

Xinwei Qiang

x2qiang@ucsd.edu | github.com/tie-pilot-qxw

Research interests: GPU systems, compilers, and distributed systems.

Publications

[*arXiv* '26] Beyond Parallel Blindness: Information Floors and Model Gaps in Block Drafting.

Xinwei Qiang, Xiang Fang, Chang Chen, Yue Guan, Yufei Ding

[*SC* '26] FluxZK: Scalable and Efficient Zero-Knowledge Proof Computation via GPU Acceleration.

Xinwei Qiang*, Liukun Yu*, Xiyu Wang, Zhengyi Li, Shixuan Sun, Jingwen Leng, Chen Chen, Jiaping Gui, Zhenzhe Zheng, Jin Dong, Minyi Guo

[*arXiv* '26] GF-DiT: Scheduling Parallelism for Diffusion Transformer Serving.

Xinwei Qiang, Yifan Hu, Shixuan Sun, Jing Yang, Han Zhao, Chen Chen, Yu Feng, Jingwen Leng, Minyi Guo

[*ICML* '26] TritonGym: A Benchmark for Agentic LLM Workflows in Triton GPU Code.

Yue Guan*, Yichen Lin*, Xu Zhao, Jianzhu Yao, **Xinwei Qiang**, Zhongkai Yu, Pramod Viswanath, Yufei Ding, Adnan Aziz

[*OSDI* '26] Syncopate: Efficient Multi-GPU AI Kernels via Automatic Chunk-Centric Compute-Communication Overlap.

Xinwei Qiang*, Yue Guan*, Zhengding Hu, Keren Zhou, Yufei Ding, Adnan Aziz

[*ICLR* '26] DASH: Deterministic Attention Scheduling for High-throughput Reproducible LLM Training.

Xinwei Qiang, Hongmin Chen, Shixuan Sun, Jingwen Leng, Xin Liu, Minyi Guo

[*SOSP* '25] Mercury: Unlocking Multi-GPU Operator Optimization for LLMs via Remote Memory Scheduling.

Yue Guan, **Xinwei Qiang**, Zaifeng Pan, Daniels Johnson, Yuanwei Fang, Keren Zhou, Yuke Wang, Wanlu Li, Yufei Ding, Adnan Aziz

* Equal contribution.

Research Appointments

Research Assistant, EPCC Lab, Shanghai Jiao Tong University

Jul 2024 – Jun 2026

Co-advised by Profs. [Jingwen Leng](#) and [Shixuan Sun](#)

Research Intern, Picasso Lab, University of California San Diego

Jul 2025 – Dec 2025

Advisor: Prof. [Yufei Ding](#)

Education

University of California San Diego — Ph.D. in Computer Science

Starting Fall 2026

Advisor: Prof. [Yufei Ding](#)

Shanghai Jiao Tong University — B.E. in Computer Science

Sept 2022 – Jun 2026

John Hopcroft Class, an honors program in computer science.

Technical Skills

Programming: C/C++, CUDA, Python, Rust, Triton