Brazil*

At present I have been involved in a research project that aims to find more efficient techniques for training large deep learning models. More specifically, we are investigating how to train models in systems with limited memory by checkpointing and restoring intermediate values. We have devised a dynamic programming algorithm for partitioning a computational graph into multiple stages. We optimize the partitioning scheme for minimizing the amount of data flowing from one stage to the next, thus lowering the footprint required to save and restore. Additionally, we compress and decompress the data using a lossless algorithm for increased efficiency. We plan to submit a research paper to a relevant conference in the following months.

TEACHING EXPERIENCE

Introduction Parallel Programming (MO644)   *2021–2022*
University of Campinas (Unicamp)   *Campinas-SP, Brazil*

During the second semester of 2021 I collaborated with Prof. Guido Araújo by preparing and lecturing classes on parallel programming. More specifically, I was responsible for the lectures related to the

CUDA programming model. The content included the CUDA programming language, NVIDIA GPU architecture and memory hierarchy. This course was offered again in the first semester of 2022.

Machine Learning Under the Hood (MO436) *2022*
University of Campinas (Unicamp) *Campinas-SP, Brazil*
During the second semester of 2022 I collaborated with Prof. Guido Araújo by preparing and lecturing classes on the Google JAX framework for machine learning and numeric computing. The lectures covered how to use the framework and, more importantly, how it works under the hood. We have explained how automatic differentiation was implemented in JAX and how to extend the library with new operations.

SKILLS AND LANGUAGE

Programming Languages

Proficient   C/C++, Python, CUDA
Familiar   Rust

Tools (Ordered by familiarity)

- Linux operating system and utilities (`grep`, `bash`, `awk`, `make`, etc);
- C/C++ compilers (`gcc` and `clang`);
- Control version system (`git`) and platforms (Github, Gitlab, etc);
- Parallel programming paradigms (OpenMP, CUDA);
- Python packages (`numpy`, `matplotlib`, `seaborn`, etc);
- Debuggers (`gdb`) and profilers (`perf`, `strace`, `ltrace`);
- LLVM compiler infrastructure.

Languages Proficiency

- **Portuguese** (Native) – comprehension: good; speaking: good; writing: good;
- **English** (Fluent) – comprehension: good; speaking: good; writing: good;
- **German** (Beginner/A1) – comprehension: basic; speaking: none; writing: basic;

EVENT ATTENDANCE

SBAC-PAD'17 ⋄ Campinas-SP, Brazil *2017*
International Symposium on Computer Architecture and High Performance Computing

8ł ERAD-SP ⋄ ICMC-USP, São Carlos-SP, Brazil *2017*
Escola Regional de Alto Desempenho do Estado de São Paulo

SBAC-PAD'19 ⋄ Campo Grande-MS, Brazil *2019*
International Symposium on Computer Architecture and High Performance Computing

CARLA'22 ⋄ Porto Alegre-RS, Brazil *2022*
Latin America High Performance Computing Conference

CARLA'23 ⋄ Cartagena, Colombia *2023*
Latin America High Performance Computing Conference

LINKS

Github	`https://github.com/leiteg`
Gitlab	`https://gitlab.com/leiteg`
Google Scholar	`https://scholar.google.com/citations?user=F6MZj_oAAAAJ`
Lattes Curriculum	`https://lattes.cnpq.br/0392351138118593`

PUBLICATIONS

[1] F. J. Affonso, G. Leite, and E. Y. Nakagawa. Dms-modeler: A tool for modeling decision-making systems for self-adaptive software domain. In *SEKE*, volume 16, pages 617–622, 2016.

[2] F. J. Affonso, G. Leite, R. A. P. Oliveira, and E. Y. Nakagawa. A framework based on learning techniques for decision-making in self-adaptive software. In *SEKE*, volume 15, 2015.

[3] C. Cardoso, H. Yviquel, G. Valarini, G. Leite, R. Ceccato, M. Pereira, A. Souza, and G. Araujo. An openmp-only linear algebra library for distributed architectures. In *2022 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW)*, pages 17–24, 2022.

[4] G. Leite. Performance evaluation of compiler optimizations in FPGA accelerators. Master's thesis, Universidade Estadual Paulista, São José do Rio Preto-SP, Brasil, Aug 2019.

[5] G. Leite, A. Baldassin, G. Araújo, and J. N. Amaral. Performance evaluation of compiler optimizations in FPGA accelerators. In *2019 Symposium on High Performance Computing Systems (WSCAD)*, Oct 2019.

[6] H. Yviquel, M. Pereira, E. Francesquini, G. Valarini, G. Leite, P. Rosso, R. Ceccato, C. Cusihualpa, V. Dias, S. Rigo, A. Souza, and G. Araujo. The openmp cluster programming model. In *Workshop Proceedings of the 51st International Conference on Parallel Processing*, ICPP Workshops '22, New York, NY, USA, 2023. Association for Computing Machinery.

Last updated: February 23, 2024.