Houxiang Ji

Contact Information

213 Coordinated Science Lab
1308 W Main St., Urbana, IL 61801

Website: https://hxji.github.io/
Email: hj14@illinois.edu

Research Interests

Computer Architecture, Memory System, Compute eXpress Link (CXL), Near Memory Processing, System for Machine Learning

Education

University of Illinois Urbana-Champaign
Ph.D. Candidate, Computer Science
Advisor: Prof. Nam Sung Kim

University of Illinois Urbana-Champaign
M.Sc., Computer Science, Dec 2021

Shanghai Jiao Tong University
B.Eng., Computer Science and Engineering, July 2018
B.Sc., Finance and Economics, July 2018

Honors and Awards

- Best Demo Award, SRC/DARPA PRISM Center Annual Review, 2023
- UIUC CS PhD Fellowship, UIUC, 2023
- Travel Grants: USENIX OSDI/ATC 2023, IEEE ISCA 2019, IEEE/ACM MICRO 2024
- Ray Ozzie Computer Science Fellowship, UIUC, 2018
- Zhiyuan Honor Degree in Computer Science, SJTU, 2018
- Outstanding Graduates, SJTU, 2018
- Zhiyan Honor Scholarship, SJTU, 2015-2017
- Excellent Undergraduate Scholarship, SJTU, 2015-2017

Research Experience

CXL: Analysis, Applications, and Utilization in Datacenters
Supervised by Prof. Nam Sung Kim
- explore CXL Type-2 device to boost hyperscaler application performance [C14]
- CXL device characterization with genius system [C11, C14]

Datacenter Memory Tax Reduction
Supervised by Prof. Nam Sung Kim
- leverage SmartNIC to reduce kernel features cost on CPU [C10]
- kernel feature acceleration with near-memory processing unit (AxDIMM)
- explore on-chip accelerator (e.g. QAT, DSA) to reduce datacenter tax [J1, J3]

Deep Learning Model Acceleration on CPU
Supervised by Prof. Josep Torrellas
- sw/hw co-design to optimize aggregation/update phase in GNN [C8]
- leverage sparsity in DNN for faster training and inference [C5, C6]
- demystify graph neural networks in recommender systems [master thesis]

Professional Experience

Capacity Engineering and Analysis, Meta
Visiting Researcher — Jun 2022 - Dec 2022
*Explored **CXL** devices in collaboration with Meta Infrastructure, focusing on their integration and performance optimization within cutting-edge computing environments.*

CPU Design, Luminous Computing
Research Intern — May 2021 - Aug 2021
Design and simulate a RISC-V-based AI CPU architecture optimized for high-bandwidth photonic data transmission.

Security and Privacy Research, Intel Lab
Research Intern — May 2020 - Aug 2020
Implement and evaluate hardware/software co-design defense schemes against speculative execution attacks using a commercial CPU simulator.

Conference & Journal Publications

[C15] **Houxiang Ji**, Minho Kim, Seonmu Oh, Daehoon Kim, Nam Sung Kim. *Para-ksm: Parallelized Memory Deduplication with Data Streaming Accelerator*, **USENIX ATC 2025**

[J3] **Houxiang Ji***, Qirong Xia*, Yang Zhou, Nam Sung Kim. *Hardware-accelerated Kernel-Space Memory Compression Using Intel QAT*, *equal contribution, **IEEE Computer Architecture Letter, Jan-June 2025**

[J2] Chihun Song, Michael Jaemin Kim, Yan Sun, **Houxiang Ji**, Kyungsan Kim, TaeKyeong Ko, Jung Ho Ahn, Nam Sung Kim. *X-PPR: Post Package Repair for CXL Memory*, **IEEE Computer Architecture Letter, Jan-June 2025**

[J1] **Houxiang Ji**, Minho Kim, Seonmu Oh, Daehoon Kim, Nam Sung Kim. *Cooperative Memory Deduplication with Intel Data Streaming Accelerator*, **IEEE Computer Architecture Letter, Jan-June 2025**

[C14] **Houxiang Ji**, Srikar Vanavasam, Yang Zhou, Qirong Xia, Jinghan Huang, Yifan Yuan, Ren Wang, Pekon Gupta, Bhushan Chitlur, Ipoom Jeong, Nam Sung Kim. *Demystifying a CXL Type-2 Device: A heterogeneous cooperative computing perspective*, **IEEE/ACM MICRO 2024**

[C13] Jinghan Huang, Jiaqi Lou, Srikar Vanavasam, Xinhao Kong, **Houxiang Ji**, Ipoom Jeong, Eun Kyung Lee, Danyang Zhuo, Nam Sung Kim. *HAL: Hardware-assisted Load Balancing for Energy-efficient SNIC-Host Cooperative Computing*, **IEEE/ACM ISCA 2024**

[C12] Chihun Song, Michael Jaemin Kim, Tianchen Wang, **Houxiang Ji**, Jinghan Huang, Ipoom Jeong, Jaehyun Park, Hwayong Nam, Minbok Wi, Jung Ho Ahn, Nam Sung Kim. *TAROT: A CXL SmartNIC-Based Defense Against Multi-bit Errors by Row Hammer Attacks*, **ACM ASPLOS 2024**

[C11] Yan Sun, Yifan Yuan, Zeduo Yu, Chihun Song, Reese Kuper, Jinghan Huang, **Houxiang Ji**, Siddharth Agarwal, Jiaqi Lou, Ipoom Jeong, Ren Wang, Jung Ho Ahn, Tianyin Xu, Nam Sung Kim. *Demystifying CXL Memory with Genuine CXL-Ready Systems and Devices*, **IEEE/ACM MICRO 2023**

[C10] **Houxiang Ji**, Yan Sun, Mark Mansi, Yifan Yuan, Jinghan Huang, Reese Kuper, Michael M. Swift, Nam Sung Kim. *STYX: Exploiting SmartNIC Capability to Reduce*

Datacenter Memory Tax, **USENIX ATC 2023**

[C9] Lihui Liu, **Houxiang Ji**, Jiejun Xu, and Hanghang Tong. *Comparative Reasoning for Knowledge Graph Fact Checking*, **IEEE BigData 2022**

[C8] Zhangxiaowen Gong, **Houxiang Ji**, Yao Yao, Christopher Fletcher, Christopher Hughes, Josep Torrellas. *Optimizing Graph Neural Networks on CPUs via Cooperative Software-Hardware Techniques*, **IEEE/ACM ISCA 2022**

[C7] Zirui Neil Zhao, **Houxiang Ji**, Adam Morrison, Darko Marinov, Josep Torrellas. *Pinned Loads: Taming Speculative Loads in Secure Processors*, **ACM ASPLOS 2022**

[C6] Zhangxiaowen Gong, **Houxiang Ji**, Christopher Fletcher, Christopher Hughes, Josep Torrellas. *SparseTrain: Leveraging Dynamic Sparsity in Software for Training DNNs on General-Purpose SIMD Processors*, **PACT 2020** [Open Source]

[C5] Zhangxiaowen Gong, **Houxiang Ji**, Christopher Fletcher, Christopher Hughes, Sara Baghsorkhi, Josep Torrellas. *SAVE: Sparsity-Aware Vector Engine for Accelerating DNN Training and Inference on CPUs*, **IEEE/ACM MICRO 2020**

[C4] Zirui Zhao, **Houxiang Ji**, Mengjia Yan, Jiyong Yu, Christopher W. Fletcher, Adam Morrison, Darko Marinov, Josep Torrellas. *Speculation Invariance (InvarSpec): Faster Safe Execution Through Program Analysis*, **IEEE/ACM MICRO 2020**

[C3] **Houxiang Ji**, Li Jiang, Tianjian Li, Naifeng Jing, Jing Ke, Xiaoyao Liang. *HUBPA: High Utilization Bidirectional Pipeline Architecture for Neuromorphic Computing*, **ASP-DAC 2019**

[C2] **Houxiang Ji**, Linghao Song, Li Jiang, Hai(Halen) Li, Yiran Chen. *ReCom: An efficient resistive accelerator for compressed deep neural networks*, **IEEE DATE 2018**

[C1] Luyu Li, **Houxiang Ji**, Chentao Wu, Jie Li, Minyi Guo. *Favorable Block First: A Comprehensive Cache Scheme to Accelerate Partial Stripe Recovery of Triple Disk Failure Tolerant Array*, **ICPP 2017**

Selected Pre-prints

[P1] **Houxiang Ji**, Yifan Yuan, Yang Zhou, Ipoom Jeong, Ren Wang, Saksham Agarwal, Nam Sung Kim. *CXL-NIC: an Efficient NIC Interface based on CXL*, Under Review

Invited Talks

Demystifying a CXL Type-2 Device: A heterogeneous cooperative computing perspective

- Shanghai Jiao Tong University ACA lab, Dec. 2024
- IEEE/ACM MICRO, Nov. 2024
- Open Compute Project Global Summit, Oct. 2024
- Open Compute Project composable memory system group talk, Mar. 2024

Memory-Intensive Kernel Features Acceleration with CXL Type-2 Device

- Open Compute Project Global Summit 2023
- Open Compute Project composable memory system group talk, Oct. 2023

STYX: Exploiting SmartNIC Capability to Reduce Datacenter Memory Tax

- USENIX ATC 2023

Professional Service

Technical Program Committee

- Great Lakes Symposium on VLSI (GLSVLSI), 2025

Conference & Journal Reviewer

- IEEE Transactions on Emerging Topics in Computing (TETC), 2023
- ACM Journal on Emerging Technologies in Computing Systems (JETC), 2024
- IEEE Computer Architecture Letters (CAL), 2024
- IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2021,2023,2025
- International Conference on Networks, Communications and Information Tech. (NCIT), 2022
- International Conference on Electrical, Computer and Energy Technologies (ICECET), 2024,2025

Artifact Evaluation Committee

- 2024: OSDI, ATC, MICRO
- 2025: ASPLOS

Community Service

- Mentor, Promoting Undergraduate Research in Engineering (PURE), UIUC, 2023
- Mentor, Undergraduate Researcher in FAST, UIUC, 2023

Teaching Experience

CS233 Computer Architecture, Teaching Assistant, UIUC Spring 2024

References

References available upon request.