CHUNWEI XIA

https://github.com/summerspringwei C.Xia@leeds.ac.uk
3.25, Sir William Henry Bragg Building, Leeds LS2 9JT, UK +44 7394140830

EDUCATION & WORKING EXPERIENCE

School of Computer Science, University of Leeds, Leeds, UK *Oct. 2024 - Now*
Lecturer/ Assistant Prof. in Intelligent System Software

School of Computer Science, University of Leeds, Leeds, UK *May. 2023 - Seq. 2024*
Research Fellow in Compiler Optimization using Machine Learning
Advisor: Prof. Dr. Zheng Wang

Institute of Computing Technology, Chinese Academic of Sciences, Beijing, China *Sep. 2016 - Jan. 2023*
Doctor of Engineering in Computer System and Architecture
Advisor: Prof. Dr. Huimin Cui and Prof. Dr. Xiaobing Feng

School of Computer Science and Technology, Tianjin University, Tianjin, China *Sep. 2012 - Jun. 2016*
Bachelor of Computer Science and Technology

HONORS AND AWARDS

- **ACM SIGHPC China Doctoral Dissertation Award, Oct. 2023 (Only three recipients of the award in China)**
- First-class Scholarship, University of Chinese Academy and Science (top 20%), 2019-2020
- Merit Student, University of Chinese Academy and Science (top 20%), 2018-2019
- Merit Student, Tianjin University (top 20%), 2013-2016
- National Encouragement Scholarship (top 10%), Tianjin University, 2013-2014

PUBLICATIONS

1. J Zhao, **C Xia**, Z Wang, Leveraging Compilation Statistics for Compiler Phase Ordering, in submitting to International Symposium on Code Generation and Optimization, **(IPDPS, Core Rank A), 2024**

2. S Qiu, **C Xia**, Z Wang, Falcon: Accelerating Tensor-train Decomposition on Graph Neural Networks, In submitting to Interna- tional Symposium on Code Generation and Optimization, **(IPDPS 2025, Core Rank A), 2024**

3. N Lin, S Wang, X Zhang, S Wang, Y He, W Zhang, B Wang, **C Xia**, W Xuan, X Chen, D Shang and Z Wang. "LSMR: Synergy Randomness in Liquid State Machine and RRAM-based Analog-digital Accelerator", ACM/IEEE International Conference on Computer-Aided Design **(ICCAD 2024, Core Rank A), 2024**

4. S Zhang, J Zhao, **C Xia**, Z Wang, Y Chen, H Cui, Introducing Compiler Semantics into Large Language Models as Programming Language Translators: A Case Study of C to x86 Assembly, The 62nd Annual Meeting of the Association for Computational Linguistics **(EMNLP 2024, Core Rank A*), 2024**

5. H Wang, Z Tang, S Tan, J Wang, Y Liu, H Fang, **C Xia**, Z Wang, Combining Structured Static Code Information and Dynamic Symbolic Traces for Software Vulnerability Prediction, the International Conference on Software Engineering **(ICSE 2024, Core Rank A*, Artifact Evaluated!**), **2023**

6. **C Xia**, J Zhao, H Cui, X Feng, Optimizing Deep Learning Inference via Global Analysis and Tensor Expression, ACM International Conference on Architectural Support for Programming Languages and Operating Systems, **(ASPLOS 2024, Core Rank A*, Artifact Evaluated!**), **2023**

7. **C Xia**, J Zhao, H Cui, X Feng, HOPE: A Heterogeneity-Oriented Parallel Execution Engine for Inference on Mobiles, High Technology Letters, **(Core Journal of China), 2022**

8. N Lin, X Chen, **C Xia**, J, Ye, X Li, ChaoPIM: A PIM-based Protection Framework for DNN Accelerators Using Chaotic Encryption, 2021 IEEE Asian Test Symposium **(ATS 2021**), **2021**

9. **C Xia**, J Zhao, H Cui, X Feng, J Xue, Dnntune: Automatic benchmarking DNN models for mobile-cloud computing, ACM Transactions on Architecture and Code Optimization **(TACO, Core Rank B)**, **2019**

10. **C Xia**, J Zhao, H Cui, X Feng, Characterizing DNN models for edge-cloud computing, 2018 IEEE International Symposium on Workload Characterization **(IISWC 2018)**, 82-83, **2018**

11. J Zhao, Y Chang, D Li, **C Xia**, H Cui, K Zhang, X Feng, On retargeting the ai programming framework to new hardware, International Conference on Network and Parallel Computing **(NPC 2018)**, 39-51, **2018**

12. **Granted Patent**: **C Xia**, J Zhao, H Cui, X Feng, Method for executing deep neural networks on heterogeneous processing units, Patent Number: ZL 2020 1 0493830.8

13. **Published Patent Application**: Y Li, **C Xia**, S Du, J Zhao, H Cui, Compilation Methods and Compiler for Cross-Operator Boundary Optimization in Deep Neural Network Inference, Patent Number: CN117742718A

14. **Published Patent Application**: Y Li, S Du, **C Xia**, J Zhao, H Cui, Deep-Neural-Network-Based Kernel Fusion Methods and Systems, Patent Number: CN117742679A

15. **Published Patent Application**: Y Li, **C Xia**, S Du, J Zhao, H Cui, Data Preprocessing Execution Methods and Systems for Artificial Intelligence Processors, Patent Number: CN117725969A

RESEARCH PROJECTS

Research on Large Language Model for Compiler Optimization ***April 2023 - present***
- LLVM performance auto-tuning for general program optimization on ARM, sponsored by Huawei, core contributor
- Modernise compiler technology through Deep Learning, funded by EPSCP1109, **technical leader**

Research on Programming and Compilation Optimization for Deep Learning Software ***Sep. 2016 - Dec. 2022***
- Design and implementation of DL compiler for edge neural processing unit, sponsored by Intel, **technical leader**
- Research on key technologies for the Ark compiler ecosystem, sponsored by HiSilicon, **technical leader**
- Programming model and compiler optimization for heterogeneous data-flow accelerators, sponsored by China National Key R&D Program, **technical leader**
- Building unified compiler infrastructure for domain-specific architectures, sponsored by NSFC, core contributor
- Building AI-native programming language and multi-level optimizations, sponsored by China National Key R&D Program, core contributor
- Redesigning programming language for Ascend AI chips, sponsored by HiSilicon, core contributor

ACADEMIC SERVICES

1. Workshop chair & poster track chair for 29th International Conference on Automation and Computing (ICAC'24)

2. Reviewer for flagship computer journals & conferences TPDS, TOCS, ICAC, CGO'22-24, PPoPP'22 and PACT'21

CONFERENCES AND PRESENTATIONS

1. Invited talk for Intel on deep learning compiler optimization, 17 June 2024

2. Invited talk for HiSilicon on optimizing tensor programs on heterogeneous hardware, 15 March 2024

3. **Oral** presentation at ASPLOS 2024, 27th April - 1st May, San Diego, US; **C Xia**, J Zhao, H Cui, X Feng, Optimizing Deep Learning Inference via Global Analysis and Tensor Expression

4. **Oral** presentation at HiPEAC 2020, January 20-22, 2020; Bologna, Italy; Dnntune: Automatic Benchmarking DNN Models for Mobile-cloud Computing

5. **Poster** presentation at HiPEAC 2024, January 17-19 2024; Munich, Germany; Optimizing Deep Learning Inference via Global Analysis and Tensor Expression

6. **Poster** presentation at IISWC 2018, Sep. 30- Oct. 2, 2018; Raleigh, North Carolina, USA; Characterizing DNN models for edge-cloud computing