Changchun Zhou

Yiheyuan Road 5, Haidian District • Beijing 10087, P.R.China
zhouchch@pku.edu.cn • (+86) 15627861177

EDUCATION

Peking University - Beijing, China 09/2018 – 12/2023

Doctor of Philosophy in Microelectronics and Solid-State Electronics

- Thesis Title: Research on On-Chip Neural Network Accelerators for 3D Understanding
- GPA: 3.6/4.0, 1/157 in Comprehensive Ranking
- Supervisor: Prof. Hailong Jiao

Sun Yat-sen University - Guangzhou, China 09/2014 – 06/2018

Bachelor of Engineering in Microelectronics Science and Engineering

- GPA: 3.8/5.0
- Final Project Supervisor: Prof. Tao Su

RESEARCH INTERESTS

- Energy-Efficient AI Accelerator for Edge Computing
- Algorithm-Hardware Co-Design

HONORS & AWARDS

- **Leo KoGuan Scholarship (1/157, ¥20, 000)**, *Peking University* 12/2023
- **Exceptional Award for Academic Innovation**, *Peking University* 12/2023
- Merit Student, *Peking University* 12/2023
- Award for Scientific Research, *Peking University* 12/2022
- Best Presentation Award, *IEEE CASS Shanghai and Shenzhen Joint Workshop* 05/2021
- Merit Student, *Peking University* 10/2019
- National Inspirational Scholarship, *Sun Yat-sen University* 10/2016
- National Inspirational Scholarship, *Sun Yat-sen University* 10/2015
- First Class Scholarship, *Sun Yat-sen University* 10/2015

PUBLICATIONS

- **C. Zhou**, M. Liu, S. Qiu, X. Cao, Y. Fu, Y. He, and H. Jiao, "Sagitta: An Energy-Efficient Sparse 3D-CNN Accelerator for Real-Time 3D Understanding," *IEEE Internet of Things Journal (**JIOT**)*, vol. 10, no. 23, pp. 20703-20717, 2023. (IF=10.6)
- **C. Zhou**, M. Liu, S. Qiu, Y. He, and H. Jiao, "An Energy-Efficient Low-Latency 3D-CNN Accelerator Leveraging Temporal Locality, Full Zero-Skipping, and Hierarchical Load Balance," in *Proc. of the IEEE/ACM Design Automation Conference (**DAC**)*, pp. 241-246, December 2021
- **C. Zhou**, Y. Fu, M. Liu, S. Qiu, G. Li, Y. He, and H. Jiao, "An Energy-Efficient 3D Point Cloud Neural Network Accelerator with Efficient Filter Pruning, MLP Fusion, and Dual-Stream Sampling," in *Proc. of the IEEE/ACM International Conference On Computer Aided Design (**ICCAD**)*, pp. 1-9, Oct. 2023
- **C. Zhou***, Y. Fu*, Y. Ma, E. Han, Y. He, and H. Jiao, "Adjustable Multi-Stream Block-Wise Farthest Point Sampling Acceleration in Point Cloud Analysis," *IEEE Transactions on Circuits and Systems II: Express Briefs (**TCAS-II**)*, 2024, Accepted
- M. Liu, **C. Zhou**, S. Qiu, Y. He, and H. Jiao, "CNN Accelerator at the Edge with Adaptive Zero Skipping and Sparsity-Driven Data Flow," *IEEE Transactions on Circuits and Systems for Video Technology(**TCSVT**)*, 2023, vol. 33, no. 12, pp. 7084-7095, Dec. 2023. (IF=8.4)
- Y. Fu, **C. Zhou**, T. Huang, E. Han, Y. He, and H. Jiao, "SoftAct: A High-Precision Softmax Architecture for Transformers Supporting Nonlinear Functions," *IEEE Transactions on Circuits and Systems for Video Technology(**TCSVT**)*, 2024, Under Review
- C. Zhang, Z. Huang, Q. Cheng, **C. Zhou**, and X. Wang, "An Energy-Efficient Configurable Coprocessor Based on 1-D CNN for ECG Anomaly Detection," in *Proc. of the IEEE International Symposium on Circuits and Systems (**ISCAS**)*, 2024, Accepted

PROJECT EXPERIENCE

Nebula: An Energy-Efficient 3D Point Cloud Neural Network Accelerator **(Tapeout)**

Project Leader & Main Contributor 02/2022 – Present

- Proposed an adjustable block-wise farthest point sampling acceleration framework, improving network accuracy by 9. 3% and reducing computations by 96.7%
- Developed a hierarchical data reuse strategy, leveraging block-wise delayed-aggregation between adjacent stages, and activation fusion across MLPs within each stage to increase the on-chip data reuse by 30.1×
- Explored a dynamic-static combined 24.8× operation skipping method, combining sampling-enabled operation skipping with efficient filter and channel pruning
- Fabricated in TSMC 28nm HPC CMOS technology with a die area of 2.0 mm×1.5 mm. Achieved effective energy efficiency (8-bit) exceeding 46 TOPS/W

Sagitta: An Energy-Efficient Sparse 3D-CNN Accelerator for Real-Time 3D Understanding **(Tapeout)**

Project Leader & Main Contributor 12/2019 – 12/2022

- Proposed a self-learning activation dropout method, which leverages spatial/temporal locality to dropout small differential activations below a learned threshold, reducing non-zero activations by 12.5×
- Developed a full-zero-skipping convolutional microarchitecture to skip both zero-value activations and zero-value weights, resulting in a speedup of 10.5×
- Explored a hierarchical load-balancing scheme, which combines fine-grained sub-task allocation with coarse-grained allocation of tasks with similar sparsity, increasing hardware utilization and improving energy efficiency by 2.2×
- Optimized hardware architecture and computation flow to enhance the effectiveness of the proposed techniques and enable the 3D convolution of differential input feature maps.
- Fabricated in UMC 55nm low-power CMOS technology with a die area of 4.2 mm×3.6 mm. Achieved effective energy efficiency of 4.5 TOPS/W for 3D U-Net

Libra: A 2.4 TOPS/W CNN Accelerator with Adaptive Zero Skipping and Sparsity-Driven Data Flow **(Tapeout)**

Module Leader 09/2020 – 06/2022

- Co-optimized the deployed CNN algorithms and the hardware architecture
- Designed an efficient storage management module for filtering out non-zero channels for PE array
- Assisted in tapeout in TSMC 28nm HPC CMOS technology, with a die area of 2.0 mm×1.35 mm

SSCNN: A 2.02 TOPS/W Sparse CNN Accelerator Skipping Invalid Activations **(Tapeout)**

Module Leader 12/2018 – 08/2020

- Designed a high-bandwidth asynchronous interface module to enable efficient off-chip storage access
- Responsible for logic synthesis to optimize area and power, as well as timing analysis for timing violation check
- Assisted in tapeout in UMC 55nm low-power CMOS technology, with a die area of 3.4 mm×2.3 mm

PATENTS

- H. Jiao, M. Liu, and **C. Zhou**, "A Convolution Operation Device Based on Systolic Array", CN Patent, ZL 202010447090.4, May 5, 2020.

PROFESSIONAL SERVICES

Engaged as a reviewer for more than 10 esteemed journals and conferences, which encompassed notable platforms such as IEEE Transactions on Circuits and Systems II (TCAS-II), IEEE Transactions on Very Large Scale Integration (VLSI) Systems, IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), IEEE International System-on-Chip Conference (SOCC)

SKILLS & TRAITS

- Flows: Architecture Exploration, RTL Implementation, Logic Synthesis, FPGA Development, Neural Network Training
- Tools: Cadence, Vivado, PyTorch, TensorFlow
- Languages: Verilog, SystemVerilog, Python, C, Shell, Makefile
- I am an easy going and self-motivated person. Personal website: https://changchun-zhou.github.io/