Dr. Jiajun Wu

Post-doc Fellow · AI Chip Center for Emerging Smart Systems

8/F, Building 17W, Hong Kong Science Park, NT, Hong Kong

jiajunwu@ust.hk | wujiajunic.cn | github.com/MaxwellWjj | linkedin.com/in/jiajun-wu-0121041a5/

Education

University of Hong Kong (HKU) *Hong Kong SAR*

PhD Candidate *2021 - 2025*

- Supervisor: Dr. Hayden Kwok-Hay So
- Research Field: *Hardware Acceleration Systems for AI Applications*
- Thesis Title: *Mixed-Precision Deep Neural Networks Acceleration: Architecture-Software Co-Design Approaches*

Huazhong University of Science and Technology (HUST) *China*

BEng, Integrated Circuit Design *2017 - 2021*

- Honored Thesis & Honored Graduate
- GPA: 3.89/4.0
- Supervisor: Prof. Chao Wang

Experience

Huazhong University of Science and Technology - Research Assistant *Wuhan, China*

Advisors: Prof. Chao Wang *2019-2021*

- Research Topic: Energy-Efficient brain-inspired computing

Zhuhai UM Science & Technology Research Institute - Research Intern *Zhuhai, China*

Advisor: Prof. Sai-Weng Sin *Jul. 2020-Oct. 2020*

- Research Topic: Analog Computing in the event sensor front-end

Singapore University of Technology and Design - Visiting Student *Singapore*

Advisor: Dr. Shaowei Lin *Jan. 2020-Feb. 2020*

- Project: Study of neuromorphic computing in machine learning algorithms

Wuxi EDA Institute, Peking University - Visiting Scholar *Wuxi, China*

Advisor: Prof. Yun Liang *Sep. 2024-Jan. 2025*

- Project: Balancing prefill and decode in LLM inference based on phase-level pipeline on FPGA

Publications

J. Wu, M. Song, J. Zhao, Y. Gao, J. Li, and H. K. -H. So. 2025. TATAA: Programmable Mixed-Precision Transformer Acceleration with a Transformable Arithmetic Architecture. ACM Trans. Reconfigurable Technol. Syst. (**TRETS**) 18, 1, Article 14 (March 2025), 31 pages.

J. Wu, M. Song, J. Zhao and H. K. -H. So, "A Case for Low Bitwidth Floating Point Arithmetic on FPGA for Transformer Based DNN Inference," 2024 IEEE International Parallel and Distributed Processing Symposium Workshops (**IPDPSW**), San Francisco, CA, USA, 2024, pp. 178-185

J. Wu, J. Zhou, Y. Gao, Y. Ding, N. Wong and H. K. -H. So, "MSD: Mixing Signed Digit Representations for Hardware-efficient DNN Acceleration on FPGA with Heterogeneous Resources," 2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (**FCCM**), Marina Del Rey, CA, USA, 2023, pp. 94-104

Y. Ding[1], **J. Wu**[1], Y. Gao, M. Wang and H. K. -H. So, "Model-Platform Optimized Deep Neural Network Accelerator Generation through Mixed-Integer Geometric Programming," 2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (**FCCM**), Marina Del Rey, CA, USA, 2023, pp. 83-93

J. Wu, Y. Zhan, Z. Peng, X. Ji, G. Yu, R. Zhao and C. Wang, "Efficient Design of Spiking Neural Network With STDP Learning Based on Fast CORDIC," in IEEE Transactions on Circuits and Systems I: Regular Papers (**TCAS-I**), vol. 68, no. 6, pp. 2522-2534, June 2021

J. Wu, X. Huang, L. Yang, J. Wang, B. Liu, Z. Wen, J. Li, G. Yu, K. -S. Chong and C. Wang, "An Energy-Efficient Deep Belief Network Processor Based on Heterogeneous Multi-Core Architecture With Transposable Memory and On-Chip Learning," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems (**JETCAS**), vol. 11, no. 4, pp. 725-738, Dec. 2021

J. Wu, X. Huang, L. Yang, L. Wang, J. Wang, Z. Liu, K. -S. Chong, S. Lin and C. Wang, "An Energy-efficient Multi-core Restricted Boltzmann Machine Processor with On-chip Bio-plausible Learning and Reconfigurable Sparsity," 2020 IEEE Asian Solid-State Circuits Conference (**A-SSCC**), Hiroshima, Japan, 2020, pp. 1-4

J. Zhou[2], **J. Wu**[2], Y. Gao, Y. Ding, C. Tao, B. Li, F. Tu, K. -T. Cheng, H. K. -H. So, and N. Wong, "DyBit: Dynamic Bit-Precision Numbers for Efficient Quantized Neural Network Inference,", in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (**TCAD**), Early Access

M. Song, **J. Wu**, Y. Ding and H. K. -H. So, "SqueezeBlock: A Transparent Weight Compression Scheme for Deep Neural Networks," 2023 International Conference on Field Programmable Technology (**ICFPT**), Yokohama, Japan, 2023, pp. 238-243

J. Xu, Y. Zhan, Y. Li, **J. Wu**, X. Ji, G. Yu, W. Jiang, R. Zhao and C. Wang, "In situ aging-aware error monitoring scheme for IMPLY-based memristive computing-in-memory systems," IEEE Transactions on Circuits and Systems I: Regular Papers (**TCAS-I**), 69(1), pp.309-321

B. Liu, Z. Wen, H. Zhu, J. Lai, **J. Wu**, H. Ping, W. Liu, G. Yu, J. Zhang, Z. Liu, H. Zeng and C. Wang "Energy-efficient intelligent pulmonary auscultation for post COVID-19 era wearable monitoring enabled by two-stage hybrid neural network," in 2022 IEEE International Symposium on Circuits and Systems (**ISCAS**), pp. 2220-2224. IEEE, 2022

Q. Wang, Y. Zhan, B. Liu, **J. Wu**, Y. Shi, G. Yu, and C. Wang. "A Reconfigurable Area and Energy Efficient Hardware Accelerator of Five High-order Operators for Vision Sensor Based Robot Systems," in 2021 IEEE International Conference on Integrated Circuits, Technologies and Applications (**ICTA**), pp. 189-190. IEEE, 2021

Selected Awards

2021 **Postgraduate Research Scholarship**, University of Hong Kong

2021 **Honored Thesis & Honored Graduate**, Huazhong University of Science and Technology

[1]Equal Contribution

[2]Equal Contribution

2019 **Second Prize in Challenge Cup College Students' Extracurricular Academic Science and Technology Works Contest**, Ministry of Education, P. R. China

2019 **Second Prize in China University Intelligent Robot Creative Competition**, China Association for Science and Technology

2019 **Third Prize in TI Cup College Students' Electronic Design Competition**, Texas Instruments, and Ministry of Education, P. R. China

2019 **Merit Student** ($5\% \sim 10\%$), Huazhong University of Science and Technology

2018 **First Prize in TI Cup Hubei Province College Students' Electronic Design Competition (TI 杯全国大学生电子设计竞赛一等奖)**, Texas Instruments, and Ministry of Education, P. R. China

Academic Presentations

November 2024. *The Case for Mixed Integer and Floating-point Acceleration of Transformer Models*. In ACCESS Seminar, AI Chip Center for Emerging Smart Systems (ACCESS), Hong Kong

May 2023. *MSD: Mixing Signed Digit Representations for Hardware-efficient DNN Acceleration on FPGA with Heterogeneous Resources*. In 2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), CA, USA

August 2021. *Energy-efficient DNN Hardware Accelerator with On-chip Learning –A Deep Belief Network Processor Case*. Invited by IEEE CASS-EDS-SSCS HUST Student Branch Chapter, Wuhan, China.

October 2020. *An Energy-efficient Multi-core Restricted Boltzmann Machine Processor with On-chip Bio-plausible Learning and Reconfigurable Sparsity*. In 2020 IEEE Asian Solid-State Circuits Conference (A-SSCC), Virtual Event.

Teaching Experience

2025 **ELEC6098,** `Electronic & Mobile Commerce`, Teaching Assistant *HKU*

2023 **ELEC3342,** `Digital system design`, Teaching Assistant *HKU*

2022 **ELEC3342,** `Digital system design`, Teaching Assistant *HKU*

2022 **ELEC6036,** `High performance computer architecture`, Teaching Assistant *HKU*

Professional Development

- Reviewer of IEEE Transactions on Circuits and Systems I: Regular Papers (TCAS-I)
- Reviewer of Microelectronics Journal
- Reviewer of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)