Behrooz Zarebavani

bzareb01@gmail.com | (+1) 437-986-5787 | behrooz-zare | BehroozZare | Toronto, Ontario

Objective: A graduate student searching for impactful problems to solve in order to accelerate the computational processes of widely used algorithms in the fields of graphics and scientific simulations.

Education

Ph.D. Computer Science | University of Toronto Sep 2020 - Ongoing
Focus: High-Performance Computing (HPC) | Supervisor: Maryam Mehri Dehnavi GPA: 3.75

M.Sc. Electrical Engineering | Sharif University of Technology Sep 2017 - Aug 2019
Focus: High-Performance Computing (HPC) | Supervisors: Matin Hashemi and Saber Salehkaleybar GPA: 4.00

B.Sc. Electrical Engineering | AmirKabir University of Technology Sep 2013 - Aug 2017
Focus: Digital Systems | Supervisors: S. Ahmad Motamedi GPA: 3.72

Research Interests

- **Graphics and Scientific Simulations**: I appreciate viewing people's creativity, and thus, I am motivated to contribute to research on graphics and scientific simulations as a tool that helps this objective.
- **Machine Learning**: My interest lies in accelerating the computational pipeline of machine learning. Additionally, I aim to utilize it to accelerate the computation of scientific simulations and graphics.
- **HPC**: I like to quickly see the results of interesting works. As a result, I apply my knowledge of hardware to develop high-performance software that efficiently utilizes the underlying hardware. This is done either through the efficient implementation of existing algorithms or by designing new ones to stretch the limits of hardware capabilities.

Publication

1. Behrooz Zarebavani, Danny M Kaufman, David IW Levin, and Maryam Mehri Dehnavi. "parth: Enabling symbolic analysis reuse for linear direct-solvers in the present of dynamic sparsity pattern via mesh-aware design.". *To be submitted to ACM SIGGRAPH*, 2024

2. Behrooz Zarebavani, Kazem Cheshmi, Bangtian Liu, Michelle Mills Strout, and Maryam Mehri Dehnavi. Hdagg: hybrid aggregation of loop-carried dependence iterations in sparse matrix computations. In *2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 1217–1227. IEEE, 2022

3. Behrooz Zarebavani, Foad Jafarinejad, Matin Hashemi, and Saber Salehkaleybar. cupc: Cuda-based parallel pc algorithm for causal structure learning on gpu. *IEEE Transactions on Parallel and Distributed Systems*, 31(3):530–542, 2019

Skills and Selected Courses

- **Skills**: C/C++, CUDA, Python, Java, LaTeX, OpenMP, git, scikit-learn
- **Parallel Computing Courses**: Parallel Processing (A+), Distributed Systems (A+), Causal Inference (A+), Advanced Computer Architecture(A+), Advanced Systems Programming(A+), Compilation Techniques for Parallel Processors (A-)
- **Machine Learning Courses**: Statistical Learning (A+), Theory of Learning (A+), Probability & Statistics (A+)
- **Graphic Courses**: Physics-based Animation (A+)

Honors and Awards

- **Top 0.1**% among more than 60000 participants in nation-wide University Entrance Exam for Master of Science. 2017
- **Qualified** for double-major program (EE and CE) at Amirkabir University of Technology 2015
- **Top 0.1**% among more than 350000 participants in nation-wide University Entrance Exam for Bachelor of Science. 2013

Research and Work Experience

Graduate Research Assistant | *University of Toronto* Sep 2020 - Present

My Ph.D. aims to build upon the objectives of my M.Sc. by delving into and applying more advanced concepts of parallel computing, not only at the hardware level but also at the algorithmic level. Additionally, I plan to apply this knowledge to the computational pipeline of optimizers used in applications such as training machine learning models and Time-integration in scientific simulations.

- **Current Project:** My current project aims to accelerate implicit time integration by incorporating machine learning and preconditioner techniques. Utilizing machine learning approaches, I am attempting to accelerate time-intensive steps in this process and to integrate this information with preconditioners to create a fast and theoretically sound optimizer. To be more specific, in each iteration of a Newton-based solver used in physics-based simulation with contact to find the positions of nodes in each frame, some parts of the simulation mesh move while many remain approximately constant. By using DNNs, we aim to reduce computation in many iterations by deactivating those parts that are likely to exhibit low movement, and focusing on the part of the linear system that define the direction toward the overall solution, leading to a smaller problem to solve in each iteration. As a result, reducing the computational overhead per iteration.

- **Mesh-aware Framework for Sparse Symbolic Analysis Reuse [1]: Parth** is a groundbreaking framework we developed for enhancing symbolic analysis in scenarios with dynamic sparsity patterns which happens in many applications such as physics-based simulations involving contact. It functions as an efficient "plug and play" module for high-performance direct solvers like Apple Accelerate, Intel MKL, and CHOLMOD, integrating with their interfaces. This innovative tool significantly boosts symbolic analysis performance, with improvements of up to 7.6x, and speedup the overall efficiency of direct solver computations by up to 2.7x per frame. Parth's effectiveness is rooted in its sophisticated ability to reuse symbolic computation data across different direct solver invocations, particularly beneficial in Newton-based optimizer iterations. Our findings and methodology will be submitted to SIGGRAPH 2024.

- **Sparse Kernels Scheduler [2]: HDagg**, our open-source scheduler, is specifically engineered to accelerate the computation of sparse kernels exhibiting loop-carried data dependencies, such as sparse triangular solve. HDagg creates computation orders to suit varying sparsity patterns, specific sparse kernels, and the hardware architecture in use. By finely tuning the balance between parallelism, locality, and synchronization, HDagg delivers exceptionally efficient computations. It notably surpasses the capabilities of existing state-of-the-art schedulers, achieving up to a 13x speedup in performance. https://github.com/BehroozZare/HDagg-benchmark

- **Resource Allocation Competition:** To secure the necessary resources for our groups' projects, we wrote a proposal to the Resource for Research Group (RRG) competition hold by Digital Research Alliance. This competition evaluates proposals and allocates resources based on their scores. Our proposal achieved a high score of 4.1 out of 5, enabling us to access resources that significantly contributed to two of my PhD publications and the publication of a NeurIPS paper by our group this year.

- **Mentorship:** I also gain experience by mentoring students through the process of research. This role has involved not just teaching research methodologies but also enabling students to navigate through the challenges inherent in academic research. The reward of this mentorship is mutual; for instance, one student found the learning process so interesting that they were inspired to document their insights on their personal webpage. Currently, I am collaborating with another student to analyze the advantages and the limitations of Monte-Carlo solvers. Our focus is on analyzing and enhancing these solvers, especially considering their emerging applications in machine learning tools for simulating complex phenomena.

Graduate Research Assistant | *Sharif University of Technology* Sep 2017 - Aug 2019

My objective for my M.Sc. was to gain experience in parallel computing and familiarize myself with GPU architecture, using these skills to accelerate emerging machine learning algorithms. To this end, my thesis and the coursework I undertook at Sharif University were centered around this objective.

- **GPU-based Causal Structure Learning Algorithm [3]:** We developed **cuPC**, a novel GPU-based parallel algorithm for uncovering causal relationships from observational data. This algorithm, designed to operate with an order-independent version of the PC algorithm, effectively handles multivariate normal distributions. cuPC is available in two variants, cuPC-E and cuPC-S, each employing a distinct parallelization approach. Our experimental results demonstrate impressive scalability across various numbers of variables, sample sizes, and graph densities. A noteworthy achievement is the reduction of runtime from over 11 hours to approximately 4 seconds on a challenging dataset, illustrating the algorithm's efficiency. On average, cuPC-E and cuPC-S realize speedups of 500x and 1300x, respectively, compared to traditional serial CPU implementations. The source code is accessible for public use. https://github.com/LIS-Laboratory/cupc

- **FPGA Accelerator for CNN Inference:** In this project, I developed and deployed a Convolutional Neural Network (CNN) inference module using FPGA (Field-Programmable Gate Array) hardware. The objective was to implement the trained CNN on the FPGA, effectively modeling the activation functions and pipelining the computations. This approach enabled the inference module to perform real-time object detection with low power consumption from a video stream at a rate of 30 frames per second (fps).

- **Predicting Stock Market Trends with a LSTM Model:** In this project, data from Iran's stock market was used to predict the trends of stock prices of two well-known companies. The project involved using Exploratory Data Analysis (EDA) techniques to clean the data and select features, and a LSTM model to train the model. My project achieved an accuracy of 92% in this endeavor.

Research Assistant | *AmirKabir University of Technology* Sep 2013 - Aug 2017

- **IoT Platform for Smart Irrigation:** In my B.Sc. thesis, I developed an Internet of Things (IoT) platform responsible for gathering environmental information, including soil and weather data, and transmitting it to a central processing unit. Later, this information was used to issue commands such as soil irrigation using a node attached to an electric valve. I designed a portable, low-power node with an ARM processor and developed a protocol on top of the Zigbee low-power transmission device, enabling the periodic transmission of critical information to the central unit for various analyses. The information was also accessible to users through a user interface. This platform was later utilized as a prototype at the Atrovan startup, becoming a part of their smart home system solution.

Intern | *Shabake Gostar Kayer* Jan 2017 - Mar 2017

- During my internship, I designed an IoT node for controlling the humidity and temperature in server rooms, as these are crucial factors for maintaining hardware in such environments. The node also supported the transmission protocol designed at Kayer company, serving as a solution for their monitoring software.

Contract | *Royan* Jun 2016 - Sep 2016

- As part of a contract with the Royan Institute, we developed and implemented a Monkey-Walking Corridor with retractable segments to monitor the rehabilitation of spinal-cord injured monkeys. The aim of this project was to provide a platform for analyzing the effectiveness different treatments for this type of injury in humans.