# Shaizeen Aga

2485 Augustine Drive,
Santa Clara, CA-95054, USA.
shaizeen.aga@amd.com
www.shaizeen.com

## Brief Biography

I am a Technical Lead and Member of Technical Staff at AMD Research where I lead research on application-driven design of accelerators and future architectures. More generally, my research interests include processor architectures, memory subsystems and security with a specific interest in near-memory accelerators.

I received my M.S. (2013) and Ph.D. (2017) from the University of Michigan, Ann Arbor where I worked on several exciting topics including in-place computing in processor caches, building secure architectures at low cost, realizing intuitive memory models efficiently and improving performance of multi-core runtimes.

My work has been published at several top-tier computer architecture venues (ISCA, MICRO, HPCA) and also at high-performance computing venues (SC). My research has won several awards at and across institutional level and I was an invited participant in Rising Stars in EECS, 2017 workshop. I am also a (co-)inventor on over twelve granted and pending US patent applications.

## Education

2013–2017 **Ph.D, Computer Science and Engineering**, *University of Michigan, Ann Arbor*.

Dissertation: Near Data Processing for Efficient and Trusted Systems

2011–2013 **MS, Computer Science and Engineering**, *University of Michigan, Ann Arbor*.

2005–2009 **BTech, Information Technology**, *College of Engineering, Pune, India*.

**Class rank 2$^{nd}$** in Information Technology Department

## Honors and Awards

June 2020 **AMD Spotlight Award**.
For outstanding work in developing analytical models and performance methodologies for accurate projections of benefits of futuristic architectural features.

Nov 2017 **CSE Nominee for Richard and Eleanor Towner Prize for Outstanding PhD Research**.
One of two students in 2017 to be nominated by the CSE Graduate Committee at University of Michigan for a college-level competition which highlights outstanding research achievements of Ph.D. students.

Aug 2017 **Rising Stars in EECS, 2017**.
An invite-only academic career workshop for nearly 70 best and brightest women scholars around the world in Electrical Engineering and Computer Science (EECS).

Dec 2016 **Best demo at Center for Future Architectures Research (CFAR) Annual Workshop**.
Awarded for **Compute Caches** which enables in-place computation in caches. This workshop showcased nearly 50 projects in computer architecture related topics from several leading institutions. **CFAR** is one of the six SRC STARnet centers sponsored by MARCO and DARPA.

Nov 2016 **1$^{st}$ place at University of Michigan CSE Graduate Students Honors Competition**.
A yearly competition which recognizes research of broad interest and exceptional quality.

2011-2013 **Recipient of K.C. Mahindra Scholarship, Bharat Petroleum Corporation Scholarship** .
Two different scholarships, awarded to select few students across India for graduate studies.

2009 **1$^{st}$ place in Parallel Computing at Imagine Cup, a worldwide student technical competition organized by Microsoft**.

## Publications

### Selected Research Publications

April 2020 **SeqPoint: Identifying Representative Iterations of Sequence-based Neural Networks**

Suchita Pati, **Shaizeen Aga**, Matt Sinclair and Nuwan Jayasena.

*In IEEE International Symposium on Performance Analysis of Systems and Software **(ISPASS)**, Boston, Massachusetts, USA.*

Sept 2019 **Co-ML: A Case for Collaborative ML Acceleration using Near-data Processing**

**Shaizeen Aga**, Nuwan Jayasena and Mike Ignatowski.

*In 5th International Symposium on Memory Systems **(MEMSYS)**, Washington, DC.*

June 2019 **InvisiPage: Oblivious Demand Paging for Enclaves**

**Shaizeen Aga** and Satish Narayanasamy.

*In 46th International Symposium on Computer Architecture **(ISCA)**, Phoenix, Arizona.*

May 2018 **MOCA: Memory Object Classification and Allocation in Heterogeneous Memory Systems**

Aditya Narayan, Tiansheng Zhang, **Shaizeen Aga**, Satish Narayanasamy and Ayse Kivilcim Coskun.

*In 32nd IEEE International Parallel and Distributed Processing Symposium **(IPDPS)**, Vancouver, British Columbia, Canada.*

June 2017 **InvisiMem: Smart Memory Defenses for Memory Bus Side Channel**

**Shaizeen Aga** and Satish Narayanasamy.

*In 44th International Symposium on Computer Architecture **(ISCA)**, Toronto, ON, Canada.*

Feb 2017 **Compute Caches**

**Shaizeen Aga**, Supreet Jeloka, Arun Subramaniyan, Satih Narayanasamy, David Blaauw, and Reetuparna Das.

*In IEEE Symposium on High Performance Computer Architecture **(HPCA)**, Austin, Texas.*

Dec 2015 **Efficiently Enforcing Strong Memory Ordering in GPUs**

Abhayendra Singh, **Shaizeen Aga** and Satish Narayanasamy.

*In International Symposium on Microarchitecture **(MICRO)**, Waikiki, Hawaii.*

Nov 2015 **CilkSpec: Optimistic Concurrency for Cilk**

**Shaizeen Aga**, Sriram Krishnamoorthy and Satish Narayanasamy.

*In International Conference for High Performance Computing, Networking, Storage and Analysis **(SC)**, Austin, TX.*

June 2015 **zFence: Data-less Coherence for Efficient Fences**

**Shaizeen Aga**, Abhayendra Singh and Satish Narayanasamy.

*In 29th International Conference on Supercomputing **(ICS)**, Newport Beach, CA.*

### Patents

2018 **Near-memory data dependent gather and packing**

**Shaizeen Aga** and Nuwan Jayasena. *Patent Pending*

2018 **Device and method for accelerating matrix multiply operations**

**Shaizeen Aga**, Nuwan Jayasena, Allen Rush and Mike Ignatowski. *Patent Pending*

2018 **Device and method for accelerating matrix multiply operations as a sum of outer products**

**Shaizeen Aga**, Nuwan Jayasena, Allen Rush and Mike Ignatowski. *Patent Pending*

2017 **Trusted computing system with enhanced memory**

Satish Narayanasamy and **Shaizeen Aga**. *Patent Pending*

2016 **Method for exploiting parallelism in task-based systems using an iteration space splitter**

Behnam Robatmili, **Shaizeen Aga**, Dario Suarez Gracia, Arun Raman, Arvind Natarajan, Gheorghe Calin Cascaval, Pablo Montesinos Ortego, Han Zhao. *US Patent 9501328*

2016 **Ordering constraint management within coherent memory systems**

**Shaizeen Aga**, Abhayendra Singh and Satish Narayanasamy. *US Patent 9367461*

## Academic and Professional Experience

Oct'19-Now **Technical Lead, Application-driven architecture design** *AMD Research*

In this role I drive new and strategic research directions for my team within AMD Research with a broad goal of application-driven design of accelerators and future architectures.

My team focuses on in-depth understanding of applications of business interest such as machine learning (CNNs, BERT, RNNs, recommendation models etc.), graphics (ray tracing), gaming and high-performance computing applications and we use this understanding to guide future accelerators and architectures research.

I am also responsible for writing proposals for external funding opportunities along with delivering key presentations and deliverables on these funding contracts.

July'19-Now **Member of Technical Staff** *AMD Research*

I developed analytical models and performance methodologies for accurate projections of benefits of futuristic architectural features. These projections fed into executive decision making at AMD and partner organizations on key projects.

I co-advised several summer co-ops on interesting architecture projects from architectural characterization of sequence-based neural networks (**ISPASS'20**), to ordering primitives for GPUs and more.

Jan'18-June'19 **Senior Software System Designer** *AMD Research*

I studied the benefits of near-data computing for machine learning workloads (**MEMSYS'19**).

Additionally, as part of AMD's PathForward Exascale research program, I studied data movement for machine learning workloads in heterogeneous systems and ways to optimize its effects. I led multiple discussions with US Department of Energy Labs as part of this effort.

Jan'12-Dec'17 **Graduate Student Research Assistant** *University of Michigan, Ann Arbor*
**Advisor:** Satish Narayanasamy

**Oblivious demand paging for enclaves:** This work (**ISCA'19**) designs a solution to mitigate page fault side channel while making OS demand paging oblivious. To do so, we customize the Oblivious RAM construct for page management context and design a novel memory partition which reduces reliance on the ORAM construct.

**Efficient secure hardware:** In this work (**ISCA'17**), I implemented a low-overhead hardware design that provides strong defenses against memory bus side-channels using compute capable 3D memories wherein we harness logic layer close to memory to implement cryptographic primitives. This design is one to two orders of magnitude lower in performance, space, energy, and memory bandwidth overhead compared to prior solutions.

**Caches as data-parallel accelerators:** This work (**HPCA'17**) transforms on-chip caches from passive to active compute units by enabling in-place computation in them. This unlocks massive data parallelism (~100X wrt SIMD processor) and saves significant amount of data movement energy (~10X wrt SIMD processor). This accelerates new-age applications, which process massive amounts of data in a data-parallel fashion considerably (1.9X performance, 2.4X energy savings).

**Efficient fences for multi-cores:** In this work (**ICS'15**), I designed and implemented an efficient fence instruction by decoupling coherence permission from data which enables stronger and more intuitive memory model like Sequential Consistency in hardware at a low cost (2.93%).

**Stronger memory models for GPUs:** Herein, we investigate memory model implications for GPUs and propose a low-overhead GPU-specific non-speculative Sequential Consistency design (**MICRO'15**).

May- Aug'14 **Interim Engineering Intern** *Qualcomm Research Silicon Valley (QRSV)*
**Manager:** *Calin Cascaval*
**Optimizing a heterogeneous benchmark:** I worked on optimizing a heterogeneous (CPU + mobile GPU) benchmark (ray tracing). Using algorithmic changes and ARM Neon instructions I attained 2-30X speedup. I also implemented an alternate algorithm which attained 40X speedup.

June-Aug'12 **Intern** *Pacific Northwest National Laboratory*
**Mentor:** *Sriram Krishnamoorthy*
**Smart runtime via optimistic concurrency:** Cilk multi-core runtime system makes it easy for programmers to express parallelism; its synchronization primitives, however, tend to be over-constrained causing poor performance. With speculation guided by a smart predictor, I improved Cilk's performance by 1.6X on 30 cores while retaining Cilk's ease of programmability (**SC'15**).

Jan-July'11 **Senior Technology Associate** *Morgan Stanley*
Jan-Dec'10 **Technology Associate** *Morgan Stanley*
**Manager:** *Vinod Alva*
**Design and development of data warehouse:** I was a part of Firm Market Risk Data Warehousing team. My work primarily involved design and development of a Data Warehouse catering to regulatory (FED/FSA) requirement and analytical needs of the risk managers, optimizing queries and building analytical (OLAP) cubes for reporting purposes. I was a key contributor to a major re-structuring project (reduced the data transformation and load cycle by 65%).
I also served as a mentor and I was involved in hiring initiatives as well.

June'08- **Intern** *NVIDIA Graphics Pvt Ltd*
Mar'09 **Manager:** *Philips Koshy*
**Accelerating true motion estimation:** I ported a True motion estimation algorithm onto the NVIDIA parallel computing platform CUDA to get 180X speedup.

## Teaching Experience

2013, 2014 **Graduate Student Instructor**, *University of Michigan, Ann Arbor*.

I taught Parallel Computer Architecture; a graduate course on recent advancements in parallel architectures.

## Other Projects

2012 **Speculatively relaxing memory model constraints by dynamic classification of cache blocks**.

Using dynamic classification of cache blocks, I relaxed memory consistency model constraints to improve performance of Sequentially Consistent hardware. This project earned **top grade in Winter 2012 class of Parallel Computer Architecture** at University of Michigan.

2011 **Design and implementation of P6 microarchitecture based core in Verilog**.

I designed and implemented the memory interface of the core and host of other components including an **Adaptive Instruction Prefetcher**. This project earned **top grade in Fall 2011 class of Computer Architecture** at University of Michigan.

2009 **Optimizing maximum likelihood method of phylogenetic tree construction algorithm**.

Using Microsoft's Task Parallel library, I improved the performance of computationally intensive Maximum Likelihood method of Phylogenetic tree construction which is used in drug design. This project won **1st place in Parallel Computing** at **Microsoft's Imagine Cup 2009**.

## Service

2020 **Program Co-chair** and **Co-organizer** for Young Architect Workshop at ASPLOS 2020.

2019-Now **Mentor** to **graduate women students** at Women in Computer Architecture Group.

2019 **Program Co-chair** and **Co-organizer** for Young Architect Workshop at HPCA 2019.

**Program Committee Member** for MICRO 2019.

2018 **External Review Committee Member** for ISCA 2019, ASPLOS 2019 and HPCA 2019.

**Reviewer** for IEEE Transactions on Computers.

**Reviewer** for IEEE Computer Architecture Letters.

2017 **Shadow Program Committee Member** for ASPLOS 2018.

2014-15 **Moderator** for the **computer architecture reading group** at University of Michigan that meets weekly to discuss recent papers from top-tier computer architecture conferences.

2015-2017 **Mentor** to **incoming graduate women students** at University of Michigan.

2007-09 **Co-ordinator** and **Member** of **Debate Club** at College of Engineering Pune, India.

2008-09 **Volunteer** with Akanksha; a NGO working for **education of under-privileged kids**.

## Presentations

2017 **Near Data Processing for Efficient and Secure Systems**

- AMD Research.
- Parallel Computing Lab (PCL), Intel Labs.
- Micro-architecture Research Lab (MRL), Intel Labs.
- Micron Technology Inc.

2017 **InvisiMem: Smart Memory Defenses for Memory Bus Side Channel**

International Symposium on Computer Architecture **(ISCA)**.

2017 **Compute Caches**

IEEE Symposium on High Performance Computer Architecture **(HPCA)**.

Center for Future Architectures Research **(CFAR)** e-Workshop.

Center for Future Architectures Research **(CFAR)** Annual Workshop.

**Rising Stars in EECS** workshop.

2015 **CilkSpec: Optimistic Concurrency for Cilk**

International conference for High Performance Computing, Networking, Storage and Analysis **(SC).**

2015 **zFence: Data-less Coherence for Efficient Fences**

International Conference on Supercomputing **(ICS)**.

## Press

2017 Shaizeen Aga - Rising Stars in EECS 2017. Link

2016 CSE Graduate Student Honors Competition Highlights Outstanding Research. Link

2016 C-FAR Student Aga Wins Graduate Honor. Link

2009 Geek Gods. Link

2009 COEP bags first spot in int'l student tech event. Link

2009 COEP students bag Microsoft Imagine Cup. Link

2009 Student team from India wins at Microsofts Imagine Cup Egypt '09. Link