# Zicong Hong

*INR 313, EPFL, 1015 Lausanne, Switzerland*
(+852)5634-9689 & (+86)159-8999-5217 | zicong.hong@epfl.ch | https://zicongs-homepage.webflow.io/

## Professional Experience

**École Polytechnique Fédérale de Lausanne** *Switerzland*
Scientist *11/2025 - Now*
- Host: Prof. Rachid Guerraoui

**The Hong Kong University of Science and Technology** *Hong Kong*
Postdoctoral Research Fellow *02/2025 - 10/2025*
- Advisor: Prof. Song Guo

**ETH Zürich** *Switzerland*
Visiting Scholar *04/2023-10/2023*
- Host: Prof. Roger Wattenhofer

## Academic Qualifications

**The Hong Kong Polytechnic University** *Hong Kong*
Doctor of Philosophy in Computer Science *09/2020 - 01/2025*
- Advisor: Prof. Song Guo and Prof. Bin Xiao

**Sun Yat-sen University** *China*
Bachelor of Engineering in Software Engineering *09/2015-07/2020*
- Advisor: Prof. Wuhui Chen

## Research Interests

AI Systems; Large Language Model Serving; Distributed Systems; Decentralized Systems

## Awards & Honors

2026 **Rising Stars in AI**, King Abdullah University of Science and Technology
2020-2024 **Hong Kong Ph.D. Fellowship (Top HK Fellowship)**, Hong Kong Research Grants Council
2024 **Best Presentation Award**, PolyU COMP 50th Anniversary Research Student Conference
2024 **Silver Medal**, Geneva Inventions Expo
2023 **The Most Appreciated Teaching Assistant Award**, Hong Kong Polytechnic University
2022 **2nd Runner-Up Award in Huawei Global Developer Competition for APAC**, Huawei
2020 **Best Paper Runner-up Award**, the 49th International Conference on Parallel Processing
2019 **Kwang-Hua Fellowship**, Kwang-Hua Educational Foundation

## Refereed Publications

### Publication Summary

- 20+ papers as first or corresponding author
- Regular publications in top-tier systems, networking, and ML venues (including SIGMOD, ASPLOS, EUROSYS, ATC, MOBICOM, VLDB, NeurIPS, WWW, NDSS, INFOCOM, TPDS, TDSC, JSAC)
- h-index: 19; citations: 2500+

### Efficient LLM Serving & Edge AI Systems:

**C1. [SIGMOD]** Jian Lin, Jiazhi Mi, **Zicong Hong#**, Haodong Wang, Qianli Liu, Haoyue Zhang, Peng Li, and Song Guo (#corresponding author). "*KVDrive: A Holistic Multi-Tier KV Cache Management System for Long-Context LLM Inference*". International Conference on Management of Data, 2026.

**C2. [EUROSYS] Zicong Hong**, Yuyan Chen, Haoyue Zhang, Peng Li, Wuhui Chen, Song Guo, and Xiaowei Shen. "*Efficient Multimodal Serving via Module Multiplexing*". European Conference on Computer Systems, 2026.

**C3. [WWW]** Zhiyuan Fang, Xingfan Yu, Yuegui Huang, **Zicong Hong#**, Yufeng Lyu, Wuhui Chen, Yue Yu, Fan Yu (#corresponding author). "*Fate: Fast Edge Inference of Mixture-of-Experts Models via Cross-Layer Gate*". The Web Conf., 2026.

**C4. [MOBICOM]** Haodong Wang, Qihua Zhou, **Zicong Hong#**, Song Guo (#corresponding author). "*$D^2$MoE: Dual Routing and Dynamic Scheduling for Efficient On-Device MoE-based LLM Serving*". Annual International Conference On Mobile Computing And Networking, 2025.

**C5. [INFOCOM]** Jinyu Chen, Wenchao Xu, **Zicong Hong#**, Song Guo, Haozhao Wang, Jie Zhang, Deze Zeng (#corresponding author). "*OTAS: An Elastic Transformer Serving System via Token Adaptation*". IEEE International Conference on Computer Communications, 2024.

**C6. [ASPLOS]** Zhiyuan Fang, Yuegui Huang, **Zicong Hong**, Yufeng Lyu, Wuhui Chen, Yue Yu, Fan Yu, Zibin Zheng. "*Klotski: Efficient Mixture-of-Expert Inference via Expert-Aware Multi-Batch Pipeline*". the ACM International Conference on Architectural Support for Programming Languages and Operating Systems, 2025.

**C7. [NeurIPS]** Sikai Bai, Haoxi Li, Jie Zhang, **Zicong Hong**, Song Guo. "*DiEP: Adaptive Mixture-of-Experts Compression through Differentiable Expert Pruning*". Conference on Neural Information Processing Systems, 2025.

### Distributed AI Systems & Serverless ML:

**C8. [INFOCOM]** Qianli Liu, Kaibin Guo, **Zicong Hong#**, Peng Li, Fahao Chen, Haodong Wang, Jian Lin, and Song Guo (#corresponding author). "*Director: Accelerating Distributed MoE Serving via Online Proactive Expert Placement*". IEEE International Conference on Computer Communications, 2026.

**C9. [INFOCOM]** Zile Wang, Qianli Liu, Kaibin Guo, Jian Lin, **Zicong Hong#**, and Song Guo (#corresponding author). "*PPAI: Enabling Personalized LLM Agent Interoperability for Collaborative Edge Intelligence*". IEEE International Conference on Computer Communications, 2026.

**C10. [EUROSYS]** Yapeng Jiang, Wuhui Chen, Ganhong Huang, Yuzhou Huang, **Zicong Hong**, Song Guo, and Yue Yu. "*Crimson: Collaborative Parameter Updates for Efficient Pipeline Training of Large Language Models*". European Conference on Computer Systems, 2026.

**C11. [INFOCOM]** Qianli Liu, **Zicong Hong#**, Peng Li, Fahao Chen, Song Guo (#corresponding author). "*Mell: Memory-Efficient Large Language Model Serving via Multi-GPU KV Cache Management*". IEEE International Conference on Computer Communications, 2025.

**C12. [EUROSYS] Zicong Hong***, Jian Lin*, Song Guo, Sifu Luo, Wuhui Chen, Roger Wattenhofer, Yue Yu (*The first two authors have equal contribution). "*Optimus: Warming Serverless ML Inference via Inter-Function Model Transformation*". European Conference on Computer Systems, 2024.

**J1. [TPDS] Zicong Hong**, Wuhui Chen, Huawei Huang, Song Guo, Zibin Zheng. "*Multi-hop Cooperative Computation Offloading for Industrial IoT-Edge-Cloud Computing Environments*". IEEE Trans. on Parallel & Distributed Systems, 2019.

**C13. [ATC]** Yuzhou Huang, Yapeng Jiang, **Zicong Hong**, Wuhui Chen, Bin Wang, Weixi Zhu, Yue Yu, Zibin Zheng. "*Obscura: Concealing Recomputation Overhead in Training of Large Language Models with Bubble-filling Pipeline Transformation*". the USENIX Annual Technical Conference, 2025.

**J2. [TON]** Fahao Chen, Peng Li, **Zicong Hong**, Zhou Su, Song Guo. "*Communication-Efficient Sparsely-Activated Model Training via Sequence Migration and Token Condensation*". IEEE Trans. on Networking, 2025.

### AI Security:

**C14. [WWW]** Enyuan Zhou, Song Guo, Zhixiu Ma, **Zicong Hong#**, Tao Guo, Peiran Dong (#corresponding author). "*Poisoning Attack on Federated Knowledge Graph Embedding*". The Web Conference, 2024.

**C15. [NIPS]** Peiran Dong, Bingjie Wang, Song Guo, Junxiao Wang, Jie Zhang, **Zicong Hong**, Xiao-Ming Wu. "*Towards Safe Concept Transfer of Multi-Modal Diffusion via Causal Representation Editing*". NeurIPS, 2024.

**C16. [ICML]** Jie Zhang, Xiaosong Ma, Song Guo, Peng Li, Wenchao Xu, Xueyang Tang, **Zicong Hong**. "*Amend to Alignment: Decoupled Prompt Tuning for Mitigating Spurious Correlation in Vision-Language Models*". International Conference on Machine Learning, 2024.

### Decentralized Systems:

**C17. [VLDB]** Enyuan Zhou, Song Guo, **Zicong Hong#**, Christian Jensen, Yang Xiao, Jinwen Liang, Dalin Zhang (#corresponding author). "*Pistis: A Decentralized Knowledge Graph Platform Enabling Ownership-Preserving SPARQL Querying*". International Conference on Very Large Data Bases, 2025.

**J3. [TDSC]** Jianting Zhang, **Zicong Hong#**, Xiaoyu Qiu, Wuhui Chen, Yufeng Zhan, and Song Guo (#corresponding author). "*Scaling Blockchain via Dynamic Sharding*". IEEE Trans. on Dependable and Secure Computing, 2025.

**J4. [TDSC]** Jinwen Liang, Song Guo, **Zicong Hong#**, Enyuan Zhou, Chuan Zhang, Bin Xiao (#corresponding author). "*SecPQ: Secure Prediction Queries on Encrypted Outsourced Databases*". IEEE Trans. on Dependable & Secure Computing, 2025.

**J5. [TPDS]** Junyuan Liang, Wuhui Chen, **Zicong Hong#**, Haogang Zhu, Wangjie Qiu, Zibin Zheng (#corresponding author). "*MoltDB: Accelerating Blockchain via Ancient State Segregation*". IEEE Trans. on Parallel & Distributed Systems, 2024.

**C18. [VLDB] Zicong Hong**, Song Guo, Enyuan Zhou, Wuhui Chen, Huawei Huang, Albert Zomaya. "*GriDB: Scaling Blockchain Database via Sharding & Off-Chain Cross-Shard Mechanism*". International Conference on Very Large Data Bases, 2023.

**C19. [INFOCOM] Zicong Hong**, Song Guo, Enyuan Zhou, Jianting Zhang, Wuhui Chen, Jinwen Liang, Jie Zhang, Albert Zomaya. "*Prophet: Conflict-Free Sharding Blockchain via Byzantine-Tolerant Deterministic Ordering*". IEEE International Conference on Computer Communications, 2023.

**J6. [JSAC] Zicong Hong**, Song Guo, Peng Li. "*Scaling Blockchain via Layered Sharding*". IEEE Journal on Selected Areas in Communications, 2022.

**C20. [DSN] Zicong Hong**, Song Guo, Rui Zhang, Peng Li, Yufeng Zhan, Wuhui Chen. "*CYCLE: Sustainable Off-Chain Payment Channel Network with Asynchronous Rebalancing*". IEEE/IFIP International Conference on Dependable Systems and Networks, 2022.

**C21. [INFOCOM] Zicong Hong**, Song Guo, Peng Li, Wuhui Chen. "*Pyramid: A Layered Sharding Blockchain System*". IEEE International Conference on Computer Communications, 2021.

**C22. [ICPP]** Jianting Zhang*, **Zicong Hong***, Yufen Zhan, Song Guo, Wuhui Chen (*The first two authors have equal contribution). "*SkyChain: A Deep Reinforcement Learning-Empowered Dynamic Sharding Blockchain System*". International Conference on Parallel Processing (**Best paper runner-up prize**), 2020.

## Books/Chapters

**B1. Zicong Hong**, Song Guo, Peng Li, and Wuhui Chen. "*Layered Sharding on Open Blockchain*". Blockchain Scalability, Springer, 97-117.

**B2.** Jianting Zhang, **Zicong Hong**, Xiaoyu Qiu, Yufeng Zhan, Song Guo, and Wuhui Chen "*Dynamic sharding: a trade-off between security and scalability*". Blockchain Scalability, Springer, 193-221.

**B3.** Hao Ren, Jjnwen Liang, **Zicong Hong**, Enyuan Zhou, and Junbao Pan. "*Application: Privacy, Security, Robustness and Trustworthiness in Edge AI*". Machine Learning on Commodity Tiny Devices, CRC Press, 161-186.

## Mentoring

**Ph.D. Student Mentoring as Postdoc (Systems and AI)**

- Qianli Liu (HKUST, admitted 2024) — distributed LLM serving systems
- Haodong Wang (HKUST, admitted 2024) — on-device MoE inference
- Jian Lin (HKUST, admitted 2024) — long-context KV cache management systems
- Zile Wang (HKUST, admitted 2025) — efficient AI model architecture
- Jiazhi Mi (HKUST, admitted 2025) — multi-GPU offloading systems

## Funding Experience & Project Leadership

- **RGC General Research Fund (GRF)** (2026–2028): "Towards Edge-Native Agentic AI Serving in Resource-Constrained Environments." *Lead proposal writer and technical lead*; HKD 1,068,006.
- **NSFC/RGC Collaborative Research Scheme (CRS)** (2025–2028): "Enabling Security Protection for Heterogeneous Blockchain Systems in Web 3.0." *Lead proposal contributor and technical co-lead*; HKD 3,564,160.
- **RGC General Research Fund (GRF)** (2025–2027): "Knowledge-Driven Edge-Native Foundation Model Serving on Collaborative Devices." *Lead proposal writer and technical lead*; HKD 1,298,709.

- **RGC Research Impact Fund (RIF)** (2024–2029): "An AI-enabled Geospatial Platform for Smart Facility Management and Smart Mobility of People with Disabilities." *Co-lead on system architecture*; HKD 4,620,000.
- **RGC General Research Fund (GRF)** (2024–2026): "Neural-enhanced Edge Perception: Representation-aware Analytics on Collaborative Devices." *Lead proposal writer and technical lead*; HKD 1,533,087.
- **Industry-Sponsored Projects**: Lead technical contributor for two MetaX Technology projects (HKD 1.2M and HKD 1.0M) and one Huawei research gift (HKD 330,000).

## Teaching

| | |
|---|---|
| Spring'24 | **Computer Systems Security**, Teaching Assistant |
| Spring'23 | **Emerging Topics in Fintech**, Teaching Assistant |
| Fall'22 | **Programming Fundamentals and Applications**, Teaching Assistant (awarded by The Most Appreciated Teaching Assistant Award in HK PolyU) |
| Spring'22 | **Big Data Analytics**, Teaching Assistant |
| Fall'21 | **Programming Fundamentals**, Teaching Assistant |
| Spring'21 | **Big Data Analytics**, Teaching Assistant |
| Fall'20 | **Object-oriented Programming**, Teaching Assistant |

## Professional Activities

Reviewer for IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Services Computing, IEEE Transactions on Computers, IEEE Journal on Selected Areas in Communications, IEEE Open Journal of the Computer Society, IEEE Transactions on Industrial Informatics, IEEE Internet of Things Journal, IEEE Transactions on Vehicular Technology, and IEEE Transactions on Knowledge and Data Engineering. Active contributor to the systems and AI research community through peer review and collaborative research initiatives.

## References

**Prof. Song Guo**
Chair Professor of Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
songguo@cse.ust.hk

**Prof. Rachid Guerraoui**
Professor
EPFL
rachid.guerraoui@epfl.ch

**Prof. Anne-Marie Kermarrec**
Professor
EPFL
anne-marie.kermarrec@epfl.ch

**Prof. Bin Xiao**
Professor of Department of Computing
The Hong Kong Polytechnic University
csbxiao@comp.polyu.edu.hk

**Prof. Albert Y. Zomaya**
Peter Nicol Russell Chair Professor of Computer Science
The University of Sydney
albert.zomaya@sydney.edu.au