Vijay Janapa Reddi, Ph.D.

Harvard Science & Engineering Complex
Room #5.305, 150 Western Avenue
Boston, MA 02134
Phone: +1 (408) 390-2790

Email: vj@eecs.harvard.edu
URL: https://seas.harvard.edu/person/vijay-janapa-reddi

Current Position

John L. Loeb Associate Professor of Engineering and Applied Sciences
John A. Paulson School of Engineering and Applied Sciences
Harvard University

Primary Research Areas

Computer Architecture; Machine Learning Systems; Autonomous Agents/Robotics

Education

2010 PH.D. in Computer Science, Harvard University
2006 M.S. in Electrical and Computer Engineering, University of Colorado—Boulder
2003 B.S. in Computer Engineering, Santa Clara University

Current and Previous Academic Positions

2019/– *Associate Professor*, Harvard University
2018/– *Adjunct Associate Professor*, The University of Texas at Austin
2017/2018 *Associate Professor*, The University of Texas at Austin
2011/2017 *Assistant Professor*, The University of Texas at Austin

Other Industry Experience

2022/- *Visiting Researcher*, Google (DeepMind)
2020/2021 *Visiting Researcher*, Google (Tensor Flow Ecosystem)
2019/2019 *Visiting Researcher*, Facebook (AR Silicon Team)
2017/2018 *Visiting Researcher*, Google (gChips)
2015/2016 *Consultant*, Intel
2015/2016 *Consultant*, Advanced Micro Devices (AMD)

2014/2014 *Consultant*, Intel
2010/2011 *Senior Design Engineer*, Advanced Micro Devices (AMD)
2009/2009 *Research Intern*, Microsoft Research
2007/2009 *Research Intern*, VMware
2003/2006 *Research Intern*, Intel

Honors and Awards

2023 MLPerf Inference selected for inclusion in ISCA@50 25-Year Retrospective
2023 IEEE Micro Top Picks in Computer Architecture (Honorable Mention)
2022 IEEE Micro Top Picks in Computer Architecture (Honorable Mention)
2021 BenchCouncil Rising Star Award, International Open Benchmark Council
2021 Deploying TinyML on HarvardX/edX: 100 Most Popular Free Online Courses, ClassCentral
2021 Best Course in AI: Tiny Machine Learning (TinyML) on HarvardX/edX, CogX Awards
2021 IEEE Micro Top Picks in Computer Architecture
2021 Best of Computer Architecture Letters (CAL), Editorial Board of IEEE CAL
2020 Programming Languages Software Award, ACM SIGPLAN
2020 Best Research Paper Award, Design Automation Conference (DAC)
2020 Google Faculty Research Award, Google
2019 Best Paper Nominee, IEEE International Symposium on Perf. Analysis of Systems and Software (ISPASS)
2019 Intl Symp. on High-Performance Computer Architecture (HPCA) Hall of Fame
2018 International Symp. on Microarchitecture (MICRO) Hall of Fame
2018 ACM SIGARCH CS TCCA Outstanding Dissertation Award (Advisee: Yuhao Zhu)
2017 IEEE Micro Top Picks in Computer Architecture
2017 Best Paper Nominee, Design Automation Conference (DAC)
2017 Google Faculty Research Award
2016 IEEE TCCA Young Computer Architect Award
2016 IEEE Micro Top Picks in Computer Architecture (Honorable Mention)
2016 Gilbreth Lectureship Honor, National Academy of Engineering (NAE)
2015 ACM SIGPLAN Most Influential PLDI Paper Award
2015 Google Faculty Research Award
2014 Best of Computer Architecture Letters (CAL) Award
2014 Best Paper Nominee, IEEE International Symposium on MicroarchitectureLow Power Electronics and Design (ISLPED)
2014 Indo-American Frontiers of Engineering, National Academy of Engineering (NAE)
2013 Google Faculty Research Award
2013 Intel Early Career Award
2012 Google Faculty Research Award
2011 IEEE Micro Top Picks in Computer Architecture
2010 IEEE Micro Top Picks in Computer Architecture
2009 Best Paper Award, International Symposium on High-Performance Computer Architecture (HPCA)
2008 John A. and Elizabeth S. Armstrong Fellowship, Harvard University

2007 Best Student Presentation, International Symposium on Code Generation and Optimization (CGO)
2006 IEEE Micro Top Picks in Computer Architecture
2005 Best Paper Award, International Symposium on MicroarchitectureMicroarchitecture (MICRO)
2003 Faculty Recognition for Technical Excellence, Santa Clara University
2003 Outstanding Undergraduate (Honorable), Computing Research Association (CRA)

University Committee Assignments

Harvard John A. Paulson School of Engineering and Applied Sciences

2023/2024 Member, Engineering Sciences Committee on Higher Degrees
2023 Member, Generative AI Steering Committee
2020 Member, Quantum Faculty Recruiting Committee
2019 Member, Robotics Faculty Recruiting Committee
2019/- Graduate Student Admissions Committee

The University of Texas at Austin

2016 Member, Faculty Recruiting Committee
2015 Member, Faculty Recruiting Committee
2014 Member, Technology in Teaching
2013 Member, Faculty Recruiting Committee
2011/2016 Graduate Student Admissions Committee

Professional Society and Major Governmental Committees

Membership

Institute of Electrical and Electronics Engineers (IEEE)
Association for Computing Machinery (ACM)

MLCommons and MLPerf

2019-/ Vice-Chair (on the Board of Directors), MLCommons (http://mlcommons.org)
2019-/ Vice President, MLCommons
2019-/ MLCommons Research Co-Chair, MLCommons
2019/2020 MLPerf Tiny Co-Chair, MLPerf (http://mlperf.org)
2018/2020 MLPerf Inference Co-Chair, MLPerf

IRDS

2020/2023 International Roadmap for Devices and Systems (IRDS)

SIGARCH

2017/2020 Associate Editor, SIGARCH Blog (http://sigarch.org/blog/)

Professional Service

General Chair

2023 Tiny Machine Learning Research Symposium (TinyML)
2022 Tiny Machine Learning Research Symposium (TinyML)
2021 Tiny Machine Learning Research Symposium (TinyML)

Program Chair

2018 IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2019 IEEE International Symposium on Workload Characterization (IISWC)
2014 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)

Program Committee

2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2023 International Symposium on Computer Architecture (ISCA)
2023 Sixth Conference on Machine Learning and Systems (MLSys)
2023 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2022 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2022 International Symposium on Computer Architecture (ISCA)
2022 Fifth Conference on Machine Learning and Systems (MLSys)
2021 ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)
2021 IEEE Micro Top Picks
2021 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2020 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2020 International Symposium on Computer Architecture (ISCA)
2020 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2019 ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS) (ERC)
2019 IEEE Micro Top Picks
2019 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2019 International Symposium on Computer Architecture (ISCA)
2019 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2018 International Symposium on Computer Architecture (ISCA) (ERC)
2018 IEEE Micro Top Picks
2018 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2018 IEEE International Symposium on Workload Characterization (IISWC)
2017 IEEE International Symposium on High-Performance Computer Architecture (HPCA) (ERC)
2017 International Symposium on Computer Architecture (ISCA)
2017 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2016 IEEE International Symposium on High-Performance Computer Architecture (HPCA) (ERC)

2016 IEEE Micro Top Picks
2016 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2016 IEEE International Symposium on Workload Characterization (IISWC)
2015 International Symposium on Computer Architecture (ISCA) (ERC)
2015 IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2015 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2015 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2015 International Symposium on Performance Analysis of Systems and Software (ISPASS)
2015 International Symposium on Principles and Practice of Parallel Computing (PPoPP)
2015 IEEE International Symposium on Workload Characterization (IISWC)
2014 IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2014 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)
2014 IEEE/ACM International Symposium on Microarchitecture (MICRO)
2013 IEEE International Symposium on Workload Characterization (IISWC)

Organizing Committee

2023 Data-centric Machine Learning Research (DMLR) Workshop (ICML)
2020 ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)
2017 Design Automation Conference
2016-2017 Workshop on Cognitive Edge Computing (CogEdge)
2016 Tutorial on Tools for Mobile Computer Architecture (MobiTools)
2015 IEEE/ACM International Symposium on Code Generation and Optimization
2016 IEEE International Symposium on Microarchitecture
2015-2016 Tutorial on Simulation and Analysis Engine (ISCA, ASPLOS, HPCA, ICS, IISWC, ISPASS)
2013 IEEE International Symposium on Workload Characterization
2012 IEEE International Symposium on Performance Analysis of Systems and Software
2013 IEEE International Symposium on Performance Analysis of Systems and Software.

Guest Editor

2023 IEEE Micro Special Issue on Tiny Machine Learning
2016 IEEE Micro Special Issue on Internet of Things
2013 IEEE Micro Special Issue on Reliability-Aware Microarchitecture Design

Local Arrangements Chair

2013 Intl. Symp. on Performance Analysis of Systems and Software (ISPASS)
2015, 2016 Workshop on Silicon Errors in Logic: System Effects (SELSE).

Publications Chair

2013 Intl. Symp. on Workload Characterization (IISWC)

Community Activities

Tiny Machine Learning Open Education Initiative, https://tinyml.seas.harvard.edu (Founder)

Hands-on Computer Science (HaCS) for Austin Independent School District (via UT Outreach), https://outreach.uteach.utexas.edu/uteach-outreach-cs-service-learning-program

Publications

As of March 4, 2024, my research publications have accumulated a total of 14628 citations with an h-index of 48, as reported by Google Scholar. The most current list of publications can be accessed at the following link: here.

Conference Publications

1 Yu-Shun Hsiao, Siva Kumar Sastry Hari, Balakumar Sundaralingam, Jason Yik, Thierry Tambe, Charbel Sakr, Stephen W. Keckler, and **Vijay Janapa Reddi**. "VaPr: Variable-Precision Tensors to Accelerate Robot Motion Planning". In: *IROS*. 2023, pp. 6304–6309. DOI: 10.1109/IROS55552.2023.10342109. URL: https://doi.org/10.1109/IROS55552.2023.10342109.

2 Yu-Shun Hsiao, Zishen Wan, Tianyu Jia, Radhika Ghosal, Abdulrahman Mahmoud, Arijit Raychowdhury, David Brooks, Gu-Yeon Wei, and **Vijay Janapa Reddi**. "MAVFI: An End-to-End Fault Analysis Framework with Anomaly Detection and Recovery for Micro Aerial Vehicles". In: *Design, Automation & Test in Europe Conference & Exhibition, DATE 2023, Antwerp, Belgium, April 17-19, 2023*. IEEE, 2023, pp. 1–6. DOI: 10.23919/DATE56975.2023.10137246. URL: https://doi.org/10.23919/DATE56975.2023.10137246.

3 Srivatsan Krishnan, Amir Yazdanbakhsh, Shvetank Prakash, Jason Jabbour, Ikechukwu Uchendu, Susobhan Ghosh, Behzad Boroujerdian, Daniel Richins, Devashree Tripathy, Aleksandra Faust, and **Vijay Janapa Reddi**. "ArchGym: An Open-Source Gymnasium for Machine Learning Assisted Architecture Design". In: *Proceedings of the 50th Annual International Symposium on Computer Architecture, ISCA 2023, Orlando, FL, USA, June 17-21, 2023*. ACM, 2023, 14:1–14:16. DOI: 10.1145/3579371.3589049. URL: https://doi.org/10.1145/3579371.3589049.

4 Mark Mazumder, Colby R. Banbury, Xiaozhe Yao, Bojan Karlas, William Gaviria Rojas, Sudnya Frederick Diamos, Greg Diamos, Lynn He, Alicia Parrish, Hannah Rose Kirk, Jessica Quaye, Charvi Rastogi, Douwe Kiela, David Jurado, David Kanter, Rafael Mosquera, Will Cukierski, Juan Ciro, Lora Aroyo, Bilge Acun, Lingjiao Chen, Mehul Raje, Max Bartolo, Evan Sabri Eyuboglu, Amirata Ghorbani, Emmett D. Goodman, Addison Howard, Oana Inel, Tariq Kane, Christine R. Kirkpatrick, D. Sculley, Tzu-Sheng Kuo, Jonas W. Mueller, Tristan Thrush, Joaquin Vanschoren, Margaret Warren, Adina Williams, Serena Yeung, Newsha Ardalani, Praveen K. Paritosh, Ce Zhang, James Y. Zou, Carole-Jean Wu, Cody Coleman, Andrew Y. Ng, Peter Mattson, and **Vijay Janapa Reddi**. "DataPerf: Benchmarks for Data-Centric AI Development". In: *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*. 2023. URL: http://papers.nips.cc/paper%5C_files/

paper/2023/hash/112db88215e25b3ae2750e9eefcded94-Abstract-Datasets%5C_and%5C_Benchmarks.html.

5 Sabrina M. Neuman, Radhika Ghosal, Thomas Bourgeat, Brian Plancher, and **Vijay Janapa Reddi**. "RoboShape: Using Topology Patterns to Scalably and Flexibly Deploy Accelerators Across Robots". In: *Proceedings of the 50th Annual International Symposium on Computer Architecture, ISCA 2023, Orlando, FL, USA, June 17-21, 2023*. ACM, 2023, 69:1–69:13. DOI: 10.1145/3579371.3589104. URL: https://doi.org/10.1145/3579371.3589104.

6 Thanh Thi Nguyen, Cuong M. Nguyen, Thien Huynh-The, Quoc-Viet Pham, Quoc Viet Hung Nguyen, Imran Razzak, and **Vijay Janapa Reddi**. "Solving Complex Sequential Decision-Making Problems by Deep Reinforcement Learning with Heuristic Rules". In: *Computational Science - ICCS 2023 - 23rd International Conference, Prague, Czech Republic, July 3-5, 2023, Proceedings, Part II*. Vol. 14074. Lecture Notes in Computer Science. Springer, 2023, pp. 298–305. DOI: 10.1007/978-3-031-36021-3_30. URL: https://doi.org/10.1007/978-3-031-36021-3%5C_30.

7 Shvetank Prakash, Tim Callahan, Joseph Bushagour, Colby R. Banbury, Alan V. Green, Pete Warden, Tim Ansell, and **Vijay Janapa Reddi**. "CFU Playground: Full-Stack Open-Source Framework for Tiny Machine Learning (TinyML) Acceleration on FPGAs". In: *IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2023, Raleigh, NC, USA, April 23-25, 2023*. IEEE, 2023, pp. 157–167. DOI: 10.1109/ISPASS57527.2023.00024. URL: https://doi.org/10.1109/ISPASS57527.2023.00024.

8 Shvetank Prakash, Tim Callahan, Joseph Bushagour, Colby R. Banbury, Alan V. Green, Pete Warden, Tim Ansell, and **Vijay Janapa Reddi**. "CFU Playground: Want a faster ML processor? Do it yourself!" In: *Design, Automation & Test in Europe Conference & Exhibition, DATE 2023, Antwerp, Belgium, April 17-19, 2023*. IEEE, 2023, pp. 1–2. DOI: 10.23919/DATE56975.2023.10137093. URL: https://doi.org/10.23919/DATE56975.2023.10137093.

9 **Vijay Janapa Reddi** and Amir Yazdanbakhsh. "Architecture 2.0: Challenges and Opportunities". In: *60th ACM/IEEE Design Automation Conference, DAC 2023, San Francisco, CA, USA, July 9-13, 2023*. IEEE, 2023, pp. 1–2. DOI: 10.1109/DAC56929.2023.10247808. URL: https://doi.org/10.1109/DAC56929.2023.10247808.

10 Zishen Wan, Nandhini Chandramoorthy, Karthik Swaminathan, Pin-Yu Chen, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "BERRY: Bit Error Robustness for Energy-Efficient Reinforcement Learning-Based Autonomous Systems". In: *60th ACM/IEEE Design Automation Conference, DAC 2023, San Francisco, CA, USA, July 9-13, 2023*. IEEE, 2023, pp. 1–6. DOI: 10.1109/DAC56929.2023.10247999. URL: https://doi.org/10.1109/DAC56929.2023.10247999.

11 Yu-Shun Hsiao, Siva Kumar Sastry Hari, Michal Filipiuk, Timothy Tsai, Michael B. Sullivan, **Vijay Janapa Reddi**, Vasu Singh, and Stephen W. Keckler. "Zhuyi: perception processing rate estimation for safety in autonomous vehicles". In: *DAC '22: 59th ACM/IEEE Design Automation Conference, San Francisco, California, USA, July 10 - 14, 2022*. ACM, 2022, pp. 289–294. DOI: 10.1145/3489517.3530445. URL: https://doi.org/10.1145/3489517.3530445.

12 Tianyu Jia, En-Yu Yang, Yu-Shun Hsiao, Jonathan J. Cruz, David Brooks, Gu-Yeon Wei, and **Vijay Janapa Reddi**. "OMU: A Probabilistic 3D Occupancy Mapping Accelerator for Real-time OctoMap at the Edge". In: *2022 Design, Automation & Test in Europe Conference & Exhibition, DATE 2022, Antwerp, Belgium, March 14-23, 2022*. IEEE, 2022, pp. 909–914. DOI: 10.23919/DATE54114.2022.9774508. URL: https://doi.org/10.23919/DATE54114.2022.9774508.

13 Srivatsan Krishnan, Zishen Wan, Kshitij Bhardwaj, Ninad Jadhav, Aleksandra Faust, and **Vijay Janapa Reddi**. "Roofline Model for UAVs: A Bottleneck Analysis Tool for Onboard Compute Characterization of Autonomous Unmanned Aerial Vehicles". In: *International IEEE Symposium on Performance Analysis of Systems and Software, ISPASS 2022, Singapore, May 22-24, 2022*. IEEE, 2022, pp. 162–174. DOI: 10.1109/ISPASS55109.2022.00023. URL: https://doi.org/10.1109/ISPASS55109.2022.00023.

14 Srivatsan Krishnan, Zishen Wan, Kshitij Bhardwaj, Paul N. Whatmough, Aleksandra Faust, Sabrina M. Neuman, Gu-Yeon Wei, David Brooks, and **Vijay Janapa Reddi**. "Automatic Domain-Specific SoC Design for Autonomous Unmanned Aerial Vehicles". In: *55th IEEE/ACM International Symposium on Microarchitecture, MICRO 2022, Chicago, IL, USA, October 1-5, 2022*. IEEE, 2022, pp. 300–317. DOI: 10.1109/MICRO56248.2022.00033. URL: https://doi.org/10.1109/MICRO56248.2022.00033.

15 Sabrina M. Neuman, Brian Plancher, Bardienus Pieter Duisterhof, Srivatsan Krishnan, Colby R. Banbury, Mark Mazumder, Shvetank Prakash, Jason Jabbour, Aleksandra Faust, Guido C. H. E. de Croon, and **Vijay Janapa Reddi**. "Tiny Robot Learning: Challenges and Directions for Machine Learning in Resource-Constrained Robots". In: *4th IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2022, Incheon, Republic of Korea, June 13-15, 2022*. IEEE, 2022, pp. 296–299. DOI: 10.1109/AICAS54282.2022.9870000. URL: https://doi.org/10.1109/AICAS54282.2022.9870000.

16 Brian Plancher, Sabrina M. Neuman, Radhika Ghosal, Scott Kuindersma, and **Vijay Janapa Reddi**. "GRiD: GPU-Accelerated Rigid Body Dynamics with Analytical Gradients". In: *2022 International Conference on Robotics and Automation, ICRA 2022, Philadelphia, PA, USA, May 23-27, 2022*. IEEE, 2022, pp. 6253–6260. DOI: 10.1109/ICRA46639.2022.9812384. URL: https://doi.org/10.1109/ICRA46639.2022.9812384.

17 Brian Plancher and **Vijay Janapa Reddi**. "TinyMLedu: The Tiny Machine Learning Open Education Initiative". In: *SIGCSE 2022: The 53rd ACM Technical Symposium on Computer Science Education, Providence, RI, USA, March 3-5, 2022, Volume 2*. ACM, 2022, p. 1159. DOI: 10.1145/3478432.3499093. URL: https://doi.org/10.1145/3478432.3499093.

18 **Vijay Janapa Reddi**, David Kanter, Peter Mattson, Jared Duke, Thai Nguyen, Ramesh Chukka, Kenneth Shiring, Koan-Sin Tan, Mark Charlebois, William Chou, Mostafa El-Khamy, Jungwook Hong, Tom St. John, Cindy Trinh, Michael Buch, Mark Mazumder, Relja Markovic, Thomas Atta-fosu, Fatih Çakir, Masoud Charkhabi, Xiaodong Chen, Cheng-Ming Chiang, Dave Dexter, Terry Heo, Guenther Schmuelling, Maryam Shabani, and Dylan Zika. "MLPerf Mobile Inference Benchmark: An Industry-Standard Open-Source Machine Learning Benchmark for On-Device AI". In: *Proceedings of Machine Learning and Systems 2022, MLSys 2022, Santa Clara, CA, USA, August 29 - Septem-*

ber 1, 2022. mlsys.org, 2022. URL: https://proceedings.mlsys.org/paper/2022/hash/7eabe3a1649ffa2b3ff8c02ebfd5659f-Abstract.html.

19 William Gaviria Rojas, Sudnya Frederick Diamos, Keertan Kini, David Kanter, **Vijay Janapa Reddi**, and Cody Coleman. "The Dollar Street Dataset: Images Representing the Geographic and Socioeconomic Diversity of the World". In: *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*. 2022. URL: http://papers.nips.cc/paper%5C_files/paper/2022/hash/5474d9d43c0519aa176276ff2c1ca528-Abstract-Datasets%5C_and%5C_Benchmarks.html.

20 Víctor Mayoral Vilches, Sabrina M. Neuman, Brian Plancher, and **Vijay Janapa Reddi**. "RobotCore: An Open Architecture for Hardware Acceleration in ROS 2". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2022, Kyoto, Japan, October 23-27, 2022*. IEEE, 2022, pp. 9692–9699. DOI: 10.1109/IROS47612.2022.9982082. URL: https://doi.org/10.1109/IROS47612.2022.9982082.

21 Zishen Wan, Aqeel Anwar, Abdulrahman Mahmoud, Tianyu Jia, Yu-Shun Hsiao, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "FRL-FI: Transient Fault Analysis for Federated Reinforcement Learning-Based Navigation Systems". In: *2022 Design, Automation & Test in Europe Conference & Exhibition, DATE 2022, Antwerp, Belgium, March 14-23, 2022*. IEEE, 2022, pp. 430–435. DOI: 10.23919/DATE54114.2022.9774562. URL: https://doi.org/10.23919/DATE54114.2022.9774562.

22 Zishen Wan, Ashwin Sanjay Lele, Bo Yu, Shaoshan Liu, Yu Wang, **Vijay Janapa Reddi**, Cong Hao, and Arijit Raychowdhury. "Robotic Computing on FPGAs: Current Progress, Research Challenges, and Opportunities". In: *4th IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2022, Incheon, Republic of Korea, June 13-15, 2022*. IEEE, 2022, pp. 291–295. DOI: 10.1109/AICAS54282.2022.9869951. URL: https://doi.org/10.1109/AICAS54282.2022.9869951.

23 Colby R. Banbury, **Vijay Janapa Reddi**, Peter Torelli, Nat Jeffries, Csaba Király, Jeremy Holleman, Pietro Montino, David Kanter, Pete Warden, Danilo Pau, Urmish Thakker, Antonio Torrini, Jay Cordaro, Giuseppe Di Guglielmo, Javier M. Duarte, Honson Tran, Nhan Tran, Wenxu Niu, and Xuesong Xu. "MLPerf Tiny Benchmark". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*. 2021. URL: https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/da4fb5c6e93e74d3df8527599fa62642-Abstract-round1.html.

24 Colby R. Banbury, Chuteng Zhou, Igor Fedorov, Ramon Matas Navarro, Urmish Thakker, Dibakar Gope, **Vijay Janapa Reddi**, Matthew Mattina, and Paul N. Whatmough. "MicroNets: Neural Network Architectures for Deploying TinyML Applications on Commodity Microcontrollers". In: *Proceedings of Machine Learning and Systems 2021, MLSys 2021, virtual, April 5-9, 2021*. mlsys.org, 2021. URL: https://proceedings.mlsys.org/paper/2021/hash/a3c65c2974270fd093ee8a9bf8ae7d0b-Abstract.html.

25 Behzad Boroujerdian, Radhika Ghosal, Jonathan J. Cruz, Brian Plancher, and **Vijay Janapa Reddi**. "RoboRun: A Robot Runtime to Exploit Spatial Heterogeneity". In: *58th ACM/IEEE Design Automation Conference, DAC 2021, San Francisco, CA, USA, December 5-9, 2021*. IEEE, 2021, pp. 829–834.

DOI: 10.1109/DAC18074.2021.9586280. URL: https://doi.org/10.1109/DAC18074.2021.9586280.

26 Michael Buch, Zahra Azad, Ajay Joshi, and **Vijay Janapa Reddi**. "AI Tax in Mobile SoCs: End-to-end Performance Analysis of Machine Learning in Smartphones". In: *IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2021, Stony Brook, NY, USA, March 28-30, 2021*. IEEE, 2021, pp. 96–106. DOI: 10.1109/ISPASS51385.2021.00027. URL: https://doi.org/10.1109/ISPASS51385.2021.00027.

27 Robert David, Jared Duke, Advait Jain, **Vijay Janapa Reddi**, Nat Jeffries, Jian Li, Nick Kreeger, Ian Nappier, Meghna Natraj, Tiezhen Wang, Pete Warden, and Rocky Rhodes. "TensorFlow Lite Micro: Embedded Machine Learning for TinyML Systems". In: *Proceedings of Machine Learning and Systems 2021, MLSys 2021, virtual, April 5-9, 2021*. mlsys.org, 2021. URL: https://proceedings.mlsys.org/paper/2021/hash/d2ddea18f00665ce8623e36bd4e3c7c5-Abstract.html.

28 Bardienus Pieter Duisterhof, Srivatsan Krishnan, Jonathan J. Cruz, Colby R. Banbury, William Fu, Aleksandra Faust, Guido C. H. E. de Croon, and **Vijay Janapa Reddi**. "Tiny Robot Learning (tinyRL) for Source Seeking on a Nano Quadcopter". In: *IEEE International Conference on Robotics and Automation, ICRA 2021, Xi'an, China, May 30 - June 5, 2021*. IEEE, 2021, pp. 7242–7248. DOI: 10.1109/ICRA48506.2021.9561590. URL: https://doi.org/10.1109/ICRA48506.2021.9561590.

29 Bardienus Pieter Duisterhof, Shushuai Li, Javier Burgués, **Vijay Janapa Reddi**, and Guido C. H. E. de Croon. "Sniffy Bug: A Fully Autonomous Swarm of Gas-Seeking Nano Quadcopters in Cluttered Environments". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2021, Prague, Czech Republic, September 27 - Oct. 1, 2021*. IEEE, 2021, pp. 9099–9106. DOI: 10.1109/IROS51168.2021.9636217. URL: https://doi.org/10.1109/IROS51168.2021.9636217.

30 Daniel Galvez, Greg Diamos, Juan Torres, Keith Achorn, Juan Felipe Cerón, Anjali Gopi, David Kanter, Max Lam, Mark Mazumder, and **Vijay Janapa Reddi**. "The People's Speech: A Large-Scale Diverse English Speech Recognition Dataset for Commercial Usage". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*. 2021. URL: https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/202cb962ac59075b964b07152d234b70-Abstract-round1.html.

31 James Gleeson, Moshe Gabel, Gennady Pekhimenko, Eyal de Lara, Srivatsan Krishnan, and **Vijay Janapa Reddi**. "RL-Scope: Cross-stack Profiling for Deep Reinforcement Learning Workloads". In: *Proceedings of Machine Learning and Systems 2021, MLSys 2021, virtual, April 5-9, 2021*. mlsys.org, 2021. URL: https://proceedings.mlsys.org/paper/2021/hash/d1fe173d08e959397adf34b1d77e88d7-Abstract.html.

32 Maximilian Lam, Gu-Yeon Wei, David Brooks, **Vijay Janapa Reddi**, and Michael Mitzenmacher. "Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix". In: *Proceedings of the 38th International Conference on Machine Learning, ICML*

2021, 18-24 July 2021, Virtual Event. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 5959–5968. URL: http://proceedings.mlr.press/v139/lam21b.html.

33 Maximilian Lam, Zachary Yedidia, Colby R. Banbury, and **Vijay Janapa Reddi**. "Precision Batching: Bitserial Decomposition for Efficient Neural Network Inference on GPUs". In: *30th International Conference on Parallel Architectures and Compilation Techniques, PACT 2021, Atlanta, GA, USA, September 26-29, 2021*. IEEE, 2021, pp. 129–141. DOI: 10.1109/PACT52795.2021.00017. URL: https://doi.org/10.1109/PACT52795.2021.00017.

34 Mark Mazumder, Colby R. Banbury, Josh Meyer, Pete Warden, and **Vijay Janapa Reddi**. "Few-Shot Keyword Spotting in Any Language". In: *Interspeech 2021, 22nd Annual Conference of the International Speech Communication Association, Brno, Czechia, 30 August - 3 September 2021*. ISCA, 2021, pp. 4214–4218. DOI: 10.21437/INTERSPEECH.2021-1966. URL: https://doi.org/10.21437/Interspeech.2021-1966.

35 Mark Mazumder, Sharad Chitlangia, Colby R. Banbury, Yiping Kang, Juan Ciro, Keith Achorn, Daniel Galvez, Mark Sabini, Peter Mattson, David Kanter, Greg Diamos, Pete Warden, Josh Meyer, and **Vijay Janapa Reddi**. "Multilingual Spoken Words Corpus". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*. 2021. URL: https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/fe131d7f5a6b38b23cc967316c13dae2-Abstract-round2.html.

36 Sabrina M. Neuman, Brian Plancher, Thomas Bourgeat, Thierry Tambe, Srinivas Devadas, and **Vijay Janapa Reddi**. "Robomorphic computing: a design methodology for domain-specific accelerators parameterized by robot morphology". In: *ASPLOS '21: 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Virtual Event, USA, April 19-23, 2021*. ACM, 2021, pp. 674–686. DOI: 10.1145/3445814.3446746. URL: https://doi.org/10.1145/3445814.3446746.

37 Muhammad Shafique, Theocharis Theocharides, **Vijay Janapa Reddi**, and Boris Murmann. "TinyML: Current Progress, Research Challenges, and Future Roadmap". In: *58th ACM/IEEE Design Automation Conference, DAC 2021, San Francisco, CA, USA, December 5-9, 2021*. IEEE, 2021, pp. 1303–1306. DOI: 10.1109/DAC18074.2021.9586232. URL: https://doi.org/10.1109/DAC18074.2021.9586232.

38 Zishen Wan, Aqeel Anwar, Yu-Shun Hsiao, Tianyu Jia, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "Analyzing and Improving Fault Tolerance of Learning-Based Navigation Systems". In: *58th ACM/IEEE Design Automation Conference, DAC 2021, San Francisco, CA, USA, December 5-9, 2021*. IEEE, 2021, pp. 841–846. DOI: 10.1109/DAC18074.2021.9586116. URL: https://doi.org/10.1109/DAC18074.2021.9586116.

39 Jingwen Leng, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, Quan Chen, Minyi Guo, and **Vijay Janapa Reddi**. "Asymmetric Resilience: Exploiting Task-Level Idempotency for Transient Error Recovery in Accelerator-Based Systems". In: *IEEE International Symposium on High Performance Computer Architecture, HPCA 2020, San Diego, CA, USA, February 22-26, 2020*. IEEE, 2020, pp. 44–57. DOI: 10.1109/HPCA47549.2020.00014. URL: https://doi.org/10.1109/HPCA47549.2020.00014.

40 Peter Mattson, Christine Cheng, Gregory F. Diamos, Cody Coleman, Paulius Micikevicius, David A. Patterson, Hanlin Tang, Gu-Yeon Wei, Peter Bailis, Victor Bittorf, David Brooks, Dehao Chen, Debo Dutta, Udit Gupta, Kim M. Hazelwood, Andy Hock, Xinyuan Huang, Daniel Kang, David Kanter, Naveen Kumar, Jeffery Liao, Deepak Narayanan, Tayo Oguntebi, Gennady Pekhimenko, Lillian Pentecost, **Vijay Janapa Reddi**, Taylor Robie, Tom St. John, Carole-Jean Wu, Lingjie Xu, Cliff Young, and Matei Zaharia. "MLPerf Training Benchmark". In: *Proceedings of Machine Learning and Systems 2020, MLSys 2020, Austin, TX, USA, March 2-4, 2020*. mlsys.org, 2020. URL: https://proceedings.mlsys.org/book/309.pdf.

41 **Vijay Janapa Reddi**, Christine Cheng, David Kanter, Peter Mattson, Guenther Schmuelling, Carole-Jean Wu, Brian Anderson, Maximilien Breughe, Mark Charlebois, William Chou, Ramesh Chukka, Cody Coleman, Sam Davis, Pan Deng, Greg Diamos, Jared Duke, Dave Fick, J. Scott Gardner, Itay Hubara, Sachin Idgunji, Thomas B. Jablin, Jeff Jiao, Tom St. John, Pankaj Kanwar, David Lee, Jeffery Liao, Anton Lokhmotov, Francisco Massa, Peng Meng, Paulius Micikevicius, Colin Osborne, Gennady Pekhimenko, Arun Tejusve Raghunath Rajan, Dilip Sequeira, Ashish Sirasao, Fei Sun, Hanlin Tang, Michael Thomson, Frank Wei, Ephrem Wu, Lingjie Xu, Koichi Yamada, Bing Yu, George Yuan, Aaron Zhong, Peizhao Zhang, and Yuchen Zhou. "MLPerf Inference Benchmark". In: *47th ACM/IEEE Annual International Symposium on Computer Architecture, ISCA 2020, Virtual Event / Valencia, Spain, May 30 - June 3, 2020*. IEEE, 2020, pp. 446–459. DOI: 10.1109/ISCA45697.2020.00045. URL: https://doi.org/10.1109/ISCA45697.2020.00045.

42 Daniel Richins, Dharmisha Doshi, Matthew Blackmore, Aswathy Thulaseedharan Nair, Neha Pathapati, Ankit Patel, Brainard Daguman, Daniel Dobrijalowski, Ramesh Illikkal, Kevin Long, David Zimmerman, and **Vijay Janapa Reddi**. "Missing the Forest for the Trees: End-to-End AI Application Performance in Edge Data Centers". In: *IEEE International Symposium on High Performance Computer Architecture, HPCA 2020, San Diego, CA, USA, February 22-26, 2020*. IEEE, 2020, pp. 515–528. DOI: 10.1109/HPCA47549.2020.00049. URL: https://doi.org/10.1109/HPCA47549.2020.00049.

43 Thierry Tambe, En-Yu Yang, Zishen Wan, Yuntian Deng, **Vijay Janapa Reddi**, Alexander M. Rush, David Brooks, and Gu-Yeon Wei. "Algorithm-Hardware Co-Design of Adaptive Floating-Point Encodings for Resilient Deep Learning Inference". In: *57th ACM/IEEE Design Automation Conference, DAC 2020, San Francisco, CA, USA, July 20-24, 2020*. IEEE, 2020, pp. 1–6. DOI: 10.1109/DAC18072.2020.9218516. URL: https://doi.org/10.1109/DAC18072.2020.9218516.

44 Wenzhi Cui, Daniel Richins, Yuhao Zhu, and **Vijay Janapa Reddi**. "Tail latency in node.js: energy efficient turbo boosting for long latency requests in event-driven web services". In: *Proceedings of the 15th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, VEE 2019, Providence, RI, USA, April 14, 2019*. ACM, 2019, pp. 152–164. DOI: 10.1145/3313808.3313823. URL: https://doi.org/10.1145/3313808.3313823.

45 Dimitris Gizopoulos, George Papadimitriou, Athanasios Chatzidimitriou, **Vijay Janapa Reddi**, Behzad Salami, Osman S. Unsal, Adrián Cristal Kestelman, and Jingwen Leng. "Modern Hardware Margins: CPUs, GPUs, FPGAs Recent System-Level Studies". In: *25th IEEE International Symposium on On-Line Testing and Robust System Design, IOLTS 2019, Rhodes, Greece, July 1-3, 2019*. IEEE, 2019, pp. 129–134. DOI: 10.1109/IOLTS.2019.8854386. URL: https://doi.org/10.1109/IOLTS.2019.8854386.

46 Matthew Halpern, Behzad Boroujerdian, Todd W. Mummert, Evelyn Duesterwald, and **Vijay Janapa Reddi**. "One Size Does Not Fit All: Quantifying and Exposing the Accuracy-Latency Trade-Off in Machine Learning Cloud Service APIs via Tolerance Tiers". In: *IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2019, Madison, WI, USA, March 24-26, 2019*. IEEE, 2019, pp. 34–47. DOI: 10.1109/ISPASS.2019.00012. URL: https://doi.org/10.1109/ISPASS.2019.00012.

47 Mark D. Hill and **Vijay Janapa Reddi**. "Gables: A Roofline Model for Mobile SoCs". In: *25th IEEE International Symposium on High Performance Computer Architecture, HPCA 2019, Washington, DC, USA, February 16-20, 2019*. IEEE, 2019, pp. 317–330. DOI: 10.1109/HPCA.2019.00047. URL: https://doi.org/10.1109/HPCA.2019.00047.

48 Yazhou Zu, Daniel Richins, Charles Lefurgy, and **Vijay Janapa Reddi**. "Fine-Tuning the Active Timing Margin (ATM) Control Loop for Maximizing Multi-core Efficiency on an IBM POWER Server". In: *25th IEEE International Symposium on High Performance Computer Architecture, HPCA 2019, Washington, DC, USA, February 16-20, 2019*. IEEE, 2019, pp. 106–119. DOI: 10.1109/HPCA.2019.00031. URL: https://doi.org/10.1109/HPCA.2019.00031.

49 Behzad Boroujerdian, Hasan Genc, Srivatsan Krishnan, Wenzhi Cui, Aleksandra Faust, and **Vijay Janapa Reddi**. "MAVBench: Micro Aerial Vehicle Benchmarking". In: *51st Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2018, Fukuoka, Japan, October 20-24, 2018*. IEEE Computer Society, 2018, pp. 894–907. DOI: 10.1109/MICRO.2018.00077. URL: https://doi.org/10.1109/MICRO.2018.00077.

50 Daniel Richins, Tahrina Ahmed, Russell M. Clapp, and **Vijay Janapa Reddi**. "Amdahl's Law in Big Data Analytics: Alive and Kicking in TPCx-BB (BigBench)". In: *IEEE International Symposium on High Performance Computer Architecture, HPCA 2018, Vienna, Austria, February 24-28, 2018*. IEEE Computer Society, 2018, pp. 630–642. DOI: 10.1109/HPCA.2018.00060. URL: https://doi.org/10.1109/HPCA.2018.00060.

51 An Zou, Jingwen Leng, Xin He, Yazhou Zu, Christopher D. Gill, **Vijay Janapa Reddi**, and Xuan Zhang. "Voltage-Stacked GPUs: A Control Theory Driven Cross-Layer Solution for Practical Voltage Stacking in GPUs". In: *51st Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2018, Fukuoka, Japan, October 20-24, 2018*. IEEE Computer Society, 2018, pp. 390–402. DOI: 10.1109/MICRO.2018.00039. URL: https://doi.org/10.1109/MICRO.2018.00039.

52 An Zou, Jingwen Leng, Xin He, Yazhou Zu, **Vijay Janapa Reddi**, and Xuan Zhang. "Efficient and reliable power delivery in voltage-stacked manycore system with hybrid charge-recycling regulators". In: *Proceedings of the 55th Annual Design Automation Conference, DAC 2018, San Francisco, CA, USA, June 24-29, 2018*. ACM, 2018, 43:1–43:6. DOI: 10.1145/3195970.3196037. URL: https://doi.org/10.1145/3195970.3196037.

53 An Zou, Jingwen Leng, Yazhou Zu, Tao Tong, **Vijay Janapa Reddi**, David M. Brooks, Gu-Yeon Wei, and Xuan Zhang. "Ivory: Early-Stage Design Space Exploration Tool for Integrated Voltage Regulators". In: *Proceedings of the 54th Annual Design Automation Conference, DAC 2017, Austin, TX, USA, June 18-22, 2017*. ACM, 2017, 1:1–1:6. DOI: 10.1145/3061639.3062268. URL: https://doi.org/10.1145/3061639.3062268.

54 Nadav Chachmon, Daniel Richins, Robert S. Cohn, Magnus Christensson, Wenzhi Cui, and **Vijay Janapa Reddi**. "Simulation and Analysis Engine for Scale-Out Workloads". In: *Proceedings of the 2016 International Conference on Supercomputing, ICS 2016, Istanbul, Turkey, June 1-3, 2016*. ACM, 2016, 22:1–22:13. DOI: 10.1145/2925426.2926293. URL: https://doi.org/10.1145/2925426.2926293.

55 Matthew Halpern, Yuhao Zhu, and **Vijay Janapa Reddi**. "Mobile CPU's rise to power: Quantifying the impact of generational mobile CPU design trends on performance, energy, and user satisfaction". In: *2016 IEEE International Symposium on High Performance Computer Architecture, HPCA 2016, Barcelona, Spain, March 12-16, 2016*. IEEE Computer Society, 2016, pp. 64–76. DOI: 10.1109/HPCA.2016.7446054. URL: https://doi.org/10.1109/HPCA.2016.7446054.

56 Mikhail Kazdagli, **Vijay Janapa Reddi**, and Mohit Tiwari. "Quantifying and improving the efficiency of hardware-based mobile malware detectors". In: *49th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2016, Taipei, Taiwan, October 15-19, 2016*. IEEE Computer Society, 2016, 37:1–37:13. DOI: 10.1109/MICRO.2016.7783740. URL: https://doi.org/10.1109/MICRO.2016.7783740.

57 Yuxi Liu, Zhibin Yu, Lieven Eeckhout, **Vijay Janapa Reddi**, Yingwei Luo, Xiaolin Wang, Zhenlin Wang, and Cheng-Zhong Xu. "Barrier-Aware Warp Scheduling for Throughput Processors". In: *Proceedings of the 2016 International Conference on Supercomputing, ICS 2016, Istanbul, Turkey, June 1-3, 2016*. ACM, 2016, 42:1–42:12. DOI: 10.1145/2925426.2926267. URL: https://doi.org/10.1145/2925426.2926267.

58 Yuhao Zhu and **Vijay Janapa Reddi**. "GreenWeb: language extensions for energy-efficient mobile web computing". In: *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2016, Santa Barbara, CA, USA, June 13-17, 2016*. ACM, 2016, pp. 145–160. DOI: 10.1145/2908080.2908082. URL: https://doi.org/10.1145/2908080.2908082.

59 Yazhou Zu, Wei Huang, Indrani Paul, and **Vijay Janapa Reddi**. "Ti-states: Processor power management in the temperature inversion region". In: *49th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2016, Taipei, Taiwan, October 15-19, 2016*. IEEE Computer Society, 2016, 55:1–55:13. DOI: 10.1109/MICRO.2016.7783758. URL: https://doi.org/10.1109/MICRO.2016.7783758.

60 Matthew Halpern, Yuhao Zhu, Ramesh Peri, and **Vijay Janapa Reddi**. "Mosaic: cross-platform user-interaction record and replay for the fragmented android ecosystem". In: *2015 IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2015, Philadelphia, PA, USA, March 29-31, 2015*. IEEE Computer Society, 2015, pp. 215–224. DOI: 10.1109/ISPASS.2015.7095807. URL: https://doi.org/10.1109/ISPASS.2015.7095807.

61 Jingwen Leng, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, and **Vijay Janapa Reddi**. "Safe limits on voltage reduction efficiency in GPUs: a direct measurement approach". In: *Proceedings of the 48th International Symposium on Microarchitecture, MICRO 2015, Waikiki, HI, USA, December 5-9, 2015*. ACM, 2015, pp. 294–307. DOI: 10.1145/2830772.2830811. URL: https://doi.org/10.1145/2830772.2830811.

62 Jingwen Leng, Yazhou Zu, and **Vijay Janapa Reddi**. "GPU voltage noise: Characterization and hierarchical smoothing of spatial and temporal voltage noise interference in GPU architectures". In: *21st IEEE International Symposium on High Performance Computer Architecture, HPCA 2015, Burlingame, CA, USA, February 7-11, 2015*. IEEE Computer Society, 2015, pp. 161–173. DOI: 10.1109/HPCA.2015.7056030. URL: https://doi.org/10.1109/HPCA.2015.7056030.

63 Yuhao Zhu, Matthew Halpern, and **Vijay Janapa Reddi**. "Event-based scheduling for energy-efficient QoS (eQoS) in mobile Web applications". In: *21st IEEE International Symposium on High Performance Computer Architecture, HPCA 2015, Burlingame, CA, USA, February 7-11, 2015*. IEEE Computer Society, 2015, pp. 137–149. DOI: 10.1109/HPCA.2015.7056028. URL: https://doi.org/10.1109/HPCA.2015.7056028.

64 Yuhao Zhu, Daniel Richins, Matthew Halpern, and **Vijay Janapa Reddi**. "Microarchitectural implications of event-driven server-side web applications". In: *Proceedings of the 48th International Symposium on Microarchitecture, MICRO 2015, Waikiki, HI, USA, December 5-9, 2015*. ACM, 2015, pp. 762–774. DOI: 10.1145/2830772.2830792. URL: https://doi.org/10.1145/2830772.2830792.

65 Yazhou Zu, Charles R. Lefurgy, Jingwen Leng, Matthew Halpern, Michael S. Floyd, and **Vijay Janapa Reddi**. "Adaptive guardband scheduling to improve system-level efficiency of the POWER7+". In: *Proceedings of the 48th International Symposium on Microarchitecture, MICRO 2015, Waikiki, HI, USA, December 5-9, 2015*. ACM, 2015, pp. 308–321. DOI: 10.1145/2830772.2830824. URL: https://doi.org/10.1145/2830772.2830824.

66 Mikhail Kazdagli, Ling Huang, **Vijay Janapa Reddi**, and Mohit Tiwari. "Morpheus: benchmarking computational diversity in mobile malware". In: *HASP 2014, Hardware and Architectural Support for Security and Privacy, Minneapolis, MN, USA, June 15, 2014*. ACM, 2014, 3:1–3:8. DOI: 10.1145/2611765.2611767. URL: https://doi.org/10.1145/2611765.2611767.

67 Jingwen Leng, Yazhou Zu, Minsoo Rhu, Meeta Sharma Gupta, and **Vijay Janapa Reddi**. "GPUVolt: modeling and characterizing voltage noise in GPU architectures". In: *International Symposium on Low Power Electronics and Design, ISLPED'14, La Jolla, CA, USA - August 11 - 13, 2014*. ACM, 2014, pp. 141–146. DOI: 10.1145/2627369.2627605. URL: https://doi.org/10.1145/2627369.2627605.

68 Yuhao Zhu and **Vijay Janapa Reddi**. "WebCore: Architectural support for mobile Web browsing". In: *ACM/IEEE 41st International Symposium on Computer Architecture, ISCA 2014, Minneapolis, MN, USA, June 14-18, 2014*. IEEE Computer Society, 2014, pp. 541–552. DOI: 10.1109/ISCA.2014.6853239. URL: https://doi.org/10.1109/ISCA.2014.6853239.

69 Jingwen Leng, Tayler H. Hetherington, Ahmed ElTantawy, Syed Zohaib Gilani, Nam Sung Kim, Tor M. Aamodt, and **Vijay Janapa Reddi**. "GPUWattch: enabling energy optimizations in GPGPUs". In: *The 40th Annual International Symposium on Computer Architecture, ISCA'13, Tel-Aviv, Israel, June 23-27, 2013*. ACM, 2013, pp. 487–498. DOI: 10.1145/2485922.2485964. URL: https://doi.org/10.1145/2485922.2485964.

70 Yuhao Zhu and **Vijay Janapa Reddi**. "High-performance and energy-efficient mobile web browsing on big/little systems". In: *19th IEEE International Symposium on High Performance Computer Architecture, HPCA 2013, Shenzhen, China, February 23-27, 2013*. IEEE Computer Society, 2013, pp. 13–24. DOI: 10.1109/HPCA.2013.6522303. URL: https://doi.org/10.1109/HPCA.2013.6522303.

71 Simone Campanoni, Timothy M. Jones, Glenn H. Holloway, **Vijay Janapa Reddi**, Gu-Yeon Wei, and David M. Brooks. "HELIX: automatic parallelization of irregular programs for chip multiprocessing". In: *10th Annual IEEE/ACM International Symposium on Code Generation and Optimization, CGO 2012, San Jose, CA, USA, March 31 - April 04, 2012*. ACM, 2012, pp. 84–93. DOI: 10.1145/2259016.2259028. URL: https://doi.org/10.1145/2259016.2259028.

72 **Vijay Janapa Reddi**. "Hardware and software co-design for robust and resilient execution". In: *2012 International Conference on Collaboration Technologies and Systems, CTS 2012, Denver, CO, USA, May 21-25, 2012*. IEEE, 2012, p. 380. DOI: 10.1109/CTS.2012.6261080. URL: https://doi.org/10.1109/CTS.2012.6261080.

73 **Vijay Janapa Reddi**, David Z. Pan, Sani R. Nassif, and Keith A. Bowman. "Robust and resilient designs from the bottom-up: Technology, CAD, circuit, and system issues". In: *Proceedings of the 17th Asia and South Pacific Design Automation Conference, ASP-DAC 2012, Sydney, Australia, January 30 - February 2, 2012*. IEEE, 2012, pp. 7–16. DOI: 10.1109/ASPDAC.2012.6165064. URL: https://doi.org/10.1109/ASPDAC.2012.6165064.

74 Peter Bailis, **Vijay Janapa Reddi**, Sanjay Gandhi, David M. Brooks, and Margo I. Seltzer. "Dimetrodon: processor-level preventive thermal management via idle cycle injection". In: *Proceedings of the 48th Design Automation Conference, DAC 2011, San Diego, California, USA, June 5-10, 2011*. ACM, 2011, pp. 89–94. DOI: 10.1145/2024724.2024745. URL: https://doi.org/10.1145/2024724.2024745.

75 **Vijay Janapa Reddi**, Svilen Kanev, Wonyoung Kim, Simone Campanoni, Michael D. Smith, Gu-Yeon Wei, and David M. Brooks. "Voltage Smoothing: Characterizing and Mitigating Voltage Noise in Production Processors via Software-Guided Thread Scheduling". In: *43rd Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2010, 4-8 December 2010, Atlanta, Georgia, USA*. IEEE Computer Society, 2010, pp. 77–88. DOI: 10.1109/MICRO.2010.35. URL: https://doi.org/10.1109/MICRO.2010.35.

76 **Vijay Janapa Reddi**, Benjamin C. Lee, Trishul M. Chilimbi, and Kushagra Vaid. "Web search using mobile cores: quantifying and mitigating the price of efficiency". In: *37th International Symposium on Computer Architecture (ISCA 2010), June 19-23, 2010, Saint-Malo, France*. ACM, 2010, pp. 314–325. DOI: 10.1145/1815961.1816002. URL: https://doi.org/10.1145/1815961.1816002.

77 Meeta Sharma Gupta, **Vijay Janapa Reddi**, Glenn H. Holloway, Gu-Yeon Wei, and David M. Brooks. "An event-guided approach to reducing voltage noise in processors". In: *Design, Automation and Test in Europe, DATE 2009, Nice, France, April 20-24, 2009*. IEEE, 2009, pp. 160–165. DOI: 10.1109/DATE.2009.5090651. URL: https://doi.org/10.1109/DATE.2009.5090651.

78 **Vijay Janapa Reddi**, Simone Campanoni, Meeta Sharma Gupta, Michael D. Smith, Gu-Yeon Wei, and David M. Brooks. "Software-assisted hardware reliability: abstracting circuit-level challenges to the software stack". In: *Proceedings of the 46th Design Automation Conference, DAC 2009, San Francisco, CA, USA, July 26-31, 2009*. ACM, 2009, pp. 788–793. DOI: 10.1145/1629911.1630114. URL: https://doi.org/10.1145/1629911.1630114.

79 **Vijay Janapa Reddi**, Meeta Sharma Gupta, Glenn H. Holloway, Gu-Yeon Wei, Michael D. Smith, and David M. Brooks. "Voltage emergency prediction: Using signatures to reduce operating margins". In: *15th International Conference on High-Performance Computer Architecture (HPCA-15 2009), 14-18 February 2009, Raleigh, North Carolina, USA*. IEEE Computer Society, 2009, pp. 18–29. DOI: 10.1109/HPCA.2009.4798233. URL: https://doi.org/10.1109/HPCA.2009.4798233.

80 Tipp Moseley, Alex Shye, **Vijay Janapa Reddi**, Dirk Grunwald, and Ramesh Peri. "Shadow Profiling: Hiding Instrumentation Costs with Parallelism". In: *Fifth International Symposium on Code Generation and Optimization (CGO 2007), 11-14 March 2007, San Jose, California, USA*. IEEE Computer Society, 2007, pp. 198–208. DOI: 10.1109/CGO.2007.35. URL: https://doi.org/10.1109/CGO.2007.35.

81 **Vijay Janapa Reddi**, Dan Connors, Robert Cohn, and Michael D. Smith. "Persistent Code Caching: Exploiting Code Reuse Across Executions and Applications". In: *Fifth International Symposium on Code Generation and Optimization (CGO 2007), 11-14 March 2007, San Jose, California, USA*. IEEE Computer Society, 2007, pp. 74–88. DOI: 10.1109/CGO.2007.29. URL: https://doi.org/10.1109/CGO.2007.29.

82 Alex Shye, Tipp Moseley, **Vijay Janapa Reddi**, Joseph Blomstedt, and Daniel A. Connors. "Using Process-Level Redundancy to Exploit Multiple Cores for Transient Fault Tolerance". In: *The 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2007, 25-28 June 2007, Edinburgh, UK, Proceedings*. IEEE Computer Society, 2007, pp. 297–306. DOI: 10.1109/DSN.2007.98. URL: https://doi.org/10.1109/DSN.2007.98.

83 Silvia M. Figueira and **Vijay Janapa Reddi**. "Topology-Based Hypercube Structures for Global Communication in Heterogeneous Networks". In: *Euro-Par 2005, Parallel Processing, 11th International Euro-Par Conference, Lisbon, Portugal, August 30 - September 2, 2005, Proceedings*. Vol. 3648. Lecture Notes in Computer Science. Springer, 2005, pp. 994–1004. DOI: 10.1007/11549468_109. URL: https://doi.org/10.1007/11549468%5C_109.

84 Chi-Keung Luk, Robert S. Cohn, Robert Muth, Harish Patil, Artur Klauser, P. Geoffrey Lowney, Steven Wallace, **Vijay Janapa Reddi**, and Kim M. Hazelwood. "Pin: building customized program analysis tools with dynamic instrumentation". In: *Proceedings of the ACM SIGPLAN 2005 Conference on Programming Language Design and Implementation, Chicago, IL, USA, June 12-15, 2005*. ACM, 2005, pp. 190–200. DOI: 10.1145/1065010.1065034. URL: https://doi.org/10.1145/1065010.1065034.

85 Tipp Moseley, Alex Shye, **Vijay Janapa Reddi**, Matthew Iyer, Dan Fay, David Hodgdon, Joshua L. Kihm, Alex Settle, Dirk Grunwald, and Daniel A. Connors. "Dynamic run-time architecture techniques for enabling continuous optimization". In: *Proceedings of the Second Conference on Computing Frontiers, 2005, Ischia, Italy, May 4-6, 2005*. ACM, 2005, pp. 211–220. DOI: 10.1145/1062261.1062296. URL: https://doi.org/10.1145/1062261.1062296.

86 Qiang Wu, Margaret Martonosi, Douglas W. Clark, **Vijay Janapa Reddi**, Dan Connors, Youfeng Wu, Jin Lee, and David M. Brooks. "A Dynamic Compilation Framework for Controlling Microprocessor Energy and Performance". In: *38th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-38 2005), 12-16 November 2005, Barcelona, Spain*. IEEE Computer Society, 2005, pp. 271–282. DOI: 10.1109/MICRO.2005.7. URL: https://doi.org/10.1109/MICRO.2005.7.

Journal Publications

1 Behzad Boroujerdian, Ying Jing, Devashree Tripathy, Amit Kumar, Lavanya Subramanian, Luke Yen, Vincent Lee, Vivek Venkatesan, Amit Jindal, Robert Shearer, and **Vijay Janapa Reddi**. "FARSI: An Early-stage Design Space Exploration Framework to Tame the Domain-specific System-on-chip Complexity". In: *ACM Trans. Embed. Comput. Syst.* 22.2 (2023), 31:1–31:35. DOI: 10.1145/3544016. URL: https://doi.org/10.1145/3544016.

2 Cansu Demirkiran, Furkan Eris, Gongyu Wang, Jonathan Elmhurst, Nick Moore, Nicholas C. Harris, Ayon Basumallik, **Vijay Janapa Reddi**, Ajay Joshi, and Darius Bunandar. "An Electro-Photonic System for Accelerating Deep Neural Networks". In: *ACM J. Emerg. Technol. Comput. Syst.* 19.4 (2023), 30:1–30:31. DOI: 10.1145/3606949. URL: https://doi.org/10.1145/3606949.

3 Alexandros Karargyris, Renato Umeton, Micah J. Sheller, Alejandro Aristizabal, Johnu George, Anna Wuest, Sarthak Pati, Hasan Kassem, Maximilian Zenk, Ujjwal Baid, Prakash Narayana Moorthy, Alexander Chowdhury, Junyi Guo, Sahil S. Nalawade, Jacob Rosenthal, David Kanter, Maria Xenochristou, Daniel J. Beutel, Verena Chung, Timothy Bergquist, James A. Eddy, Abubakar Abid, Lewis Tunstall, Omar Sanseviero, Dimitrios Dimitriadis, Yiming Qian, Xinxing Xu, Yong Liu, Rick Siow Mong Goh, Srini Bala, Victor Bittorf, Sreekar Reddy Puchala, Biagio Ricciuti, Soujanya Samineni, Eshna Sengupta, Akshay Chaudhari, Cody Coleman, Bala Desinghu, Gregory F. Diamos, Debo Dutta, Diane Feddema, Grigori Fursin, Xinyuan Huang, Satyananda Kashyap, Nicholas D. Lane, Indranil Mallick, Pietro Mascagni, Virendra Mehta, Cassiano Ferro Moraes, Vivek Natarajan, Nikola Nikolov, Nicolas Padoy, Gennady Pekhimenko, **Vijay Janapa Reddi**, G. Anthony Reina, Pablo Ribalta, Abhishek Singh, Jayaraman J. Thiagarajan, Jacob Albrecht, Thomas Wolf, Geralyn Miller, Huazhu Fu, Prashant Shah, Daguang Xu, Poonam Yadav, David Talby, Mark M. Awad, Jeremy P. Howard, Michael Rosenthal, Luigi Marchionni, Massimo Loda, Jason M. Johnson, Spyridon Bakas, and Peter Mattson. "Federated benchmarking of medical artificial intelligence with MedPerf". In: *Nat. Mac. Intell.* 5.7 (2023), pp. 799–810. DOI: 10.1038/S42256-023-00652-2. URL: https://doi.org/10.1038/s42256-023-00652-2.

4 Thanh Thi Nguyen and **Vijay Janapa Reddi**. "Deep Reinforcement Learning for Cyber Security". In: *IEEE Trans. Neural Networks Learn. Syst.* 34.8 (2023), pp. 3779–3795. DOI: 10.1109/TNNLS.2021.3121870. URL: https://doi.org/10.1109/TNNLS.2021.3121870.

5 Shvetank Prakash, Matthew Stewart, Colby R. Banbury, Mark Mazumder, Pete Warden, Brian Plancher, and **Vijay Janapa Reddi**. "Is TinyML Sustainable?" In: *Commun. ACM* 66.11 (2023), pp. 68–77. DOI: 10.1145/3608473. URL: https://doi.org/10.1145/3608473.

6 **Vijay Janapa Reddi** and Boris Murmann. "Special Issue on TinyML". In: *IEEE Micro* 43.6 (2023), pp. 7–10. DOI: 10.1109/MM.2023.3322048. URL: https://doi.org/10.1109/MM.2023.3322048.

7 Pete Warden, Matthew Stewart, Brian Plancher, Sachin Katti, and **Vijay Janapa Reddi**. "Machine Learning Sensors". In: *Commun. ACM* 66.11 (2023), pp. 25–28. DOI: 10.1145/3586991. URL: https://doi.org/10.1145/3586991.

8 Sathwika Bavikadi, Abhijitt Dhavlle, Amlan Ganguly, Anand Haridass, Hagar Hendy, Cory E. Merkel, **Vijay Janapa Reddi**, Purab Ranjan Sutradhar, Arun Joseph, and Sai Manoj Pudukotai Dinakarrao. "A Survey on Machine Learning Accelerators and Evolutionary Hardware Platforms". In: *IEEE Des. Test* 39.3 (2022), pp. 91–116. DOI: 10.1109/MDAT.2022.3161126. URL: https://doi.org/10.1109/MDAT.2022.3161126.

9 Srivatsan Krishnan, Max Lam, Sharad Chitlangia, Zishen Wan, Gabriel Barth-Maron, Aleksandra Faust, and **Vijay Janapa Reddi**. "QuaRL: Quantization for Fast and Environmentally Sustainable Reinforcement Learning". In: *Trans. Mach. Learn. Res.* 2022 (2022). URL: https://openreview.net/forum?id=xwWsiFmUEs.

10 Behzad Boroujerdian, Hasan Genc, Srivatsan Krishnan, Bardienus Pieter Duisterhof, Brian Plancher, Kayvan Mansoorshahi, Marcelino Almeida, Wenzhi Cui, Aleksandra Faust, and **Vijay Janapa Reddi**. "The Role of Compute in Autonomous Micro Aerial Vehicles: Optimizing for Mission Time and Energy Efficiency". In: *ACM Trans. Comput. Syst.* 39.1-4 (2021), 3:1–3:44. DOI: 10.1145/3511210. URL: https://doi.org/10.1145/3511210.

11 Mark D. Hill and **Vijay Janapa Reddi**. "Accelerator-level parallelism". In: *Commun. ACM* 64.12 (2021), pp. 36–38. DOI: 10.1145/3460970. URL: https://doi.org/10.1145/3460970.

12 Srivatsan Krishnan, Behzad Boroujerdian, William Fu, Aleksandra Faust, and **Vijay Janapa Reddi**. "Air Learning: a deep reinforcement learning gym for autonomous aerial robot visual navigation". In: *Mach. Learn.* 110.9 (2021), pp. 2501–2540. DOI: 10.1007/S10994-021-06006-6. URL: https://doi.org/10.1007/s10994-021-06006-6.

13 Jingwen Leng, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, Yazhou Zu, and **Vijay Janapa Reddi**. "Erratum to "Predictive Guardbanding: Program-Driven Timing Margin Reduction for GPUs"". In: *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 40.6 (2021), p. 1272. DOI: 10.1109/TCAD.2021.3058491. URL: https://doi.org/10.1109/TCAD.2021.3058491.

14 Jingwen Leng, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, Yazhou Zu, and **Vijay Janapa Reddi**. "Predictive Guardbanding: Program-Driven Timing Margin Reduction for GPUs". In: *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 40.1 (2021), pp. 171–184. DOI: 10.1109/TCAD.2020.2992684. URL: https://doi.org/10.1109/TCAD.2020.2992684.

15 Brian Plancher, Sabrina M. Neuman, Thomas Bourgeat, Scott Kuindersma, Srinivas Devadas, and **Vijay Janapa Reddi**. "Accelerating Robot Dynamics Gradients on a CPU, GPU, and FPGA". In: *IEEE Robotics Autom. Lett.* 6.2 (2021), pp. 2335–2342. DOI: 10.1109/LRA.2021.3057845. URL: https://doi.org/10.1109/LRA.2021.3057845.

16 **Vijay Janapa Reddi**, Christine Cheng, David Kanter, Peter Mattson, Guenther Schmuelling, and Carole-Jean Wu. "The Vision Behind MLPerf: Understanding AI Inference Performance". In: *IEEE Micro* 41.3 (2021), pp. 10–18. DOI: 10.1109/MM.2021.3066343. URL: https://doi.org/10.1109/MM.2021.3066343.

17 An Zou, Huifeng Zhu, Jingwen Leng, Xin He, **Vijay Janapa Reddi**, Christopher D. Gill, and Xuan Zhang. "System-level Early-stage Modeling and Evaluation of IVR-assisted Processor Power Delivery System". In: *ACM Trans. Archit. Code Optim.* 18.4 (2021), 52:1–52:27. DOI: 10.1145/3468145. URL: https://doi.org/10.1145/3468145.

18 Srivatsan Krishnan, Zishen Wan, Kshitij Bhardwaj, Paul N. Whatmough, Aleksandra Faust, Gu-Yeon Wei, David Brooks, and **Vijay Janapa Reddi**. "The Sky Is Not the Limit: A Visual Performance Model for Cyber-Physical Co-Design in Autonomous Machines". In: *IEEE Comput. Archit. Lett.* 19.1 (2020), pp. 38–42. DOI: 10.1109/LCA.2020.2981022. URL: https://doi.org/10.1109/LCA.2020.2981022.

19 Peter Mattson, Hanlin Tang, Gu-Yeon Wei, Carole-Jean Wu, **Vijay Janapa Reddi**, Christine Cheng, Cody Coleman, Greg Diamos, David Kanter, Paulius Micikevicius, David A. Patterson, and Guenther Schmuelling. "MLPerf: An Industry Standard Benchmark Suite for Machine Learning Performance". In: *IEEE Micro* 40.2 (2020), pp. 8–16. DOI: 10.1109/MM.2020.2974843. URL: https://doi.org/10.1109/MM.2020.2974843.

20 An Zou, Jingwen Leng, Xin He, Yazhou Zu, Christopher D. Gill, **Vijay Janapa Reddi**, and Xuan Zhang. "Voltage-Stacked Power Delivery Systems: Reliability, Efficiency, and Power Management". In: *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 39.12 (2020), pp. 5142–5155. DOI: 10.1109/TCAD.2020.2969607. URL: https://doi.org/10.1109/TCAD.2020.2969607.

21 Jingwen Leng, Alper Buyuktosunoglu, Ramon Bertran, Pradip Bose, and **Vijay Janapa Reddi**. "Asymmetric Resilience for Accelerator-Rich Systems". In: *IEEE Comput. Archit. Lett.* 18.1 (2019), pp. 83–86. DOI: 10.1109/LCA.2019.2917898. URL: https://doi.org/10.1109/LCA.2019.2917898.

22 Daniel Richins, Dharmisha Doshi, Matthew Blackmore, Aswathy Thulaseedharan Nair, Neha Pathapati, Ankit Patel, Brainard Daguman, Daniel Dobrijalowski, Ramesh Illikkal, Kevin Long, David Zimmerman, and **Vijay Janapa Reddi**. "AI Tax: The Hidden Cost of AI Data Center Applications". In: *ACM Trans. Comput. Syst.* 37.1-4 (2019), 3:1–3:32. DOI: 10.1145/3440689. URL: https://doi.org/10.1145/3440689.

23 Ting-Wu Chin, Chia-Lin Yu, Matthew Halpern, Hasan Genc, Shiao-Li Tsao, and **Vijay Janapa Reddi**. "Domain-Specific Approximation for Object Detection". In: *IEEE Micro* 38.1 (2018), pp. 31–40. DOI: 10.1109/MM.2018.112130335. URL: https://doi.org/10.1109/MM.2018.112130335.

24 **Vijay Janapa Reddi**, Hongil Yoon, and Allan Knies. "Two Billion Devices and Counting". In: *IEEE Micro* 38.1 (2018), pp. 6–21. DOI: 10.1109/MM.2018.011441560. URL: https://doi.org/10.1109/MM.2018.011441560.

25 Peter Bailis, Jean Yang, **Vijay Janapa Reddi**, and Yuhao Zhu. "Research for practice: web security and mobile web computing". In: *Commun. ACM* 60.1 (2017), pp. 50–53. DOI: 10.1145/2980989. URL: https://doi.org/10.1145/2980989.

26 Hasan Genc, Yazhou Zu, Ting-Wu Chin, Matthew Halpern, and **Vijay Janapa Reddi**. "Flying IoT: Toward Low-Power Vision in the Sky". In: *IEEE Micro* 37.6 (2017), pp. 40–51. DOI: 10.1109/MM.2017.4241339. URL: https://doi.org/10.1109/MM.2017.4241339.

27 Yuhao Zhu and **Vijay Janapa Reddi**. "Optimizing General-Purpose CPUs for Energy-Efficient Mobile Web Computing". In: *ACM Trans. Comput. Syst.* 35.1 (2017), 1:1–1:31. DOI: 10.1145/3041024. URL: https://doi.org/10.1145/3041024.

28 Yuhao Zhu, **Vijay Janapa Reddi**, Robert Adolf, Saketh Rama, Brandon Reagen, Gu-Yeon Wei, and David M. Brooks. "Cognitive Computing Safety: The New Horizon for Reliability / The Design and Evolution of Deep Learning Workloads". In: *IEEE Micro* 37.1 (2017), pp. 15–21. DOI: 10.1109/MM.2017.2. URL: https://doi.org/10.1109/MM.2017.2.

29 Yazhou Zu, Wei Huang, Indrani Paul, and **Vijay Janapa Reddi**. "Ti-States: Power Management in Active Timing Margin Processors". In: *IEEE Micro* 37.3 (2017), pp. 106–114. DOI: 10.1109/MM.2017.68. URL: https://doi.org/10.1109/MM.2017.68.

30 **Vijay Janapa Reddi** and Hyesoon Kim. "On the Internet of Things". In: *IEEE Micro* 36.6 (2016), pp. 5–7. DOI: 10.1109/MM.2016.92. URL: https://doi.org/10.1109/MM.2016.92.

31 Jean Yang, **Vijay Janapa Reddi**, Yuhao Zhu, and Peter Bailis. "Research for Practice: Web Security and Mobile Web Computing". In: *ACM Queue* 14.4 (2016), p. 80. DOI: 10.1145/2984629.3005356. URL: https://doi.org/10.1145/2984629.3005356.

32 Yuhao Zhu, Matthew Halpern, and **Vijay Janapa Reddi**. "The Role of the CPU in Energy-Efficient Mobile Web Browsing". In: *IEEE Micro* 35.1 (2015), pp. 26–33. DOI: 10.1109/MM.2015.8. URL: https://doi.org/10.1109/MM.2015.8.

33 Yuhao Zhu, Aditya Srikanth, Jingwen Leng, and **Vijay Janapa Reddi**. "Exploiting Webpage Characteristics for Energy-Efficient Mobile Web Browsing". In: *IEEE Comput. Archit. Lett.* 13.1 (2014), pp. 33–36. DOI: 10.1109/L-CA.2012.33. URL: https://doi.org/10.1109/L-CA.2012.33.

34 **Vijay Janapa Reddi**. "Reliability-Aware Microarchitecture Design". In: *IEEE Micro* 33.4 (2013), pp. 4–5. DOI: 10.1109/MM.2013.87. URL: https://doi.org/10.1109/MM.2013.87.

35 **Vijay Janapa Reddi** and David M. Brooks. "Resilient Architectures via Collaborative Design: Maximizing Commodity Processor Performance in the Presence of Variations". In: *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 30.10 (2011), pp. 1429–1445. DOI: 10.1109/TCAD.2011.2163635. URL: https://doi.org/10.1109/TCAD.2011.2163635.

36 **Vijay Janapa Reddi**, Svilen Kanev, Wonyoung Kim, Simone Campanoni, Michael D. Smith, Gu-Yeon Wei, and David M. Brooks. "Voltage Noise in Production Processors". In: *IEEE Micro* 31.1 (2011), pp. 20–28. DOI: 10.1109/MM.2010.104. URL: https://doi.org/10.1109/MM.2010.104.

37 **Vijay Janapa Reddi**, Benjamin C. Lee, Trishul M. Chilimbi, and Kushagra Vaid. "Mobile processors for energy-efficient web search". In: *ACM Trans. Comput. Syst.* 29.3 (2011), 9:1–9:39. DOI: 10.1145/2003690.2003693. URL: https://doi.org/10.1145/2003690.2003693.

38 **Vijay Janapa Reddi**, Simone Campanoni, Meeta Sharma Gupta, Michael D. Smith, Gu-Yeon Wei, David M. Brooks, and Kim M. Hazelwood. "Eliminating voltage emergencies via software-guided code transformations". In: *ACM Trans. Archit. Code Optim.* 7.2 (2010), 12:1–12:28. DOI: 10.1145/1839667.1839674. URL: https://doi.org/10.1145/1839667.1839674.

39 **Vijay Janapa Reddi**, Meeta Sharma Gupta, Glenn H. Holloway, Michael D. Smith, Gu-Yeon Wei, and David M. Brooks. "Predicting Voltage Droops Using Recurring Program and Microarchitectural Event Activity". In: *IEEE Micro* 30.1 (2010), p. 110. DOI: 10.1109/MM.2010.25. URL: https://doi.org/10.1109/MM.2010.25.

40 Alex Shye, Joseph Blomstedt, Tipp Moseley, **Vijay Janapa Reddi**, and Daniel A. Connors. "PLR: A Software Approach to Transient Fault Tolerance for Multicore Architectures". In: *IEEE Trans. Dependable Secur. Comput.* 6.2 (2009), pp. 135–148. DOI: 10.1109/TDSC.2008.62. URL: https://doi.org/10.1109/TDSC.2008.62.

41 Qiang Wu, Margaret Martonosi, Douglas W. Clark, **Vijay Janapa Reddi**, Dan Connors, Youfeng Wu, Jin Lee, and David M. Brooks. "Dynamic-Compiler-Driven Control for Microprocessor Energy and Performance". In: *IEEE Micro* 26.1 (2006), pp. 119–129. DOI: 10.1109/MM.2006.9. URL: https://doi.org/10.1109/MM.2006.9.

42 **Vijay Janapa Reddi**, Dan Connors, and Robert S. Cohn. "Persistence in dynamic code transformation systems". In: *SIGARCH Comput. Archit. News* 33.5 (2005), pp. 69–74. DOI: 10.1145/1127577.1127591. URL: https://doi.org/10.1145/1127577.1127591.

ArXiV Preprints

1 Matthew Stewart, Emanuel Moss, Pete Warden, Brian Plancher, Susan Kennedy, Mona Sloane, and **Vijay Janapa Reddi**. "Materiality and Risk in the Age of Pervasive AI Sensors". In: *CoRR* abs/2402.11183 (2024). DOI: 10.48550/ARXIV.2402.11183. arXiv: 2402.11183. URL: https://doi.org/10.48550/arXiv.2402.11183.

2 Cansu Demirkiran, Rashmi Agrawal, **Vijay Janapa Reddi**, Darius Bunandar, and Ajay Joshi. "Leveraging Residue Number System for Designing High-Precision Analog Deep Neural Network Accelerators". In: *CoRR* abs/2306.09481 (2023). DOI: 10.48550/ARXIV.2306.09481. arXiv: 2306.09481. URL: https://doi.org/10.48550/arXiv.2306.09481.

3 Yu-Shun Hsiao, Siva Kumar Sastry Hari, Balakumar Sundaralingam, Jason Yik, Thierry Tambe, Charbel Sakr, Stephen W. Keckler, and **Vijay Janapa Reddi**. "VaPr: Variable-Precision Tensors to Accelerate Robot Motion Planning". In: *CoRR* abs/2310.07854 (2023). DOI: 10.48550/ARXIV.2310.07854. arXiv: 2310.07854. URL: https://doi.org/10.48550/arXiv.2310.07854.

4 Srivatsan Krishnan, Amir Yazdanbakhsh, Shvetank Prakash, Jason Jabbour, Ikechukwu Uchendu, Susobhan Ghosh, Behzad Boroujerdian, Daniel Richins, Devashree Tripathy, Aleksandra Faust, and **Vijay Janapa Reddi**. "ArchGym: An Open-Source Gymnasium for Machine Learning Assisted Architecture Design". In: *CoRR* abs/2306.08888 (2023). DOI: 10.48550/ARXIV.2306.08888. arXiv: 2306.08888. URL: https://doi.org/10.48550/arXiv.2306.08888.

5 Maximilian Lam, Jeff Johnson, Wenjie Xiong, Kiwan Maeng, Udit Gupta, Yang Li, Liangzhen Lai, Ilias Leontiadis, Minsoo Rhu, Hsien-Hsin S. Lee, **Vijay Janapa Reddi**, Gu-Yeon Wei, David Brooks, and G. Edward Suh. "GPU-based Private Information Retrieval for On-Device Machine Learning Inference". In: *CoRR* abs/2301.10904 (2023). DOI: 10.48550/ARXIV.2301.10904. arXiv: 2301.10904. URL: https://doi.org/10.48550/arXiv.2301.10904.

6 Luis Oala, Manil Maskey, Lilith Bat-Leah, Alicia Parrish, Nezihe Merve Gürel, Tzu-Sheng Kuo, Yang Liu, Rotem Dror, Danilo Brajovic, Xiaozhe Yao, Max Bartolo, William Gaviria Rojas, Ryan Hileman, Rainier Aliment, Michael W. Mahoney, Meg Risdal, Matthew Lease, Wojciech Samek, Debojyoti Dutta, Curtis G. Northcutt, Cody Coleman, Braden Hancock, Bernard Koch, Girmaw Abebe Tadesse, Bojan Karlas, Ahmed M. Alaa, Adji Bousso Dieng, Natasha F. Noy, **Vijay Janapa Reddi**, James Zou, Praveen K. Paritosh, Mihaela van der Schaar, Kurt D. Bollacker, Lora Aroyo, Ce Zhang, Joaquin Vanschoren, Isabelle Guyon, and Peter Mattson. "DMLR: Data-centric Machine Learning Research - Past, Present and Future". In: *CoRR* abs/2311.13028 (2023). DOI: 10.48550/ARXIV.2311.13028. arXiv: 2311.13028. URL: https://doi.org/10.48550/arXiv.2311.13028.

7 Alicia Parrish, Hannah Rose Kirk, Jessica Quaye, Charvi Rastogi, Max Bartolo, Oana Inel, Juan Ciro, Rafael Mosquera, Addison Howard, Will Cukierski, D. Sculley, **Vijay Janapa Reddi**, and Lora Aroyo. "Adversarial Nibbler: A Data-Centric Challenge for Improving the Safety of Text-to-Image Models". In: *CoRR* abs/2305.14384 (2023). DOI: 10.48550/ARXIV.2305.14384. arXiv: 2305.14384. URL: https://doi.org/10.48550/arXiv.2305.14384.

8 Shvetank Prakash, Matthew Stewart, Colby R. Banbury, Mark Mazumder, Pete Warden, Brian Plancher, and **Vijay Janapa Reddi**. "Is TinyML Sustainable? Assessing the Environmental Impacts of Machine Learning on Microcontrollers". In: *CoRR* abs/2301.11899 (2023). DOI: 10.48550/ARXIV.2301.11899. arXiv: 2301.11899. URL: https://doi.org/10.48550/arXiv.2301.11899.

9 Matthew Stewart, Pete Warden, Yasmine Omri, Shvetank Prakash, Joao Santos, Shawn Hymel, Benjamin Brown, Jim MacArthur, Nat Jeffries, Brian Plancher, and **Vijay Janapa Reddi**. "Datasheets for Machine Learning Sensors". In: *CoRR* abs/2306.08848 (2023). DOI: 10.48550/ARXIV.2306.08848. arXiv: 2306.08848. URL: https://doi.org/10.48550/arXiv.2306.08848.

10 Víctor Mayoral Vilches, Jason Jabbour, Yu-Shun Hsiao, Zishen Wan, Alejandra Martínez-Fariña, Martiño Crespo-Álvarez, Matthew Stewart, Juan Manuel Reina-Muñoz, Prateek Nagras, Gaurav Vikhe, Mohammad Bakhshalipour, Martin Pinzger, Stefan Rass, Smruti Panigrahi, Giulio Corradi, Niladri Roy, Phillip B. Gibbons, Sabrina M. Neuman, Brian Plancher, and **Vijay Janapa Reddi**. "RobotPerf: An Open-Source, Vendor-Agnostic, Benchmarking Suite for Evaluating Robotics Computing System Performance". In: *CoRR* abs/2309.09212 (2023). DOI: 10.48550/ARXIV.2309.09212. arXiv: 2309.09212. URL: https://doi.org/10.48550/arXiv.2309.09212.

11 Zishen Wan, Nandhini Chandramoorthy, Karthik Swaminathan, Pin-Yu Chen, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "BERRY: Bit Error Robustness for Energy-Efficient Reinforcement Learning-Based Autonomous Systems". In: *CoRR* abs/2307.10041 (2023). DOI: 10.48550/ARXIV.2307.10041. arXiv: 2307.10041. URL: https://doi.org/10.48550/arXiv.2307.10041.

12 Jason Yik, Soikat Hasan Ahmed, Zergham Ahmed, Brian Anderson, Andreas G. Andreou, Chiara Bartolozzi, Arindam Basu, Douwe den Blanken, Petrut Bogdan, Sander M. Bohté, Younes Bouhadjar, Sonia M. Buckley, Gert Cauwenberghs, Federico Corradi, Guido de Croon, Andreea Danielescu, Anurag Reddy Daram, Mike Davies, Yigit Demirag, Jason Eshraghian, Jeremy Forest, Steve B. Furber, Michael Furlong, Aditya Gilra, Giacomo Indiveri, Siddharth Joshi, Vedant Karia, Lyes Khacef, James C. Knight, Laura Kriener, Rajkumar Kubendran, Dhireesha Kudithipudi, Gregor Lenz, Rajit Manohar, Christian Mayr, Konstantinos P. Michmizos, Dylan R. Muir, Emre Neftci, Thomas Nowotny, Fabrizio Ottati, Ayça Özcelikkale, Noah Pacik-Nelson, Priyadarshini Panda, Pao-Sheng Sun, Melika Pay-

vand, Christian Pehle, Mihai A. Petrovici, Christoph Posch, Alpha Renner, Yulia Sandamirskaya, Clemens JS Schaefer, André van Schaik, Johannes Schemmel, Catherine D. Schuman, Jae-sun Seo, Sumit Bam Shrestha, Manolis Sifalakis, Amos Sironi, Kenneth Michael Stewart, Terrence C. Stewart, Philipp Stratmann, Guangzhi Tang, Jonathan Timcheck, Marian Verhelst, Craig M. Vineyard, Bernhard Vogginger, Amirreza Yousefzadeh, Biyan Zhou, Fatima Tuz Zohora, Charlotte Frenkel, and **Vijay Janapa Reddi**. "NeuroBench: Advancing Neuromorphic Computing through Collaborative, Fair and Representative Benchmarking". In: *CoRR* abs/2304.04640 (2023). DOI: 10.48550/ARXIV.2304.04640. arXiv: 2304.04640. URL: https://doi.org/10.48550/arXiv.2304.04640.

13 Behzad Boroujerdian, Ying Jing, Amit Kumar, Lavanya Subramanian, Luke Yen, Vincent Lee, Vivek Venkatesan, Amit Jindal, Robert Shearer, and **Vijay Janapa Reddi**. "FARSI: Facebook AR System Investigator for Agile Domain-Specific System-on-Chip Exploration". In: *CoRR* abs/2201.05232 (2022). arXiv: 2201.05232. URL: https://arxiv.org/abs/2201.05232.

14 Javier M. Duarte, Nhan Tran, Benjamin Hawks, Christian Herwig, Jules Muhizi, Shvetank Prakash, and **Vijay Janapa Reddi**. "FastML Science Benchmarks: Accelerating Real-Time Scientific Edge Machine Learning". In: *CoRR* abs/2207.07958 (2022). DOI: 10.48550/ARXIV.2207.07958. arXiv: 2207.07958. URL: https://doi.org/10.48550/arXiv.2207.07958.

15 Yu-Shun Hsiao, Siva Kumar Sastry Hari, Michal Filipiuk, Timothy Tsai, Michael B. Sullivan, **Vijay Janapa Reddi**, Vasu Singh, and Stephen W. Keckler. "Zhuyi: Perception Processing Rate Estimation for Safety in Autonomous Vehicles". In: *CoRR* abs/2205.03347 (2022). DOI: 10.48550/ARXIV.2205.03347. arXiv: 2205.03347. URL: https://doi.org/10.48550/arXiv.2205.03347.

16 Shawn Hymel, Colby R. Banbury, Daniel Situnayake, Alex Elium, Carl Ward, Mat Kelcey, Mathijs Baaijens, Mateusz Majchrzycki, Jenny Plunkett, David Tischler, Alessandro Grande, Louis Moreau, Dmitry Maslov, Artie Beavis, Jan Jongboom, and **Vijay Janapa Reddi**. "Edge Impulse: An MLOps Platform for Tiny Machine Learning". In: *CoRR* abs/2212.03332 (2022). DOI: 10.48550/ARXIV.2212.03332. arXiv: 2212.03332. URL: https://doi.org/10.48550/arXiv.2212.03332.

17 Tianyu Jia, En-Yu Yang, Yu-Shun Hsiao, Jonathan J. Cruz, David Brooks, Gu-Yeon Wei, and **Vijay Janapa Reddi**. "OMU: A Probabilistic 3D Occupancy Mapping Accelerator for Real-time OctoMap at the Edge". In: *CoRR* abs/2205.03325 (2022). DOI: 10.48550/ARXIV.2205.03325. arXiv: 2205.03325. URL: https://doi.org/10.48550/arXiv.2205.03325.

18 Srivatsan Krishnan, Natasha Jaques, Shayegan Omidshafiei, Dan Zhang, Izzeddin Gur, **Vijay Janapa Reddi**, and Aleksandra Faust. "Multi-Agent Reinforcement Learning for Microprocessor Design Space Exploration". In: *CoRR* abs/2211.16385 (2022). DOI: 10.48550/ARXIV.2211.16385. arXiv: 2211.16385. URL: https://doi.org/10.48550/arXiv.2211.16385.

19 Srivatsan Krishnan, Zishen Wan, Kshitij Bhardwaj, Ninad Jadhav, Aleksandra Faust, and **Vijay Janapa Reddi**. "Roofline Model for UAVs: A Bottleneck Analysis Tool for Onboard Compute Characterization of Autonomous Unmanned Aerial Vehicles". In: *CoRR* abs/2204.10898 (2022). DOI: 10.48550/ARXIV.2204.10898. arXiv: 2204.10898. URL: https://doi.org/10.48550/arXiv.2204.10898.

20 Hyoukjun Kwon, Krishnakumar Nair, Jamin Seo, Jason Yik, Debabrata Mohapatra, Dongyuan Zhan, Jinook Song, Peter Capak, Peizhao Zhang, Peter Vajda, Colby R. Banbury, Mark Mazumder, Liangzhen Lai, Ashish Sirasao, Tushar Krishna, Harshit Khaitan, Vikas Chandra, and **Vijay Janapa Reddi**. "XR-Bench: An Extended Reality (XR) Machine Learning Benchmark Suite for the Metaverse". In: *CoRR* abs/2211.08675 (2022). DOI: 10.48550/ARXIV.2211.08675. arXiv: 2211.08675. URL: https://doi.org/10.48550/arXiv.2211.08675.

21 Maximilian Lam, Michael Mitzenmacher, **Vijay Janapa Reddi**, Gu-Yeon Wei, and David Brooks. "Tabula: Efficiently Computing Nonlinear Activation Functions for Secure Neural Network Inference". In: *CoRR* abs/2203.02833 (2022). DOI: 10.48550/ARXIV.2203.02833. arXiv: 2203.02833. URL: https://doi.org/10.48550/arXiv.2203.02833.

22 Mark Mazumder, Colby R. Banbury, Xiaozhe Yao, Bojan Karlas, William Gaviria Rojas, Sudnya Frederick Diamos, Greg Diamos, Lynn He, Douwe Kiela, David Jurado, David Kanter, Rafael Mosquera, Juan Ciro, Lora Aroyo, Bilge Acun, Sabri Eyuboglu, Amirata Ghorbani, Emmett D. Goodman, Tariq Kane, Christine R. Kirkpatrick, Tzu-Sheng Kuo, Jonas Mueller, Tristan Thrush, Joaquin Vanschoren, Margaret Warren, Adina Williams, Serena Yeung, Newsha Ardalani, Praveen K. Paritosh, Ce Zhang, James Zou, Carole-Jean Wu, Cody Coleman, Andrew Y. Ng, Peter Mattson, and **Vijay Janapa Reddi**. "DataPerf: Benchmarks for Data-Centric AI Development". In: *CoRR* abs/2207.10062 (2022). DOI: 10.48550/ARXIV.2207.10062. arXiv: 2207.10062. URL: https://doi.org/10.48550/arXiv.2207.10062.

23 Sabrina M. Neuman, Brian Plancher, Bardienus Pieter Duisterhof, Srivatsan Krishnan, Colby R. Banbury, Mark Mazumder, Shvetank Prakash, Jason Jabbour, Aleksandra Faust, Guido C. H. E. de Croon, and **Vijay Janapa Reddi**. "Tiny Robot Learning: Challenges and Directions for Machine Learning in Resource-Constrained Robots". In: *CoRR* abs/2205.05748 (2022). DOI: 10.48550/ARXIV.2205.05748. arXiv: 2205.05748. URL: https://doi.org/10.48550/arXiv.2205.05748.

24 Shvetank Prakash, Tim Callahan, Joseph Bushagour, Colby R. Banbury, Alan V. Green, Pete Warden, Tim Ansell, and **Vijay Janapa Reddi**. "CFU Playground: Full-Stack Open-Source Framework for Tiny Machine Learning (tinyML) Acceleration on FPGAs". In: *CoRR* abs/2201.01863 (2022). arXiv: 2201.01863. URL: https://arxiv.org/abs/2201.01863.

25 Víctor Mayoral Vilches, Sabrina M. Neuman, Brian Plancher, and **Vijay Janapa Reddi**. "RobotCore: An Open Architecture for Hardware Acceleration in ROS 2". In: *CoRR* abs/2205.03929 (2022). DOI: 10.48550/ARXIV.2205.03929. arXiv: 2205.03929. URL: https://doi.org/10.48550/arXiv.2205.03929.

26 Zishen Wan, Aqeel Anwar, Abdulrahman Mahmoud, Tianyu Jia, Yu-Shun Hsiao, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "FRL-FI: Transient Fault Analysis for Federated Reinforcement Learning-Based Navigation Systems". In: *CoRR* abs/2203.07276 (2022). DOI: 10.48550/ARXIV.2203.07276. arXiv: 2203.07276. URL: https://doi.org/10.48550/arXiv.2203.07276.

27 Zishen Wan, Ashwin Sanjay Lele, Bo Yu, Shaoshan Liu, Yu Wang, **Vijay Janapa Reddi**, Cong Hao, and Arijit Raychowdhury. "Robotic Computing on FPGAs: Current Progress, Research Challenges, and Opportunities". In: *CoRR* abs/2205.07149 (2022). DOI: 10.48550/ARXIV.2205.07149. arXiv: 2205.07149. URL: https://doi.org/10.48550/arXiv.2205.07149.

28 Pete Warden, Matthew Stewart, Brian Plancher, Colby R. Banbury, Shvetank Prakash, Emma Chen, Zain Asgar, Sachin Katti, and **Vijay Janapa Reddi**. "Machine Learning Sensors". In: *CoRR* abs/2206.03266 (2022). DOI: 10.48550/ARXIV.2206.03266. arXiv: 2206.03266. URL: https://doi.org/10.48550/arXiv.2206.03266.

29 Colby R. Banbury, **Vijay Janapa Reddi**, Peter Torelli, Jeremy Holleman, Nat Jeffries, Csaba Király, Pietro Montino, David Kanter, Sebastian Ahmed, Danilo Pau, Urmish Thakker, Antonio Torrini, Pete Warden, Jay Cordaro, Giuseppe Di Guglielmo, Javier M. Duarte, Stephen Gibellini, Videet Parekh, Honson Tran, Nhan Tran, Wenxu Niu, and Xuesong Xu. "MLPerf Tiny Benchmark". In: *CoRR* abs/2106.07597 (2021). arXiv: 2106.07597. URL: https://arxiv.org/abs/2106.07597.

30 Behzad Boroujerdian, Radhika Ghosal, Jonathan J. Cruz, Brian Plancher, and **Vijay Janapa Reddi**. "RoboRun: A Robot Runtime to Exploit Spatial Heterogeneity". In: *CoRR* abs/2108.13354 (2021). arXiv: 2108.13354. URL: https://arxiv.org/abs/2108.13354.

31 Cansu Demirkiran, Furkan Eris, Gongyu Wang, Jonathan Elmhurst, Nick Moore, Nicholas C. Harris, Ayon Basumallik, **Vijay Janapa Reddi**, Ajay Joshi, and Darius Bunandar. "An Electro-Photonic System for Accelerating Deep Neural Networks". In: *CoRR* abs/2109.01126 (2021). arXiv: 2109.01126. URL: https://arxiv.org/abs/2109.01126.

32 Bardienus Pieter Duisterhof, Shushuai Li, Javier Burgués, **Vijay Janapa Reddi**, and Guido C. H. E. de Croon. "Sniffy Bug: A Fully Autonomous Swarm of Gas-Seeking Nano Quadcopters in Cluttered Environments". In: *CoRR* abs/2107.05490 (2021). arXiv: 2107.05490. URL: https://arxiv.org/abs/2107.05490.

33 Daniel Galvez, Greg Diamos, Juan Ciro, Juan Felipe Cerón, Keith Achorn, Anjali Gopi, David Kanter, Maximilian Lam, Mark Mazumder, and **Vijay Janapa Reddi**. "The People's Speech: A Large-Scale Diverse English Speech Recognition Dataset for Commercial Usage". In: *CoRR* abs/2111.09344 (2021). arXiv: 2111.09344. URL: https://arxiv.org/abs/2111.09344.

34 James Gleeson, Srivatsan Krishnan, Moshe Gabel, **Vijay Janapa Reddi**, Eyal de Lara, and Gennady Pekhimenko. "RL-Scope: Cross-Stack Profiling for Deep Reinforcement Learning Workloads". In: *CoRR* abs/2102.04285 (2021). arXiv: 2102.04285. URL: https://arxiv.org/abs/2102.04285.

35 Yu-Shun Hsiao, Zishen Wan, Tianyu Jia, Radhika Ghosal, Arijit Raychowdhury, David Brooks, Gu-Yeon Wei, and **Vijay Janapa Reddi**. "MAVFI: An End-to-End Fault Analysis Framework with Anomaly Detection and Recovery for Micro Aerial Vehicles". In: *CoRR* abs/2105.12882 (2021). arXiv: 2105.12882. URL: https://arxiv.org/abs/2105.12882.

36 Alexandros Karargyris, Renato Umeton, Micah J. Sheller, Alejandro Aristizabal, Johnu George, Srini Bala, Daniel J. Beutel, Victor Bittorf, Akshay Chaudhari, Alexander Chowdhury, Cody Coleman, Bala Desinghu, Gregory F. Diamos, Debo Dutta, Diane Feddema, Grigori Fursin, Junyi Guo, Xinyuan Huang, David Kanter, Satyananda Kashyap, Nicholas D. Lane, Indranil Mallick, Pietro Mascagni, Virendra Mehta, Vivek Natarajan, Nikola Nikolov, Nicolas Padoy, Gennady Pekhimenko, **Vijay Janapa Reddi**, G. Anthony Reina, Pablo Ribalta, Jacob Rosenthal, Abhishek Singh, Jayaraman J. Thiagarajan, Anna Wuest, Maria Xenochristou, Daguang Xu, Poonam Yadav, Michael Rosenthal, Massimo Loda, Jason M. Johnson, and Peter Mattson. "MedPerf: Open Benchmarking Platform for Medical

Artificial Intelligence using Federated Evaluation". In: *CoRR* abs/2110.01406 (2021). arXiv: 2110.01406. URL: https://arxiv.org/abs/2110.01406.

37 Srivatsan Krishnan, Thierry Tambe, Zishen Wan, and **Vijay Janapa Reddi**. "AutoSoC: Automating Algorithm-SOC Co-design for Aerial Robots". In: *CoRR* abs/2109.05683 (2021). arXiv: 2109.05683. URL: https://arxiv.org/abs/2109.05683.

38 Srivatsan Krishnan, Zishen Wan, Kshitij Bhardwaj, Paul N. Whatmough, Aleksandra Faust, Sabrina M. Neuman, Gu-Yeon Wei, David Brooks, and **Vijay Janapa Reddi**. "Machine Learning-Based Automated Design Space Exploration for Autonomous Aerial Robots". In: *CoRR* abs/2102.02988 (2021). arXiv: 2102.02988. URL: https://arxiv.org/abs/2102.02988.

39 Maximilian Lam, Gu-Yeon Wei, David Brooks, **Vijay Janapa Reddi**, and Michael Mitzenmacher. "Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix". In: *CoRR* abs/2106.06089 (2021). arXiv: 2106.06089. URL: https://arxiv.org/abs/2106.06089.

40 Mark Mazumder, Colby R. Banbury, Josh Meyer, Pete Warden, and **Vijay Janapa Reddi**. "Few-Shot Keyword Spotting in Any Language". In: *CoRR* abs/2104.01454 (2021). arXiv: 2104.01454. URL: https://arxiv.org/abs/2104.01454.

41 Brian Plancher, Sabrina M. Neuman, Radhika Ghosal, Scott Kuindersma, and **Vijay Janapa Reddi**. "GRiD: GPU-Accelerated Rigid Body Dynamics with Analytical Gradients". In: *CoRR* abs/2109.06976 (2021). arXiv: 2109.06976. URL: https://arxiv.org/abs/2109.06976.

42 **Vijay Janapa Reddi**, Greg Diamos, Pete Warden, Peter Mattson, and David Kanter. "Data Engineering for Everyone". In: *CoRR* abs/2102.11447 (2021). arXiv: 2102.11447. URL: https://arxiv.org/abs/2102.11447.

43 **Vijay Janapa Reddi**, Brian Plancher, Susan Kennedy, Laurence Moroney, Pete Warden, Anant Agarwal, Colby R. Banbury, Massimo Banzi, Matthew Bennett, Benjamin Brown, Sharad Chitlangia, Radhika Ghosal, Sarah Grafman, Rupert Jaeger, Srivatsan Krishnan, Maximilian Lam, Daniel Leiker, Cara Mann, Mark Mazumder, Dominic Pajak, Dhilan Ramaprasad, J. Evan Smith, Matthew Stewart, and Dustin Tingley. "Widening Access to Applied Machine Learning with TinyML". In: *CoRR* abs/2106.04008 (2021). arXiv: 2106.04008. URL: https://arxiv.org/abs/2106.04008.

44 Zishen Wan, Aqeel Anwar, Yu-Shun Hsiao, Tianyu Jia, **Vijay Janapa Reddi**, and Arijit Raychowdhury. "Analyzing and Improving Fault Tolerance of Learning-Based Navigation Systems". In: *CoRR* abs/2111.04957 (2021). arXiv: 2111.04957. URL: https://arxiv.org/abs/2111.04957.

45 Colby R. Banbury, **Vijay Janapa Reddi**, Max Lam, William Fu, Amin Fazel, Jeremy Holleman, Xinyuan Huang, Robert Hurtado, David Kanter, Anton Lokhmotov, David A. Patterson, Danilo Pau, Jae-sun Seo, Jeff Sieracki, Urmish Thakker, Marian Verhelst, and Poonam Yadav. "Benchmarking TinyML Systems: Challenges and Direction". In: *CoRR* abs/2003.04821 (2020). arXiv: 2003.04821. URL: https://arxiv.org/abs/2003.04821.

46 Colby R. Banbury, Chuteng Zhou, Igor Fedorov, Ramon Matas Navarro, Urmish Thakker, Dibakar Gope, **Vijay Janapa Reddi**, Matthew Mattina, and Paul N. Whatmough. "MicroNets: Neural Network Architectures for Deploying TinyML Applications on Commodity Microcontrollers". In: *CoRR* abs/2010.11267 (2020). arXiv: 2010.11267. URL: https://arxiv.org/abs/2010.11267.

47 Robert David, Jared Duke, Advait Jain, **Vijay Janapa Reddi**, Nat Jeffries, Jian Li, Nick Kreeger, Ian Nappier, Meghna Natraj, Shlomi Regev, Rocky Rhodes, Tiezhen Wang, and Pete Warden. "TensorFlow Lite Micro: Embedded Machine Learning on TinyML Systems". In: *CoRR* abs/2010.08678 (2020). arXiv: 2010.08678. URL: https://arxiv.org/abs/2010.08678.

48 Maximilian Lam, Zachary Yedidia, Colby R. Banbury, and **Vijay Janapa Reddi**. "Quantized Neural Network Inference with Precision Batching". In: *CoRR* abs/2003.00822 (2020). arXiv: 2003.00822. URL: https://arxiv.org/abs/2003.00822.

49 George Papadimitriou, Athanasios Chatzidimitriou, Dimitris Gizopoulos, **Vijay Janapa Reddi**, Jingwen Leng, Behzad Salami, Osman S. Unsal, and Adrián Cristal Kestelman. "Exceeding Conservative Limits: A Consolidated Analysis on Modern Hardware Margins". In: *CoRR* abs/2006.01049 (2020). arXiv: 2006.01049. URL: https://arxiv.org/abs/2006.01049.

50 **Vijay Janapa Reddi**, David Kanter, Peter Mattson, Jared Duke, Thai Nguyen, Ramesh Chukka, Kenneth Shiring, Koan-Sin Tan, Mark Charlebois, William Chou, Mostafa El-Khamy, Jungwook Hong, Michael Buch, Cindy Trinh, Thomas Atta-fosu, Fatih Çakir, Masoud Charkhabi, Xiaodong Chen, Jimmy Chiang, Dave Dexter, Woncheol Heo, Guenther Schmuelling, Maryam Shabani, and Dylan Zika. "MLPerf Mobile Inference Benchmark: Why Mobile AI Benchmarking Is Hard and What to Do About It". In: *CoRR* abs/2012.02328 (2020). arXiv: 2012.02328. URL: https://arxiv.org/abs/2012.02328.

51 Daniel Richins, Dharmisha Doshi, Matthew Blackmore, Aswathy Thulaseedharan Nair, Neha Pathapati, Ankit Patel, Brainard Daguman, Daniel Dobrijalowski, Ramesh Illikkal, Kevin Long, David Zimmerman, and **Vijay Janapa Reddi**. "AI Tax: The Hidden Cost of AI Data Center Applications". In: *CoRR* abs/2007.10571 (2020). arXiv: 2007.10571. URL: https://arxiv.org/abs/2007.10571.

52 Behzad Boroujerdian, Hasan Genc, Srivatsan Krishnan, Wenzhi Cui, Marcelino Almeida, Kayvan Mansoorshahi, Aleksandra Faust, and **Vijay Janapa Reddi**. "MAVBench: Micro Aerial Vehicle Benchmarking". In: *CoRR* abs/1905.06388 (2019). arXiv: 1905.06388. URL: http://arxiv.org/abs/1905.06388.

53 Behzad Boroujerdian, Hasan Genc, Srivatsan Krishnan, Bardienus Pieter Duisterhof, Brian Plancher, Kayvan Mansoorshahi, Marcelino Almeida, Wenzhi Cui, Aleksandra Faust, and **Vijay Janapa Reddi**. "The Role of Compute in Autonomous Aerial Vehicles". In: *CoRR* abs/1906.10513 (2019). arXiv: 1906.10513. URL: http://arxiv.org/abs/1906.10513.

54 Bardienus Pieter Duisterhof, Srivatsan Krishnan, Jonathan J. Cruz, Colby R. Banbury, William Fu, Aleksandra Faust, Guido C. H. E. de Croon, and **Vijay Janapa Reddi**. "Learning to Seek: Autonomous Source Seeking with Deep Reinforcement Learning Onboard a Nano Drone Microcontroller". In: *CoRR* abs/1909.11236 (2019). arXiv: 1909.11236. URL: http://arxiv.org/abs/1909.11236.

55 Matthew Halpern, Behzad Boroujerdian, Todd W. Mummert, Evelyn Duesterwald, and **Vijay Janapa Reddi**. "One Size Does Not Fit All: Quantifying and Exposing the Accuracy-Latency Trade-off in Machine Learning Cloud Service APIs via Tolerance Tiers". In: *CoRR* abs/1906.11307 (2019). arXiv: 1906.11307. URL: http://arxiv.org/abs/1906.11307.

56 Mark D. Hill and **Vijay Janapa Reddi**. "Accelerator-level Parallelism". In: *CoRR* abs/1907.02064 (2019). arXiv: 1907.02064. URL: http://arxiv.org/abs/1907.02064.

57 Srivatsan Krishnan, Behzad Boroujerdian, William Fu, Aleksandra Faust, and **Vijay Janapa Reddi**. "Air Learning: An AI Research Platform for Algorithm-Hardware Benchmarking of Autonomous Aerial Robots". In: *CoRR* abs/1906.00421 (2019). arXiv: 1906.00421. URL: http://arxiv.org/abs/1906.00421.

58 Srivatsan Krishnan, Sharad Chitlangia, Maximilian Lam, Zishen Wan, Aleksandra Faust, and **Vijay Janapa Reddi**. "Quantized Reinforcement Learning (QUARL)". In: *CoRR* abs/1910.01055 (2019). arXiv: 1910.01055. URL: http://arxiv.org/abs/1910.01055.

59 Peter Mattson, Christine Cheng, Cody Coleman, Greg Diamos, Paulius Micikevicius, David A. Patterson, Hanlin Tang, Gu-Yeon Wei, Peter Bailis, Victor Bittorf, David Brooks, Dehao Chen, Debojyoti Dutta, Udit Gupta, Kim M. Hazelwood, Andrew Hock, Xinyuan Huang, Bill Jia, Daniel Kang, David Kanter, Naveen Kumar, Jeffery Liao, Guokai Ma, Deepak Narayanan, Tayo Oguntebi, Gennady Pekhimenko, Lillian Pentecost, **Vijay Janapa Reddi**, Taylor Robie, Tom St. John, Carole-Jean Wu, Lingjie Xu, Cliff Young, and Matei Zaharia. "MLPerf Training Benchmark". In: *CoRR* abs/1910.01500 (2019). arXiv: 1910.01500. URL: http://arxiv.org/abs/1910.01500.

60 Thanh Thi Nguyen and **Vijay Janapa Reddi**. "Deep Reinforcement Learning for Cyber Security". In: *CoRR* abs/1906.05799 (2019). arXiv: 1906.05799. URL: http://arxiv.org/abs/1906.05799.

61 **Vijay Janapa Reddi**, Christine Cheng, David Kanter, Peter Mattson, Guenther Schmuelling, Carole-Jean Wu, Brian Anderson, Maximilien Breughe, Mark Charlebois, William Chou, Ramesh Chukka, Cody Coleman, Sam Davis, Pan Deng, Greg Diamos, Jared Duke, Dave Fick, J. Scott Gardner, Itay Hubara, Sachin Idgunji, Thomas B. Jablin, Jeff Jiao, Tom St. John, Pankaj Kanwar, David Lee, Jeffery Liao, Anton Lokhmotov, Francisco Massa, Peng Meng, Paulius Micikevicius, Colin Osborne, Gennady Pekhimenko, Arun Tejusve Raghunath Rajan, Dilip Sequeira, Ashish Sirasao, Fei Sun, Hanlin Tang, Michael Thomson, Frank Wei, Ephrem Wu, Lingjie Xu, Koichi Yamada, Bing Yu, George Yuan, Aaron Zhong, Peizhao Zhang, and Yuchen Zhou. "MLPerf Inference Benchmark". In: *CoRR* abs/1911.02549 (2019). arXiv: 1911.02549. URL: http://arxiv.org/abs/1911.02549.

62 Ethan Shaotran, Jonathan J. Cruz, and **Vijay Janapa Reddi**. "GLADAS: Gesture Learning for Advanced Driver Assistance Systems". In: *CoRR* abs/1910.04695 (2019). arXiv: 1910.04695. URL: http://arxiv.org/abs/1910.04695.

63 Thierry Tambe, En-Yu Yang, Zishen Wan, Yuntian Deng, **Vijay Janapa Reddi**, Alexander M. Rush, David Brooks, and Gu-Yeon Wei. "AdaptivFloat: A Floating-point based Data Type for Resilient Deep Learning Inference". In: *CoRR* abs/1909.13271 (2019). arXiv: 1909.13271. URL: http://arxiv.org/abs/1909.13271.

64 Ting-Wu Chin, Chia-Lin Yu, Matthew Halpern, Hasan Genc, Shiao-Li Tsao, and **Vijay Janapa Reddi**. "Domain Specific Approximation for Object Detection". In: *CoRR* abs/1810.02010 (2018). arXiv: 1810.02010. URL: http://arxiv.org/abs/1810.02010.

65 Mikhail Kazdagli, Ling Huang, **Vijay Janapa Reddi**, and Mohit Tiwari. "EMMA: A New Platform to Evaluate Hardware-based Mobile Malware Analyses". In: *CoRR* abs/1603.03086 (2016). arXiv: 1603.03086. URL: http://arxiv.org/abs/1603.03086.

Workshop Papers

1 Jayashree Mohan, Dhathri Purohith, Matthew Halpern, Vijay Chidambaram, and **Vijay Janapa Reddi**. "Storage on Your SmartPhone Uses More Energy Than You Think". In: *9th USENIX Workshop on Hot Topics in Storage and File Systems, HotStorage 2017, Santa Clara, CA, USA, July 10-11, 2017*. USENIX Association, 2017. URL: https://www.usenix.org/conference/hotstorage17/program/presentation/mohan.

2 Alex Shye, Matthew Iyer, Tipp Moseley, David Hodgdon, Dan Fay, **Vijay Janapa Reddi**, and Daniel A. Connors. "Analysis of path profiling information generated with performance monitoring hardware". In: *9th Annual Workshop on Interaction between Compilers and Computer Architectures, INTERACT-9 2005, San Francisco, California, USA, February 13, 2005*. IEEE Computer Society, 2005, pp. 34–43. DOI: 10.1109/INTERACT.2005.3. URL: https://doi.org/10.1109/INTERACT.2005.3.

3 Alex Shye, Matthew Iyer, **Vijay Janapa Reddi**, and Daniel A. Connors. "Code coverage testing using hardware performance monitoring support". In: *Proceedings of the Sixth International Workshop on Automated Debugging, AADEBUG 2005, Monterey, California, USA, September 19-21, 2005*. ACM, 2005, pp. 159–163. DOI: 10.1145/1085130.1085151. URL: https://doi.org/10.1145/1085130.1085151.

4 **Vijay Janapa Reddi**, Alex Settle, Daniel A. Connors, and Robert S. Cohn. "PIN: a binary instrumentation tool for computer architecture research and education". In: *Proceedings of the 2004 workshop on Computer architecture education - Held in conjunction with the 31st International Symposium on Computer Architecture, WCAE@ISCA 2004, Munich, Germany, June 19, 2004*. ACM, 2004, p. 22. DOI: 10.1145/1275571.1275600. URL: https://doi.org/10.1145/1275571.1275600.

Talks

10/2023 "Architecture 2.0" Open Compute Project (OCP), Virtual
09/2023 "Architecture 2.0" MIT Industry-Academia Partnership
07/2023 "Architecture 2.0" Design Automation Conference (DAC), Virtual
06/2023 "The Parameter and Chip Wars" Vail Computer Elements Workshop, Virtual
06/2023 "Adopting AI: With Power Comes Responsibility" Panel, FDCAI, Virtual
04/2023 "TinyML" IEEE International Symposium on Low-Power and High-Speed Chips (COOL Chips), Virtual
04/2023 "The Parameter and Chip Wars: Shifting the Focus from Model-centric to Data-centric AI" MICRON, Virtual
04/2023 "NeuroBench: Advancing Neuromorphic Computing through Collaborative and Rigorous Benchmarking" NICE, Virtual
11/2022 "TinyML" Urban Sensor Networks Workshop Panel, Virtual
10/2022 "ML Metrics: The Past, Present, and Future of Benchmarking ML Systems, Datasets, and Use Cases" Specialization with Benchmarks for Emerging Applications (MICRO), Virtual
09/2022 "Tiny Machine Learning" Chips & Compilers Symposium, MLSys '22, Virtual
09/2022 "MLPerf & DataPerf" The Autonomous Vehicle Computing Consortium (AVCC), Virtual
09/2022 "Benchmarking FastML Systems" Fast ML for Science Workshop
08/2022 "Tiny Machine Learning: A System-level Perspective" ACM/IEEE International Symposium on Low

Power Electronics and Design, Virtual. **(Keynote)**

08/2022 "The Vision Behind MLPerf and DataPerf" Monterey Data Conference, Virtual

08/2022 "DataPerf: Benchmarks for Data-centric AI Development" The Future of Data-Centric AI, Snorkel.ai, Virtual

07/2022 "Tiny Machine Learning: Challenges and Opportunities" Design Automation Conference, ROAD4NN Workshop, Virtual. **(Keynote)**

07/2022 "The Future of Smart Cities is Tiny and Bright" ACM International Conference on Future Energy Systems (ACM e-Energy), Virtual. **(Keynote)**

06/2022 "Machine Learning Metrics" HiPEAC AccML Workshop

05/2022 "Democratizing TinyML" Rutgers Efficient AI (REFAI) Seminar, Virtual

04/2022 "Tiny Machine Learning (TinyML) for Domain-Specific Systems" International Workshop on Domain Specific System Architecture (DOSSA-4), Virtual

03/2022 "IoT 2.0: The Era of Intelligence on Things" Design Automation and Test in Europe, Virtual

01/2022 "Tiny Machine Learning" Accelerated Machine Learning Workshop, co-located with HiPEAC 2022, Virtual

12/2021 "Machine Learning's Future is Tiny & Bright" AICON GWANGJU, Virtual. **(Keynote)**

12/2021 "Democratizing TinyML" Globecom Workshop on Sustainable Environmental Sensing Systems, Virtual. **(Keynote)**

11/2021 "Tiny Machine Learning" EdukCircle International Convention on Engineering and Computer Technology, Virtual

11/2021 "Tiny Machine Learning (TinyML) for Robotics" Conference on Robot Learning, Virtual. **(Keynote)**

11/2021 "Democratizing TinyML: Generalization, Standardization and Automation" Workshop on Hardware and Algorithms for Learning On-a-chip (HALO) workshop, ICCAD conference, Virtual. **(Keynote)**

11/2021 "Democratizing TinyML: Generalization, Standardization and Automation" Multi-DNN Workshop, co-located with MICRO, Virtual. **(Keynote)**

11/2021 "Data for TinyML" Data for AI Summit @Google (internal), Virtual. **(Keynote)**

10/2021 "Widening Access to Applied Machine Learning with TinyML" IEEE Global Humanitarian Technology Conference, Virtual. **(Keynote)**

10/2021 "The Vision Behind MLPerf" SiFive Engineering Forum, Virtual

10/2021 "The Vision Behind MLPerf" Samsung AI Cambridge, Santa Clara

10/2021 "Democratizing TinyML" MICRO 2021 Workshop on Systems for Multi-DNN Workloads, Virtual. **(Keynote)**

09/2021 "The Vision Behind MLPerf" Tensorrent, Virtual

07/2021 "Tiny Machine Learning" Workshop on Artificial Intelligence, Machine Learning, & Computational Intelligence, Virtual

07/2021 "The Vision Behind MLPerf" STMicroelectronics, Virtual

03/2021 "tinyMLPerf: Benchmarking Ultra-low-power Systems" Tiny Machine Learning Summit, San Francisco

03/2021 "tinyMLPerf: Benchmarking Ultra-low-power Systems" "Machine Learning at the Edge," Workshop co-located with Design Automation Conference

06/2020 "The Vision Behind MLPerf" AMD Tech Talk, Austin

03/2020 "tinyMLPerf: Benchmarking Ultra-low-power Systems" Tiny Machine Learning Summit

02/2020 "The Vision Behind MLPerf" International Solid-State Circuits Conference (ISSCC), San Francisco

02/2020 "MLPerf Inference" Machine Learning Systems Workshop, Santa Clara

09/2019 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators in Cloud and Edge Computing" Taiwan Semiconductor Manufacturing Company (TSMC)

09/2019 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators in Cloud and Edge Computing" Taiwan AI Labs. **(Keynote)**

09/2019 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators in Cloud and Edge Computing" Synopsis SNUG, Taiwan

04/2019 "Ten Commandments for Mobile Computer Architecture" Workshop on Infrastructure and Methodology for SoC-level Performance and Power Modeling, co-located with ASPLOS

03/2019 "The Vision Behind MLPerf (mlperf.org)" Intel VSSAD

03/2019 "Evaluating Resiliency in End-to-end Learning for Autonomous Machines" The 15th Workshop on Silicon Errors in Logic – System Effects

03/2019 "Autonomous Aerial Computing Machines" International Workshop on Performance Analysis of Machine Learning Systems

12/2018 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators, and ML cloud and edge platforms" IEEE BigBench co-located with the IEEE Big Data Conference. **(Keynote)**

12/2018 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators in Cloud and Edge Computing" The Forum of Turing Centers, Shanghai Jiao Tong University. **(Keynote)**

12/2018 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators in Cloud and Edge Computing" Boston Area Computer Architecture Workshop. **(Keynote)**

11/2018 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators, and ML cloud and edge platforms" Samsung Advanced Computing Lab (ACL)

10/2018 "The Vision Behind MLPerf: A broad ML benchmark suite for measuring the performance of ML software frameworks, ML hardware accelerators, and ML cloud and edge platforms" Samsung Austin Research Center (SARC)

10/2018 "Mobile Robotics for Computer Architects" First Annual Workshop on Domain Specific System Architecture co-located with International Symposium on Microarchitecture (MICRO). **(Keynote)**

04/2018 "Aerial Computing: Challenges and Opportunities for Hardware and Software Architects Designing Flying Systems" IBM T. J. Watson

03/2018 "Architecting for Big Data Analytics: Think Dubai rather than Venice" Workshop on BigData Benchmarks, Performance, Optimization and Emerging Hardware (co-located with ASPLOS)

02/2017 "Architecture Support for Scripting from Mobile to Cloud" Stanford University, Palo Alto

05/2016 "Watt-Wise-Web://Architecting for Responsiveness and Energy-Efficiency" The University of Chicago, Chicago–IL

05/2016 "Mobile CPU Evolution: The Past, the Present, and the Future" Rice University – TexasWISE Keynote, Houston

05/2016 "Microarchitectural Implications of Event-driven Programming" Northwestern, Chicago–IL
05/2016 "Microarchitectural Implications of Event-driven Programming" Intel Santa Clara–CA
05/2016 "Microarchitectural Implications of Event-driven Programming" AMD Austin–TX
03/2016 "Programming the Web of Things: Why Architects Should Care" Sensors to Cloud Architectures Workshop, Barcelona. **(Keynote)**
02/2016 "From Moore's Law to Moore's Crawl: Architecting the Next-Generation of Mobile Computing Devices" University of Washington, Seattle–WA
02/2016 "From Moore's Law to Moore's Crawl: Architecting the Next-Generation of Mobile Computing Devices" National Academy of Engineering (NAE) Annual Event, Irvine–CA
12/2015 "Programming the Web of Things" Workshop on Internet of Things (IoT) held in conjunction with International Symposium on Microarchitecture, Hawaii
12/2015 "End of the Road for My CAREER" Workshop on Negative Outcomes, Post-mortems, and Experiences (NOPE) held in conjunction with International Symposium on Microarchitecture, Hawaii
11/2015 "Watt-Wise Web: Architecting for a Responsive and Energy-Efficient Mobile Web" Texas A&M University
10/2015 "Watt-Wise Web: Architecting for a Responsive and Energy-Efficient Mobile Web" Google Faculty Summit
10/2015 "Watt-Wise Web: Architecting for a Responsive and Energy-Efficient Mobile Web" Georgia Tech University
09/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" National Taiwan University, Taiwan
09/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" Mediatek, Taiwan
09/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" Academia Sinica, Taiwan
09/2015 "Mobile CPU Evolution: The Past, the Present, and the Future" Taiwan Application Processor Union – Mobile SoC Summer Course, Taiwan
06/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" Duke University, Raleigh–NC
06/2015 "GPU Voltage Guardband Management to Achieve Exascale Energy-Efficiency" AMD Austin–TX
05/2015 "Voltage Noise in Multicore Processors" Intel, Portland
05/2015 "GPU Voltage Guardband Management to Achieve Exascale Energy-Efficiency" Intel, Portland
04/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" Qualcomm, Raleigh–NC
04/2015 "Mobile CPU Evolution: The Past, the Present, and the Future" Microsoft, Seattle–WA
03/2015 "What Users Want and What Hardware Provides: Bridging the Gap Between User Quality of Experience (QoE) and Mobile Device Trends" Facebook, Menlo Park–CA
02/2015 "Mobile CPU Evolution: The Past, the Present, and the Future" Intel Santa Clara–CA
11/2014 "Watt-Wise Web: Architecting for a Responsive and Energy-Efficient Mobile Web" Univ. of Michigan
06/2014 "Simulators are Perfect, Authors are Oracles, Users are Innocent" Workshop on Duplicating, Deconstructing and Debunking (WDDD) held in conjunction with International Symposium on Computer Architecture

06/2014 "Architecting for the Mobile Web: Where We've Been, Where We're Heading, and What We Need to Address" Parallelism in Mobile Platforms (PRISM) held in conjunction with International Symposium on Computer Architecture

05/2014 "Mobile Processor Architectures: Design Implications and Challenges for Energy Efficiency" Indo-American Frontiers of Engineering (IAFOE), Mysore–India

05/2014 "Hardware and Software Co-Design for Robust and Resilient Execution" International Conference on Integrated Circuit Design and Technology (ICICDT), Austin

03/2014 "Architectural Support for the Interactive Mobile Web" Samsung Austin–TX

03/2014 "Architectural Support for the Interactive Mobile Web" ARM Austin–TX

02/2014 "Robust and Resilient Systems from the Bottom-Up: Circuits, Architecture and Software Integration" ISSCC Forum, San Francisco–CA

02/2014 "Architectural Support for the Interactive Mobile Web" Intel Austin–TX

02/2013 "Toward High-Performance and Energy-Efficient Mobile Web Browsing" Qualcomm, Santa Clara–MA

08/2012 "Toward High-Performance and Energy-Efficient Mobile Web Browsing" Intel Austin–TX

08/2012 "Toward High-Performance and Energy-Efficient Mobile Web Browsing" AMD Austin–TX

10/2010 "Web Search Using Small Cores" AMD, Boxborough–MA

07/2010 "Web Search Using Small Cores" SeaMicro Santa Clara–CA

07/2010 "Web Search Using Small Cores" Intel Hudson–MA

07/2010 "Web Search Using Small Cores" IBM T. J. Watson Labs, Hawthorne–NY

07/2010 "Web Search Using Small Cores" HP Labs, Palo Alto–CA

07/2010 "Web Search Using Small Cores" Google, Palo Alto–CA

07/2010 "Web Search Using Small Cores" Facebook, Palo Alto–CA

07/2010 "Software-Assisted Hardware Reliability" Intel, Portland

07/2010 "Software-Assisted Hardware Reliability" IBM T. J. Watson Labs, Yorktown–NY

06/2010 "Web Search Using Small Cores" Amazon, Seattle–WA

06/2010 "Software-Assisted Hardware Reliability" Microsoft Research, Redmond–WA

03/2010 "Software-Assisted Hardware Reliability" Intel Santa Clara–CA

03/2010 "Software-Assisted Hardware Reliability" AMD Austin–TX

03/2007 "Persistent Code Caching" Intel Santa Clara–CA

Books

1 **Vijay Janapa Reddi**. *Machine Learning Systems with TinyML*. (Accessed on 03/04/2024). 2023. URL: https://harvard-edge.github.io/cs249r_book/.

2 **Vijay Janapa Reddi** and Meeta Sharma Gupta. *Resilient Architecture Design for Voltage Variation*. Synthesis Lectures on Computer Architecture. Morgan & Claypool Publishers, 2013. ISBN: 9781608456376. DOI: 10.2200/S00486ED1V01Y201303CAC022. URL: https://doi.org/10.2200/S00486ED1V01Y201303CAC022.

Technical Reports

2010 V. Janapa Reddi, B. Lee, T. Chilimbi, K. Vaid. "Web Search Using Small Cores: Quantifying the Price of Efficiency," in Microsoft Research Tech. Report, June.

Theses

2010 V. Janapa Reddi. "Software-Assisted Hardware Reliability: Enabling Aggressive Timing Speculation Using Run-Time Feedback from Hardware and Software," Ph.D. Thesis, School of Engineering and Applied Sciences, Harvard University.

2005 V. Janapa Reddi. "Deploying Dynamic Code Transformation in Modern Computing Environments," M.S. Thesis, Department of Electrical and Computer Engineering, University of Colorado.

2003 V. Janapa Reddi. "Heterogeneous Networks of Workstations Across Wide Area Networks," B.S. Thesis, Department of Electrical and Computer Engineering, Santa Clara University.

Patents

2005 R. Cohn, T. Moseley, and V. Janapa Reddi. "System and method to instrument references to shared memory." U.S. Patent Application 11/143,130, filed June 1, 2005.

2012 N. Kim, J. O'Connor, M. Schulte, and V. Janapa Reddi. "Method and apparatus for power reduction during lane divergence." U.S. Patent Application 13/605,460, filed September 6, 2012.

2015 V. Janapa Reddi, M. Gupta, G. Holloway, G. Wei, M. D. Smith, and D. Brooks. "Adaptive event-guided system and method for avoiding voltage emergencies." U.S. Patent 8,949,666, issued February 3, 2015

Teaching

Harvard University

Sp' 2024 COMPSCI 141: Computing Hardware
Fa' 2023 COMPSCI 249R: Tiny Machine Learning
Sp' 2023 COMPSCI 141: Computing Hardware
Fa' 2022 COMPSCI 249R: Tiny Machine Learning
Sp' 2021 COMPSCI 141: Computing Hardware
Fa' 2020 COMPSCI 249R: Tiny Machine Learning
Fa' 2019 COMPSCI 249R: Autonomous Machines

The University of Texas at Austin

Sp' 2016 EE 319K: Introduction to Embedded Systems
Sp' 2015 EE 319K: Introduction to Embedded Systems
Fa' 2014 EE382V: Code Generation and Optimization

Sp' 2014 EE 319K: Introduction to Embedded Systems
Fa' 2013 EE382V: Code Generation and Optimization
Sp' 2013 EE 319K: Introduction to Embedded Systems
Fa' 2012 EE382V: Dynamic Compilation
Sp' 2012 EE382V: Code Generation and Optimization
Fa' 2011 EE382V: Dynamic Compilation

Students

Current PhD students

2023/- Jeffrey Ma
2023/- Oishii Banerjee
2022/- Ikechukwu Uchendu
2022/- Jason Jabbour
2022/- Jason Yik
2022/- Jessica Quaye
2021/- Emma Chen
2021/- Mark Mazumder
2020/- Shvetank Prakash
2019/- Colby Banbury
2019/- Max Lam
2019/- Radhika Ghosal
2019/- Yu-shun Hsiao
2018/- Srivatsan Krishnan

Graduated PhD students

2018/2022 Brian Plancher,
PhD Thesis: *"GPU Acceleration for Real-time, Whole-body, Nonlinear Model Predictive Control,"*
First Job: Assistant Professor at Barnard College at University of Columbia

2014/2022 Behzad Boroujerdian,
PhD Thesis: *"Agile Development of Domain-Specific Solutions for Emerging Mobile Systems,"*
First Job: Deep Learning Researcher, NVIDIA

2014/2022 Daniel Richins,
PhD Thesis: *"Bottlenecks in Big PhD Thesis: Data Analytics and AI Applications and Opportunities for Improvement,"*
First Job: Instructor, Brigham Young University

2013/2018 Yazhou Zu,
PhD Thesis: *"Active Timing Margin Management to Improve Microprocessor Power Efficiency,"*

First Job: Software Engineer, Google

2011/2016 Jingwen Leng,
PhD Thesis: *"Guardband Management in Heterogeneous Architectures,"*
First Job: Assistant Professor at Shanghai Jiao Tong University (CSE)

2011/2016 Yuhao Zhu,
PhD Thesis: *"Energy-Efficient Mobile Web Computing,"*
First Job: Assistant Professor at Univ. of Rochester (CS)

Postdoctoral Associates

2020/- Matthew Stewart
2020/2023 Sabrina Neuman, Assistant Professor at Boston University

M.S. students

2019/2021 Jonathan Cruz, Harvard
2015/2017 Wenzhi Cui, UT Austin
2014/2017 Matthew Halpern, UT Austin
2011/2013 Aditya Srikanth, UT Austin
2011/2013 Ankita Garg, UT Austin

Bio

Vijay Janapa Reddi is an Associate Professor at Harvard University, Vice President, and Founding Member of MLCommons (mlcommons.org), a nonprofit organization devoted to accelerating machine learning (ML) innovation for all. He co-chairs the MLCommons Research organization and sits on the board of directors of MLCommons. He co-led the development of the MLPerf Inference benchmark for IoT, mobile, edge, and datacenter applications. Before moving to Harvard, he was an Associate Professor at The University of Texas at Austin's Department of Electrical and Computer Engineering. He specializes in developing mobile and edge computing platforms, as well as the Internet of Things. His work is largely based on runtime systems, computer architecture, and applied machine learning methods. Numerous accolades and awards have been awarded to Dr. Janapa-Reddi, including the Gilbreth Lecturer Honor from the National Academy of Engineering (NAE) in 2016, the IEEE TCCA Young Computer Architect Award (2016), the Intel Early Career Award (2013), the Google Faculty Research Awards in 2012, 2013, 2015, 2017, and 2020, the Best Papers at the 2020 Design Automation Conference (DAC), the 2005 International Symposium on Microarchitecture (MICRO), and the 2009 International Symposium on High-Performance Computer Architecture (HPCA). Additionally, he has won various honors and awards, including IEEE Top Picks in Computer Architecture (2006, 2010, 2011, 2016, 2017, 2022, 2023). The MICRO and HPCA Halls of Fame include him (inducted in 2018 and 2019, respectively). He is strongly devoted to expanding access to applied machine learning for STEM, diversity, and the application of AI for social good. To merge

embedded systems and machine learning, he developed the Tiny Machine Learning (TinyML) series on edX, a massive open online course (MOOC) that thousands of students worldwide can access and audit for free. Additionally, he oversaw the Austin Hands-on Computer Science (HaCS) program, which the Austin Independent School District used to teach CS to students in grades K-12. Dr. Janapa-Reddi holds degrees in computer science from Harvard University, electrical and computer engineering from the University of Colorado at Boulder, and computer engineering from Santa Clara University. Dr. Janapa-Reddi's life's passion is dedicated to helping individuals and teams succeed while making the world a better place, one bit at a time.