YAKUN SOPHIA SHAO

Assistant Professor
EECS, UC Berkeley
Cory 570, Berkeley CA, 94720

ysshao@eecs.berkeley.edu
+1-617-955-6894
https://people.eecs.berkeley.edu/~ysshao/

RESEARCH INTERESTS

Domain-Specific Systems, Machine Learning Accelerators, Design Methodology, Hardware Prototyping

EDUCATION

2009-2016 **Harvard University**, Cambridge, MA
Ph.D. in Computer Science.
Advisors: Professors David Brooks and Gu-Yeon Wei

2009-2014 **Harvard University**, Cambridge, MA
Master of Science in Computer Science.

2005-2009 **Zhejiang University**, Hangzhou, Zhejiang, China
Bachelor of Electrical Engineering

PROFESSIONAL EXPERIENCE

2019-Present **Assistant Professor**
EECS Department, UC Berkeley, Berkeley, CA

2018-2019 **Senior Research Scientist**
NVIDIA Research, Santa Clara, CA

2016-2018 **Research Scientist**
NVIDIA Research, Santa Clara, CA

2009-2016 **Research Assistant**
Harvard University, Cambridge, MA

Summer 2015 **Research Intern**
IBM T.J. Watson Research Center, Yorktown Heights, NY

Summer 2014 **Research Intern**
IBM T.J. Watson Research Center, Yorktown Heights, NY

Summer 2012 **Research Intern**
Intel Labs, Santa Clara, CA

AWARDS AND HONORS

2017 ACM Doctoral Dissertation Award Harvard Nominee

2015-2016 IBM Ph.D. Fellowship

2015 Siebel Scholar, Class of 2015 - awarded annually for academic excellence and demonstrated leadership to 85 top students from the world's leading graduate schools

2014 IEEE Micro's Top Picks in Computer Architecture

2014 Best in Session Award, SRC TECHCON

2014 Rising Stars in EECS Workshop Invited Participant

REFEREED PUBLICATIONS

2019 **MAGNet: A Modular Accelerator Generator for Neural Networks**
Rangharajan Venkatesan, Yakun Sophia Shao, Miaorong Wang, Jason Clemons, Steve Dai, Matthew Fojtik, Ben Keller, Alicia Klinefilter, Nathaniel Pinckney, Yanqing Zhang, Brian Zimmer, William J. Dally, Joel S. Emer, Stephen W. Keckler, Brucek Khailany
International Conference on Computer Aided Design (ICCAD), November 2019

Simba: Scaling Deep-Learning Inference with Multi-Chip-Module-Based Architecture
Yakun Sophia Shao, Jason Clemons, Rangharajan Venkatesan, Brian Zimmer, Matthew Fojtik, Ted Jiang, Ben Keller, Alicia Klinefelter, Nathaniel Pinckney, Priyanka Raina, Stephen G Tell, Yanqing Zhang, William J. Dally, Joel S. Emer, C. Thomas Gray, Brucek Khailany, Stephen W. Keckler
International Symposium on Microarchitecture (MICRO), October 2019

A 0.11pJ/Op, 0.32-128 TOPS, Scalable Multi-Chip-Module-based Deep Neural Network Accelerator with Ground-Reference Signaling in 16nm
Brain Zimmer, Rangharajan Venkatesan, Yakun Sophia Shao, Jason Clemons, Matthew Fojtik, Nan Jiang, Ben Keller, Alicia Klinefilter, Nathaniel Pinckney, Priyanka Raina, Stephen G. Tell, Yanqing Zhang, William J. Dally, Joel S. Emer, C. Thomas Gray, Stephen W. Keckler, Brucek Khailany
VLSI Symposium on Circuits, June 2019

SNAP: A 1.67-21.55 TOPS/W Sparse Neural Acceleration Processor for Unstructured Sparse Deep Neural Network Inference
Jie-Fang Zhang, Ching-En Lee, Chester Liu, Yakun Sophia Shao, Stephen W. Keckler, Zhengya Zhang
VLSI Symposium on Circuits, June 2019

Buffets: An Efficient and Composable Storage Idiom for Explicit Decoupled Data Orchestration
Michael Pellauer, Yakun Sophia Shao, Jason Clemons, Neal Crago, Kartik Hegde, Rangharajan Venkatesan, Stephen W. Keckler, Christopher W. Fletcher, Joel Emer
International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), April 2019.

Timeloop: A Systematic Approach to DNN Accelerator Evaluation
Angshuman Parashar, Priyanka Raina, Yakun Sophia Shao, Yu-Hsin Chen, Victor A. Ying, Anurag Mukkara, Rangharajan Venkatesan, Brucek Khailany, Stephen W. Keckler, Joel Emer
International Symposium on Performance Analysis of Systems and Software (ISPASS), March 2019.

2018 **A Modular Digital VLSI Flow for High-Productivity SoC Design**
Brucek Khailany, Matthew Fojtik, Alicia Klinefelter, Evgeni Krimer, Michael Pellauer, Nathaniel Pinckney, Haoxing Ren, Yakun Sophia Shao, Rangharajan Venkatesan, Yanqing Zhang, Brian Zimmer
Design Automation Conference (DAC), March 2018.

Stitch-X: An Accelerator Architecture for Exploiting Unstructured Sparsity in Deep Neural Networks
Ching-En Lee, Yakun Sophia Shao, Jie-Fang Zhang, Angshuman Parashar, Joel Emer, Stephen W. Keckler, Zhengya Zhang
SysML Conference, February 2018.

Assisting High-Level Synthesis Improve SpMV Benchmark Through Dynamic Dependence Analysis
Rafael Garibotti, Brandon Reagen, Yakun Sophia Shao, Gu-Yeon Wei and David Brooks
IEEE Transactions on Circuits and Systems II: Express Briefs, 2018.

2017 **Using Dynamic Dependence Analysis to Improve the Quality of High-Level Synthesis Designs**
Rafael Garibotti, Brandon Reagen, Yakun Sophia Shao, Gu-Yeon Wei and David Brooks
International Symposium on Circuits and Systems (ISCAS), May 2017.

2016 **Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin**
Yakun Sophia Shao, Sam Xi, Viji Srinivasan, Gu-Yeon Wei and David Brooks
International Symposium on Microarchitecture (MICRO), October 2016.

2015 **Toward Cache-Friendly Hardware Accelerators**
Yakun Sophia Shao, Sam Xi, Viji Srinivasan, Gu-Yeon Wei and David Brooks
HPCA Sensors and Cloud Architectures Workshop (SCAW), Feb 2015.

The Aladdin Approach to Accelerator Design and Modeling
Yakun Sophia Shao, Brandon Reagen, Gu-Yeon Wei and David Brooks
IEEE Micro, May-June 2015.

2014 **MachSuite: Benchmarks for Accelerator Design and Customized Architectures**
Brandon Reagen, Robert Adolf, Yakun Sophia Shao, Gu-Yeon Wei and David Brooks
International Symposium on Workload Characterization (IISWC), Oct 2014.

Aladdin: A Pre-RTL, Power-Performance Accelerator Simulator Enabling Large Design Space Exploration of Customized Architectures
Yakun Sophia Shao, Brandon Reagen, Gu-Yeon Wei and David Brooks
International Symposium on Computer Architecture (ISCA), June 2014.
Selected as one of the Top Picks in Computer Architecture in 2014
Nominated by SIGARCH for publication in the Communications of the ACM (CACM)

2013 **Energy Characterization and Instruction-Level Energy Model of Intel's Xeon Phi Processor**
Yakun Sophia Shao and David Brooks
International Symposium on Low Power Electronics and Design (ISLPED), Sept 2013.

Quantifying Acceleration: Power/Performance Trade-offs of Application Kernels in Hardware
Brandon Reagen, Yakun Sophia Shao, Gu-Yeon Wei and David Brooks
International Symposium on Low Power Electronics and Design (ISLPED), Sept 2013.

2013 **ISA-Independent Workload Characterization and its Implications for Specialized Architectures**
Yakun Sophia Shao and David Brooks
International Symposium on Performance Analysis of Systems and Software (ISPASS), April 2013.

2010 **Power, Performance and Portability: System Design Considerations for Micro Air Vehicle Applications**
Yakun Sophia Shao, Judson Porter, Michael J. Lyons, Gu-Yeon Wei and David Brooks
Advanced Computer Architecture and Compilation for Embedded Systems (ACACES), July 2010

DISSERTATION AND BOOK

2016 **Design and Modeling of Specialized Architectures**
Yakun Sophia Shao
Ph.D. Dissertation, Harvard University, May 2016.

2015 **Research Infrastructures for Hardware Accelerators**

Yakun Sophia Shao and David Brooks
Synthesis Lectures on Computer Architecture, Morgan & Claypool Publishers, Nov 2015.

TUTORIALS AND SPECIAL CLASSES

2016 **Rapid Exploration of Accelerator-Rich Architectures: Automation from Concept to Prototyping,**
Yakun Sophia Shao with David Brooks, Jason Cong, Zhenman Fang, Gu-Yeon Wei and Sam Xi
International Symposium on Microarchitecture (MICRO), October 2016.

2016 **Aladdin and gem5-Aladdin: Research Infrastructures for Specialized Architectures,**
Yakun Sophia Shao with David Brooks, Gu-Yeon Wei and Sam Xi
International Symposium on Workload Characterization (IISWC), September 2016.

2015 **Rapid Exploration of Accelerator-Rich Architectures: Automation from Concept to Prototyping,**
Yakun Sophia Shao with David Brooks, Yu-Ting Chen, Jason Cong, Zhenman Fang, Brandon Reagen, Glenn Reinman, Gu-Yeon Wei and Sam Xi
International Symposium on Computer Architecture (ISCA), June 2015.

2015 **Research Infrastructures for Accelerator-centric Architectures,**
Yakun Sophia Shao with David Brooks, Mark Hempstead, Brandon Reagen and Gu-Yeon Wei
International Symposium on High Performance Computer Architecture (HPCA), Feb 2015.

2014 **Research Infrastructures for Accelerator-centric Architectures,**
Yakun Sophia Shao with David Brooks, Brandon Reagen, Kevin Skadron, Liang Wang, and Gu-Yeon Wei
International Symposium on Computer Architecture (ISCA), June 2014.

PROFESSIONAL SERVICE

Editing

2018 Guest Editor, IEEE Micro Special Issue on Hardware Acceleration, November/December 2018

Technical Program Committees

2020 Systems and Machine Learning (SysML)
International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)

2019 International Symposium on Microarchitecture (MICRO)
The First Young Architect Workshop (YArch)
International Symposium on High-Performance Computer Architecture (HPCA) Industry Session
International Symposium on High-Performance Computer Architecture (HPCA) External

2018 International Symposium on Microarchitecture (MICRO)
Design Automation Conference (DAC)
International Symposium on High-Performance Computer Architecture (HPCA) Industry Session
International Symposium on High-Performance Computer Architecture (HPCA) External
International Conference on Supercomputing (ICS) External

2017 International Symposium on Computer Architecture (ISCA)

Organizing Committees

2020 Finance Chair of International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)
2019 SIGARCH Visioning Workshop on Agile and Open Hardware Design for Next-Generation Computing
Area Chair of International Conference on Artificial Intelligence Circuits and Systems (AICAS)
Tutorial Chair of International Symposium on Computer Architecture (ISCA)
2018 Registration Chair of International Symposium on Workload Characterization (IISWC)
2017- Web Director of ACM Special Interest Group on Microarchitecture (SIGMICRO)
Web Co-Chair of Women in Computer Architecture (WICARCH)

Journal Review

ACM Journal on Emerging Technologies in Computing Systems
ACM Transactions on Architecture and Code Optimization
GTC Poster Reviewer for AI Application Deployment/Inference
Communications of the ACM (CACM)
IEEE Computer Architecture Letters
IEEE Micro
IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems
IEEE Transactions on Computers

TEACHING EXPERIENCE

2013 CS247r Advanced Topics in Computer Architecture, Teaching Fellow
2013 CS246 Advanced Computer Architecture, Teaching Fellow
2011 CS141 Computing Hardware, Teaching Fellow

OPEN-SOURCE SOFTWARE

Aladdin: A pre-RTL, power-performance-area simulator for fixed-function accelerators. [GitHub]
gem5-Aladdin: An SoC simulator. [GitHub] [Users Group]
LLVM-Tracer: An LLVM optimization pass to print a dynamic LLVM IR trace. [GitHub]
MachSuite: A benchmark suite for accelerators. [GitHub]
MatchLib: A SystemC/C++ library of commonly-used hardware components for HLS. [GitHub]
Timeloop: A design space exploration tool for DNN accelerators. [GitHub]
WIICA: An ISA-independent workload characterization to for accelerators. [GitHub]