

# Lei Zhao

PhD of Computer Science

Department of Computer Science, University of Pittsburgh  
210 S. Bouquet Street, Pittsburgh, PA 15260, USA

☎ (+1) 412-708-4514 | ✉ leizhao@cs.pitt.edu | 🏠 <https://leizhaocs.github.io>

## Research

**Hardware level:** Computer Architecture, In-Memory Computing, Emerging Memory Technology, Accelerator Design

**Software level:** Deep Learning, Secure Machine Learning, Machine Learning Framework Development

## Education

### University of Pittsburgh

PH.D. OF COMPUTER SCIENCE

- Co-advised by Youtao Zhang and Jun Yang

*Pittsburgh, PA, U.S.*

*August 2014 - April 2022*

### Northwestern Polytechnical University

MASTER OF COMPUTER SCIENCE

*Xi'an, Shaanxi, China*

*August 2011 - April 2014*

### Northwestern Polytechnical University

BACHELOR OF SOFTWARE ENGINEERING

*Xi'an, Shaanxi, China*

*August 2007 - July 2011*

## Work Experience

### Meta, Inc., USA

REALITY LABS

*Postdoc Research Scientist*

*January 2022 - Present*

- In-memory computing accelerator design for codec avatars in AR/VR. (One paper submitting, one patent filing)
- AR/VR codec avatars deep learning model optimization for in-house systolic array accelerator.
- AR/VR codec avatars deep learning model deployment on SOC in existing head mounted devices.

### National Key Lab of Process Optimization and Intelligent Decision, Ministry of Education, China

SMART MEDICAL SYSTEM TEAM

*Research Intern*

*July 2018 - August 2018*

- Maintaining our previously developed privacy-preserving neural network training framework.
- Convert and optimize our DNN model into a privacy-preserving friendly fashion.

### National Key Lab of Process Optimization and Intelligent Decision, Ministry of Education, China

SMART MEDICAL SYSTEM TEAM

*Research Intern*

*August 2017 - September 2017*

- Developed a privacy-preserving neural network training framework in C++ with CUDA acceleration.

## Projects

### Machine Learning Accelerators

DESIGNER AND CONDUCTOR

*University of Pittsburgh*

*April 2017 - Present*

- Design ASIC accelerators based on existing or emerging memory technologies to improve performance and energy efficiency of machine learning computations with a focus on model security and user privacy.
- Three conference papers in [ICCAD'17] [ICS'19] [DAC'20] and one U.S. patent.

### Privacy-Preserving Medical Data Analysis

RESEARCH ASSISTANT

*Hefei University of Technology*

*August 2017 - August 2018*

- Evaluate Neural Networks on medical data with homomorphic encryption to protect patient's private information.
- One journal paper in ACM Transactions on Internet Technology (TOIT).

## Privacy-Preserving Neural Network Framework

DESIGNER AND DEVELOPER

University of Pittsburgh

August 2017 - September 2017

- A neural network framework that targets at privacy-preserving and architecture research.
- Written in C++ and CUDA. Support fixed-point computing, stochastic-computing and homomorphic encryption.
- Open sourced on github: <https://github.com/leizhaocs/ArchNet>.

## Emerging Memory Design

DESIGNER AND CONDUCTOR

University of Pittsburgh

September 2014 - April 2016

- Circuit level optimization of non-volatile memory for machine learning acceleration.
- Developed a cycle-accurate out-of-order architecture simulator based on MIPS ISA. Open sourced on github: <https://github.com/leizhaocs/Monichi>.
- Four conference papers in [ICCD'15] [ISQED'17] [NVMSA'17] [ICCAD'17] and one journal paper in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD).

## Publications

---

### CONFERENCE PROCEEDINGS

- **Lei Zhao**, Yuecheng Li, Jae-sun Seo, H. Ekin Sumbul, Edith Beigne, and Dawei Wang. (2023). ReARVR: A ReRAM-Based DNN Accelerator for Mobile Devices. Design Automation Conference (Poster Session). (DAC'23)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2022). A DNN Protection Solution for PIM Accelerators With Model Compression. IEEE Computer Society Annual Symposium on VLSI. (ISVLSI'22)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2022). SRA: A Secure ReRAM-Based DNN Accelerator. Design Automation Conference. (DAC'22)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2021). Flipping Bits to Share Crossbars in ReRAM-Based DNN Accelerator. International Conference on Computer Design. (ICCD'21)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2020). SCA: A Secure CNN Accelerator for both Training and Inference. Design Automation Conference. (DAC'20)
- **Lei Zhao**, Quan Deng, Youtao Zhang, and Jun Yang. (2019). RFacc: A 3D ReRAM Associative Array based Random Forest Accelerator. International Conference on Supercomputing. (ICS'19)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2017). AEP: An Error-bearing Neural Network Accelerator for Energy Efficiency and Model Protection. International Conference On Computer Aided Design. (ICCAD'17)
- Wen Wen, **Lei Zhao**, Youtao Zhang, and Jun Yang. (2017). Speeding Up Crossbar Resistive Memory by Exploiting In-memory Data Patterns. International Conference On Computer Aided Design. (ICCAD'17)
- **Lei Zhao**, Youtao Zhang, and Jun Yang. (2017). Mitigating Shift-Based Covert-Channel Attacks in Racetrack Last Level Caches. Non-Volatile Memory Systems and Applications Symposium. (NVMSA'17)
- **Lei Zhao**, Lei Jiang, Youtao Zhang, Nong Xiao, and Jun Yang. (2017). Constructing Fast and Energy Efficient 1TnR based ReRAM Crossbar Memory. International Symposium on Quality Electronic Design. (ISQED'17)
- Xianwei Zhang, **Lei Zhao**, Youtao Zhang, and Jun Yang. (2015). Exploit Common Source-Line to Construct Energy Efficient Domain Wall Memory based Caches. International Conference on Computer Design. (ICCD'15)

### JOURNAL ARTICLES

- Zijie Yue, Shuai Ding, **Lei Zhao**, Youtao Zhang, Zehong Cao, M. Tanveer, Alireza Jolfaei, and Xi Zheng. (2020). Privacy-preserving Time Series Medical Images Analysis Using a Hybrid Deep Learning Framework. ACM Transactions on Internet Technology.
- Wen Wen, **Lei Zhao**, Youtao Zhang, and Jun Yang. (2019). Exploiting In-memory Data Patterns for Performance Improvement on Crossbar Resistive Memory. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.

## Patents

---

- Youtao Zhang, **Lei Zhao**, and Jun Yang. (2019). **System and method of deploying an artificial neural network on a target device**. U.S. Patent, 11,531,877 B2.
- Shuai Ding, **Lei Zhao**, Shanlin Yang, Hao Wang and Zijie Yue. (2020). **The architecture, image processing method and process of an artificial intelligence chip for medical endoscope**. Chinese. Patent, CN108055454B.

## Skills

---

<b>Programming</b>	C/C++, Python, Java, CUDA
<b>Algorithms</b>	Deep Learning, Reinforcement Learning, Homomorphic Encryption
<b>Tools</b>	Gem5, DRAMSim2, numpy, PyTorch, Homomorphic Encryption Libraries(SEAL, HEAAN), Linux/Unix

## Honors & Awards

---

2021	<b>Best Paper Candidate</b> , 39th International Conference on Computer Design (ICCD)	<i>Virtual Conference</i>
2017	<b>Best Paper Candidate</b> , 18th International Symposium on Quality Electronic Design (ISQED)	<i>Santa Clara, CA, USA</i>
2014	<b>Outstanding Master Degree Thesis</b> , Graduation Commencement of Northwestern Polytechnical University	<i>Xi'an, Shaanxi, China</i>