

Zhenhao He

H319, Building STF, Stampfenbachstrasse 114, 8092 Zürich, Switzerland

☎ (+41) 786860225 | ✉ zhe@inf.ethz.ch | 🏠 Homepage | 📄 Google Scholar

Research Interest

My research interest lies in advancing data processing systems within data centers, with a primary research focus on the convergence of heterogeneous hardware, distributed computing, and data center networking. My work involves crafting networking abstractions—such as TCP, RDMA, and MPI—specifically tailored for hardware accelerators to prototype future systems with smart-NICs and in-network processors. Additionally, I contribute to the development of system infrastructures that efficiently orchestrate heterogeneous clusters, ensuring optimal performance for large-scale applications, particularly in the realm of machine learning.

Education

PhD Candidate in Computer Science

SYSTEMS GROUP, ETH ZURICH

Zurich Switzerland

Mar. 2019 - Exp. Jun. 2024

- Thesis: Enabling Networking for Distributed Computation on FPGA Clusters
- Advisor: Gustavo Alonso
- Committee Members: Gustavo Alonso, Michaela Blott, Jason Cong, Zsolt István

Master of Science in Information Technology and Electrical Engineering

SYSTEMS GROUP, ETH ZURICH

Zurich Switzerland

Sep. 2016 - Oct. 2018

- Advisor: Gustavo Alonso
- Master Thesis: Virtual Channel for Enabling Distributed FPGA Applications

Bachelor of Science in Electronic Engineering

POLITECNICO DI MILANO

Milan Italy

Sep. 2015 - Jul. 2016

- Double Degree Program with Tongji University

Bachelor of Science in Electronic and Information Engineering

TONGJI UNIVERSITY

Shanghai China

Sep. 2012 - Jul. 2015

- Outstanding Graduates of Tongji University

Publications

[arXiv'23] ACCL+: an FPGA-Based Collective Engine for Distributed Applications

Dec. 2023

ZHENHAO HE, DARIO KOROLIJA, YU ZHU, BENJAMIN RAMHORST, TRISTAN LAAN, LUCIAN PETRICA, MICHAELA BLOTT, GUSTAVO ALONSO

ArXiv Preprint (under submission)

[ASPLOS'22] Enzian: an Open, General, CPU/FPGA Platform for OS Research

Feb. 2022

DAVID COCK, ABISHEK RAMDAS, DANIEL SCHWYN, MICHAEL GIARDINO, ADAM TUROWSKI, ZHENHAO HE, NORA HOSSLE, DARIO KOROLIJA, MELISSA LICCIARDELLO, KRISTINA MARTSENKO, RETO ACHERMANN, GUSTAVO ALONSO, TIMOTHY ROSCOE

ACM International Conference on Architectural Support for Programming Languages and Operating Systems

[H2RC'21] ACCL: FPGA-Accelerated Collectives over 100 Gbps TCP-IP

Nov. 2021

ZHENHAO HE, DANIELE PARRAVICINI, LUCIAN PETRICA, KENNETH O'BRIEN, GUSTAVO ALONSO, MICHAELA BLOTT

International Workshop on Heterogeneous High-performance Reconfigurable Computing

[FPL'21] EasyNet: 100 Gbps Network for HLS

Aug. 2021

ZHENHAO HE, DARIO KOROLIJA, GUSTAVO ALONSO

International Conference on Field-Programmable Logic and Applications

[FPL'21] Distributed Recommendation Inference on FPGA Clusters

Aug. 2021

YU ZHU, ZHENHAO HE, WENQI JIANG, KAI ZENG, JINGREN ZHOU, GUSTAVO ALONSO

International Conference on Field-Programmable Logic and Applications

[KDD'21] FleetRec: Large-Scale Recommendation Inference on Hybrid GPU-FPGA Clusters

Aug. 2021

WENQI JIANG*, **ZHENHAO HE***, SHUAI ZHANG, KAI ZENG, LIANG FENG, JIANSONG ZHANG, TONGXUAN LIU, YONG LI, JINGREN ZHOU, CE ZHANG, GUSTAVO ALONSO (* INDICATES EQUAL CONTRIBUTION)
SIGKDD Conference on Knowledge Discovery and Data Mining

[MLSys'21] MicroRec: Accelerating Deep Recommendation Systems to Microseconds by Hardware and Data Structure Solutions

Apr. 2021

WENQI JIANG, **ZHENHAO HE**, SHUAI ZHANG, THOMAS B PREUßER, KAI ZENG, LIANG FENG, JIANSONG ZHANG, TONGXUAN LIU, YONG LI, JINGREN ZHOU, CE ZHANG, GUSTAVO ALONSO
Conference on Machine Learning and Systems

[TC'21] Shuhai: A Tool for Benchmarking High Bandwidth Memory on FPGAs

Apr. 2021

HONGJING HUANG, ZEKE WANG, JIE ZHANG, **ZHENHAO HE**, CHAO WU, JUN XIAO, GUSTAVO ALONSO
IEEE Transactions on Computers

[SoCC'20] Specializing the Network for Scatter-Gather Workloads

Oct. 2020

CATALINA ALVAREZ, **ZHENHAO HE**, GUSTAVO ALONSO, ANKIT SINGLA
ACM Symposium on Cloud Computing

[FPGA'20] BiS-KM: Enabling Any-Precision K-Means on FPGAs

Feb. 2020

ZHENHAO HE, ZEKE WANG, GUSTAVO ALONSO
ACM/SIGDA International Symposium on Field-Programmable Gate Arrays

[FPL'18] A Flexible K-Means Operator for Hybrid Databases

Aug. 2018

ZHENHAO HE, DAVID SIDLER, ZSOLT ISTVÁN, GUSTAVO ALONSO
International Conference on Field-Programmable Logic and Applications

Research Experience

PhD Candidate

Zurich Switzerland

SYSTEMS GROUP, ETH ZURICH

Mar. 2019 - Present

- Designed FPGA-based machine learning operators, such as a flexible K-Means hardware accelerator, tailored for hybrid databases and accommodating various runtime parameters, including computation precision.
- Developed EasyNet, a high-level synthesis library enabling a 100 Gbps TCP/IP network stack on FPGAs, fostering seamless integration of FPGAs into data centers and empowering developers to construct in-network accelerators.
- Developed an FPGA-based smart-NIC featuring hardware TCP/IP offloading to enhance scatter-gather efficiency in cloud computing, enabling parallel request scattering and incorporating a smart scheduler to prevent in-cast issues during response gathering.
- Developed ACCL, an open-source library facilitating FPGA-accelerated collective communication in distributed applications, offering flexibility with support for TCP/RDMA. ACCL serves as both a collective offload engine for distributing CPU-based applications and a crucial, efficient component for designing fully FPGA-based applications.
- Conducted distributed deep learning recommendation inference across FPGA clusters and seamlessly extended to encompass heterogeneous FPGA, CPU, and GPU clusters, utilizing either EasyNet or ACCL.
- Contributed to the Enzian project, a research computer developed by ETH Systems Group, where the CPU is cache-coherently connected to an FPGA. Specifically, I played a key role in enhancing the hardware network support for the Enzian FPGA.

Research Intern

Dublin Ireland (Remote)

AMD XILINX RESEARCH LABS

Mar. 2021 - Sep. 2021

- Designed an open-source FPGA accelerated collective offload engine with 100 Gbps hardware TCP backend.

Research Assistant

Zurich Switzerland

INSTITUTE OF ELECTROMAGNETIC FIELDS, ETH ZURICH

Sep. 2016 - Apr. 2017

- Designed and simulated a novel structure of photonic-plasmonic beam splitter in subwavelength scale realizing low loss transmission.

Invited Talks and Tutorials

[Tutorial] Network and Memory Abstractions on FPGAs for Distributed Applications

Sep. 2023

SPEAKERS AND ORGANIZERS: **ZHENHAO HE**, DARIO KOROLIJA, GUSTAVO ALONSO

Tutorial co-located with International Conference on Field-Programmable Logic and Applications

- [Talk] Enabling Networking for Distributed Data Processing on Heterogeneous Clusters** Jul. 2023
Huawei Future Database Seminar co-located with the ACM Europe Summer School on Data Science
- [Tutorial] Enabling Networking for Distributed Applications on FPGA Clusters** Feb. 2023
SPEAKERS AND ORGANIZERS: **ZHENHAO HE**, DARIO KOROLJICA, LUCIAN PETRICA, MICHAELA BLOTT, GUSTAVO ALONSO
Tutorial co-located with ACM/SIGDA International Symposium on Field-Programmable Gate Arrays
- [Talk] Introduction to XACC-ETHZ cluster** Aug. 2021
Workshop on DevOps for Cloud FPGA Platforms at the International Conference on Field-Programmable Logic and Applications
- [Talk] VNx and EasyNet** Jun. 2021
AMD Xilinx Tech Talk Series
- [Talk] EasyNet: 100Gbps Network for HLS** Jan. 2021
AMD Xilinx HACC School

Scholarship & Awards

- 2021 **AMD Xilinx XACC Outstanding Researcher Awards** AMD Xilinx
- 2016 **Outstanding Graduates** Tongji University
- 2015 **National Scholarship** Chinese Educational Bureau
- 2015 **First Prize of Learning Scholarship** Tongji University
- 2015 **Chinese Government Scholarship for Studying Abroad** Chinese Educational Bureau
- 2014 **Siemens Scholarship** Tongji University

Community Services

- 2023 **Reviewer** IEEE Transactions on Parallel and Distributed Systems
- 2023 **Reviewer** IEEE Transactions on Neural Networks and Learning Systems
- 2023 **Shadow Program Committee** ACM European Conference on Computer Systems

Teaching

- 2023 Fall **Master's Course: Big Data** Head Teaching Assistant ETH Zurich
- 2023 Spring **Master's Course: Big Data for Engineers** Head Teaching Assistant ETH Zurich
- 2022 Fall **Master's Course: Big Data** Head Teaching Assistant ETH Zurich
- 2022 Spring **Master's Course: Big Data for Engineers** Head Teaching Assistant ETH Zurich
- 2021 Fall **Master's Course: Information System for Engineers** Head Teaching Assistant ETH Zurich
- 2021 Spring **Master's Course: Information System** Head Teaching Assistant ETH Zurich
- 2020 Fall **Master's Course: Information System for Engineers** Teaching Assistant ETH Zurich
- 2020 Spring **Seminar: Hardware Acceleration for Machine Learning** Teaching Assistant ETH Zurich

Supervision Experience

- 2023 Fall **Master Thesis** *Multi-word Compare-and-Swap on FPGAs*, Floris Westermann ETH Zurich
- 2023 Fall **Semester Project** *Distributed Machine Learning with ACCL*, De Gendt Antoine & Streich Georg ETH Zurich
- 2023 Fall **Bachelor Thesis** *Optimize a Network-Attached Deduplication Accelerator*, Niksch Philipp ETH Zurich
- 2023 Spring **Semester Project** *Data Deduplication with Near-Storage Acceleration*, Jiayong Li ETH Zurich
- 2023 Spring **Semester Project** *Benchmarking and Optimizing EasyNet over HACC*, Mihai Zorca (Now at Optiver) ETH Zurich
- 2023 Spring **Semester Project** *Benchmark Cryptography Offload on FPGAs*, Arnet Colin ETH Zurich
- 2023 Spring **Master Thesis** *Optimized BFT Consensus For Smart-NIC Offload*, Petr Gebauer (Now at AIM Software) ETH Zurich
- 2021 Fall **Semester Project** *Aggregation Group-by on FPGAs*, Yu Zhu (Now as PhD in Systems Group, ETH) ETH Zurich
- 2020 Fall **Master Thesis** *Distributed Group-by Aggregation with FPGA Cluster*, Zhifei Yang (Now at Tiktok UK) ETH Zurich
- 2020 Fall **Semester Project** *Recommendation Inference Acceleration with FPGA Cluster*, Yu Zhu ETH Zurich