AmirHossein **Sojoodi**

PhD Candidate · HPC Researcher · GPU Software Engineer

Kingston, Ontario, Canada

Available upon request | amir.sojoodi@gmail.com | amirsojoodi.github.io | amirsojoodi | amirsojoodi

Summary

As a researcher, interested in parallel algorithm design, middleware performance optimization, and application acceleration. Focused on improving GPU-aware communication in OpenMPI and UCX/UCC libraries. As an instructor, with 25+ courses TA'ed and taught, skilled in course design, development, and delivery, with a strong commitment to student success. As a software engineer, experienced in parallel-processing platforms/tools and programming frameworks, including CUDA, OpenMP, MPI, Pthreads, etc.

Education

Queen's University — *Kingston, Canada*

Ph.D. in Electrical and Computer Engineering — *Jan. 2020 - Jan. 2025*

- **Research Area**: High-performance GPU-aware communication in hybrid clusters, supervised by Dr. A. Afsahi

Shiraz University — *Shiraz, Iran*

M.S. in Software Engineering — *Sep. 2012 - Aug. 2015*

Shiraz University — *Shiraz, Iran*

B.S. in Software Engineering — *Sep. 2007 - Aug. 2012*

Professional Experience

CUDA Developer, part-time — *Kingston, Canada*

RE Grant Consulting Company — *Oct. 2024 - Apr. 2025*

- Extend the Rockport Networks Inc. (CERIO) CUDA microbenchmark suite to provide more insights into the performance of their systems.
- Maintain and update this C++ CMake project to adhere with new use cases and requirements, such as Windows development environments.
- Utilize NVIDIA management library (NVML) to monitor the GPU performance and power utilization.

GPU Software Engineer, part-time — *Kingston, Canada*

Distributive Co. — *Sep. 2022 - Sep. 2024*

- Researched and developed CUDA, WebGPU, and WebAssembly solutions for the Distributed Compute Protocol.
- Achieved a +700x speedup in a WebGPU-based combinatorial math problem (LeGendre Pairs) compared to the CPU implementation.
- Designed and developed a WebGPU microbenchmark for Distributed Compute Protocol (DCP) systems.
- Prepared and delivered technical presentations and reports for demos to potential investors.

CUDA Developer, part-time — *Kingston, Canada*

RE Grant Consulting Company — *Oct. 2023 - Mar. 2024*

- Designed and implemented a CUDA microbenchmark suite to analyze the performance of Rockport Networks Inc. (CERIO) systems.
- Developed a comprehensive C++ CMake project incorporating stress and speed tests for various data transfer types, sizes, and scenarios.
- Consistently met project deadlines and delivered high-quality code under tight schedules.

Research Intern, part-time — *Seattle, US (Remote)*

Microsoft Research (MSR) — *Jun. 2022 - Aug. 2022*

- Performed a comparative study on the methods to utilize GPU in Numpy-based python frameworks.
- Profiled and analyzed model-parallelism Deep Learning framework named Distributed Deep Learning (DistDL).
- Accelerated DistDL utilizing CuPy and Numba libraries to get over 2x performance improvement.

Course Design and Development Specialist, part-time — *Kingston, Canada*

Engineering Teaching and Learning Team, Queen's University — *Sep. 2020 - Apr. 2022*

- Produced and edited videos, and created animations to enhance virtual lectures for the Faculty of Engineering and Applied Science courses.
- Automated the course evaluation process using Python and Queen's survey platform, streamlining the compilation of evaluations.
- Updated and ported the Courses' Learning Objectives (CLOs) to meet the latest requirements.

Research Assistant and System Administrator, part-time — *Shiraz, Iran*

HPC Group, CSE Department, Shiraz University — *Sep. 2018 - Nov. 2019*

- Designed and implemented a GPU-aware big-data processing framework based on Apache Ignite.
- Demonstrated a 200x speedup in Genetic Algorithm (GA) optimization using the GPU-based Apache Ignite.
- Mentored and supported undergraduate/graduate members of the group.
- Setup, organized, and maintained group's GPU cluster, with services like Apache Hadoop, Spark, Tez, and Ignite.

System Administrator and Java Backend Developer, full-time — *Shiraz, Iran*

Aria Hamrah Samaneh — *Apr. 2016 - Feb. 2017*

- Developed and maintained interactive data visualization dashboards using Tableau, enhancing decision-making processes.
- Configured and managed the company's Tableau server for Shiraz City Hall offices, ensuring high availability and performance.
- Implemented robust Java-based backend services to support various company projects.

Developer and Research Assistant, full-time — *Shiraz, Iran*

Information and Communication Technology Center (ICTC), Shiraz University — *Oct. 2015 - Apr. 2016*

- Studied and developed Business Intelligence (BI) solutions for Shiraz University IT center.
- Developed and localized BI solutions based on Microsoft Power BI.

Graduate Research Fellow and System Administrator, part-time — *Shiraz, Iran*

HPC Group, CSE Department, Shiraz University — *Sep. 2012 - Aug. 2015*

- Designed and implemented a GPU-enabled framework based on Apache Tez (MapReduce-based task graph), with a speedup of 4x.
- Performed a comparative study on GPU processing frameworks and libraries in Java programming language.
- Managed and administered the HPC group's GPU servers and XenServer virtual machines.
- Setup and maintained various services such as Nexus repository manager, Apt Cacher, Squid, etc.

Publications

[1] Amirhossein Sojoodi, Ali Farazdaghi, Hamed Sharifian, Ryan E Grant, Ahmad Afsahi, "Collaborative Bandwidth-Efficient Intra-Node Allreduce", *Proceedings of the International Workshop on Accelerators and Hybrid Emerging Systems (AsHES)*, pp. 1–4, DOI: 10.1109/IPDPSW66978.2025.00016 — 2025

[2] Hamed Sharifian, Amirhossein Sojoodi, Ahmad Afsahi, "A Topology- and Load-Aware Design for Neighborhood Allgather", *Proceedings of the IEEE International Conference on Cluster Computing (CLUSTER)*, pp. 1–12, DOI: 10.1109/CLUSTER59578.2024.00019 — 2024

[3] Amirhossein Sojoodi, Yiltan Hassan Temucin, Ahmad Afsahi, "Enhancing Intra-Node GPU-to-GPU Performance in MPI + UCX through Multi-Path Communication", *Proceedings of the International Workshop on Extreme Heterogeneity Solutions (ExHET)*, pp. 1–6, DOI: 10.1145/3642961.3643800 — 2024

[4] Yıltan Hassan Temucin, Whit Schonbein, Scott Levy, Amirhossein Sojoodi, Ryan E Grant, Ahmad Afsahi, "Design and Implementation of MPI-Native GPU-Initiated MPI Partitioned Communication", *Proceedings of the workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis (SC-W)*, pp. 1–12, DOI: 10.1109/SCW63240.2024.00065 — 2024

[5] Pedram Alizadeh, Amirhossein Sojoodi, Yiltan Hassan Temucin, Ahmad Afsahi, "Efficient Process Arrival Pattern Aware Collective Communication for Deep Learning", *Proceedings of the European MPI Users' Group Meeting (EuroMPI)*, pp. 68–78, DOI: 10.1145/3555819.3555857 — 2022

[6] Philipp A. Witte, Russell J. Hewett, Kumar Saurabh, AmirHossein Sojoodi, Ranveer Chandra, "SciAI4Industry – Solving PDEs for industry-scale problems with deep learning", *arXiv* (2022), pp. 1–11, DOI: 10.48550/arXiv.2211.12709 — 2022

[7] Yiltan Hassan Temucin, Amirhossein Sojoodi, Pedram Alizadeh, Ahmad Afsahi, "Efficient Multi-Path NVLink / PCIe-Aware UCX based Collective Communication for Deep Learning", *Proceedings of the IEEE Symposium on High-Performance Interconnects (HOTI)*, pp. 1–10, DOI: 10.1109/HOTI52880.2021.00018 — 2021

[8] Yiltan Hassan Temucin, Amirhossein Sojoodi, Pedram Alizadeh, Benjamin W Kitor, Ahmad Afsahi, "Accelerating Deep Learning using Interconnect-Aware UCX Communication for MPI Collectives", *IEEE Micro* (2021), pp. 1–9, DOI: 10.1109/MM.2022.3148670 — 2021

[9] Majid Salimi Beni, Amirhossein Sojoodi, Farshad Khunjush, "A GPU-Enabled Extension for Apache Ignite to Facilitate Running Genetic Algorithms", *Proceedings of the International Symposium on Computer Architecture and Digital Systems (CADS)*, pp. 1–8, DOI: 10.1109/CADS50570.2020.9211857 — 2020

[10] Amirhossein Sojoodi, Majid Salimi Beni, Farshad Khunjush, "Ignite-GPU: a GPU-enabled in-memory computing architecture on clusters", *Journal of Supercomputing* (2020), pp. 1–28, DOI: 10.1007/s11227-020-03390-z — 2020

Honors & Awards

2024	**Best Paper Award**, International Workshop on Extreme Heterogeneity Solutions (ExHET)	*Edinburgh, UK*
2020	**Parya Scholarship**, Parya Trillium Foundation	*Canada*
2019	**Best T.A.**, CSE Department students' poll	*Shiraz University*
2016	**9th Place**, National IoT Hackathon	*IUST, Tehran*
2015	**Silver Medal (with B. Ahmadi and M. R. Katebzadeh)**, 7th National JavaChallenge	*Sharif University*
2012	**Gold Medal in Team Section**, Chess Games, South of Iran Universities	*Shiraz University*
2010	**5th Place (with M. Asadi)**, Students Competitions (Ms Pacman Intelligent Controller)	*IEEE CIG (Online)*
2010	**5th Place (with S. Kazemi and M. Saeedi)**, Kashan 2nd International Programming contest (ACM)	*Kashan University*

Selected Teaching Experience

Course Instructor

2016 **Introduction to Object Oriented Programming with Java**, Instructor — *Shiraz University*
2013 **Software Engineering Lab**, Instructor — *Shiraz University*
2012 **Software Engineering Lab**, Instructor — *Shiraz University*
2012 **Software Engineering Lab**, Instructor — *Shiraz University*

Teaching Assistant

2024 **ELEC 278 - Fundamentals of Information Structure**, TA of Dr. Tom Dean — *Queen's University*
2024 **ELEC 374 - Digital Systems Engineering**, TA of Dr. Ahmad Afsahi — *Queen's University*
2023 **ELEC 379 - Introduction to Algorithms**, TA of Dr. Naraig Manjikian — *Queen's University*
2018 **GPU Programming**, TA of Dr. Farshad Khunjush — *Shiraz University*
2014 **Multicore Programming**, TA of Dr. Farshad Khunjush — *Shiraz University*
2012 **Operating Systems**, TA of Dr. Mohammadreza Moosavi — *Shiraz University*

Related Volunteer Experience

Author and Developer — *amirsojoodi.github.io/posts*

Technical Blog/Reports — *2015 - Present*

- Author technical posts/reports about my research, projects, and experiences to share my knowledge with the community.
- Published over 70 posts so far, covering various topics in computer science and software engineering.

Member — *Queen's University*

MPI Forum — *2022 - Present*

- Participated in the MPI Forum meetings to discuss the updates of the MPI standard.
- Contributed to the discussions in bi-weekly MPI hybrid working group meetings.

Open Source Contributor — *Distributive Co.*

Google — *2023 - 2024*

- Contributed to Google Dawn, an open-source and cross-platform implementation of the WebGPU standard, used in Google Chrome.
- My contributions included fixing building issues on Linux using CMake.

PhD Representative — *Queen's University*

Graduate ECE Student Council — *2023 - 2024*

- Liaised graduate students concerns to the Electrical and Computer Engineering department during departmental monthly meetings.

Chair and Executive Member — *Shiraz University*

Students' Scientific Group, CSE Department — *2011 - 2014*

- Held more than 10 technical seminars and workshops for students.
- Scheduled courses and their exams for each semester, in coordination with the Computer Science and Engineering department.

Network Specialist — *Shiraz University*

CSE Network Administration Group — *2011 - 2012*

- Supported and maintained the CSE department's network infrastructure, which included more than 60 workstations and 4 servers.
- Fixed and troubleshooted network and workstation software/hardware issues for students and faculty.

Selected Presentations

Workshop Presenter — *Kingston, Canada*

Distributive Book Club Meetings — *Aug. 2023*

- Presented **"Application Optimization Techniques"** for the LeGendre Pairs algorithm on GPUs.
- Discussed the challenges and optimizations, including shared-memory utilizaion, coalesced memory access, warp divergence, etc.
- Presented to Distributive Co. employees for 1 hour. (Slides I & Slides II)

Workshop Presenter — *Kingston, Canada*

Distributive Book Club Meetings — *Oct. 2023*

- Presented **"Introduction to GPUs"**, their architecture, their memory hierarchy, and their memory access patterns.
- Covered the CUDA programming model, code profiling and performance tuning of the GPU applications. (Slides I & Slides II)

Skills

Programming C, C++, Python, JavaScript, Rust, TXL, Matlab, Java, Assembly, and Shell
Platforms/APIs/Libs CUDA, OpenMP, MPI, UCX, WebGPU, WGSL, Pthreads, MapReduce, PyTorch, NumPy/CuPy, Apache Ignite
Misc. Tools Git, Perf, Docker, Valgrind, NVIDIA Nsight Tools, Arm DDT, Auto Tools, CMake, Nexus, LaTeX
Languages Farsi (maternal), English (fluent), French (basic)