Kangqi Zhang

zhkangqi@umich.edu | github | Personal website | www.linkedin.com
Undergraduate Student, University of Michigan, Ann Arbor, USA

Education

University of Michigan 2024/08 – 2026/05(expected)
- Bachelor of Science in Computer Engineering, GPA: 4.0/4.0

Shanghai JiaoTong University 2022/08 – 2026/08(expected)
- Bachelor of Science in Mechanical Engineering, Core GPA: 3.65/4.0

Coursework: Computer Architecture, Parallel computer architecture, Microarchitecture, Parallel GPU Programming, Operating Systems, Scalable Systems, Digit Integrated Circuit, Data Structures and Algorithms

Research Experience

RL Driven Turn-Restricted Adaptive Routing for Chiplet-Based System, Advisor:**Nathaniel Bleier** 2024/09 – Present
- Developed a fast turn addition algorithms and baseline for arbitrary topologies.
- Applied Deep Q-Learning to learn optimal turn sets under different traffic patterns.
- Evaluated with Gem5 Garnet and PARSEC workloads, demonstrating robust latency reduction and scalability.

Long context LLM inference acceleration with ANNS search, Advisor: **Jishen Zhao, Zhijian Liu** 2025/05 – Present
- Proposed a CPU–GPU heterogeneous ANN search pipeline for efficient KV-cache retrieval.
- Exploited inter-layer query similarity by reusing clusters across adjacent Transformer layers, forming a cross-layer search pipeline.

Three-dimensional Multi-physics Modeling of Li-ion Batteries, Advisor: **Yanming Wang** 2023/02 – 2024/02
- Develop a comprehensive three-dimensional multi-physics model for lithium-ion batteries with mechanical and thermal aspects.
- Successfully predicted battery aging behavior, enabling a more accurate assessment of battery performance and longevity.

Course Project

EECS 470 Out-of-Order RISC-V Superscalar Processor 2025/02 – 2025/05
- Proposed framework for RTL design on MIPS R10K microarchitecture with RV32-IM ISA.
- Developed advanced feature includes N-way superscalar execution, early branch resolution, early tag broadcast, prefetching, non-blocking associative caches and a cycle accurate simulator.
- Integrated advanced branch prediction subsystems: TAGE, perceptron, and CNN-based predictors, along with a branch target buffer (BTB) and return address stack (RAS).
- Achieved an average CPI of 1.5 under 10ns clock period. Final system ranked in the top 2 of the course.

EECS 573 Adaptive Feature Selection for Heterogeneous Core Scheduling 2025/09– Present
- Extended GEM5/QEMU to extract comprehensive MICA features and modeled realistic migration costs
- Designed an information-theoretic feature selection framework to identify Pareto-optimal MICA subsets balancing accuracy and overhead.

EECS 471 GPU Optimization of CNN Forward Pass
- Implemented the forward pass of a convolutional neural network (CNN) layer in CUDA with performance-critical kernels.
- Applied shared memory tiling, register-level GEMM transformations, and loop unrolling, reducing runtime from 13.5s to 0.11s.

Intellectual Property & Publications

H.Jin, J.Yang, Y.Liu, B.Lyu, K.Zhang, N.Bleier. Mozart: An Ecosystem-Accelerator Codesign Framework for Composable Heterogeneous Chiplet Based Neural Network Accelerators. ASPLOS 2026, under review.

PyBatSim V1.0: Multi-physics Simulation Software for Li-ion Cells
Co-developer, Software Copyright (China National Copyright Administration) 2024/10

Teaching

Teaching assistant for ECE 2150 (Intro to Electric Circuits) 2024/9-2024/12

Honors and Awards

- Summer Undergraduate Research in Engineering (SURE) Stipend, University of Michigan 2025/05
- Dean's Honor List at College of Engineering, University of Michigan 2024/12, 2025/05

Skills

- Programming language: C/C++, Python, CUDA, Verilog, System Verilog, Mathematica, MATLAB, R
- Architecture Simulators & Tools: gem5, QEMU
- Machine Learning Frameworks: PyTorch, HuggingFace Transformers