

Gregory Bolet

✉ gbolet@vt.edu 🌐 gregbolet 🌐 greg-bolet 🔗 people.cs.vt.edu/gbolet

EDUCATION

VIRGINIA TECH | *PhD student in Computer Science (CS)* Fall 2020 – Now

- GPA: 3.88/4.0, PhD Advisor: Kirk W. Cameron
- New Horizons Graduate Scholar (NHGS)

FRANKLIN & MARSHALL COLLEGE | *BA in CS & Math* Fall 2016 – Spring 2020

- GPA: 3.75/4.0, Computer Science and Mathematics Double Major
- Gates Millennium and Posse Miami Scholarships

RESEARCH

VIRGINIA TECH & LAWRENCE LIVERMORE NATIONAL LABORATORY (LLNL)

Blacksburg, VA & Livermore, CA (remote)

LLMs for GPU Performance Analysis

August 2024 – Now

- Collaborating with **Gal Oren**, **Giorgis Georgakoudis**, **Harshitha Menon**, **Konstantinos Parasyris**, and **Niranjan Hasabnis** in applying LLMs to automatically predict the runtime performance characteristics of CUDA and OpenMP GPU kernels from source code.
- Our most recent publication was accepted to the HPDC 2025 AI4Sys Workshop.

OpenMP Program Autotuning with Global Optimization Strategies August 2023 – May 2024

- Collaborated with **Giorgis Georgakoudis** and **Konstantinos Parasyris** to compare convergence times of several automatic optimization strategies for hyperparameter tuning of CPU-based OpenMP programs.
- Paper published at the IPDPS 2024 iWAPT Workshop.

Apollo Autotuner

May 2021 – Aug 2023

- Focused on tuning OpenMP codes online/offline at the end-to-end and region levels for both the CPU and GPU.
- Integrated PAPI performance counters, expanding Apollo's autotuner metrics beyond just timing regions; offering cache-based, memory, and FLOP/INTOP metrics.
- Implemented Bayesian Optimization (BO) modeling support to Apollo, exposing users to fast and strong autotuning models beyond the offered decision tree models.

AMD – GRADUATE RESEARCH INTERN

Seattle, WA

Deep Learning Recommendation Model (DLRM) Optimization

Feb 2022 – May 2022

- Worked with John Kalamatianos, Varun Agrawal, and Marko Scrbak on optimizing AMD's DLRM Embedding Bag (EBag) implementation in PyTorch.
- DLRM EBag is a sparse parallel table reduction operation written in OpenMP, where my strong initial analysis of the performance problems have led to a joint publication in IEEE Micro (see publications)

FRANKLIN & MARSHALL COLLEGE – UNDERGRADUATE RESEARCH ASSISTANT Lancaster, PA
Neural Acceleration of Cholesky Factorization May 2020 – Dec 2022

- Applied GRU and LSTM networks in PyTorch to predict Cholesky matrix factorization nonzero patterns for sparse matrices, achieving a 30.3x speedup over traditional serial methods.

Hybrid Cholesky Factorization Implementation Jan 2018 – May 2020

- Worked with Dr. Joshua Booth in creating a proof-of-concept for a Cholesky matrix factorization-based hybrid solver combining direct and iterative methods for sparse linear systems of equations.

PUBLICATIONS

- [1] G. Bolet, G. Georgakoudis, H. Menon, K. Parasyris, N. Hasabnis, H. Estes, K. W. Cameron, and G. Oren, “Can large language models predict parallel code performance?” *arXiv preprint arXiv:2505.03988*, 2025.
- [2] G. Bolet, G. Georgakoudis, K. Parasyris, K. W. Cameron, D. Beckingsale, and T. Gamblin, “An exploration of global optimization strategies for autotuning openmp-based codes,” in *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 741–750, 2024.
- [3] K. Nair, A.-C. Pandey, S. Karabannavar, M. Arunachalam, J. Kalamatianos, V. Agrawal, S. Gupta, A. Sirasao, E. Delaye, S. Reinhardt, R. Vivekanandham, R. Wittig, V. Kathail, P. Gopalakrishnan, S. Pareek, R. Jain, M. T. Kandemir, J.-L. Lin, G. G. Akbulut, C. R. Das, and G. Bolet, “Parallelization strategies for dlm embedding bag operator on amd cpus,” *IEEE Micro*, pp. 1–8, 2024.
- [4] J. D. Booth and G. S. Bolet, “Neural acceleration of graph based utility functions for sparse matrices,” *IEEE Access*, vol. 11, pp. 31619–31635, 2023.
- [5] J. Booth and G. Bolet, “Javelin: A scalable implementation for sparse incomplete lu factorization,” in *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 461–470, 2019.

WORK EXPERIENCE

LAWRENCE LIVERMORE NATIONAL LABORATORY (LLNL) Livermore, CA (remote)
Graduate Research Intern Summer 2021, Summer 2022, Fall 2023

- (see LLNL under Research section)

VIRGINIA TECH Blacksburg, VA
Graduate Teaching Assistant (GTA) Aug 2020 – Dec 2020

- GTA for Dr. Wu Feng’s CS4234 Parallel Computing course.
- Created homework assignments, graded weekly homework submissions, held weekly office hours (5 hours/week), and maintained lecture participation counts.
- Reinforced parallel programming concepts and workflows through the use of pThreads, OpenMP, CUDA, and OpenACC.

FRANKLIN & MARSHALL COLLEGE Lancaster, PA
Computer Science 1 (CS1) Lab Assistant Aug 2019 – Dec 2019

- Guided students in completing lab assignments designed to reinforce introductory Python programming concepts dealing with control flow, data structures, and algorithms.

LAWRENCE BERKELEY NATIONAL LABORATORY (LBNL)

Berkeley, CA

SULI Summer Intern

Jun 2019 – Aug 2019

- Collaborated with facility engineer Mark Friedrich in creating an SQL database, NodeJS server, and Bootstrap web interface that enabled engineers to maintain a catalog of their energy systems assets while automatically notifying them of their software vulnerabilities against the NIST Vulnerabilities Database (NVD)

MENTORING

POWERPACK PROJECT

Virginia Tech

Instrumented physical desktop hardware (e.g: CPU, GPU, motherboard) with shunt resistors to record power consumption profiles of AI/ML workloads.

Fall 2024 – Now Mentoring **Nitya Ganta** and **Bruno Zegada** to complete their CS undergraduate degrees; guided them on using PowerPack to profile and model the full-system power usage of modern AI workloads.

Fall 2023 – Spring 2024 Mentored **Cesar Smokowski** in completing his master's degree by profiling and comparing the power consumption of AI/ML workloads on AMD and NVIDIA GPU hardware.

SEEMORE KINETIC SCULPTURE

Virginia Tech

The SeeMore sculpture is a 15ft tall cylindrical collection of 256 Raspberry Pi computers, designed to visualize the communication phases amongst nodes in a distributed computing system running MPI-based scientific codes.

Fall 2024 Worked closely with undergraduate **Mallika Pamula** to design a 3D version of the sculpture in the Godot Game Engine, complete with animations so viewers could see the entire sculpture at-a-glance. Demoed a small 30-node cluster at SC24 on the showroom floor.

Fall 2023 – Spring 2024 Acted in the role of a Project Manager, guiding a team of undergraduate students to update and upgrade the SeeMore kinetic sculpture, adding and programming LED matrices to each node so as to visualize the compute done on each node. Trained **Hayden Estes** in the role of project manager.

LACEY KINETIC SCULPTURE

Virginia Tech

LACEY consisted of styluses attached to moving robotic arms that would interact with 30 tablets. LACEY demonstrated blockchain mining and distributed ledger concepts for general audiences.

Fall 2021 – Summer 2022: Mentored masters student **Eles Jones** and undergraduate **Skylar Liang** in designing control and visualization software for the sculpture.

April 2022: Presented LACEY at the *Smithsonian Museum* in Washington D.C. for the VT Accelerate Festival.

POSTERS

- iSeeMore: Design of a 256-Node RPi Cluster to Visualize LLM Computation Through Light and Movement for Mass Audiences
SC24 Student Research Competition Nov 19, 2024
- Online Tuning of CUDA Kernels using Bayesian Optimization
IPDPS PhD Forum Poster Session May 29, 2024
- Automagically Tuning the Execution of Parallel Programs
LLNL Summer Poster SLAM Aug 03, 2021
- On Improving The Security of Industrial Control Systems at LBNL
F&M Fall Poster Session Oct 25, 2019
LBNL Summer Poster Session Aug 07, 2019
F&M Summer Research Fair Apr 12, 2019

PRESENTATIONS

- An Exploration of Global Optimization Strategies for Autotuning OpenMP-based Codes
iWAPT Workshop (IPDPS 2024) May 24, 2024
- LACE: A Robotic Sculpture to Visualize Blockchain Computing
VT Accelerate Festival @ Smithsonian Nat'l Museum of Natural History Apr 8-11, 2022
Virginia Tech Science Festival Nov 10, 2021
- An Introduction To Cholesky Matrix Factorization
F&M Hackman Summer Scholar Meetup Jun 06, 2018
- Troubling Transoms! Exploring Multiple Solutions to the Same Problem
EPaDel Mathematics Conference Apr 01, 2017
- Drawing Perspective Letters Using Geometry
EPaDel Mathematics Conference Feb 02, 2017

JOURNAL / CONFERENCE REVIEWS

- Transactions on Parallel and Distributed Systems (TPDS)
Manuscript Reviewer Sept 2022
- Journal of Parallel and Distributed Computing (JPDC)
Manuscript Reviewer Nov 2021

SCHOLARSHIPS & AWARDS

2025 VT GPSS Travel Fund Grant (for HPDC)
2025 VT CS Department Student Conference Travel Grant (for HPDC)
2024 VT CS Department Student Conference Travel Grant (for IPDPS, and SC24)
2024 IPDPS Student Conference Travel Grant
2022 Eleanor Davenport Leadership Fund Scholarship from VT
2020 New Horizons Graduate Scholarship to attend Virginia Tech (graduate studies)
2016 Gates Millennium and Posse Miami scholarships to attend Franklin & Marshall College

VOLUNTEERING

- Supercomputing Conference (SC)

Student Volunteer	Nov 2024
Student Volunteer	Nov 2023
Student Volunteer	Nov 2022
Student Volunteer	Nov 2021