Haiyu Mao

Homepage: https://hybol1993.github.io/ Email: maohaiyu1993@gmail.com

EDUCATIONAL BACKGROUND

ETH Zurich Zurich, Switzerland Postdoc researcher in Computer Science; 07.2020 – Present

Advisor: Prof. Onur Mutlu

Tsinghua University Ph.D. in Computer Science and Technology;

8eijing, China
08.2015 – 07.2020

Advisor: Prof. Jiwu Shu; GPA: 3.7/4.0 (Rank:24/95)

Northeastern University

Bachelor in Software Engineering;

Shenyang, China
08. 2011 – 07.2015

GPA: 90/100 (Rank: 1/201)

SELECTED PRACTICAL EXPERIENCE

Memory System Design for AI/ML Accelerators

ETH Zurich, Research Scholar, 01.2020 - Present

- + Contributing to the Semiconductor Research Corporation (SRC) program led by Prof. Onur Mutlu by conducting research on non-volatile memory (NVM) system design for machine learning accelerators.
- + Designed a durable and energy-efficient NVM system for in-memory neural network training by utilizing low-quality NVM cells to store insignificant weight values of neural networks.

Acceleration of Genome Analysis on Raw Signal Data

ETH Zurich, Postdoc, 11.2021 - Present

Phone: +41 76-830-6068

- + Designing a CAM-based accelerator along with the algorithm for analyzing a large amount of raw digit-signal data.
- + Analyzed the bottleneck of GPU when it is executing genome analysis on raw-signal data.

In-Memory Genome Analysis

ETH Zurich, Postdoc, 09.2020 - 11.2021

- + Pioneered an in-memory acceleration of genome analysis through a tight integration of basecalling and read mapping.
- + Designed a tight pipeline to minimize data movement by eliminating the need for storing intermediate results.
- + Proposed an early-rejection mechanism to stop inefficient computation based on the timely feedback from quality control and read mapping stages.

Durable NVM-based PIM for Training Neural Networks

Tsinghua University, PhD, 10.2018 - 07.2020

- + Studied the write behavior of updating the weight matrix in the NVM-based PIM array during neural network training.
- + Proposed a scheme for long-lived PIM by (a) leveraging the characteristics of NVM cells and (b) combining the inherent fault-tolerance characteristic of the neural networks and their particular weight updating behaviors.

Analysis of The Data Collected by FAST

National Astronomical Observatories, Intern, 06.2019 - 08.2019

- + Pre-processed and analyzed the astronomical data collected by the Five-hundred-meter Aperture Spherical Telescope (FAST) located in a natural basin in Pingtang County, Guizhou, China.
- + Realized a web crawler to collect the published astronomical data and compared with the data collected by FAST.

Efficient NVM-based PIM for Machine Learning Applications Tsinghua Univ

Tsinghua University, PhD, 04.2017 - 10.2018

- + Put forward a data reshaping mechanism that removes zero-related computations, along with a structured data mapping mechanism to save both storage capacity and communication bandwidth in PIM.
- + Designed a 3D reconfigurable interconnect fabric in PIM to radically shorten the routing paths.

Secure and Efficient Non-Volatile Memories

Tsinghua University, PhD, 12.2015 - 04.2017

- + Devised a scheme that reveals the information of the address mapping through read latency difference between row-offer hit and miss, whereby an effective wear-out attack on the physical data location can be conducted.
- + Proposed a countermeasure that prolongs the lifetime of NVM compared with the state-of-the-art wear-leveling scheme, while only introducing trivial hardware overhead.

Secure and Efficient Non-Volatile Memories

Peking University, Intern, 01.2015 - 12.2015

+ Realized a scratchpad-memory-based data placement scheme to reduce the movement of read/write ports in Racetrack Memory by leveraging the genetic algorithm.

CV-Haiyu Mao 1 Last update: Oct.09, 2022

PUBLICATIONS

- 1. **Mao H**, Alser M, Sadrosadati M, Firtina C, Baranwal A, Cali DC, Manglik A, Alserr NA, and Mutlu O, "GenPIP: In-Memory Acceleration of Genome Analysis by Tight Integration of Basecalling and Read Mapping", in *International Symposium on Microarchitecture*, (**MICRO**), October 2022.
- 2. Alser M, Lindegger J, Firtina C, Almadhoun N, **Mao H**, Singh G, Gomez-Luna J, and Mutlu O, "<u>From Molecules to Genomic Variations: Accelerating Genome Analysis via Intelligent Algorithms and Architectures</u>", in *Computational and Structural Biotechnology Journal*, (**CSBJ**), August 2022.
- 3. Ghiasi NM, Park J, Mustafa H, Kim J, Gollwitzer A, Olgun A, **Mao H**, Firtina C, Cali DS, Alserr NA, Ausavarungnirun R, Vijaykumar N, Alser M, and Mutlu O, "GenStore: An In-Storage Processing System for Genome Sequence Analysis", in *ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (**ASPLOS**), March 2022.
- 4. **Mao H**, Shu J, Song M, and Li T, "LrGAN: A Compact PIM-based GAN Architecture with Low Energy Consumption", in *IEEE Transactions on Computers* (**TC**), 2021.
- 5. **Mao H**, Shu J, Li F, and Liu Z, "The Development of Processing In Memory", in *SCIENTIA SINICA Informationis*, (**SSI**), (In Chinese), 2020.
- 6. Yang F, Chen Y, **Mao H**, Lu Y, and Shu J, "Libra: An Efficient and Fast Recoverable System for Secure Non-Volatile Memory", in *ACM Transactions on Storage* (**TOS**), 2020.
- 7. **Mao H**, and Shu J, "3D Memristor Array Based Neural Network Processing in Memory Architecture", in Journal of Computer Research and Development, (In Chinese), 2019.
- 8. Yang F, Lu Y, Chen Y, **Mao H**, and Shu J, "No Compromises: Secure NVM with Crash Consistency, Write-Efficiency and High-Performance", in *Design Automation Conference* (**DAC**), June 2019.
- 9. **Mao H**, Song M, Li T, Dai Y, and Shu J, "LerGAN: A Zero-Free, Low Data Movement and PIM-Based GAN Architecture", *in International Symposium on Microarchitecture* (**MICRO**), October 2018.
- 10. **Mao H**, Zhang X, Sun G, and Shu J, "Protect Non-Volatile Memory from Wear-Out Attack Based on Timing Difference of Row Buffer Hit/Miss", *in Conference on Design, Automation & Test in Europe* (**DATE**), March 2017.
- 11. **Mao H**, Zhang C, Sun G, and Shu J, "Exploring Data Placement in Racetrack Memory Based Scratchpad Memory", in Non-Volatile Memory System and Applications Symposium (NVMSA), August 2015.

SELECTED AWARDS

Outstanding Ph.D. Graduate in Beijing (**Top 5 in the department**)

National Scholarship for Ph.D. (**2.5**%)

Outstanding Undergraduate in Shenyang (**0.26**%)

Outstanding Pioneer Student (**0.5**%, **three times**)

National Scholarship (**1**%, **three times**)

Ministry of National Education of China, 2019

Northeastern University, 2012/2013/2014

Ministry of National Education of China, 2012/2013/2014

LANGUAGE SKILLS

Chinese: Native speaker

English: Professional working proficiency

TECHNICAL SKILLS

• Programming Languages:

Expert: Python, C; Proficient: C++, Java; Prior experience: Javascript, Go, MPI, OpenMP, CUDA, Matlab

- Frameworks: Caffe, TensorFlow, PyTorch, Hadoop, Spark
- Simulators: Gem5, DRAMSim, NVSim, NVmain, CACTI

SELECTED SERVICES AND TEACHING ASSISTANCE

ETH Future Computing Laboratory

Group Associate, 2021/2022

- + Proposed "PIM-based Acceleration of Genome Analysis" to EFCL Funding Committee as a principal investigator.
- + Organized seminars (e.g. invited Fabrice Devaux, the CTO of UPMEM, to introduce UPMEM DPU Architecture).

Computer Architecture Courses

Teaching Assistant, 2020/2021

+ Developed homework/exams; Prepared and managed teaching materials; Answered questions from students.

SAFARI Project & Seminars courses

Mentor, 2021/2022

+ Proposed and supervised practical PIM projects; Helped students to code the simulation of PIM architectures.

$\boldsymbol{Conference\ Paper\ Reviews\ (DATE)}$

PC member, 2022

+ Served the Technical Program Committee (TPC) of the Design, Automation and Test in Europe Conference (DATE).

+ Helped prepare and distribute the call for papers, review papers, participate in the TPC meeting to discuss the acceptance of papers.

Journal Paper Reviews (ACM Computing Surveys, IEEE Micro, TC)

Reviewer, 2021/2022