

Yi Liu

1156 High Street, Santa Cruz, CA, United States, 95064

 yliu634@ucsc.edu

 +1(831)266-5469

 <https://yliu634.github.io>

EDUCATION

University of California Santa Cruz (UCSC)

Ph.D. program on Computer Science Engineering, GPA: 3.84/4.0

Department of Computer Science Engineering

Santa Cruz, CA

Sep. 2020 - Present

University of Science and Technology of China (USTC)

B.E. in Electrical Information Engineering

Department of Electronic Information and Communication Engineering

Hefei, Anhui, China

Sep. 2016 - July 2020

EMPLOYMENT

Baidu Inc.

AI networks R&D Intern

Beijing, China

June 2025 - Sep. 2025

- Worked on LLM deployment using the SGLang engine. Contributed to developing attention-FFN disaggregation techniques for LLM serving. Gained experience with popular LLM models (e.g., DeepSeek) deployment and collective communication mechanisms.

Amazon Web Services Inc. (AWS)

Software Development Engineering Intern

Sunnyvale, CA

June 2023 - Sep. 2023

SKILLS

- Programming Languages:** C/C++, Java, Python, MATLAB, SQL.
- Tools/Framework:** SGLang, DeepEP Low_latency, RDMA, P4 DataPlane Programming, LevelDB, RocksDB, FatCache, Redis, Ceph, MySQL.

SELECTED PUBLICATIONS

Efficient Vector Search on Disaggregated Memory with d-HNSW

Y. Liu*, F. Fang*, C. Qian. (*Equal contribution)

ACM Workshop on Hot Topics in Storage and File Systems (**HotStorage 2025**).

Outback: Fast and Communication-efficient Index for Key-Value Store on Disaggregated Memory.

Y. Liu, M. Xie, S. Shi, H. Litz, C. Qian.

International Conference on Very Large Data Bases (**VLDB 2025**).

Smash: Flexible, Fast, and Resource-efficient Placement and Lookup of Distributed Storage.

Y. Liu, S. Shi, M. Xie, H. Litz, C. Qian.

Proceedings of the ACM on Measurement and Analysis of Computing Systems (**SIGMETRICS 2023**).

Scalable, Fast, and Low-memory Table Lookups for Network Applications with one CRC-8.

Y. Liu, S. Shi, R. Zhou, Y. Gan, C. Qian.

IEEE International Conference on Network Protocols (**ICNP 2024**).

PROJECTS

Efficient Vector Search on Disaggregated Memory with d-HNSW.

Santa Cruz, CA

University of California Santa Cruz

Sep. 2024 - April. 2025

- d-HNSW is a disaggregated vector search engine for RDMA-based remote memory systems. It disaggregates the HNSW index and reduces network overhead through cached lightweight indexes, RDMA-friendly data layouts, and batched data loading. Experiments show up to 117× lower latency with 0.87 recall on SIFT1M@1.

Outback: Fast indexing for remote data accessing with RDMA.

Santa Cruz, CA

University of California Santa Cruz

Sep. 2023 - Oct. 2024

- Outback aims to achieve high throughput on remote KV stores with disaggregated memory systems. We designed a compact hashing algorithm that makes Outback serve as a communication-efficient indexing framework for fast data access. Our scheme achieves 1.05-5.03× throughput compared with state-of-the-art baselines.