Jiahan Xie
Phone: (607)280-7683
Email: jxie84@ucsc.edu
Website: https://jiahanxie353.github.io/

Education

- **UC Santa Cruz** — Santa Cruz, CA
  PhD student in Computer Science — *Expected 2030*
- **Cornell University** — Ithaca, NY
  Master of Science in Computer Science — *Aug 2023 - May 2025*
- **Cornell University** — Ithaca, NY
  Bachelor of Science in Computer Science and Environmental Engineering — ***Magna Cum Laude***
- **Joint Degree Program** with **Zhejiang University** — Hangzhou, China
  Bachelor of Engineering in Environmental Engineering

Internship Experience

- **Machine Learning Compiler Engineer** — Jun 2025 - Now
  Cerebras Systems — *Sunnyvale, CA*
  - Applied liveness analysis and profiling tools to optimize memory scheduling and visualization on Perfetto
  - Built a two-pass compile workflow to reduce wafer layout selection optimization process for large MoE models
- **Software Engineer** — Jun 2023 - Aug 2023
  Orenda Power Inc. — *New York City, NY*
  - Built a power grid database with real-time processing and designed a REST API for IoT control operations
- **Machine Learning Compiler Engineer** — May 2022 - Aug 2022
  Deep Ivy Inc. — *Remote*
  - Developed framework converters and a graph compiler to accelerate ML model transformations

Research Experience

- **Graduate Researcher** — Sep 2025 - Now
  UC Santa Cruz *MASC* Group, led by José Renau
  - Building an AI hardware agent from RTL code generation, verification, debugging, and tapeout.
- **Graduate Researcher** — Sep 2023 - May 2025
  Cornell *Capra* Group, led by Adrian Sampson
  - Built and optimized a PyTorch-to-FPGA compiler flow, achieving 1.7× performance over AMD Vitis
- **Undergraduate Researcher** — Nov 2021 - May 2023
  Cornell *PEESE* Group, led by Fengqi You
  - Created a multi-agent reinforcement learning model for smart city energy management

Talks and Publications

"Global Instruction Selection for Scalable Vectors." Students Technical Talk at LLVM Developer Conference 2024.

Xie, J., Ajagekar, A., You, F. "Multi-Agent Attention-Based Deep Reinforcement Learning for Demand Response in Grid-Responsive Buildings." *Applied Energy.* Link

Project Experience

- **Distributed Multi-GPU Influence Maximization (CUDA)** — March 2025 - May 2025
  - Implemented multi-GPU distributed algorithms in CUDA with MPI for large-scale influence maximization, focusing on GPU-parallel SelectSeeds optimization and scalability analysis across SNAP datasets.
- **Multidecree Paxos (Distributed Systems)** — Oct 2024 - Nov 2024
  - Built a multidecree Paxos system ensuring linearizable consensus and fault tolerance under failures, with leader election, AMO semantics, and majority-based decision making.
- **Systolic Array Design for Binarized Matrix Multiplication (Dataflow architecture)** Oct 2024 - Nov 2024
  - Designed and optimized an FPGA systolic array for binarized matmul using Allo, achieving a 15.2× speedup
- **Global Instruction Selection for RISC-V Vector Extension (LLVM)** — Nov 2023 - Oct 2024
  - Extended GISel for RISC-V vectors, enabling scalable vector support for SAXPY lowering from C to assembly
- **Compiler Development and Optimization for Bril (Advanced Compilers, C++)** Aug 2023 - Dec 2023
  - Built a Bril backend with compiler optimizations (LVN, DCE, LICM), achieving a 10.7% benchmark speedup
- **Pipelined RISC-V Processor with Cache (Computer Architecture, RTL Design)** Sep 2023 - Nov 2023
  - Implemented pipelined RISC-V processors with stalling, bypassing, and instruction/data caches

Coding Languages

- C++, Python, C, JAVA, OCaml