Jiajun Li

lijiajun@ict.ac.cn
(+86) 137-1641-2674
Room 719, Kexueyuan South Rd. No.6
Beijing, China

Research Interests

High-performance and energy-efficient computing. Specially, I focus on the following areas: (i) Specialized hardware; (ii) Parallel programming models and runtime; (iii) Performance optimization and modeling.

Education

Institute of Computing Technology, Chinese Academy of Scienses, Beijing Sep. 2013 – Present

Ph.D Candidate in Computer Science, expected June 2019
Advisor: Prof. Xiaowei Li

Tsinghua University, Beijing Sep. 2009 – July 2013

B.S. in Control Science and Engineering

Work Experience

Yusur Technology Co., Ltd, Beijing Aug. 2018 – Present

Co-founder& Software Director

- Research and development of software (including compiler, runtime, driver) for application-specific computing architecture solutions.

Research Experience

I. Enegy-efficient Neural Network Accelerator Architecture, ICT, CAS Sep. 2014 – Dec. 2016

Research Assistant NSFC, C/C++/Python

- Instruction Set Architecture (ISA), compiler framework for neural network accelerators
- Simulator for neural network accelerators (DeepWater). Link: https://github.com/ysalpha/DeepWater
- High performance and energy efficient neural network accelerator designs [Papers: TODAES'18, DATE'18a, DATE'18b, JOLPE'18, ASPDAC'19]

II. Quantitative financial toolbox, ICT, CAS Sep. 2016 – Aug. 2017

Research Assistant Individual project, C/C++/Python/Matlab

- Strategy development, data management, portfolio management in quantitative investment.
- Multi-factor model evaluation module, CTA strategy evaluation module and real-time risk control module

III. Software Stack& Toolchain for KPU, YUSUR Tech Jun. 2017 – Dec. 2018

Research Assistant NSFC, C/C++/Python/CUDA/Flex& Bison

- Architecture and micro-architecture design of a dedicated accelerator (KPU) for time series analysis [Papers: DAC'19]
- Computing graph-based programming& execution model, customized SDK& driver for KPU
- Application development based on customized SDK, including KPU-powered DBMS, multi-factor model evluation system, and real-time risk-control system.

Publications

[TODAES'18] **Jiajun Li**, Guihai Yan, Wenyan Lu *et al.* SynergyFlow: An Elastic Accelerator Architecture Supporting Batch Processing of Large-Scale Deep Neural Networks. ACM Trans Des Autom Electron Syst. 2018;24(1):1-27.

[DATE'18a] **Jiajun Li**, Guihai Yan, Wenyan Lu *et al.* CCR: A concise convolution rule for sparse neural network accelerators. Design, Automation & Test in Europe Conference & Exhibition (DATE); 2018: IEEE.

[DATE'18b] **Jiajun Li**, Guihai Yan, Wenyan Lu *et al.* SmartShuttle: Optimizing off-chip memory accesses for deep learning accelerators. 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE); 2018: IEEE.

[ASPDAC'19] **Jiajun Li**, Guihai Yan, Wenyan Lu *et al.* TNPU: an efficient accelerator architecture for training convolutional neural networks. Proceedings of the 24th Asia and South Pacific Design Automation Conference; 2019: ACM.

[JOLPE'18] Xiaowei Li, **Jiajun Li**, Guihai Yan. Optimizing Memory Efficiency for Deep Convolutional Neural Network Accelerators. Journal of Low Power Electronics. 2018;14(4):496-507.

[HPCA'17] Wenyan Lu, Guihai Yan, **Jiajun Li** *et al.* Flexflow: A flexible dataflow accelerator architecture for convolutional neural networks. High Performance Computer Architecture (HPCA), 2017 IEEE International Symposium on; 2017: IEEE.

[DAC'19] Shijun Gong, **Jiajun Li**, Wenyan Lu *et al.* ShuntFlow: An Efficient and Scalable Dataflow Accelerator Architecture for Streaming Applications. In 2019 56th ACM/EDAC/IEEE Design Automation Conference (DAC). 2019: IEEE.

[Book chapter] Guihai Yan, **Jiajun Li**, Xiaowei Li. Chapter "Modelling Many-core Architectures" in Book "Many-Core Computing: Hardware and software", Published by IET

Jiajun Li, Shuhao Jiang, Guihai Yan *et al.* SqueezeFlow: A Sparse CNN Accelerator Exploiting Concise Convolution Rules. (submitted to IEEE Trans. on Computers, first round, major revision)

Participated Grants

National Natural Science Foundation of China, Grant No. 61872336 2018–present

Research on design method of software definition dedicated computing architecture

National Natural Science Foundation of China, Grant No. 61521092 2016–2018

Research on computer architecture and design method of ultra-parallel and high efficiency energy design

National Natural Science Foundation of China, Grant No. 61572470 2016–2018

Efficient and cooperative power consumption management method for large-scale heterogeneous computing system

National Natural Science Foundation of China, Grant No. 61532017 2015–2017

Theory and method of fault detection in integrated circuits

National Natural Science Foundation of China Grant No. 61100016 2013–2014

Power consumption analysis and management optimization of heterogeneous multicore processors

Skills

- Programming Language: Python, C, C++, Matlab
- Deep Learning Framework: TensorFlow
- Database: Oracle, MySQL, PostgreSQL

Invited Lectures

Design, Automation & Test in Europe Mar. 22, 2018

Present my work on off-chip memory access optimization for deep learning accelerators Dresden, Germany

Design, Automation & Test in Europe Mar. 22, 2018

Present my work on sparse neural network accelerator design Dresden, Germany

Asia and South Pacific Design Automation Conference Jan. 23, 2019

Present my work on accelerator for training deep neural networks Tokyo, Japan

Honors and Awards

Endeavourers Scholarship, Tsinghua University	2012
The First Prize Scholarship, University of Chinese Academy of Scienses	2017
Outstanding papers on the 2nd selection of excellent scientific papers for CAST	2017
Merit Student, Institute of Computing Technology	2018
The Huawei Doctoral Scholarship, Institute of Computing Technology	2018

Language

- English: **CET-6**, Fluent English (both written and spoken)
- Japanese: JLPT-N3 level