

# YAKUN SOPHIA SHAO

---

Assistant Professor

EECS, UC Berkeley

Cory 570, Berkeley CA, 94720

Email: [ysshao@berkeley.edu](mailto:ysshao@berkeley.edu)

Website: <https://people.eecs.berkeley.edu/~ysshao/>

## RESEARCH INTERESTS

---

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

## EDUCATION

---

2009-2016 **Harvard University**, Cambridge, MA  
Ph.D. in Computer Science.

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**  
University of California, 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

---

- 2021 Dr. Sudhakar Yalamanchili Award for Contribution to Modeling and Simulation in Computer Architecture
- 2021 Paper selected as a DAC Best Paper Candidate
- 2020 Google Research Recognition for Technical Leadership and Achievements in Systems Research
- 2020 SK Hynix Faculty Fellow
- 2020 Facebook Research Award
- 2020 Paper selected as a CACM Research Highlight (Nominated by ACM SIGMICRO)
- 2020 Two papers selected as the IEEE Micro Top Picks in Computer Architecture Honorable Mentions
- 2019 Best Paper Award, International Symposium on Microarchitecture (MICRO)
- 2017 ACM Doctoral Dissertation Award Harvard Nominee
- 2015 IBM Ph.D. Fellowship
- 2015 Siebel Scholar
- 2014 IEEE Micro's Top Picks in Computer Architecture
- 2014 Best in Session Award, SRC TECHCON
- 2014 Rising Stars in EECS Workshop Invited Participant

## PUBLICATIONS

---

- 2021 **Gemmini: Enabling Systematic Deep-Learning Architecture Evaluation via Full-Stack Integration**  
Hasan Genc, Seah Kim, Alon Amid, Ameer Haj-Ali, Vighnesh Iyer, Pranav Prakash, Jerry Zhao, Daniel Grubb, Harrison Liew, Howard Mao, Albert Ou, Colin Schmidt, Samuel Steffl, John Wright, Ion Stoica, Jonathan Ragen-Kelley, Krste Asanovic, Borivoje Nikolic, Yakun Sophia Shao  
*Design Automation Conference (DAC), December 2021*  
☆ **DAC Best Paper Candidate**
- A 16mm<sup>2</sup> 106.1 GOPS/W Heterogeneous RISC-V Multi-Core Multi-Accelerator SoC in Low-Power 22nm FinFET**  
Abraham Gonzalez, Jerry Zhao, Ben Korpan, Hasan Genc, Colin Schmidt, John Wright, Ayan Biswas, Alon Amid, Farhana Sheikh, Anton Sorokin, Sirisha Kale, Mani Yalamanchi, Ramya Yarlagadda, Mark Flannigan, Larry Abramowitz, Elad Alon, Yakun Sophia Shao, Krste Asanovic, and Bora Nikolic  
*IEEE European Solid-State Circuit Conference (ESSCIRC), September 2021*
- CoSA: Scheduling by Constrained Optimization for Spatial Accelerators**  
Qijing Huang, Minwoo Kang, Grace Dinh, Thomas Norell, Aravind Kalaiah, James Demmel, John Wawrzynek, Yakun Sophia Shao  
*International Symposium on Computer Architecture (ISCA), June 2021*
- Simba: Scaling Deep-Learning Inference with Chiplet-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  
*Communications of the ACM (CACM), June 2021*
- Vertically Integrated Computing Labs Using Open-Source Hardware Generators and Cloud-Hosted FPGAs**  
Alon Amid, Albert Ou, Krste Asanovic, Yakun Sophia Shao, Borivoje Nikolic

*IEEE International Symposium on Circuits and Systems (ISCAS), May 2021*

**Memory-Efficient Hardware Performance Counters with Approximate-Counting Algorithms**

Jingyi Xu, Sehoon Kim, Borivoje Nikolic, Yakun Sophia Shao

*IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), March 2021*

**SNAP: An Efficient 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

*IEEE Journal of Solid-State Circuits (JSSC), February 2021*

2020 **Chipyard: Integrated Design, Simulation, and Implementation Framework for Custom SoCs**

Alon Amid, David Biancolin, Abraham Gonzalez, Daniel Grubb, Sagar Karandikar, Harrison Liew, Albert Magyar, Howard Mao, Albert Ou, Nathan Pemberton, Paul Rigge, Colin Schmidt, John Wright, Jerry Zhao, Yakun Sophia Shao, Krste Asanovic, Borivoje Nikolic

*IEEE Micro Special Issue on Agile and Open-Source Hardware, July/August 2020*

**NeuroVectorizer: End-to-End Vectorization with Deep Reinforcement Learning**

Ameer Haj-Ali, Nesreen K. Ahmed, Ted Willke, Yakun Sophia Shao, Krste Asanovic, Ion Stoica

*International Symposium on Code Generation and Optimization (CGO), February 2020*

**A 0.32-128 TOPS, Scalable Multi-Chip-Module-based Deep Neural Network Inference Accelerator with Ground-Referenced Signaling in 16nm**

Brian 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

*IEEE Journal of Solid-State Circuits (JSSC), Jan 2020*

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 Klinefilter, 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*

☆ **MICRO Best Paper Award**

☆ **Top Picks in Computer Architecture Honorable Mentions in 2019**

☆ **Research Highlight in Communications of the ACM (CACM)**

**A 0.11pJ/Op, 0.32-128 TOPS, Scalable Multi-Chip-Module-based Deep Neural Network Accelerator with Ground-Referenced 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*

**Buffers: 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*

☆ **Top Picks in Computer Architecture Honorable Mentions in 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, 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, 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, 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, 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, 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, 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, David Brooks

*International Symposium on Computer Architecture (ISCA), June 2014*

☆ **Top Picks in Computer Architecture in 2014**

2013 **Energy Characterization and Instruction-Level Energy Model of Intel's Xeon Phi Processor**

Yakun Sophia Shao, 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, 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, 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, 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, David Brooks

*Synthesis Lectures on Computer Architecture*, Morgan & Claypool Publishers, Nov 2015.

## PATENTS

---

**Efficient Neural Network Accelerator Dataflows**

*US Patent App. 16/672,918*, Filed Nov 2019.

**Scalable Multi-Die Deep Learning System**

*US Patent App. 16/517,431*, Filed July 2019.

**Deep Neural Network Accelerator with Fine-Grained Parallelism Discovery**

*US Patent App. 15/929,093*, Filed Jan 2019.

## OPEN-SOURCE SOFTWARE

---

**Aladdin**: A pre-RTL, power-performance-area simulator for fixed-function accelerators. [[GitHub](#)]

**Chipyard**: An integrated design, simulation, and implementation framework for custom SoCs. [[GitHub](#)]

**CoSA**: A constrained-optimization-based scheduler for spatial accelerators. [[GitHub](#)]

**gem5-Aladdin:** An SoC simulator. [\[GitHub\]](#) [\[Users Group\]](#)  
**Gemmini:** A systolic array generator for deep-learning architecture. [\[GitHub\]](#)  
**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. [\[GitHub\]](#)  
**ONNXRuntime-RISCV:** ONNXRuntime support for RISCV-based Accelerators. [\[GitHub\]](#)  
**Timeloop:** A design space exploration tool for DNN accelerators. [\[GitHub\]](#)  
**WIICA:** An ISA-independent workload characterization to for accelerators. [\[GitHub\]](#)

## TEACHING EXPERIENCE

---

- 2021 **EE290-2 Hardware for Machine Learning (Spring)**, Instructor, UC Berkeley  
Enrollment: 21  
Course Evaluation: 5.4/7  
Instructor Evaluation: 6.5/7
- 2020 **EECS151/251A Introduction to Digital Design and Integrated Circuits (Fall)**, Instructor, UC Berkeley  
Enrollment: 79  
Course Evaluation: 5.6/7 (151), 6.0/7 (251A)  
Instructor Evaluation: 5.2/7 (151), 6.0/7 (251A)
- EE290-2 Hardware for Machine Learning (Spring)**, Instructor, UC Berkeley  
Enrollment: 39  
Course Evaluation: 6.0/7  
Instructor Evaluation: 6.3/7
- 2019 **EECS151/251A Introduction to Digital Design and Integrated Circuits (Fall)**, Instructor, UC Berkeley  
Enrollment: 72  
Course Evaluation: 6.3/7 (151), 5.8/7 (251A)  
Instructor Evaluation: 6.2/7 (151), 5.8/7 (251A)
- 2013 **CS247r Advanced Topics in Computer Architecture**, Teaching Fellow, Harvard University  
2013 **CS246 Advanced Computer Architecture**, Teaching Fellow, Harvard University  
2011 **CS141 Computing Hardware**, Teaching Fellow, Harvard University

## TUTORIALS AND SPECIAL CLASSES

---

- 2021 **Gemmini: Enabling Systematic Deep-Learning Architecture Evaluation via Full-Stack Integration**,  
Yakun Sophia Shao with Hasan Genc, Seah Kim, Dima Nikiforov, Krste Asanovic, and Bora Nikolic  
*International Symposium on Workload Characterization (IISWC)*, November 2021.
- 2021 **Next-Generation Deep-Learning Accelerators: From Hardware to System**,  
Yakun Sophia Shao  
*VLSI Symposia on Technology and Circuits (VLSI)*, June 2021.
- 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.

## ADVISING

---

### Ph.D. Students

Prashanth Ganesh, Sep 2021-  
Roger Hsiao, Sep 2021-  
Hansung Kim, Sep 2020-  
Seah Kim (with Nikolic), Jan 2020-  
Vadim Nikiforov (with Nikolic), Sep 2020-  
Jingyi Xu (with Nikolic), Sep 2020-

### Ph.D. Student Collaborators

Alon Amid (advisor A. Asanovic and B. Nikolic), Jan 2021-  
Hasan Genc (advisor K. Asanovic), Jan 2020-  
Qijing (Jenny) Huang (advisor J. Wawrzynek), Sep 2019-

### 5-th Year M.S. and Undergraduate Students

Kareem Ahmad, 2020-2021  
Charles Hong (Apple Masters Scholarship), 2020-  
Thomas Norell (2nd Place in CGO'2020 Student Research Competition), 2019-2020  
Avinash Nandakumar (Apple Masters Scholarship), 2019-  
Patrick Wang (1st Place in PACT'2020 Student Research Competition), 2019-2020  
Jingyi Xu (→ Ph.D. Student at UC Berkeley), 2019-2020

### Ph.D. Qual Committee

2021 Nathan Pemberton (advisor R. Katz)  
2020 Benyuanyi Liu (advisor A. Niknejad)

Adelson Chua (advisor R. Muller)  
Qijing (Jenny) Huang (advisor J. Wawrzynek)  
Ameer Haj Ali (advisors I. Stoica and K. Asanovic)

### **Ph.D. Dissertation Committee**

2021 Qijing (Jenny) Huang (advisor J. Wawrzynek)

### **PROFESSIONAL SERVICE**

---

#### **Editing**

2020 Guest Editor, IEEE Micro Special Issue on Commercial Products, November/December 2020  
2018 Guest Editor, IEEE Micro Special Issue on Hardware Acceleration, November/December 2018

#### **Program Committees**

2022 International Solid-State Circuits Conference (ISSCC)  
Machine Learning and Systems (MLSys)  
International Symposium on Computer Architecture (ISCA)  
2021 International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)  
Machine Learning and Systems (MLSys)  
International Symposium on Computer Architecture (ISCA)  
Design Automation Conference (DAC)  
IEEE Micro Top Picks in Computer Architecture  
2020 Symposium on High Performance Chips (Hot Chips)  
International Symposium on Computer Architecture (ISCA)  
Machine Learning and Systems (MLSys)  
International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)  
2019 International Symposium on Microarchitecture (MICRO)  
2018 International Symposium on Microarchitecture (MICRO)  
Design Automation Conference (DAC)  
2017 International Symposium on Computer Architecture (ISCA)

#### **External/Session Program Committees**

2022 International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) External  
International Symposium on High-Performance Computer Architecture (HPCA) External  
2021 International Symposium on Microarchitecture (MICRO) External  
2020 International Symposium on Microarchitecture (MICRO) External  
International Symposium on Computer Architecture (ISCA) Industry Session  
2019 International Symposium on High-Performance Computer Architecture (HPCA) External  
International Symposium on High-Performance Computer Architecture (HPCA) Industry Session

- International Symposium on Computer Architecture (ISCA) External  
The First Young Architect Workshop (YArch)
- 2018 International Symposium on High-Performance Computer Architecture (HPCA) External  
International Symposium on High-Performance Computer Architecture (HPCA) Industry Session

### **Journal Review**

IEEE Journal of Solid State Circuits  
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

### **Grant Review**

- 2021 Department of Energy (DOE) Proposal  
2020 Natural Sciences and Engineering Research Council of Canada (NSERC) Proposal  
National Science Foundation (NSF) Proposal

### **Organizing Committees**

- 2022 Co-Chair of the Next-Generation Circuit Designer Workshop at ISSCC  
2020 Co-Organizer of the Rising Stars in EECS  
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 Co-Chair of Women in Computer Architecture (WICARCH)  
Web Director of ACM Special Interest Group on Microarchitecture (SIGMICRO)

### **UC Berkeley Service**

- 2019- Graduate Student Admissions Committee, EECS  
Future Direction Committee, EECS