Updated: 5/3/26

Benjamin Berg
ben@cs.unc.edu
781-248-9467

Personal Research Interests

Parallel Scheduling, Caching, Queueing Theory, Performance Modeling of Computer Systems

Education

Carnegie Mellon University, School of Computer Science August 2016-August 2022
Ph.D. student, Computer Science Department
Advised by Mor Harchol-Balter

Duke University, Trinity School of Arts & Sciences Graduated May 2013
Major in Computer Science, minors in Economics and Math

Professional Experience

Assistant Professor, Computer Science Department, UNC Chapel Hill January 1, 2023-Present

Nominated Assistant Professor, Computer Science Department, UNC Chapel Hill July 2022-December2022

Research Assistant, Computer Science Department, Carnegie Mellon University August 2016-August 2022
Working with Ph.D. advisor Mor Harchol-Balter on applications of queuing theoretic models to the analysis of computer systems. Served as teaching assistant for *Performance Modeling of Computer Systems* graduate course. Current research projects include:

- Optimal scheduling of parallelizable jobs in multicore systems to minimize average latency
- Server-side caching for web services to minimize tail latency of web requests

Developer and Researcher, Caching Team, Facebook Summer 2019
Analyzed caching workloads and developed models for workload generation tools used to benchmark caching applications and hardware at Facebook.

Developer and Researcher, Production Kernel Team, Google Summer 2018
Worked on linux scheduler for the production kernel used on Google Cloud VM hosting servers.

Developer and Researcher, Life Long Kindergarten Group, MIT Media Lab August 2015 – August 2016
Developed new features for the Scratch programming language. Performed research based on usage data from the Scratch online community. Worked with researchers at the MIT Media Lab and University of Washington to model trends in collaboration in the Scratch online community. Described the influences of website design changes on patterns of user collaboration.

Technology Associate, Statistical Arbitrage, Susquehanna International Group August 2013 – August 2015
Worked on the statistical arbitrage desk, which specializes in low latency trading of securities. Worked on the core of the low latency trading platform, as well as scalable monitoring, alerting, and process scheduling tools. Worked with SIG assistant traders to complete quantitative analysis of trading opportunities related to long dated options on U.S. equities.

Honors

2025 Google Junior Faculty Award: Systems/ML
2021 SOSP Best Paper Award
2019 Facebook Fellow in Compute Storage and Efficiency. Selected from over 900 applicants.
2018 NSF GRFP Honorable Mention.
2013, Magna Cum Laude, Duke University

2013, Thesis resulting in graduation with High Distinction in Computer Science, Duke University

2013, Alex Vasilos Memorial Award in Computer Science for outstanding research, coursework, and contributions as a teaching assistant, Duke University

2013, Phi Beta Kappa

Publications

Work in submission

Sara McAllister, Benjamin Berg, et. al. *Scaling the IO wall with Declarative IO*. (In submission, 2026)

Zhouzi Li, Cindy Zhu, Arpan Mukhopadhyay, Mor Harchol-Balter, Benjamin Berg. *BOA Constrictor: Squeezing Performance out of GPUs in the Cloud via Budget-Optimal Allocation*. (In submission, 2026)

Zhouzi Li, Mor Hachol-Balter, Benjamin Berg. *Mean field optimal Core Allocation across Malleable jobs.* (In submission, 2026)

Peer Reviewed Publications

Zhongrui Chen, Adityo Anggraito Diletta Olliaro, Andrea Marin, Marco Ajmone Marsan, Benjamin Berg, Izzy Grosof, *Improving Nonpreemptive Multiserver Job Scheduling with Quickswap*. IFIP Performance 2025.

Zhongrui Chen, Isaac Grosof, and Benjamin Berg. *Improving Multiresource Job Scheduling with Markovian Service Rate Policies.* SIGMETRICS 2025.

Sara McAllister, Yucong Wang, Benjamin Berg, Daniel S. Berger, George Amvrosiadis, Nathan Beckmann, and Greg Ganger*, FairyWREN: A Sustainable Cache for Emerging Write-Read-Erase Flash Interfaces.* OSDI 2024.

Zhongrui Chen, Isaac Grosof, and Benjamin Berg. *Simple Policies for Multiresource Job Scheduling.* MAMA 2024.

Zhouzi Li, Benjamin Berg, Arpan Mukhopadhyay, Mor Harchol-Balter. How to Rent GPUs on a Budget. EPEW 2024.

Berg, Benjamin, Benjamin Moseley, Weina Wang, and Mor Harchol-Balter. *Asymptotically Optimal Scheduling of Multiple Parallelizable Job Classes*. arXiv preprint arXiv:2404.00346 2024 (*in minor revision for Stochastic Systems)*.

Ghosh, Bineet, Clara Hobbs, Shengjie Xu, Don Smith, James H. Anderson, P. S. Thiagarajan, Benjamin Berg, Parasara Sridhar Duggirala, and Samarjit Chakraborty. *Statistical verification of autonomous system controllers under timing uncertainties*. Real-Time Systems 2024.

Benjamin Berg. *A Principled Approach to Parallel Job Scheduling*. Diss. Carnegie Mellon University Pittsburgh, PA, 2022.

Sara McAllister, Benjamin Berg, Julian Tutuncu-Macias, Juncheng Yang, Sathya Gunasekar, Jimmy Lu, Daniel Berger, Nathan Beckmann, and Gregory R. Ganger. *Kangaroo: Theory and Practice of Caching Billions of Tiny Objects on Flash.* Transactions on Storage 2022.

Benjamin Berg, Justin Whitehouse, Benjamin Moseley, Weina Wang, Mor Harchol-Balter. *The Case for Phase-Aware Scheduling of Parallelizable Jobs*. IFIP Performance 2021.

Sara McAllister, Benjamin Berg, Julian Tutuncu-Macias, Juncheng Yang, Sathya Gunasekar, Jimmy Lu, Daniel Berger, Nathan Beckmann, and Gregory R. Ganger. *Kangaroo: Caching Billions of Tiny Objects on Flash.* SOSP 2021. **Awarded best paper**.

Benjamin Berg, Mor Harchol-Balter. *Optimal Scheduling of Parallel Jobs with Unknown Service Requirements.* Handbook of Research on Methodologies and Applications of Supercomputing, 2021.

Benjamin Berg, Daniel Berger, Sara McAllister, Isaac Grosof, Sathya Gunasekar, Jimmy Lu, Michael Uhlar, Jim Carrig, Nathan Beckmann, Mor Harchol-Balter, Greg Ganger.
The CacheLib Caching Engine: Design and Experiences at Scale. OSDI 2020.

Benjamin Berg, Rein Vesilo and Mor Harchol-Balter.
heSRPT: Parallel Scheduling to Minimize Mean Slowdown. IFIP Performance 2020, PEVA.

Benjamin Berg, Mor Harchol-Balter, Ben Moseley, Weina Wang, Justin Whitehouse.
Optimal Resource Allocation for Elastic and Inelastic Jobs. SPAA '20.

Benjamin Berg, Rein Vesilo, Mor Harchol-Balter.
heSRPT: Optimal Scheduling of Parallel Jobs with Known Sizes. MAMA 2019.

Daniel S. Berger, Benjamin Berg, Timothy Zhu, Siddhartha Senn, and Mor Harchol-Balter.
RobinHood: Tail Latency Aware Caching -- Dynamic Reallocation from Cache-Rich to Cache-Poor.
OSDI 2018.

Benjamin Berg, Jan-Pieter Dorsman, and Mor Harchol-Balter. *Towards optimality in parallel scheduling.* ACM Sigmetrics 2018, Proceedings of the ACM on Measurement and Analysis of Computing Systems, Vol. 1.

The Case for Dynamic Cache Partitioning for Tail Latency. Poster presented at NSDI 2017. Daniel S. Berger, Benjamin Berg, Timothy Zhu, and Mor Harchol-Balter.

Talks and Workshops

Benjamin Berg. *How to Rent GPUs on a* Budget. Invited talk at workshop: *Managing Specialized and Heterogeneous Architectures*. Simons Institute, Berkeley, CA. November 2025

Invited to workshop: *Algorithms for Memory Management*. Simons Institute, Berkeley, CA. October 2025

Benjamin Berg. Optimal Scheduling of Elastic and Inelastic Jobs. INFORMS Applied Probability Society Conference. Atlanta, Georgia. June 2025

Benjamin Berg. Asymptotically Optimal Scheduling of Multiple Parallelizable Job Classes. MAPSP Workshop. Kolding, Denmark. June 2024.

Benjamin Berg. *Optimal Scheduling of Elastic and Inelastic Jobs.* Systems Lunch at UNC. April 2024.

Benjamin Berg. *Optimal Scheduling of Elastic and Inelastic Jobs.* ACO Seminar at Carnegie Mellon University. December 2023.

Benjamin Berg. *Optimal Scheduling of Elastic and Inelastic Jobs.* STOR Colloquim at UNC Chapel Hill. November 2023.

Benjamin Berg. *Optimal Parallel Scheduling For Elastic and Inelastic Jobs.* INFORMS 2021.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at Georgia Tech, April 2021.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at Caltech, April 2021.

Benjamin Berg. *The CacheLib Caching Engine: Design and Experiences at Scale*. Talk presented at Columbia University, April 2021.

Benjamin Berg. *Optimal Parallel Scheduling For Elastic and Inelastic Jobs.* INFORMS 2020.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Invited talk at Red Hat Research Day, September 2020.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at UC Berkeley RISE Lab, March 2020.

Benjamin Berg. *Optimal Parallel Scheduling: From EQUI to heSRPT.* Talk presented at USC, Los Angeles, March 2019

Benjamin Berg. Optimal Parallel Scheduling of Jobs with Known Sizes. INFORMS 2019.

Benjamin Berg. *RobinHood: Tail Latency Aware Caching -- Dynamic Reallocation from Cache-Rich to Cache-Poor.* Talk presented at UMass Amherst, November 2018.

Benjamin Berg. *Towards optimality in parallel job scheduling*. Talk presented at INFORMS 2017.

Benjamin Berg. *Towards optimality in parallel job scheduling*. Talk presented at IFORS 2017.

Teaching

University of North Carolina at Chapel Hill

Spring 2026 – COMP 690, Performance Modeling of Computer Systems

Enrollment: 6

Fall 2025 – COMP 421, Files and Databases

Enrollment: 89

Fall 2024 – COMP 790, Performance Modeling of Computer Systems
Enrollment: 5

Spring 2023 – COMP 790, Computer Systems Design: A Quantitative Approach
Enrollment: 5

Fall 2022 – COMP 790, Performance Modeling of Computer Systems
Enrollment: 5

Advising

Ph.D Students

Xiaolong Huang, 2025-Present
Zhongrui Chen, 2022-Present

REU Students

Joshua Harrell, Summer 2024

Thesis Committee

Shengjie Xu, 2024-Present
Zelin Tong, 2023-Present
Shareef Ahmed, 2023-2025

Grants

Title: NSF-CIF Towards optimal scheduling for parallelizable machine learning training workloads
PI: Benjamin Berg
Co-PI: Mor Harchol-Balter, Weina Wang
Agency: NSF-CISE/CCF (CIF)
Status: Awarded
Dates: June 2024-June 2027
Amount: $363,266
Support: 33% SU
Award No.: 2403195
PM: Alfred Hero

Title: NSF-III High-Performance Scheduling for Modern Database System
PI: Benjamin Berg
Co-PI: Mor Harchol-Balter
Agency: NSF-CISE/IIS (III)
Status: Awarded
Dates: April 2024-April 2027
Amount: $275,000
Support: 16% SU
Award No.: 2322974
PM: Judith Cushing

Title: Google Junior Faculty Award in Systems/ML
Amount: $100,000
Source: Google
Status: Awarded
Support: Unrestricted gift

Professional Service

SIGMETRICS Shadow PC Chair, 2026
SIGMETRICS Program Committee, 2026
MLSys Program Committee, 2026
The Web Conference (WWW) reviewer, 2026
MLSys Program Committee, 2025
SIGMETRICS Student Research Competition Program Committee, 2025
SIGMETRICS Program Committee, 2025
NSF-OE ad-hoc reviewer, 2024.
2023-2024 UNC Chapel Hill Graduate Admissions Committee
2024 The Web Conference (WWW) reviewer
2023 SOSP Student Research Competition and Poster Session reviewer
2022-2023 UNC Chapel Hill Graduate Admissions Committee
2021-2022 Carnegie Mellon Tartan Scholars Mentor