

Gerasimos Gerogiannis

gg24@illinois.edu

gergerog.github.io

4238, Thomas M. Siebel Center

201 N. Goodwin Avenue, Urbana, IL, 61801, USA

RESEARCH INTERESTS

- Computer architecture and high-performance computing.
- Accelerator-centric heterogeneous architectures for machine learning, graph analytics, and scientific workloads.
- Machine learning and its application to systems and computer architecture.

EDUCATION

University of Illinois at Urbana-Champaign (UIUC)

Urbana, IL, USA

PhD - Electrical and Computer Engineering

Aug. 2021 - present

- Thesis: 'Towards Accelerator-Centric Computing'
- Advisor: Prof. Josep Torrellas
- GPA: 4.0/4.0

University of Patras

Patras, Greece

Diploma in Electrical and Computer Engineering (BSc + MSc)

Sep. 2016 - June 2021

- Thesis: 'Reinforcement Learning for Task Offloading in Next Generation Networks: Algorithms and Hardware Acceleration'
- Thesis Supervisor: Prof. Alexios Birbas
- GPA: 9.83/10.00 (first of class)

AWARDS AND HONORS

MICRO PhD Forum

IEEE TCuARCH, 2025

- Selected to present my PhD research at the PhD Forum of the International Symposium on Microarchitecture.

Wen-mei W. Hwu Award

UIUC, ECE Department, 2025

- For demonstrating research expertise in the area of Computer Engineering.

Rambus Computer Engineering Fellowship

UIUC, ECE Department, 2024

- For demonstrating outstanding research performance in the area of Computer Engineering.

IEEE Micro Top Picks from Computer Architecture Conferences Top Pick

IEEE, 2024

- Awarded to the 12 most significant papers in computer architecture published in the previous year based on novelty and potential for long-term impact. Top Pick awarded for paper: 'Micro-Armed Bandit: Lightweight & Reusable Reinforcement Learning for Microarchitecture Decision-Making'.

IEEE Micro Top Picks from Computer Architecture Conferences Honorable Mention

IEEE, 2024

- Awarded to the 12 most significant papers in computer architecture published in the previous year based on novelty and potential for long-term impact. Honorable Mention awarded for paper: 'SPADE: A Flexible and Scalable Accelerator for SpMM and SDDMM'.

Conference Travel Grants for ISCA/MICRO/HPCA/ASPLOS

IEEE/ACM, 2023-2025

Student Excellence Award

University of Patras, Greece, 2021

- For the highest GPA among all 2020-2021 graduates of the ECE Department of the University of Patras.

Excellence Award

State Scholarships Foundation (IKY), Greece, 2021

- For the highest GPA among all 2020-2021 graduates of the ECE Department of the University of Patras.

Skouras Foundation Award

Skouras Foundation, Greece, 2020

- For the highest GPA among all undergraduate students in the ECE Department of the University of Patras.

Freshman Excellence Award

University of Patras, Greece, 2016

- For the highest grade in the National University Entrance Exams (Panhellenic Exams) among the freshmen students in the ECE Department of the University of Patras.

Distinctions in National High-school Student Competitions

2013-2016

- Distinctions in International Physics Olympiad, National Student Physics, Chemistry, and Math Competitions.

- [C12] '*GRANII: Selection and Ordering of Primitives in GRaph Neural Networks using Input Inspection*'
Damitha Lenadora, Vimarsh Sathia, **Gerasimos Gerogiannis**, Serif Yesil, Josep Torrellas, and Charith Mendis
In Proceedings of the International Symposium on Code Generation and Optimization (CGO), February 2026.
- [C11] '*NetSparse: In-Network Acceleration of Distributed Sparse Kernels*'
Gerasimos Gerogiannis, Dimitrios Merkouriadis, Charles Block, Annus Zulfiqar, Filippos Tofalos, Muhammad Shahbaz, and Josep Torrellas
In Proceedings of the International Symposium on Microarchitecture (MICRO), October 2025.
- [C10] '*DECA: A Near-Core LLM Decompression Accelerator Grounded on a 3D Roofline Model*'
Gerasimos Gerogiannis, Stijn Eyerman, Evangelos Georganas, Wim Heirman, and Josep Torrellas
In Proceedings of the International Symposium on Microarchitecture (MICRO), October 2025.
- [C9] '*Micro-MAMA: Multi-Agent Reinforcement Learning for Multicore Prefetching*'
Charles Block, **Gerasimos Gerogiannis** and Josep Torrellas
In Proceedings of the International Symposium on Microarchitecture (MICRO), October 2025.
- [C8] '*COGNATE: Learning-Based Acceleration of Sparse Tensor Programs on Emerging Hardware*'
Chamika Sudusinghe, **Gerasimos Gerogiannis**, Damitha Lenadora, Charles Block, Josep Torrellas, and Charith Mendis
In Proceedings of the International Conference on Machine Learning (ICML), July 2025.
- [C7] '*MeshSlice: Efficient 2D Tensor Parallelism for Distributed DNN Training*'
Hyoungwook Nam, **Gerasimos Gerogiannis**, and Josep Torrellas
In Proceedings of the International Symposium on Computer Architecture (ISCA), June 2025.
- [C6] '*Distributed-Memory Parallel Algorithms for Sparse Matrix and Sparse Tall-and-Skinny Matrix Multiplication*'
Isuru Ranawaka, Md Taufique Hussain, Charles Block, **Gerasimos Gerogiannis**, Josep Torrellas, and Ariful Azad
In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC), November 2024.
- [C5] '*Two-Face: Combining Collective and One-Sided Communication for Efficient Distributed SpMM*'
Charles Block*, **Gerasimos Gerogiannis***, Charith Mendis, Ariful Azad, and Josep Torrellas
*Co-first authors, order is alphabetical.
In Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), April 2024.
- [C4] '*HotTiles: Accelerating SpMM with Heterogeneous Accelerator Architectures*'
Gerasimos Gerogiannis, Sriram Aananthakrishnan, Josep Torrellas, and Ibrahim Hur
In Proceedings of the International Symposium on High Performance Computer Architecture (HPCA), March 2024.
- [C3] '*Micro-Armed Bandit: Lightweight & Reusable Reinforcement Learning for Microarchitecture Decision-Making*'
Gerasimos Gerogiannis and Josep Torrellas
In Proceedings of the International Symposium on Microarchitecture (MICRO), October 2023.
Selected as a Top Pick in IEEE Micro Top Picks from Computer Architecture Conferences.
- [C2] '*SPADE: A Flexible and Scalable Accelerator for SpMM and SDDMM*'
Gerasimos Gerogiannis, Serif Yesil, Damitha Lenadora, Dingyuan Cao, Charith Mendis, and Josep Torrellas
In Proceedings of the International Symposium on Computer Architecture (ISCA), June 2023.
Selected as a Honorable Mention in IEEE Micro Top Picks from Computer Architecture Conferences.
- [C1] '*Characterizing the Scalability of Graph Convolutional Networks on Intel® PIUMA*'
Matthew Adiletta, Jesmin Jahan Tithi, Emmanouil-Ioannis Farsarakis, **Gerasimos Gerogiannis**, Robert Adolf, Robert Benke, Sidharth Kashyap, Samuel Hsia, Kartik Lakhotia, Fabrizio Petrini, Gu-Yeon Wei, and David Brooks
In Proceedings of the International Symposium on Performance Analysis of Systems and Software (ISPASS), April 2023.

JOURNAL PUBLICATIONS

- [J2] '*Practical Online Reinforcement Learning for Microprocessors with Micro-Armed Bandit*'
Gerasimos Gerogiannis and Josep Torrellas
IEEE Micro Magazine, Top Picks in Computer Architecture Special Issue, July-August 2024.
- [J1] '*Deep Reinforcement Learning Acceleration for Real-Time Edge Computing Mixed Integer Programming Problems*'
Gerasimos Gerogiannis, Michael Birbas, Aimilios Leftheriotis, Eleftherios Mylonas, Nikolaos Tzanis, and Alexios Birbas
IEEE Access, January 2022.

WORKSHOP PUBLICATIONS (WITH PROCEEDINGS)

- [W1] '*Automated Data Selection for Efficient Cost Model Training to Optimize Sparse Matrix Kernels on Emerging Hardware Accelerators*'
Chamika Sudusinghe, **Gerasimos Gerogiannis**, Damitha Lenadora, Charles Block, Josep Torrellas, and Charith Mendis
In Proceedings of the Exploration in AI Today Workshop at ICML 2025.

US PATENTS (FILED)

- [P4] '*Unified Transfer Engine for Near-Core Compute Accelerators*'
Gerasimos Gerogiannis, Stijn Eyerman, and Wim Heirman
- [P3] '*Speculative Invocation of Accelerators in Out-of-Order Core Pipelines*'
Gerasimos Gerogiannis, Stijn Eyerman, and Wim Heirman
- [P2] '*Efficient On-chip Memory Bandwidth Configurations and Distribution for Artificial Intelligence (AI) Workloads*'
Wim Heirman, Stijn Eyerman, and **Gerasimos Gerogiannis**
- [P1] '*Methods and Apparatus for a Machine Learning Model Decompression Accelerator*'
Gerasimos Gerogiannis, Stijn Eyerman, Wim Heirman, and Evangelos Georganas
US Patent App. 18/927,638 November 2024.

WORK EXPERIENCE

- | | |
|--|---|
| Intel Corporation
<i>CPU Architecture & Performance Intern</i> | January 2024 - December 2025
(remote) Champaign, IL, USA |
| <ul style="list-style-type: none">• CPU architecture redesign for emerging machine learning applications such as large language models.• Research conducted in internship led to publication in MICRO'25, one paper under review, and 4 US patents. | |
| Intel Corporation
<i>Accelerator Architecture & Performance Intern</i> | May 2023 - August 2023
(remote) Champaign, IL, USA |
| <ul style="list-style-type: none">• Heterogeneous accelerator architectures to accelerate sparse matrix - dense matrix multiplication.• Research conducted in internship led to publication in HPCA'24. | |
| Intel Corporation
<i>Accelerator Architecture & Performance Intern</i> | May 2022 - August 2022
(remote) Champaign, IL, USA |
| <ul style="list-style-type: none">• Synergies between hardware and software to maximize performance on the Intel PIUMA accelerator.• Research conducted in internship led to publication in ISPASS'23. | |
| i-acoma group, University of Illinois at Urbana-Champaign
<i>Research Assistant</i> | August 2021 - present
Champaign, IL, USA |

TALKS

- [T6] '*Accelerator-Centric Computing*'
PhD Forum at the 58th International Symposium on Microarchitecture, October 2025.
- [T5] '*Sparsity at Scale: Towards Efficient Distributed Sparse Accelerators*'
SRC TECHCON, September 2024.
- [T4] '*Micro-Armed Bandit: Lightweight & Reusable Reinforcement Learning for Microarchitecture Decision-Making*'
Intel Archfest, May 2024.
- [T3] '*Micro-Armed Bandit: Lightweight & Reusable Reinforcement Learning for Microarchitecture Decision-Making*'
ACE Center for Evolvable Computing Theme Meeting, December 2023.
- [T2] '*Domain-Specific Hardware and Software for Mixed Sparse-Dense Algebra at Scale*'
ACE Center for Evolvable Computing Theme Meeting, August 2023.
- [T1] '*Accelerators for Irregular Applications*'
IBM-Illinois Discovery Accelerator Institute Annual Meeting, March 2023.

SELECTED PROJECTS

- MiniSPADE** 2022-2023
- Designed the microarchitecture and taped-out a simplified ASIC prototype of the accelerator described in our ISCA'23 paper using TSMC 65nm technology.

MENTORING AND TEACHING EXPERIENCE

- Formal Training** Mavis Future Faculty Fellows (MF3) Program, UIUC, 2025 – present
- Received formal training in both mentoring and teaching as part of the MF3 program.
- Graduate Student Mentoring** UIUC, 2022 – present
- Dimitrios Merkouriadis*, PhD student at UIUC: In-Network Acceleration for Sparse Kernels, 2024 - present.
 - Filippos Tofalos*, PhD student at UIUC: In-Network Acceleration for Sparse Kernels, 2024 - 2025.
 - Chamika Sudusinghe*, PhD student at UIUC: Accelerator Performance Modeling and Tuning, 2023 - present.
 - Charles Block*, PhD student at UIUC: Algorithms for Distributed Sparse Kernels, 2022 - 2024.
- Private Tutor** Patras, Greece, 2016 – 2020
- Tutor** Prosimo Education Center, Patras, Greece, 2020 – 2021
- Gave lectures and tutored undergraduate students in preparation for their exams in over 15 electrical and computer engineering subjects.

GRANT WRITING EXPERIENCE

- NSF** 2023 – 2028
- PPoSS: LARGE: General-Purpose Scalable Technologies for Fundamental Graph Problems* Amount: \$5,000,000
- Lead PI: Josep Torrellas, number of PIs: 10
 - Assisted in the preparation and writing of the NSF Grant proposal, focusing on hardware support for scalable graph algorithms and on the hardware-software interaction.

PROFESSIONAL MEMBERSHIPS

- Student Member** May 2023 – present
- ACM

REVIEWER SERVICE

- IEEE Transactions on Computers** 2025
- Invited Reviewer

TECHNICAL SKILLS

Programming Languages: C/C++, Python, Matlab, Shell/Bash scripting, Perl, VHDL, Verilog.
Parallel Programming: OpenMP, MPI, Posix Threads, CUDA.
Frameworks/Libraries: Vivado, Vivado HLS, Cadence EDA tools, TensorFlow, PyTorch, Deep Graph Library.
Microarchitectural Simulators and Tools: SST, Sniper, gem5, ChampSim, Accel-Sim, DRAMSim, CACTI, McPAT.
Development Tools: git, CMake.