

# Bin Gao

☎ +65-84997688 • ✉ [bingao.cs@gmail.com](mailto:bingao.cs@gmail.com) • [in](#) [csbingao](#)  
🌐 [csbingao](#) • 🌐 [gaobin.me](#)

## Education

<b>National University of Singapore</b> <i>PhD, Computer Science</i>	<b>Singapore</b> <i>2021–present</i>
<b>Huazhong University of Science and Technology</b> <i>Master, Computer Architecture</i>	<b>Wuhan, China</b> <i>2017–2020</i>
<b>Huazhong University of Science and Technology</b> <i>Bachelor, Computer Science and Technology</i>	<b>Wuhan, China</b> <i>2013–2017</i>

## Research Interests

Computer Architecture, Computer Systems, Machine Learning System, Mobile Edge Computing

## Experience

<b>National University of Singapore</b> <i>Research Assistant</i>	<b>Singapore</b> <i>Dec. 2020–Aug. 2021</i>
<b>Key Lab of Services Computing Technology and System</b> <i>Research Assistant</i>	<b>Wuhan, China</b> <i>2017–2020</i>

## Publications (\*corresponding author)

- Conferences .....
- [1] **Bin Gao**, Zhehui Wang, Zhuomin He, Tao Luo, Weng-Fai Wong, Zhi Zhou, "IMI: In-memory Multi-job Inference Acceleration for Large Language Models", in the Proceedings of International Conference on Parallel Processing (ICPP), Gotland, Sweden, 2024
  - [2] **Bin Gao**, Zhuomin He, Puru Sharma, Qingxuan Kang, Djordje Jevdjic, Junbo Deng, Xingkun Yang, Zhou Yu, Pengfei Zuo, "Cost-Efficient Large Language Model Serving for Multi-turn Conversations with CachedAttention", in the Proceedings of USENIX Annual Technical Conference (ATC), Santa Clara, CA, USA, 2024
  - [3] **Bin Gao**, Qingxuan Kang, Hao-Wei Tee, Kyle Timothy Ng Chu, Alireza Sanaee, Djordje Jevdjic, "Scalable and Effective Page-table and TLB management on NUMA Systems", in the Proceedings of USENIX Annual Technical Conference (ATC), Santa Clara, CA, USA, 2024.
  - [4] Puru Sharma, Gary Goh Yipeng, **Bin Gao**, Longshen Ou, Dehui Lin, Deepak Sharma, Djordje Jevdjic, "DNA Storage Toolkit: A Modular End-to-End DNA Data Storage Codec and Simulator", in the Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), Indianapolis, Indiana, 2024."
  - [5] Qingyuan Wang, **Bin Gao\***, Zhi Zhou, Fei Xu, Chenghao Ouyang. "DAG-Aware Optimization for Geo-Distributed Data Analytics", in the Proceedings of International Conference on Parallel

Processing (ICPP), 2023.

[6] **Bin Gao**, Hao-Wei Tee, Alireza Sanaee, Soh Boon Jun, Djordje Jevdjic, "OS-level Implications of Using DRAM Caches in Memory Disaggregation", in the Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2022.

[7] Guohao Ying, Xin He, **Bin Gao**, Xiaowu Chu, Bo Han. "EAGAN: Efficient Multi-objective Evolutionary Architecture Search For GANs", in the Proceedings of ECCV, 2021.

[8] **Bin Gao**, Zhi Zhou , Fangming Liu, Fei Xu, "Winning at the Starting Line: Joint Network Selection and Service Placement for Mobile Edge Computing", in the Proceedings of IEEE INFOCOM, Paris, France, 2019.

Journals.....

[1] **Bin Gao**, Zhi Zhou, Fangming Liu, Fei Xu, Bo Li, "An Online Framework for Joint Network Selection and Service Placement in Mobile Edge Computing", IEEE Transaction on Mobile Computing (TMC), 2021

[2] Kongyang Zhao, **Bin Gao**, Zhi Zhou. "Cost-Effective Task Scheduling for Effective Task Scheduling for Collaborative Cross Collaborative Cross-Edge Analytics", ZTE Communications, 2021.

Workshops and Posters.....

[1] Yunkai Liang, **Bin Gao**, Pengfei Zuo, Zhi Zhou, Xu Chen, "PipeDecode: Efficient LLM Inference using Pipelines within Decoding", in USENIX Symposium on Operating Systems Design and Implementation (OSDI), Santa Clara, CA, US, 2024.

[2] **Bin Gao**, Hejing Li, Jialin Li, and Antoine Kaufmann, "Improving Disaggregated System Evaluation with Modular End-to-End Simulation", in Workshop On Resource Disaggregation and Serverless Computing (WORDS), 2022.

## Teaching

---

(1) NUS CS2106: Operating System

## Service

---

(1) SOSP 2021: Student Volunteer

(2) EuroSys 2022: Artifact Evaluation Committee

(3) OSDI/ATC 2022: Artifact Evaluation Committee

(4) TMC: Reviewer

## Honors

---

**Student Grant**

USENIX ATC

May 2024

**Silver Prize, "Internet+" Innovation Entrepreneurship Competition**

Ministry of Education of the People's Republic of China

Oct. 2022

**Excellent Graduate**

Huazhong University of Science and Technology

June 2020

**Excellent Graduate***Huazhong University of Science and Technology**June 2020***National Scholarship***Ministry of Education of the People's Republic of China**Oct. 2019***Merit Student***Huazhong University of Science and Technology**Oct. 2019***Excellent Student Cadre***Huazhong University of Science and Technology**Oct. 2018***First-class Academic Scholarship***Huazhong University of Science and Technology**2017-2020***Excellent Graduate***Huazhong University of Science and Technology**June 2017***National Encouragement Scholarship***Ministry of Education of the People's Republic of China**Oct. 2016***Languages**

---

**Chinese:** Native**English:** Business Level**Hobbies**

---

○ Music, Cycling, History