Yuhao Zhu

The University of Texas at Austin
Electrical and Computer Engineering
201 E 24th St, Austin, TX 78712

+1 (512) 672-9292
yzhu@utexas.edu
http://yuhaozhu.com/

RESEARCH INTERESTS

I am interested in designing and building better hardware and software systems to make next-generation edge and cloud computing fast, energy-efficient, intelligent, and safe.

In my dissertation work, I have designed and prototyped systems that address the fundamental trade-off between energy-efficiency and user QoS experience in mobile Web computing via architecture, runtime, and programming language co-design. (Project website)

My recent research interests also include safe and efficient machine learning on edge devices and complementing machine and human intelligence in a systematic, programmable way.

EDUCATION

The University of Texas at Austin *2010 - present*
Ph.D. candidate (post defense), Electrical and Computer Engineering
Advisor: Vijay Janapa Reddi

Beihang University, Beijing, China *2006 - 2010*
B.S., Computer Science and Engineering
Advisor: Yangdong Deng

HONORS & RECOGNITIONS

- Google Ph.D. Fellowship, 2016
- IEEE Micro Top Picks of Computer Architecture (Honorable Mention), 2016
- Qualcomm Innovation Fellowship Finalist, 2015
- Best of Computer Architecture Letters Award, 2014
- UT Austin Microelectronics and Computer Development Fellowship, 2011-2012
- Outstanding Undergraduate Thesis Award, 2010

PROFESSIONAL POSITIONS

Harvard University
- Research Fellow — Aug 2016 - Present

The University of Texas at Austin
- Graduate Researcher — Aug 2010 - Present
- Teaching Assistant — Fall 2010, Spring 2011, Spring 2014

Tsinghua University
- Undergraduate Researcher — June 2009 - May 2010

Google Inc.
- Software Engineering Intern (Lead of Flow API of the Catapult project) — Summer 2015

AMD Research Lab
- Research Intern — Summer 2012, Summer 2013

STMicroelectronics
- Co-op Engineer — Summer 2011

SELECTED PRESS COVERAGE

- Smarter Systems Make Web Surfing on Your Phone Less of a Battery Drain, IEEE Spectrum
- Engineers Have a Plan to Make the Web More Energy Efficient, eeDesignIt
- New 'GreenWeb' Tools Aim to Create an Energy-efficient Web, Science Daily

PUBLICATIONS

Invited Articles

- Yuhao Zhu, Vijay Janapa Reddi, Robert Adolf, Saketh Rama, Brandon Reagen, Gu-Yeon Wei, David M. Brooks
  Cognitive Computing Safety: The New Horizon for Reliability / The Design and Evolution of Deep Learning Workloads
  IEEE Micro Special Issue on Cognitive Architectures, Jan/Feb 2017, 37(1):15-21

- Peter Bailis, Jean Yang, Vijay Janapa Reddi, Yuhao Zhu
  Research for Practice: Web Security and Mobile Web Computing (Website)
  Communications of the ACM (CACM), Jan 2017, 60(1): 50-53
  Also appears in *ACM Queue*, July/Aug 2016, 14(4):80-95.

Journal Articles

- Yuhao Zhu, Vijay Janapa Reddi
  Optimizing General-Purpose CPUs for Energy-Efficient Mobile Web Computing
  ACM Transactions on Computer Systems (TOCS), March 2017, 35(1):1

- Yuhao Zhu, Matthew Halpern, Vijay Janapa Reddi
  The Role of the CPU in Energy-Efficient Mobile Web Browsing
  IEEE Micro Special Issue on Mobile Systems, Jan/Feb 2015, 35(1):26-33

- Yuhao Zhu, Aditya Srikanth, Jingwen Leng, Vijay Janapa Reddi
  Exploiting Webpage Characteristics for Energy-Efficient Mobile Web Browsing
  Computer Architecture Letters (CAL), Oct 2012, 13(1):33-36
  Awarded Best of Computer Architecture Letter in 2014

- Yuhao Zhu, Bo Wang, Yangdong Deng
  Massively Parallel Logic Simulation with GPUs
  ACM Transactions on Design Automation of Electronic Systems (TODAES), June 2011, 16(3):29

Conference Papers

- Yuhao Zhu, Vijay Janapa Reddi
  GreenWeb: Language Extensions for Energy-Efficient Mobile Web Computing
  PLDI 2016

- Matthew Halpern, Yuhao Zhu, Vijay Janapa Reddi
  Mobile CPU's Rise to Power: Quantifying the Impact of Generational Mobile CPU Design Trends on Performance, Energy, and User Satisfaction
  HPCA 2016

- Yuhao Zhu, Daniel Richins, Matthew Halpern, Vijay Janapa Reddi
  Microarchitectural Implications of Event-driven Server-side Web Applications
  MICRO 2015
  IEEE Micro Top Picks of Computer Architecture (Honorable Mention) in 2016

- Yuhao Zhu, Matthew Halpern, Vijay Janapa Reddi
  Event-based Scheduling for Energy-Efficient QoS (eQoS) in Mobile Web Applications
  HPCA 2015

- Matthew Halpern, Yuhao Zhu, Ramesh Peri, Vijay Janapa Reddi
  Mosaic: Cross-Platform User-Interaction Record and Replay Tool for the Fragmented Android Ecosystem
  ISPASS 2015

- Yuhao Zhu, Vijay Janapa Reddi
  WebCore: Architectural Support for Mobile Web Browsing
  ISCA 2014

- Chen Zhou, Xiaofei Wang, Weichao Xu, Yuhao Zhu, Vijay Janapa Reddi, Chris Kim
  Estimation of Instantaneous Frequency Fluctuation in a Fast DVFS Environment Using an

Empirical BTI Stress-Relaxation Model
IRPS 2014

- Yuhao Zhu, Vijay Janapa Reddi
  High-Performance and Energy-Efficient Mobile Web Browsing on Big/Little Systems
  HPCA 2013
- Yuhao Zhu, Yangdong Deng, Yubei Chen
  Hermes: An Integrated CPU/GPU Microarchitecture for IP Routing
  DAC 2011
- Bo Wang, Yuhao Zhu, Yangdong Deng
  Distributed Time, Conservative Parallel Logic Simulation on GPUs
  DAC 2010

Book Chapters

- Yangdong Deng, Yuhao Zhu, Bo Wang
  Asynchronous Parallel Logic Simulation on Modern Graphics Processors
  GPU Solutions to Multi-scale Problems in Science and Engineering, 2013
- Yangdong Deng, Xiaomemg Jiao, Shuai Mu, Kang Kang, Yuhao Zhu
  NPGPU: Network Processing on Graphics Processing Units
  Theoretical and Mathematical Foundations of Computer Science, 2011

Other Significant Writeups

- Yuhao Zhu, Gu-Yeon Wei, David Brooks
  Cloud No Longer a Silver Bullet, Edge to Rescue
- Matthew Halpern, Yuhao Zhu, Vijay Janapa Reddi
  Generational Mobile CPU Design Trends and Its Impact on Performance, Energy, and User Satisfaction
  Summary of the HPCA 2016 paper
- Daniel Richins, Yuhao Zhu, Matthew Halpern, Vijay Janapa Reddi
  Locality Lost: Unlocking the Performance of Event-driven Servers
  Summary of the MICRO 2015 paper
- Yuhao Zhu, Vijay Janapa Reddi
  WebCore: An Architectural Substrate for Enabling High-Performance and Energy-Efficient Mobile Web
  Summary of the ISCA 2014 paper
- Yuhao Zhu, Vijay Janapa Reddi
  Energy-Conscious QoS in the Mobile Web with Hardware Heterogeneity
  Summary of the HPCA 2013 paper
- Yuhao Zhu, Yubei Chen, Yangdong Deng
  A Heterogeneous Packet Processing Architecture and Its Analytical Performance Model
  Technical report of the DAC 2011 paper

ACADEMIC TALKS

- **The Watt Wise Web**
  Texas A&M University, Jan 2017, Teleseminar
  Boston Area Architecture Workshop (BARC), Jan 2017, Cambridge, MA
  ARM Research, Jan 2017, Austin, TX
  UT Austin School of Information, Feb 2017, Austin, TX
- GreenWeb: Language Extensions for Energy-Efficient Mobile Web Computing
  PLDI 2016, June 2016, Santa Barbara, CA
- The Human Processing Unit (HPU) as a New Approximate Computing Substrate
  WAX 2016 co-located with ASPLOS 2016, April 2016, Atlanta, GA

- Scalable End-to-end Quality Control in Approximate Computing
  WAX 2016 co-located with ASPLOS 2016, April 2016, Atlanta, GA
- Microarchitectural Implications of Event-driven Server-side Web Applications
  MICRO 2015, December 2015, Waikiki, HI (lightening version)
- Exploiting Webpage Characteristics for Energy-Efficient Mobile Web Browsing
  Best of CAL 2014 presented at HPCA 2015, February 2015, San Francisco, CA
- Event-Based Scheduling for Energy-Efficient QoS (eQoS) in Mobile Web Applications
  HPCA 2015, February 2015, San Francisco, CA
- WebCore: Architectural Support for Mobile Web Browsing
  ISCA 2014, June 2014, Minneapolis, MN (lightening version)
  Intel invited talk, July 2014, Austin, TX
- High-Performance and EnergyEfficient Mobile Web Browsing on Big/Little Systems
  HPCA 2013, February 2013, Shenzhen, China
  UT Austin Programming Language Lunch Seminar, September 2012, Austin, TX
  AMD Research Lab, August 2012, Austin, TX
- Hermes: An Integrated CPU/GPU Microarchitecture for IP Routing
  DAC 2011, June 2011, San Diego, CA

RESEARCH EXPERIENCE

The University of Texas at Austin (*Graduate*)

Energy-Efficient Mobile Web Computing *8/2011-present*

Language Support for Energy-Efficient Mobile Web (*PLDI 2016*)
- GreenWeb is one of the first programming language extensions that express end-user QoS to guide the underlying runtime system for energy-efficient mobile Web computing.

Energy-aware Mobile Web Runtime (*CAL 2012, HPCA 2013, HPCA 2015*)
- Event-based Scheduling (EBS) is a new scheduling mechanism that leverages the fundamental event-driven execution model of mobile applications for energy optimizations.

Web-Specific Mobile Processor Architecture (*ISCA 2014*)
- WebCore is a Web-specific mobile processor architecture customized and specialized for key Web computation kernels and communication patterns.

Understanding the Role of CPU in Mobile (Web) Computing (*IEEE Micro 2015, HPCA 2016*)
- Quantitatively demonstrated the importance of CPU to the overall mobile Web performance and energy-efficiency. Research focuses on the compute side is thus warranted.

Cloud Computing of the Future *1/2015-present*

Microarchitecture for Scripted Event-driven Web Services (*MICRO 2015*)
- Devised microarchitectural mechanisms to improve managed-language-based, event-driven server performance by mitigating fundamental front-end bottlenecks.

Taming Tail Latency in Event-driven Web Services (In Submission)
- Event Dependency Graph is a runtime framework that identifies the root-causes of, and helps mitigate, tail latency in event-driven servers.

CPU Microarchitecture (with Yale N. Patt) *2/2011-7/2011*

Microarchitecture Support for Speculative Loop-level SIMDization
- Quantified the upper bound of loop level parallelism in the form of SIMD in SPEC CPU benchmark. Proposed hardware mechanisms for dynamically vectorizing loops.

Tsinghua University (*Undergraduate*)

GPU Programming and Microrchitecture (with Yangdong Deng) *6/2009-7/2010*

Hermes: An Integrated CPU/GPU Microarchitecture for IP Routing (*DAC 2011*)

- Designed and prototyped a CPU+GPU heterogeneous IP router microarchitecture that adaptively balances packet delay and overall throughput.

Parallel Gate Level Logic Simulation on GPUs (*DAC 2010, TODAES 2011*)

- Design and prototyped a gate-level logic simulator on GPUs that outperformed Synopsys' sequential simulator by one order of magnitude
- Introduced dynamic paging into GPGPU programming to mitigate the limited memory capacity of GPUs. A similar implementation is later introduced into the CUDA runtime.

SOFTWARE ARTIFACTS

GreenWeb (https://codereview.chromium.org/1835303002/)

- Language and runtime for energy-efficient mobile Web.

AutoGreen (https://github.com/yuhao/AutoGreen)

- Framework to annotate Web applications with GreenWeb language extensions.

Node Benchmark (https://github.com/nodebenchmark/benchmarks)

- Node.js-based event-driven server workload suite.

Domino.js (https://github.com/yuhao/domino.js)

- DOM instrumentation and optimization framework.

COMMUNITY SERVICES

Tiny Transactions on Computer Science (TinyToCS) (http://tinytocs.org/)

- Program Committee Chair, Volume IV
- Program Committee Member, Volume III

MobiTools (http://mobitools.ece.utexas.edu/)

- Workshop/Tutorial on *Infrastructure for Mobile Computer Architecture Research*. Co-located with ISCA 2016 (~30 academia and industry attendance)
- Co-founder and organizer. Hosted the entire workshop.

Women in Engineering Program (WEP) (http://www.engr.utexas.edu/wep)

- Speaker at high school-focused summer camps with an emphasis on inspiring high school female about engineering (~75 high school senior women)
- Mentored two female sophomores in the Graduates Linked with Undergraduates in Engineering (GLUE) Program. Hannah Peeler, one of my mentees, won the GLUE scholarship.

Emerging Topics in Computer System Design Reading Group (https://goo.gl/gzT4Fw)

- Co-founder. Moderate the weekly seminar (~15 graduate students and faculty members.)

Program Committee

- ISCA 2017: ACM/IEEE International Symposium on Computer Architecture (ERC)
- HCOMP 2016: AAAI Human Computation and Crowdsourcing
- CGO-PPoPP Artifacts Evaluation 2015, 2016

Organizing Committee

- Co-chair: Sensors to Cloud Architectures Workshop (SCAW) co-located with HPCA 2017
- Co-organizer: MobiTools Workshop co-located with ISCA 2016

Solicited Reviewer

- TACO (2016), ESL (2015), DAC (2011, 2012), TODAES (2011)

TEACHING EXPERIENCE

Teaching Assistant

- Lower Division, Introduction to Embedded Systems, with Prof. Jonathan W. Valvano
- Upper Division, Computer Architecture, with Prof. Yale N. Patt
- Graduate Level, Dynamic Compilation, with Prof. Vijay Janapa Reddi

Guest Lectures

- *Model Checking*, Advanced Complier, Fall 2010
- *Garbage Collection*, Dynamic Compilation, Spring 2014
- *Compilation for Power-Efficiency* , Dynamic Compilation, Spring 2014
- *Mobile CPU Design Trends*, Computer Architecture: User System Interplay, Spring 2016
- *Architectural Support for Mobile Web*, Mixed-Signal IC Design, Harvard University, Fall 2016

MENTORING

- Hannah Peeler, Undergraduate student at UT Austin, 2016
- Janna Tulabot, Undergraduate student at UT Austin, 2016
- Matthew Halpern, Ph.D. student at UT Austin, 2013 - present
- Wenzhi Cui, Ph.D. student at UT Austin, 2015 - present

REFERENCES

Vijay Janapa Reddi (Advisor)
Assistant Professor
Electrical and Computer Engineering Department
The University of Texas at Austin
`vj@ece.utexas.edu`
+1 (408) 390-2790

Scott Mahlke
Professor
Electrical Engineering and Computer Science Department
University of Michigan
`mahlke@umich.edu`
+1 (734) 936-1602

Matt Welsh
Computer Scientist & Engineering Manager
Google, Inc.
`mdw@mdw.la`
+1 (617) 819-4504

David Brooks
Haley Family Professor
School of Engineering and Applied Sciences
Harvard University
`dbrooks@seas.harvard.edu`
+1 (617) 495-3989

Christine Julien
Professor, Temple Foundation Endowed Faculty Fellowship No. 7
Electrical and Computer Engineering Department
The University of Texas at Austin
`c.julien@mail.utexas.edu`
+1 (512) 232-5671