Raghav Gupta

Website | GitHub | LinkedIn
raghavgupta@berkeley.edu | 424.272.7363

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY | PhD in Computer Science
Aug 2023 - Present
Advisor: Prof. Borivoje Nikolic

Coursework:
Architectures and Systems for Warehouse-Scale Computers (A)

UNIVERSITY OF CALIFORNIA, BERKELEY | BS in Electrical Engineering and Computer Science
Aug 2019 - May 2023 | GPA: 3.95/4.0
High Honors | Dean's List

Coursework:
Computer Architecture and Engineering (A) • Operating Systems and System Programming (A) • Introduction to Digital Design and Integrated Circuits (A+) • 22nm SoC for IoT (A) • Advanced Topics in Computer Systems (A-) • Programming Languages and Compilers (A)

RESEARCH INTERESTS: COMPUTER ARCHITECTURE

Hyperscale Architecture and Systems • Hardware Design Methodology

RESEARCH EXPERIENCE

SLICE LAB AT UC BERKELEY | Student Researcher
Jan 2021 – Present | Berkeley, CA | Website

Hybrid Simulation/Emulation | Collaboration with AMD/Xilinx | Sep 2024 - Present
- Building a hybrid simulation/emulation platform atop FireSim to enable design verification and early DSE
- Initial prototype allows system bus modules to be simulated in C++ while the rest of the SoC is emulated on an FPGA
- Initial results show O(10 MHz) simulation frequency when running Linux+Coremark and snooping on SoC memory requests in software

Multi-level Simulation | Oct 2023 - Apr 2024
- Helped build a multi-level simulator trading off speed and fidelity in a hierarchy of simulators for DSE on large workloads
- Used program analysis to identify critical portions of an execution trace to execute at high fidelity
- Investigated warm-up strategies to approximate cache state

Extending FirePerf to Userspace | Original Paper | Poster | Userspace Paper
- FirePerf is an FPGA-accelerated full-system hardware/software performance profiling utility built atop FireSim
- Provides high-fidelity out-of-band introspection with call stack reconstruction and Flame Graph generation
- Extended FirePerf to support userspace profiling by instrumenting Rocket Chip trace port, obtaining DWARF + hex information, constructing a user binary search space and matching traces to it

FireMarshal Networking | Original Paper
- FireMarshal is an open-source software workload management tool that automates bootable workload image generation and evaluation on FireSim, QEMU, and Spike platforms
- Implemented running multi-node workloads concurrently in QEMU
- Implemented a network interface using Virtual Distributed Ethernet for multi-node workloads emulated in QEMU

WORK EXPERIENCE

AMD/XILINX | Research and Advanced Development Intern
Jun 2024 – Dec 2024 | San Jose, CA

Hybrid Simulation/Emulation
- Implemented and evaluated two strategies to build a fast and flexible hybrid simulation/emulation platform
- Ported an internal co-simulation framework and an open-source emulation framework to an internal emulation machine

OMNISTRATE | Software Engineering Intern
May 2023 – July 2023 | Remote

Monitoring for Stateful Containerized Cloud Applications
- Developed monitoring infrastructure for stateful containers using the sidecar pattern
- Implemented support for application-specific probes and process metadata checks to trigger alerts and fail-over
- Improved reliability at scale for a public database provider's DBaaS offering

NVIDIA | Power Architect Intern
May 2022 – Aug 2022 | Santa Clara, CA

CPU Power Modeling for Gaming Workloads
- Developed a strategy for application-specific CPU power modeling on production silicon
- Targeted characteristics of gaming workloads and isolated static and dynamic power costs
- Automated data collection, processing, visualization, modeling and summary statistics

TEACHING EXPERIENCE

CS 152/252A - COMPUTER ARCHITECTURE AND ENGINEERING | Teaching Assistant
Spring 2024 | Website

- Led overall organization of a course catered towards 130+ engineering graduates and upper-division undergraduates
- Meticulously designed and debugged problems for exams and homework assignments
- Helped students learn about pipelining, OoO core design, the memory hierarchy, virtual memory, branch prediction, etc., in weekly office hours and by answering on the course Q/A forum

EE290C - 22NM SOC FOR IOT | Teaching Assistant
Fall 2022 - Spring 2023

- Developed teaching resources for RISC-V SoC tapeout, on topics such as the use of RTL generators (especially custom accelerators) and verification in simulation, that were shared as part of Intel's University Shuttle Program
- Managed and supported 3 teams across 2 chips as they designed specialized modules for sparse-dense matrix multiplication, near-cache compute, and Extended Kalman Filter

EECS 151/251A - INTRODUCTION TO DIGITAL DESIGN AND INTEGRATED CIRCUITS | Teaching Assistant
Fall 2022 | Website

- Taught 1 weekly office hour and 1 weekly lab section using Xilinx PYNQ FPGAs for Berkeley's upper division VLSI course catered towards undergraduates and graduates
- Helped students learn hardware design methods and principles using Verilog, VCS, and waveforms as they implemented FSMs, audio synthesis circuits, UART modules and FIFOs on FPGAs
- Supported students in applying architectural and microarchitectural concepts to design a 3+ stage pipelined RISC-V datapath with L1 data cache, UART/MMIO and optimize it for timing
- Designed a memory controller lab with specification and testbenches as an introduction to working with synchronous memories

EECS16A - DESIGNING INFORMATION DEVICES AND SYSTEMS I | Lab Teaching Assistant and Head Lab Content/Development
Summer 2020 – Spring 2022 | Website

- Taught a weekly lab section of 50+ engineering undergraduates for Berkeley's introductory EECS course for 5 semesters
- Assisted students in building a single pixel camera, resistive and capacitive touchscreens, and an acoustic positioning system
- Trained lab staff and managed lab content for 1000+ students, revamped existing labs, developed new labs and managed the shift to online labs while achieving the requisite learning goals.

CS61C - GREAT IDEAS IN COMPUTER ARCHITECTURE | Teaching Assistant
Summer 2021 | Website

- Taught 2 discussion sections and 1 office hour weekly for Berkeley's introductory computer architecture and systems course
- Developed content and infrastructure for projects on C programming and Logisim datapath design
- Helped design labs on Logisim datapath design, pipelining, and caches

PROJECTS

BEARLY ML: SOC FOR MACHINE LEARNING | EE290C

Jan 2022 – May 2022 | Poster

- Designed, verified, and taped-out a RISC-V SoC with ML/DSP accelerators in Intel 16 technology using the Chipyard environment
- Member of 5-person team that developed a sparse-dense matrix multiplication accelerator using the Rocket Custom Coprocessor (RoCC) interface
- Tightly-coupled with a 5 stage, in-order Rocket core with 16KB L1 DCache, 4KB L1 ICache
- Wrote Chisel RTL, performed verification in functional simulation with unit-testing and integration testing with bare-metal C tests, and resolved floor-planning Design Rule Violations
- Maximum frequency: 500 MHz, Achievable Throughput: 2.6 uint8 GOPS

PINTOS | Projects for CS162

Jan 2022 – May 2022

- As a 3-person team, implemented the following features in an x86-based educational OS running as a VM in QEMU/Bochs with extensive debugging in GDB
- Supported user program execution with syscall argument passing, process control syscalls (halt, exec, wait, exit), basic file operation syscalls, and a floating point unit
- Implemented multithreading with an efficient sleep timer, strict priority scheduler with priority donation, and a simplified version of the user-level pthread library and user-level synchronization
- Implemented a Unix FFS-like filesystem with buffer cache, extensible files and support for subdirectories and relative paths

PUBLICATIONS AND PRESENTATIONS

- Whangbo, Joonho, Edwin Lim, Chengyi Lux Zhang, Kevin Anderson, Abraham Gonzalez, **Raghav Gupta,** Nivedha Krishnakumar, et al. **"FireAxe: Partitioned FPGA-Accelerated Simulation of Large-Scale RTL Designs,"** in Proceedings of the 2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA), 2024.
- Chi, Yufeng, Franklin Huang, **Raghav Gupta,** Ella Schwarz, Jennifer Zhou, Reza Sajadiany, Animesh Agrawal, et al. **"A Heterogeneous RISC-V SoC for ML Applications in Intel 16 Technology."** Poster presented at HotChips 2023, Stanford, CA, August 27-28, 2023.

AWARDS

- EECS Oustanding TA Award (2020 - 2021)
- CS61C Project - RISCV CPU in Logisim - Apple Design Award (Fall 2020)
- EECS151 Project - RISCV Core on FPGA - Apple Design Award (Fall 2021)
- Edward Frank Kraft Award (Fall 2019)
- Times of India - Spark Scholarship (2018)

ORGANIZATIONS

ETA KAPPA NU | EECS HONOR SOCIETY

Jan 2021 - May 2023

Tutoring Officer

Responsible for managing activities of the tutoring committee, such as organizing and running exam review sessions, planning daily office hours, and supporting learning resources, in Spring 2022.

DeCal Assistant Officer

Responsible for facilitating "Going Down the EECS Stack", a course that explores the different fields within the EECS major, in Fall 2021.

CURIOSITY | Tutor and Co-founder

May 2020 - Aug 2020

- Tutored high school students on STEM topics and SAT prep during Summer 2020
- Raised ₹15,000 and donated all funds to COVID-19 and flood relief

SKILLS

Python • C/C++ • Git • Linux/Bash • Verilog • Chisel • Go • RISC-V • CUDA • Java • Vivado • Docker • Kubernetes • ROS