YUNJIE PAN

(734) 747-3150 ⋄ panyj@umich.edu ⋄ pyjhzwh.github.io

EDUCATION

University of Michigan, Ann Arbor Aug 2019 - Present

Ph.D. in Computer Science and Engineering, GPA: 4.0/4.0
Advisor: Prof. Scott Mahlke

Zhejiang University Aug 2015 - June 2019

B.Eng. in Electrical Engineering, GPA: 91.04/100, Rank: 1/121

GRADUATE COURSEWORK

Computer Architecture - Computer Architecture - EECS 470 (A+), Parallel Computer Architecture - EECS 570 (A), Advanced Compilers - EECS 583 (A+), Microarchitecture - EECS 573 (A)
Machine Learning - Machine Learning - EECS 545 (A), Matrix Methods for Machine Learning - EECS 551 (A)

WORK EXPERIENCE

Meta, Inc. May 2022- Aug 2022
Research Intern, Pytorch Compiler Mentor: Jason Ansel

· Enabled Torchinductor to codegen from triton template for Convolutional Layer and Fully-connected layer;
· Fused pointwise operations (like relu, batchnorm, etc.) to Convolutional and FC layer;
· Autotuning framework to choose the best kernel to run conv/mm among cudnn and triton, improved end-to-end performance on Torchbenchmark by 15%.

Twitter, Inc. June 2021- Aug 2021
Engineering Intern, TwitterVMTeam Manager: Y.S. Ramakrishna

· Integrated Super-Level Parallelism algorithm into Graal compiler (a Java JIT compiler) 20.3 version.
· Improved the cost model of Autovectorization, co-optimized the pre-Autovectorization phase to further speedup several benchmarks including scimark2, ionut-jvm-perforamnce and java-matrix-benchmark.

RESEARCH EXPERIENCE

Accurate and Fast Estimation of Temporal Motifs using Path Sampling Sept 2022 - Aug 2023
CSE Department, University of Michigan, Ann Arbor

· Designed an algorithm, TEACUPS, that addresses the temporal motif count problem using a novel technique of temporal path sampling.
· Proved that TEACUPS is an unbiased estimator with provable concentration behavior.
· Achieved one-to-four orders of magnitude speedup over current algorithms and exhibits near-linear runtime scalability.
· Achieved a relative error of less than 1% for most temporal connected 4-vertex motifs, with an error below 8% for all motifs evaluated

BITSET: An Energy Efficient Software-Hardware Co-design solution for Convolutional Neural Network Jan 2020 - Aug 2022
CSE Department, University of Michigan, Ann Arbor Advisor: Prof. Scott Mahlke

- Proposed an early termination technique to terminate computation of convolution operations early while maintaining the high accuracy of CNNs in a bit-serial manner.
- Designed a novel architecture that supports early activation termination operations, which also allows fully-variable weight bit-precision. It yeilds 1.74x speedup and 1.51x energy efficiency over a state-of-art CNN accelerator UNPU.

Distribution-driven neural network quantization without training Oct 2018 - May 2019
Institute of VLSI design, Zhejiang University, China Advisor: Dr. Kejie Huang

- Developed a novel hardware-friendly method to quantize the weight value of the neural network in the log domain based on the double-peak distribution of pruned CNN weights.
- Explored parameters to balance granular distortion and overload distortion and adjusted the precision of each layer according to their tolerance for precision loss.
- Explored a simple approach to incrementally quantized both data and weights and then compensated the precision loss by updating the values of the remaining weights.

COURSEWORK PROJECTS

4-way Superscalar R10K Out-of-Order Processor Fall 2019
Computer Architecture (EECS 470), Prof. Ronald G. Dreslinski

Designed and implemented a synthesizable four-way superscalar Out-of-Order processor in Verilog HDL with speculative LSQ, instruction prefetching and post-store retirement buffer, and developed graphical debugging tool.

Shadowclone: Thwarting and Detecting DOP Attacks with Stack Layout Randomization and Canary Fall 2019
Advanced Compilers (EECS 583), Prof. Scott Mahlke

Developed a compile time stack layout randomization scheme- Shadowclone -to thwart and detect DOP attacks effectively, and implemented with LLVM. Shadowclone generates randomized clones of vulnerable target functions and randomly selects one copy of clones to execute during runtime.

TEACHING EXPERIENCE

Advanced Compilers (EECS 583) Fall 2020, Fall 2021, Fall 2022
Computer Science and Engineering, University of Michigan

- Hold office hours for the compiler class of more than 100 students 3 times a week.
- Prepare LLVM homework template for compiler optimization, plan review slides, grade assignments, and come up with exams.

SKILLS

Programming Languages	C++, Python, Verilog, Bash
Tools & Frameworks	PyTorch, TensorFlow, LLVM, VCS, Graal

HONORS AND AWARDS

National Scholarship (rank 1/330)	*Oct 2016*
First-class Scholarship for Outstanding Students (top 3% of school)	*Nov 2016*
Zhejiang Provincial Government Scholarship (top 3% of school)	*Nov 2017*
Second-class Scholarship for Outstanding Students (top 10% of school)	*Nov 2017*
Wang Guosong Scholarships(The highest honor of the scholarship of EE Department)	*Oct 2018*