Murali Annavaram

Lloyd Hunt Chair Professor
Ming Hsieh Department of Electrical & Computer Engineering
Thomas Lord Department of Computer Science
University of Southern California
Los Angeles, CA 90089-2562
TEL: (213) 740-3299
annavara@usc.edu
http://scip-lab.usc.edu

Professional Experience

- 7/2023-Current: **Lloyd Hunt Chair in Electrical and Computer Engineering**, Ming Hsieh Electrical and Computer Engineering Department, Thomas Lord Department of Computer Science (courtesy), University of Southern California
- 8/2022-7/2023: **Rukmini Gopalakrishnachar Chair Visiting Professor,** Indian Institute of Sciences, Bangalore, India.
- 4/2022- 7/2023t: **Dean's Professor**, Ming Hsieh Electrical and Computer Engineering Department, Thomas Lord Department of Computer Science, University of Southern California
- 12/2017- Current: **Professor**, Ming Hsieh Electrical and Computer Engineering Department, Thomas Lord Department of Computer Science, University of Southern California
- 4/2012- 12/2017: **Assoc. Professor**, Ming Hsieh Electrical and Computer Engineering Department, Thomas Lord Department of Computer Science, University of Southern California
- 8/2007- 4/2012: **Asst. Professor**, Ming Hsieh Electrical and Computer Engineering Department, Thomas Lord Department of Computer Science (courtesy), University of Southern California

- 4/2010-12/2017: **Robert G. and Mary G. Lane Early Career Chair Professor**, Ming Hsieh Electrical and Computer Engineering Department, University of Southern California
- 2/2020-2/2021: Consulting Faculty Scientist, Facebook
- 8/2014-12/2014: Visiting faculty on Sabbatical, Caltech
- 2/2007-8/2007: Faculty Researcher, Nokia Research Labs
- 8/2001-2/2007: Senior Research Scientist, Intel
- 5/1995-8/1995: Summer Intern, AT&T Bell Labs.
- 9/1993-7/1994: Software Engineer, Wipro Systems, India.

Education

- The University of Michigan, Ann Arbor, MI; 1996-2001
  Ph.D. Computer Science and Engineering
  (Dissertation Title: Prefetch Mechanisms that Acquire and Exploit Application Specific Knowledge; Advisor: Prof. Edward Davidson)
- Colorado State University, Fort Collins, CO; 1994-1996
  M.S, Computer Science and Engineering
- National Institute of Technology, Warangal, India; 1989-1993
  B.Tech, Computer Science

Research Summary

Murali Annavaram is the Lloyd Hunt Chair Professor in the Ming-Hsieh Department of Electrical and Computer Engineering at the University of Southern California. He also holds a joint appointment with the Thomas Lord department of Computer Science. He held the Rukmini Gopalakrishnachar Visiting Chair Professorship at the Indian Institute of Science, Bangalore. He is the founding Director of the USC-Meta Center for Research and Education in AI and Learning. He is also the architecture thrust leader for the DISCoVER NSF Expeditions in Computing Center at USC. For his numerous publications, he was inducted into the hall of fame at three top-tier computer architecture conference venues, ISCA, HPCA and MICRO. He held the Dean's Chair and Robert G. and Mary G. Lane Early Career Chair professorships at USC in the past.

His research focuses on:

- ◊ Superconductive computing
- ◊ Energy efficiency through heterogeneous computing
- ◊ Reliability of high performance computing
- ◊ Bandwidth efficient big data computing,
- ◊ Runtime systems design to enable dispersed computing
- ◊ Hardware-assisted secure and private machine learning
- ◊ Building innovative sensor data collection platforms to improve human performance.

Prior to his appointment at USC he worked at Intel Microprocessor Research Labs from 2001 to 2007 working on energy efficient system design and die-stacking

architectures. In 2007 he was a visiting researcher at the Nokia Research Center, Palo Alto working on mobile phone-based wireless traffic sensing using virtual trip lines.

Murali received the NSF CAREER award in 2010 and an IBM Faculty Partnership award in 2008, Facebook faculty award in 2019. He received the Stevens Institute's Innovation Inside curriculum award for jointly developing a mobile systems design course.

Murali co-authored Parallel Computer Organization and Design, a widely used textbook to teach both the basic and advanced principles of computer architecture. This book is used in graduate computer architecture courses in several reputed universities around the world.

Murali received the Ph.D. degree in Computer Engineering from the University of Michigan, Ann Arbor, in 2001. He is a Senior Member of ACM. He is an IEEE Fellow

Honors and Awards

- 12/2023: ACM Distinguished Member
- 7/2023: Lloyd Hunt Chair in Electrical Power
- 8/2022: Rukmini Gopalakrishnachar Chair Visiting Professorship, Indian Institute of Science
- 4/2022: Dean's Professorship in ECE and CS departments
- 4/2022: Architecture thrust leader of the Discover Expeditions grant from NSF (largest computing grant given at the NSF CISE division)
- 3/2022: Frontiers in Technology Distinguished Speaker, UC Merced
- 12/ 2021: Founding and Inaugural Director, USC-Meta Center for Research and Education in AI and Learning
- 11/2021: IEEE Fellow
- 11/2021: Keynote speaker at the King Fahd University of Petroleum & Minerals annual KIXX forum
- 10/2021: NeurIPS 2021 Outstanding Reviewer Award
- 12/2020: Best paper award for FedML.AI infrastructure
- 11/2019: Best paper and Best Student paper finalist nominations, Supercomputing 2019 conference
- 11/2018: Keynote speaker for Swedish Multicore Symposium
- 6/2017: Inducted into the ACM ISCA hall of fame for publishing eight papers at the ISCA conference (one of 81 researchers over the last 44 years)
- 3/2016: Patent on improving mobile system's battery efficiency licensed
- 12/2015: Inducted into the ACM SIGMICRO hall of fame for publishing eight papers at the MICRO conferences (one of 47 researchers over the last 48 years)
- 7/2014: Keynote speaker at the 2014 IEEE International On-Line Test Symposium, Spain.
- 6/2013: IEEE Micro Top Picks Award for the paper titled "KnightShift-Scaling Energy Proportionality Wall Through Server-level Heterogeniety."
- 8/2011: Holder of Robert G. and Mary G. Lane Early Career Chair 2011-Present
- 7/2010: NSF CAREER award 2010

- 4/2012: Best Paper Nomination at the 2012 IEEE International Symposium on Workload Characterization (IISWC) for his work "Wireless Body Area Networks: Where Does the Energy Go?"
- 10/2009: Best Paper Award at the 2009 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS) for his joint work on "Optimal Time-Resource Allocation for Activity-Detection via Multimodal Sensing."
- 10/2009: Winner of the 2009 Body Computing Award for project titled The KNOWME Network: Energy-Efficient Activity-Detection for Pediatric Obesity
- 8/2008: Senior Member of Association of Computing Machinery (ACM)
- 8/2008: Senior Member of Institute of Electrical and Electronics Engineers (IEEE)
- 8/2008: IBM Faculty Partnership Award 2008.
- 12/2007: USC Stevens Institute's Innovation Inside Curriculum Award for developing EE579 course on mobile application development at USC
- 3/2007: Nokia Exemplary Team Achievement award (highest award at the Nokia research labs)
- 5/2002, 5/2005: Two Intel divisional recognition awards (highest awards at the corporate technology division at Intel).
- 8/1996-8/1997: University of Michigan EECS Department Fellowship.
- 1989: 32nd Rank in AP Engineering and Medical Common Entrance Test.
- 1987: 14th Rank in State-wide High School Graduation Exam, India.
- 1987-1993: National Merit Scholarship, India.

Consulting Experience

- June 2014-Oct 2015. Testifying expert in the damages phase of **Warf**-vs-Apple. Testified in a jury trial. Deposed by the defendant's counsel.
- Nov 2015-May 2017. ITC matter on various haptic patents **Immersion**-vs-Apple. Delivered expert report and witness statements.
- Aug 2015-May 2017. **Futurelink**-vs-Intel ITC case related to various patents regarding server designs. Delivered expert report and a supplemental expert report and was deposed.
- Oct 2016-June 2018. **Qualcomm**-vs-Apple ITC case on a patent associated with GPUs. Delivered expert report and a supplemental expert report on infringement and validity of the patent. Deposed by the opposing counsel. Testified at ITC.
- June 2018-Feb 2019. **Qualcomm**-vs-Apple district case on a patent associated with GPUs. Delivered expert report and a supplemental expert report on infringement and validity of the patent. Deposed by the opposing counsel. Testified at jury trial.
- Feb 2018-Dec 2018. **NVIDIA**-vs-ZiiLabs ITC case on a patent associated with GPUs. Delivered expert report and a supplemental expert report on non-infringement and invalidity of the patent. Deposed by the opposing counsel.
- Sept 2019-Dec 2022. **PACT**-vs-Intel IPR proceedings. Submitted expert declaration.

- Dec 2019-June 2023. **XMTT**-vs-Intel IPR proceedings. Submitted expert declaration and deposed.
- Dec 2019-June 2023. **XMTT**-vs-Intel District case on a patent. Submitted expert report and deposed.
- June 2020-Oct 2020. **Samsung**-vs-KIPB. District case on a patent. Submitted expert report and deposed.
- Aug 2020-Dec 2020. **Samsung**-vs-Arbor. District case on a patent. Submitted expert report and deposed.
- Sept 2019-Nov 2022. **VLSI**-vs-Intel case on a patent dispute. Expert report submitted and deposed. Testified at trials in Texas.
- June 2021-June 2022. **Netlist**-vs-Google case on a patent dispute.
- June 2021-April 2023. **Williams**-vs-Apple case on a trade secret dispute.
- July 2022-Aug 2024. **Rivos**-vs-Apple case on a trade secret dispute.
- May 2023-Dec 2024, **Qualcomm**-vs-ARM case on a licensing dispute. Submitted expert declaration, deposed and testified at trial.
- Aug 2024-Current, **Mediatek**-vs-Daedalus IPR proceedings.
- Dec 2024-Current, **NXP/Qualcomm**-vs-Redstone IPR proceedings.

Research Projects and Support

Annavaram's research is supported through the following peer-reviewed research grants and industrial research gifts.

Active:

1. VMWARE, "Enhancing Privacy Through Hardware Trusted Execution Environments," M. Annavaram (PI, $265K, July 2022-July 2025)
2. NSF, "SHF:Small:ML Accelerator Cohort Architecture," M. Annavaram (PI, 600K), June 2022-June 2025.
3. NSF, Expeditions in Computing, "DISCOVER: Design and Integration of Superconductive Computation for Ventures beyond Exascale Realization," M. Annavaram (co-PI, ~300K/year), May 2022-Aug 2029. Total grant amount ~15M total funds.
4. USC-Meta Center for Research and Education in AI and Learning. M. Annavaram (Director), Dec 2021-Dec 2026 (~5M total funds).
5. DARPA FastNICs: "DIAMOND: Distributed Training of Massive Models at Bandwidth Frontiers," M. Annavaram (Co-PI ~750K), May 1st 2020-May 1st 2024, Total Grant Amount $2,062,540.

Completed:

1. Avast/Borsat/Intel Gift, "Secure and Privacy Preserving Machine Learning: Foundations and Scalable System Design", M. Annavaram (Co-PI ~50K/year), Nov 2020 – Nov 2023.
2. IARPA SuperTools: "ColdFlux: CAD Methodologies and Tools for Single Flux Quantum Based Superconductive Electronics," M. Annavaram (Co-PI,

~$500,000), M. Pedram (PI). May 1st 2017- Dec 31st 2022, Total Grant Amount ~$13M.

3. NSF, "CNS – 2002874/200282: Collaborative Research: MLWiNS: A Coding-Centric Approach to Robust, Secure, and Private Distributed Learning over Wireless," M. Annavaram (Co-PI~100K), Aug 1st 2020-Dec 31st 2021, Total Grant Amount (~300K)
4. DARPA Dispersed Computing Grant: "ApaC: Adaptive Pricing and Coding for Dispersed Computing, " M. Annavaram (Co-PI -33%, ~$1,406,782), S. Avestimehr (PI-33%), B. Krishnamachari (Co-PI-33%), April 1st 2017- April 1st 2021, Total Grant Amount $ 4,217,754.
5. Facebook Gift, "Machine Inference at the Edge," M. Annavaram (PI), May 1st 2020-April 30th 2021, Total Gift Amount 100,000.
6. NSF, "SHF:Small Semantically Aware Graph Storage and Processing," M. Annavaram (PI-75%), A. Ortega (Co-PI, 25%), Aug 15th 2017 – Aug 14th 2021, Total Grant Amount $400,000.
7. Facebook Gift, "Privacy Preserving Machine Inference," M. Annavaram, Nov 1st 2019-Nov 1st 2020, Total Gift Amount 50,000.
8. Samsung: "Energy Efficient Graph Processing on SSDs," M. Annavaram (PI), , June 15th 2017 – June 14th 2018, Total Grant Amount $100,000.
9. IARPA: "TILES: Tracking Individual Performance with Sensors," M. Annavaram (Co-PI, ~$500,000), S. Narayanan (PI). May 1st 2017- Oct 31st 2020, Total Grant Amount ~$12M.
10. Ming-Hsieh Institute Mini Grant, "Creating the USC Technology Enabled Aging Mind Center (TEAMc)", M. Annavaram (co-PI), U. Mitra (PI), $8,996.
11. NSF CAREER Grant# 0954211, "From Nonstop-Monitoring to Nano-ISA: An Adaptive Multi-Dimensional Framework for Processor Reliability", M. Annavaram (PI), Funded, August 1, 2010-July 31, 2017, $427,664.
12. DARPA PERFECT Grant: "Embedded POWER Optimized Systems Using Near and Super-threshold Computing Fabric (EMPOWER), " M. Annavaram (Co-PI -33%), M. Pedram (PI-33%), K. Roy (Co-PI-33%), Phase 3 funded Aug 1st 2015-Oct 1st 2016, Total Grant Amount $2,155,102.
13. NSF Grant# 1219186, "SHF:Small: Benchmarking of Transient and Intermittent Errors and Their Application to Microarchitecture", M. Annavaram (PI-50%), M. Dubois (Co-PI-50%), Sept 1, 2012 – August 31 2017, $400,000.
14. DARPA PERFECT Grant: "Low-power and Error-resilient Digital Components Realized in Deeply-scaled CMOS (LEDRA), " M. Annavaram (Co-PI -0%), M. Fritze (PI), Funded Phase 1 for ~$1.5M Dec 1st 2012-Dec 1st 2014, Total Grant Amount: ~$8M. Grant money returned due to publication restrictions.
15. DARPA IRIS Grant: "Techniques for Estimating Reliability in COTS Ics (TERCI), " M. Annavaram (Co-PI -10%), March 1st 2013-March 1st 2014, $125,000.
16. NSF Grant# 0954211, "IEEE International Symposium on Workload Characterization (IISWC) Student Subsidy Proposal", M. Annavaram (PI), Jan 1, 2011-Dec 31, 2011, $5,000.

17. NSF Grant# 0831545, "A Game Theoretic Framework for Privacy Preservation in Community-Based Mobile Applications", M. Annavaram (PI-50%), B. Krishnamachari (Co-PI-50%), October 1, 2008 – September 30, 2011, $249,999.
18. NSF REU Supplement Grant# 0939625, "A Game Theoretic Framework for Privacy Preservation in Community-Based Mobile Applications", M. Annavaram (PI-50%), B. Krishnamachari (Co-PI-50%), June 1, 2009 – September 30, 2011, $16,000.
19. NSF Grant# 0834798, "A Holistic Design Approach to Reliability Using 3D Stacked Monitoring", M. Annavaram (PI), Sept 1, 2008 – August 31 2011, $402,904.
20. NSF REU Supplement Grant# 0939623, "A Holistic Design Approach to Reliability Using 3D Stacked Monitoring", M. Annavaram (PI), July 7, 2009 – August 31 2011, $16,000.
21. NSF Grant# 0834799, "Trade-offs Between Static Power, Performance and Reliability in Future Chip Multiprocessors", M. Annavaram (PI-50%), M. Dubois (Co-PI-50%), September 1, 2008 – August 31, 2010, $200,000.
22. NIH/National Center for Minority Health and Health Disparities, Grant# 53-4503-6037, "Mobile device biomonitoring to prevent and treat obesity in underserved minority youth (KnowMe Study)" M. Annavaram (Principal-30%), D. Metz (Principal-30%), U.Mitra (Co-PI-10%), S. Narayanan (Co-PI-10%), G. Sukhatme (Co-PI-10%), N. Medvidovic (Co-Principal-10%), May 1, 2008 – April 30, 2010, $948,348.
23. Qualcomm Research Grant, "Mobile Metabolic Health Monitoring", M. Annavaram (Co-PI -10%), D. Metz (Co-PI-10%), U.Mitra (Co-PI-10%), S. Narayanan (PI-50%), G. Sukhatme (Co-PI-10%), August 1, 2008 – July 31, 2009, $50,000.
24. Oracle Equipment Grant. Virtex-5 OpenSPARC Research Platform Boards, August 2011, $16,000.
25. Qualcomm Equipment Grant. SnapDragon Mobile Development Platform Boards, MSM8660, August 2011, $13,500.
26. Nokia Equipment Grant. 50 Lumia900 Phones, Jan 2013, $40,000.
27. Annenberg Micro Seminar series grant. Aug 2011, $2,500.
28. Rose Hills Foundation Science and Engineering Fellowship grant. June 2011, $2,000.
29. Ming-Hsieh Institute Mini Grant, "The Smartphone In Your Pocket – Refreshing The Hardware Of EE579", M. Annavaram, Jan 2011, $8,996.
30. Nokia Research Gift, "Mobile Energy Efficiency", M. Annavaram (PI), September 1, 2007, $50,000.
31. Nokia Research Gift, "Mobile Systems for Health", M. Annavaram (PI), May 1, 2008, $50,000.
32. Nokia Research Gift, "Mobile Systems for Health", M. Annavaram (PI), April 1, 2009, $50,000.
33. IBM Faculty Award Researc Gift, "3D Stacking for Reliability", M. Annavaram (PI), August 2008, $25,000.
34. USC Startup Package, August 15th 2007, $70,000 Equipment, $50,000 Research Expenses, $120,000 RA Support, 4 Summer Months Salary

Student Advising

Current Phd Students
1. Haipeng Zha
2. Tingting Tang
3. Abdulla Alshabanah
4. Yongqin Wang
5. Rachit Rajat
6. Tara Renduchintala
7. James Flemings
8. Lei Gao
9. Chaoyi Jiang
10. Jonghyun Lee
11. Hossein Entezari Zarch

Phd Students Graduated	Graduation Date	First Employment
1. Kimish Patel (co-advised with Prof. Pedram)	Sept 2010	Nvidia
2. Yi Wang (co-advised with Prof. Krishnamachari)	March 2011	Juniper
3. Jinho Suh (co-advised with Prof. Dubois)	Jan 2012	Intel
4. Bardia Zandian	Nov 2012	Qualcomm
5. Waleed Dweik	Dec 2014	Asst. Professor, University of Jordan, Amman
6. Daniel Wong	June 2015	Assoc. Professor, UC, Riverside
7. Hyeran Jeon	June 2015	Asst. Professor, San Jose State University
8. Sangwon Lee	Dec 2015	VMWare
9. Mohammad Abdel-Majeed	Feb 2016	Asst. Professor, University of Jordan, Amman
10. Qiumin Xu	May 2017	Google
11. Abdulaziz Tabbakh	Jan 2018	Asst. Professor, KAUST, Saudi Arabia.
12. Gunjae Koo	May 2018	Asst. Professor, Korea Univ, Korea

13. Kiran Matam	Dec 2019	Facebook
14. Krishna Narra	Aug 2020	Google
15. Hanieh Hashemi	Aug 2023	Apple
16. Keshav Balasubramanian	Jan 2024	Nvidia

MS Students Advised	Graduation Date	First Employment
1. Sabyasachi Ghosh	June 2012	Riverbed Networks
2. Suk Hun Kang	June 2010	Samsung
3. Kumar Dabbiru	June 2012	Google
4. Melina Demertzi	June 2013	Oracle
5. Zhifeng Lin	Aug 2019	Google

BS Students Mentored	Graduation Date	First Employment
1. Thomas Punihaole	June 2010	UCLA, MS
2. Willy Long	June 2013	Duke, PhD
3. Justin Kuang	June 2014	Stanford, MS
4. Zhifeng Lin	June 2013	USC, PhD
5. Francisco Romero	June 2015	Stanford, PhD
6. Julia Chen	June 2014	Oracle

Phd Student Thesis Committee Member
1. Richard Hankins
2. Weirong Jiang
3. Hua Liu
4. Jianwei Chen
5. Mehrtash Manochehri

Teaching and Course Development at USC

- EE579: Wireless and Mobile Networks Design and Laboratory. Developed this course from ground-up to focus on challenges associated with mobile phone design challenges and usage models. The course covers issues related to energy efficiency in mobile technologies and data privacy issues.
- EE599: Energy Efficiency and Reliability in Information and Communication Technologies. Developed this new experimental course focusing on topics

related to energy proportional computing, datacenter power usage efficiency, server reliability and methods to tradeoff energy efficiency and reliability.
- EE357, EE457 and EE557: Enhanced computer architecture course sequence to bring parallel architecture and software development issues into mainstream architecture courses.

Textbook Co-Authorship

- Parallel Computer Organization and Design. Michel Dubois, Murali Annavaram and Per Stenström. Cambridge Press. Published in August 2012.

Conference Organizational Responsibilities:

- 6/2024: Technical program chair of International Conference on Supercomputing (ICS 2024)
- 12/2021: Technical program chair of IEEE 34th International Symposium on Computer Architecture and High-Performance Computing
- 2/2020: Technical program chair of HPCA 2021
- 6/2018: General co-chair of ISCA 2018
- 4/2016: Track Chair: Green Computing, Design Automation and Test in Europe (DATE), 2016.
- 6/2014: Vice chair computer architecture track, High Performance Computing (HiPC), 2014

Other Professional Activities and Service

- Associate Editor
  - Journal of Parallel and Distributed Computing
  - ACM Transactions on ACM Transactions on Design Automation of Electronic Systems (2011-2014)
- Journal Reviewer
  - ACM Transactions on Computers
  - ACM Transactions on Embedded Computing Systems
  - ACM Computing Surveys
  - ACM Transactions on Architecture and Code Optimization
  - IEEE Top Picks in Microarchitecture 2010, 2015, 2016, 2017.
- Program committee member (including external review committee memberships)
  - IEEE/ACM Architectural Support for Programming Languages and Operating Systems (ASPLOS) multiple years during 2018-2024
  - NeurIPS multiple years during 2020-2024
  - Sigmetrics multiple years during 2016-2024
  - International Symposium on Computer Architecture (ISCA), multiple years during 2007-2024
  - Dependable Systems and Networks (DSN), multiple years during 2015-2022

    - Design Automation and Test in Europe (DATE), 2015, 2016
    - IEEE/ACM International Symposium on Microarchitecture (MICRO) multiple years during 2007-2024
    - IEEE High Performance Computer Architecture (HPCA) multiple years during 2007-2024
    - International Parallel and Distributed Processing Systems (IPDPS) multiple years during 2007-2024.
    - IEEE International Symposium on Performance Analysis of Software and Systems (ISPASS) 2011, 2012, 2018
    - International Conference on Mobile Computing, Applications, and Services (MobiCASE) 2011
    - IEEE Workshop of Energy Efficient Design (WEED) 2011, 2012
    - IEEE International Symposium on Workload Characterization (IISWC) 2010, 2012, 2014, 2015, 2016
    - IEEE/ACM International Conference on Parallel Architecture and Compilation Techniques (PACT) 2010
    - International Conference on Mobile Computing, Applications, and Services (MobiCASE) 2009, 2010, 2011
- Program Track chair
    - Workshop on Exascale Evaluation and Research Techniques (EXERT) 2010, 2011
    - Workshop on Duplicating, Deconstructing, and Debunking (WDDD) 2011, 2012, 2013. 2014, 2015
- Local arrangements chair HPCA 2004, Registration chair for HPCA 2005 and HPCA 2006, Registration chair for PACT 2010, Finance Chair for IISWC 2010.
- Reviewer for IEEE Micro Special Issue on Hardware/Software Interactions 2008.
- Multiple NSF proposal review panels

Graduate Student Experience

- Research Assistant: *Jan 1998 - Aug 2001, The University of Michigan.*
    - I worked as a research assistant under Prof. Edward Davidson working on my Ph.D. thesis. My dissertation research focused on reducing cache miss stalls of irregular applications such as databases by hardware and software assisted prefetching.
- Teaching Assistant: *Sept 1996 – Dec 1997, The University of Michigan.*
    - Teaching assistant for a programming and introductory data structures course (one semester) and a graduate level operating systems course (two semesters). Job responsibilities included leading weekly class discussions, designing homework for the operating systems class and grading homework and exams.
- Research Assistant: *May 1996 - Aug 1996, Colorado State University*
    - I worked as a research assistant under Prof. Walid Najjar in the Pebbles research group at Colorado State University. I enhanced a

Machine Independent Dataflow Code (MIDC) format that is used by a Sisal compiler for automatic generation of blocking threads from Sisal programs. I also enhanced a cycle accurate multithreaded machine simulator, called ID, to execute the blocking threads. This simulation infrastructure is used for evaluating the memory system performance of multithreaded machines (more details are available in conference paper [16]).

- Teaching Assistant: *August 1994 – May 1996, Colorado State University*
  - Teaching assistant for introductory personal computing course (two semesters), introductory programming language course (one semester) and automata theory course (one semester). Job responsibilities included leading weekly class discussions, lab in-charge and grading homework and exams.

Refereed Publications

Short note on Conference Publications: Here I am paraphrasing information from three relevant articles (1) "Evaluating Computer Scientists and Engineers For Promotion and Tenure," by Prof. David Patterson, *et Al.* (2) "Committee on Academic Careers for Experimental Computer Scientists," by National Research Council (3) "Judging the Impact of Conference and Journal Publications in Computer Architecture," University of California, San Diego Computer Engineering Research Guidelines. In my research area, primarily Computer Architecture, our top tier publication venues are conferences where the conference paper quality meets or exceeds journal publication. Conferences in our field exceed journals in selectivity, visibility, circulation, number of reviews, expertise of reviewers and program committee members. Our conferences have about 5 reviewers and each submitted paper is 20-25 pages in length. Our conference papers represent high quality finished research as they report full evaluations with detailed simulations or models. The program committee consists of about 25 world-renowned researchers and the acceptance rates are less than 30%; I listed acceptance rates where I was able to get the data. Hence, I will first list the refereed full conference papers with acceptance rates wherever available, followed by journals related to my inter-disciplinary work and finally workshop papers. In the list below * represents authors are my PhD students either advised solely by me or co-advised. Underline represents the primary senior author. If multiple primary senior authors they are all underlined in the author list.

Conference Papers

1. **[PETS]** C. Harth-Kitzerow*, A. Suresh, Y. Wang*, H. Yalame, G. Carle, and **M. Annavaram**. High-Throughput Secure Multiparty Computation with an Honest Majority in Various Network Settings In *proceedings of the 25th annual Privacy Enhancing Technologies Symposium (PETS)* July 2025.
2. **[ASPLOS]** Y. Wang*, R, Rajat*, and **M. Annavaram.** MPC-Pipe: Accelerating multi-party computing through pipelining. . In *proceedings of the Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, April 2025**.**

3. **[RECSYS]** K. Balasubramanian*, A. Alsabnah*, E. Markowitz, G. Steeg and **M. Annavaram.** Biased User History Synthesis for Personalized Long-Tail Item Recommendation. In *proceedings of 18th ACM Conference on Recommender Systems*, Oct 2024.
4. **[ACL]** J. Flemings*, **M. Annavaram**. Differentially private knowledge distillation via synthetic text generation. In *2024 Annual Conference of the Association for Computational Linguistics***.** Aug 2024.
5. **[NAACL]** J. Flemings*, M. Razaviyayn, **M. Annavaram**. Differentially private next token generation. In *2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics***.** June 2024.
6. **[NAACL]** L. Gao*, Y. Nie, T. Tang*, S. Avestimehr, **M. Annavaram**. Ethos: Rectifying Language Models in Orthogonal Parameter Space. In *2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics***.** June 2024.
7. **[PETS]** T. Tang*, Y. Niu, S. Avestimehr, and **M. Annavaram**. Edge Private Graph Neural Networks with Singular Value Perturbation. In *proceedings of the 24th annual Privacy Enhancing Technologies Symposium (PETS)* July 2024. (Acceptance rate 99/456, 22%).
8. **[MICRO]** H. Zha*, S. Tannu and **M. Annavaram**. SuperBP: Design Space Exploration for SFQ Branch Predictors. In *proceedings of 55th IEEE/ACM International Symposium on Microarchitecture*, October 2023. (Acceptance rate 101/424, 24%).
9. **[ISCA]** Y. Wang*, R, Rajat*, and **M. Annavaram.** LAORAM: Look Ahead ORAM for Recommendation Model Privacy. . In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2023 (Acceptance rate 79/373, 21%).
10. **[MICRO]** R, Rajat*, Y. Wang*, and **M. Annavaram.** PageORAM: An Efficient DRAM Page Aware ORAM. In *proceedings of 55th IEEE/ACM International Symposium on Microarchitecture*, October 2022. (Acceptance rate 83/369, 22%).
11. **[NAACL]** E.S. Markowitz, K. Balasubramanian*, M. Mirtaheri, **M. Annavaram**, A. Galstyan, G.V. Steeg. StATIK: Structure and Text for Inductive Knowledge Graph Completion. In *2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics***.** July 2022.
12. **[ISPASS]** Y. Wang*, E. Suh, W. Xiong, **M. Annavaram** and H. Lee. Characterization of MPC-based Private Inferences for Transformer-based Models. In *Proceedings of the International Symposium on Performance Analysis of Systems and Software,* May 2022.
13. **[ICASSP]** T. Feng, H. Hashemi*, **M. Annavaram and S. Narayanan.** Enhancing Privacy Through Domain Adaptive Noise Injection for Speech Emotion Recognition. In *proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2022.
14. **[IPDPS]** T. Tang*, H. Hashemi*, R. Ali, S. Avestimehr, and **M. Annavaram.** AVCC: Adaptative Verifiable Coded Computing. In *proceedings of International Conference on Parallel and Distributed Processing Systems*, May 2022 (Acceptance rate 46/474 10% Round1 acceptance).

15. **[AAAI]** E. Ceyani, C. He, K. Balasubramaniyan, **M. Annavaram,** and S. Avestimehr. SpreadGNN: : Serverless Multi-task Federated Learning for Molecular Graphs. In *proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)*, Feb 2022. (Acceptance rate 1349/9251, 15%).
16. **[HPCA]** H. Zha*, N. Katam, M. Pedram and **M. Annavaram**. HiPerRF: A Dual-Bit Dense Storage SFQ Register File. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2022. (Acceptance rate 54/260, 21%).
17. **[NSDI]** A. Eisenman, K. Matam, S. Ingram, D. Mudigere, R, Krishnamoorthi, K. Nair, M. Smelyanskiy and **M. Annavaram**. Check-N-Run: A Checkpointing System for Training Deep Learning Recommendation Models. In *Proceedings of the 19th USENIX Symposium on Networked Systems Design and Implementation (NSDI)* April 2022.
18. **[MICRO]** H. Hashemi, Y. Wang, and **M. Annavaram.** DarKnight: An Accelerated Framework for Privacy and Integrity Preserving Deep Learning Using Trusted Hardware. In *proceedings of 54th IEEE/ACM International Symposium on Microarchitecture*, October 2021. (Acceptance rate 94/423, 22%).
19. **[RECSYS]** K. Balasubramanian*, A. Alsabnah*, J. Choe* and **M. Annavaram.** cDLRM: Look Ahead Caching for Scalable Training of Recommendation Models. In *proceedings of 15th ACM Conference on Recommender Systems*, Oct 2021 (Acceptance rate 49/267, 18%).
20. **[CLOUD]** K. Narra*, Z.Lin*, Y. Wang*, K. Balasubramanian*, and **M. Annavaram.** Origami Inference: Private Inference Using Hardware Enclaves. In *proceedings of IEEE International Conference on Cloud Computing (Short paper)*, Sept 2021 (Acceptance rate 23.8%).
21. **[IPDPS]** K. Matam, H. Hashemi, and **M. Annavaram.** MultiLogVC: Efficient Out-of-Core Graph Processing Framework on Flash Storage. In *proceedings of International Conference on Parallel and Distributed Processing Systems*, May 2021.
22. **[NeurIPS]** C. He, **M. Annavaram**, and S. Avestimehr. Group Knowledge Transfer: Federated Learning of Large CNNs at the Edge. In *Advances in Neural Information Processing Systems*. (Acceptance rate 1900/9454, 20%). Dec 2020.
23. **[ICDCS]** K. Narra, Z.Lin, S. Avestimehr, G. Ananthanarayanan, and **M. Annavaram.** Collage Inference: Using Coded Redundancy for Lowering Latency Variation in Distributed Image Classification Systems/ In *proceedings of International Conference on Distributed Computing Systems)*, July 2020 (Acceptance rate 105/584, 18%).
24. **[SC]** K. Narra*. Z. Lin, M. Kiamari, S. Avestimehr, and **M. Annavaram.** Distributed Matrix Multiplication Using Speed Adaptive Coding. In *proceedings of SuperComputing (SC)*, Nov 2019 (Acceptance rate 72/344, 21%). **Best Paper and Best Student Paper Finalist Nomination.**
25. **[ICS]** Q. Xu*. H. Naghibijouybari, S. Wang, N. Abu-Ghazaleh, and **M. Annavaram.** GPUGuard: Mitigating Contention Based Side and Covert Channel Attacks on GPUs In *proceedings of the International Symposium on SuperComputing (ICS)*, June 2019 (Acceptance rate 45/167, 27%).

26. **[ISCA]** Y. Oh, G. Koo*, **M. Annavaram,** W. Ro. Linebacker: Preserving Victim Cache Lines in Idle Register Files of GPUs. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2019 (Acceptance rate 62/365, 17%).
27. **[ISCA]** K. Matam*, G. Koo*, H. Zha*, H. Tseng, and **M. Annavaram**. GraphSSD: Graph Semantics Aware SSD. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2019 (Acceptance rate 62/365, 17%).
28. **[NIPS]** M. Yu, Z. Lin*, K. Narra*, S. Li, Y. Li, N. Kim, A. Schwing, **M. Annavaram**, and S. Avestimehr. GradiVeQ: Vector Quantization for Bandwidth-Efficient Gradient Aggregation in Distributed CNN Training. In *Advances in Neural Information Processing Systems*, pp. 5129-5139. 2018. (Acceptance rate 1011/4856, 21%).
29. **[IPDPS]** G. Koo*, H. Jeon, Z. Liu, N. Kim and **M. Annavaram**. CTA-Aware Prefetching and Scheduling. In *proceedings of the Symposium on Parallel Distributed Computing,* May, 2018 (Acceptance rate 38/461, 8%).
30. **[HPCA]** A. Tabbakh*, X. Qian and **M. Annavaram**. G-TSC: Timestamp Based Coherence for GPUs. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2018. (Acceptance rate 54/260, 21%).
31. **[MICRO]** G. Koo*, K. Matam*, Te I, K. Narra*, Jing Li, H. Tseng, S. Swanson, and **M. Annavaram**. Flash-Summarizer: Trading Communication with Computing Near Storage. In *proceedings of the IEEE International Symposium on Microarchitecture,* Oct 2017. (Acceptance rate 61/325, 19%)
32. **[ISCA]** G. Koo*, Y. Oh, W. Ro, and **M. Annavaram**. Access Pattern-Aware Cache Management for Improving Data Utilization in GPU. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2017 (Acceptance rate 54/322, 17%).
33. **[IPDPS]** A. Tabbakh*, X. Qian and **M. Annavaram**. Power Efficient Sharing-Aware GPU Data Management. In *proceedings of the Symposium on Parallel Distributed Computing,* June 2017.
34. **[IPDPS]** B. Li, J. Sun, **M. Annavaram** and N.S. Kim. Elastic-Cache: GPU Cache Architecture for Efficient Fine- and Coarse-Grained Cache-Line Management. In *proceedings of the Symposium on Parallel Distributed Computing,* June 2017.
35. **[HPCA]** M. Abdel-Majeed*, H. Jeon*, A.S. Bejestan, M. Pedram, and **Murali Annavaram**. Pilot Register File: Energy Efficient Register File for GPUs. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2017. (Acceptance rate 50/224, 22%)
36. **[HPCA]** Z. Liu, S. Gilani, **Murali Annavaram,** and N.S. Kim. G-Scalar: Cost-effective generalized scalar execution architecture for power-efficient GPUs. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2017. (Acceptance rate 50/224, 22%)
37. **[ISCA]** Q. Xu*, H. Jeon*, K. Kim*, W. Ro, **M. Annavaram**. Efficient Intra-SM Slicing through Dynamic Resource Partitioning for GPU Multiprogramming. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2016. (Acceptance rate 57/291, 20%)

38. **[ISCA]** M. Yoon*, S. Lee*, K. Kim*, G. Koo*, W. Ro and **M. Annavaram**. Virtual Thread: Maximizing Thread-Level Parallelism beyond GPU Scheduling Limit. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2016. (Acceptance rate 57/291, 20%)
39. **[ISCA]** Y. Oh, K. Kim*, M. Yoon*, J. Park, Y. Park, W. Ro, and **M. Annavaram**. APRES: Improving Cache Efficiency by Exploiting Load Characteristics on GPUs. In *proceedings of the International Symposium on Computer Architecture (ISCA)*, June 2016. (Acceptance rate 57/291, 20%)
40. **[ICS]** M. Abdel-Majeed*, D. Wong*, J. Kuang and **M. Annavaram**. Origami: Folding Warps for Energy Efficient GPUs. In *proceedings of the International conference on Supercomputing (ICS)*, June 2016. (Acceptance rate 43/183, 23%)
41. **[HPCA]** D. Wong*, N. Kim and **M. Annavaram**. Warped-Approximation: Using Value Similarity for approximate computing on GPUs. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, March 2016. (Acceptance rate 53/240, 23%)
42. **[HPCA]** K. Kim*, S. Lee, M. Yoon, G. Koo, W. Ro and **M. Annavaram**. Warped-Preexecution: A GPU Prexecution Approach to Improve Latency Hiding. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, March 2016. (Acceptance rate 53/240, 23%)
43. **[MICRO]** H. Jeon*, G. Ravi, N.S. Kim and **M. Annavaram**. GPU register file virtualization. In proceedings of the *IEEE International Symposium on Microarchitecture,* Dec 2015. (Acceptance rate 61/283, 22%)
44. **[IISWC]** G. Koo*, H. Jeon* and **M. Annavaram**. Revealing Critical Loads and Hidden Data Locality in GPGPU applications. In proceedings of the *IEEE International Symposium on Workload Characterization (IISWC),* October 2015.
45. **[IISWC]** D. Wong*, J. Chen* and **M. Annavaram**. Traveling to the Edge of Energy Proportionality. In proceedings of the *IEEE International Symposium on Workload Characterization (IISWC),* October 2015.
46. **[DSN]** M. Abdel-Majeed*, W. Dweik*, H. Jeon* and **M. Annavaram**. Warped-RE: Low-Cost Fault Detection and Correction in GPUs. In proceedings of the *International Conference on Dependable Systems and Networks (DSN),* June 2015.
47. **[ISCA]** S. Kim, K. Kim, G. Koo, H. Jeon, W. Ro, and **M. Annavaram**. Warped-Compression: Enabling Power Efficient GPUs Through Register Register Compression. In proceedings of the *International Symposium on Computer Architecture (ISCA),* June 2015.
48. **[ITC]** H. Jeon*, G. Loh and **M. Annavaram**. " RAS Support for Wide-I/O Stacked DRAM". In proceedings of the *IEEE International Test Conference (ITC),* Nov 2014.
49. **[IISWC]** Q. Xu*, H. Jeon*, and **M. Annavaram**. " Characterization and analysis of GPGPU design for Graph Applications". In proceedings of the *IEEE International Symposium on Workload Characterization (IISWC),* Oct 2014.
50. **[PACT]** Q. Xu* and **M. Annavaram**. "PATS: Pattern Aware Scheduling and Power Gating for GPGPUs". In proceedings of the *International Conference on Parallel Architecture and Compilation Techniques (PACT),* Aug 2014.

51. **[DSN]** W. Dweik*, M. Abdel-Majeed* and **M. Annavaram**. "Warp-Shield: Tolerating Hard Faults in GPGPUs". In proceedings of the *International Conference on Dependable Systems and Networks (DSN),* June 2014.
52. **[GOMAC]** W. Dweik* and **M. Annavaram**. SignTest: Signature-based Adaptive Periodic Testing. I*n Government Microcircuit Applications and Critical Technology Conference, (GOMACTech),* April 2014.
53. **[GOMAC]** W. Dweik*, M. Dubois and **M. Annavaram**. Reliability-Aware Exceptions: Tolerating Intermittent Faults in Microprocessor Array Structures. In *Government Microcircuit Applications and Critical Technology Conference, (GOMACTech),* April 2014.
54. **[DATE]** W. Dweik*, M. Dubois and **M. Annavaram**. Reliability-Aware Exceptions: Tolerating Intermittent Faults in Microprocessor Array Structures. In proceedings of the *Design Automation and Test in Europre (DATE),* March 2014.
55. **[HPCA]** D. Wong *, and **M. Annavaram**. Implications of High Energy Proportional Servers on Cluster-wide Energy Proportionality. In *proceedings of* the *International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2014.
56. **[MICRO]** M. Abdel-Majeed*, D. Wong* and **M. Annavaram**. Warped Gates: Gating Aware Scheduling and Power Gating for GPGPUs. In *proceedings of the International Symposium on Microarchitecture (MICRO)*, Dec 2013.
57. **[DSN]** J. Suh*, M. Dubois and **M. Annavaram**. PHYS: Profiled-Hybrid Sampling for Soft Error Reliability Benchmarking. In proceedings of the *International Conference on Dependable Systems and Networks (DSN),* June 2013.
58. **[HPCA]** M. Abdel-Majeed*, and **M. Annavaram**. Warped Register File: A Power Efficient Register File for GPGPUs. In *proceedings of the International Conference on High-Performance Computer Architecture (HPCA)*, Feb 2013.
59. **[MICRO]** H. Jeon* and **M. Annavaram**. Warped-DMR: Light-weight Error Detection for GPGPUs. In *proceedings of the International Symposium on Microarchitecture (MICRO)*, Dec 2012.
60. **[MICRO]** D. Wong* and **M. Annavaram**. Scaling the energy proportionality wall through server-level heterogeneity. In *proceedings of the International Symposium on Microarchitecture (MICRO)*, Dec 2012.
61. **[IISWC]** S. Lee* and **M. Annavaram**. Wireless Body Area Networks: Where Does the Energy Go?. In *proceedings of the IEEE International Symposium on Workload Characterization (IISWC),* Nov 2012. **(Best Paper Award Nominee)**
62. **[IISWC]** M. Demertzi*, B.Zandian*, R. Rojas* and **M. Annavaram**. Benchmarking Instruction Set Architecture for Non-Transient Errors. In proceedings of the *IEEE International Symposium on Workload Characterization (IISWC),* Nov 2012.
63. **[DSN]** B. Zandian* and **M. Annavaram**. Software-based Infield Wearout Monitoring for Synchronous Digital Chips. In proceedings of the *International Conference on Dependable Systems and Networks (DSN) – Fast Abstracts,* June 2012.

64. **[DSN]** W. Dweik* and **M. Annavaram**. Signature-based Online Periodic Fault Tolerance for Microprocessors. In proceedings of the *International Conference on Dependable Systems and Networks (DSN) – Fast Abstracts,* June 2012.
65. **[SECON]** Y.Wang*, B. Krishnamachari, and **M. Annavaram**. Semi-Markov State Estimation and Policy Optimization for Energy Efficient Mobile Sensing. In *proceedings of 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012*.
66. **[ISQED]** M.Abdel-Majeed*, S.Chen and **M. Annavaram**. A case for 3D stacked analog circuits in high-speed sensing systems. In *proceedings of the IEEE International Symposium on Quality Electronics Design (ISQED),* March 2012.
67. **[HPCA]** J. Suh*, **M. Annavaram** and M. Dubois. MACAU: A Markovian Model for Reliability Evaluation of Caches Under SBUs and MBUs. In proceedings of the *International Symposium on High Performance Computer Architecture (HPCA)*, Feb 2012 (Acceptance rate 36/210, 17%).
68. **[IISWC]** M.Demertzi*, **M. Annavaram** and M. Hall. Analyzing the effect of compiler optimizations on application reliability. In *proceedings of the 2011 IEEE International Symposium on Workload Characterization (IISWC),* Nov 2011.
69. **[EMBC]** S. Kim, M. Li, S. Lee*, U. Mitra, A. Emken, D. Spruijt-Metz, **M. Annavaram**, S. Narayanan. High-level Descriptions of Real-Life Physical Activities Using Latent Topic Modeling of Multimodal Sensor Signals. In proceedings of the *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* Aug 2011.
70. **[DSN]** B.Zandian* and **M. Annavaram**. Cross Layer Resilience Using Wearout Aware Design Flow. In *proceedings of the International Conference on Dependable Systems and Networks (DSN),* pages 279-290, June 2011 (Acceptance rate 26/148, 18%).
71. **[ISCA]** M. Manocheri*, **M. Annavaram** M. Dubois. CPCC: Correctable Parity Protected Cache. In *proceedings of* the *International Symposium on Computer Architecture (ISCA)*, pages 223-234, June 2011 (Acceptance rate 40/208, 19%).
72. **[SIGMETRICS]** J. Suh*, M. Manocheri, **M. Annavaram** and M. Dubois. Soft Error Benchmarking of L2 Caches with PARMA. In *proceedings of the International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, pages 85-96, June 2011 (Acceptance rate 26/177, 15%).
73. **[HPCA]** N. Madan*, A. Buyuktosunoglu, P. Bose and **M. Annavaram**. Guarded Power Gating in a Multi-Core Setting. In *proceedings of the International Symposium on High Performance Computer Architecture (HPCA)*, pages 291-300, Feb 2011 (Acceptance rate 46/238, 19%).
74. **[MICRO]** J. Chen, K. Dabbiru*, D. Wong*, **M. Annavaram** and M. Dubois. Adaptive and Speculative Slack Simulations of CMPs on CMPs. In *proceedings of the International Symposium on Microarchitecture (Micro)*, pages 523-534, Dec 2010. (Acceptance rate: 45/245, 18%)
75. **[DSN]** B. Zandian*, W. Dweik*, S. Kang*, T. Punihaole* and **M. Annavaram**. Continuous Reliability Monitoring Using Adaptive Critical Path Testing. In *proceedings of the International Conference on Dependable Systems and Networks (DSN),* pages 151-160, July 2010 (Acceptance rate: 39/170, 23%)

76. **[IPSN]** Y.Wang*, B. Krishnamachari, Q. Zhao and **M. Annavaram**. Markov-Optimal Sensing Policy for User State Estimation in Mobile Devices. In *proceedings of 9th International Conference on Information Processing in Sensor Networks 2010 - IP Track (IPSN-2010)*, pages 268-278, Apr 2010 (Acceptance rate: 20/117, 17%).
77. **[ITC]** T-Y. Hsieh, M. Breuer, **M. Annavaram**, S. Gupta, K-J. Lee. Tolerance of Performance-degrading Faults for Effective Yield Improvement. In *Proceedings of the 40th Annual International Test Conference (ITC)*, Pages 1-10, Nov 2009.
78. **[MOBICASE]** Y.Wang*, B. Krishnamachari, Q. Zhao and **M. Annavaram**. Towards the Tradeoff between Energy Efficiency and User State Estimation Accuracy in Mobile Sensing. In *Proceedings of the International Conference on Mobile Computing, Applications, and Services (MOBICASE),* Oct 2009.
79. **[EMBC]** G. Thatte, M. Li, A. Emken, U. Mitra, S. Narayanan, **M. Annavaram** and D. Spruijt-Metz. Energy-Efficient Multihypothesis Activity-Detection for Health-Monitoring Applications. *In Proceedings of the 31st Annual International Conference of IEEE Engineering in Medicine and Biology (EMBC09)*, September 2009.
80. **[ICPP]** J. Chen, **M. Annavaram** and M. Dubois. Exploiting Simulation Slack to Improve Parallel Simulation Speed. In *Proceedings of the 2009 International Conference on Parallel Processing (ICPP),* pages 371-378, Sept 2009. (Acceptance Rate: 30%)
81. **[DCOSS]** G. Thatte, V. Rozgic, M. Li, S. Ghosh*, U. Mitra, S. Narayanan, **M. Annavaram**, and D. Spruijt-Metz. Optimal allocation of time-resources for multi hypothesis activity-level detection. In *Proceedings of the 5th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS'09)*, June 2009. **(Best Paper Award)**
82. **[MOBISYS]** Y. Wang*, J. Lin, **M. Annavaram**, Q. A. Jacobson, J. Hong, B. Krishnamachari and N. Sadeh. A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition. In *proceedings of 7th Annual International Conference on Mobile Systems, Applications and Services (MobiSys)*, pages 179-192, June 2009 (Acceptance Rate, 26/128, 20%).
83. **[SIGMETRICS]** J. Chen, **M. Annavaram**, M. Dubois. Slacksim: : A Platform for Parallel Simulations of CMPs on CMPs. In *proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, June 2009. (Acceptance Rate: 26%)
84. **[MSWIM]** H. Liu, B. Krishnamachari and **M. Annavaram**. Game theoretic approach to location sharing with privacy in a community-based mobile safety application. In *Proceedings of the 11th international Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems* MSWiM '08. Pages 229-238, October 2008.
85. **[MOBISYS]** B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, A.M. Bayen, **M. Annavaram**, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *Proceedings of the 6th international Conference on Mobile Systems, Applications, and Services*, pages 15-28, June, 2008. (Acceptance Rate: 22/123, 18%)

86. **[HPCA] M. Annavaram**, E. Grochowski, and P. Reed. Implications of Device Timing Variability on Full Chip Timing. In *Proceedings of the 13th International High-Performance Computer Architecture*, pages 37-45, Feb 2007. (Acceptance rate 32/184, 17%)
87. **[MICRO]** B. Black, **M. Annavaram**, N. Brekelbaum, J. DeVale, L. Jiang, G.H. Loh, D. McCauley, P. Morrow, D.W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. Shen, and C. Webb. Die Stacking (3D) Microarchitecture. In *Proceedings of the 39th International Symposium on Microarchitecture*, pages 469-479, Dec 2006 (Acceptance Rate :42/174, 24%). (B.Black and M. Annavaram are senior authors with equal contributions)
88. **[ISCA] M. Annavaram**, E. Grochowski, J. Shen. Mitigating Amdahl's Law Through EPI Throttling. In *Proceedings of the 32nd International Symposium on Computer Architecture*, pages 298-309, June 2005 (Acceptance Rate 45/194, 23%).
89. **[MICRO] M. Annavaram**, R. Rakvic, M. Polito, J. Bouguet, R. Hankins, B. Davies. The Fuzzy Correlation between Code and Performance Predictability. In *Proceedings of the 37th International Symposium on Microarchitecture*, pages 93-104, Dec 2004 (Acceptance Rate: 29/158, 18%).
90. **[MICRO]** R. Hankins*, T. Diep, **M. Annavaram**, B. Hirano, H. Eri, H. Nueckel, and J. Shen. Scaling and Characterizing Database Workloads: Bridging the Gap between Research and Practice. In *Proceedings of the 36th International Symposium on Microarchitecture*, pages 151-162, Dec 2003 (Acceptance Rate: 35/134, 26%)
91. **[ICCD] M. Annavaram**, T. Diep and J. Shen. Branch Behavior of a Commercial OLTP Workload on Intel IA32 Processors. In *Proceedings of the International Conference on Computer Design*, pages 242-248, Sept 2002.
92. **[ISCA] M. Annavaram**, J. M. Patel and E. S. Davidson. Data Prefetching by Dependence Graph Precomputation. In *Proceeding of the 28th International Symposium on Computer Architecture*, pages 52-61, July 2001 (Acceptance Rate 24/163, 15%). (E.S. Davidson is Annavaram's PhD Thesis advisor)
93. **[HPCA] M. Annavaram**, J. M. Patel and E. S. Davidson. Call Graph Prefetching for Database Applications. In *Proceedings of the Seventh International High-Performance Computer Architecture*, pages 281-290, Jan 2001 (Acceptance Rate: 26/110, 24%)
94. **[ISPASS] M. Annavaram**, G. S. Tyson and E. S. Davidson. Instruction Overhead and Data Locality Effects in Superscalar Processors. In *Proceedings of the International Symposium on Performance Analysis of Systems and Software,* pages 95-100, April 2000.
95. **[IPDPS] M. Annavaram**, W. Najjar. Comparison of Two Storage Models in Data-Driven Multithreaded Architectures. In *proceedings of the Symposium on Parallel Distributed Computing,* pages 100-110, October 1996. (W. Najjar is MS Thesis advisor)

Journal Papers

1. **[IEEE TAS}** C. Fourie et. Al., 2023. Results from the ColdFlux Superconductor Integrated Circuit Design Tool Project. *IEEE Transactions on Applied Superconductivity*.
2. **[Optics]** A. Minoofar, N, Karapetyan, A, Almaiman, H. Zhou, H. Song, K. Zou, W. Ko, M. Ramakrishnan, M. Annavaram, J.L. Habif, M. Tur, and A.E. Willner, Demonstration of optically-assisted reconfigurable average of two 20-Gbaud 4-phase-encoded data channels using nonlinear wave mixing. *Opt. Lett. 48, 4617-4620 (2023).*
3. **[IEEE TAS]** M. Pedram, M. Annavaram et al., ColdFlux Superconducting EDA and TCAD Tools Project: Overview and Progress. *IEEE Transactions on Applied Superconductivity,* Dec 2018.
4. **[IEEE VLSI]** K. Matam*, M. Abdel-Majeed*, M. Annavaram. Efficient automatic parallelization of a single GPU program for a multiple GPU system. *IEEE Journal of VLS-Integration. Dec 2018.*
5. **[IEEE TOC]** Y. Oh, K. Kim, M. Yoon, J. Park, Y. Park, M. Annavaram, W. Ro**.** Adaptive Cooperation of Prefetching and Warp Scheduling on GPUs. *IEEE Transactions on Computer, Oct 2018*
6. **[IEEE TOC]** S. Lee, K. Kim, G. Koo*, H. Jeon*, **M. Annavaram** and W. Ro. Improving Energy Efficiency of GPUs through Data Compression and Compressed Execution. IEEE Transactions on Computers, Vol. 66, no. 5, Oct 2016.
7. **[IEEE TOC]** M. Manocheri*, **M. Annavaram** M. Dubois. CPCC: Correctable Parity Protected Cache. IEEE Transactions on Computers, Vol.63, no. 10, Oct 2014.
8. **[MICRO TOPPICKS]** D. Wong* and **M. Annavaram**. Scaling the energy proportionality wall through server-level heterogeneity. In *IEEE MICRO TOPPICKS*, Vol 33, no.3, June 2013.
9. **[IEEE TOC]** K. Patel*, **M. Annavaram and M. Pedram**. NFRA: Generalized Network Flow Based Resource Allocation for Hosting Centers. In *IEEE Transactions on Computers,* Vol 62, no.9, 2013.
10. **[COMM]** U. Mitra, A. Emken, S. Lee, M. Li, V. Rozgic, G. Thatte, H. Vathsangam, D. Zois, M. Levorato, **M. Annavaram**, S, Narayanan and D. Spruijt-Metz, G Sukhatme. KNOWME: A Case Study in Wireless Body Area Sensor Network Design. In *IEEE Communication Magazine,* Vol 50, no. 5, March 2012.
11. **[JPAH]** A. Emken, M. Li, G. Thatte, S. Lee*, **M. Annavaram**, U. Mitra, S, Narayanan and D. Spruijt-Metz. Recognition of Physical Activities in Overweight Hispanic Youth Using KNOWME Networks. In *Journal of Phyiscal Activity and Helath,* Vol.9, no.3*,* March 2012.
12. **[ToMC]** B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, A.M. Bayen, **M. Annavaram**, and Q. Jacobson. Enhancing Privacy and Accuracy in Probe Vehicle-Based Traffic Monitoring via Virtual Trip Lines. In *IEEE Transactions on Mobile Computing,* Vol. 11, no.5, May 2012.
13. **[TECS]** G. Thatte, M. Li, S. Lee*, A. Emken, S. Narayanan, U. Mitra, D. Spruijt-Metz and **M. Annavaram**. KNOWME: An Energy-Efficient and Multimodal Body Area Sensing System for Physical Activity Monitoring. *ACM Transactions*

in Embedded Computing Systems (TECS), Special Issue on Wireless Health Systems. Vol. 11, no. S2, August 2012.

14. **[TSP]** G. Thatte, M. Li, S. Lee*, A. Emken, **M. Annavaram**, S. Narayanan, D. Spruijt-Metz and U. Mitra. Optimal Time-Resource Allocation for Energy-Efficient Physical Activity Detection. *IEEE Transactions on Signal Processing,* Vol. 59, no.4, April 2011.
15. **[LNCS]** N. Madan*, A. Buyuktosunoglu, P. Bose, and, **M. Annavaram**. Guarded Power Gating in a Multi-Core Setting. *Lecture Notes in Computer Science*, Springer Publication, Vol 6162, Feb 2011.
16. **[LNCS]** S. Ghosh*, M. Redekopp and **M. Annavaram**. KnightShift: Shifting the I/O Burden in Datacenters to Management Processor for Energy Efficiency. *Lecture Notes in Computer Science*, Springer Publication, Vol 6161, 2010.
17. **[ToNSRE]** M. Li, V. Rozgic, G. Thatte, S. Lee*, A. Emken, **M. Annavaram**, U. Mitra, D. Spruijt-Metz and S, Narayanan. Multimodal Physical Activity Recognition by Fusing Temporal and Cepstral Information. *Transactions on Neural Systems & Rehabilitation Engineering,* Vol. 18, no.4*,* 2010.
18. **[CAN]** J. Chen, **M. Annavaram**, and M. Dubois. SlackSim: a platform for parallel simulations of CMPs on CMPs. *SIGARCH Computer Architecture News*, Vol. 37, no.2, July 2009.
19. **[TOCS] M. Annavaram**, J. M. Patel and E. S. Davidson. Call Graph Prefetching for Database Applications. *ACM Transactions on Computer Systems,* Vol. 21, no.4, Nov 2003.

Workshop and Invited Papers

1. **[SEAS]** C. Jiang*, L. Gao*, **M. Annavaram**. Efficient LLM Inference with I/O-Aware Partial KV Cache Recomputation. *SEAS Workshop @ AAAI 2025*, Feb 2025.
2. **[ENLSP]** L. Gao*, A. Ziashahabi. Y. Nue, S. Avestimehr, **M. Annavaram**. Enabling Resource-Efficient On-Device Fine-Tuning of LLMs Using Only Inference Engines. *ENLSP workshop @ NeurIPS 2024,* Dec 2024.
3. **[SATA]** J. Flemings*, B. Ziang, Z. Takhirov, **M. Annavaram.** Characterizing Context Memorization and Hallucination of Language Models, *SATA Workshop @ NeurIPS 2024,* Dec 2024.
4. **[ReML]** L. Gao*, Y. Niu, T. Tang*, S. Avestimehr, **M. Annavaram**. Ethos: Rectifying Language Models in Orthogonal Parameter Space. *ReML workshop@AAAI 2024,* Spotlight Presentation, Feb 2024.
5. **[PPAI]** J. Flemings*, M. Razaviyayn, **M. Annavaram**. Differentially Private Prediction of Large Language Models. Accepted for *PPAI workshop@AAAI 2024,* Feb 2024.
6. **[DPML]** H. Hashemi*, Y. Wang*, **M. Annavaram**. Privacy and Integrity Preserving Training Using Trusted Hardware, *DPML workshop@ICLR 2021, May 2021.*
7. **[SSML]** H. Hashemi*, Y. Wang*, C. Guo, **M. Annavaram**. Byzantine-Robust and Privacy-Preserving Framework for FedML, *SSML workshop@ICLR 2021, May 2021.*

8. **[DPML]** C. He, K. Balasubamanian, E. Ceyani, Y. Ron, P. Zhao, J. Huang, **M. Annavaram, and S. Avestimehr.** FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks, *DPML workshop@ICLR 2021,* May 2021.
9. **[GNNSys]** C. He, K. Balasubamanian, E. Ceyani, Y. Ron, P. Zhao, J. Huang, **M. Annavaram, and S. Avestimehr.** FedGraphNN: A Federated Learning System and Benchmark for Graph Neural Networks, *GNNSys workshop@MLSys 2021,* March 2021.
10. **[SPICYFL]** C. He, S. Li, …,**M. Annavaram, and S. Avestimehr.** FedML: A Research Library and Benchmark for Federated Machine Learning, *SpicyFL workshop, Dec* 2020.
11. **[CWWMCA]** H. Hashemi, and **M. Annavaram.** DarKnight: Privacy protected machine learning, *6th Career Workshop for Women and Minorities in Computer Architecture),* Oct 2020.
12. **[CWWMCA]** T. Tan, and **M. Annavaram.** Scalability of coded machine learning, *6th Career Workshop for Women and Minorities in Computer Architecture),* Oct 2020.
13. **[NAS]** C. He, **M. Annavaram, and S. Avestimehr.** Towards non-IID and invisible data with FedNAS: Federated Deep Learning via Neural Architecture Search, *Workshop on Neural Architecture Search and Beyond for Representation Learning,* Aug 2020.
14. **[EUCAS]** N. Katam, H. Zha, M. Pedram and **M. Annavaram.** Steroid: A High Performance SFQ Architecture Using Dual-bit Boolean Logic, In *Proceedings of the 14th European Conference on Applied Superconductivity (**EUCAS**),* Sept 2019.
15. **[EUCAS]** N. Katam, H. Zha, M. Pedram and **M. Annavaram.** High Capacity Fluxon Storage for Microprocessor Design. In *Proceedings of the 14th European Conference on Applied Superconductivity (**EUCAS**),* Sept 2019.
16. **[ISEC] M. Annvaraum**, P. Beerel, S. Gupta, S. Nazarem, and M. Pedram. Progress Towards azn Open-Source Front End CAD Flow for RSFQ. In *proceedings of 17th International Superconductive Electronics Conference,* July 2019.
17. **[CodML]** K. Narra, Z. Lin, G. Ananthanarayana, S. Avestimeher, and **M. Annavaram.** Collage-CNN: Using Coded Redundancy for Low Variance Distributed Image Classification, In *Proceedings of the Coded Machine Learning Workshop (**CodML**),* July 2019.
18. **[GPUDEPEND]** W. Dweik*, M. Abdel-Majeed* and **M. Annavaram**. Tolerating Hard Faults in GPGPUs. In the *International Workshop on Dependable GPU Computing*, March 2014.
19. **[WEED]** D. Wong* and **M. Annavaram**. Evaluating A Prototype KnightShift-enabled Server. In the *Workshop on Energy Efficient Design*, June 2012.
20. **[MOBS]** J.Chen, K. Dabbiru*, **M. Annavaram** and M. Dubois. Adaptive and Speculative Slack Simulations of CMPs on CMPs. In the Sixth Annual Workshop on Modeling, Benchmarking and Simulation *(MOBS)*, June 2010.

21. **[MOBS]** J. Suh*, **M. Annavaram** and M. Dubois. Soft Error Benchmarking for L2 Cache with PARMA. In the Sixth Annual Workshop on Modeling, Benchmarking and Simulation *(MOBS)*, June 2010.
22. **[WEED]** S. Ghosh*, M. Redekopp and **M. Annavaram**. KnightShift: Shifting the I/O Burden in Datacenters to Management Processor for Energy Efficiency. In the *Workshop on Energy Efficient Design*, June 2010.
23. **[WEED]** N. Madan*, A. Buyuktosunoglu, P. Bose, and, **M. Annavaram**. Guarded Power Gating in a Multi-Core Setting. In the *Workshop on Energy Efficient Design*, June 2010.
24. **[CENS]** G. Thatte, M. Li, A. Emken, U. Mitra, S. Narayanan, **M. Annavaram** and D. Spruijt-Metz. Energy-Efficient Activity-Detection via Multihypothesis Testing for Pediatric Obesity. In the *7th Annual CENS Research Review*, October 2009.
25. **[mHealth]** D. Spruijt-Metz, S. Narayanan, U. Mitra, G. Sukhatme, M. Li, G. Thatte, A. Emken, S. Lee, H. Vathsangam and **M. Annavaram**. KNOWME Networks: Mobile Device Biomonitoring to Prevent and Treat Obesity in Underserved Minority Youth. In *the mHealth Summit*, October 2009.
26. **[BIC]** S. Lee*, **M. Annavaram**, G. Thatte, V. Rozgic, M.Li, U. Mitra, S. Narayanan, A. Emken, and D. Spruijt-Metz. Sensing for Obesity: KNOWME Implementation and Lessons for an Architect. In Workshop on Biomedicine in Computing: Systems, Architectures, and Circuits (BiC), June 2009.
27. **[ICDAM]** D. Spruijt-Metz, M. Li, G. Thatte, G. Sukhatme, **M. Annavaram**, S. Ghosh, V. Rozgic, U. Mitra, N. Medvidovic, B. Belcher, and S. Narayanan. Differentiating physical activity modalities in youth using heartbeat waveform shape and differences between adjacent waveforms. In *proceedings of the 7th International Conference on Diet and Activity Methods (ICDAM 7)*, June 2009.
28. **[BodyNets]** G. Thatte, V. Rozgic, M. Li, S. Ghosh*, U. Mitra, S. Narayanan, **M. Annavaram**, and D. Spruijt-Metz. Optimal time-resource allocation for activity-detection via multimodal sensing. In *proceedings of the Fourth International Conference on Body Area Networks*, April 2009.
29. **[URBANSENSE]** **M. Annavaram**, N. Medvidovic, U. Mitra, S. Narayanan, G. Sukhatme, Z. Meng, S. Qiu, R. Kumar, G. Thatte, and D. Spruijt-Metz. Multimodal sensing for pediatric obesity applications. In *Proceedings of UrbanSense08,* November 2008.
30. **[MODUS]** **M. Annavaram**, Q. Jacobson. HangOut: A Privacy Preserving Location Based Social Networking Service. In *Workshop on Mobile Devices and Urban Sensing*, **(Invited Paper)** April 2008.
31. **[ISPD]** **M. Annavaram**, E. Grochowski, and P. Reed. Implications of Device Timing Variability on Full Chip Timing. In *Proceedings of the International Symposium on Physical Design*, **(Invited Paper)** April 2008.
32. **[CAECW]** R. Hankins*, **M. Annavaram**, B. Hirano, J. Patel and J. Shen. Comparing OLTP Scaling Behavior on Intel® Xeon™ and Itanium® 2 Processors. In *the seventh Workshop on Computer Architecture Evaluation using Commercial Workloads*, Feb 2004.
33. **[MEM]** P. Kundu, **M. Annavaram**, T. Diep and J. Shen. A Case for Shared Instruction Cache on Chip Multiprocessors Running OLTP. In *the Memory*

Performance: Dealing With Applications, Systems and Architectures Workshop, Sep 2003.

34. **[CAECW]** T. Diep, **M. Annavaram**, B. Hirano and J. Shen. Analyzing Performance Characteristics of OLTP Cached Workloads by Linear Interpolation. In *the sixth Workshop on Computer Architecture Evaluation using Commercial Workloads*, Feb 2003.
35. **[WWC]** J. Rupley II, **M. Annavaram**, J. DeVale, T. Diep and B. Black. Comparing and Contrasting a Commercial OLTP Workload with CPU2000 on IPF. In *the fifth Workshop on Workload Characterization*, Nov 2002.
36. **[EPIC]** R. Rakvic, E. Grochowski, B. Black, **M. Annavaram**, T. Diep and J. Shen. Performance Advantage of the Register Stack in Intel Itanium Processors. In *the second Workshop on Explicitly Parallel Instruction Computing Architecture and Compilers*, Nov 2002.

Patents

1. US-2020310690-A1, *Dynamic near-data processing control mechanism based on computer resource availability on solid-state disk platforms,* Murali Annavaram, Gunjae Koo, Kiran Kumar Matam, Hung-Wei Tseng. Filed 10/2017.
2. US-9008735-B2, *Runtime selection of most energy-efficient approach for services requested by mobile applications*, Murali Annavaram, Sangwon Lee. Issued 4/2015.
3. US-7437581-B2, *Method and apparatus for varying energy per instruction according to the amount of available parallelism, Edward Grochowski, John Shen, Hong Wang, Doron Orenstein, Gad S Sheaffer, Ronny Ronen, Murali Annavaram.* Issued 10/2008.
4. US-2012131366-A1, *Load balancing for multi-threaded applications via asymmetric power throttling*, Ryan Rakvic, Hankins Richard A, Ed Grochowski, Hong Wang, Murali Annavaram, Poulsen David K, Sanjiv Shah, John Shen, Gautham Chinya. Issued 9/2014.
5. US-8631290-B2, Automated *detection of and compensation for guardband degradation during operation of clocked data processing circuit*, Bardia Zandian, Murali Annavaram. Issued 1/2014.
6. US-2010052730-A1, *Method and apparatus for late timing transition detection*, Edward Grochowski, Chris Wilkerson, Lu Shih-Lien L, Murali Annavaram. Issued 2/2012.
7. HK-1121257-A1, *Prefetching from a dynamic random access memory to a static random access memory*, Bryan P Black, Murali Annavaram, Donald W Mccauley, John P Devale. Issued 10/2011.
8. WO-2009068970-A1, *Methods, apparatuses, and computer program products for traffic data aggregation using virtual trip lines and a combination of location and time based measurement triggers in gps-enabled mobile handsets*, Jacobson Quinn, Hoh Baik, Murali Annavaram, David Sutter. Issued 3/2011.
9. US-2009143966-A1, *Methods, apparatuses, and computer program product for traffic data aggregation using virtual trip lines and gps-enabled mobile handsets, Quinn* Jacobson, Baik Hoh, Murali Annavaram. Issued 10/2010.

10. US-7480838-B1, *Method, system and apparatus for detecting and recovering from timing errors,* Chris Wilkerson, Shih-Lien L. Lu, Edward Grochowski, Murali Annavaram. Issued 1/2009.
11. US-2007220207-A1, *Transferring data from stacked memory*, Bryan Black, Murali Annavaram, Paul Reed. Issued 9/2007.
12. US-2004049666-A1 *Method and apparatus for variable pop hardware return address stack,* Murali Annavaram, Trung Diep, John Shen. Issued 3/2004.