WENLONG WANG

+1(612)-978-1482 wang9467@umn.edu

EDUCATION

Ph.D. in Computer Science and Engineering 2019 - 2026(Expected)
University of Minnesota, Twin Cities. Supervised by Professor David Hung-Chang Du. GPA: 3.76/4.0

BS in Computer Science and Engineering 2015 - 2019
Hong Kong University of Science and Technology. First Honor. GPA: 3.67/4.3

WORKING EXPERIENCE

Research Intern | **Futurewei Storage Lab**, *Santa Clara, CA* June 2022 - Aug 2022
Mentors: Dr. Chun Liu and Dr. Nelson Liao

- Researched updatable Learned Indexes, identifying performance impacts of order maintenance in leaf nodes and resolving slow performance in Single-Producer-Multi-Consumer (SPMC) situations.
- Designed and implemented a bucketed approach for leaf nodes, improving efficiency by up to 2.05x over existing works and laying the groundwork for a future publication.
- Extended the established framework of Learned Indexes to the string key index domain, broadening its applicability.

Research Intern | **Alibaba Group (U.S.) Inc**, *Sunnyvale, CA* June 2021 - Aug 2021
Mentor: Dr. Sheng Qiu

- Researched optimizing the Flash Translation Layer (FTL) mapping on open-channel SSD during garbage collection in log-structured files (e.g., blob files in RocksDB) to reduce data reorganization overhead in physical space.
- Constructed the open-channel SSD benchmarking environment using FEMU and benchmarked the initial performance using BlobDB(RocksDB), which was used to guide future optimizations.

RESEARCH INTEREST

Distributed/Disaggregated data infrastructure, Machine learning with systems, Large-scale KV stores, Learned indexes, Emerging storage technologies (NVM, ZNS, hybrid SSD, DNA storage), LLM serving/caching optimization.

PUBLICATIONS AND RESEARCH PROJECTS

LearnedKV: Integrating LSM and Learned Index for Superior Performance on SSD

- Designed and implemented a large-scale, novel tiered key-value store (over 4K lines of C++, excluding RocksDB) that seamlessly integrates a Log-Structured Merge (LSM) tree with a Learned Index on storage, including a non-blocking conversion mechanism.
- Benchmarked our implementation with YCSB workloads and the results show that our work outperforms state-of-the-art solutions by up to 1.32x in read requests and 1.31x in write performance.
- **Publication (First author)**: Preprint: https://arxiv.org/pdf/2406.18892. Full paper to be submitted to SIGMOD 2025.

BLI: a High-performance Bucket Learned Index with Concurrency support

- Designed and implemented an updatable in-memory learned index that adopts a "globally sorted, locally unsorted" approach by replacing linear sorted arrays with unsorted Buckets and optimized tree height (with 6K lines of C++).
- Benchmarked our implementation with SOSD workloads and the results show that our work achieves up to 4.42x better throughput than state-of-the-art learned indexes, with even more gains under multi-threaded conditions.
- **Publication (Second author)**: Poster selected in 2023 UMN CS&E Research Showcase. Full paper to be submitted to SIGMOD 2025.

VL-DNA: Enhancing DNA Storage Capacity with Variable Payload (Strand) Lengths

- Proposed and implemented a new scheme that used variable strand length to split primer-payload collisions and further enhanced the capacity of single tube DNA storage ranging from 18.27% to 19x (with 7K lines of C++).
- **Publication (Second author)**: Preprint: https://arxiv.org/pdf/2403.14204. Selected in DAC 2023 WIP poster session.

ACADEMIC SERVICES AND AWARDS

- OSDI/ATC 2024 Artifact Evaluation Committee (AEC).
- MSST 2024 Reviewer.
- ATC 2024 Student Grant.

SKILLS

- Programming Languages: C/C++, Python, JavaScript, HTML, Matlab
- Others: Linux/Unix, RocksDB, LevelDB, FEMU, HDFS, Ray, Git, Data Structures and Algorithms