Bingyi Zhang

bingyizh@usc.edu
(571)-443-9187

EDUCATION

Fudan University — Shanghai, China
B.Eng., Micro Electronics — *Sept. 2013 - Jun. 2017*

Fudan University — Shanghai, China
Master of Engineering, Integrated circuit engineering — *Sept. 2017 - Jun. 2019*

University of Southern California — Los Angeles, USA
Ph.D. of Computer Engineering, Ming Hsieh Dept of ECE — *Sept. 2019 - present*

Research Interest

Parallel and distributed computing, Computer architecture, Reconfigurable computing, Graph-based machine learning, Digital signal processing, Very large-scale integrated circuit (VLSI), Edge computing

PUBLICATION

- Soundarya Jayaraman, **Bingyi Zhang**, Viktor Prasanna, Hypersort: High-performance Parallel Sorting on HBM-enabled FPGA, the International Conference on Field Programmable Technology (FPT 2022)
- Yi-Chien Lin, **Bingyi Zhang**, Viktor Prasanna, "Accelerating GNN Training on CPU+Multi-FPGA Heterogeneous Platform", The Latin America High Performance Computing (CARLA), 2022
- **Bingyi Zhang**, Hanqing Zeng, Viktor Prasanna, Low-latency Mini-batch GNN Inference on CPU-FPGA Heterogeneous Platform, 29th IEEE international conference on high performance computing, data, & analytics (HiPC 2022)
- **Bingyi Zhang**, Akhilesh Jaiswal, Clynn Mathew, Ravi Teja Lakkireddy, Ajey P. Jacob, Sasindu Wijeratne, Viktor Prasanna, Modeling the Energy Efficiency of GEMM using Optical Random Access Memory, 2022 IEEE High Performance Extreme Computing Virtual Conference (HPEC 2022)
- **Bingyi Zhang**, Rajgopal Kannan, Viktor Prasanna, Carl Busart, Accurate, Low-latency, Efficient SAR Automatic Target Recognition on FPGA, International Conference on Field Programmable Logic and Applications (FPL), 2022.
- Hongkuan Zhou, **Bingyi Zhang (equal contribution)**, Rajgopal Kannan, Viktor Prasanna, Carl Busart, Model-Architecture Co-Design for High Performance Temporal GNN Inference on FPGA, The 36th IEEE Internatioal Parallel and Distributed Processing Symposium. (IPDPS 2022).
- **Bingyi Zhang**, Hanqing Zeng, and Viktor Prasanna, DecGNN: A Framework for Mapping Decoupled GNN Models on CPU-FPGA Heterogeneous Platform, The 2022 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. (FPGA 2022).
- Yi Chien Lin, **Bingyi Zhang (equal contribution)**, Viktor Prasanna, "HP-GNN: Generating High Throughput GNN Training Implementation on CPU-FPGA Heterogeneous Platform", The 30th ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2022).
- Madhav Aggarwal, **Bingyi Zhang (equal contribution)**, Viktor Prasanna. Performance of Local Push Algorithms for Personalized PageRank on Multi-core Platforms. 28th IEEE international conference on high performance computing, data, & analytics (HiPC 2021)

- Yi Chien Lin; **Bingyi Zhang**; Prasanna, Viktor. GCN Inference Acceleration using High-Level Synthesis. 2021 IEEE High Performance Extreme Computing Virtual Conference (HPEC 2021).
- **Bingyi Zhang**; Sanmukh R. Kuppannagari; Kannan, Rajgopal ; Prasanna, Viktor. BoostGCN: Efficient Neighbor-Sampling-based GNN Trainingon CPU-FPGA Heterogeneous Platform. 2021 IEEE High Performance Extreme Computing Virtual Conference (HPEC 2021).
- **Bingyi, Zhang**; Kannan, Rajgopal ; Prasanna, Viktor. BoostGCN: A Framework for Optimizing GCN Inference on FPGA. FCCM, 2021.
- **Bingyi Zhang**, Hanqing Zeng, Viktor Prasanna, "A Framework for Optimizing GCN Inference on FPGA", The 29th ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2021).
- **Bingyi, Zhang**; Zeng, Hanqing; Prasanna, Viktor, Hardware acceleration of large scale gcn inference, The 31st IEEE International Conference on Application-specific Systems, Architectures and Processors, 2020
- **Bingyi Zhang**, Hanqing Zeng, Viktor Prasanna, "Accelerating Large Scale GCN inference on FPGA", The 28th IEEE International Symposium On Field-Programmable Custom Computing Machines. 2020.
- **Bingyi Zhang**, Xin Li, Jun Han, Xiaoyang Zeng, "MiniTracker: A Lightweight CNN-based system for Visual Object Tracking on Embedded Device", International Conference on Digital Signal Processing, 2018.
- **Bingyi Zhang**, Jun Han, Zhize Huang, Jianwei Yang, Xiaoyang Zeng, "A Realtime and Hardware-efficient Processor for Skeleton-based Action Recognition with Lightweight Convolutional Neural Network", IEEE Transaction on Circuits and Systems II: Express Briefs
- J. Yin, J. Han, C. Wang, **Bingyi Zhang** and X. Zeng, "A Skeleton-based Action Recognition System for Medical Condition Detection," 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 2019, pp. 1-4.

Award

University of Southern California Annenberg fellowship

Work Experience

University of Southern California Los Angeles, USA

- Research assistant for Professor Viktor K. Prasanna *Sept. 2019 - present*
- Teaching assistant for CSCI-570: Analysis of Algorithms *Sept. 2020 - May 2021*
- Teaching assistant for CSCI-356: Introduction to Computer Systems *Feb. 2022 - Dec. 2022*

Advanced Micro Devices (AMD) *Sept. 2016 - Mar. 2017*

BIOS Design Intern

- Working to build the BIOS Test Suite Program to verify the correctness of the BIOS, using C++ and C#.
- Developing a tool to check the code style of BIOS Test Suite Program using Perl.

SYNERGISTIC ACTIVITIES

Organization:

- **Local Arrangements Chair**; The 31th IEEE International Symposium On Field-Programmable Custom Computing Machines (FCCM)

- **Program Committee Member**; The 16th International Conference on Solid-State and Integrated Circuit Technology (ICSICT 2022)
- **Program Committee Member**; 14th International Conference on ASIC (ASICON 2021)

Reviewer Experience:

- **Reviewer**: International Conference on Field Programmable Technology (FPT), 2022.
- **Reviewer**: Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), 2021-2022;
- **Reviewer**: Transactions on Computers (TC), 2022
- **Reviewer**: The 30th ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2022)
- **Reviewer**: International Symposium on Computer Architecture (ISCA), 2021;
- **Reviewer**: IEEE Transaction on Circuits and Systems II: Express Briefs (TCAS II), 2020-2022;
- **Reviewer**: The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC), 2022;
- **Reviewer**: 33rd IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2022;
- **Reviewer**: 32rd IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2021;
- **Reviewer**: Future Generation Computer Systems, 2022;

PROFESSIONAL SKILLS

Expertise: Parallel computing, machine learning, digital circuit design, digital signal processing.
Software programming: C, C++, Python, Matlab, Perl
Hardware design:

- *Hardware description languages (HDL)*: Vivado HLS, Verilog.
- *Design automation Tool (EDA)*: Vivado, Design compiler, IC compiler, Cadence,

ACADEMIC COMPETITION

China Graduate Mathematical Modeling Competition *Nov. 2017*

Result: Second Prize (Top 10%)

- Personal contribution: algorithm design and software development

First National Undergraduate Integrated Circuit Innovation and Entrepreneurship Competition of China *Feb. 2017*

Topic: Neural network accelerator. Result: Rank 3rd out of 150+ teams

- Personal contribution: Team leader, algorithm designer and software developer
- Collaboratively design a neural network accelerator

Third National Undergraduate Integrated Circuit Innovation and Entrepreneurship Competition of China *Feb. 2019*

Topic: Hardware acceleration for convolutional neural networks. Result: Second price out of 500+ teams

- Personal contribution: hardware developer and software developer.
- Collaboratively build a hardware accelerator which can achieve 14 fps for VGG16 Net on FPGA.