Jieming Yin

Curriculum Vitae

http://jiemingyin.github.io/

RESEARCH INTEREST

Computer architecture, heterogeneous multi-core architectures, interconnection networks, 2.5D/3D integration, near-data computing, machine learning, artificial intelligence

EDUCATION

2009 - 2015 **University of Minnesota, Twin Cities**.
Ph.D. in Computer Science
Dissertation: *Time-Division-Multiplexing Based Hybrid-Switched NoC for Heterogeneous Multi-core Systems*
Committee: Antonia Zhai, Sachin S. Sapatnekar, Pen-Chung Yew, Anand Tripathi

2004 - 2008 **Harbin Institute of Technology, China**.
B.Eng. in Electrical Engineering
Thesis: *Point-to-point Scalar Operand Networks for Clustered Superscalar Processor*

EMPLOYMENT

08/2020 - Present **Lehigh University**, Bethlehem, PA.
Assistant Professor

12/2016 - 08/2020 **AMD Research**, Bellevue, WA.
Member of Technical Staff Silicon Design Engineer

03/2015 - 12/2016 **AMD Research**, Bellevue, WA.
Postdoc Researcher

05/2013 - 08/2013 **Qualcomm Inc.**, San Diego, CA.
Summer Intern

09/2009 - 02/2015 **University of Minnesota, Twin Cities**, Minneapolis, MN.
Research Assistant

PUBLICATIONS

Conference Papers

[1] **Jieming Yin**, Antonia Zhai. *In-Network Memory Access Ordering for Heterogeneous Multicore Systems*. 14th IEEE/ACM International Symposium on Networks-on-Chip (NOCS), Virtual Conference, September, 2020. (**Best Paper Award**)

[2] Srikant Bharadwaj, **Jieming Yin**, Bradford M. Beckmann, Tushar Krishna. *Kite: A Family of Heterogeneous Interposer Topologies Enabled via Accurate Interconnect Modeling*. 57th Annual Design Automation Conference (DAC), Virtual Conference, July, 2020.

[3] **Jieming Yin**, Subhash Sethumurugan, Yasuko Eckert, Alan Smith, Chintan Patel, Eric Morton, Mark Oskin, Natalie Enright Jerger, Gabriel H. Loh. *Experiences with ML-Driven Design: A NoC Case Study*. 26th IEEE International Symposium on High-Performance Computer Architecture (HPCA), San Diego, CA, February 2020.

[4] Shuai Che, **Jieming Yin**. *Northup: Divide-and-Conquer Programming in Systems with Heterogeneous Memories and Processors*. 33rd IEEE International Parallel & Distributed Processing Symposium (IPDPS), Rio de Janeiro, Brazil, May 2019.

[5] **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*. 45th International Symposium on Computer Architecture (ISCA), Los Angeles, CA, June 2018. (**Featured in IEEE Spectrum**)

[6] 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*. 24th IEEE International Symposium on High-Performance Computer Architecture (HPCA), Vienna, Austria, February 2018.

[7] 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*. 44th International Symposium on Computer Architecture (ISCA), Toronto, CA, June 2017.

[8] **Jieming Yin**, Onur Kayiran, Matthew Poremba, Natalie Enright Jerger, Gabriel H. Loh. *Efficient Synthetic Traffic Models for Large, Complex SoCs*. 22nd International Symposium on High Performance Computer Architecture (HPCA), Barcelona, Spain, March 2016.

[9] **Jieming Yin**, Pingqiang Zhou, Sachin S. Sapatnekar, Antonia Zhai. *Energy-Efficient Time-Division Multiplexed Hybrid-Switched NoC for Heterogeneous Multicore Systems*. 28th IEEE International Parallel & Distributed Processing Symposium (IPDPS), Phoenix, Arizona, USA, May 2014.

[10] **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*. International Symposium on Low Power Electronics and Design (ISLPED), Redondo Beach, USA, August 2012.

[11] Pingqiang Zhou, **Jieming Yin**, Antonia Zhai, Sachin S. Sapatnekar. *NoC Frequency Scaling with Flexible-Pipeline Routers*. International Symposium on Low Power Electronics and Design (ISLPED), Fukuoka, Japan, August 2011.

Journal Article

[12] Xiangwei Cai, **Jieming Yin**, Pingqiang Zhou. *An orchestrated NoC prioritization mechanism for heterogeneous CPU-GPU systems*. Integration, Volume 65, March 2019.

Workshop Paper

[13] **Jieming Yin**, Yasuko Eckert, Shuai Che, Mark Oskin, and Gabriel H. Loh. *Toward More Efficient NoC Arbitration: A Deep Reinforcement Learning Approach*. The 1st International Workshop on AI-assisted Design for Architecture (AIDArc), Los Angeles, CA, June 2018.

Technical Report

[14] Pingqiang Zhou, **Jieming Yin**, Antonia Zhai, Sachin S. Sapatnekar. *NoC Design and Performance Optimization*. SRC TECHCON, 2011.

US Patents

[15] US Patent 10042774. Shuai Che, **Jieming Yin**. *Method and apparatus for masking and transmitting data*. Granted Aug. 2018.

[16] US Patent 10097091. Wei Huang, Yasuko Eckert, Xudong An, Muhammad Shoaib Bin Altaf, **Jieming Yin**. *Setting operating points for circuits in an integrated circuit chip*. Granted Oct. 2018.

[17] US Patent 10719441. **Jieming Yin**, Yasuko Eckert, Matthew Poremba, Steven Raasch, Doug Hunt. *Using Predictions of Outcomes of Cache Memory Access Requests for Contolling Whether A Request Generator Sends Memory Access Requests To A Memory In Parallel With Cache Memory Access Requests*. Granted Jul. 2020.

[18] US Patent App 15/922,875. Shuai Che, **Jieming Yin**. *Reconfigurable prediction engine for general processor counting*. Filed Mar. 2018.

[19] US Patent App 15/948,795. Tony Gutierrez, Sergey Blagodurov, Scott Moe, Xianwei Zhang, **Jieming Yin**, Matt Sinclair. *Selecting a precision level for executing a workload in an electronic device*. Filed Apr. 2018.

[20] US Patent App 16/176,903. Shuai Che, **Jieming Yin**. *Architecture for deep Q-learning*. Filed Oct. 2018.

[21] US Patent App 16/224,739. Mohamed Ibrahim, Onur Kayiran, Yasuko Eckert, **Jieming Yin**. *A Mechanism for Dynamic Latency-Bandwidth Trade-off for Efficient Broadcasts/Multicasts*. Filed Dec. 2018.

[22] 190256-US-NP. **Jieming Yin**, Subhash Sethumurugan, Yasuko Eckert. *Cache management based on reuse distance*. Filed Oct. 2019.

[23] 190350-US-NP. **Jieming Yin**, Subhash Sethumurugan, Yasuko Eckert. *Cache line re-reference interval prediction using physical page address*. Filed Dec. 2019.

[24] 190351-US-NP. **Jieming Yin**, Subhash Sethumurugan, Yasuko Eckert. *Cache management based on access type priority*. Filed Dec. 2019.

[25] 190066-US-NP. Onur Kayiran, **Jieming Yin**, Yasuko Eckert. *Look-ahead teleportation for reliable computing in multi-SIMD quantum processor*. Filed Feb. 2020.

PRESENTATIONS

09/2020 "In-Network Memory Access Ordering for Heterogeneous Multicore Systems." Presented at International Symposium on Networks-on-Chip (NOCS), Virtual Conference.

03/2020 "In-Package Interconnection Networks in the Era of Exascale and Beyond."

- Virginia Tech, Blacksburg, VA.
- Illinois Institute of Technology, Chicago, IL.
- Lehigh University, Bethlehem, PA.
- College of William and Mary, Williamsburg, VA.

02/2020 "Experiences with ML-Driven Design: A NoC Case Study." Presented at International Symposium on High-Performance Computer Architecture (HPCA), San Diego, CA.

04/2019 "Exploiting Machine Learning Insights for NoC Design." Presented to the Department of Energy (DOE), Bellevue, WA.

09/2018 "Better NoCs through Machine Learning." Presented to the Department of Energy (DOE), Austin, TX.

06/2018 "Modular Routing Design for Chiplet-based Systems." Presented at International Symposium on Computer Architecture (ISCA), Los Angeles, CA.

06/2018 "Toward More Efficient NoC Arbitration: A Deep Reinforcement Learning Approach." Presented at International Workshop on AI-assisted Design for Architecture (AIDArc), Los Angeles, CA.

02/2015 "Designing Energy-Efficient NoCs for Heterogeneous Multicore Systems." Invited talk. Presented at Nvidia, Santa Clara, CA.

05/2014 "Energy-Efficient Time-Division Multiplexed Hybrid-Switched NoC for Heterogeneous Multicore Systems." Presented at 28th IEEE International Parallel & Distributed Processing Symposium (IPDPS), Phoenix, AZ.

07/2012 "Energy Efficient Non-Minimal Path On-chip Interconnection Network for Heterogeneous Multicore Systems." Presented at International Symposium on Low Power Electronics and Design (ISLPED), Redondo Beach, CA.

08/2011 "NoC Frequency Scaling with Flexible-Pipeline Routers." Presented at International Symposium on Low Power Electronics and Design (ISLPED), Fukuoka, Japan.

FUNDING

01/2017 - 12/2019 **Pathforward Program**, *Department of Energy (DOE)*.
Contributed to proposal and deliverable report writing. Working on Component and Memory Integration work package, one of the largest work packages in Pathforward.

01/2015 - 12/2016 **Fastforward 2 Program**, *Department of Energy (DOE)*, Total amount $\$32,000,000$.
Contributed to deliverable report writing. Worked on Component Integration project.

RESEARCH EXPERIENCE

12/2016 - 08/2020 **AMD Research**, Member of Technical Staff Silicon Design Engineer.
- 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, propose new system topologies and routing algorithms.
- Utilize machine learning techniques to speed up design process and improve hardware design.

03/2015 - 11/2016 **AMD Research**, Postdoc Researcher.
- 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 communication architecture that guarantees quality-of-service.
- Optimize coherence protocols for exascale heterogeneous processors containing CPUs and GPUs.

05/2013 - 08/2013 **Qualcomm Inc.**, Summer Intern.
- Bring up a simulation platform containing CPU and GPU with shared memory space.
- 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.

08/2009 - 02/2015 **University of Minnesota, Twin Cities**, Research Assistant.
- NoC-Enhanced 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.

PROFESSIONAL ACTIVITIES

Committee
- ERC, International Conference on Parallel Architectures and Compilation Techniques (PACT), 2020
- TPC, IEEE/ACM International Symposium on Networks-on-Chip (NOCS), 2020
- TPC, IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2020
- TPC, Workshop on General Purpose Processing Using GPU (GPGPU), 2020
- TPC, Design Automation Conference (DAC), 2020
- TPC, Design Automation Conference (DAC), 2019

- TPC, IEEE/ACM International Symposium on Networks-on-Chip (NOCS), 2019
- Finance Chair, IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2019
- TPC, IEEE International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2019
- Submissions Chair, IEEE International Symposium on Workload Characterization (IISWC), 2017
- TPC, IEEE International Symposium on Workload Characterization (IISWC), 2017
- TPC, IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC), 2016

Editor
- Guest Editor, Journal of Signal Processing Systems special issue on Application-specific Systems, Architectures and Processors (JSPS)

Journal Reviewer
- ACM Transactions on Architecture and Code Optimization (TACO)
- IEEE Transactions on Computers (TC)
- IEEE Transactions on Computers Special Issue on "Communications for Many-Core Processors and Accelerators"
- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- IEEE Computer Architecture Letters (CAL)
- IEEE Transactions on Very Large Scale Integration Systems(TVLSI)
- Journal of Systems Architecture (JAS)
- EURASIP Journal on Embedded Systems

ADVISING AND TEACHING EXPERIENCE

Student Mentoring
- Subhash Sethumurugan, PhD student, University of Minnesota, Twin Cities
  Project: Exploiting Machine Learning Insights for Cache Replacement Policy
- Zhifeng Lin, PhD student, University of Southern California
  Project: Exploring QoS in Interposer-based Systems
- Xiangwei Cai, Master student, ShanghaiTech University
  Project: NoC Prioritization Mechanism for Heterogeneous CPU-GPU Systems

Teaching
- Fall 2020, ECE 350/450-014, Interconnection Networks forMany-core Architectures, Lehigh University
- Lab Lecturer, TA, Introduction to Operating Systems, University of Minnesota
- Lecturer, TA, Introduction to Compilers, University of Minnesota
- TA, Operating Systems, University of Minnesota
- TA, Advanced Computer Architecture, University of Minnesota
- Guest Lecturer, Advanced Topics in Computer Architecture, University of Minnesota

PRESS COVERAGE

06/2018 AMD Tackles Coming "Chiplet" Revolution With New Chip Network Scheme, **IEEE Spectrum**