

Giorgis Georgakoudis

Principal Compiler and Systems Engineer | LLVM, GPU, OpenMP, Performance Portability

📍 Pleasanton, CA | ✉ georgakoudis1@llnl.gov | ☎ +1 925-495-9544 | 🌐 ggeorgakoudis | 📷 ggeorgakoudis

Summary

Principal compiler and systems engineer with 10+ years of experience building **compiler**, **runtime**, and **GPU performance** infrastructure across **LLVM/MLIR**, **OpenMP**, **CUDA/HIP**, **Numba**, and **HPC** software stacks. Track record leading technical programs and building optimization tooling for production scientific applications.

Core Skills

Languages: C++, C, Python

Compiler & Runtime: LLVM/Clang, MLIR, OpenMP, Numba, code generation, runtime systems, JIT optimization

Parallel & GPU: CUDA, HIP, GPU offload, performance portability, autotuning, heterogeneous computing

Systems & Tooling: Linux, benchmarking, profiling, performance analysis, software workflows

Experience

Principal Computer Scientist (previously Computer Scientist, Postdoctoral Researcher), 2018 – present
Lawrence Livermore National Laboratory – Livermore, CA 7 years 5 months
Lead compiler, runtime, and performance engineering work for large-scale scientific computing software.

- Lead development of **Proteus**, a programmable **LLVM/MLIR** JIT optimization framework for host and GPU kernels targeting **CUDA** and **HIP**, used in production HPC physics applications.
- Co-lead **PyOMP**, an OpenMP-for-Python system built on **Numba**, extending parallel and offload support in a widely used Python compiler ecosystem.
- Contribute to **Mneme**, a GPU kernel record/replay and autotuning workflow built around LLVM IR artifacts and Proteus-based optimization experiments.
- Contributed to OpenMP compilation and the OpenMP offloading runtime.
- Serve as **Principal Investigator** for a **\$2.7M / 3-year** project on just-in-time optimization for HPC applications.
- Chair the **OpenMP Python Language Subcommittee**, contributing to a de-facto industry standard for parallel programming and Python-facing OpenMP support.

Research Fellow/Assistant, Queen’s University Belfast – Belfast, UK 2013 – 2018
Built runtime and systems software for streaming analytics, energy-aware execution, OpenMP runtimes, and heterogeneous architectures. 5 years 11 months

- Developed throughput optimizations in **OpenMP** runtimes, resilience-aware adaptation, and transprecision techniques for heterogeneous systems.

HPC Engineer, Simtec Software and Services – Greece 2011 – 2013
Supported engineering simulation workloads on Linux-based HPC infrastructure; procured and managed a **\$300K** Linux cluster for CFD simulations. 1 year 9 months

Selected Projects

Proteus 2024-present
Programmable JIT compilation and optimization framework using LLVM and MLIR for host and GPU kernels, portable across CUDA and HIP.

- Used in production-level HPC applications for physics simulations.
- Supports compiler-guided specialization, runtime adaptation, and portable GPU optimization workflows.

PyOMP

2024-present

OpenMP for Python built on Numba, bringing parallel and offload capabilities into a major Python compiler ecosystem.

- Extends Python productivity with compiler- and runtime-level control for HPC workloads.
- Connects OpenMP compilation, Python developer workflows, and GPU execution.

Mneme

2024-present

Record/replay and autotuning workflow for GPU kernels, built around LLVM IR artifacts and Proteus-based optimization experiments.

- Isolates GPU kernels for experimentation with compiler transformations and launch-parameter tuning.
- Extends practical autotuning workflows for performance-sensitive HPC applications.

Education

University of Thessaly, PhD in Computer Engineering – Greece

2010 – 2017

- Dissertation: Scheduling and Performance Characterization on Heterogeneous Computing Systems

University of Thessaly, MS in Computer Engineering – Greece

2008 – 2009

University of Thessaly, Diploma in Computer Engineering – Greece

2002 – 2007

Selected Publications

Proteus: Portable Runtime Optimization of GPU Kernel Execution with Just-In-Time Compilation

2025

G. Georgakoudis, K. Parasyris, D. Beckingsale
10.1145/3696443.3708939 (CGO 2025)

Programming GPUs with OpenMP and Python

2025

G. Georgakoudis, T. A. Anderson, S. Archibald, B. R. de Supinski, T. G. Mattson
10.1007/978-3-032-06343-4_14 (IWOMP 2025, Best Paper Award)

Enhancing Performance Through Control-Flow Unmerging and Loop Unrolling on GPUs

2024

A. Murtovi, *G. Georgakoudis*, K. Parasyris, et al.
10.1109/CGO57630.2024.10444819 (CGO 2024)

Scalable Tuning of (OpenMP) GPU Applications via Kernel Record and Replay

2023

K. Parasyris, *G. Georgakoudis*, et al.
10.1145/3581784.3607098 (SC 2023, Best Paper Finalist)

Recognition

- Best paper awards (IWOMP 2025, IWOMP 2021, IWOMP 2020, DATE 2020, COSMIC 2013)
- LLNL Director's Excellence in Publication (2024)
- LLNL Achievement Award (2024, 2020)
- SC'23 Best Paper Finalist
- SC'21 Best Reproducibility Award