Moein Khazraee

Senior Architect, NVIDIA

moein@csail.mit.edu | moeinkhazraee.com | mkhazraee | mkhazraee

Education

Ph.D. Computer Science and Engineering UC San Diego *2013 - 2020*

- Dissertation: "Reducing the development cost of customized hardware acceleration for cloud infrastructure"
- Advisor: Aaron Schulman, from June 2017
- Advisor: Michael Taylor, before his move to the University of Washington.

Micro MBA Rady School of Management, UC San Diego *2017 - 2018*

- Technology Management and Entrepreneurism Certificate Program

M.Sc. Computer Science UC San Diego *2013 - 2016*

- Specialization: Computer Architecture, Advisor: Michael Taylor

B.Sc. Electrical Engineering Sharif University of Technology *2008 - 2013*

- Specialization: Digital Systems

Research Experience

Senior Architect *Aug. 2022 - Present*

NVIDIA

- As an applied scientist I work on the intersection of networking and high performance computing and machine learning workloads.

Postdoctoral Scholar, Mentor: Manya Ghobadi *Jun. 2020 - Present*

Massachusetts Institute of Technology

- Hardware framework for Silicon Photonics (SiP) - Ongoing work: Developing a simulation framework to model SiP links between hardware accelerators. SiP is a rising physical layer technology for networks that offers Tbps bandwidth. I believe this is a key step to better understand and deploy SiP within our systems , and also the groundwork for optimization across the network stack.
- TopoOpt, Network topology customization for Machine Learning (ML): Developed a cross-layer optimizer for parallelization strategy and network topology that significantly reduces the training time at the same cost. Large scale ML clusters further benefit from this approach, and we show an implementation with currently-available commodity hardware. *[NSDI 2023]*

Research Assistant, Advisor: Aaron Schulman *Jul. 2017 - Jun. 2020*

UC San Diego

- Rosebud, Framework for FPGA-based network middle-boxes: Developed the first 200 Gbps framework for FPGA smartNICs that significantly reduces the development effort by using packet abstraction to separate hardware and software developments , while at the same time provides parallelization abstractions to increase throughput and hit line rate performance. *[ASPLOS 2023]*
- SparSDR, Framework for sparsity proportional Software Defined Radios: Developed a framework that significantly reduces the required bandwidth and processing power, by offloading channelization and activity detection to hardware. User can select the desired bands, and process the data only when there is an active transmission. This framework unlocks the commodity modules to become wide-band, and enables real-time processing on a low-end processor without losing universality of SDRs. *[Mobisys 2019]*
- SweepSense, Low-cost full spectrum monitoring: Developed a customized hardware accelerator that enables sweeping the frequencies, or more technically chirp a receiver's local oscillator. Alongside a modification to the circuit of a commodity radio module, this enabled sweeping the full 5GHz. Based on a novel algorithm our system detects the communication type, at 500 times lower cost . *[NSDI 2019]*

Software Engineer Intern, Mentor: Partha Ranganathan *Apr. 2017 - Jul. 2017*

Google, Sunnyvale

- Initial steps for the Google VCU chip: Collaborated with a group of engineers to do due diligence for use of ASIC chips for YouTube transcoding workloads . Also I was involved in evaluating the impact of a category of video encoding filters, and analyzing whether it should be implemented on the chip.

Research Assistant, Advisor: Michael Taylor *Sep. 2013 - Apr. 2017*

UC San Diego

- Moonwalk, the first Model in academia for Non-Recurring Engineering (NRE) costs of developing ASICs: Developed a model and showed that technology node selection is a major tool for managing ASIC Cloud NRE, which is the main challenge for ASIC Clouds. This allows the designer to trade off an accelerator's excess energy efficiency and performance for lower total cost . *[ASPLOS 2017]*
- ASIC Clouds, Framework for building a TCO-optimal data center from ASIC chips: Specialized a fully customized hardware accelerator chip, customized circuit board, custom-designed cooling and power delivery systems, with tailored external memory and network. The goal is to optimize the Total Cost of Ownership, which determines the balance between performance and energy efficiency . *[ISCA 2016]*
- Promises and Perils in 3D Architecture Elaborated the opportunities for more efficient accelerators using 3D IC technology, in terms of performance and power. *[Masters comprehensive exam]*
- Vanilla bean, an Application-specific Instruction set Processor (ASIP): Designed and evaluated performance and power trade-offs for an ASIP, which was later used for the undergraduate computer architecture course .

Grants & Fellowships

2021	**Collaborator**, NSF Collaborative Research (SHF Medium) Grant for **$1,200,000**. "Spatial Multi-Tenant Neural Acceleration for Next Generation Datacenters".
2017-18	**Recipient**, Institute of Global Entrepreneurship fellowship, UC San Diego
2015-16	**Recipient**, Qualcomm Fellow-Mentor-Advisor (FMA) Fellowship
2013	**Recipient**, UC San Diego Computer Science and Engineering departmental fellowship

Service

CSE Diversity, Equity, and Inclusion Committee UC San Diego *Jan. 2017 - Jun. 2020*

- I was part of the initial effort to make a DEI committee specific to Computer Science and Engineering, with the goal of better understanding the challenges and to foster diversity within our department.
- We divided the DEI committee into 3 parts: Culture, Image and Recruitment. I was part of the Culture committee, trying to improve the culture of diversity within the department.
- I was also a member of the organizing committee for the **2019 Celebration of Diversity** day-long event.

Cultural Iranian Student Association UC San Diego *Feb. 2014 - Sep. 2016*

- The goal of this student organization is to help incoming Iranian students in settling down and getting to know more people with similar backgrounds, and to feel included in the university community.
- I was the organization **president** in 2016, where we organized a student orientation session, and several celebrations to introduce our culture to non-Iranians, as well as to connect the student body to resources outside the university.

Head and Executive Director of the Promotion Committee Sharif University of Technology *2010 - 2012*

- Sharif Cup 2012, the first Sharif Open Robotics Competition, Sep. 26-28, 2012
- Smart Grid Conference, Oct. 19-21, 2010

Review Activities *2014 - Present*

- Journal/Conference Reviewer: IEEE TPDS, IEEE/ACM ToN, Computing Surveys
- Member of SIGCOMM artifact evaluation committee for ACM CoNEXT 2019

Publications

M. Khazraee, A. Forencich, G. Papen, A. Snoeren, A. Schulman *ACM ASPLOS, 2023*
"Rosebud: Making FPGA-accelerated Middlebox Development More Pleasant" 25% acceptance

W. Wang, M. Khazraee, Z. Zhong., Z. Jia, D. Mudigere, Y. Zhang, A. Kewitsch, M. Ghobadi *USENIX NSDI, 2023*
"TOPOOPT: Optimizing the Network Topology for Distributed DNN Training" 18% acceptance

Y. Guddeti, R. Subbaraman, M. Khazraee, A. Schulman, D. Bharadia *ACM GetMobile, 2020*
"Towards Low-Cost, Ubiquitous High-Time Resolution Sensing for Terrestrial Spectrum"

M. B. Taylor, L. Vega, M. Khazraee, I. Magaki, S. Davidson, D. Richmond *Communications of the ACM, 2020*
"ASIC Clouds: Specializing the Datacenter for Planet-Scale Applications"

M. Khazraee, Y. Guddeti, S.Crow, A. Snoeren, K. Levchenko, D. Bharadia, A. Schulman *ACM Mobisys, 2019*
"SparSDR: Sparsity-proportional Wideband SDRs" 23% acceptance

Y. Guddeti, R. Subbaraman, M. Khazraee, A. Schulman, D. Bharadia *USENIX NSDI, 2019*
"SweepSense: Sensing 5 GHz in 5 Milliseconds with Low-cost SDRs" 11% acceptance

S. Xie, S. Davidson, I. Magaki, M. Khazraee, L. Vega, L. Zhang, M. B. Taylor *ACM SIGOPS OSR, 2018*
"Extreme Datacenter Specialization for Planet-Scale Computing: ASIC Clouds"

M. Khazraee, L. V Gutierrez, I. Magaki, M. B. Taylor — *IEEE Micro top picks, 2017*
"Specializing a planet's computation: ASIC Clouds"

M. Khazraee, L. Zhang, L. V Gutierrez, M. B. Taylor — *ACM ASPLOS, 2017*
"Moonwalk: NRE optimization in ASIC Clouds, or, accelerators will use old silicon" — 18% acceptance

I. Magaki*, M. Khazraee*, L. V Gutierrez, M. B. Taylor (* equal contribution) — *ACM/IEEE ISCA, 2016*
"ASIC clouds: specializing the datacenter" — 19% acceptance

M. Khazraee (Masters comprehensive exam) — *UCSD, 2016*
"Promises and Perils in 3D Architecture"

M. Khazraee, A. R. Zamani, M. Hallajian, S. P. Ehsani, H. A. Moghaddam, A. Parsafar, M. Shabany — *IEEE ISCAS, 2013*
"A novel hardware implementation for joint heart rate, respiration rate, and gait analysis applied to Body Area Network."

M. R. Homaeinezhad, M. Khazraee, M. Khazraee. — *IJIE, 2012*
"An Open-Source High Speed C++/MEX Framework for the Detection and Delineation of Long Duration Ambulatory Holter ECG Events: HSEDF."

Patents

- US20200326362A1, High resolution spectrum monitoring
  Y. R. Guddeti, D. Bharadia, M. Khazraee, A. Shalev, R. V. Subbaraman

Teaching Experience

Guest Lecturer UC San Diego — *Apr. 2019*
- Taught the session about reliable transmission for CSE 123, computer networks. In this lecture I presented the hardware architecture for error detection to give students some insight on the underlying hardware.

Graduate Teaching Assistant UC San Diego — *Fall 2014*
- Teaching Assistant for Principals of Computer Architecture graduate course with 140 students. One of the two TAs in charge of homework, office hours, and the final course project.

Co-Lead Sharif University of Technology — *Winter 2013*
- Took the lead role in major overhaul of the Microprocessor Lab, and a member of the design group for the new course, "Embedded Systems' Laboratory".

Teacher Allame Helli 5 (NODET) Junior High School — *2012 - 2013*
- Taught the "Computer Algorithms" summer course, and the "Web Programming" course for the academic year. (NODET: National Organization for Development of Exceptional Talents)

Instructor Sharif University of Technology — *2011 - 2013*
- Instructor for hardware programming session, part of the "Logic Circuits and Digital Systems" course required for all undergraduate students (∼200 per academic year). I was the instructor for every faculty member offering the course **for three consecutive semesters, to continuously update and enhance the syllabus.**

Teaching Assistant Sharif University of Technology — *2011-2013*
- "Microprocessor Systems' Design"
- "Advanced Microprocessor Systems' Design"
- "Probability and Statistics for Engineering"
- "Object Oriented Programming"

Co-Lead Sharif University of Technology — *Winter 2011*
- Revised experiments and instructions for "Structure of Computer and Microprocessors Laboratory".

Laboratory Teaching Assistant Sharif University of Technology — *2010-2013*
- "Structure of Computer and Microprocessors" (4 semesters)
- "Logic Circuits and Digital Systems" (2 semesters)
- "FPGA and ASIC Systems' Design"
- "Microprocessors Laboratory"
- "Principles of Electronics"