

Jack Kosaian

jackkosaian.github.io \diamond jkosaian@cs.cmu.edu

Education

Carnegie Mellon University

Aug. 2017 - Present

Ph.D. in Computer Science

Advisor: Rashmi Vinayak

Thesis topic: Resource-efficiency and reliability for machine learning systems

University of Michigan, Ann Arbor

Sept. 2013 - Dec. 2016

B.S.E. in Computer Science & Engineering

Awards

NSF Graduate Research Fellowship (2017)

Angell Scholar (2015)

Branstrom Prize (2014)

Publications

Conference

Arithmetic-Intensity-Guided Fault Tolerance for Neural Network Inference on GPUs

[Jack Kosaian](#), K. V. Rashmi

SC 2021 (to appear)

Boosting the Throughput and Accelerator Utilization of Specialized CNN Inference Beyond Increasing Batch Size

[Jack Kosaian](#), Amar Phanishayee, Matthai Philipose, Debadeepta Dey, K. V. Rashmi

ICML 2021

Parity Models: Erasure-Coded Resilience for Prediction Serving Systems

[Jack Kosaian](#), K. V. Rashmi, Shivaram Venkataraman

ACM SOSP 2019

Vantage: Optimizing Video Upload for Time-shifted Viewing of Social Live Streams

Devdeep Ray, [Jack Kosaian](#), K. V. Rashmi, Srinivasan Seshan

ACM SIGCOMM 2019

EC-Cache: Load-Balanced, Low-Latency Cluster Caching with Online Erasure Coding

K. V. Rashmi, Mosharaf Chowdhury, [Jack Kosaian](#), Ion Stoica, and Kannan Ramchandran

USENIX OSDI 2016

Journal

Learning-Based Coded Computation

[Jack Kosaian](#), K. V. Rashmi, Shivaram Venkataraman

IEEE Journal on Selected Areas in Information Theory, 2020

Preprints

ECRM: Efficient Fault Tolerance for Recommendation Model Training via Erasure Coding

Kaige Liu*, [Jack Kosaian](#)*, K. V. Rashmi

arXiv:2104.01981

*Equal contribution

Talks

Parity Models: Erasure-Coded Resilience for Prediction Serving Systems

- ACM Symposium on Operating Systems Principles (SOSP 19), October 2019
- Foundations of Cloud and ML Infrastructure (Guest Lecture), October 2019

Resilient ML Inference via Erasure Coding

- Parallel Data Lab Retreat, November 2019
- Parallel Data Lab Retreat, October 2018

Teaching Experience

CMU 15-712: Advanced Operating Systems and Distributed Systems Teaching Assistant	Spring 2021
CMU 15-440: Distributed Systems Head Teaching Assistant	Spring 2020
UofM EECS 370: Introduction to Computer Organization Teaching Assistant	Fall 2015, Winter 2016

Outreach

CMU SCS Creative Technology Nights

- Assisted in teaching STEM concepts to middle school girls in the Pittsburgh area

Industry Experience

Microsoft Research <i>Research Intern</i> Mentor: Amar Phanishayee	June 2021 - Sep. 2021 <i>Remote (COVID-19)</i>
Microsoft Research <i>Research Intern</i> Mentor: Amar Phanishayee - Researched strategies to improve the performance of in-house DNNs on various accelerators	May 2019 - Aug. 2019 <i>Redmond, WA</i>
Google <i>Software Engineering Intern</i> BigQuery team - Analyzed performance and scalability bottlenecks of high-throughput read/write API	May 2017 - July 2017 <i>Seattle, WA</i>
Google <i>Software Engineering Intern</i> gVisor team - Explored hardware virtualization extensions for efficient sandboxing	May 2016 - Aug. 2016 <i>Mountain View, CA</i>
Epic Systems <i>Software Development Intern</i> - Developed dashboard for physicians to explore changes in patient health	May 2015 - Aug. 2015 <i>Madison, WI</i>