Po-An Tsai

January 2021

CONTACT poant@nvidia.com https://research.nvidia.com/person/po-an-tsai

EDUCATION

Massachusetts Institute of Technology
Ph.D. in Computer Science, minor in *Optimization Methods* **September 2013 – June 2019**
S.M. in Computer Science **June 2015**
Advisor: Professor Daniel Sanchez
GPA: 4.92/5.0 in Computer Architecture, Machine Learning, Computer Security, Geometric Algorithms

National Taiwan University
B.Sc. in Electrical Engineering **June 2012**
GPA: 3.96/4.0 in GPU Programming, VLSI Design, Algorithms, Data Structures, OS, Computer Networks

SKILLS AND TOOLS

Languages: C, C++, Python, Java, Verilog, Matlab, bash, SQL
Libraries: CUDA, OpenCL, matplotlib
Tools: Git, Intel Pin, Timeloop, Zsim, ModelSim, Altera Quantus II, IC Encounter

WORK EXPERIENCE

NVIDIA Research, Westford MA
Research Scientist **July 2019 – Current**

I develop architectures to address the emerging demands of computer vision and machine learning algorithms. This task requires understanding and analyzing the interplay between hardware, software, and algorithms. Specifically, I work on a flexible tensor accelerator that accelerates a wider range of tensor algorithms than conventional accelerators. To evaluate designed accelerators, I use and contribute to an open-source analytical modeling tool (Timeloop+Accelergy) for rapid evaluation of DNN accelerators.

I also collaborate with teams across the company, spanning software, research, engineering, and product groups, and publish original research and speak at conferences and events.

MIT Computer Science & Artificial Intelligence Lab, Cambridge MA
Research Assistant **September 2013 – June 2019**

My Ph.D. research focuses on reducing data movement in computer systems to improve their performance and energy efficiency. I designed new memory hierarchies, developed algorithms for data placement and workload scheduling, and co-designed hardware/software to optimize systems.

Across my projects, I prototyped ideas extending Zsim, a C++, Intel Pin-based open-sourced multicore simulator. I leveraged latest commodity hardware features (e.g., Intel CAT) and profiled workloads using hardware performance counters. I made essential changes throughout the software stack, including applications (e.g., key-value store, graph analytics) and runtime/compiler in Maxine, a Java-based research JVM.

VMware, Palo Alto CA
Ph.D. Intern **June 2015 – August 2015**

Distributed Resource Management Team. Worked on a VM scheduler that performs multi-dimensional resource balancing and traffic engineering. Proposed a randomized and graph-clustering-based algorithm and evaluated it using a trace-driven simulator written in Python. My algorithm reduces the runtime overhead by 10× while improving utilization by 5% and was publicly released in 2016 and filed as a US patent.

NTU-IBM Collaborative Project, Taiwan
Undergraduate Research Assistant **March 2012 – September 2012**

Built a primary simulation tool for FPGA acceleration. Designed a verification environment for FPGA-assisted medical image processing. Programmed FPGAs using both Verilog and OpenCL.

PATENT

Resource-Based Virtual Computing Instance Scheduling US 15283274
Po-An Tsai, Sahan Gamage, and Rean Griffith.

RECENT PUBLICATIONS

Mind Mappings: Enabling Efficient Algorithm-Accelerator Mapping Space Search
Kartik Hegde, **Po-An Tsai**, Sitao Huang, Vikas Chandram, Angshuman Parashar, and Christopher W. Fletcher. **ASPLOS-26**, April 2021.

Safecracker: Leaking Secrets through Compressed Caches
Po-An Tsai, Andres Sanchez, Christopher W. Fletcher, and Daniel Sanchez. **ASPLOS-25**, March 2020.