YINTAO HE

(+86)-188-1051-1960 ⋄ heyintao19z@ict.ac.cn ⋄ https://yintao-he.github.io/resume/

EDUCATION

University of Chinese Academy of Sciences Sep. 2019 - Present
Ph.D. Student in Computer System and Architecture *Beijing, CN*

- Advisor: Prof. Huawei Li and Prof. Ying Wang
- Focus: In-Memory Computing, Energy-Efficient Accelerators, Deep Learning
- Overall GPA: 3.73/4.00

Nankai University Sep. 2015 - June 2019
Bachelor of Electronic Science and Technology (Rank 3/35) *Tianjin, CN*

PUBLICATIONS

- **Yintao He**, Ying Wang, Xiandong Zhao, Huawei Li, Xiaowei Li, "Towards State-Aware Computation in ReRAM Neural Networks," to appear in IEEE/ACM Proceedings of Design, Automation Conference (DAC 2020).
- **Yintao He**, Ying Wang, Yongchen Wang, Huawei Li, Xiaowei Li, "An Agile Precision-Tunable CNN Accelerator based on ReRAM," in IEEE/ACM International Conference On Computer Aided Design (ICCAD 2019).

RESEARCH EXPERIENCE

Towards State-Aware Computation in ReRAM Neural Networks Jul. 2019 - Dec. 2019
Institute of Computing Technology, Chinese Academy of Science *Beijing, CN*

- We propose a state-aware Signed ReRAM accelerator architecture for the regularized neural networks, which could effectively reduce the LRS rate in the crossbar without affecting the computation results.
- We propose a state-aware model training method for ReRAM-based BNNs, so that more weights can be represented by low-power HRS cells on the signed crossbars at a negligible loss of accuracy.
- The evaluation shows that our design reduces operation current effectively, and achieves 47% ReRAM computation in the best case and 85.85% SA energy saving on average with minor accuracy loss.

An Agile Precision-Tunable CNN Accelerator based on ReRAM Nov. 2018 - Apr. 2019

- We devise a novel NN training algorithm to train a mixed-precision neural network. It requires only one single set of parameters to work adaptively in different precision modes for the same NN architecture.
- We propose a ReRAM-based approximate accelerator design for agile precision-tunable Neural Network, which could achieve on-line tradeoff of computation efficiency and accuracy as the system demands.
- Experimental results show that ReRAM-based accelerator with the mixed-precision neural network reduces 58.3%-62.47% of area consumption compared with storing multiple models on chip.

SELECTED AWARD

- 2020: DAC 2020 Young Fellow Member
- 2020: UCAS Merit Student
- 2019: Outstanding Graduation Thesis (Top 5 in the department)
- 2019: NKU Merit Student
- 2018: The First Prize Scholarship