Jieming Yin

Personal Website: http://jiemingyin.github.io Email: Jieming.Yin@amd.com

EDUCATION

Ph.D in Computer Science

University of Minnesota, Twin Cities, 2009 – 2015

Thesis: *Time-Division-Multiplexing Based Hybrid-Switched NoC for Heterogeneous Multicore Systems*

B.Eng. in Electrical Engineering

Harbin Institute of Technology, China, 2004 – 2008

Thesis: *Point-to-point Scalar Operand Networks for Clustered Superscalar Processor*

COMPUTER SKILLS

Programming Languages	C/C++; Python, Perl; CUDA, OpenCL; VerilogHDL.
Operating Systems	LINUX, UNIX, and Windows.
Tools/Simulators	PyTorch, Tensorflow; Vivado; gem5, and gpgpu-sim Simulator; Cacti Power model; Pin Tools; R, Matlab; SQL.

EXPERIENCE

MTS Silicon Design Engineer **Dec. 2016–Present**

AMD Research (Bellevue, WA)

- Work on PathForward program–Department of Energy's Exascale Computing Project (ECP) for the development of the nation's first exascale supercomputers.
- SoC performance modeling and evaluation for in-package communication networks and memory networks.
- Optimize cache replacement/bypassing policies for high-performance CPUs.
- Utilize machine learning techniques to improve hardware design.

Postdoc Researcher **Mar. 2015–Dec. 2016**

AMD Research (Bellevue, WA)

- Work on FastForward 2 program–A jointly-funded collaboration between DOE Office of Science and the National Nuclear Security Administration (NNSA) focused on accelerating the research and development of critical technologies needed to enable exascale computing.
- Design novel in-package communication architecture that guarantees quality-of-service.
- Optimize coherence protocols for exascale heterogeneous processor that contains both CPUs and GPUs.

Summer Intern **May. 2013–Aug. 2013**

Qualcomm, Multimedia Architecture Team (San Diego, CA)

- Bring up a heterogeneous simulation platform containing CPU and GPU with shared memory space (SVM).
- Enable ARM and OpenCL support for running GPU applications on the simulation platform.
- Evaluate SVM efficiency and on-chip communication traffic caused by data transfer and coherence messages.

Academic Simulator Development

gem5-gpu: A heterogeneous CPU-GPU simulation infrastructure (https://gem5-gpu.cs.wisc.edu/wiki/**)**

- Contribute to support ARMv7 32-bit ISA.

Research Assistant in University of Minnesota, Twin Cities **Sep. 2009–Mar. 2015**

- Memory Consistency for Heterogeneous Multicore Systems.
- Co-design of Computation and Communication for High-throughput Data-parallel Accelerators.
- Energy Efficient Time-Division Multiplexed (TDM) Hybrid-Switched NoC.
- Integrated Layout and Architectural Design for Multicore Platforms.
- Parallelization of Single-Threaded Programs Using Speculative Slicing.

PATENT APPLICATION

[1] US Patent 10042774. *Method and apparatus for masking and transmitting data*. Granted August 7, 2018.

[2] US Patent 10097091. *Setting operating points for circuits in an integrated circuit chip*. Granted Oct 9, 2018.

[3] 170358-US-NP. *Reconfigurable prediction engine for general processor counting*. Filed March 15, 2018.

[4] 170377-US-NP. *Selecting a precision level for executing a workload in an electronic device*. Filed April 9, 2018.

[5] 180169-US-NP. *Architecture for deep Q-learning*. Filed Oct 30, 2018.

[6] 180283-US-NP. *A Mechanism for Dynamic Latency-Bandwidth Trade-off for Efficient Broadcasts/Multicasts*. Filed Dec 12, 2018.

[7] 180465-US-NP. *Reducing the impact of Speculative DRAM Reads through the use of confidence prediction*. Filed Feb 1, 2019.

SELECTED PUBLICATION

[1] Shuai Che, **Jieming Yin**. *Northup: Divide-and-Conquer Programming in Systems with Heterogeneous Memories and Processors*. IPDPS 2019.

[2] **Jieming Yin**, Zhifeng Lin, Onur Kayiran, Matthew Poremba, Muhammad Shoaib Bin Altaf, Natalie Enright Jerger, Gabriel H. Loh. *Modular Routing Design for Chiplet-based Systems*. ISCA 2018.

[3] **Jieming Yin**, Yasuko Eckert, Shuai Che, Mark Oskin, and Gabriel H. Loh. *Toward More Efficient NoC Arbitration: A Deep Reinforcement Learning Approach*. AIDArc 2018.

[4] Anthony Gutierrez, Bradford Beckmann, Alexandru Dutu, Joseph Gross, John Kalamatianos, Onur Kayiran, Michael LeBeane, Matthew Poremba, Brandon Potter, Sooraj Puthoor, Mark Wyse, **Jieming Yin**, Akshay Jain, Tim Rogers, Xianwei Zhang, Matt Sinclair. *Lost in Abstraction: Pitfalls of Analyzing GPUs at the Intermediate Language Level*. HPCA 2018.

[5] Matthew Poremba, Itir Akgun, **Jieming Yin**, Onur Kayiran, Yuan Xie, Gabriel H. Loh. *There and Back Again: Optimizing the Interconnect in Networks of Memory Cubes*. ISCA 2017.

[6] **Jieming Yin**, Onur Kayiran, Matthew Poremba, Natalie Enright Jerger, Gabriel H. Loh. *Efficient Synthetic Traffic Models for Large, Complex SoCs*. HPCA 2016.

[7] **Jieming Yin**, Pingqiang Zhou, Sachin S. Sapatnekar, Antonia Zhai. *Energy-Efficient Time-Division Multiplexed Hybrid-Switched NoC for Heterogeneous Multicore Systems*. IPDPS 2014.

[8] **Jieming Yin**, Pingqiang Zhou, Anup P. Holey, Sachin S. Sapatnekar, Antonia Zhai. *Energy-Efficient Non-Minimal Path On-chip Interconnection Network for Heterogeneous Multicore Systems*. ISLPED 2012.

[9] Pingqiang Zhou, **Jieming Yin**, Antonia Zhai, Sachin S. Sapatnekar. *NoC Frequency Scaling with Flexible-Pipeline Routers*. ISLPED 2011.