Frank Li Fan

+1 (469) 855-0299 | lifrankfan@gmail.com | lifrankfan.github.io | linkedin.com/in/lifrankfan | github.com/lifrankfan

Education

University of Illinois Urbana-Champaign Urbana, IL
M.S. in Electrical and Computer Engineering *Expected May 2028*
Advisor: Jian Huang; **Research interests:** TPU architecture, performance modeling, world models
Coursework: Computer System Analysis, Hardware and Software for AI, Simulation and Modeling

University of Illinois Urbana-Champaign Urbana, IL
B.S. in Computer Engineering with Honors *May 2026*
Coursework: SoC Design (Graduate Level), Computer Architecture, FPGA Design, RTOS/OS, Parallel Programming

Experience

Systems Platform and Intelligence Lab (SystemX) Urbana, IL
Graduate Researcher (TPU Performance Modeling); Advisor: Jian Huang *Aug 2026 – Present*
- Developing an analytical performance model for Google TPUs to estimate execution latency, memory bandwidth utilization, and compute efficiency, targeting custom Pallas kernels and ordinary JAX and TorchTPU operations.

Skyworks Solutions Nashua, NH
Silicon R&D Intern *May 2026 – Aug 2026*
- Architected and prototyped an RTL digital calibration subsystem for a high-precision IC, reducing worst-case phase error by 83% across modeled operating variation.
- Developed a cycle-level behavioral reference model and automated UVM/UVMF verification environment to check functional correctness and performance in simulation.

Rivian Urbana, IL
SoC Intern *Jan 2025 – May 2026*
- Brought up the System Control Processor on RAP1 silicon for the Rivian Safety Island; ported and optimized Zephyr RTOS, developed device drivers, and verified boot.
- Automated provisioning across OSPI, UFS, firmware images, and Linux, and built the accompanying CI/CD pipeline.

FAST Lab, Coordinated Science Lab Urbana, IL
Research Assistant (Vector Database Acceleration); Advisor: Nam Sung Kim *May 2025 – May 2026*
- Designed a pipelined ZFP floating-point decompressor in SystemVerilog on an Intel Agilex 7 FPGA for vector search; achieved an 8.2× decompression cycle-count speedup over a CPU baseline.
- Integrated the decoder as a CXL Type 2 device using CXL.cache for coherent host-memory reads, CXL.mem for device-memory output, and CXL.io for control.

FAST Lab, Coordinated Science Lab Urbana, IL
Research Assistant (DPU Compression) *Feb 2024 – Aug 2024*
- Optimized a DOCA compression pipeline on NVIDIA BlueField-3 DPUs, reducing execution time by 50% in tested DNN workloads.

IMPACT Research Group Urbana, IL
Research Assistant (GPU Kernels); Advisor: Wen-mei Hwu *Sep 2022 – Sep 2023*
- Developed CUDA kernels for sparse matrix multiplication (SpGEMM) in graph neural network workloads, achieving up to 200× speedup over naive CPU implementations.

Selected Projects

Systolic Accelerator for Genomic Profile-HMM Search May 2026
- Designed a profile-stationary systolic array for the HMMER3 Plan7 Viterbi filter on Xilinx Artix-7, with multi-pass profile folding and cross-pass state buffering for profiles longer than the array.
- Projected 4.2× end-to-end pipeline speedup when composed with the HAVAC SSV accelerator.

SOPERR: Out-of-Order RISC-V Processor Apr 2025
- Designed a two-way superscalar processor with Tomasulo scheduling, register renaming, a reorder buffer, split load/store queues, and gshare branch prediction; placed top 10 of 40 teams in the 2025 Optiver Computer Organization & Design Competition.

Technical Skills

Languages: Python, C++, C, CUDA, SystemVerilog, Assembly (RISC-V, ARM)
ML Frameworks and Compilers: JAX, Pallas/Mosaic, TorchTPU, Compiler IR Analysis
Architecture and Performance: Tensor Processing Unit (TPU) Architecture, Graphics Processing Unit (GPU) Architecture, Microarchitecture, Performance Modeling, Workload Profiling, Cycle-Level Modeling, Hardware-Software Co-design
Systems and Development: Linux, Git, CI/CD, Zephyr RTOS, Device Drivers, Numerical Validation
Hardware: Register-Transfer Level (RTL) Design, FPGA Design, SoC Design, Design Verification (DV), CXL