JIAN MENG

https://mengjian0502.github.io/
(+1) 503 810 8292 ⋄ jmeng15@asu.edu

EDUCATION

Arizona State University *August 2019 - Present*
Doctor of Philosophy Overall GPA: 3.83/4.0
School of Electrical, Computer and Energy Engineering

Portland State University *September 2016 - June 2019*
Bachelor of Science (Received June 2019) Overall GPA: 3.61/4.0; Graduate level GPA: 3.75/4.0
Department of Electrical and Computer Engineering

RESEARCH EXPERIENCE

Seo Lab, Arizona State University **August 2019 - Present**
Advisor: Jae-sun Seo

- Design a novel dynamic sparsification algorithm for contrastive self-supervised learning. As one of the first studies in this area, the research analyzes different feature sparsification strategies with rigorous experiments. The proposed algorithm skips the uninformative features during contrastive learning without hurting the trainability of the networks, achieves up to 15% CIFAR-10 accuracy improvements compared to the prior works.
- Collaborated with Georgia Institute of Technology, focusing on the robustness enhancement of the DNN model with the given analog non-ideality (e.g., thermal variations). The research improves the robust RRAM IMC-based DNN inference where >30% CIFAR-10 accuracy and >60% TinyImageNet accuracy is recovered against temperature variations.
- Design a hardware-aware algorithm for a 28nm taped-out DNN training accelerator with FP8 precision. The designed algorithm exploits structured sparsity in both weights and feature maps, along with backward gradient skipping.
- Collaborated with ARM research team, investigated the compact MobileNet model compression along with the on-chip FPGA CNN accelerator design. The research compressed the MobileNets model down to 4-bit along with over 95% element sparsity. The proposed CNN accelerator design achieved 3.01 TOPS throughput with 0.37 ms latency on ImageNet-224 datasets.
- A system-level RRAM accelerator design incorporates low precision model with structured sparsity. The proposed software-hardware co-design achieved up to 7×/6× energy/area reduction with ResNet/VGG models for CIFAR-10 dataset with minimum accuracy degradation.

Teuscher Lab, Portland State University **January 2018 - June 2019**

- Design an energy-efficient architectures for mobile radiation detection platforms. Employed deep neural network and Unscented Kalman Filter for source localization and radiation pattern recognition.

WORKING EXPERIENCE

Teaching Assistant: Arizona State University Jan. 2022 - May 2022
Neuromorphic Computing Hardware Design

System Engineer: Kilby Labs of Texas Instrument Jun. 2021 - Aug. 2021

- **End-to-end compiler design for deep neural network acceleration.** Design and test a Pytorch-based compiler for deploying low-precision deep neural networks to the in-memory-computing-based accelerator. Given the pre-trained DNN model, the designed tool can automatically generate the C code for hardware deployment and allocates the computation resource.

Teaching Assistant: Portland State University Sep. 2018 - Jun. 2019

· ECE 221/2/3 (Circuit Analysis) and ECE 510 (Mathematical Foundation of Machine Learning)

SELECTED PUBLICATIONS

Conference Publications

[**DATE'22**] Fan Zhang, Li Yang, **Jian Meng**, Jae-sun Seo, Yu Cao and Deliang Fan, XST: A Crossbar Column-wise Sparse Training for Efficient Continual Learning, IEEE Design, Automation & Test in Europe (DATE) [**Best IP (Interactive Presentations) Paper Award**]

· [**CVPR'22**] **Jian Meng**, Li Yang, Jinwoo Shin, Deliang Fan, and Jae-sun Seo, "Contrastive Dual Gating: Learning Sparse Features With Contrastive Learning," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[**IRPS'22**] **Jian Meng**, Injune Yeo, Wonbo Shim, Li Yang, Deliang Fan, Shimeng Yu, and Jae-sun Seo "Sparse and Robust RRAM-based Efficient In-memory Computing for DNN Inference" (IRPS), 2022.

[**ESSCIRC'22**] Shreyas K. Venkataramanaiah, Jian Meng, Han-Sok Suh, Injune Yeo, Jyotishman Saikia, Sai Kiran Cherupally, Yichi Zhang, Zhiru Zhang, and Jae-sun Seo, A 28nm 8-bit Floating-Point Tensor Core based CNN Training Processor with Dynamic Activation/Weight Sparsification, IEEE European Solid-State Circuits Conference (ESSCIRC), 2022.

[**FPT'21**] Han-sok Suh, **Jian Meng**, Ty Nguyen, Shreyas K. Venkataramanaiah, Vijay Kumar, Yu Cao, and Jae-sun Seo, Algorithm-Hardware Co-Optimization for Energy-Efficient Drone Detection on Resource-Constrained FPGA, IEEE ICFPT, 2021.

[**FPL'21**] **Jian Meng**, Shreyas Kolala Venkataramanaiah, Chuteng Zhou, Patrick Hansen, Paul Whatmough and Jae-sun Seo, "FixyFPGA: Efficient FPGA Accelerator for Deep Neural Networks with High Element-Wise Sparsity and without External Memory Access", International Conference on Field Programmable Logic and Applications (FPL), 2021.

[**ISCAS'21**] **Jian Meng**, Li Yang, Xiaochen Peng, Shimeng Yu, Deliang Fan, Jae-sun Seo, "Structured Pruning of RRAM Crossbars for Efficient In-Memory Computing Acceleration of Deep Neural Networks", IEEE International Symposium on Circuits and Systems (ISCAS), 2021.

[**IRPS'21**] Wonbo Shim, **Jian Meng**, Xiaochen Peng, Jae-sun Seo, and Shimeng Yu, "Impact of Multilevel Retention Characteristics on RRAM based DNN Inference Engine" (IRPS), 2021

[**DATE'21**] Jyotishman Saikia, Shihui Yin, Bo Zhang, **Jian Meng**, Mingoo Seok and Jae-sun Seo, "Modeling and Optimization of SRAM-based In-Memory Computing Hardware Design," IEEE Design, Automation, and Test in Europe (DATE), 2021.

[**INTERSPEECH'20**] Deepak Kadetotad, **Jian Meng**, Visar Berisha, Chaitali Chakrabarti, and Jae-sun Seo, Compressing LSTM Networks with Hierarchical Coarse-Grain Sparsity, INTERSPEECH.

Journal Publications

[**IEEE SSCM**] Jae-sun Seo, Jyotishman Saikia, **Jian Meng**, Wangxin He, Han-sok Suh, Anupreetham, Yuan Liao, Ahmed Hasssan, and Injune Yeo, Advances in Digital vs. Analog AI Accelerators, IEEE Solid-State Circuits Magazine, 2022

[**IEEE MICRO**] **Jian Meng**, Wonbo Shim, Li Yang, Deliang Fan, Shimeng Yu, and Jae-sun Seo, Temperature-Resilient RRAM-based In-Memory Computing for DNN Inference, IEEE Micro, 2021 (**Invited & Presented in IBM Research AI Hardware Forum**)

[**IEEE JETCAS**] Arnab Neelim Mazumder, **Jian Meng**, Hasib-Al Rashid, Utteja Kallakuri, Xin Zhang, Jae-sun Seo, Tinoosh Mohsenin, "A Survey on the Optimization of Neural Network Accelerators for Micro-AI On-Device Inference", IEEE JETCAS, 2021

[**IEEE TCAS-II**] **Jian Meng**, Li Yang, Xiaochen Peng, Shimeng Yu, Deliang Fan, Jae-sun Seo, "Structured Pruning of RRAM Crossbars for Efficient In-Memory Computing Acceleration of Deep Neural Networks". IEEE TCAS-II, 2021.