HAO WANG

+852 68758509 ◇ Hong Kong ◇ 1996.11.14
hwangdv@connect.ust.hk ◇ whwh1996.github.io

EDUCATION

PhD Candidate in Computer Science, Hong Kong University of Science and Technology 2019.8 - now
– Advised by Prof. Kai Chen in intelligent System and networkING Lab (iSING Lab).
– Research interests include on Network & AI System.

BSc in Computer Science, Peking University 2015.9 - 2019.7
– Awarded outstanding graduate.

EXPERIENCES

Phd Program of HKUST in iSING Lab 2019.9 - now, Hong Kong
– Advised by Prof. Kai Chen in iSING Lab. Focus on AI centric Networking.
– Led the MLT project, designing network transport optimized for distributed machine learning. As the first author, published in NSDI 2024. MLT introduced bounded-loss tolerance, priority queuing & selective dropping and adopted per-packet load balancing. Widely cited by top-tier conferences and journals such as NSDI, EuroSys, and ToN.
– Led the DSA project, implementing preemptive scheduling in switch memory for in-network gradient aggregation in multiple AI training jobs. Published as the first author in ICNP 2023. DSA adopted an end-network co-design approach to support gradient aggregation on resource-constrained programmable switches, with memory management and reliability mechanisms. Widely cited by conferences and journals like ToN and INFOCOM.
– Led the INAaaS project, designing a general-purpose in-network computing system tailored for multi-tenant cloud environments. Currently under review at INFOCOM 2025.
– Led the MFS project, designing an efficient LLM inference serving system for model families. The system enhances resource utilization and reduces latency for serving multiple LLMs. Currently under review at NSDI 2025.
– Contributed to development, network setup, and operations for the Hong Kong government-funded TRS project - Turing AI Computing Cloud (HKD 30M), managing two clusters with 160+ GPUs. Co-authored a paper based on this project accepted at ASPLOS 2025 (top-tier conference in computer architecture).

UG Visiting Internship Program of HKUST in iSING Lab 2019.3 - 2019.6, Hong Kong
– Work in iSING Lab, supervised by Prof. Kai Chen. Learn on applying AI to data center network.

Research Intern in Naturali (an AI Startup Corporation) 2018.1 - 2018.6, Beijing
– Reproduce and test latest NLP research work.
– Write TF operators and benchmark for the company's toolkit.

Student Research Assistant in the Network and Dig Data Lab of PKU 2017.4 - 2019.3, Beijing
– Research on data streams, big data algorithms, network measurement under the advisement of Prof. Tong Yang.
– Contributed to three research projects focused on optimizing algorithms for measuring Frequencies, Cardinality, and Multi-set Membership Queries in Data Streams.
– Led a project called "Relative Chinese Words Cloud", where we apply HeavyGuardian to generate words cloud from a billion-scale web dataset provided by Skynet of Peking University.

PUBLICATIONS

Towards Domain-Specific Network Transport for Distributed DNN Training NSDI 2024
Hao Wang, Han Tian, Jingrong Chen, Xinchen Wan, Jiacheng Xia, Gaoxiong Zeng, Wei Bai, Junchen Jiang, Yong Wang, and Kai Chen

Accelerating Neural Recommendation Training with Embedding Scheduling *(Sixth Author)* NSDI 2024
Chaoliang Zeng, Xudong Liao, Xiaodian Cheng, Han Tian, Xinchen Wan, Hao Wang, and Kai Chen

Preemptive Switch Memory Usage to Accelerate Training Jobs with Shared INA — ICNP 2023
Hao Wang, Yuxuan Qin, ChonLam Lao, Yanfang Le, Wenfei Wu, and Kai Chen

AutoByte: Automatic Configuration for Optimal DNN Scheduling — INFOCOM 2022
Yiqing Ma, Hao Wang, Yiming Zhang, and Kai Chen

QCluster: Clustering Packets for Flow Scheduling — WWW 2022
Tong Yang, Jizhou Li, Yikai Zhao, Kaicheng Yang, Hao Wang, Jie Jiang, Yinda Zhang, and Nicholas Zhang

Tacc: A Full-stack Cloud Computing Infrastructure for Machine Learning Tasks — Arxiv 2021
Kaiqiang Xu, Xinchen Wan, Hao Wang, Zhenghang Ren, Xudong Liao, Decang Sun, Chaoliang Zeng, and Kai Chen

RAT: Resilient Allreduce Tree for Distributed Machine Learning — APNet 2020
Xinchen Wan, Hong Zhang, Hao Wang, Shuihai Hu, Junxue Zhang, and Kai Chen

FID-sketch: an Accurate Sketch to Store Frequencies in Data Streams — WWW 2019
Tong Yang, Haowei Zhang, Hao Wang, Muhammad Shahzad, Xue Liu, Qin Xin, and Xiaoming Li

ID Bloom Filter: Achieving Faster Multi-Set Membership Query in Network Applications — ICC 2018
Peng Liu, Hao Wang*, Siang Gao, Tong Yang, Lei Zou, Lorna Uden, and Xiaoming Li*

Fine-grained Probability Counting: Refined Loglog Algorithm — BigComp 2018
Lun Wang, Zekun Cai, Hao Wang, Jie Jiang, Tong Yang, Bin Cui, and Xiaoming Li

ACTIVITIES

- Web chair of Asia-Pacific Workshop on Networking (APNet) — 2021-2024
- Vice president of MSC (Microsoft Student Club) of Peking University. — 2017.9 - 2018.7

TEACHING & TUTORING

- TA, HKUST COMP 3511 – Operating Systems — Spring 2020, Spring 2021, Fall 2022
- TA, HKUST COMP 5621 – Computer Networks — Fall 2021

HONORS & AWARDS

- Travel Grants, ICNP — 2023
- Postgraduate Fellowship, Hong Kong University of Science and Technology — 2019 - now
- Academic Excellence Recognition, Peking University — 2017
- Peking University Intramural ACM Third Prize — 2016, 2017
- Peking University Guanghua Scholarship — 2016
- Merit Student, Peking University — 2016

SKILLS

Languages	Mandarin, English
Programming	C/C++, Python, Linux, FPGA, Network Simulator, Java, Html, SQL
Mathematics	Data Structure and Algorithm, Machine Learning, Big Data, Statistics
Soft Skills	Self-study, Paper Writing, Project Management, Presentation, Teamwork