

Rajesh Shashi Kumar

✉ rajesh.shashikumar@arm.com ♦ [linkedin.com/in/rajeshwisc](https://www.linkedin.com/in/rajeshwisc) ♦ [rajesh-s.github.io](https://github.com/rajesh-s)

EDUCATION

M.S. Electrical and Computer Engineering, <i>Specialization: Computer Systems & Architecture</i> University of Wisconsin-Madison, Madison, WI, USA	09/08/2021 - 05/31/2023 Grade: 3.94/4
B.E. Electronics & Communication Engineering PES Institute of Technology, Bangalore, Karnataka, India	08/01/2013 - 05/31/2017 Grade: 9.11/10

WORK EXPERIENCE

ARM <i>Senior Engineer</i>	06/20/2023 - Present Austin, Texas, USA
<ul style="list-style-type: none">Conducting architectural research to improve scalability and efficiency of datacenter systems for emerging workloads.Specialized in hardware-software co-design for distributed heterogeneous memory systems for AI/ML.Designed novel, scalable architectures for interconnects and memory coherence in next-generation datacenter systems.Developed analytical models, characterization tools, and simulation frameworks for architecture performance evaluation.Investigated workloads: NGINX, Vector Packet Processing, Linux Kernel Atomics, XGBoost, OpenMP, LLM inference	
Advanced Micro Devices Research <i>Research Co-op</i>	05/23/2022 - 08/26/2022 Austin, Texas, USA
Investigated efficient mapping of sparse linear algebra operations in graph workloads at-scale for GPU architectures.	
Qualcomm <i>Network on Chip Design Engineer</i>	09/03/2019 - 08/11/2021 Bangalore, India
<ul style="list-style-type: none">Designed multi-core hardware cache coherent interconnects. My role involved close collaboration with performance, processor, power, physical design, & memory subsystem teams.Extensive experience optimizing RTL and architecture for advanced nodes, managing system performance trade-offs, performance analyses and debugging for mobile and automotive SoCs.	
Analog Devices <i>Design Engineer</i>	07/03/2017 - 08/14/2019 Bangalore, India
<ul style="list-style-type: none">Designed hardware interconnects, offload engines, peripheral IP & memory protection as a part of the Digital Signal Processors team. Proven track record in evaluating design feasibility and closing timing at high operational frequencies.Contributed to ECO root cause analysis, silicon bring-up, and formal verification of boot flow security schemes.	
Analog Devices <i>Performance Verification Intern</i>	01/01/2017 - 06/20/2017 Bangalore, India
Performance analysis of DSP accelerators and system interconnects. Developed an in-house tool for flow-based stimulus generation and capturing SoC interconnect performance metrics.	

ACADEMIC RESEARCH EXPERIENCE

Heterogeneous Architectures Lab <i>Advisor: Prof. Matthew Sinclair</i>	Aug 2022 - May 2023 Madison, WI
Efficient multi-chiplet GPU coherence with intelligent acquires and releases to promote inter-kernel data reuse.	
Heterogeneous Architectures Lab <i>Advisors: Prof. Shivaram Venkataraman & Prof. Matthew Sinclair</i>	Jan 2022 - May 2022 Madison, WI
Characterization of spatio-temporal effects of DVFS on performance variability in multi-tenant GPU systems for ML. Link .	
National Institute of Technology <i>Advisor: Prof. Santanu Sarkar</i>	May 2015 - July 2015 Rourkela, India
Design of process, voltage, temperature independent CMOS bandgap voltage and current reference circuits.	

PUBLICATION

- **CPElide: Efficient Multi-Chiplet GPU Implicit Synchronization**
Preyesh Dalmia, Rajesh Shashi Kumar, Matthew Sinclair
57th IEEE/ACM International Symposium on Microarchitecture (MICRO 2024)
- **Contention Aware Machine Learning: Understanding CPU ML Inference Scalability Bottlenecks**
Josh Minor, Minjun Wu, Rajesh Shashi Kumar, Shashank Hegde, Eric Van Hensbergen
Internal Research Paper, Arm, Texas, USA, 2024

PATENTS

- **Cache Synchronization for Chiplet Accelerators.** Preyesh Dalmia, Rajesh Shashi Kumar, Matthew Sinclair.
Publication No.: US20240320158A1. Status: Grant notification received in November, 2024.

TEACHING

Department of Electrical & Computer Engineering, Wisconsin-Madison

Teaching Assistant

August 2022 - May 2023

Madison, WI

- ECE/CS532: Matrix methods in Machine Learning (Fall 2022)
- ECE210: Introductory Experience in Electrical & Computer Engineering (Spring 2023)
- ECE342: Circuit Analysis (Spring 2023)

Department of Computer Sciences, Wisconsin-Madison

Teaching Assistant

August 2021 - May 2022

Madison, WI

- CS200: Object-oriented Programming I (Fall 2021)
- CS300: Object-oriented Programming II (Spring 2022)

TECHNICAL SKILLS

Programming Languages & Runtimes

Python, C++, C, CUDA, HIP, System Verilog, Bash
OpenMP, MPI, PyTorch

Performance Profiling

Nsight Systems, Streamline, VTune, Perf

Architectural simulators

gem5, GPGPUsim, Interconnect Workbench

Interconnect Protocols

CHI, CHI-C2C ACE, AXI, PCIe

Digital Logic Design Tools

Xcelium, VCS, ModelSim, Spyglass, Vivado, JasperGold

Implementation Tools

(Synth:DC,Genus,Yosys)(STA:Tempus)(LEC:Conformal)

Productivity Tools

Docker, SLURM, Platform LSF, Git, ~~TeX~~

ACADEMIC PROJECTS

- Improving TLB performance in GPUs using locality prefetching. [Link](#).
- Implementation and analysis of bidirectional-exchange and ring primitives for allreduce using MPI. [Link](#).
- Deepspeech training and inference performance profiling on Nvidia P100. [Link](#).
- Fine-grained GPU cache bypassing in hardware for GPGPU workloads. [Link](#).
- Design of an operational trans-conductance amplifier using the gm/ID methodology to meet the EEG biosignal pre-processing feature extraction requirements as a part of Bachelor's thesis under Prof. Partha Ray.

AWARDS

Spot Award: Analog Devices

2018

- Awarded in Dec 2018, for root cause analysis and ECO implementation to fix critical clock infrastructure issues affecting preboot, which was much needed for the tapeout confidence of a DSP SoC.
- Awarded in May 2018, for an efficient Interconnect design in NIC400 from scratch resulting in significant performance improvement within the area and power budget with no impact to the project schedule.

IEEE IBM Watson Student Showcase: Top five winning projects worldwide

December 2015

Awarded for design of an NLP application to aid in verbal skill development using IBM Watson

Publication: *IEEE Computer Issue No.01 - Jan. (2016 vol.49)* [Link](#)