Bowen Wu

bowenwu.cs@gmail.com bowen.wu@inf.ethz.ch wubowen-cs

Education

ETH Zürich **2023 – 2027 (Expected)**
PhD Candidate in Computer Science *Zürich, Switzerland*

- Supervised by Prof. Gustavo Alonso and Prof. Ana Klimovic.
- Optimize databases for heterogeneous computing platforms, including CPUs, GPUs, FPGAs, and advanced networking technologies (e.g., Infiniband, RoCE).
- Develop operators and query optimizations for single-GPU, GPU clusters, and CPU-GPU hybrid environments.
- Explore software abstractions to enhance efficiency and reduce costs in query processing across heterogeneous platforms.

ETH Zürich **2020 – 2023**
MSc in Computer Science with distinction (GPA: 5.78/6) *Zürich, Switzerland*

The Chinese University of Hong Kong **2016 – 2020**
BSc in Computer Science, First Class Honor (major-GPA: 3.85/4) *Hong Kong, China*

Work Experiences

Microsoft Gray Systems Lab **May 2024 – Aug 2024**
Research Intern *Redmond, USA*

- Researched and optimized distributed multi-GPU SQL query acceleration.
- Profiled and modeled heterogeneous interconnects (e.g., NVLink, PCIe, Ethernet) in GPU clusters to assess their impact on GPU-accelerated query execution.
- Proposed a novel approach to optimize data exchange over heterogeneous interconnects, significantly enhancing query execution efficiency in GPU clusters. A patent for this technique has been applied for.

Cablex AG (Swisscom) **Aug 2022 – Oct 2022**
Database Analyst (Part-time) *Zürich, Switzerland*

- Designed a unified relational data model to enable comprehensive data analysis across multiple business sections.
- Leveraged cloud technologies to digitize and modernize operational workflows.

Amazon Web Service **Sep 2021 – Feb 2022**
Software Engineer Intern *Berlin, Germany*

- Optimized the query rewriter for Redshift, petabyte-scale cloud data warehouse in AWS.
- Proposed a collection of rewriting algorithms for nested SQL queries and implemented them in Redshift using both C++ and a proprietary domain-specific language.
- Improved the performance significantly with my proposed algorithms on the tera-byte scale TPC-DS benchmark.

Publications

- **Bowen Wu***, Wei Cui*, Carlo Curino, Matteo Interlandi, Rathijit Sen "Terabyte-Scale Analytics in the Blink of an Eye", https://arxiv.org/abs/2506.09226, in review.
- Marko Kabić, **Bowen Wu**, Jonas Dann, Gustavo Alonso "Powerful GPUs or Fast Interconnects: Analyzing Relational Workloads on Modern GPUs", VLDB 2026.
- Vasilis Mageirakos, **Bowen Wu**, Gustavo Alonso "Cracking Vector Search Indexes", VLDB 2026.
- **Bowen Wu**, Dimitrios Koutsoukos, Gustavo Alonso, "Efficiently Processing Joins and Grouped Aggregations on GPUs,", SIGMOD 2025.
- **Bowen Wu***, Chenyu Jiang*, Sanghamitra Dutta, Pulkit Grover, "An Information-Theoretic Measure for Enabling Category Exemptions with an Application to Filter Bubbles," BIAS@ECIR2021.
- Hongzhi Chen, **Bowen Wu**, Shiyuan Deng, et al., "High Performance Distributed OLAP over Property Graphs with Grasper," SIGMOD 2020 demo track.

Invited Talks

- **Huawei Edinburgh Research Center** 3/18/2025
- **IBM Research Center** 3/3/2025
- **Microsoft Gray Systems Lab** 6/4/2024

Selected Past Projects

Accelerate Relational Databases using GPU **Sep 2022 – Mar 2023**
Master thesis, advised by Prof. Gustavo Alonso *ETH Zürich*

- Developed a GPU-accelerated data analytics system based on vectorized execution.
- Implemented a rich collection of relational operators on the GPU with CUDA, e.g., grouped aggregation and join.
- Evaluated the efficiency and effectiveness of the vectorized execution model for GPU-accelerated databases.
- Achieved up to 2.5x speedup over a state-of-the-art GPU data processing system on the TPC-H benchmark.

Barrelfish Operating System **Feb 2022 – Jun 2022**
Advanced Operating System group project *ETH Zürich*

- Learned in-depth about the vast operating system design space, such as micro-kernel, capabilities, and scheduling.
- Implemented core functions of Barrelfish operating system for armv8 architecture on QEMU and i.MX 8x SoC, which include memory management and paging, multi-threading, process spawning, multi-core support, intra-/inter-core communication, and a FAT-32 file system. Contributed ∼6000 lines of C code individually.
- Redesigned the intra-/inter-core communication to be non-blocking with queues, multi-threading, and (de-)multiplexing.

Teaching

- **Data Management Systems** - Head Teaching Assistant & Invited Speaker Autumn 2024
- **Hardware Acceleration for Data Processing Seminar** - Invited Speaker Autumn 2024
- **Data Modeling and Databases** - Teaching Assistant Spring 2024, 2025
- **Information Systems for Engineers** - Teaching Assistant Autumn 2023

Technical Skills

Languages:	C/C++, CUDA, Python, Shell Script
Systems:	PyTorch, PostgreSQL, SQL Server
Frameworks:	MPI, NCCL, CUB, CuCollection, libcudf, Torch
Tools:	NSight System, NSight Compute
Topics:	GPU Programming, Databases, Distributed Systems, Cloud Computing, Systems for ML