Zhijing Li

Third-year Phd Student · Cornell University

(+1) 607-279-1527 | zl679@cornell.edu | https://tissue3.github.io/ | tissue3 | zhijingli94

Education

Cornell University *Sep. 2018 - Present*

Doctor of Philosophy in Electrical and Computer Engineering *GPA: 3.9/4.0*

- Research Interest: Programming Languages, Deep Learning Accelerator, Computer Architecture
- Committee: Adrian Sampson (chair), Christopher De Sa, Christina Delimitrou

Shanghai Jiao Tong University (SJTU) *Sep. 2014 - Aug. 2018*

Bachelor of Science in Electrical and Computer Engineering *GPA: 3.5/4.0*

- Advisor: Weikang Qian

Publications & Patents

Optimizing JPEG Quantization for Classification Networks (Workshop Paper). *ReCoML @ MLSys 2020*

Zhijing Li, Christopher De Sa, Adrian Sampson.

A Compiler Infrastructure for Accelerator Generators (Submitted). *Submitted to ASPLOS 2021*

Rachit Nigam, Samuel Thomas, **Zhijing Li**, Adrian Sampson.

Predictable Accelerator Design with Time-Sensitive Affine Types. *PLDI 2020*

Rachit Nigam, Sachille Atapattu, Samuel Thomas, **Zhijing Li**, Theodore Bauer, Yuwei Ye, Apurva Koti, Adrian Sampson, Zhiru Zhang.

Accurate Operation Delay Prediction for FPGA HLS using Graph Neural Networks. *ICCAD 2020*

Ecenur Ustun, Chenhui Deng, Debjit Pal, **Zhijing Li**, and Zhiru Zhang.

Simultaneous Area and Latency Optimization for Stochastic Circuits by D Flip-flop Insertion. *TCAD 2019*

Zhijing Li, Zhao Chen, Yili Zhang, Zixin Huang, and Weikang Qian.

A high-accuracy approximate adder with correct sign calculation. *Integration 2017*

Junjun Hu, **Zhijing Li**, Meng Yang, Zixin Huang, and Weikang Qian.

Android mobile terminal detecting method based on optimized Laplacian feature extraction and support vector machine (SVM) for resisting replay attacks *Pub No. 105913024A*

Yanfeng Sun, Xingghao Jiang, Zepeng Wang, **Zhijing Li**, Jialong Li.

Experience

Xilinx, Inc. *San Jose, CA*

Research Intern, Compiler Engineer *Jun. 2020 - Aug. 2020*

Developed EQueue dialect, a structure-control hybrid intermediate language (IL) that models hardware property and hierarchy, explicit memory movements and concurrency between distributed processors, using MLIR (a compiler built on LLVM). Proven the expressiveness of the EQueue dialect by implementing control flows on different machine learning accelerator designs. Built a fast, easily deployed performance simulator on energy and latency that models concurrency and memory movements for EQueue dialect programs using LLVM.

Cornell University *NY, United States*

Magnolia *Mar. 2020 - Present*

Developed Magnolia, an end-to-end domain specific language (DSL), for reconfigurable acceleration of machine learning training that balances specification with flexibility. The language facilitates easily adding different hardware abstractions, quick reconfiguration without FPGA synthesis, and control flow expressiveness on hardware structure using Scala and Verilog.

Dahlia *Sep. 2019 - Mar. 2020*

Joined Dahlia team to develop Dahlia, a high-level synthesis (HLS) language that guarantees predictable hardware through affine types and region-based capabilities implemented by Scala. Developed the loop dependency pass for Dahlia compiler that assures the unrolling factor constraints on loop-carried dependency, improving design space exploration (DSE) efficiency. Designed experiments to compare Dahlia with Spatial, a DSL with automated DSE tools, and proven the vitality of Dahlia. The work is published in **PLDI 2020**.

Calyx *Sep. 2019 - Aug. 2020*

Joined Calyx team to develop Calyx, a general intermediate language (IL) for compiling DSLs, e.g. Dahlia, to custom spatial architectures and Calyx compiler that implements a modular pass system for Calyx compiler using Rust. Implemented Calyx backend to generate finite-state machines from the control flow and emit synthesizable RTL descriptions. Gained 5.3× speedup on systolic array over highly optimized commercial HLS toolchain. The work is submitted in **ASPLOS 2021**.

Optimizing JPEG for Neural Networks *Feb. 2019 - Nov. 2019*

Redesigned Joint Photographic Experts Group (JPEG) algorithm to adapt to classification neural network that outperformed existing methods that target minimal image distortion or human visual system. Attempted several approaches including Bayesian Optimization and Multi-Armed Bandit to tune the quantization table. Developed a simple while efficient sorted random sampling method that exceeds the performance of the standard JPEG Q-table on compression rate by up to 200%. The work is published in **ReCoML workshop at MLSys 2020**.

Intel, Asia-Pacific Research and Development Ltd. *Shanghai, China*

Software Engineer Intern, Distributed System Engineer *June. 2017 - Sep. 2017*

Joined the HDCS (hyper- converged distributed storage) team to optimize the performance of Ceph, a distributed storage system. Proposed and implemented a key-value cache on client node for Ceph to speed up the performance on client nodes. Performed scalability tests on Ceph using Cetune, a cloud storage benchmarking and profiling tool.

Shanghai Jiao Tong University *Shanghai, China*

Optimizing Stochastic Circuits *Mar 2016 – May 2017*

Optimized DFF insertion on stochastic circuits using integer linear programming (ILP) method. Reduced long computation latency by 14.3% and overhead of DFFs by 48.1% compared to state of the art method. The work is published in **TCAD 2019**.

Approximate Computing *Sep 2015 – Jun 2016*

Developed a novel approximate adder that takes small power consumption and short delay, while minimizing the maximal computation error and guaranteeing the correct sign bit. Reduced power-delay product by 32% compared to carry-lookahead adder. Proven the importance of sign correction in applications including mean filter, edge detection, and k-means clustering. The work is published in **Intergration, the VLSI Journal 2017**.

Replay Attack Prevention *Sep 2015 – Mar 2016*

Developed a fast Laplacian feature extraction algorithm and applied to support vector machine (SVM) to prevent replay attack that achieves lower power consumption. Trained the SVM with CASIA Face Anti-Spoofing Database and transplanted the inference to Android device with Android NDK. Published the work with **an patent (105913024A)**.

Honors & Awards

2017	**Fung Scholar Program Sponsorship**	*Hong Kong University*
2015	**Distinguished Achievement in Extracurricular Activity Award**	*SJTU*
2015	**Dean's List**	*SJTU*
2014	**Dean's List**	*SJTU*
2015	**"Three Good" Student**	*SJTU*
2011	**SJTU Outstanding Scholarship**	*SJTU*
2015	**JI Best Presenter Award**	*SJTU*
2013	**2nd Award, National Chemistry Contest**	*Shanghai*
2012	**2nd Award, National Chemistry Contest**	*Shanghai*
2013	**Best leadership for volunteering work**	*No.2 High School Of East China Normal University*

Skills

Programming C/C++, Python, Scala/Java, Rust, Haskell, Ocaml, Verilog, HTML/CSS, SQL

DevTools GCC/Makefile/Cmake, GDB, Git, Vim