MOSHARAF CHOWDHURY

Computer Science and Engineering
University of Michigan
4820 BBB, 2260 Hayward Street
Ann Arbor, MI 48109-2121, USA

+1 (734) 764-4148
mosharaf@umich.edu
https://www.mosharaf.com
https://www.symbioticlab.org

RESEARCH INTERESTS

My research enables application-infrastructure symbiosis across different layers of software and hardware stacks in wide-area, datacenter-scale, and rack-scale computing in the context of Big Data and AI/ML applications.

EDUCATION

Ph.D. in Computer Science, 2015
Electrical Engineering and Computer Sciences
University of California, Berkeley
Dissertation Title: Coflow: A Networking Abstraction for Distributed Data-Parallel Applications
Advisor: Ion Stoica

Master of Mathematics, 2009
Cheriton School of Computer Science
University of Waterloo

Bachelor of Science and Engineering, 2007
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology (BUET)

EMPLOYMENT

Associate Professor, 2022–Present
Assistant Professor, 2016–2022
Electrical Engineering and Computer Science
University of Michigan

Visiting Researcher, 2013–2014
Facebook, Inc.

Research Intern, Summer 2011
Microsoft Research, Redmond

Research Intern, Summer 2010
Microsoft Research, Cambridge, UK

Software Engineer, 2005–2006
Information Engineers and Consultants Bangladesh Ltd.

HONORS & AWARDS

David E. Liddle Research Excellence Award, 2025

Google Research Scholar Award, 2023

Meta Systems Research Award, 2022

Meta Networking Research Award, 2022

Morris Wellman Faculty Development Professorship, 2021

NSF CAREER Award, 2019

VMware Early Career Faculty Award, 2019

Google Faculty Research Award, 2016

ACM SIGCOMM Doctoral Dissertation Award, 2015

Facebook Fellowship, 2012

UC Berkeley Graduate Fellowship, 2009

Outstanding Teaching Assistant Award, University of Waterloo, 2009

Cheriton Scholarship, 2009

University of Waterloo Alumni Gold Medal, Finalist, 2009

University of Waterloo Graduate Entrance Scholarship, 2007

PAPER AWARDS

Top Picks, Honorable Mention, IEEE MICRO, 2024

Featured Article, IEEE Access, 2022

Best Paper Award, ACM SOSP ResilientFL, 2021

Distinguished Artifact Award, USENIX OSDI, 2021

Best Paper Award, USENIX ATC, 2020

Best Paper Award, ACM SIGMOD GRADES-NDA, 2018

Best Paper Award, USENIX NSDI, 2012

Community Award, Honorable Mention, USENIX NSDI, 2012

AWARDS TO STUDENTS AND/OR PROJECTS

Dennis M. Ritchie Doctoral Dissertation Award, Honorable Mention, 2024 → Fan Lai

UM CSE DEI Service Award, 2024 → Jiachen Liu

Mozilla Technology Fund AI and Environmental Justice Award, 2024 → Zeus

David J. Kuck CSE Dissertation Prize, 2023 → Fan Lai

Machine Learning and Systems Rising Stars, 2023 → Fan Lai

Machine Learning and Systems Rising Stars, 2023 → Jiachen Liu

Richard and Eleanor Towner Prize, 2023 → Fan Lai

Carbon Hack 22, Runner-up, 2022 → Zeus

ACM SIGMOD Systems Award, 2022 → Apache Spark

Meta Ph.D. Fellowship, Finalist, 2022 → Fan Lai

Meta Ph.D. Fellowship, Finalist, 2022 → Hasan Al Maruf

NSF Graduate Research Fellowship, 2017 → Jack Kosaian

IMPACT

SymbioticLab projects are open-source and available via **https://symbioticlab.org/open-source/**. Many have led to adoption and large-scale deployments in academia and industry.

MAJOR RESEARCH PROJECTS

Cornstarch, 2022–Present
Cornstarch is the first distributed multimodal AI platform that is designed from the ground up to support training, fine tuning, and inference of large multimodal AI models.

Zeus, 2021–Present
Zeus is the first energy optimization software platform for AI workloads that minimizes the energy consumption for a given time budget.

FedScale, 2019–2024
FedScale is a federated learning and analytics platform that enables distributed privacy-preserving AI without copying raw user data into the cloud.

Salus, 2017–2022
Multi-scale GPU resource management solutions for inference, training, and hyperparameter tuning of AI workloads both for single GPUs and GPU clusters.

Infiniswap, 2016–2022
Infiniswap is the first memory disaggregation solution that allows unmodified applications to use remote memory (almost) at line speed and ushered in the era of practical disaggregation.

Coflow, 2010–2016
Coflow generalizes traditional point-to-point flow-based communication in computer networks and enables application-aware data-parallel networking for Big Data and AI workloads.

Apache Spark, 2009–2014
Apache Spark is the industry-leading unified computation platform for Big Data and AI/ML workloads with massive adoption.

ViNEYard, 2008-2012
ViNEYard is an extensible framework for allocating resources in software-defined networking (SDN), network function virtualtion (NFV), multi-tenant datacenters, 5G network slicing, and edge computing.

PRODICTIONALIZATION IN THE INDUSTRY

- **Oobleck** (SOSP'23) → Google Gemini 1.0 Ultra training.
- **AdaEmbed** (OSDI'23) → Meta recommendation engines.
- **Zeus** (NSDI'23) → PyTorch Ecosystem.
- **TPP** (ASPLOS'23) → Meta infrastructure and Linux kernel v5.18.
- **Aequitas** (SIGCOMM'22) → Google cloud networking infrastructure.
- **FedScale** (ICML'22) → LinkedIn Data and AI Foundations.
- **Oort** (OSDI'21) → Cisco Flame client selection.
- **EC-Cache** (OSDI'16) → Amazon AWS Lambda storage and caching system.
- **Sinbad** (SIGCOMM'13) → Facebook HDFS codebase.
- **HARP** (SIGCOMM'12) → Microsoft Bing datacenter server provisioning.
- **Apache Spark** (NSDI'12) → Leading solution in the $100B Data + AI industry.
- **Orchestra** (SIGCOMM'11) → Apache Spark's default broadcast algorithm.

PUBLICATIONS

28,500+ citations; h-index = 43; 32 papers with 100+ citations [Google Scholar, April 2025]

Authors in **bold** are students advised by me.

REFEREED FULL CONFERENCE PAPERS

- **J. Liu**, F. Lai, **D. Ding**, **Y. Zhang**, M. Chowdhury, Venn: Resource Management Across Federated Learning Jobs, MLSys 2025. Acceptance Rate: NN.NN%
- P. Tholoniat, K. Kostopoulou, M. Chowdhury, A. Cidon, R. Geambasu, M. Lécuyer, Junfeng Yang, DPack: Efficiency-Oriented Privacy Budget Scheduling, ACM EuroSys 2025. Acceptance Rate: NN.NN%
- S. Srivastava, A. Mittal, A. Agrawal, V. Moskovich, M. Brevard, M. Chowdhury, INFA-FinOps for Cloud Data Integration, IEEE BigData 2024. Acceptance Rate: 35.19%
- P. T. J. Kon, **J. Liu**, Y. Qiu, W. Fan, T. He, L. Lin, H. Zhang, Q. M. Park, G. S. Elengikal, Y. Kang, A. Chen, M. Chowdhury, M. Lee, X. Wang, IaC-Eval: A Code Generation Benchmark for Infrastructure-as-Code Programs, NeurIPS 2024. Acceptance Rate: 25.3%
- **J. Chung**, **Y. Gu**, **I. Jang**, **L. Meng**, N. Bansal, M. Chowdhury, Reducing Energy Bloat in Large Model Training, ACM SOSP 2024. Acceptance Rate: 17.34%
- Y. Zhong, D. Berger, C. Waldspurger, I. Agarwal, R. Agarwal, F. Hady, K. Kumar, M. Hill, M. Chowdhury, A. Cidon, Managing Memory Tiers with CXL in Virtualized Environments, USENIX OSDI 2024. Acceptance Rate: 17.38%
- **Y. Zhu**, **J. Liu**, M. Chowdhury, **F. Lai**, FedTrans: Efficient Federated Learning via Multi-Model Transformation, MLSys 2024. Acceptance Rate: 22.02%
- **Y. Zhang**, X. Zhang, G. Ananthanarayanan, A. Iyer, Y. Shu, V. Bahl, Z. M. Mao, M. Chowdhury, Vulcan: Automatic Query Planning for Live ML Analytics, USENIX NSDI 2024. Acceptance Rate: 18.64%
- **J. Liu**, **F. Lai**, **Y. Dai** A. Akella, H. Madhyastha, M. Chowdhury, Auxo: Efficient Federated Learning via Scalable Client Clustering, ACM SoCC 2023. Acceptance Rate: 31%
- **I. Jang**, **Z. Yang**, Z. Zhang, X. Jin, M. Chowdhury, Oobleck: Resilient Distributed Training of Large Models Using Pipeline Templates, ACM SOSP 2023. Acceptance Rate: 18.78%
- **F. Lai**, W. Zhang, R. Liu, W. Tsai, X. Wei, Y. Hu, S. Devkota, J. Huang, J. Park, X. Liu, Z. Chen, E. Wen, P. Rivera, J. You, J. Chen, M. Chowdhury, AdaEmbed: Adaptive Embedding for Large-Scale Recommendation Models, USENIX OSDI 2023. Acceptance Rate: 21.18%
- **H. A. Maruf**, Y. Zhong, H. Wang, M. Chowdhury, A. Cidon, C. Waldspurger, Memtrade: Marketplace for Disaggregated Memory Clouds, ACM SIGMETRICS 2023. Acceptance Rate: 16.52%
- E. Wang, A. Kannan, Y. Liang, B. Chen, M. Chowdhury, FLINT: A Platform for Federated Learning Integration, MLSys 2023. Acceptance Rate: 22%
- Y. Wang, D. Sun, K. Chen, **F. Lai**, M. Chowdhury, Egeria: Efficient DNN Training with Knowledge-Guided Layer Freezing, ACM EuroSys 2023. Acceptance Rate: 16.12%
- **J. You***, **J. Chung***, M. Chowdhury (*Equal contribution), Zeus: Understanding and Optimizing GPU Energy Consumption of DNN Training, USENIX NSDI 2023. Acceptance Rate: 17.14%
- **F. Lai**, **Y. Dai**, H. V. Madhyastha, M. Chowdhury, ModelKeeper: Accelerating DNN Training via Automated Training Warmup, USENIX NSDI 2023. Acceptance Rate: 17.14%
- **H. A. Maruf**, H. Wang, A. Dhanotia, J. Weiner, N. Agarwal, P. Bhattacharya, C. Petersen, M. Chowdhury, S. Kanaujia, P. Chauhan, TPP: Transparent Page Placement for CXL-Enabled Tiered Memory, ACM ASPLOS 2023. Acceptance Rate: 21.4% **IEEE MICRO Top Picks (Honorable Mention)**
- P. Das, S. Srivastava, V. Moskovich, A. Chaturvedi, A. Mittal, Y. Xiao, M. Chowdhury, CDI-E: An Elastic Cloud Service for Data Engineering, VLDB 2022. Acceptance Rate: 36.67%
- **Y. Zhang**, G. Kumar, N. Dukkipati, X. Wu, P. Jha, M. Chowdhury, A. Vahdat, Aequitas: Admission Control for Performance-Critical RPCs in Datacenters, ACM SIGCOMM 2022. Acceptance Rate: 19.57%

- **F. Lai**, **Y. Dai**, **S. S. V. Singapuram**, **J. Liu**, **X. Zhu**, H. V. Madhyastha, M. Chowdhury, FedScale: Benchmarking Model and System Performance of Federated Learning at Scale, ICML 2022. Acceptance Rate: 21.94%
- **Y. Zhang**, **Y. Tan**, B. Stephens, M. Chowdhury, Justitia: Software Multi-Tenancy in Hardware Kernel-Bypass Networks, USENIX NSDI 2022. Acceptance Rate: 19.4%
- Y. Lee*, **H. A. Maruf***, A. Cidon, M. Chowdhury, K. G. Shin (*Equal contribution), Hydra: Resilient and Highly Available Remote Memory, USENIX FAST 2022. Acceptance Rate: 21.54%
- Z. Yu, **C. Hu**, J. Wu, X. Sun, V. Braverman, M. Chowdhury, Z. Liu, X. Jin, Programmable Packet Scheduling with a Single Queue, ACM SIGCOMM 2021. Acceptance Rate: 22.82%
- **F. Lai**, **X. Zhu**, H. V. Madhyastha, M. Chowdhury, Oort: Efficient Federated Learning via Guided Participant Selection, USENIX OSDI 2021. Acceptance Rate: 18.79%. **Distinguished Artifact Award**
- **J. You**, J. Wu, X. Jin, M. Chowdhury, Ship Compute or Ship Data? Why Not Both?, USENIX NSDI 2021. Acceptance Rate: 15.99%
- **P. Yu***, **J. Liu***, M. Chowdhury (*Equal contribution), Fluid: Resource-Aware Hyperparameter Tuning Engine, MLSys 2021. Acceptance Rate: 23.53%
- Z. Yu, **Y. Zhang**, V. Braverman, M. Chowdhury, X. Jin, NetLock: Fast, Centralized Lock Management Using Programmable Switches, ACM SIGCOMM 2020. Acceptance Rate: 21.6%
- **H. A. Maruf**, M. Chowdhury, Effectively Prefetching Remote Memory with Leap, USENIX ATC, 2020. Acceptance Rate: 18.68%. **Best Paper Award**
- T. N. Le, X. Sun, M. Chowdhury, Z. Liu, AlloX: Compute Allocation in Hybrid Clusters, ACM EuroSys 2020. Acceptance Rate: 18.38%
- **P. Yu**, M. Chowdhury, Salus: Fine-Grained GPU Sharing Primitives for Deep Learning Applications, MLSys 2020. Acceptance Rate: 19.2%
- **F. Lai**, **J. You**, **X. Zhu**, H. V. Madhyastha, M. Chowdhury, Sol: Fast Distributed Computation Over Slow Networks, USENIX NSDI 2020. Acceptance Rate: 18.36%
- M. Uluyol, A. Huang, A. Goel, M. Chowdhury, H. V. Madhyastha, Near-Optimal Latency Versus Cost Tradeoffs in Geo-Distributed Storage, USENIX NSDI 2020. Acceptance Rate: 18.36%
- M. Chowdhury, S. Khuller, M. Purohit, S. Yang, **J. You**, Near Optimal Coflow Scheduling in Networks, ACM SPAA 2019. Acceptance Rate: 33%
- **J. Gu**, M. Chowdhury, K. G. Shin, Y. Zhu, M. Jeon, J. Qian, H. Liu, C. Guo, Tiresias: A GPU Cluster Manager for Distributed Deep Learning, USENIX NSDI 2019. Acceptance Rate: 14.76%
- A. P. Iyer, L. E. Li, M. Chowdhury, I. Stoica, Mitigating the Latency-Accuracy Tradeoff in Mobile Data Analytics Systems, ACM MobiCom 2018. Acceptance Rate: 22.46%
- K. Mahajan, M. Chowdhury, A. Akella, S Chawla, Dynamic Query Re-Planning Using QOOP, USENIX OSDI 2018. Acceptance Rate: 18.29%
- D. Y. Yoon, M. Chowdhury, B. Mozafari, Distributed Lock Management with RDMA: Decentralization Without Starvation, ACM SIGMOD 2018. Acceptance Rate: 19.52%
- H. Zhang, J. Zhang, W. Bai, K. Chen, M. Chowdhury, Resilient Datacenter Load Balancing in the Wild, ACM SIGCOMM 2017. Acceptance Rate: 14.4%
- **J. Gu**, Y. Lee, **Y. Zhang**, M. Chowdhury, K. G. Shin, Efficient Memory Disaggregation with Infiniswap, USENIX NSDI 2017. Acceptance rate: 18.04%
- R. Grandl, M. Chowdhury, A. Akella, G. Ananthanarayanan, Altruistic Scheduling in Multi-Resource Clusters, USENIX OSDI 2016. Acceptance rate: 18.08%
- K. V. Rashmi, M. Chowdhury, **J. Kosaian**, I. Stoica, K. Ramchandran, EC-Cache: Load-Balanced, Low-Latency Cluster Caching with Online Erasure Coding, USENIX OSDI 2016. Acceptance Rate: 18.08%
- H. Zhang, L. Chen, B. Yi, K. Chen, M. Chowdhury, Y. Geng, CODA: Toward Automatically Identifying and Scheduling Coflows in the Dark, ACM SIGCOMM 2016. Acceptance Rate: 17.33%

- M. Chowdhury, Z. Liu, A. Ghodsi, I. Stoica, HUG: Multi-Resource Fairness for Correlated and Elastic Demands, USENIX NSDI 2016. Acceptance Rate: 19.74%
- M. Chowdhury, I. Stoica, Efficient Coflow Scheduling Without Prior Knowledge, ACM SIGCOMM 2015. Acceptance Rate: 15.83%
- M. Chowdhury, Y. Zhong, I. Stoica, Efficient Coflow Scheduling with Varys, ACM SIGCOMM 2014. Acceptance Rate: 18.99%
- M. Chowdhury, S. Kandula, I. Stoica, Leveraging Endpoint Flexibility in Data-Intensive Clusters, ACM SIGCOMM 2013. Acceptance Rate: 15.83%
- L. Popa, G. Kumar, M. Chowdhury, A. Krishnamurthy, S. Ratnasamy, I. Stoica, FairCloud: Sharing The Network In Cloud Computing, ACM SIGCOMM 2012. Acceptance Rate: 13.62%
- P. Bodik, I. Menache, M. Chowdhury, P. Mani, D. Maltz, I. Stoica, Surviving Failures in Bandwidth-Constrained Datacenters, ACM SIGCOMM 2012. Acceptance Rate: 13.62%
- M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, I. Stoica, Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, USENIX NSDI 2012. Acceptance Rate: 17.75%. **Best Paper Award** | **Community Award (Honorable Mention)**
- M. Chowdhury, M. Zaharia, J. Ma, M. I. Jordan, I. Stoica, Managing Data Transfers in Computer Clusters with Orchestra, ACM SIGCOMM 2011. Acceptance Rate: 14.35%
- N. Butt, M. Chowdhury, R. Boutaba, Topology-Awareness and Reoptimization Mechanism for Virtual Network Embedding, IFIP NETWORKING 2010. Acceptance Rate: 23.76%
- N. M. M. K. Chowdhury, M. R. Rahman, R. Boutaba, Virtual Network Embedding with Coordinated Node and Link Mapping, IEEE INFOCOM 2009. Acceptance Rate: 19.65%
- N. M. M. K. Chowdhury, F.-E Zaheer, R. Boutaba, iMark: An Identity Management Framework for Network Virtualization Environment, IFIP/IEEE IM 2009. Acceptance Rate: 32.8%

JOURNAL ARTICLES

- S. Qi, C. Jin, M. Chowdhury, Z. Liu, X. Liu, X. Jin, Pyxis: Scheduling Mixed Tasks in Disaggregated Datacenters, IEEE Transactions on Parallel and Distributed Systems, 35(9):1536–1550, 2024.
- N. Shi, **F. Lai**, R. Kontar, M. Chowdhury, Fed-ensemble: Ensemble Models in Federated Learning for Improved Generalization and Uncertainty Quantification, IEEE Transactions on Automation Science and Engineering, 21(3):2792–2803, 2024.
- Z. Wan, X. Wang, C. Liu, S. Alam, Y. Zheng, **J. Liu**, Z. Qu, S. Yan, Y. Zhu, Q. Zhang, M. Chowdhury, M. Zhang, Efficient Large Language Models: A Survey, Transactions on Machine Learning Research, 2024
- K. N. Gilley, L. Baroudi, M. Yu, I. Gainsburg, N. Reddy, C. Bradley, C. Cislo, M. L. Rozwadowski, C. A. Clingan, M. S. DeMoss, T. Churay, K. Birditt, N. Colabianchi, M. Chowdhury, D. Forger, J. Gagnier, R. F. Zernicke, J. L. Cunningham, S. M. Cain, M. Tewari, and S. W. Choi, Risk Factors for COVID-19 in College Students Identified by Physical, Mental, and Social Health Reported During the Fall 2020 Semester: Observational Study Using the Roadmap App and Fitbit Wearable Sensors, JMIR Mental Health, 9(2):e34645, 2022.
- R. Kontar, N. Shi, X. Yue, S. Chung, E. Byon, M. Chowdhury, J. Jin, W. Kontar, N. Masoud, M. Noueihed, C. E. Okwudire, G. Raskutti, R. Saigal, K. Singh, Z. Ye, The Internet of Federated Things (IoFT), IEEE Access, 9:156071–156113, 2021. **Featured Article**
- V. Gupta, T. M. Braun, M. Chowdhury, M. Tewari, S. W. Choi, A Systematic Review of Machine Learning Techniques in Hematopoietic Stem Cell Transplantation (HSCT), Sensors, 20(21), 6100, 2020
- F. Samuel, M. Chowdhury, R. Boutaba, PolyViNE: Policy-based Virtual Network Embedding Across Multiple Domains, Journal of Internet Services and Applications, 4(6):1–23, 2013
- M. Chowdhury, M. R. Rahman, R. Boutaba, ViNEYard: Virtual Network Embedding Algorithms with Coordinated Node and Link Mapping, IEEE/ACM Transactions on Networking (ToN), 20(1):206–219, 2012

- N. M. M. K. Chowdhury, R. Boutaba, A Survey of Network Virtualization, Computer Networks, 54(5):862–876, 2010
- N. M. M. K. Chowdhury, R. Boutaba, Network Virtualization: State of the Art and Research Challenges, IEEE Communications Magazine, 47(7):20–26, 2009
- D. T. Ahmed, N. M. M. K. Chowdhury, M. M. Akbar, Admission Control Algorithm for Multimedia Server: A Hybrid Approach, International Journal of Computers and Applications, 29(4):414–419, 2007

REFEREED WORKSHOP PAPERS

- **J. Lu**, Y. Xiao, S. Chakraborty, S. Fu, Y. Ji, A. Chen, M. Chowdhury, N. Rao, S. Ratnasamy, X. Wang, OpenInfra: A Co-simulation Framework for the Infrastructure Nexus, HotInfra 2024.
- **S. S. V. Singapuram**, **C. Hu**, **F. Lai**, **C. Zhang**, M. Chowdhury, Flamingo: A User-Centric System for Fast and Energy-Efficient DNN Training on Smartphones, DistributedML 2023.
- Y. Qiu, P. T. J. Kon, J. Xing, Y. Huang, H. Liu, X. Wang, P. Huang, M. Chowdhury, A. Chen, Cloudless Computing: Simplifying Cloud Management with Infrastructure Clarity, ACM HotNets 2023.
- **H. A. Maruf**, M. Chowdhury, Memory Disaggregation: Open Challenges in the Era of CXL, HotInfra 2023.
- **Z. Yang**, **L. Meng**, **J. Chung**, M. Chowdhury, Chasing Low-Carbon Electricity for Practical and Sustainable DNN Training, Climate Change AI 2023.
- T. Anderson, A. Belay, M. Chowdhury, A. Cidon, I. Zhang, Treehouse: A Case For Carbon-Aware Datacenter Software, HotCarbon 2022.
- **F. Lai**, **Y. Dai**, **X. Zhu**, H. Madhyastha, M. Chowdhury, FedScale: Benchmarking Model and System Performance of Federated Learning, ACM SOSP ResilientFL 2021. **Best Paper Award**
- H. Zhang, K. Chen, M. Chowdhury, Pas de Deux: Shape the Circuits, and Shape the Apps Too!, ACM APNet 2018.
- **F. Lai**, M. Chowdhury, H. Madhyastha, To Relay or Not to Relay for Inter-Cloud Transfers?, USENIX HotCloud 2018.
- A. P. Iyer, A. Panda, M. Chowdhury, A. Akella, S. Shenker, I. Stoica, Monarch: Gaining Command on Geo-Distributed Graph Analytics, USENIX HotCloud 2018.
- X. Sun, T. N. Le, M. Chowdhury, Z. Liu, Fair Allocation of Heterogeneous and Interchangeable Resources, ACM SIGMETRICS MAMA 2018.
- A. P. Iyer, A. Panda, S. Venkataraman, M. Chowdhury, A. Akella, S. Shenker, I. Stoica, Bridging the GAP: Towards Approximate Graph Analytics, ACM SIGMOD GRADES-NDA 2018. **Best Paper Award**
- **Y. Zhang**, **J. Gu**, Y. Lee, M. Chowdhury, K. G. Shin, Performance Isolation Anomalies in RDMA, ACM SIGCOMM KBNets 2017.
- L. Nguyen*, **P. Yu***, M. Chowdhury (*Equal contribution), No! Not Another Deep Learning Framework, ACM HotOS-XVI 2017.
- M. Chowdhury, I. Stoica, Coflow: A Networking Abstraction for Cluster Applications, ACM HotNets-XI 2012.
- G. Kumar, M. Chowdhury, S. Ratnasamy, I. Stoica, A Case for Performance-Centric Network Allocation, USENIX HotCloud 2012.
- M. Chowdhury, F. Samuel, R. Boutaba, PolyViNE: Policy-based Virtual Network Embedding Across Multiple Domains, ACM SIGCOMM VISA 2010.
- M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, Spark: Cluster Computing with Working Sets, USENIX HotCloud 2010.
- N. M. M. K. Chowdhury, M. M. Akbar, M. Kaykobad, DiskTrie: An Efficient Data Structure Using Flash Memory for Mobile Devices, WALCOM 2007.

INVITED PAPERS

- T. Anderson, A. Belay, M. Chowdhury, A. Cidon, I. Zhang, Treehouse: A Case For Carbon-Aware Datacenter Software, ACM SIGENERGY Energy Informatics Review, 3(3), 2023
- **H. A. Maruf**, M. Chowdhury, Memory Disaggregation: Advances and Open Challenges, ACM SIGOPS Operating Systems Review, 57(1):29–37, 2023
- **J. Gu**, Y. Lee, **Y. Zhang**, M. Chowdhury, K. G. Shin, Decentralized Memory Disaggregation Over Low-Latency Networks, USENIX ;login:, 42(4):42–48, 2017
- M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, I. Stoica, Fast and Interactive Analytics over Hadoop Data with Spark, USENIX ;login:, 37(4):45–51, 2012

NON-REFEREED TECHNICAL REPORTS/ARXIV PREPRINTS

- **I. Jang**, **R. Lu**, N. Bansal, A. Chen, M. Chowdhury, Cornstarch: Distributed Multimodal Training Must Be Multimodality-Aware , arXiv:2503.11367, March 2025
- P. T. J. Kon, **J. Liu**, Q. Ding, Y. Qiu, Z, Yang, Y. Huang, J. Srinivasa, M. Lee, M. Chowdhury, Ang Chen, Curie: Toward Rigorous and Automated Scientific Experimentation with AI Agents, arXiv:2502.16069, February 2025
- **S. He**, **I. Jang**, M. Chowdhury, Mordal: Automated Pretrained Model Selection for Vision Language Models, arXiv:2502.00241, February 2025
- **J. Lu***, **Y. Zhang***, H. A. Maruf, M. Park, **Y. Tang**, F. Lai, M. Chowdhury, Mercury: QoS-Aware Tiered Memory System, arXiv:2412.08938, December 2024
- **J. Chung**, N. Talati, M. Chowdhury, Toward Cross-Layer Energy Optimizations in AI Systems, arXiv:2404.06675, August 2024
- **J. Liu**, **Z. Wu**, **J. Chung**, F. Lai, M. Lee, M. Chowdhury, Andes: Defining and Enhancing Quality-of-Experience in LLM-Based Text Streaming Services, arXiv:2404.16283, April 2024
- **Y. Zhu**, **J. Liu**, M. Chowdhury, **F. Lai**, FedTrans: Efficient Federated Learning Over Heterogeneous Clients via Model Transformation, arXiv:2404.13515, April 2024
- **J. Chung**, M. Chowdhury, Toward Cross-Layer Energy Optimizations in Machine Learning Systems, arXiv:2404.06675, April 2024
- **J. Liu**, **F. Lai**, **D. Ding**, **Y. Zhang**, M. Chowdhury, Venn: Resource Management Across Federated Learning Jobs, arXiv:2312.08298, December 2023
- **J. Chung**, **Y. Gu**, **I. Jang**, **L. Meng**, N. Bansal, M. Chowdhury, Perseus: Removing Energy Bloat from Large Model Training, arXiv:2312.06902, December 2023
- Z. Wan, X. Wang, C. Liu, S. Alam, Y. Zheng, **J. Liu**, Z. Qu, S. Yan, Y. Zhu, Q. Zhang, M. Chowdhury, M. Zhang, Efficient Large Language Models: A Survey, arXiv:2312.03863, December 2023
- **I. Jang**, **Z. Yang**, Z. Zhang, X. Jin, M. Chowdhury, Oobleck: Resilient Distributed Training of Large Models Using Pipeline Templates, arXiv:2309.08125, September 2023
- **H. A. Maruf**, M. Chowdhury, Memory Disaggregation: Advances and Open Challenges, arXiv:2305.03943, May 2023
- **Z. Yang**, **L. Meng**, **J. Chung**, M. Chowdhury, Chasing Low-Carbon Electricity for Practical and Sustainable DNN Training, arXiv:2303.02508, March 2023
- E. Wang, A. Kannan, Y. Liang, B. Chen, M. Chowdhury, FLINT: A Platform for Federated Learning Integration, arXiv:2302.12862, February 2023
- P. Tholoniat, K. Kostopoulou, M. Chowdhury, A. Cidon, R. Geambasu, M. Lécuyer, Junfeng Yang, Packing Privacy Budget Efficiently, arXiv:2212.13228, December 2022
- **J. Liu**, **F. Lai**, **Y. Dai**, A. Akella, Harsha V. Madhyastha, M. Chowdhury, Auxo: Heterogeneity-Mitigating Federated Learning via Scalable Client Clustering, arXiv:2210.16656, October 2022

- **P. Yu**, **Y. Qiu**, X. Jin, M. Chowdhury, Orloj: Predictably Serving Unpredictable DNNs, arXiv:2209.00159, September 2022
- **J. You***, **J. Chung***, M. Chowdhury (*Equal contribution), Zeus: Understanding and Optimizing GPU Energy Consumption of DNN Training, arXiv:2208.06102, August 2022
- **S. S. V. Singapuram**, **F. Lai**, **C. Hu**, M. Chowdhury, Swan: A Neural Engine for Efficient DNN Training on Smartphone SoCs, arXiv:2206.04687, June 2022
- **H. A. Maruf**, H. Wang, A. Dhanotia, J. Weiner, N. Agarwal, P. Bhattacharya, C. Petersen, M. Chowdhury, S. Kanaujia, P. Chauhan, TPP: Transparent Page Placement for CXL-Enabled Tiered Memory, arXiv:2206.02878, June 2022
- **J. Gu**, M. Chowdhury, A. Akella, K. G. Shin, Elastic Model Aggregation with Parameter Service, arXiv:2204.03211, April 2022
- Y. Wang, D. Sun, K. Chen, **F. Lai**, M. Chowdhury, Efficient DNN Training with Knowledge-Guided Layer Freezing, arXiv:2201.06227, January 2022
- T. Anderson, A. Belay, M. Chowdhury, A. Cidon, I. Zhang, Treehouse: A Case For Carbon-Aware Datacenter Software, arXiv:2201.02120, January 2022
- R. Kontar, N. Shi, X. Yue, S. Chung, E. Byon, M. Chowdhury, J. Jin, W. Kontar, N. Masoud, M. Noueihed, C. E. Okwudire, G. Raskutti, R. Saigal, K. Singh, Z. Ye, The Internet of Federated Things (IoFT): A Vision for the Future and In-Depth Survey of Data-Driven Approaches for Federated Learning, arXiv:2111.05326, November 2021
- **H. A. Maruf**, Y. Zhong, H. Wang, M. Chowdhury, A. Cidon, C. Waldspurger, Memtrade: A Disaggregated-Memory Marketplace for Public Clouds, arXiv:2108.06893, August 2021
- N. Shi, **F. Lai**, R. A. Kontar, M. Chowdhury, Fed-ensemble: Improving Generalization through Model Ensembling in Federated Learning, arXiv:2107.10663, July 2021
- **F. Lai**, **Y. Dai**, **X. Zhu**, M. Chowdhury, FedScale: Benchmarking Model and System Performance of Federated Learning, arXiv:2105.11367, May 2021
- **F. Lai**, **X. Zhu**, H. V. Madhyastha, M. Chowdhury, Oort: Efficient Federated Learning via Guided Participant Selection, arXiv:2010.06081, May 2021
- Y. Lee, **H. A. Maruf**, A. Cidon, M. Chowdhury, K. G. Shin, Mitigating the Performance-Efficiency Tradeoff in Resilient Memory Disaggregation, arXiv:1910.09727, October 2020
- T. N. Le, X. Sun, M. Chowdhury, Z. Liu, BoPF: Mitigating the Burstiness-Fairness Tradeoff in Multi-Resource Clusters, arXiv:1912.03523, December 2019
- **H. A. Maruf**, M. Chowdhury, Effectively Prefetching Remote Memory with Leap, arXiv:1911.09829, November 2019
- M. Chowdhury, S. Khuller, M. Purohit, S. Yang, **J. You**, Near Optimal Coflow Scheduling in Networks, arXiv:1906.06851, June 2019
- **Y. Zhang**, **Y. Tan**, B. Stephens, M. Chowdhury, RDMA Performance Isolation With Justitia, arXiv:1905.04437, May 2019
- **J. You**, M. Chowdhury, Terra: Scalable Cross-Layer GDA Optimizations, arXiv:1904.08480, April 2019
- **P. Yu**, M. Chowdhury, Salus: Fine-Grained GPU Sharing Primitives for Deep Learning Applications, arXiv:1902.04610, February 2019
- A. P. Iyer, I. Stoica, M. Chowdhury, L. E. Li, Fast and Accurate Performance Analysis of LTE Radio Access Networks, arXiv:1605.04652, May 2016
- M. Chowdhury, R. Agarwal, V. Sekar, I. Stoica, A Longitudinal and Cross-Dataset Study of Internet Latency and Path Stability, UC Berkeley Technical Report UCB/EECS-2014-172, October 2014
- M. Chowdhury, I. Stoica, Coflow: An Application Layer Abstraction for Cluster Networking, UC Berkeley Technical Report UCB/EECS-2012-184, August 2012

- S. Agarwal, M. Chowdhury, D. Joseph, I. Stoica, Lattice: A Scalable Layer-Agnostic Packet Classification Framework, UC Berkeley Technical Report UCB/EECS-2011-96, August 2011
- M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, I. Stoica, Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, UC Berkeley Technical Report UCB/EECS-2011-82, July 2011
- M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, Spark: Cluster Computing with Working Sets, UC Berkeley Technical Report UCB/EECS-2010-53, May 2010

THESES

- N. M. M. K. Chowdhury, Coflow: A Networking Abstraction for Distributed Data-Parallel Applications, Doctoral Dissertation, University of California, Berkeley, 2015. **ACM SIGCOMM Doctoral Dissertation Award**
- N. M. M. K. Chowdhury, Identity Management and Resource Allocation in the Network Virtualization Environment, Master's Thesis, University of Waterloo, 2009. **School Nominee for the University of Waterloo Alumni Gold Medal**
- N. M. M. K. Chowdhury, A Study of the Hybrid Admission Control Algorithm for Multimedia Server, Bachelor's Thesis, Bangladesh University of Engineering and Technology, 2006

PATENT

- P. Bodik, I. Menache, P. Winkler, G. Foxman, N. M. M. K. Chowdhury, Management of Datacenters for Fault Tolerance and Bandwidth, Application # US 13/489,207, Microsoft Corporation, 2013

RESEARCH GRANTS & GIFTS

Total Raised $10.87M
PI Share $5.12M

FEDERAL GRANTS

	Source	Title	Role	Total
2021	NSF	Systems Support for Federated Learning	Lead PI	1.200M
2021	NSF & VMware	Foundations of Clean and Balanced Datacenters: TreeHouse	Co-PI	3.000M
2019	NSF	Towards Enabling Optimal Performance-Cost Tradeoffs in Distributed Storage	Co-PI	1.120M
2019	NSF	Multi-Scale GPU Resource Management for AI Applications	Single PI	463K
2019	NSF	CAREER: End-to-End Network Design for Unified Memory Disaggregation	Single PI	578K
2016	NSF	A Cross-Layer Approach Toward Low-Latency Data-Parallel Applications in Rack-Scale Computing	Lead PI	825K
2016	NSF	Enabling Application-Level Performance Predictability in Public Clouds	Lead PI	450K
2016	NSF	Enabling Flexible and High Performance Big Data Analytics Over Geo-Distributed Clouds	Co-PI	800K

OTHER GRANTS & GIFTS

	Source	Title	Role	Total
2024	Google	Large-Scale Systems for Machine Learning	Single PI	80K
2024	UM OVPR	Infrastructure Management Under the Lens of Computing	Co-PI	135K
2024	Cisco	Efficient Multi-Modal Inference	Single PI	150K
2024	Ford	Federated Learning for Connected Vehicle Energy Management	Single PI	200K
2024	Google	Gemma Academic Program	Single PI	5K
2024	Mozilla	Zeus: Deep Learning Energy Measurement and Optimization	Single PI	50K
2024	Salesforce	Gift	Single PI	20K
2023	UM MICDE	Resilient Distributed Training of Large Models	Single PI	75K
2023	Ford	Federated Learning for Connected Vehicle Energy Management	Single PI	25K
2023	Google	Fault-Tolerant Distributed Execution of Large DNNs	Single PI	60K
2022	Meta	Quality-of-Service (QoS) in Tiered Memory	Single PI	76K
2022	Meta	SmartNIC-Driven Scheduling in AI/ML Clusters	Single PI	50K
2022	UM CoE	Privacy-Preserving Machine Learning	Co-PI	45K
2022	Nvidia	Equipment Gift	Single PI	4K
2021	Cisco	MLOps for Edge AI	Single PI	328K
2021	UM CSE	Morris Wellman Faculty Development Professorship	Single PI	40K
2021	UM Biosciences	Enhancing Mechanisms of Human Resilience for Student Success and Well-Being	Co-PI	667K
2020	Google	Gift	Single PI	50K
2020	KLA	Gift	Single PI	50K
2019	VMware	Early Career Faculty Award	Single PI	50K
2018	Alibaba	InfiniFS: Microsecond-Latency I/O Over Disaggregated Memory	Single PI	100K
2018	Alibaba	Consus: Erasure-Coded Zettabyte Storage Across Geo-Distributed Datacenters	Lead PI	100K
2017	Chelsio	Equipment Gift	Single PI	7K
2016	Google	Application-Network Symbiosis in Geo-Distributed Analytics	Lead PI	55K

SELECTED PRESENTATIONS

PLENARY TALKS AND DISTINGUISHED LECTURES

- Plenary Speaker, Sustainable Cloud Computing and AI, Climate Week, Columbia University
  "Toward Energy-Optimal AI Systems", September 2024
- Colloquium Speaker, Computer and Information Science, UM Dearborn
  "Energy-Efficient AI Systems", December 2023
- Colloquium Speaker, Computer Science and Engineering, Pennsylvania State University
  "Data Science Without Data Collection", November 2022
- Colloquium Speaker, School of Computer Science, University of Windsor, Canada
  "Toward Practical Federated Learning", January 2022
- Keynote, The 2nd Workshop on Distributed Machine Learning (DistributedML'21)
  "Toward Practical Federated Learning", December 2021
- Keynote, The 2021 Google Workshop on Federated Learning and Analytics
  "Systems Support for Federated Computation", November 2021
- Keynote, The 9th IEEE International Conference on Cloud Networking (IEEE CloudNet 2020)
  "Practical Memory Disaggregation: A Case Study in Network-Informed Data Systems Design", November 2020
- Keynote, The 2nd Workshop on Hot Topics in Video Analytics and Intelligent Edges (HotEdgeVideo'20)
  "Multi-Scale GPU Resource Management for Deep Learning", August 2020

PANELS AND FIRESIDE CHATS

- Panelist, Sustainable Cloud Computing and AI, Climate Week, Columbia University
  "Sustainable AI", September 2024
- Panelist, ACES Mini-Symposium, MICDE, University of Michigan
  "Algorithms, AI, and Advanced Computer Architectures for Efficient Scientific Applications", October 2024

SELECTED RECENT TALKS

Toward Energy-Optimal AI Systems

- November 2024: University of Texas, Austin
- October 2024: Meta Open Innovation AI Research Community Workshop, London
- June 2024: Mozilla Technology Fund Summit, Amsterdam
- August 2023: Bangladesh University of Engineering and Technology

Practical Federated Learning // Data Science Without Data Collection

- July 2023: Simons Workshop on Federated & Collaborative Learning
- May 2023: Brave Software
- April 2023: OSU AIoT Seminar Series, Ohio State University
- November 2022: MIDAS Data Science and AI Summit, University of Michigan, Ann Arbor
- November 2022: Open Data Science Conference (ODSC West)
- October 2022: Qualcomm AI Research
- July 2022: IEEE LANMAN 2022
- July 2022: Microsoft Research, Redmond
- June 2022: Cisco Research Federated Learning Summit
- March 2022: LinkedIn

- December 2021: MLSys Seminar, Stanford University

Auxo: Heterogeneity-Mitigating Federated Learning via Scalable Client Clustering

- November 2022: Google Workshop on Federated Learning and Analytics

Multi-Scale GPU Resource Management

- April 2021: KLA Corporation
- March 2021: CS Graduate Seminar, KAUST

Practical Memory Disaggregation

- March 2021: MAKI Workshop on Emerging Use-Cases and Technologies for In-Network Computing
- September 2020: CSE Faculty Seminar, University of Michigan, Ann Arbor
- June 2020: BUET ACM Chapter Seminar, Bangladesh University of Engineering and Technology
- September 2019: Wayne State University
- August 2018: Microsoft Research Asia
- August 2018: Xidian University, Xi'an, China
- July 2018: TTIC Summer Workshop: Data Center Scheduling from Theory to Practice
- July 2018: Datacenter Networking Workshop, University of Wisconsin, Madison
- June 2018: Nokia Bell Labs
- November 2017: Rice University
- June 2017: NSF CSR PI Meeting, Orlando, FL
- December 2016: University of Southern California

Application-Infrastructure Symbiosis

- January 2018: Google
- May 2016: Bangladesh University of Engineering and Technology
- April 2016: The 60th HPC User Forum, Tucson, AZ

Recent Advances in Coflow-Based Networking

- June 2017: DIMACS Workshop on Algorithms for Data Center Networks
- November 2016: EECS 589 Guest Lecture, University of Michigan, Ann Arbor
- July 2016: Dagstuhl Seminar 16281 on Network Latency Control in Data Centres

PROFESSIONAL SERVICE

LEADERSHIP ROLES

Workshop Co-Chair

NSF NeTS PI Meeting	2023
NSF Workshop on Next Generation Cloud Research Infrastructure	2019
Second Asia-Pacific Workshop on Networking (APNet)	2018

Posters Co-Chair

USENIX NSDI	2017

Publicity Co-Chair

Third Asia-Pacific Workshop on Networking (APNet)	2019

Steering Committee Member

HotCarbon Workshop on Sustainable Computer Systems	2024

PROGRAM COMMITTEES

Conference Program Committee Member

USENIX OSDI	2021, 2023, 2025
USENIX NSDI	2017, 2019, 2021, 2025
ACM SIGCOMM	2017–2019, 2021–2022, 2025
MLSys	2023–2024
ACM CoNEXT	2016
IEEE ICDCS	2018, 2020
IEEE BigData	2019
NeurIPS Datasets and Benchmarks Track	2021
International Conference on Networking, Systems and Security (NSysS)	2017, 2019
International Conference on Computer and Information Technology (ICCIT)	2016

Workshop & Symposia Program Committee Member

Workshop on Sustainable Computer Systems (HotCarbon)	2022, 2024
Workshop on Resource Disaggregation (WORD)	2019
Asia-Pacific Workshop on Networking (APNet)	2017, 2019
ACM SIGCOMM Workshop on Kernel-Bypass Networks (KBNets)	2017
ACM SIGCOMM Posters and Demos	2016
ACM CoNEXT Student Workshop	2015

REFEREE SERVICE

Funding Agency Panelist

National Science Foundation (NSF)	2016, 2016, 2018, 2019, 2021, 2023, 2025
Department of Energy (DOE)	2024

UNIVERSITY SERVICE

Steering Committee, Institute for Energy Solutions, University of Michigan	2024–2025
GenAI Task Force Research Committee College of Engineering, University of Michigan	2024
CSE Graduate Student Organization Advising, EECS, University of Michigan	2023–2025
CSE Masters Advising, EECS, University of Michigan	2023–2025
CS-ENG Undergraduate Advising, EECS, University of Michigan	2016–2022
Distinguished Lecture Series / Seminar Series, EECS, University of Michigan	2019–2022
NextProf Nexus Workshop, College of Engineering, University of Michigan	2020–2021

K-12 Outreach, EECS, University of Michigan 2018–2019

CS Undergraduate Honors and Awards, EECS, University of Michigan 2017–2018

Graduate Admissions Committee, EECS, University of Michigan 2015–2017

TEACHING

LEAD INSTRUCTOR

University of Michigan

Term	Course #	Course Title	Class Size	Evaluation
Winter 2025	EECS 489	Computer Networks	143	N.NN/5.00
Fall 2024	CSE 585	Advanced Scalable Systems	54	4.80/5.00
Winter 2024	EECS 489	Computer Networks	138	4.30/5.00
Winter 2024	EECS 598	Systems for Generative AI	37	4.90/5.00
Fall 2021	EECS 489	Computer Networks	132	4.70/5.00
Winter 2021	EECS 598	Systems for AI	34	4.70/5.00
Fall 2020	EECS 489	Computer Networks	121	4.70/5.00
Winter 2020	EECS 598	Systems for AI	18	4.80/5.00
Fall 2019	EECS 489	Computer Networks	116	4.40/5.00
Winter 2019	EECS 598	Big Data Systems and Applications	15	4.80/5.00
Fall 2018	EECS 489	Computer Networks	103	4.30/5.00
Fall 2017	EECS 598	Big Data Systems and Applications	34	4.80/5.00
Winter 2017	EECS 489	Computer Networks	72	4.61/5.00
Fall 2016	EECS 582	Advanced Operating Systems	23	4.82/5.00
Winter 2016	EECS 582	Advanced Operating Systems	18	4.61/5.00
		Overall Average		**4.66/5.00**

ADVISING

CURRENT PH.D. STUDENTS

Jiachen Liu	2020–Present
Jae-Won Chung	2021–Present
Insu Jang	2021–Present
Shiqi He	2023–Present
Runyu Lu (w/ Ang Chen)	2024–Present
Ruofan Wu	2024–Present
Kevin Xue (w/ Ryan Huang)	2024–Present
Jeff Ma	2024–Present

GRADUATED PH.D. STUDENTS

		Dissertation Title	Year
1	**Juncheng Gu** (w/ Kang G. Shin)	Efficient Resource Management for Deep Learning Clusters	2021
2	**Peifeng Yu**	Application-Aware Scheduling in Deep Learning Software Stacks	2022
3	**Jie You**	Toward Practical Application-Aware Big Data Systems	2022
4	**Hasan Al Maruf**	Practical Memory Disaggregation	2023
5	**Fan Lai**	Minimalist Systems for Pervasive Machine Learning	2023
6	**Yiwen Zhang**	Quality of Service for Performance-Critical Cloud Applications	2024

MASTER'S AND UNDERGRADUATE ADVISEES

I have also worked with 60+ students at master's, undergrad, and high school levels on various research projects. 15 of them went on to join top-tier Ph.D. programs.