HEXU ZHAO

+1 (929) 689-7240 ◇ New York, NY
hz3496@nyu.edu ◇ Homepage ◇ LinkedIn

EDUCATION

Courant Institute, New York University 2023 – Present
Ph.D. in Machine Learning Systems, advised by Prof. Jinyang Li and Prof. Aurojit Panda

Tsinghua University (Honored Yao Class) 2019 – 2023
Bachelor of Computer Science

RESEARCH INTERESTS

My research focuses on machine learning systems, with particular emphasis on systems for Gaussian Splatting. From a systems perspective, I leverage both distributed computation and GPU kernel optimization techniques to enable scalable training and rendering. From a workload perspective, I develop optimizations for both static and dynamic scene reconstruction at unprecedented scales.

PEER-REVIEWED PUBLICATIONS

CLM: Removing the GPU Memory Barrier for 3D Gaussian Splatting
Hexu Zhao*, Xiwen Min*, Xiaoteng Liu, Moonjun Gong, Yiming Li, Ang Li, Saining Xie, Jinyang Li, Aurojit Panda
ASPLOS 2026 (* Equal Contribution)

On Scaling Up 3D Gaussian Splatting Training
Hexu Zhao, Haoyang Weng, Daohan Lu, Ang Li, Jinyang Li, Aurojit Panda, Saining Xie
ICLR 2025 Oral (Top 0.5%) | GitHub (600+ stars)

On Optimizing the Communication of Model Parallelism
Yonghao Zhuang*, Hexu Zhao*, Lianmin Zheng, Zhuohan Li, Eric P. Xing, Qirong Ho, Joseph E. Gonzalez, Ion Stoica, Hao Zhang
MLSys 2023 (* Equal Contribution)

Fully Hyperbolic Neural Networks
Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, Jie Zhou
ACL 2022

MANUSCRIPTS UNDER REVIEW

Scaling Point-based Differentiable Rendering for Large Scale 3D Reconstruction
Hexu Zhao, Xiaoteng Liu, Xiwen Min, Jianhao Huang, Youming Deng, Yanfei Li, Ang Li, Jinyang Li, Aurojit Panda
Under Review

Development of a Doctor-in-the-loop Interpretation Framework for Insulin Titration in Diabetes
Haowei He*, Zhen Ying*, Biao Li, Yujuan Fan, Ping Wang, Jiaping Lu, Liming Wu, Hexu Zhao, Xiaoying Li, Yang Yuan, Ying Chen
Under Review (* Equal Contribution)

ACADEMIC SERVICE

Reviewer: IEEE Transactions on Parallel and Distributed Systems (TPDS) 2024

SELECTED EXPERIENCE

Research Scientist Intern, NVIDIA Spatial Intelligence Lab May 2025 – Now
Advisors: Francis Williams, Ken Museth

- Developing scalable point transformer system for processing large-scale 3D perception and generation tasks.

Research Scientist Intern, Microsoft DeepSpeed Team May 2024 – Aug 2024
Advisors: Guanhua Wang, Olatunji Ruwase

- Fused GEMM and collective communication into a single kernel for fine-grained overlapping at GPU warp-level using tile-based GEMM and GPU-initiated communication (IPC, NVSHMEM).
- Achieved up to 19% speedup over non-fused baseline after kernel-level optimization.

Research Assistant, MBZUAI Feb 2022 – Oct 2022
Advisors: Prof. Eric Xing, Prof. Hao Zhang, and Lianmin Zheng

- Contributed to the `Alpa` distributed ML system project.
- Designed a faster cross-mesh resharding communication pattern, achieving 10x speedup in synthetic benchmarks and 10–50% throughput improvement for GPT-3 and U-Transformer training.

Investment Analyst Intern, ZhenFund Venture Capital Jun 2021 – Sep 2021
Mentor: Emma Yin (Partner of ZhenFund)

- Performed due diligence and market analysis on early-stage AI startups to support investment decisions.
- Sourced founders from Tsinghua University research labs; organized entrepreneurship networking events to foster community engagement.

AWARDS AND HONORS

Henry Mitchell MacCracken PhD Fellowship, New York University Aug 2023 – Present
Gold Medal, China National Olympiad in Informatics (NOI) Aug 2018
Self-taught and became the first NOI gold medalist from Zibo, Shandong.

ACADEMIC REFEREES

Prof. Jinyang Li jinyang@cs.nyu.edu
Prof. Aurojit Panda apanda@cs.nyu.edu