Gauri Joshi

http://andrew.cmu.edu/user/gaurij
Email: gaurij@andrew.cmu.edu

Employment

Carnegie Mellon University — Sept 2017 – present
Assistant Professor of Electrical and Computer Engineering — Pittsburgh PA

IBM T. J. Watson Research Center — July 2016 – Aug 2017
Research Staff Member — Yorktown Heights NY

Education

Massachusetts Institute of Technology — Sept 2012– Jun 2016

Ph.D in Electrical Engineering & Computer Science, GPA 5.0/5.0 (minor in Finance)
Thesis: Using Redundancy to Reduce Latency in Cloud Systems
Committee: Gregory Wornell (advisor), Emina Soljanin, Devavrat Shah

Massachusetts Institute of Technology — Sept 2010– Jun 2012

S.M. in Electrical Engineering & Computer Science, GPA 5.0/5.0
Thesis: On Playback Delay in Streaming Communication — **Best Thesis Award**
Advisors: Gregory Wornell, Yuval Kochman

Indian Institute of Technology Bombay — Jul 2005–Jun 2010

B. Tech & M. Tech in Electrical Engineering; GPA 9.77/10.0 — **Institute Gold Medal**
Thesis: On Relay-Assisted Cellular Networks
Advisor: Abhay Karandikar

Selected Awards and Honors

NSF CAREER Award 2021

Best Paper Award at ACM SIGMETRICS 2020

NSF CRII Research Initiation Award 2019

IBM Faculty Research Award 2018

William Martin Memorial Award for **Best Masters Thesis** in Computer Science, MIT 2012

Morris Joseph Levin Award for **Outstanding Oral Thesis Presentation**, MIT 2012

Claude E. Shannon Research Assistantship, MIT 2015–2016

Schlumberger **Faculty for the Future Fellowship** 2011–2015

Irwin and Joan Jacobs Presidential Fellowship, MIT 2010–2011

Institute Gold Medal for highest GPA in the undergraduate class, IIT Bombay 2010

Best student in Communications & Signal processing Award, IIT Bombay 2009

Selected among top 50 students in India for the International Chemistry Olympiad camp 2005

PhD Student Advising

Ankur Mallick	Fall 2017-present
Jianyu Wang	Fall 2017-present
Samarth Gupta (co-advised with Prof. Osman Yağan)	Fall 2017-present
Yae Jee Cho	Fall 2019-present
Tuhinangshu Choudhury (co-advised with Prof. Weina Wang)	Fall 2019-present
Ting-Wu (Rudy) Chin (co-advised with Prof. Diana Marculescu)	Spring 2020-present
Ahmet Inci (co-advised with Prof. Diana Marculescu)	Spring 2020-present
Divyansh Jhunjhunwala	Fall 2020-present

Teaching and Mentorship

Course Instructor

- 18-461/661: Introductory Machine Learning for Engineers, Carnegie Mellon — Spring 2019, 2020
- 18-847F: Foundations of Cloud and ML Infrastructure, Carnegie Mellon — Fall 2017, 2018, 2019

EECS REFS, Conflict Management and Mentorship in EECS, MIT — Jan 2014 - Jun 2016

Complete a 40-hour training course in conflict management and mediation, MIT — Jan 2014

Completed a semester-long Graduate Teaching Certificate Program offered by MIT — Spring 2013

Institute Student Mentor for freshmen, IIT Bombay — 2008–2010

Selected Publications (Journal and Peer-Reviewed Conference Papers)

[27] A. Mallick, U. Sheth, G. Palanikumar, M. Chaudhari, and G. Joshi, "Rateless Codes for Near-Perfect Load Balancing in Distributed Matrix-Vector Multiplication", *Proceedings of ACM SIGMETRICS*, June 2020, arXiv:1804.10331 **Best Paper Award**

[26] G. Joshi, Y. Liu, E. Soljanin, "On the Delay-Storage Trade-off in Coded Distributed Storage Systems", *IEEE Journal on Selected Areas of Communications*, volume 32, number 5, May 2014, arXiv:1305.3945

[25] G. Joshi, E. Soljanin, G. Wornell, "Efficient Redundancy Techniques to Reduce Latency in Cloud Systems", *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, volume 2, issue 2, May 2017, arXiv:1508.03599

[24] D. Wang, G. Joshi, G. Wornell, "Efficient Straggler Replication in Parallel Computing ", *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, 2019 arXiv:1503.03128

[23] S. Gupta, G. Joshi, and O. Yagan, "Correlated Multi-armed Bandits with a Latent Random Source" *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2020, arXiv:1808.059045

[22] J. Wang, H. Liang, and G. Joshi, "Overlap Local-SGD: An Algorithmic Approach to Hide Communication Delays in Distributed SGD", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2020

[21] X. Zhang, J. Wang, G. Joshi, and C. Joe-Wong, "Machine Learning on Volatile Instances", *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, Apr 2020

[20] J. Wang, A. Sahu, G. Joshi, and S. Kar, "MATCHA: Speeding up Decentralized SGD via Matching Decomposition Sampling", *Proceedings of the Neural Information Processing Systems (NeurIPS) Federated Learning workshop*, Dec 2019, **Distinguished Student Paper Award**, arXiv:1905.09435

[19] J. Wang and G. Joshi, "Cooperative SGD: A unified Framework for the Design and Analysis of Communication-Efficient SGD Algorithms", *Proceedings of the International Conf. on MachineLearning (ICML) CodML workshop*, Jun 2019, arXiv:1808.07576

[18] A. Mallick and G. Joshi, "Rateless Codes for Distributed Computations with Sparse Compressed Matrices", *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Jul 2019

[17] A. Mallick, M. Chaudhari, G. Joshi, "Fast and Efficient Distributed Matrix-Vector Multiplication Using Rateless Fountain Codes", *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.

[16] J. Wang, G. Joshi, "Adaptive Communication Strategies to Achieve the Best Error-Runtime Trade-off in Local-Update SGD", *SysML Conference*, Mar 2019.

[15] S. E. Anderson, A. Johnston, G. Joshi, G. L. Matthews, C. Mayer, and E. Soljanin, "Service Capacity Region of Content Access from Erasure Coded Storage", *Information Theory Workshop (ITW)*, Nov 2018.

[14] S. Gupta, G. Joshi, O. Yagan"Active Distribution Learning from Indirect Samples", *Allerton Conference on Communication, Control and Computing*, Oct 2018.

[13] S. Dutta, G. Joshi, S. Ghosh, P. Dube, P. Nagpurkar, "Slow and Stale Gradients Can Win the Race: Error-Runtime Trade-offs in Distributed SGD", *International Conference on Artificial Intelligence and Statistics (AISTATS)*, Apr 2018.

[12] G. Joshi, "Synergy via Redundancy: Boosting Service Capacity via Adaptive Task Replication", *ACM/IFIP Performance*, Nov 2017.

[11] M. Aktas, S. E. Anderson, A. Johnston, G. Joshi, S. Kadhe, G. L. Matthews, C. Mayer, and E. Soljanin, "On the Service Capacity Region of Accessing Erasure Coded Content", *Allerton Conference on Communication, Control and Computing*, Oct 2017.

[10] G. Joshi, "Boosting Service Capacity via Adaptive Task Replication", *ACM Sigmetrics MAMA workshop*, June 2017.

[9] G. Joshi, E. Soljanin, G. Wornell, "Efficient Task Replication to Reduce Latency in Cloud Systems", *Allerton Conference on Communication, Control and Computing*, Sep 2015.

[8] G. Joshi, E. Soljanin, G. Wornell, "Queues with Redundancy: Latency-Cost Analysis", *SIGMETRICS Workshop on Mathematical Modeling and Analysis*, Jun 2015.

[7] D. Wang, G. Joshi, G. Wornell, "Using Straggler Replication to Reduce Latency in Large-scale Parallel Computing", *SIGMETRICS Workshop on Distributed Cloud Computing*, Jun 2015.

[6] K. Mahadaviani, A. Khisti, G. Joshi, G. Wornell, "Playback Delay in Streaming Communication with Feedback *International Symposium on Information Theory (ISIT)*, Jun 2015.

[5] G. Joshi, Y. Kochman, G. Wornell, "Throughput-Smoothness Trade-offs in Multicasting of an Ordered Packet Stream", *International Symposium on Network Coding (NetCod)*, Jun 2014.

[4] D. Wang, G. Joshi, G. Wornell, "Efficient Job Replication for Fast Response Times in Parallel Computation", *ACM SIGMETRICS*, Jun 2014.

[3] G. Joshi, Y. Kochman, G. Wornell, "The Effect of Block-wise Feedback on the Throughput-Delay Trade-off in Streaming", *INFOCOM Workshop on Communication and Networking Techniques for*

Contemporary Video, Apr 2014.

[2] G. Joshi, E. Soljanin, "Round-robin Overlapping Generations Coding for Fast Content Download", *International Symposium on Information Theory (ISIT)*, Jul 2013.

[1] G. Joshi, Y. Kochman, G. Wornell, "On Playback Delay in Streaming Communication", *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Jul 2012.

Invited Seminar/Conference Talks (not corresponding to contributed papers)

Speeding up Distributed SGD via Communication-Efficient Model Aggregation

University of Michigan, MI	Mar 2021
University of Virginia, VA	Oct 2020
University of Texas at Austin, TX	Mar 2020
ITA Workshop, UCSD, San Diego, CA	Feb 2020
ML Seminar, Univ. of Washington, Seattle WA	Oct 2019
EE Seminar, Harvard SEAS, Cambridge MA	Oct 2019
RLE Seminar, Massachusetts Inst. Tech, Cambridge MA	Oct 2019

Correlated Multi-armed Bandit Algorithms

INFORMS Annual Meeting, Seattle WA	Oct 2019

Fast Distributed Machine Learning with Slow and Stale Updates

SILO Seminar, University of Wisconsin-Masion	Nov 2018
INFORMS Annual Meeting, Phoenix AR	Nov 2018
Google Inc., Pittsburgh PA	May 2018
Georgia Institute of Technology	Mar 2018
Information Theory and Application Workshop, UCSD	Feb 2018

Efficient Redundancy Techniques to Reduce Latency in Cloud Systems

Rutgers University, New Brunswick NJ	Mar 2017
Indian Institute of Technology Bombay, Mumbai	Dec 2016
Tata Institute of Fundamental Research, Mumbai	Dec 2016
Texas A&M University TX	Oct 2016
Carnegie Mellon University, Pittsburgh PA	Mar 2016
University of Southern California, Los Angeles, CA	Mar 2016
IBM T. J. Watson Research Center, Yorktown Heights NY	Feb 2016
Princeton University, EE Departmental Seminar, NJ	Feb 2016
Graduation Day, Information Theory and Applications Workshop UCSD	Feb 2016
IBM Research Student Workshop on Cloud and Data Services, NY	Dec 2015
Rising Stars in EECS, academic workshop for women, MIT	Nov 2015

Using Coding to Reduce Delay in Content Access

Three-lecture series on Coding Theory at Clemson University	Mar 2018
Joint Mathematics Meeting, Atlanta	Jan 2017

Mobile Networked Systems Group, MIT	Sept 2014
Google Headquarters, Mountain View, CA	Aug 2014
University of California Berkeley CA	Aug 2014
DIMACS Workshop on Green Data Storage, Rutgers NJ	Dec 2013

Throughput-Smoothness Trade-offs in Streaming Communication

BIRS Workshop on Mathematical Coding Theory for Streaming	Oct 2015
University of Toronto, Canada	May 2014
Indian Institute of Technology (IIT) Bombay	Aug 2012

Selected Professional Service and Leadership

Editorials and Organizing Roles

Special Session organizer at the Asilomar Conference	2020
Session Chair and Judge for Best Student Presentation Award, ITA Workshop UCSD	2020
Associate Editor of the IEEE Open Journal on Signal Processing	2020–
Area Chair, International Conf. on Machine Learning (ICML)	2020
Publications/Submissions Chair, ACM MobiHoc	2020
Publicity Chair, MLSys Conference	2020
ICML workshop on redundancy techniques in machine learning	2019
General Co-Chair, Scalable Deep Learning at IPDPS	2019
Machine Learning in Science and Engineering, ECE Track (24 invited talks)	2019

Technical Program Committee

International Conference on Machine Learning (ICML)	2019
Information Theory Workshop (ITW)	2020
ACM MobiHoc Conference	2019, 2020
International Symposium on Information Theory (ISIT)	2018, 2019
Machine Learning and Systems (MLSys) Conference	2019, 2020

Reviewer

IEEE Transactions on Communications, IEEE Transactions on Information Theory, IEEE Journal of Selected Areas of Communications, IEEE Transactions on Wireless Communications, IEEE Communication Letters, IEEE Wireless Communication Letters, IEEE/ACM Transactions on Networking, Journal of Operations Research, Performance Evaluation, IEEE Transactions on Services Computing, ACM Transactions on Modeling and Performance Evaluation of Computer Systems. International Symposium on Information Theory (ISIT), Information Theory Workshop (ITW), International Conference on Communications (ICC), Conference on Learning Theory (COLT), International Conference on Machine Learning (ICML), Neural Information Processing Systems (NeurIPS)