Dr. Zhenhua Zhu

Postdoc Researcher
Room 4-205, Rohm Building, E.E. Dept., Tsinghua University, Beijing, 100084, P. R. China
ICDC, CYT Building, ECE Dept., HKUST, Hong Kong SAR, P. R. China
(+86)178-8883-3495
zhuzhenhua@mail.tsinghua.edu.cn, eezhuzh@ust.hk, zhu-z14@outlook.com
https://scholar.google.com/citations?user=Y3IiQmQAAAAJ

EDUCATION

Department of Electronic Engineering, Tsinghua University Aug. 2018 – Jan. 2024
Ph.D. (**with honors**) in Department of Electronic Engineering
Dissertation: *Research on Design Methodology of Memory-Centric Computer Architectures* (**Excellent Doctoral Dissertation Award of Tsinghua University**)
Advisor: Prof. Yu Wang

Department of Electronic Engineering, Tsinghua University Aug. 2014 – Jul. 2018
B.S. in Electronic Engineering, Grade Ranking: 42/245, Outstanding Thesis
Thesis: *Mixed size crossbar-based RRAM CNN accelerator with overlapped mapping method* (**Excellent BS Thesis Award of Tsinghua University**)
Advisor: Prof. Yu Wang

PROFESSIONAL EXPERIENCE

Hong Kong University of Science and Technology Jul. 2024 – Now
Visiting Scholar, with Prof. Yuan Xie

Department of Electronic Engineering, Tsinghua University Feb. 2024 – Now
Postdoc Researcher, with Prof. Yu Wang

Hong Kong University of Science and Technology & AI Chip Center for Emerging Smart Systems Apr. 2023 – Jun. 2023
Visiting Researcher, with Prof. Tim Cheng and Prof. Fengbin Tu

University of California, Santa Barbara Jun. 2017 – Aug. 2017
Visiting Researcher, with Prof. Yuan Xie

RESEARCH & SERVICE SUMMARY

Dr. Zhenhua Zhu has been devoted to the research of computer architecture, Processing-In-Memory, and Near Memory Computing. He has published/accepted 40 academic papers in IEEE TCAD, DAC, ISCA, MICRO, ICCAD, ASPLOS, and DATE, with Google Scholar citations more than 1000. He has received Best Paper Nomination in DATE 2023 and the Outstanding Ph.D. Dissertation Award of Tsinghua University. He serves as a reviewer for DAC, ISVLSI, IEEE TCAD, IEEE TVLSI, ACM TODAES, ACM TECS, and AICAS. He also serves as the TPC secretary of ASP-DAC 2025.

PUBLICATION

✧ **First author:**

[HPCA-25] Tongxin Xie*, **Zhenhua Zhu***, Bing Li, Yukai He, Cong Li, Guangyu Sun, Huazhong Yang, Yuan Xie, Yu Wang, "UniNDP: A Unified Compilation and Simulation Tool for Near DRAM Processing Architecture", to appear in International Symposium on High-Performance Computer Architecture (HPCA), IEEE, 2025. **(Contributed equally**, my contribution: project leader & paper writing**)**

[ICCAD-24] Lidong Guo*, **Zhenhua Zhu***, Tengxuan Liu, Xuefei Ning, Shiyao Li, Guohao Dai, Huazhong Yang, Wangyang Fu and Yu Wang, "Towards Floating Point-Based Attention-Free LLM: Hybrid PIM with Non-Uniform Data Format and Reduced Multiplications", in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, IEEE/ACM, 2024. **(Contributed equally**, my contribution: hardware architecture design & paper writing**)**

[IEEE TCAD] **Zhenhua Zhu**, Hanbo Sun, Tongxin Xie, Yu Zhu, Guohao Dai, Lixue Xia, Dimin Niu, Xiaoming Chen, X. Sharon Hu, Yu Cao, Yuan Xie, Huazhong Yang, Yu Wang, "MNSIM 2.0: A Behavior-Level Modeling Tool forProcessing-In-Memory Architectures", in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, IEEE, 2023.

[DAC-23] **Zhenhua Zhu**, Jun Liu, Guohao Dai, Shulin Zeng, Bing Li, Huazhong Yang and Yu Wang, "Processing-In-Hierarchical-Memory Architecture for Billion-Scale Approximate Nearest Neighbor Search", in *Design Automation Conference (DAC)*, IEEE/ACM, 2023.

[DAC-23] Shiyao Li*, **Zhenhua Zhu***, Yu Zhu, Qingpeng Zhu, Jiangwei Zhang, Wenxiu Sun, Guohao Dai, Fei Qiao, Huazhong Yang and Yu Wang, "Memory-Efficient and Real-Time SPAD-based dToF Imaging with Spatial and Statistical Correlation", in *Design Automation Conference (DAC)*, IEEE/ACM, 2023. **(Contributed equally**, my contribution: dToF hardware design & paper writing**)**

[MICRO-23] Shulin Zeng*, **Zhenhua Zhu***, Jun Liu, Haoyu Zhang, Guohao Dai, Zixuan Zhou, Shuangchen Li, Xuefei Ning, Yuan Xie, Huazhong Yang, Yu Wang, "DF-GAS: a Distributed FPGA-as-a-Service Architecture towards Billion-Scale Graph-based Approximate Nearest Neighbor Search", in *56th IEEE/ACM International Symposium on Microarchitecture*, 2023. **(Contributed equally**, my contribution: architecture design & paper writing**)**

[ISCA-22] Guohao Dai*, **Zhenhua Zhu***, Tianyu Fu, Chiyue Wei, Banyan Wang, Xiangyu Li, Yuan Xie, Huazhong Yang, Yu Wang, "DIMMining: Pruning-Efficient and Parallel Graph Mining on DIMM-based Near-Memory-Computing", in *The International Symposium on Computer Architecture (ISCA), 2022*, ACM/IEEE, 2022. **(Contributed equally**, my contribution: architecture and hardware design & paper writing**)**

[DATE-22] Yu Zhu*, **Zhenhua Zhu***, Guohao Dai, Kai Zhong, Huazhong Yang, Yu Wang, "Exploiting Parallelism with Vertex-Clustering in Processing-In-Memory-based GCN Accelerators", in *Design, Automation and Test in Europe Conference,* 2022. **(Contributed equally**, my contribution: architecture design & paper writing**)**

[ASPDAC-21] Hanbo Sun*, **Zhenhua Zhu***, Yi Cai*, Shulin Zeng, Kaizhong Qiu, Yu Wang, Huazhong Yang, "Reliability-Aware Training and Performance Modeling for Processing-In-Memory Systems", in *the 26th Asia and South Pacific Design Automation Conference(ASP-DAC 2021)*, IEEE, 2021. **(Contributed equally**, my contribution: performance modeling tool design**)**

[ASPDAC-20] Hanbo Sun*, **Zhenhua Zhu***, Yi Cai, Xiaoming Chen, Yu Wang, Huazhong Yang, "An Energy-Efficient Quantized and Regularized Training Framework for Processing-In-Memory Accelerators", in *the 25th Asia and South Pacific Design Automation Conference(ASP-DAC 2020)*, 2020. **(Contributed equally**, my contribution: idea proposal, performance evaluation & paper writing**)**

[GLSVLSI-20] **Zhenhua Zhu**, Hanbo Sun, Kaizhong Qiu, Lixue Xia, Gokul Krishnan, Guohao Dai, Dimin Niu, Xiaoming Chen, X. Sharon Hu, Yu Cao, Yuan Xie, Yu Wang, Huazhong Yang, "MNSIM 2.0: A Behavior-Level Modeling Tool for Memristor-based Neuromorphic Computing Systems", in *Great Lakes Symposium on VLSI (GLSVLSI)*, 83-88, 2020.

[DAC-19] **Zhenhua Zhu**, Hanbo Sun, Yujun Lin, Guohao Dai, Lixue Xia, Song Han, Yu Wang, Huazhong Yang, "A Configurable Multi-Precision CNN Computing Framework Based on Single Bit RRAM", in *Design Automation Conference*, IEEE, 2019.

[ICCAD-19] **Zhenhua Zhu**, Mingyuan Ma, Jialong Liu, Liying Xu, Xiaoming Chen, Yuchao Yang, Yu Wang and Huazhong Yang, "A General Logic Synthesis Framework for Memristor-based Logic Design", in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, IEEE/ACM, 2019.

[ICCAD-18] **Zhenhua Zhu**, Jilan Lin, Ming Cheng, Lixue Xia, Hanbo Sun, Xiaoming Chen, Yu Wang and Huazhong Yang, "Mixed Size Crossbar based RRAM CNN Accelerator with Overlapped Mapping Method", in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2018.

✧ **Corresponding author:**

[HPCA-25] Tongxin Xie*, **Zhenhua Zhu***, Bing Li, Yukai He, Cong Li, Guangyu Sun, Huazhong Yang, Yuan Xie, Yu Wang, "UniNDP: A Unified Compilation and Simulation Tool for Near DRAM Processing Architecture", to appear in International Symposium on High-Performance Computer Architecture (HPCA), IEEE, 2025.

[DATE-24] Tongxin Xie, Tianchen Zhao, **Zhenhua Zhu**, Xuefei Ning, Bing Li, Guohao Dai, Huazhong Yang and Yu Wang, "DyPIM: Dynamic-inference-enabled Processing-In-Memory Accelerator", in *Design, Automation and Test in Europe Conference*, IEEE, 2024.

[ICCAD-24] Lidong Guo*, **Zhenhua Zhu***, Tengxuan Liu, Xuefei Ning, Shiyao Li, Guohao Dai, Huazhong Yang, Wangyang Fu and Yu Wang, "Towards Floating Point-Based Attention-Free LLM: Hybrid PIM with Non-Uniform Data Format and Reduced Multiplications", in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, IEEE/ACM, 2024.

✧ **Other papers:**

Journal:

[IEEE TCAD] Yuan Shuai, Weifeng He, Zhenhua Zhu, Fangxin Liu, Zhuoran Song, Guohao Dai, Guanghui He, Yanan Sun, "HyCTor: A Hybrid CNN-Transformer Network Accelerator with Flexible Weight/Output Stationary Dataflow and Multi-Core Extension", in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD),* IEEE, 2024.

[IEEE TCSVT] Shiyao Li, **Zhenhua Zhu**, Hanbo Sun, Xuefei Ning, Guohao Dai, Yiming Hu, Huazhong Yang, Yu Wang, "Toward High-accuracy and Real-time Two-stage Small Object Detection on FPGA" , in I*EEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, IEEE, 2024.

[IEEE TCAD] Minhui Zou, **Zhenhua Zhu**, Tzofnat Greenberg-Toledo, Orian Leitersdorf, Jiang Li, Junlong Zhou, Yu Wang, Nan Du, Shahar Kvatinsky, TDPP: 2-D Permutation-Based Protection of Memristive Deep Neural Networks, in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1-1, IEEE, vol. 43, no. 3, pp. 742-755, March 2024.

[IEEE TCAD] Hanbo Sun, **Zhenhua Zhu**, Chenyu Wang, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang, "Gibbon: An Efficient Co-Exploration Framework of NN model and Processing-In-Memory Architecture", in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1-1, IEEE, 2023.

[IEEE TCAD] Kai Zhong, Shulin Zeng, Wentao Hou, Guohao Dai, **Zhenhua Zhu**, Xuecang Zhang, Shihai Xiao, Huazhong Yang, Yu Wang, "CoGNN: An Algorithm-Hardware Co-Design Approach to Accelerate GNN Inference with Mini-Batch Sampling", in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 4883-4896, 42,12, IEEE, 2023.

[IEEE MNANO] Mingyuan Ma*, Yu Zhu*, **Zhenhua Zhu**, Rui Yuan, Jialong Liu, Liying Xu, Yuchao Yang, and Yu Wang, "Efficient In-Memory AES Encryption Implementation Using a General Memristive Logic: Surmounting the data movement bottleneck" , in *IEEE Nanotechnology Magazine*, 24 - C3, 16,2, IEEE, 2022

[IEEE TCAD] Kai Zhong, Xuefei Ning, Guohao Dai, **Zhenhua Zhu**, Tianchen Zhao, Shulin Zeng, Yu Wang, Huazhong Yang, "Exploring the Potential of Low-bit Training of Convolutional Neural Networks", in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, IEEE, 2022.

[ACM TODAES] Xuefei Ning, Guangjun Ge, Wenshuo Li, **Zhenhua Zhu**, Yin Zheng, Xiaoming Chen, Zhen Gao, Yu Wang, Huazhong Yang, "FTT-NAS: Discovering Fault-Tolerant Neural Architecture", in *ACM TODAES*, ACM, 2021.

[IEEE TCAD] Ming Cheng, Lixue Xia, **Zhenhua Zhu**, Yi Cai, Yuan Xie, Yu Wang, Huazhong Yang, "TIME: A Training-in-memory Architecture for RRAM-based Deep Neural Networks" *, to appear in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2019.

[AMT] Liying Xu, Rui Yuan, **Zhenhua Zhu**, Keqin Liu, Zhaokun Jing, Yimao Cai, Yu Wang, Yuchao Yang, Ru Huang, "Memristor-Based Efficient In-Memory Logic for Cryptologic and Arithmetic Applications", in *Advanced Materials Technologies*, 1900212, 4,7, Wiley, 2019.

Conference:

[ASPLOS-24] Kai Zhong, **Zhenhua Zhu**, Guohao Dai, Hongyi Wang, Xinhao Yang, Haoyu Zhang, Jin Si, Qiuli Mao, Shulin Zeng, Ke Hong, Genghan Zhang, Huazhong Yang and Yu Wang, "FEASTA: A Flexible and Efficient Accelerator for Sparse Tensor Algebra in Machine Learning" , in 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, ACM, 2024.

[DAC-24] Hongyi Wang, Kai Zhong, Haoyu Zhang, Shulin Zeng, **Zhenhua Zhu**, Xinhao Yang, Shuang Wang, Guohao Dai, Huazhong Yang and Yu Wang, "DySpMM: From Fix to Dynamic for Sparse Matrix-Matrix Multiplication Accelerators" , in Design Automation Conference (DAC), ACM/IEEE, 2024.

[DAC-24] Chenyu Wang, Zhen Dong, Daquan Zhou, **Zhenhua Zhu,** Yu Wang, Jiashi Feng, Kurt Keutzer, "EPIM: Efficient Processing-In-Memory Accelerators based on Epitome", in Design Automation Conference (DAC), ACM/IEEE, 2024.

[DAC-23] Yu Zhu, **Zhenhua Zhu**, Guohao Dai, Fengbin Tu, Hanbo Sun, Kwang-Ting Cheng, Huazhong Yang, Yu Wang, "PIM-HLS: An Automatic Hardware Generation Tool for Heterogeneous Processing-In-Memory-based Neural Network Accelerators", in *Design Automation Conference (DAC)*, IEEE/ACM 2023.

[DATE-23] Tianyu Fu*, Chiyue Wei***, Zhenhua Zhu,** Shang Yang, Zhongming Yu, Guohao Dai, Huazhong Yang, and Yu Wang , "CLAP: Locality Aware and Parallel Triangle Counting with Content Addressable Memory" , in *Design, Automation and Test in Europe Conference*, IEEE, 2023.

[DATE-23] Hanbo Sun, Tongxin Xie, **Zhenhua Zhu,** Guohao Dai, Huazhong Yang and Yu Wang, "Minimizing Communication Conflicts in Network-On-Chip based Processing-In-Memory Architecture", in *Design, Automation and Test in Europe Conference*,

IEEE, 2023.

[HPCA-23] Jiangwei Zhang, Chong Wang, **Zhenhua Zhu,** Donald Kline Jr, Alex K. Jones, Huazhong Yang, Yu Wang, "Realizing Extreme Endurance Through Fault-aware Wear Leveling and Improved Tolerance", in *the 29th IEEE International Symposium on High-Performance Computer Architecture (HPCA-29)*, Montreal, QC, Canada, 2023.

[MDM-22] Jun Liu, **Zhenhua Zhu**, Jingbo Hu, Hanbo Sun, Li Liu, Lingzhi Liu, Guohao Dai, Huazhong Yang, Yu Wang, "Optimizing Graph-based Approximate Nearest Neighbor Search: Stronger and Smarter", in *International Conference on Mobile Data Management*, 2022.

[DAT-E22] Hanbo Sun*, Chenyu Wang*, **Zhenhua Zhu**, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang, "Gibbon: Efficient Co-Exploration of NN Model and Processing-In-Memory Architecture", in *Design, Automation and Test in Europe Conference*, 2022.

[AICAS-21] Kaizhong Qiu, **Zhenhua Zhu**, Yi Cai, Hanbo Sun, Yu Wang, Huazhong Yang, "MNSIM-TIME: Performance Modeling Framework for Training-In-Memory Architectures", in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2021.

[ICCAD-21] Yitu Wang, **Zhenhua Zhu**, Mingyuan Ma, Fan Chen, Guohao Dai, Yu Wang, Hai "Helen" Li, Yiran Chen, "ReRec: In-ReRAM Acceleration with Access-Aware Mapping for Personalized Recommendation", in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 1-9, 2021.

[DATE-20] Minhui Zou, **Zhenhua Zhu**, Yi Cai, Junlong Zhou, Chengliang Wang, Yu Wang, "Security Enhancement for RRAM Computing System through Obfuscating Crossbar Row Connections", in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020.

[ASPDAC-19] Jilan Lin, **Zhenhua Zhu**, Yu Wang, Yuan Xie, "Learning the Sparsity for ReRAM: Mapping and Pruning Sparse Neural Network for ReRAM based Accelerator", in *Proceedings of the 24th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2019.

[ICECS-19] Jialong Liu, Mingyuan Ma, **Zhenhua Zhu**, Yu Wang, Huazhong Yang, "HDC-IM: Hyperdimensional Computing In-Memory Architecture based on RRAM", *in IEEE International Conference on Embedded Circuit and System,* 2019.

[DATE-18] Jilan Lin, Lixue Xia, **Zhenhua Zhu**, Hanbo Sun, Yi Cai, Hui Gao, Ming Cheng, Xiaoming Chen, Yu Wang and Huazhong Yang, "Rescuing Memristor-based Computing with Non-linear Resistance Levels", in *DATE 2018*, 407-412, 2018.

[ASPDAC-18] Yi Cai, Tianqi Tang, Lixue Xia, Ming Cheng, **Zhenhua Zhu**, Yu Wang, Huazhong Yang, "Training Low Bitwidth Convolutional Neural Networks on RRAM", in *Proceedings of the 23rd Asia and South Pacific Design Automation Conference (ASP-DAC)*, 117-122, 2018.

[DAC-17] Ming Cheng, Lixue Xia, **Zhenhua Zhu**, Yi Cai, Yuan Xie, Yu Wang, Huazhong Yang, "TIME:A Training-in-memory Architecture for Memristor-based Deep Neural Network", in *Design Automation Conference (DAC)*, 26:1-26:6, 2017.

INVITED TALK, WORKSHOP & TUTORIAL

Co-Organizer of **Workshop** on LLM Acceleration using PIM Architecture @ ESWEEK (Raleigh, USA) 2024

Co-Organizer of **Special Session** on Making LLM Faster, Stronger, and More Efficient @ IEEE SOCC (Dresden, Germany) 2024

Tutorial Speaker, When Sparse Computing Meets Near-Data Processing @ ASP-DAC 2024 (Seoul, South Korea) 2024

Co-Organizer of **Workshop** on Design and Optimization Tool for PIM Architecture @ ESWEEK (Hamburg, Germany) 2023

Invited Talk, Towards Energy-Efficient Memory-Centric Computer Architectures: A Design Tool Perspective @ ESWEEK Workshop on Memory and Storage Computing 2024

Invited Talk, Towards Energy-Efficient PIM Architectures for AI 2.0 Era @ CCF HPC China (Wuhan, China) 2024

Invited Talk, Design Methodology for Energy-Efficient Memory-Centric Computer Architecture @ CCF CFTC (Changzhou, China) 2023

SCHOLARSHIP

Shuimu Tsinghua Scholar Program 2024

Tsinghua Integrated Second-Class Scholarship 2022

Tsinghua Dept. of EE Outstanding Teaching Assistant First-Class Award 2021

Tsinghua Integrated Second-Class Scholarship 2021

Tsinghua Dept. of EE Outstanding Teaching Assistant Special Award 2020

National Scholarship for Graduates	2020
Tsinghua Integrated First-Class Scholarship	2019
Tsinghua EE Department Future Scholar Award	2018
Academic Excellence Award	2017
Tsinghua - Liyuan Scholarship for Encourage	2014, 2015, 2016
Tsinghua Academic Progress Scholarship	2015, 2016
Tsinghua Social Work Scholarship	2015, 2017
Friends of Tsinghua - Yicong Huang's Couple Scholarship	2016