Dr. Omais Shafi Pandith

Email:	omais.shafi@gmail.com
Website:	omais-shafi.github.io
Google Scholar:	Google Scholar
Phone:	+44-7436438165

About Me

CPU Architect and Computer Systems Researcher with over nine years of experience spanning processor architecture, AI systems, SoC performance analysis, RISC-V architectures, Trusted Execution Environments (TEEs), secure processor design, and heterogeneous computing. I currently work as a CPU Architect at Arm, where I focus on workload characterization, microarchitectural performance analysis, and performance optimization for next-generation Arm processors.

Previously, I held senior architecture and research positions at the Technology Innovation Institute (TII), Qualcomm, and IIT Delhi, leading research in processor architecture, hardware security, memory systems, GPU-based machine learning inference, and edge AI systems. My work has resulted in publications in premier computer architecture and systems venues, including PACT, DAC, DATE, IEEE TDSC, ACM JETC, ACM TOSN, and the Journal of Systems Architecture.

My research interests lie at the intersection of computer architecture, AI systems, hardware security, performance modeling, memory hierarchy optimization, and hardware–software co-design. I am particularly interested in developing efficient AI systems by combining advances in computer architecture, high-performance computing, and modern machine learning techniques for scalable and energy-efficient intelligent computing.

Research Highlights

- 10 peer-reviewed publications in computer architecture, hardware security, and AI systems.
- Publications in premier conferences and journals, including PACT, DAC, DATE, IEEE TDSC, ACM TOSN, ACM JETC, IISWC, and JSA.
- Reviewer for HPCA, DAC, DATE, IEEE TCAD, IEEE TDSC, IPDPS, and JSA.
- Invited speaker at DATE, RISC-V Summit USA, ACM COMPASS, PACT, IISWC, and NYU Abu Dhabi.

Education

2017–2021 **Doctor of Philosophy (Ph.D.) in Computer Science**
Department of Computer Science and Engineering
Indian Institute of Technology Delhi, India
Advisor: Prof. Rijurekha Sen — CGPA: 8.45/10

2012–2016 **Bachelor of Technology in Information Technology**
Department of Information Technology
National Institute of Technology Srinagar, India
CGPA: 8.91/10

Work Experience

1. Arm Ltd., Cambridge, United Kingdom — **Jan 2026 – Present**

CPU Architect

- Analyze representative workloads on Arm big and LITTLE CPU cores to identify microarchitectural performance bottlenecks.
- Drive performance optimization through workload characterization, pipeline analysis, and cache and memory subsystem evaluation.
- Analyze real-world application workloads to identify microarchitectural performance bottlenecks and optimization opportunities.
- Collaborate with CPU architecture teams to evaluate and optimize future microarchitectural features.

- Develop performance insights that guide architectural design decisions for next-generation Arm processors.

2. Technology Innovation Institute (TII), Abu Dhabi, UAE **May 2023 – Nov 2025**

Senior Researcher

- Led architectural research on RISC-V processors with a focus on performance, security, and Trusted Execution Environments (TEEs).
- Developed and integrated RISC-V architectural extensions and processor customizations using the Tejas cycle-accurate simulator.
- Performed workload characterization and architectural performance analysis of representative mission-critical and embedded computing applications to identify optimization opportunities.
- Collaborated with CPU architecture and security teams to evaluate the performance and security implications of emerging microarchitectural features during early design stages.
- Conducted architectural exploration and performance modeling to identify bottlenecks and guide optimization of secure processor designs.
- Authored peer-reviewed research on secure processor architectures, memory authentication, and RISC-V Trusted Execution Environments in leading computer architecture venues.

3. Qualcomm Technologies, Bangalore, India **Dec 2021 – Apr 2023**

Senior Performance Architect

- Served as the performance platform architect for a next-generation Snapdragon mobile SoC, leading the bring-up and integration of CPU, memory, multimedia, and interconnect subsystems into the pre-silicon transaction-level performance model.
- Evaluated system performance across CPU, memory, display, camera, video, modem, and NoC subsystems using representative multimedia and application workloads.
- Identified architectural bottlenecks through transaction-level performance modeling and proposed optimizations before silicon availability.
- Collaborated with cross-functional architecture teams to evaluate new SoC features, system-level trade-offs, and performance across diverse use cases.
- Optimized LPDDR4/LPDDR5 memory subsystems, ARM SMMU, QoS mechanisms, and multimedia pipelines.

4. Indian Institute of Technology Delhi **2017 – 2021**

Researcher

- Conducted research in secure processor architecture, memory hierarchy optimization, and heterogeneous computing.
- Designed and evaluated architectural proposals using cycle-accurate simulation frameworks.
- Performed performance analysis of secure processors and GPU-based machine learning inference workloads.
- Mentored junior researchers and contributed to multiple publications in top-tier architecture venues.
- Developed novel techniques for reducing performance overheads in Intel SGX and secure processor memory systems.

5. Indian Institute of Science (IISc), Bangalore **Dec 2014 – Mar 2015**

Research Intern

- Developed machine learning algorithms for LandSAT image analysis and remote sensing applications.

6. Indian Institute of Technology Delhi **Jan 2017 – Dec 2021**

Teaching Assistant

Research Interests

Computer Architecture, AI Systems, Machine Learning Systems, Hardware Security, Trusted Execution Environments, Heterogeneous Computing, Performance Modeling, Memory Hierarchy Optimization, Edge AI, Hardware–Software Co-design.

Selected Research Contributions

- IceEdge: Thermal-Aware AI Inference Serving on Edge Platforms
  Designed **IceEdge**, a runtime framework for efficient multi-tenant deep learning inference on heterogeneous CPU–GPU edge platforms. The framework jointly optimizes inference latency, throughput, and thermal behavior, enabling reliable deployment of AI applications in resource-constrained smart-city environments. Published in **ACM Transactions on Sensor Networks (TOSN)**.

- Deep Learning Inference Optimization on Embedded GPUs
  Conducted a comprehensive performance characterization of TensorRT and other DNN inference optimization frameworks on NVIDIA edge GPUs. Identified sources of non-determinism, latency variation, and accuracy differences across deployment platforms, providing practical insights for real-time and safety-critical AI systems. Published in **IISWC 2021** with an extended version in **ACM JETC**.

- Efficient Page Fault Handling for Intel SGX
  Developed **SGXFault**, a hardware-assisted page fault optimization technique for Intel SGX enclaves. Leveraged cache lockdown and predictive page prefetching to significantly reduce enclave page fault overheads for memory-intensive secure applications. Published in **IEEE Transactions on Dependable and Secure Computing (TDSC)**.

- Efficient Memory Authentication for Secure Processors
  Designed novel integrity tree optimizations that reduce both authentication latency and metadata storage overhead in secure processors. Proposed efficient cache management mechanisms for integrity counters, resulting in publications at **DAC 2021** and the **Journal of Systems Architecture (JSA)**.

- Reducing Context Switching Overheads in Intel SGX
  Proposed **SecSched**, a Cuckoo filter-based scheduling mechanism that minimizes costly TLB flushes during enclave transitions, substantially improving the performance of OS-intensive secure workloads. Published at **PACT 2020**.

Research Software and Technical Contributions

- **Tejas Cycle-Accurate Simulator**
  - Developed and integrated RISC-V architectural extensions for processor architecture research and customer-specific evaluations.
  - Designed and evaluated microarchitectural features for secure processors, memory systems, and performance optimization.
- **Performance Analysis and Modeling**
  - Performed workload characterization and microarchitectural performance analysis for Arm CPUs to identify performance bottlenecks and optimization opportunities.
  - Conducted pre-silicon performance evaluation of Qualcomm Snapdragon SoCs using transaction-level performance models across CPU, memory, multimedia, and interconnect subsystems.
  - Developed representative workloads and performance studies to evaluate CPU, memory, cache, and NoC behavior across diverse system configurations.
- **AI Systems and Edge Computing**
  - Developed runtime frameworks and evaluation methodologies for GPU-based deep learning inference on heterogeneous edge platforms.
  - Characterized TensorRT and DNN compiler optimizations for NVIDIA embedded GPUs to improve AI inference performance and reliability.
- **Secure Processor Research**
  - Designed architectural optimizations for Intel SGX and RISC-V Trusted Execution Environments, including memory authentication, integrity protection, and page fault optimization.
- **Smart Snapper (Freelance Technical Lead)**

- Led the technical development of Smart Snapper, coordinating the design and implementation of its web, Android, and iOS platforms.
- Defined system architecture, decomposed product requirements into engineering tasks, and coordinated implementation across multiple developers.
- Oversaw feature integration, technical reviews, deployment planning, and overall software quality throughout the project lifecycle.
- Served as the primary technical liaison with the client, translating functional requirements into engineering deliverables.

Technical Skills

- **Computer Architecture:** ARM, RISC-V, SoC Architecture, CPU Microarchitecture, Secure Processors, Intel SGX, Trusted Execution Environments, Memory Systems, Performance Modeling, Tejas Simulator
- **Programming:** C, C++, Python, Java
- **AI Systems and Machine Learning:** NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch
- **Development Tools:** Linux, Git, Docker

Selected Publications

Conference Publications

- **Omais Shafi**, Janibul Bashir. *SecSched: Flexible Scheduling in Secure Processors.* **International Conference on Parallel Architectures and Compilation Techniques (PACT)**, 2020.
- **Omais Shafi**, Ismi Abidi. *CuckoOnsai: An Efficient Memory Authentication Using Amalgam of Cuckoo Filters and Integrity Trees.* **Design Automation Conference (DAC)**, 2021.
- **Omais Shafi**, Chinmai Rai, Gayathri Ananthanarayanan, Rijurekha Sen. *Demystifying TensorRT: Characterizing Neural Network Inference Engines on NVIDIA Edge Devices.* **IEEE International Symposium on Workload Characterization (IISWC)**, 2021.
- **Omais Shafi**, Sachin Kumar Chauhan, Gayathri Ananthanarayanan, Rijurekha Sen. *DynCNN: Application Dynamism and Ambient Temperature-Aware Neural Network Concurrency Controller for Edge Devices.* **ACM COMPASS**, 2022.
- **Omais Shafi**, Rafail Psiakis, Johanna Toivannen. *EMAClave: An Efficient Memory Authentication for RISC-V Enclaves.* **Design, Automation and Test in Europe (DATE)**, 2024.

Journal Publications

- **Omais Shafi**, Janibul Bashir. *FreqCounter: Efficient Cacheability of Encryption and Integrity Tree Counters in Secure Processors.* **Journal of Systems Architecture (JSA)**.
- **Omais Shafi**. *SGXFault: An Efficient Page Fault Handling for SGX Enclaves.* **IEEE Transactions on Dependable and Secure Computing (TDSC)**.
- **Omais Shafi**, Mohammad Khalid, Amarjeet Saini, Gayathri Ananthanarayanan, Rijurekha Sen. *Repercussions of Using DNN Compilers on Edge GPUs for Real-Time and Safety-Critical Systems: A Quantitative Audit.* **ACM Journal on Emerging Technologies in Computing Systems (JETC)**.
- **Omais Shafi**, Mohammad Khalid Pandith, Arpan Gujarati, Astha Mehta, Rijurekha Sen. *IceEdge: Thermal-Aware Machine Learning Inference Serving for Emerging Edge Applications.* **ACM Transactions on Sensor Networks (TOSN)**.
- **Omais Shafi**, Mohammad Khalid Pandith. *EXACT: EXecution-Aware Cache and InTegrity Protection for Trusted Execution Environments.* **Under Review**.

Professional Development

- Machine Learning Specialization, Andrew Ng (DeepLearning.AI / Coursera)

- Neural Networks, 3Blue1Brown

Invited Talks and Conference Presentations

Mar 2024	*EMAClave: An Efficient Memory Authentication for RISC-V Enclaves*, presented at DATE 2024, Valencia, Spain.
Nov 2023	*EMAClave: An Efficient Memory Authentication for RISC-V Enclaves*, presented at the RISC-V Summit USA 2023, San Diego, USA.
Jun 2022	*DynCNN: Application Dynamism and Ambient Temperature-Aware Neural Network Concurrency Controller for Edge Devices*, presented at ACM COMPASS 2022, Seattle, USA.
Dec 2021	*CuckoOnsai: An Efficient Memory Authentication Using Cuckoo Filters and Integrity Trees*, presented at DAC 2021.
Nov 2021	*Demystifying TensorRT: Characterizing Neural Network Inference Engines on NVIDIA Edge Devices*, presented at IISWC 2021.
Aug 2021	*SecSched: Flexible Scheduling in Secure Processors*, invited research talk at New York University Abu Dhabi.
Oct 2020	*SecSched: Flexible Scheduling in Secure Processors*, presented at PACT 2020.
Dec 2019	*SecSched: Flexible Scheduling in Secure Processors*, presented at the IIT Delhi PhD Symposium.

Awards and Honors

2026	DAC PhD Forum Best PhD Thesis Nominee.
2026	ACM SIGDA Travel Grant to present research at the DAC PhD Forum.
2022	ACM Travel Grant to present research at ACM COMPASS, Seattle, USA.
2021	Outstanding Teaching Assistant Award, Data Structures and Algorithms, IIT Delhi.
2021	ACM-India / IARCS Travel Grant to attend the Design Automation Conference (DAC), USA.
2019	Travel Grant to attend the Cryptographic Hardware and Embedded Systems (CHES) Conference, USA.
2016	Graduated 3rd in the B.Tech class (60 students), National Institute of Technology Srinagar.
2011	Merit Scholarship for securing 97.2% in Higher Secondary School.
2009	Merit Scholarship for securing 95% in Secondary School.

Professional Service

- Reviewer for HPCA, DAC, DATE, IEEE TCAD, IPDPS, IEEE TDSC, JSA.

Teaching Experience

Teaching Assistant, Indian Institute of Technology Delhi (2017–2021)

- Five years of teaching experience supporting undergraduate and graduate courses in Computer Science.
- Conducted tutorials, laboratory sessions, programming assignments, and student mentoring across multiple core courses.
- Received the **Outstanding Teaching Assistant Award (2021)** for the Data Structures and Algorithms course.
- Courses taught:
  - Computer Architecture
  - Data Structures and Algorithms
  - Digital Logic and System Design
  - Cloud Computing
  - Introduction to Computer Science
  - Software Laboratory

References

1. Prof. Rijurekha Sen
Associate Professor
riju@cse.iitd.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

2. Dr. Gayathri Ananathanarayanan
Assistant Professor
gayathri@iitdh.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Dharwad

3. Prof. Amitabha Bagchi
Professor
bagchi@cse.iitd.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

4. Dr. Venkata Koppula
Assistant Professor
kvenkata@cse.iitd.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

5. Prof. Mohammad Ahsan Chishti
Associate Professor
ahsan@nitsri.ac.in
Department of Computer Science
National Institute of Technology Srinagar

6. Dr. Janibul Bashir
Assistant Professor
janibbashir@nitsri.ac.in
Department of Information Technology
National Institute of Technology Srinagar

7. Pragnajit Roy
SoC Power and Performance Architect
pragnajit.datta.roy@gmail.com
Company name: Aptiv Bangalore

8. Dr. Mohammad Khalid Pandit
Assistant Professor
mkhalid@nith.ac.in
Department of Computer Science
National Institute of Technology Hamirpur