Mariam Rakka, PhD

rakka.mariam@icloud.com | mariamrakka.github.io

in linkedin.com/in/mariam-rakka | ♦ scholar.google.com/citations?user=Fb9_7EMAAAAJ&hl=en

Austin, TX, USA

SUMMARY

Computer architecture researcher focused on **systems for machine learning** and **in-memory computing**, with a PhD (UCI, 06/2025). I bring hands-on experience modeling and scaling **multi-chiplet heterogeneous systems** in **gem5**, conducting workload-driven **performance calibration** and **kernel-level analysis** (C/C++/Python), and leading **software-hardware co-design** efforts that resulted in two patent filings and first-author publications (e.g., TPAMI, JETC, TETC, DATE). I aim to advance **high-performance, energy-efficient ML acceleration** on server-class architectures through architecture research, workload characterization, and cross-functional collaboration.

INDUSTRY EXPERIENCE

- **Arm, Inc.** _06/2025 – Present_
  Senior Research Engineer Austin, Texas
  ○ Modeling **multi-chiplet heterogeneous systems** in **gem5** to evaluate scalability and performance across diverse workloads, including machine learning kernels
  ○ Designing and executing **performance calibration** studies; collecting memory bandwidth, cache hit rates, and latency distributions to diagnose bottlenecks and inform system-level tradeoffs
  ○ Performing **kernel-level analysis** (C/C++/Python) of instruction mix, vectorization, memory access locality, and synchronization patterns, and proposing software–hardware optimizations aligned with internal roadmaps
  ○ Building automated **build/run/post-processing pipelines** (Python/Bash) to ensure reproducibility and scale, and collaborating via Git
  ○ Authoring internal reports and presenting findings to architecture/research teams to drive design decisions

- **Arm, Inc.** _06/2024 – 09/2024_
  Computer Architecture Research Intern Austin, Texas
  ○ Characterized a diverse set of **machine learning workloads**, using **perf** and **gem5** to profile instruction mixes, memory access patterns, and microarchitectural bottlenecks on Arm-based CPU systems
  ○ Identified key performance bottlenecks (e.g., vectorization inefficiencies, memory bandwidth limitations) and **conceptualized a specialized SIMD-capable hardware accelerator** to target these critical computational kernels
  ○ Estimated **hardware resource requirements**, including area, cache utilization, and memory bandwidth using internal modeling tools, and explored the integration of the accelerator into the Arm pipeline
  ○ Co-designed **software-hardware interfaces and execution flows** to maximize offloading efficiency and ensure compatibility with existing Arm toolchains, demonstrating significant performance improvements over baseline
  ○ _Contributed to two patent filings based on the developed architectural techniques (already filed, not public yet)_

- **Arm, Inc.** _06/2023 – 09/2023_
  Computer Architecture Research Intern Austin, Texas
  ○ Investigated a **tightly coupled CPU architecture extension** aimed at accelerating data analytics workloads on Arm-based systems and contributed to early-stage architectural evaluation and design discussions
  ○ Deployed and characterized **Apache Hadoop and Spark** benchmark suites on a multi-node cluster to exercise diverse data processing kernels representative of real-world analytics pipelines
  ○ Performed detailed **bottleneck analysis** using **perf** and **Java Flight Recorder**, identifying instruction-level and memory access inefficiencies and outlining optimization opportunities enabled by the proposed extension

- **Arm, Inc.** _06/2022 – 09/2022_
  Machine Learning Research Intern Austin, Texas
  ○ Designed and developed a **machine learning–based automated design space exploration framework** to efficiently identify high-performance, low-cost configurations of Arm processor designs, targeting complex trade-offs
  ○ Implemented and integrated **reinforcement learning**, **deep neural networks**, and **active learning** pipelines in Python for iterative exploration and developed robust data preprocessing, feature engineering, and postprocessing modules to enable scalable, statistically sound evaluations
  ○ Interfaced the ML framework with an **Arm hardware simulator** through automated Bash workflows, enabling end-to-end generation, simulation, and analysis of thousands of design points without manual intervention

ACADEMIC EXPERIENCE

- **University of California, Irvine** *10/2020 – 05/2025*
  Graduate Student Researcher Irvine, CA
  - Led the effort to propose a reconfigurable in-memory accelerator for Large Language Model inference in collaboration with KAUST
  - Surveyed mixed-precision neural networks, which resulted in a first-author publication in TPAMI
  - Proposed an analytical simulator for an in-memory Hyperdimensional Reinforcement Learning accelerator, resulting in a first-author journal publication in JETC
  - Implemented an accurate functional simulator for faster Decision Tree inference using Python and MATLAB, resulting in a first-author journal publication in TETC

- **King Abdullah University of Science and Technology** *04/2024 – 05/2024*
  Visiting Student Researcher Thuwal, Saudi Arabia
  - Developed a novel analytical simulator for a reconfigurable in-memory Convolutional Neural Network inference accelerator, resulting in a first-author paper under review
  - Conducted a software-hardware co-design for a proposed in-memory inference accelerator for modules in Large Language Models using Python, resulting in a first-author paper published in DATE
  - Met and discussed new potential machine learning acceleration projects with collaborators at KAUST

- **University of California, Irvine** *09/2021 – 03/2022, 01/2025 – 03/2025*
  Teaching Assistant (Introduction to Python Programming/Advanced C Programming/Intro to Digital Systems) Irvine, CA
  - Taught students how to solve mathematical formulations, process images using data structures in Python/C, and digital design principles
  - Prepared assignments and exams and delivered weekly online and in-person discussion sessions and labs.
  - Met regularly with the course Professors to present progress and developed Bash scripts for grading

- **University of California, Irvine** *07/2019 – 09/2019*
  Hardware Engineering Intern Irvine, CA
  - Implemented a hardware-based real-time temperature tracking solution on the Nexys Video FPGA
  - Utilized the DRP interface of the XADC soft core to tap into registers where temperature data is stored and customized a FIFO core to store different instances of temperature data
  - Designed a DAC module and displayed the analog temperature value on the OLED of the FPGA
  - Acquired knowledge about the RISC-V processor, the Information Processing Factory Project, Vivado tools, and different FPGA boards
  - Extended the internship as a senior-year project collaboration between UCI and AUB and developed dedicated hardware modules for the Trace Abstraction Layer of the Information Processing Factory to enable real-time non-intrusive on-chip FPGA system verification using VHDL and C, which resulted in a second-author publication in VLSI-SoC and another publication in DATE

- **American University of Beirut** *01/2018 – 06/2020*
  Undergraduate Student Researcher Beirut, Lebanon
  - Explored the energy, latency, and quality of various state-of-the-art Resistive RAM-based Ternary CAM implementations, which resulted in a first-author publication in TCAS-II
  - Improved the efficiency of rare-fail event estimation statistical methodologies by proposing novel hybrid algorithm designs and testing them
  - Implemented the proposed algorithms in MATLAB/PERL/HSPICE and evaluated them on 16nm SRAM designs, resulting in two publications in ISCAS

EDUCATION

- **University of California, Irvine** *09/2020 – 06/2025*
  MS/PhD in Electrical and Computer Engineering Irvine, CA
  - PhD in Electrical and Computer Engineering **conferred 06/2025**; Dissertation: *In-Memory Computing to Accelerate Machine Learning*
  - MS in Electrical and Computer Engineering **conferred 06/2022**; Thesis: *Resistive Content Addressable Memory Design for Decision Tree Acceleration*
  - GPA: 4.00/4.00

- **American University of Beirut** *08/2016 – 05/2020*
  Bachelor of Computer and Communications Engineering Beirut, Lebanon
  - Graduated with high distinction
  - Minored in Mathematics and Business Administration
  - Cumulative GPA: 88.67/100

Google Scholar metrics (as of 10/07/2025): 88 citations; h-index = 4.

[C.1] Mariam Rakka, et al. (2025). **SoftmAP: Software-Hardware Co-design for Integer-Only Softmax on Associative Processors**. *2025 Design, Automation & Test in Europe Conference & Exhibition (DATE).*

[T.1] Mariam Rakka. (2025). **In-Memory Computing to Accelerate Machine Learning**. *University of California, Irvine.*

[S.1] Mariam Rakka, et al. (2025). **Mixed-Precision Quantization for Language Models: Techniques and Prospects**. *ArXiv preprint arXiv:2510.16805.*

[J.1] Mariam Rakka, et al. (2024). **A Review of State-of-the-Art Mixed-Precision Neural Network Frameworks**. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[J.2] Mariam Rakka, et al. (2024). **HDRLPIM: A Simulator for Hyper Dimensional Reinforcement Learning based on Processing In Memory**. *ACM Journal on Emerging Technologies in Computing Systems.*

[S.2] Mariam Rakka, et al. (2024). **BF-IMNA: A Bit Fluid In-Memory Neural Architecture for Neural Network Acceleration**. *ArXiv preprint, arXiv:2411.01417.*

[L.1] Walaa Amer, Mariam Rakka, et al. (2024). **FPonAP: Implementation of Floating Point Operations on Associative Processors**. *IEEE Embedded Systems Letters.*

[J.3] Mariam Rakka, et al. (2023). **DT2CAM: A Decision Tree to Content Addressable Memory Framework**. *IEEE Transactions on Emerging Topics in Computing.*

[C.2] Walaa Amer, Mariam Rakka, et al. (2023). **Hardware Implementation and Evaluation of an Information Processing Factory**. *2023 IFIP/IEEE 31st International Conference on Very Large Scale Integration (VLSI-SoC).*

[C.3] Nora Sperling, Alex Bendrick, Dominik Stöhrmann, Rolf Ernst, Bryan Donyanavard, Florian Maurer, Oliver Lenke, Anmol Surhonne, Andreas Herkersdorf, Walaa Amer, Caio Batista de Melo, Ping-Xiang Chen, Quang Anh Hoang, Rachid Karami, Biswadip Maity, Paul Nikolian, **Mariam Rakka**, Dongjoo Seo, Saehanseul Yi, Minjun Seo, Nikil Dutt, and Fadi Kurdahi. (2025). **Information Processing Factory 2.0 – Self-Awareness for Autonomous Collaborative Systems**. *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE).*

[T.2] Mariam Rakka. (2022). **Resistive Content Addressable Memory Design for Decision Tree Acceleration**.

[C.4] Mariam Rakka, et al. (2021). **Importance Splitting Sample Point Reuse for Efficient Memory Yield Estimation**. *2021 IEEE International Symposium on Circuits and Systems (ISCAS).*

[J.4] Mariam Rakka, et al. (2020). **Design exploration of sensing techniques in 2T-2R resistive ternary CAMs**. *IEEE Transactions on Circuits and Systems II: Express Briefs.*

[C.5] Mariam Rakka, et al. (2020). **Hybrid importance splitting importance sampling methodology for fast yield analysis of memory designs**. *2020 IEEE International Symposium on Circuits and Systems (ISCAS).*

PROJECTS

• **Canny Edge Decoder** *01/2021 – 04/2021*
Tools/Platforms: C, SpecC, RISC-V
◦ Designed a multi-threaded version of the Canny Edge Decoder

◦ Migrated the sequential C code to a RISC-V Virtual Platform with LCD and Camera device drivers' support

◦ Completed a parallel, pipelined system-level design of the Canny Edge Decoder, and received an A for the projects as part of EECS222: Embedded System Modeling and EECS226: Embedded System Software

• **Novel 4-Bit ALU** *01/2019 – 05/2019*
Tools: Cadence
◦ Proposed a novel efficient 4-Bit ALU design using modified Shannon theorem

◦ Reviewed related works in literature

◦ Designed the adder using 90nm technology, and performed schematic-level simulations

◦ Demonstrated that the ALU is more compact, faster, and more power efficient when compared to the CMOS-based design, and received an A for the project as part of EECS412L: VLI Computer Aided Design Lab

• **Quantum Cellular Automata Literature Review** *01/2019 – 05/2019*
Tools: Research
◦ Conducted an in-depth **literature review on Quantum Cellular Automata (QCA)** as a nanotechnology computing paradigm, focusing on fundamental device principles, layout structures, and clocking mechanisms.

◦ Analyzed and compared multiple **QCA logic gate implementations** (e.g., majority gates, inverters, XOR structures), highlighting trade-offs in area efficiency, signal propagation, and fabrication feasibility.

◦ Delivered a well-structured written report and presentation as part of a VLSI course, **earning a full grade** for technical depth, clarity, and critical analysis.

SKILLS

- **Programming Languages:** C, C++, Python, MATLAB, SpecC, HDL (Verilog/VHDL), Gem5, OrCAD PSpice, HSPICE, Perl, Shell Scripting, LaTeX
- **Machine Learning Models:** Large Language Models, Convolutional Neural Networks, Decision Trees, Reinforcement Learning
- **Other Tools:** Perf, Java Flight Recorder, Flame Graphs, Visual Studio Code, Cadence, Github, Overleaf, Microsoft Office programs
- **Specialized Area:** In-memory computing, Hardware acceleration, Statistical analysis methodologies
- **Soft Skills:** Research, Teambuilding, Leadership, Written and verbal communication, Time management

CERTIFICATIONS AND AWARDS

- **iREDEFINE Fellow** *03/2025*
  National Science Foundation
- **10-week Hardware Design Program** *05/2023*
  VLSI System Design
- **High Level Synthesis Tutorial – DAC 2021** *01/2022*
  Cadence Design Systems
- **DAC'21 Young Fellow** *12/2021*
  Design Automation Conference
- **Best Computer Hardware System Project** *05/2020*
  Maroun Semaan Faculty of Engineering and Architecture
- **Dean's Honor List** *09/2016 – 06/2020*
  Maroun Semaan Faculty of Engineering and Architecture
- **One-year PhD Fellowship (Merit-Based)** *09/2020 – 09/2021*
  University of California, Irvine

PROFESSIONAL SERVICE & MENTORING

- **Peer Reviewer** *2020 – 2025*
  Journals & Conferences
  - 2026 Design Automation Conference (The Chips to Systems Conference)
  - IEEE Transactions on Circuits and Systems for Artificial Intelligence
  - 2026 Design, Automation and Test in Europe Conference
  - IEEE Circuits and Systems Magazine
  - 22nd International Conference on Hardware/Software Codesign and System Synthesis
  - IEEE 66th International Midwest Symposium on Circuits and Systems
  - IEEE Transactions on Circuits and Systems I: Regular Papers
  - ACM Transactions on Architecture and Code Optimization
  - IEEE Transactions on Computers
- **Mentor, Techotolia Robotics (Team #7748)** *07/2025 – Present*
  FIRST Robotics Competition
  - Mentoring a **high-school** FRC team via regular Zoom meetings
  - Providing technical guidance, planning, and materials preparation to support competition readiness
  - Applying for funding on behalf of the team to help them cover registration costs and material costs
- **Mentor** *08/2018*
  All Girls Code
  - Mentored a hackathon where girls developed apps and made websites tackling tech and health
  - Participated in organizing the hackathon and answering questions, and connected with other mentors

VOLUNTEERING EXPERIENCE

- **Youth Member** *10/2018 – 05/2019*
  Lebanese Red Cross
  - Underwent training courses such as prevention from sexually transmitted diseases and disaster and preparedness
  - Engaged in activities and helped in organizing some that are targeted toward the elderly segment in Lebanon
- **Tutor** *07/2018 – 08/2018*
  BASSMA
  - Helped children as a part of the Night School Program sponsored by Touch Lebanon
  - Supported public school students in the 7th, 8th, and 9th grades, who are drawn from underprivileged families
- **Writer** *10/2018 – 06/2020*
  AUB Outlook
  - Published Arabic articles about different social issues
  - Met with the team of writers and editors on weekly basis

ADDITIONAL INFORMATION

Languages: English (Fluent), Arabic (Fluent), French (Elementary Proficiency)
Interests: Writing articles, Reading novels, Watching crime docu-series, Hiking, Chasing sunsets, Weightlifting

Additional references are available upon request