technische universität dortmund

computer science 12

Bachelor/Master Thesis

Exploring the Trade-offs between Accuracy and Reliability in In-Storage Vector Search

Yun-Chih Chen
Prof. Dr. Jian-Jia Chen

Otto-Hahn Str. 16
Technische Universität Dortmund
Email: yunchih.chen@tu-dortmund.de
January 24, 2024

Vector search is an essential algorithm in modern search engines, recommendation systems, and retrieval-augmented large language models [1]. Managing billions of vectors requires a large vector index, which is usually stored on disk and loaded into memory on demand [2]. However, due to their wide and scattered access patterns, these systems have high disk I/O demands. With the introduction of in-storage computing, it's now possible to integrate specialized circuits into flash memory chips to perform vector distance computations directly at the location where the vectors are stored.

However, data stored in flash memory can have errors, and while traditional storage demands perfect accuracy, vector search systems can tolerate some inaccuracies. Your role in this thesis is to explore the balance between accuracy and reliability in these systems, particularly focusing on how data errors impact the precision of search results.

In this thesis, your mission is to investigate and optimize the trade-offs in vector search accuracy and reliability. You'll dive into the world of in-storage computing, understand how it works with NAND flash memory, and explore the state-of-the-art vector index structures like HNSW.

You will learn the fundamentals of NAND-flash-based in-storage computing and reliability models. Based on this knowledge, you will investigate the balance of accuracy and reliability when performing vector search in NAND flash. Finally, you will use simulations to validate your findings and theoretical models.

Required Skills:

- Foundation in probability and statistical analysis.
- Analytical skills in algorithm complexity analysis.
- Experiences in C++ or Python programming for implementing and testing algorithms

Acquired Skills after the thesis:

- Hands-on experience with sophisticated data structures designed for efficient vector search, including their design, implementation, and optimization techniques.
- Deep knowledge of in-storage computing, focusing on NAND-flash NVM systems.
- Skills to connect theoretical models with real-world system performance, especially in systems where memory errors are common.

References:

[1] Jang et al., CXL-ANNS: Software-Hardware Collaborative Memory Disaggregation and Computation for Billion-Scale Approximate Nearest Neighbor Search, ATC 2023

[2] Subramanya et al., DiskANN: Vector Search for Web Scale Search and Recommendation, NeurIPS 2019

[3] Wang et al., In-Storage Acceleration of Graph-Traversal-Based Approximate Nearest Neighbor Search, ArXiv 2023