Liang Geng

Curriculum Vitae

geng.161@osu.edu — Google Scholar — Linkedin

EDUCATION

Ph.D. Candidate
Jan, 2022 - Dec, 2026 (Expected)
The Ohio State University, USA
Computer Science and Engineering

Master of Engineering
Sep, 2016 - Jan, 2019
Northeastern University, China
Computer Science and Engineering

Bachelor of Engineering
Sep, 2012 - June, 2016
Liaoning Technical University, China
Software Engineering

EMPLOYMENT

Graduate Research Associate
Jan 2022 - Dec 2025
The Ohio State University, Computer Science and Engineering Department

Software Engineer Intern
Jun 2025 - Aug 2025
Wherobots, Spatial Database Group

- Developed a GPU-accelerated spatial join library based on my previous research projects [3, 2] that exploit NVIDIA RT cores to accelerate the filtering stage of spatial queries.
- Designed and implemented a GPU-accelerated RelateEngine to support diverse geometries and combinatorial predicates.

Senior Development Engineer (Full Time)
Jun 2020 - Dec 2021
Alibaba Group, DAMO Academy

- Researched and developed state-of-the-art graph computing systems.
- Maintained and optimized major open-source projects, including libgrape-lite and GraphScope.
- Directed the implementation and performance evaluation of key internal research initiatives.

Software Engineer (Full Time)
Jan 2019 - May 2020
Chehaoduo Group, Data Platform Division

- Engineered and optimized toolchains for enterprise data warehouses, including robust SQL parsers and ETL job management systems.

Visiting Research Associate
Dec 2017 - Sep 2018
The Ohio State University, Computer Science and Engineering Department

- Architected and developed SEP-Graph [11], a highly efficient hybrid graph computing framework.

REPRESENTATIVE PROJECTS

RayBooster: RT-Accelerated Spatial Query Engine
Jun 2025 - Aug 2025
Wherobots, Inc

- Architected and integrated the **first Ray Tracing (RT) core acceleration layer** into Apache SedonaDB, optimizing spatial join performance for production-grade geospatial databases.
- Engineered a GPU-optimized Structure of Arrays (SoA) storage layout to replace Well-Known Binary (WKB) formats, enabling high-performance random access and memory throughput.
- Developed a **Monolithic Bounding Volume Hierarchy (BVH)** indexing strategy that encodes geometry IDs into the Z-axis, eliminating the scalability overhead of managing millions of micro-indexes.
- Designed **RelateEngine**, a unified topological descriptor utilizing RT cores to compute the *DE-9IM model*, supporting complex combinatorial spatial predicates across diverse geometry types
- Implemented a memory-aware execution scheduler to handle massive datasets, providing robust resource management and preventing Out-of-Memory failures.
- Delivered a **5.8**× speedup on the *SpatialBench* suite and reduced operational cloud costs by **58.1%** by repurposing idle GPU RT units for spatial analytics.

X-HD: RT-Accelerated Hausdorff Distance Computation[1]
Jan 2025 - Dec 2025
The Ohio State University

- Developed a general-purpose Hausdorff Distance (HD) algorithm leveraging NVIDIA Ray Tracing (RT) cores to accelerate spatial data analysis for large-scale datasets.
- Engineered a custom grid-partitioned Bounding Volume Hierarchy (BVH) traversal strategy, reducing tree traversal intensity and optimizing hardware resource utilization.
- Implemented HD Estimators to prune non-contributing data points, significantly reducing redundant computations during the k-NN search phase.
- Optimized hardware throughput by developing a hybrid scheduling model, selectively offloading distance computations from RT shaders to CUDA kernels to resolve load imbalances.
- Achieved a **5.3**× speedup over the industry-standard *ITK* library and up to **6.4**× faster performance compared to existing GPU-optimized solutions.

LibRTS: A High-Performance RT-Accelerated Spatial Index Library [2]
May 2023 - Aug 2024
The Ohio State University

- Designed and implemented a **general-purpose spatial index library** that repurposes **NVIDIA Ray Tracing (RT) cores** to accelerate non-rendering workloads like point and range queries.
- Developed a **mutable indexing framework** capable of adapting to dynamic datasets, bridging the gap between specialized RT hardware and general spatial data processing.
- Formulated complex spatial queries into **hardware-native BVH traversal problems**, abstracting the specialized RT programming model into a developer-friendly API.
- Engineered a **load-balancing mechanism** to optimize GPU utilization, ensuring high throughput across diverse query types (point-in-polygon, contains, and intersects).
- Outperformed industry-standard CPU and GPU spatial libraries by up to **93.2**× **for range queries** and **11.7**× **for point queries**.

RayJoin: Real-Time RT-Accelerated Spatial Join Framework [3]
Aug 2022 - Jan 2024
The Ohio State University

- Developed **RayJoin**, a high-performance framework utilizing **NVIDIA RT Cores** to achieve unprecedented real-time speeds for Line Segment Intersection (LSI) and Point-in-Polygon (PIP) queries.
- Engineered a **high-precision spatial join framework** that overcomes hardware-level floating-point limitations, ensuring rigorous analytical accuracy despite underlying hardware constraints.
- Optimized BVH (Bounding Volume Hierarchy) construction pipelines, significantly reducing index buildup latency while maintaining peak query execution performance.
- Resolved critical bottlenecks in conventional plane-sweeping and tree-based algorithms, enabling the processing of **millions of polygons in under 460ms**.
- Achieved performance speedups ranging from **3.0× to 28.3×** over state-of-the-art GPU-optimized methods, establishing a new benchmark for real-time spatial analytics.

RR-Compound: RDMA-Enabled gRPC Framework [4]
Jan 2022 - May 2023
The Ohio State University

- Engineered **RR-Compound**, a high-performance RPC framework that integrates **Remote Direct Memory Access (RDMA)** as an internal transport for **gRPC**, achieving sub-millisecond latencies for data center workloads.
- Designed **BPEV (Busy-Pooling + Event)**, a novel connection management method that decouples RDMA resource handling from gRPC internals, enhancing scalability across massive connection counts.
- Optimized the gRPC asynchronous framework to operate within constrained RDMA-registered memory, balancing high-throughput buffer management with strict memory-space utilization.
- Designed the system as a **seamless replacement** for standard gRPC, requiring zero code changes for existing production applications (e.g., TensorFlow, KV-Stores).
- Outperformed the state-of-the-art *mRPC* in micro-benchmarks, delivering a **14.77% increase in throughput** and a **42.55% reduction in latency**.
- Validated performance in real-world deployments, achieving a **2.35×** throughput boost for *TensorFlow* and *KV-Store* applications compared to standard gRPC over IPoIB.

GPU support for libgrape-lite
Nov 2020 - Dec 2021
Alibaba Group

- Engineered a **GPU-acceleration layer** for *libgrape-lite*, Alibaba's high-performance C++ library for distributed parallel graph processing.
- Designed and implemented **in-device-memory graph representations**, enabling large-scale graph structures to be maintained and processed directly in GPU VRAM for reduced latency.
- Developed **state-of-the-art load balancing strategies** to optimize GPU throughput for irregular graph topologies, mitigating thread divergence and hardware underutilization.
- Integrated **multi-GPU communication** protocols utilizing **NVIDIA NCCL**, enabling efficient data synchronization and scaling across multi-node GPU clusters.

SEP-Graph: Hybrid graph computing method [11]
Jan 2018 - Sep 2018
The Ohio State University

- Developed **SEP-Graph**, a high-performance framework that optimizes graph processing by automatically switching between execution modes (**Sync/Async**), communication mechanisms (**Push/Pull**), and traversal schemes (**Data-driven/Topology-driven**).

- Engineered a **shortest-path execution model** that dynamically reconfigures algorithm parameters per iteration to minimize execution time based on real-time graph characteristics.
- Implemented hardware-specific optimizations for NVIDIA Pascal and Volta architectures, addressing thread divergence and memory coalescing issues inherent in irregular graph workloads.
- Achieved significant performance gains over state-of-the-art frameworks, reducing execution time by up to **45.8× compared to Groute** and **39.4× compared to Gunrock**.
- Validated system scalability and efficiency across multiple GPU generations (GTX 1080, P100, and V100) using diverse graph datasets.

PUBLICATIONS

[1] **Geng, Liang** and Yuan, Zhehu and Lee, Rubao and Wang, Fusheng and Zhang, Xiaodong. "X-HD: Fast Hausdorff Distance Computation with Ray Tracing". In: *Proceedings of the ACM International Conference on Supercomputing 2026 (to appear)*. 2026.

[2] **Geng, Liang** and Lee, Rubao and Zhang, Xiaodong. "LibRTS: A Spatial Indexing Library by Ray Tracing". In: *Proceedings of the 30th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming*. 2025.

[3] **Geng, Liang** and Lee, Rubao and Zhang, Xiaodong. "RayJoin: Fast and Precise Spatial Join". In: *Proceedings of the ACM International Conference on Supercomputing*. 2024.

[4] **Geng, Liang** and Wang, Hao and Meng, Jingsong and Fan, Dayi and Ben-Romdhane, Sami and Pichumani, Hari Kadayam and Phegade, Vinay and Zhang, Xiaodong. "RR-Compound: RDMA-Fused gRPC for Low Latency and High Throughput With an Easy Interface". In: *IEEE Transactions on Parallel and Distributed Systems* (2024).

[5] Gong, Shufeng and Tian, Chao and Yin, Qiang and Wang, Zhengdong and Yu, Song and Zhang, Yanfeng and Yu, Wenyuan and **Geng, Liang** and Fu, Chong and Yu, Ge and Zhou, Jingren. "Ingress: an automated incremental graph processing system". In: *The VLDB Journal* (2024).

[6] Meng, Ke and **Geng, Liang** and Li, Xue and Tao, Qian and Yu, Wenyuan and Zhou, Jingren. "Efficient Multi-GPU Graph Processing with Remote Work Stealing". In: *2023 IEEE 39th International Conference on Data Engineering*. 2023.

[7] Fan, Wenfei and **Geng, Liang** and Jin, Ruochun and Lu, Ping and Tugay, Resul and Yu, Wenyuan. "Linking Entities across Relations and Graphs". In: *2022 IEEE 38th International Conference on Data Engineering*. 2022.

[8] Xiao, Mengbai and Wang, Hao and **Geng, Liang** and Lee, Rubao and Zhang, Xiaodong. "An RDMA-enabled In-memory Computing Platform for R-tree on Clusters". In: *ACM Transactions on Spatial Algorithms and Systems* (2022).

[9] Gong, Shufeng and Tian, Chao and Yin, Qiang and Yu, Wenyuan and Zhang, Yanfeng and **Geng, Liang** and Yu, Song and Yu, Ge and Zhou, Jingren. "Automating incremental graph processing with flexible memoization". In: *Proceedings of the VLDB Endowment* (2021).

[10] Wang, Qiange and Zhang, Yanfeng and Wang, Hao and **Geng, Liang** and Lee, Rubao and Zhang, Xiaodong and Yu, Ge. "Automating incremental and asynchronous evaluation for recursive aggregate data processing". In: *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2020.

[11] Wang, Hao and **Geng, Liang** and Lee, Rubao and Hou, Kaixi and Zhang, Yanfeng and Zhang, Xiaodong. "Sep-graph: finding shortest execution paths for graph processing under a hybrid framework on GPU". In: *Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming*. 2019.

[12] Xiao, Mengbai and Wang, Hao and **Geng, Liang** and Lee, Rubao and Zhang, Xiaodong. "Catfish: Adaptive RDMA-enabled r-tree for low latency and high throughput". In: *2019 IEEE 39th International Conference on Distributed Computing Systems*. 2019.

[13] Zhang, Simon and Xiao, Mengbai and Guo, Chengxin and **Geng, Liang** and Wang, Hao and Zhang, Xiaodong. "Hypha: a framework based on separation of parallelisms to accelerate persistent homology matrix reduction". In: *Proceedings of the ACM International Conference on Supercomputing*. 2019.

Grants & AWARDS

Presidential Fellowship
One-year support for Ph.D. Dissertation
Nov, 2025
The Ohio State University

CSE Graduate Student Research Poster Exhibition
Best Poster Award
Apr, 2024
The Ohio State University

The 3rd National University Cloud Computing Application Innovation Competition
First Place Award
Apr, 2017
Nanjing, China