



Ivris B. Raymond

Ph.D. Candidate - Computer Science & Engineering

Email: ivris@umich.edu
University of Michigan

Education

- **University of Michigan** Ann Arbor, MI
Ph.D in Computer Science & Engineering *Expected August 2028*
– Advisor: Ronald G. Dreslinski
- **University of Michigan** Ann Arbor, MI
M.S. Computer Science & Engineering *December 2025*
- **University of Arkansas** Fayetteville, AR
B.S. in Computer Engineering *May 2023*
– GPA: 4.0
– Honors Thesis: Digital Simulations of Memristors Towards Integration with Reconfigurable Computing

Employment

- **Hands-on Training in Semiconductors** Ann Arbor, MI
Lecturer & Lab Instructor *2025 - 2026*
A collaborative effort between several faculty and students at the University of Michigan to provide education on semiconductor design, manufacturing, and applications of semiconductors (control systems, AI, etc.) to high school students.
– Gave lectures introducing students to a wide variety of topics in embedded system programming, computer architecture, and ML.
– Developed some instructional material for this initiative. Some material I developed is available online at arenile.github.io/hotis-docs-site.
- **University of Arkansas** Fayetteville, AR
Research/Teaching Assistant *Summer 2021 - Spring 2022, Fall 2022 - Spring 2023*
– Designed hardware for an OBD-II interface device and assisted in implementing an RTOS for this device.
- **Phillips 66** Houston, TX
IT Intern *Summer 2022*
– Identified communication and reporting issues in diagnostic software and architected a holistic dashboarding solution for reporting these issues in third-party software. This solution provides a usable problem notification and identification mechanism to Analysts with non-technical backgrounds.
– Implemented a stop-gap solution for catching these problems until the long-term solution is able to be implemented by the development teams.

Selected Publications

- **GLEAM: Graph-based Learning through Efficient Aggregation in Memory** DATE 2025
Andrew McCrabb, Ivris Raymond, Valeria Bertacco
GLEAM is a processing-near-memory accelerator for Graph Neural Networks (GNNs) that achieves up to a 10× speedup by leveraging a bank-level processing unit to perform partial aggregations near memory.
– I designed the GLEAM architecture and performed the theoretical analysis of its performance along with the area analysis.
– I handled most of the graph data preparation for benchmarks that were run to evaluate GLEAM's performance.

Research Projects

- **World Models for Microarchitectural Design Space Exploration** Summer 2026 - Current
Ivris Raymond, Ron Dreslinski

With the rising adoption of ML for design space exploration, current approaches to simulating designs with meaningful fidelity have become a substantial bottleneck for the number of designs that ML-based approaches can evaluate when exploring the design space. This project examines whether world models may offer a faster alternative to traditional approaches to detailed microarchitectural simulation while maintaining enough fidelity to converge on good designs.

- Designing a transformer-based world model for microarchitectural state capable of modeling how execution of a given workload affects state across various microarchitectural configurations.
- Conducted a motivational study on the need for faster simulators to support emerging ML-based techniques for design space exploration in computer architecture.

- **ML for Maintaining Obsolete Digital Systems**

Spring 2025 - Spring 2026

Ivris Raymond, Sitota Mersha, Valeria Bertacco

Computing systems in high-reliability deployments are often qualified for those deployments at system-level granularity while failures often occur at subsystem-level. Obsolescence poses a high risk to systems of this manner, as failing components often necessitate a complete system redesign or at least re-qualification, which is itself expensive and time-consuming. This project examines whether LLMs can be used to ease this burden by engineering drop-in replacements for failing, obsolete digital components based on the behavior of the original component and the broader design specifications of the system.

- Implemented an agent orchestration system designed to support LLM-based agents in writing RTL for obsolete digital components with hierarchical structure to improve legibility for human engineers.
- Designed and implemented a Retrieval Augmented Generation (RAG) system to provide additional context based on existing system documentation and the functional behavior of the obsolete digital component.

Selected Engineering Projects

- **Applying Graph Algorithms for MoE Token Batching**

Coursework — Advanced Operating Systems (CSE 582)

Fall 2024

Current LLMs often leverage Expert Parallelism to improve inference throughput. This poses challenges for routing tokens between physical devices while also batching effectively for throughput optimization, particularly for long context inputs. We applied a graph-based algorithm in this project and measured its effectiveness for routing tokens among physical devices assuming a model was deployed using expert parallelism.

- **Efficient Memory Access Monitoring and Bug Detection**

Coursework — Advanced Compilers (EECS 583)

Fall 2023

Many existing tools for memory tracing and bug detection rely on dynamic program analysis. For this project, we implemented ideas presented in Spindle (Wang, et al.) to perform memory tracing and bug detection using static analysis, achieving constant execution time and trace size scaling with array element count compared to the increasing time and size scaling of tools like Intel Pin and Valgrind. Our implementation is available on GitHub at github.com/EECS-583-Group-4/final-project.git.

- **RTOS Driven Autonomous Vehicle**

Coursework — Embedded Systems (CSCE 4114)

Fall 2022

Developed an autonomous vehicle capable of navigating around a track of reflective tape using two simple light sensors using an FPGA board. The system implemented a MicroBlaze SoC on the FPGA and utilized the FreeRTOS Kernel for scheduling.

- **Exploration of Rowhammering Techniques**

Coursework — Embedded Systems Security (CSCE 5013)

Spring 2022

Conducted a survey of various techniques for rowhammering and their implementation schemes, then produced a report on how a synthesis of new techniques may be capable of breaking some of the finalists in the NIST PQC Standardization Process by serving as the fault-injection mechanism.

- **Power Usage of AES Power Analysis Attack Mitigation Strategies**

Coursework — Low Power Digital Systems (CSCE 4233)

Spring 2022

Analyzed the power usage characteristics of a microcontroller performing AES encryptions across a variety of AES implementations. These implementations used many variations and combinations of masking and hiding techniques to improve their resiliency against power analysis attacks.

Selected Awards

- **2023** NSF Graduate Research Fellowship
- **2023** University of Michigan Rackham Engineering Award
- **2023** University of Arkansas, Computer Engineering Outstanding Senior
- **2019 - 2022** University of Arkansas Chancellor's List
- **2021** University of Arkansas Honors College Undergraduate Research Grant
- **2019** University of Arkansas National Merit Scholar

Other Technical Experience

- **PyTorch** ML training, particularly for transformer-based world models
- **gem5** CPU microarchitecture simulation
- **PydanticAI**, **FastMCP**, **LlamaIndex** Building applications that apply LLM agents and RAG
- **LLVM** Constructing compiler passes for static analysis of workload characteristics
- **Vulkan** Implementing real-time 3D rendering engines