

Hongzheng Tian

hongzheng@uci.edu | +1-951-202-2793 | Irvine, CA
hongzhengtian.com | GitHub | LinkedIn | Google Scholar

Education

UC Irvine — Ph.D. in Computer Engineering Jun. 2027 (expected)
UC Irvine — M.S. in Electrical and Computer Engineering, 2022
Northwestern Polytechnical University — B.Eng., 2020
Microelectronics Science and Engineering

Research & Engineering Experience

UC Irvine — CoVA: MLIR Compiler for CPU/GPU Computing Jun. 2023 – Present
Doctoral research; project initiated during the 2023 HPE internship

- **Compiler pipeline:** Built a numerical Python compiler integrating custom MLIR passes with upstream bufferization and lowering infrastructure for CPU LLVM/OpenMP and NVIDIA GPU LLVM/NVVM backends.
- **Parallelism & storage:** Implemented loop and region storage analyses using access partitions and alias information to prove loop independence and enable GPU-resident buffers where safe across control flow.
- **GPU mapping & reductions:** Mapped independent loop domains to GPU thread blocks; implemented fixed-group reduction trees with results independent of physical thread count for supported reductions.
- **CUDA runtime:** Implemented per-invocation stream reuse, temporary-buffer reuse, and analysis-guided synchronization to avoid unnecessary waits between eligible kernel launches.
- **Validation:** Checked numerical correctness with NPBench; evaluated compiler changes through paired runs and ablations, recording CPU/NUMA placement to identify measurement sensitivity.
- SC24 ACM Student Research Competition poster.

HeteroTask — Distributed Heterogeneous Task Runtime Jan. 2025 – Apr. 2026

- **Offline scheduling:** Independently built a DAG execution framework with task-variant profiling, memory-aware device/variant selection, communication-cost modeling, and buffer lifetime planning.
- **Distributed execution:** Implemented a C++/MPI runtime to execute planned task placements, dependency transfers, and buffer allocation/free operations; scheduling concepts later informed CoVA.

Hewlett Packard Enterprise — Research Associate Intern Jun. – Sep. 2024

HeteroBench: Multi-kernel Benchmarks for Heterogeneous Systems

- **Benchmark framework:** Independently implemented a CPU/GPU/FPGA suite using Python, OpenMP, CUDA, and HLS implementations for evaluating heterogeneous HPC workloads.
- **Build & run automation:** Implemented a JSON-configured Python driver with Jinja-generated Makefiles for backend-specific builds, benchmark execution, profiler integration, and experiment logging.
- ICPE 2025: **Best Paper Award – Artifact Track.** Code: HewlettPackard/HeteroBench.

Hewlett Packard Enterprise — Research Associate Intern Jun. – Sep. 2023

- Developed CoVA's initial MLIR dialects and passes separating program semantics from target-specific lowering.

UC Irvine — PyAIE: Python Programming for Versal AI Engines Sep. 2022 – Jun. 2023

- **End-to-end code generation:** Independently implemented a Python-based framework generating AI Engine kernel code, host code, and configuration scripts for AMD Versal ACAP.
- DAC 2023 Late Breaking Results; LATTE 2023 presentation.

Technical Skills

Languages: C, C++, Python, Verilog HDL

Compilers & parallel computing: MLIR, LLVM, OpenMP, CUDA, MPI, HLS

Platforms & tools: CPUs, GPUs, FPGAs, AMD Versal AI Engines, Xilinx design tools