Linjian Ma

+1 217 979 7114 ⋄ lma16@illinois.edu ⋄ linjianma.github.io
github/Linkedin: linjianma

RESEARCH INTERESTS

Numerical analysis	numerical linear algebra, tensor decompositions, tensor networks, randomized algorithms, numerical optimizations
High performance computing	parallel algorithms, communication-avoiding algorithms, scalable mathematical systems
Machine learning	recommendation systems, large language models

EMPLOYMENT

- **Meta Platforms** September 2023 -
  Staff Research Scientist at Facebook MRS
  Topic: *Recommendation systems foundation model co-design*

- **Meta Platforms** May 2022 - August 2022
  Software Engineer Intern at PyTorch Distributed
  Topic: *Improved auto wrapping policy for PyTorch Fully Sharded Data Parallel (FSDP)*
  - Implemented a new FSDP wrapping policy based on the parameter execution ordering
  - Integrated a compiler based tracing technique from torch.fx module in FSDP
  - Up to 65% speed-ups compared to existing wrapping policies on both vision and NLP models with 8 to 175 billion parameters

- **Center for Computational Quantum Physics, Flatiron Institute** June 2021 - August 2021
  Research Associate, Advisor: *Miles Stoudenmire* and *Matthew Fishman*
  Topic: *Automatic differentiation systems for tensor networks*

- **Lawrence Berkeley National Laboratory** May 2020 - August 2020
  Research Intern, Advisor: *Chao Yang*
  Topic: *Low-rank approximation in simulations of quantum algorithms*

- **Wave Computing & Berkeley AI Research (BAIR)** May 2019 - August 2019
  Machine Learning Intern
  Topic: *Compressing large scale neural networks based on second-order information*
  - Applied mixed-precision quantization on BERT guided by second order information
  - Proposed a new quantization scheme, named group-wise quantization, to alleviate accuracy degradation
  - Investigated the bottlenecks in BERT quantization

EDUCATION BACKGROUNDS

University of Illinois Urbana-Champaign August 2019 - August 2023
PhD, Computer Science, Advisor: *Edgar Solomonik* GPA: 3.97/4.0
Area: Scientific Computing
Thesis: *Towards efficient algorithms and systems for tensor decompositions and tensor networks*

University of California, Berkeley August 2018 - May 2019
MEng, Computer Science, Advisor: *Michael Mahoney* Major GPA: 3.94/4.0
Track: Data Science & Systems
Capstone project: *Second-order optimization of neural network learning*

University of Illinois at Urbana-Champaign August 2015 - May 2018
MS, Mechanical Engineering, Advisor: *N.R. Aluru* GPA: 3.97/4.0
Concentration: Computational Science and Engineering
Thesis: *A multiscale model for the oxide ion conducting and proton conducting solid oxide cells*

Zhejiang University August 2012 - June 2016
BE, Energy Engineering, Advisor: *Tao Wang and Zhongyang Luo* GPA: 3.95/4.0
Graduate with Honors, Chu Kochen Honors College

HONORS AND AWARDS

Mavis Future Faculty Fellow, UIUC 2021-2022

Kenichi Miura Award, UIUC 2021

Student Travel Award, SIAM-CSE21, SIAM-LA21, NeurIPS 22 2021-2022

Kuck Computational Science & Engineering Scholarship, UIUC 2020

Computer Science Gene Golub Fellowship, UIUC 2019

Graduate with Honor, ZJU 2016

Meritorious Winner, Mathematical Contest In Modeling (MCM) 2015

National Scholarship for Undergraduate, ZJU 2014

The First Class Scholarship for Outstanding Students, ZJU 2013-2014

The First Prize in China Undergraduates Mathematical Contest 2013

PRESENTATIONS

First author presentations Workshop on Sparse Tensor Computations,
NeurIPS 2022, CUNY quantum computing and tensor network symposium,
SIAM'PP 2022, SIAM'LA 2021, IPDPS 2021,
SIAM'CSE 2021, PACT 2020,
SIAM'PP 2020, Berkeley'SCseminar 2019, USNCCM 2017

Posters NeurIPS 2021, SIAM'PP 2020, AAAI 2020

PUBLICATIONS

[1] **Linjian Ma**, Matthew Fishman, Miles Stoudenmire, Edgar Solomonik, Approximate Contraction of Arbitrary Tensor Networks with a Flexible and Efficient Density Matrix Algorithm, *Quantum*, 2024. [link]

[2] Louis Schatzki, **Linjian Ma**, Edgar Solomonik, and Eric Chitambar, Tensor Rank and Other Multipartite Entanglement Measures of Graph States, *Physics Review A,* 2024. [link]

[3] **Linjian Ma** and Edgar Solomonik, Cost-efficient Gaussian Tensor Network Embeddings for Tensor-structured Inputs, *Conference on Neural Information Processing Systems (NeurIPS'22)*, 2022. [link]

[4] **Linjian Ma** and Chao Yang, Low Rank Approximation in Simulations of Quantum Algorithms, *Journal of Computational Science,* 2022. [link]

[5] **Linjian Ma** and Edgar Solomonik, Accelerating Alternating Least Squares for Tensor Decomposition by Pairwise Perturbation, *Numerical Linear Algebra with Applications (NLA)*, 2022. [link]

[6] **Linjian Ma** and Edgar Solomonik, Fast and Accurate Randomized Algorithms for Low-rank Tensor Decompositions, *Conference on Neural Information Processing Systems (NeurIPS'21)*, 2021. [link]

[7] Navjot Singh, **Linjian Ma**, Hongru Yang, and Edgar Solomonik, Comparison of Accuracy and Scalability of Gauss-Newton and Alternating Least Squares for CP Decomposition, *SIAM Journal on Scientific Computing (SISC),* 2021. [link]

[8] **Linjian Ma** and Edgar Solomonik, Efficient Parallel CP Decomposition with Pairwise Perturbation and Multi-sweep Dimension Tree, *International Parallel and Distributed Processing Symposium (IPDPS'21)*, 2021. [link]

[9] **Linjian Ma***, Jiayu Ye*, and Edgar Solomonik, AutoHOOT: Automatic High-Order Optimization for Tensors, *International Conference on Parallel Architectures and Compilation Techniques (PACT'20)*, 2020. [link]

[10] Sheng Shen, Zhen Dong, Jiayu Ye, **Linjian Ma**, Zhewei Yao, Amir Gholami, Michael W. Mahoney, and Kurt Keutzer, Q-BERT: Hessian Based Ultra Low Precision Quantization of BERT, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'20),* 2020. [link]

[11] **Linjian Ma***, Gabe Montague*, Jiayu Ye*, Zhewei Yao, Amir Gholami, Kurt Keutzer, and Michael W. Mahoney, Inefficiency of K-FAC for Large Batch Size Training, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'20),* 2020. [link]

[12] **Linjian Ma**, Pikee Priya, and N. R. Aluru, A Multiscale Model for Electrochemical Reactions in LSCF Based Solid Oxide Cells, *Journal of the Electrochemical Society*, 2018. [link]

PREPRINTS AND TECHNICAL REPORTS

[1] Zhenan Shao, **Linjian Ma**, Bo Li, Diane Beck, Leveraging the Human Ventral Visual Stream to Improve Neural Network Robustness, *arXiv:2405.02564*, 2024. [link]

[2] **Linjian Ma**, Towards Efficient Algorithms and Systems for Tensor Decompositions and Tensor Networks, *PhD thesis, University of Illinois Urbana-Champaign*, 2023. [link]

[3] Zhewei Yao, **Linjian Ma**, Sheng Shen, Kurt Keutzer, and Michael W. Mahoney, MLPruning: A Multilevel Structured Pruning Framework for Transformer-based Models, *arXiv:2105.14636,* 2021. [link]

[4] **Linjian Ma**, A Multiscale Model for the Oxide Ion Conducting and Proton Conducting Solid Oxide Cells, *MS thesis, University of Illinois Urbana-Champaign*, 2018. [link]

SERVICES

Teaching Assistant	CS 450 Numerical Analysis (Fall 2020)
	CS 554 Parallel Numerical Algorithms (Fall 2021)
Reviewer	SuperComputing 2023, SPAA 2023, SDM 2024,
	ICML 2022-2023, NeurIPS 2022-2023, ICLR 2024, LoG 2022-2023,
	AISTATS 2023-2024, IJCAI 2023, KDD 2023, ACM-TOMS,
	SIAM Journal on Matrix Analysis and Applications (SIMAX),
	International Journal of Data Science and Analytics,
	Transactions on Machine Learning Research (TMLR), IEEE-TPAMI
	Numerical Linear Algebra with Applications

SKILLS

Programming Languages	Python, C/C++, Julia, Go, Matlab, CUDA
ML Frameworks	Pytorch, TensorFlow