Arun Subramaniyan

Email : arunsub@umich.edu

Research interests: Near-Data Computing, Memory Systems, HW/SW Co-design

Education

- **University of Michigan** — Ann Arbor, MI
  PhD. Candidate, Computer Science and Engineering; Advisor: Reetuparna Das; GPA: 4.0/4.0 — *Aug. 2016 – present*
- **Birla Institute of Technology and Science - Pilani (BITS-Pilani)** — Pilani, India
  B.E.(Hons) in Electrical and Electronics; GPA: 9.62/10.0 — *Aug. 2011 – July. 2015*

Awards

- **IEEE Micro Top Picks, 2018.** Awarded to Neural Cache which re-purposes caches into a data-parallel accelerator for deep neural networks.
- **UM Precision Health Scholar Award, 2018.** One of 12 recipients of university-wide precision health initiative.
- **Rackham International Student Fellowship, 2017.** This is a university-wide award given to 25 international graduate students for outstanding academic performance.
- **CFAR Annual Review Workshop Demo Winner, 2016.** Awarded as co-author of Compute Caches which enables in-place computation in caches. This workshop showcased nearly 50 projects in computer architecture related topics from several leading institutions.
- **Best Paper Award, CODES+ISSS, 2016.** Awarded to R^{2}Cache, a reconfigurable last-level cache architecture to improve reliability to soft errors.
- **DAAD WISE Scholarship, 2014.** Awarded the DAAD Wise Scholarship for a three month research internship at a public-funded German Institution.

Experience

- **University of Michigan** — Ann Arbor, MI
  Graduate Student Research Assistant, Advisor: Reetuparna Das — *Aug 2016 - present*
  - **Hardware Acceleration for Precision Health (ISCA'18, VLSI'20, ISCA'21)**: This work enables FPGA-accelerated gene sequencing analysis in the cloud (e.g., AWS EC2 F1) while producing clinically relevant results. Towards this end we redesigned software tools and developed new data structures and algorithms to accelerate the key time consuming computations in the BWA-GATK pipeline: read alignment and variant calling. In collaboration with Intel, we developed a genomics benchmark suite and are working on designing a domain-specific architecture for genomics.
    Impact: Our accelerated software implementation has been validated and integrated into popular read alignment tool BWA-MEM2 (ert branch)
  - **In-Memory Automata Processing (ISCA'17, MICRO'17, MICRO'18)**: In this project we leverage enumerative parallelization techniques to accelerate finite state automata computation on Micron's Automata Processor. Further, we developed techniques to re-purpose the last level caches in general purpose processors for automata processing using novel SRAM-based interconnects. To support parsing of recursively nested data, we also developed a custom datapath for deterministic pushdown automata computation in cache.
  - **Bit-Serial Arithmetic in Caches: Neural Network Acceleration (HPCA'17, ISCA'18)**: In this work, we leverage bitline computing techniques to transform the passive last-level cache real-estate in modern processors into massive bit-serial computational units that can accelerate deep neural networks. One of the key insights from this work was that caches can compute at a much higher frequency, since in-situ computation requires only SRAM lookups and does not incur the overheads of a traditional cache access.
- **Microsoft Research** — Redmond, WA
  Research Intern, Systems Research Group, Manager: Bill Bolosky — *May 2019 - Aug 2019, June 2020 - Aug 2020*
  - **Fast and accurate gene sequencing analysis with SNAP**: Worked on improving the accuracy of the SNAP aligner. SNAP's aligner is 2–4× faster than BWA-MEM2, 3–6× faster than Bowtie2 and 20–30× faster than Novoalign on on the Genome-in-a-Bottle and Illumina Platinum Genome samples.
    Impact: Proposed improvements contributed to the release of SNAP 1.0.0 and a manuscript is under preparation
- **Karlsruhe Institute of Technology** — Karlsruhe, DE
  Student Research Assistant, Advisor: Muhammad Shafique — *Jan 2015 - Jul 2015*
  - **Reliability Adaptive Cache Hierarchies in Multi-Cores (CODES+ISSS'15, TECS'16, DATE'17)**: In this work we showed that the soft error vulnerability of the cache hierarchy is highly dependent on the cache access patterns of different applications and the parameters of different cache levels and proposed analytical models to simultaneously reconfigure multiple cache levels to improve their reliability against soft errors.
- **Nvidia Graphics Pvt Ltd.** — Bangalore, IN
  Architect, Manager: Jay Gupta — *Aug 2015 - Aug 2016*
  - **Tegra SoC Memory Controller Architecture**: My focus was on designing high-throughput QoS-aware arbitration schemes for the various crossbars in the Xavier SoC memory controller and developing performance models. I was also involved in post-silicon memory subsystem performance studies and parameter tuning.

Publications (conferences)

- **Accelerated Seeding for Genome Sequence Alignment with Enumerated Radix Trees**: Arun Subramaniyan, Jack Wadden, Kush Goliya, Nathan Ozog, Xiao Wu, Satish Narayanasamy, David Blaauw, Reetuparna Das, ISCA'2021 *(to appear)*
- **GenomicsBench: A Genomics Benchmark Suite**: Arun Subramaniyan, Yufeng Gu, Timothy Dunn, Somnath Paul, Md Vasimuddin, Sanchit Misra, David Blaauw, Satish Narayanasamy, Reetuparna Das, ISPASS'2021 *(to appear)*
- **17.3 GCUPS Pruning-Based Pair-Hidden-Markov-Model Accelerator for Next-Generation DNA Sequencing**: Xiao Wu, Arun Subramaniyan, Zhehong Wang, Satish Narayanasamy, Reetu Das, and David Blaauw, VLSI'2020.
- **A 2.46 M reads/s Genome Sequencing Accelerator using a 625 Processing-Element Array**: Zhehong Wang, Tianjun Zhang, Daichi Fujiki, Arun Subramaniyan, Xiao Wu, Makoto Yasuda, Satoru Miyoshi, Masaru Kawaminami, Reetuparna Das, Satish Narayanasamy, David Blaauw, CICC'2020
- **Memory Bandwidth-Aware Seeding For Genome Sequencing**: Arun Subramaniyan, Jack Wadden, Kush Goliya, Xiao Wu, Satish Narayanasamy, David Blaauw, Reetuparna Das American Society for Human Genetics Meeting, ASHG'2019 (Poster)
- **A compute SRAM with bit-serial integer/floating-point operations for programmable in-memory vector acceleration**: Jingcheng Wang, Xiaowei Wang, Charles Eckert, Arun Subramaniyan, Reetuparna Das, David Blaauw, and Dennis Sylvester, ISSCC'2019
- **ASPEN: A Scalable In-SRAM Architecture for Pushdown Automata**: Kevin Angstadt, Arun Subramaniyan, Elaheh Sadredini, Reza Rahimi, Kevin Skadron, Westley Weimer, Reetuparna Das, MICRO'2018
- **GenAx: A Genome Sequencing Accelerator**: Daichi Fujiki*, Arun Subramaniyan*, Tianjun Zhang*, Yu Zeng, Reetuparna Das, David Blaauw, Satish Narayanasamy, ISCA'2018. *equal contribution
- **Neural Cache: Bit-Serial In-Cache Acceleration of Deep Neural Networks**: Charles Eckert, Xiaowei Wang, Jingcheng Wang, Arun Subramaniyan, Ravi Iyer, Dennis Sylvester, David Blaauw, Reetuparna Das, ISCA'2018.
  IEEE Micro Top Picks
- **Cache Automaton**: Arun Subramaniyan, Jingcheng Wang, Ezhil R.M. Balasubramanian, David Blaauw, Dennis Sylvester, Reetuparna Das, MICRO'2017
- **Parallel Automata Processor**: Arun Subramaniyan, Reetuparna Das, ISCA'2017
- **Compute Caches**: Shaizeen Aga, Supreet Jeloka, Arun Subramaniyan, Satish Narayanasamy, David Blaauw, Reetuparna Das, HPCA'2017
- **Soft Error-Aware Architectural Exploration for Designing Reliability Adaptive Cache Hierarchies in Multi-Cores**: Arun Subramaniyan, Semeen Rehman, Muhammad Shafique, Akash Kumar, Jörg Henkel, DATE'2017
- **R^{2}Cache: Reliability-Aware Reconfigurable Last-Level Cache Architecture for Multi-Cores**: Florian Kriebel, Arun Subramaniyan, Semeen Rehman, Segnon Jean Bruno Ahandagbe, Muhammad Shafique, Jörg Henkel, CODES+ISSS'2015
  Best Paper Award

Publications (journals)

- **A High-Throughput Pruning-based Pair-Hidden-Markov-Model Hardware Accelerator for Next-Generation DNA Sequencing**: Xiao Wu, Arun Subramaniyan, Zhehong Wang, Satish Narayanasamy, Reetuparna Das, and David Blaauw, IEEE Solid-State Circuits Letters, 2020.
- **A 2.46M reads/s Seed-Extension Accelerator for Next-Generation-Sequencing using a String-Independent PE Array**: Zhehong Wang, Tianjun Zhang, Daichi Fujiki, Arun Subramaniyan, Xiao Wu, Makoto Yasuda, Satoru Miyoshi, Masaru Kawaminami, Reetuparna Das, Satish Narayanasamy, David Blaauw, IEEE Journal of Solid-State Circuits, 2020
- **A 28-nm Compute SRAM With Bit-Serial Logic/Arithmetic Operations for Programmable In-Memory Vector Computing**: Jingcheng Wang, Xiaowei Wang, Charles Eckert, Arun Subramaniyan, Reetuparna Das, David Blaauw, and Dennis Sylvester, IEEE Journal of Solid-State Circuits, 2019
- **Reliability-aware adaptations for shared last-level caches in multi-cores**: Florian Kriebel, Semeen Rehman, Arun Subramaniyan, Segnon Jean Bruno Ahandagbe, Muhammad Shafique, and Jörg Henkel, ACM Transactions on Embedded Computing Systems, 2016

Service

- **Reviewer:** Reviewed papers for ACM TACO (Transactions on Architecture and Code Optimization) 2019, IEEE Transactions on Computers (TC) 2020.
- **Artifact Evaluation Committee:** ASPLOS 2020, PLDI 2020.
- **Shadow Program Committee:** Eurosys 2021.
- **Laboratory of Geometry LoG(M), University of Michigan:** Mentored an undergraduate team and helped design an ion-channel mesh generation package.
- **CS KickStart:** Co-organizer for the CS KickStart hardware lab (2017, 2018), a workshop aimed at improving gender diversity in computer science through increased female enrollments.
- **Lunch & Lab with a Grad:** Mentored undergraduate students and familiarized them with the graduate school application process.

Technical Skills

- **Languages:** C, C++, Python, Verilog HDL
- **Software tools:** BWA-MEM, Picard Tools, GATK, Murphi, LLVM
- **Architectural simulators:** gem5, Snipersim
- **Version Control:** git, Perforce