

ZHIHAN XU

Phone: (+1) 610-938-4632 ◊ Email: zhihanxu@usc.edu

Homepage: zhihanxu.github.io

Google Scholar ◊ Github ◊ LinkedIn

EDUCATION

University of Southern California (USC)

Aug. 2023 - Now

Ph.D. in Computer Engineering

GPA: 3.87/4.0

· Working in FPGA/Parallel Computing Lab, advised by Professor Viktor K. Prasanna

University of Pennsylvania (UPenn)

May 2023

M.S. in Electrical Engineering

GPA: 3.84/4.0

· Outstanding Teaching Assistant (Penn Engineering)

University of Electronic Science and Technology of China (UESTC)

May 2020

B.Eng. in Electrical Engineering

GPA: 3.7/4.0

· Outstanding Graduate Student (University-level Distinction)

RESEARCH STATEMENT

My research interests broadly lie in computer system architecture and customized accelerator design. My recent works leverage reconfigurable computing technology (i.e., FPGAs) to accelerate homomorphic encryption and its applications, e.g., privacy-preserving machine learning.

EMPLOYMENT EXPERIENCE

Research Assistant, USC

Aug. 2023 - Now

Supervisor: Prof. Viktor K. Prasanna

- Contribute to two NSF-funded Projects: (i) Portable Library for Homomorphic Encrypted Machine Learning on FPGA Accelerated Cloud Cyberinfrastructure; (ii) Accelerating Privacy Preserving Deep Learning for Real-time Secure Applications
- Research outcomes published at FCCM '24, ISFPGA '25, FPL '25, HPEC '25, ISFPGA '26, FPL '26, TRETs '26 (All first author)

GPU Architecture Intern, Nvidia Corporation (US)

May 2026 - Aug. 2026

- Develop an agentic workflow for GPU functional model design, implementation, and verification.
- The workflow improves GPU design efficiency, reducing the effort required of hardware engineers.

GPU Design Verification Intern, Nvidia Corporation (US)

May 2022 - Aug. 2022

- Internship at Streaming Multiprocessor (SM) Design Verification Team focusing on testlist optimization
- Designed and deployed Baysian Optimization flow on testlists to improve test coverage. The algorithm explores the knob space and tunes knob values with a test result feedback loop to improve the hit rate
- Implemented NNs trained with random tests, accelerating the BayOpt flow for optimized testlists

Research Assistant, Shanghai Jiao Tong University

Sep. 2020 - Jul. 2021

Supervisors: Prof. Li Jiang, Prof. Zhezhi He

- Neural network compression and design space exploration for the FPGA-based CNN accelerator
- FPGA acceleration of CNN with heterogeneous cores, coupled with mixed-precision quantization
- Research outcome published at FPL '22

MENTORING AND TEACHING

Research Mentor, IUSSTF-Viterbi Program

May 2025 - Jul. 2025

Supervisor: Prof. Viktor K. Prasanna

- Mentored undergraduate Ashwajit Singh for FPGA acceleration of the number theoretic transform
- Research outcome accepted by HPEC '25

Teaching Assistant, USC

Jan. 2024 - May 2024

Lecturer: Gandhi Puvvada

- EE354L Introduction to Digital Circuits

Teaching Assistant, UPenn

Jan. 2023 - May 2023

Lecturer: Dr. Tania Khanna

- ESE5310 Digital Signal Processing

PROFESSIONAL SERVICES

Journal/Conference Peer Review

- International Workshop on Heterogeneous High-performance Reconfigurable Computing (H²RC '23)
- IEEE Transactions on Emerging Topics in Computing (TETC '24, '25)
- IEEE Transactions on Computers (TC '24)
- IEEE High Performance Extrem Computing Conference (HPEC '24)
- IEEE Transactions on Very Large Scale Integration Systems (TVLSI '24, '25, '26)
- IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD '25)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS '25, '26)
- ACM Transactions on Architecture and Code Optimization (TACO '25)

Conference Organization

- Proceedings Chair of IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC '24, '25)

ACHIEVEMENTS

Outstanding Teaching Award, awarded by Penn Engineering

May 2023

Outstanding Graduate Student (Distinction), awarded by UESTC

Dec. 2019

Outstanding Student Leadership Scholarship, awarded by UESTC

Dec. 2019

Outstanding Student Scholarship (top 10%) , awarded by UESTC

2017, 2018, 2019

Honorable Mention of Mathematical Contest in Modeling (MCM)

2019

PUBLICATIONS

- [1] Xu Z, Kannan R, Prasanna VK. S²MM: Scalable and Secure Matrix Multiplication with Homomorphic Encryption on FPGA. In ACM Transactions on Reconfigurable Technology and Systems (TRETs '26) (in press)
- [2] Xu Z*, Adivarahan J*, Kannan R, Prasanna VK. FPGA Acceleration of Fully Homomorphic Encryption with Adaptive Key Switching. In 2026 36th IEEE International Conference on Field-Programmable Logic and Applications (FPL '26)
- [3] Xu Z, Kannan R, Prasanna VK. HERA: A Bandwidth-efficient Accelerator for Fully Homomorphic Encryption on HBM-enabled FPGA. In Proceedings of the 2026 ACM/SIGDA International Symposium on Field Programmable Gate Arrays (ISFPGA '26)
- [4] Singh A*, Xu Z*, Prasanna VK. NTT-SAA: Exploring NTT Acceleration with 2-D Systolic Array Architecture on FPGAs. In 2025 IEEE High Performance Extreme Computing Conference (HPEC '25) (* equal contribution)
- [5] Xu Z, Kannan R, Prasanna VK. FAME: FPGA Acceleration of Secure Matrix Multiplication with Homomorphic Encryption. In 2025 35th IEEE International Conference on Field-Programmable Logic and Applications (FPL '25)
- [6] Xu Z, Ye T, Kannan R, Prasanna VK. FAST: FPGA Acceleration of Fully Homomorphic Encryption with Efficient Bootstrapping. In Proceedings of the 2025 ACM/SIGDA International Symposium on Field Programmable Gate Arrays (ISFPGA '25)
- [7] Xu Z, Yang Y, Kannan R, Prasanna VK. Bandwidth Efficient Homomorphic Encrypted Discrete Fourier Transform Acceleration on FPGA. In 2024 IEEE 32nd Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM '24)
- [8] Gong Y*, Xu Z*, He Z, Zhang W, Tu X, Liang X, Jiang L. N3h-core: Neuron-designed neural network accelerator via fpga-based heterogeneous computing cores. In Proceedings of the 2022 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (ISFPGA '22) (* equal contribution)
- [9] Xu Z, Ansari S, Abdulghani AM, Imran MA, Abbasi QH. IoT enabled smart security framework for 3D printed smart home. In 2020 IEEE International Conference on Smart Internet of Things (SmartIoT '20)
- [10] Wang H, Wang G, Xu Z, Lei W, Zhang S. High-and low-level feature enhancement for medical image segmentation. Cham: Springer International Publishing. In 2019 International Workshop on Machine Learning in Medical Imaging (MLMI '19)

OTHER PUBLICATIONS (EXTENDED ABSTRACTS AND POSTERS)

- [1] Xu, Zhihan; Yang, Yang; Prasanna, Viktor (2025). Portable Library for Homomorphic Encrypted Machine Learning on FPGA Accelerated Cloud Cyberinfrastructure. figshare. Poster. <https://doi.org/10.6084/m9.figshare.29648843.v1>
(2024) <https://doi.org/10.6084/m9.figshare.26540695.v1>
(2023) <https://doi.org/10.6084/m9.figshare.24201921.v3>