

WENQI JIANG

PhD Student

Department of Computer Science
ETH Zurich

STF G222
Stampfenbachstrasse 114
8092 Zurich, Switzerland

+41 076 585 8978
wenqi.jiang@inf.ethz.ch
<https://wenqijiang.github.io/>

RESEARCH INTERESTS

My research is centered around designing computer systems in the era of heterogeneous hardware. To be more specific, I am exploring how reconfigurable hardware (FPGAs) can be applied and deployed in data centers efficiently. Such research ranges from application-specific designs such as recommendation systems and information retrieval systems to infrastructure development such as distributed frameworks and virtualization on the cloud.

EDUCATION

ETH Zurich, Switzerland 2021~2025 (*Expected*)

PhD in Computer Science

Affiliated with the Systems Group

Concentration in Data Processing on Heterogeneous Hardware

Advisor: Prof. Gustavo Alonso

Columbia University, USA 2018~2020

Master in Electrical Engineering

Concentration in Data-Driven Analysis and Computation

Advisor: Prof. Luca Carloni

Overall GPA: 4.0/4.0 (top 5%)

Huazhong University of Science and Technology, China 2014~2018

Bachelor in Automation

Concentration in Pattern Recognition

Overall GPA: 3.7/4.0

PROFESSIONAL APPOINTMENTS

Alibaba Cloud Sep. 2019 ~ Dec. 2019

Database Development Intern

Shenzhen, China

I was responsible to develop the high-availability module for ADB-V, a vector similarity search engine (also known as vector database) developed by Alibaba.

AWARDS AND HONORS

The Master's Award of Excellence (top 5%), Columbia University	2021
Outstanding Graduate, HUST	2018
Scholarship for Excellent Academic Performance, HUST	2015

CONFERENCE PUBLICATIONS

- [1] **Wenqi Jiang**, Shigang Li, Yu Zhu, Johannes de Fine Licht, Zhenhao He, Runbin Shi, Cedric Renggli, Shuai Zhang, Theodoros Rekatsinas, Torsten Hoefer, and Gustavo Alonso, "Co-design Hardware and Algorithm for Vector Search." *The International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2023)*, Denver, USA, Nov 12–17, 2023.
- [2] Yu Zhu, Zhenhao He, **Wenqi Jiang**, Kai Zeng, Jingren Zhou, and Gustavo Alonso, "Distributed Recommendation Inference on FPGA Clusters." *31th International Conference on Field-Programmable Logic and Applications (FPL 2021)*, Virtual, Aug 30 - Sep 03, 2021.
- [3] **Wenqi Jiang***, Zhenhao He*, Shuai Zhang, Kai Zeng, Liang Feng, Jiansong Zhang, Tongxuan Liu, Yong Li, Jingren Zhou, Ce Zhang, and Gustavo Alonso, "FleetRec: Large-Scale Recommendation Inference on Hybrid GPU-FPGA Clusters." *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2021)*, August 14-18, virtual, 2021.
- [4] **Wenqi Jiang**, Zhenhao He, Shuai Zhang, Thomas B. Preußer, Kai Zeng, Liang Feng, Jiansong Zhang, Tongxuan Liu, Yong Li, Jingren Zhou, Ce Zhang, and Gustavo Alonso, "MicroRec: Efficient Recommendation Inference by Hardware and Data Structure Solutions." *Proceedings of the 4th Conference on Machine Learning and Systems (MLSys 2021)*, Virtual, April 5-9, 2021.

OTHER PUBLICATIONS

- [1] **Wenqi Jiang**, Dario Korolija, and Gustavo Alonso, "Data Processing with FPGAs on Modern Architectures." *International Conference on Management of Data (SIGMOD 2023 Tutorial)*.
- [2] Shuai Zhang, Yi Tay, **Wenqi Jiang**, Da-cheng Juan, and Ce Zhang, "Switch Spaces: Learning Product Spaces with Sparse Gating." *Under Submission*.
- [3] Shaoxiong Ji, **Wenqi Jiang**, Anwar Walid, and Xue Li, "Dynamic Sampling and Selective Masking for Communication-Efficient Federated Learning." *arXiv preprint*, 2020.

TALKS

Modern Search Engines on Specialized Hardware

ETH Zurich	Sept. 2022
------------	------------

Efficient Recommendation Inference on Heterogeneous Hardware

AMD	March 2022
-----	------------

ETH Zurich	March 2021
------------	------------

FleetRec: Large-Scale Recommendation Inference on Hybrid GPU-FPGA Clusters

SIGKDD Conference	Aug. 2021
MicroRec: Efficient Recommendation Inference by Hardware and Data Structure Solutions	
MLSys Conference	April 2021
ETH Zurich	June 2020

PROFESSIONAL SERVICE

Reviewer

IEEE Micro	2021
------------	------

LANGUAGES

English: fluent

Mandarin (Chinese): native

SKILLS

Programming Languages: C/C++, Python, OCaml, System Verilog

Platforms & Frameworks: Vivado HLS, CUDA, OpenCL, OpenCV, TensorFlow, PyTorch, Keras, Spark Streaming, Apache Beam, MySQL, PostgreSQL

REFERENCES

Available upon request