Arghya Bhattacharya, Ph.D.

Seattle, WA, USA || arghyabh@google.com || +1 (934) 777-9896 || https://www3.cs.stonybrook.edu/~argbhattacha/

Summary

- With 8+ years of experience in academic research and industry, currently building large-scale Bare Metal cloud compute infrastructure for Google Cloud Platform (GCP) Compute Engine.
- With 5+ years of experience in developing scalable software solutions, with a focus on **algorithm and performance engineering**, specifically, specialized in designing **cache- and memory-efficient algorithms** tailored for shared-memory and cloud systems. Additionally, experienced in crafting **batch-parallel graph-processing algorithms** optimized for processing large, time-evolving sparse graphs.
- Experienced in designing **ML-advised online algorithms**, i.e., algorithms with predictions. Furthermore, contributed to an **explainable graph neural network** for high-energy nuclear physics applications.

Skills

- **Technical Skills:** External-memory algorithm, Parallel algorithm, High-performance computing (HPC), Dynamic graph algorithm, Machine learning (ML) advice, ML explainability, Computer networks, File system fragmentation, Distributed computing, Version management (GitHub).
- **In-depth Programming Skills:** C/C++ (including Multithreading tools: std::thread, posix thread, OpenMP), Bash (Memory management: control group; Virtualization: QEMU, KVM; Network: iperf, netcat, FTP, Ethernet/InfiniBand networks, TCP/IP stack, RDMA, RoCE, OFED, and peer-to-peer data transfer systems), Python (Libraries: NumPy, Pandas, Scipy; ML: Keras, PyTorch, Scikit-learn, XGBoost, Docker Cvxpy; Data visualization: Matplotlib, Seaborn, Plotly; Web development: Streamlit, Flask), GitHub (Version control), GoogleSQL, Matlab, Latex, Markdown, HTML, CSS, JavaScript.

Education

- **Stony Brook University**, Dept. of Computer Science — Stony Brook, NY, USA
  Ph.D. in Computer Science, Advisor: Prof. Michael A. Bender — Sep. 2018 – Aug. 2024
  - **Thesis: Going Beyond Worst-Case: A Study on Cache Adaptivity and Machine Learning Advice.**
    Committee members: Dr. Michael A. Bender, Dr. Rezaul A. Chowdhury (chair), Dr. Joseph S. B. Mitchell, and Dr. Helen Xu.
  - My thesis sheds light on an algorithm design principle and an empirical framework design to evaluate when programs are cache- and memory-adaptive for external memory applications, i.e., programs maintain performance despite memory fluctuations typical in multicore, multithreaded, shared-memory, and cloud systems. It also shows how to redesign traditional online algorithms for several online problems, e.g., rent-or-buy problems, scheduling in dark, parallel paging with augmentation by single and multiple ML oracles.
  - **Graduate Courses:** Analysis of Algorithms, Computer Networks, Discrete Maths, Data Science, Introduction to Computer Vision, Theory of Database Systems, Medical Imaging.
  - **Academic Services:**
    - **Teaching Assistant:** Theory of Database Systems (CSE 532), System Fundamentals - II (CSE 320), Fundamentals of Information Technology (ISE 218).
    - **Mentored** high-school students for 4 years through High School Women in Science and Engineering (HS-WISE).
    - **Subreviewer** for ESA'23, SPAA'22, SPAA'23, SPAA'24, IPDPS'23, SEA'23, APOCS'23 and Reviewer for Soft Computing, Springer and Neural Computing and Applications, Springer.
- **Jadavpur University**, Dept. of Electrical Engg. — Kolkata, WB, India
  B.E. in Electrical Engineering, Advisor: Prof. Debangshu Dey — Jul. 2012 – May. 2016
  - **Related Coursework**: Advanced Instrumentation (special paper), Digital Signal Processing, Numerical Analysis and Computer Programming, Reliability Engineering, Signals and Systems, Circuit Theory, Control System Engineering.
  - Awarded the Jagadis Bose National Science Talent Search (JBNSTS) Senior Scholarship 2012.

Selected Work Experience (Check my Webpage for more details)

- **Google**, Google Cloud Platform Compute, Google Cloud — Seattle, WA, USA
  Software Engineer (L4), Manager: Vicky Xu — Sep. 2024 – Present
  - Engineering Google Cloud Platform (GCP) solutions to develop, maintain, and operationalize Bare Metal cloud solutions for demanding applications. The Google Computing Engine (GCE) Node Bare Metal servers allow users to run specialized high-performance computing (HPC) workloads, e.g. third-party virtualization software, applications with low-level access to the server, computationally demanding AI workloads, etc. Our team manages the regionally distributed cluster of servers with high-performance connection with a ultra-low-latency (ULL) network fabric.
  - Proven track record of spearheading CI/CD pipelines and overall DevOps for industry-first hardware shapes (ARM-based C4A and GPU-based A4X Max). Adept at leading cross-functional teams to build integrated, LLM-powered agentic AI tools that drastically reduce on-call resolution times and drive organizational productivity.
  - Built and owned comprehensive internal dashboards for observability to monitor the capacity, health, and availability of the Bare Metal server fleet, directly improving daily data-driven operational decisions.
- **Google**, Playback Services, YouTube — San Bruno, CA, USA
  Software Engineering Ph.D. Intern, Host: Dr. Jie (Jeremy) Zhou, Dr. Kevin Chen — May 2023 – Aug. 2023
  - Collaborated closely with the playback experience, media algorithms, transcoding, and edge streaming teams and extended UVQ to YouTube Shorts, and evaluated multiple ML models through A/B testing to comprehend subjective perception of video quality and learn from contextual factors (such as content type, device type, geolocation, and other streaming objectives).

- **Lawrence Berkeley National Lab**, PASSION Lab, Performance and Algorithms Research Group — Berkeley, CA, USA
  Visiting Student Research Intern, Host: Dr. Helen Xu — Sep 2023 – Dec 2023
  - **Dynamic GAP:** Expanded the experimental framework of the GAP@Berkeley project to incorporate dynamic graph algorithms. Designed batch-parallel algorithms for processing large sparse time-evolving graphs and conducted a comparative study with the top node-wise parallel static algorithm using this framework.
- **Nokia Bell Labs**, Network Systems and Security Research Lab, Bell Labs Core Research — Murray Hill, NJ, USA
  Cloud and Networking Intern, Host: Dr. Edward Grinshpun, Chuck Payette — Jun. 2022 – Aug. 2022
  - Researched out-of-band machine learning-based prediction-enhanced congestion control algorithms tailored for low-latency, high-volume, variable-bitrate applications, such as live video streaming, within 5G wireless network systems.
- **Stony Brook University**, Center of Excellence Wireless and Information Technology (CEWIT) — Stony Brook, NY, USA
  SPIR (Strategic Partnership for Industrial Resurgence) Intern, Host: Dr. Dantong Yu — Jan 2024 – Aug. 2024
  - **Explainable graph learning:** Collaborated with researchers from Brookhaven and Los Alamos national labs and designed an explainable graph neural network (GNN) to detect *heavy flavor decays* in real-time for the sPHENIX project, a high-energy nuclear physics experiment at Brookhaven National Laboratory.
  - **Terabits data transfer:** Collaborated with engineers at Sunrise Technology Inc. incubated at the Center of Excellence Wireless and Information Technology (CEWIT) on the development of a low-power embedded system (NVIDIA Jetson devices serving as DPUs) enabling high-speed data transfer (100 Gbps) across distributed file system nodes (CephFS) over the internet using RDMA protocol RoCEv2, optimizing performance for high-performance computing applications (funded by and in collaboration with the DoE). **[demo]**.
- **Stony Brook University**, Algorithms Lab, Dept. of Computer Science — Stony Brook, NY, USA
  Research Project Assistant, Advisor: Prof. Michael A. Bender, Prof. Rezaul A. Chowdhury — Jun. 2019 – Dec. 2022
  - **Cache-efficient algorithms:** Designed an empirical framework to evaluate when programs are cache- and memory-adaptive for external memory applications, i.e., programs maintain performance despite memory fluctuations typical in multicore, multithreaded, shared-memory, and cloud systems. Provided an algorithm design principle and reference implementation for cache-adaptive algorithms to solve fundamental problems such as **matrix multiplication**, **sorting**, and dynamic programming algorithm for the **longest common subsequence** (LCS) **[ESA'22 paper]**. **[code]**
  - **ML-advised algorithms:** Redesigned traditional online algorithms for **rent-or-buy problems** with augmentation by single and multiple ML oracles accounting for arbitrarily fluctuating discounts on the rent of the resource **[WALCOM'22 paper and TCS@Elsevier article]**. Subsequent work revisited conventional online decision-making problems, such as **list access problem [SIGMETRICS'23 poster]**, **job scheduling problem**, **green and parallel paging problem**.
  - **Filesystem aging:** Collaborated with a large group of researchers, engineers, and professors to evaluate the degradation of read performance across five production filesystems (`ext4`, `btrfs`, `xfs`, `zfs`, and `f2fs`), owing to poor data layout and disk fullness. This evaluation, conducted using microbenchmarks and application-level fragmentation benchmarks, demonstrated that the degradation could be mitigated with a B^{ε}-tree-based write-optimized in-kernel filesystem like `BetrFS` **[arXiv article]**. **[code]**

Selected Scientific Communications (Google Scholar has the complete list)

[1] **Arghya Bhattacharya**, Helen Xu, Michael A. Bender, Rezaul A. Chowdhury, et al., "When Are Cache-Oblivious Algorithms Cache Adaptive? A Case Study of Matrix Multiplication and Sorting," *30th Annual European Symposium on Algorithms* ***(ESA'22)***. **[paper] [slides]**

[2] **Arghya Bhattacharya**, Rathish Das, "Machine Learning Advised Algorithms for the Ski Rental Problem with a Discount," ***Theoretical Computer Science, Elsevier*** (the conference version was presented at *16th International Conference and Workshops on Algorithms and Computation* ***(WALCOM'22)***). **[article] [paper] [talk] [slides]**

[3] **Arghya Bhattacharya**, Dwaipayan Choudhury, and Debangshu Dey, "Edge-enhanced Bi-dimensional empirical mode decomposition based emotion recognition using fusion of feature set," ***Soft Computing, Springer*** (2018) 22: 889–903 (the conference version was presented at *First IEEE Conference on Control, Measurement and Instrumentation* ***(CMI'16)***). **[article] [paper]**