

Wei Cheng

✉ wei.cheng@duke.edu | 🏠 weicheng14159.github.io | 📄 github.com/WeiCheng14159 | 🔗 linkedin.com/in/WeiCheng14159 | 🎓 Wei Cheng

Research Interests

Hardware-software co-design, Next-generation wireless system, Quantum control system, Non-volatile memory, AI accelerators

Education

Duke University

Ph.D. Student in Electrical and Computer Engineering

- Advisor: *Prof. Tingjun Chen, Prof. Yiran Chen*
- Paper [1] published in **WiNTECH'24**

Durham, NC

Sep 2023 - Now

National Cheng Kung University

M.S. in Computer Science and Information Engineering

- Advisor: *Prof. Ing-Chao Lin*
- Paper [2] published in **ICCAD'22**, Paper [3] published in **TCAS-I**
- Thesis: A Dual-Addressing Graph Processing Accelerator with Vertex Coalescing

Taiwan

Sep 2020 - Jun 2022

University of Hong Kong

B.E. in Computer Engineering

- Advisor: *Prof. Cho-Li Wang*
- Thesis: The performance optimization on TensorFlow framework on Mobile GPU devices using OpenCL

Hong Kong

Sep 2014 - Jun 2018

Research Experiences

Research Assistant, Electrical and Computer Engineering, Duke University

SPEAR: Software-defined Python-Enhanced RFSoc for Wideband Radio Applications [**WiNTECH'24**], supervised by *Prof. Tingjun Chen*

- Design a real-time wideband RF system based on Xilinx's RFSoc platform.

Durham, NC

Sep 2023 - Now

Research Assistant, Institute of Information Science, Academia Sinica

Graph processing on dual-addressing memory [**ICCAD'22**], supervised by *Prof. Yuan-Hao Chang*

- Design a graph processing accelerator for dual-addressing non-volatile memory (RCNM).

Taiwan

Feb 2021 - Jun 2023

Research Assistant, Computer Architecture & IC Design Lab, NCKU

CNN accelerator with CLIP-Q network quantization on FPGA [**TCAS-I**], supervised by *Prof. Ing-Chao Lin*

- Design a hardware-software co-designed CNN accelerator based on the CLIP-Q network quantization algorithm.

Taiwan

Sep 2020 - Jun 2022

Summer Intern, TCL Corporate Research (HK) Co., Ltd

Testing Structure from Motion (SfM), Simultaneous localization and mapping (SLAM) algorithms

Hong Kong

Summer 2016

Publications

- [1] W. Cheng, Z. Gao, and T. Chen, "SPEAR: Software-defined Python-Enhanced RFSoc for Wideband Radio Applications," in *Proc. ACM WiNTECH'24*, Washington D.C., 2024.
- [2] W. Cheng, C.-F. Wu, Y.-H. Chang, and I.-C. Lin, "GraphRC: Accelerating Graph Processing on Dual-Addressing Memory with Vertex Merging," in *Proc. of the 41st IEEE/ACM Int. Conf. on Comput.-Aided Des.*, San Diego CA, 2022, pp. 1–9. [Online]. Available: <https://dl.acm.org/doi/10.1145/3508352.3549408>
- [3] W. Cheng, I.-C. Lin, and Y.-Y. Shih, "An Efficient Implementation of Convolutional Neural Network With CLIP-Q Quantization on FPGA," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 69, no. 10, pp. 4093–4102, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9849674/>

Teachings

- | | | |
|------|--|---------------|
| 2022 | Teaching Assist , AI Accelerator with Good Performance Course (reserved for faculty members and graduate students only) taught by Prof. H. T. Kung with 100+ people | <i>Taiwan</i> |
| 2021 | Teaching Assist , Computer Organization Course (Undergraduate level) with 100+ students | <i>Taiwan</i> |
| 2021 | Teaching Assist , Deep Learning IC Design Course (Graduate level) | <i>Taiwan</i> |

Honors & Awards

2022	Best Dissertation Award , IEEE Taipei Section	Taiwan
2022	Best Master Thesis Award , IEEE Tainan Section	Taiwan
2022	Master Thesis Award , Institute of Information and Computing Machinery (IICM)	Taiwan
2014-2018	HKU Foundation Scholarship for International Students ,	Hong Kong
2016-2017	Reaching Out Award scholarship from HKSAR gov. ,	Hong Kong
2018	Final Year Project Merit Award , IEEE Hong Kong	Hong Kong

Projects

SPEAR: A 1.25 GHz bandwidth real-time RF testbed

Durham, NC

Duke University

Sep 2023 - Now

- SPEAR is an SDR platform based on the Xilinx RFSoc ZCU216 evaluation board capable of supporting real-time streaming of signals with a bandwidth of up to 1.25 GHz employing the direct RF radio architecture.
- It comes with an DSP pipeline written in Python and reaches the 3GPP EVM requirement of 3.5% with 256QAM modulation.
- Github: <https://github.com/functions-lab/SPEAR>

A SW/HW co-designed RISC-V CNN accelerator for mask detection

Taiwan

National Cheng Kung University

Sep 2021 - Dec 2021

- RISC-V core consists of: pipelined RV32I core, I-cache/D-cache, AXI bus, DMA, DRAM/ROM controller, Interrupt manager.
- It also handles the booting sequence, data movements, the control of acceleration unit, and system interrupts.
- Apply network compression and quantization on a mask detection CNN model.
- Inference the compressed CNN model on an HW acceleration unit.
- Github: https://github.com/WeiCheng14159/VSD_CNN_accelerator

Contribute to ria-jit (an open source RISC-V to x86 binary translator)

Taiwan

National Cheng Kung University

Sep 2020 - Dec 2020

- Expose and fix a divide by zero bug with RISC-V compliance tests.

Contribute to srv32 (an open source 3-stage pipeline RV32IM core)

Taiwan

National Cheng Kung University

Sep 2020 - Dec 2020

- Verify and contribute RV32C instructions to the existing implementation.

Other Experiences

Taiwan Semiconductor Research Institute

Taiwan

Cell-based Digital IC Tapeout

2021

- Design, tapeout, and verification of an UMC 0.18 um process chip

University of Hong Kong

Hong Kong

Class representative for Computer Engineering major students

2015 - 2016, 2017 - 2018

Skills

Programming	Python, C++/C, Verilog/SystemVerilog
AI Frameworks	PyTorch, TensorFlow
EDA tools	NCSim, Design Compiler, IC Compiler/Innovus, Virtuoso
Miscellaneous	Linux, Shell, \LaTeX , Markdown, Git