Zihan Wu

PhD Candidate, Electrical Engineering, City University of Hong Kong
wzh4464@gmail.com | (+852) 9810 6427 | https://scholar.zihanng.shop | zihan-wu-ustc
wzh4464

SUMMARY

PhD candidate in Electrical Engineering with strong C++ and high-performance computing (HPC) background. Experienced in low-latency systems, distributed computing, and algorithmic optimization with quantifiable performance gains (83% reduction, 100x speedup). Eager to apply advanced algorithmic and systems knowledge to quantitative research, trading infrastructure, or high-frequency systems.

EDUCATION

City University of Hong Kong, Ph.D. in Electrical Engineering — **Hong Kong SAR, China**
2020 – 2025 (Expected)

University of Science and Technology of China, B.Sc. in Physics & Applied Mathematics — **Hefei, China**
2015 – 2020

EXPERIENCE

Huawei Technologies, Research Intern — **Hong Kong**
Developing LLM optimization techniques for enterprise AI with focus on performance efficiency — 2025 – Present

City University of Hong Kong, Research Assistant — **Hong Kong**
Leading work on high-performance ML algorithms and low-latency optimization systems — 2024 – Present

PROJECTS

C++, Distributed Systems, Performance Engineering — **X-Shard: Distributed Transaction Engine**
2023 – 2024

- Reduced transaction latency by 37% via cross-shard scheduling and commit path optimization
- Optimized transaction execution path in a sharded blockchain by minimizing branch mispredictions via path classification and memory-aligned data layout
- Implemented $O(n)$ threshold signature scheme for fast consensus under network partition
- Published: IEEE Trans. on Parallel and Distributed Systems, 2024

C++, MPI, Optimization — **High-Performance Co-Clustering System**
2020 – 2024

- Built safe and efficient Rust orchestrator calling C++ HPC core via FFI; ensured memory safety while maintaining nanosecond-level latency
- Integrated Rust front-end with C++ computational backend using FFI for safe and efficient co-clustering orchestration
- Achieved 83% runtime reduction on $800K+$ document datasets via parallelized co-clustering (MPI)
- Built real-time ellipse detection system with sub-millisecond response time
- Published in IEEE SMC 2024 and IEEE TIM

C++, PyTorch, CUDA — **LMEraser: Fast Transformer Optimization**
2023 – Present

- Accelerated model unlearning by 100x using adaptive prompt tuning and memory optimization
- Optimized 86M+ parameter transformer models for deployment without performance loss
- To appear at AISTATS 2025

SKILLS

Programming: C++ (Advanced), Python, Rust, C, MATLAB
HPC & Systems: MPI, OpenMP, CUDA, Distributed Systems, Real-Time Processing
Quantitative: Algorithm Design, Statistical Modeling, Optimization, Computational Geometry
Tools: PyTorch, TensorFlow, OpenCV, Linux

AWARDS

HK PhD Fellowship (Top 5% acceptance) — 2020–2024
National Encouragement Scholarship (Top 2%) — 2017–2018
Patent: Physical Activity Assessment System: *HK30081186* — 2023