Yunxin Sun

Email: martin.yunxin.sun@gmail.com

Website: yunxin-sun.com

Computer Science Ph.D. Program

Fall 2022

RESEARCH INTERESTS

I am interested in **computer systems** and their interactions with computer architecture and core database systems. In particular, I am interested in doing research in: ***(i)*** machine learning for systems, ***(ii)*** next-generation storage systems, and ***(iii)*** multi-objective data centers.

EDUCATION BACKGROUND

M.Sc. in Computer Science   Zurich, Switzerland
ETH Zurich   Sep. 2020 - Aug. 2022 (expected)

- Major in Data Management Systems, Minor in Programming Languages and Software Engineering
- GPA: 5.37/6

B.Eng. in Software Engineering   Shanghai, China
Tongji University   Sep. 2015 - Jul. 2020

- GPA: 88.88/100
- Ranking: Top 5%

RESEARCH EXPERIENCE

A Learning-Based Flash Translation Layer (FTL) for Solid-State Drives
Advisor: Prof. Dr. Jian Huang, PlatformX Lab   *UIUC*

- We built a learning-based FTL that could learn address mapping at runtime. It indexed a continuous access segment by training a linear-regression model. We coordinated this process with garbage collection (GC) so that it incurred negligible overhead.
- I proposed an LSM-tree-like data structure to accelerate update/lookup operations and tolerate machine learning misprediction. I implemented 40% of the project, including the LSM-tree-like data structure, tailored cache policies, and various other baselines.
- This learning-based FTL saves the memory footprint of the mapping table by up to 2.9x and improves the tail latency by 1.6x on average.
- This work has been submitted to USENIX FAST 2022 and is currently under review.

Fair and Deterministic I/O Scheduling in Modern NVMe Sets Solid State Drives
Advisor: Prof. Dr. Onur Mutlu & Dr. Jisung Park, SAFARI Research Group   *ETH Zurich*

- We designed a fair I/O scheduling policy for NVMe Sets specification that offers performance isolation. We divided the program and erase (P/E) operations into multiple substeps. Therefore, the SSD controller could alternately schedule each P/E substep and read request, preventing a program or erase operation from blocking a read request for too long.
- I implemented the project based on *MQSim*, a state-of-the-art simulator for SSD. I worked closely with my mentor to develop the research idea.
- The write tail latency improves up to 30% while not affecting read tail latency.
- We plan to submit this work to USENIX ATC 2022.

Execution Time Prediction For Serverless Workload
Advisor: Prof. Dr. Ana Klimovic, EASLab   *ETH Zurich*

- We used several machine learning models including linear regression and gradient boosting to estimate and predict the execution time of serverless functions. We trained the model using the information we collected, including the CPU usage, the memory footprint, the network traffic, and the input function parameters such as the size of the input image.

- I came up with a taxonomy to differentiate between different serverless workloads. I implemented 60% of the project, including setting up the pipeline to collect data and train the model.
- On average, our model achieves a root mean squared error (RMSE) of 0.49s.
- We released our findings as a technical report.

Publications

- **Yunxin Sun**, Tamar Shinar, and Craig Schroeder. "Effective time step restrictions for explicit MPM simulation." In Computer Graphics Forum, vol. 39, no. 8, pp. 55-67. 2020.

- Josh Urban Davis, Jun Gong, **Yunxin Sun**, Parmit Chilana, and Xing-Dong Yang. "Circuitstyle: A system for peripherally reinforcing best practices in hardware computing." In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, pp. 109-120. 2019.

- Guokai Zhang, Ye Luo, Dandan Zhu, Yixuan Xu, **Yunxin Sun**, and Jianwei Lu. "Spatial pyramid dilated network for pulmonary nodule malignancy classification." In 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3911-3916. IEEE, 2018.

Other Manuscripts

- "A Learning-Based Flash Translation Layer for Solid-State Drives". In submission to USENIX FAST 2022.

- "Execution Time Prediction For Serverless Workload". Technical Report.

Teaching and Mentoring Experience

- Mentored two undergraduate students in the course "*Seminar on Modern SSDs*". Met with them weekly to check the progress of the course project.

- Gave a lecture about *MQSim*, a state-of-the-art simulator for Solid State Drives, in the course "*Seminar on Modern SSDs*".

Volunteer Experience

- ACM SIGGRAPH/Eurographics Symposium on Computer Animation 2019 Student Volunteer

Skills

- **Programming Languages**: C, C++, SQL, Rust, Python, Assembly Language, Verilog, HLS, OCaml.
- **Tools & Software**: Git, CMake, perf, eBPF, PyTorch, GDB, Intel SGX, Shell, gem5, FPGA, LaTeX.

Honors and Awards

- Best Paper Honorable Mention at the ACM SIGGRAPH Symposium on Computer Animation (SCA) 2020.

Reference

- Prof. Dr. Jian Huang, *UIUC*, jianh@illinois.edu

- Dr. Jisung Park, *ETH Zurich*, jisung.park@safari.ethz.ch

- Prof. Dr. Ana Klimovic, *ETH Zurich*, aklimovic@ethz.ch