Jie Zhao

Curriculum Vitae

College of Computer Science and Electronic Engineering
Hunan University, Lushan Road (S)
Yuelu District, Changsha, 410082, China
✉ *jiezhao@hnu.edu.cn*
https://yaozhujia.github.io/

Research Interests

Jie Zhao is a Full Professor of computer sciences in the College of Computer Science and Electronic Engineering of Hunan University. He has been working on compilation techniques since he was a graduated student. He devotes to high-performance code generation for different architectures using compilation approaches and strives to bridge the gap between high-level programming models and underlying computer architectures. Specifically, his research interests and experiences span across:

- systems for machine learning
- code generation and optimization
- numerical program analysis
- high level synthesis

Employments

2024.01–present **Full Professor**, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

2023.09–2023.12 **Associate Professor**, *Cloud and Big Data System Laboratory*, School of Information, Renmin University of China, Beijing, China.

2019.01–2023.07 **Lecturer**, *State Key Laboratory of Mathematical Engineering and Advanced Computing*, School of Cyberspace Security, Information Engineering University, Zhengzhou, China.

Experiences

2023.08–2024.07 **Senior Consultant**, *Huawei Technologies Co., Ltd.*, Beijing, China.

2021.04–2023.03 **Senior Consultant**, *Stream Computing Co., Ltd.*, Hangzhou, China.

2019.02–2021.12 **Senior Consultant**, *Huawei Technologies Co., Ltd.*, Hangzhou, China.

2020.04–2021.03 **Visiting Scholar**, *Alibaba Group*, Beijing, China.

2018.05–2018.09 **Consultant**, *Huawei Technologies France SASU*, Paris, France.

Education

2015.03–2018.12 **PhD in Mathematics**, *INRIA & École Normale Supérieure*, 75005 Pairs, France, PhD Thesis: A Combined Language and Polyhedral Approach for Heterogeneous Parallelism. Supervised by Prof. Albert Cohen

2012.09–2016.06 **PhD in Computer Science**, *National Digital Switching System Engineering & Technological Research Center (NDSC), Information Engineering University*, 450001, Zhengzhou, China, Phd thesis: Data Dependence Analysis in Parallelizing Compilers.
Supervised by Prof. Rongcai Zhao

2009.09–2012.06 **MPhil in Computer Science**, *National Digital Switching System Engineering & Technological Research Center (NDSC), Information Engineering University*, 450001, Zhengzhou, China, Master thesis: Research on Optimization Technologies of Parallel Compilation for Distributed Memory Architecture.
Supervised by Prof. Rongcai Zhao

2005.08–2009.07 **BEng**, *Department of Computer Science and Technology, Tsinghua University*, 100084, Beijing, China.

Grants

Huawei 2025 **Principle Investigator**, *Auotmatic Search for Fine-Grained Computation-Communication Overlapping Opportunities*, Huawei Technologies, No. TC20251120013, 2025.12-2026.12.
927,000 CNY

Huawei 2024 **Principle Investigator**, *Intelligent Compilation and Optimization Techniques for Supernode Parallelism Strategies*, Huawei Technologies, No. TC20241115006, 2025.01-2026.12.
2,286,600 CNY

NSFC 2024 **Principle Investigator**, *Research on Key Technologies for the Integrated Architecture for Communication, Perception, and Computation in Edge Supercomputing*, National Natural Science Foundation of China, No. U24A20235, 2025.01-2028.12.
2,600,000 CNY

NSFC 2024 **Principle Investigator**, *Design Automation of Integrated Circuits for Intelligent Chips*, National Natural Science Foundation of China, No. T2422007, 2025.01-2027.12.
2,000,000 CNY

NSFC 2021 **Principle Investigator of the SKL-MEAC part**, *Deep Learning and Tensor Compilers based on the Polyheral Model*, National Natural Science Foundation of China, No. U20A20226, 2021.01-2024.12.
2,600,000 CNY in total for the joint (with Tsinghua University and Beijing OneFlow Research.) project, and 800,000 CNY for the SKL-MEAC part

NSFC 2018 **Researcher**, *Analysis and Optimization of the Precision of Mathematical Functions on Domestic Processors*, National Natural Science Foundation of China, No. 61802434, 2019.01-2021.12.
250,000 CNY

NSFC 2017 **Principle Investigator**, *Polyhedral Compilation Techniques for Heterogeneous Architectures*, National Natural Science Foundation of China, No. 61702546, 2018.01-2020.12.
240,000 CNY

Awards

2025 Distinguished YueLu Scholar (B Class) of the Hunan University

2025 Research Fellow of Beijing Academy of Artificial Intelligence

2023 ACM SIGHPC China Rising Star

2023 Outstanding Young Scholar (A Class) of Renmin University of China

2020 IEEE/ACM MICRO-53 Best Paper Nominees

2019 SKL-MEAC Outerstanding Young Teacher

2019 HPC China 2019 Outerstanding Paper Award

2016 SIGPLAN PLMW Scholarship for OOPSLA 2016

2014 NDSC Scholarship

2013 NDSC's Excellent Master Dissertation Award, 1st place

Languages

Mandarin	**Mothertongue**	*Daily use*
English	**Flent**	*Conversationally fluent*
French	**Basic**	*Basic words and phrases only*

Books

TUP Book **Jie Zhao** and Baoliang Li, *Polyhedral Compilation Theory and its Practice in Deep Learning (in Chinese)*, Tsinghua University Press, 1 November, 2022, 293 pages.

Conference Publications

ASPLOS 2025 Zhanyuan Di, Leping Wang, En Shao, Zhaojia Ma, Ziyi Ren, Feng Hua, Lixian Ma, **Jie Zhao**, Guangming Tan, and Ninhui Sun, *Optimizing Deep Learning Inference Efficiency through Block Dependency Analysis*, In Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2025), Volume 2, 30 March–3 April, 2025, Rotterdam, Netherlands, pages 719–733. (acceptance rate: 160/912=17.5%).

CC 2025 Shaobai Yuan, Jihong He, Yihui Xie, Feng Wang, and **Jie Zhao**, *Post-Link Outlining for Code Size Reduction*, In Proceedings of the 34th ACM SIGPLAN International Conference on Compiler Construction (CC 2025), 1–2 March, 2025, Las Vegas, NV, USA, pages 154–166. (acceptance rate: 17/56=30.4%).

ISSTA 2024 Jinchen Xu, Mengqi Cui, Fei Li, Zuoyan Zhang, Hongru Yang, Bei Zhou, and **Jie Zhao**, *Arfa: an Agile Regime-based Floating-point Optimization Approach for Rounding Errors*, In Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2024), 16-20 September, 2024, Vienna, Austria, 1516-1528. (acceptance rate: 143/694=20.6%).

OSDI 2024 Yi Zhai, Sijia Yang, Keyu Pan, Renwei Zhang, Shuo Liu, Chao Liu, Zichun Liu, Jianmin Ji, **Jie Zhao**, Yu Zhang, and Yanyong Zhang, *Enabling Tensor Language Model to Assist in Generating High-Performance Tensor Programs for Deep Learning*, In Proceedings of the 18th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2024), 10-12 July, 2024, Santa Clara, CA, USA, pages 289-305. (acceptance rate: 44/282=15.6%).

PPoPP 2024 Jinchen Xu, Guanghui Song, Bei Zhou, Fei Li, Jiangwei Hao, and **Jie Zhao**, *A Holistic Approach to Automatic Mixed-Precision Code Generation and Tuning for Affine Programs*, In Proceedings of 29th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP'24), 02-06 March, 2024, Edinburgh, UK, 55-67. (acceptance rate: 32/153=20.9%).

ASE 2023 Zuoyan Zhang, Bei Zhou, Jiangwei Hao, Hongru Yang, Mengqi Cui, Yuchang Zhou, Guanghui Song, Fei Li, Jinchen Xu, and **Jie Zhao**, *Eiffel: Inferring Input Ranges of Significant Floating-point Errors via Polynomial Extrapolation*, In Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering (ASE 2023), 11-15 September, 2023, Kirchberg, Luxembourg, pages 1441-1453. (acceptance rate: 134/629=21.3%).

OSDI 2023 **Jie Zhao**, Siyuan Feng, Xiaoqiang Dan, Fei Liu, Chengke Wang, Sheng Yuan, Wenyuan Lv, and Qikai Xie, *Effectively Scheduling Computational Graphs of Deep Neural Networks toward Their Domain-Specific Accelerators*, In Proceedings of the 17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2023), 10-12 July, 2023, Boston, MA, USA, pages 719-737. (acceptance rate: 50/255=19.6%).

MLSys 2023 Yijin Li, Jiacheng Zhao, Qianqi Sun, Haohui Mai, Lei Chen, Wanlu Chao, Yanfan Chen, Zhicheng Li, Ying Liu, Xiyuan Zhang, Xiyu Shi, **Jie Zhao**, Jingling Xue, Huimin Cui, and Xiaobing Feng, *Sirius: Harvesting Whole-Program Optimization Opportunities for DNNs*, In Proceedings of the 6th Conference on Machine Learning and Systems (MLSys 2023), 4-8 June, 2023, Miami, Florida, USA, pages 377-393. (acceptance rate: 46/207=22.2%).

PACT 2022 **Jie Zhao**, Cédric Bastoul, Yanzhi Yi, Jiahui Hu, Wang Nie, Renwei Zhang, Zhen Geng, Chong Li, Thibaut Tachon, and Zhiliang Gan, *Parallelizing Neural Network Models Effectively on GPU by Implementing Reductions Atomically*, In Proceedings of the 31st International Conference on Parallel Architectures and Compilation Techniques (PACT 2022), 10-12 October, 2022, Chicago, Illinois, USA, pages 451-466. (acceptance rate: 40/118=33.9%).

ICPP 2022 Xiaohan Tao, Yu Zhu, Boyang Wang, Jinlong Xu, Jianmin Pang, and **Jie Zhao**, *Automatically Generating High-performance Matrix Multiplication Kernels on the Latest Sunway Processor*, In Proceedings of the 51st International Conference on Parallel Processing (ICPP 2022), 29 August-1 September, 2022, Online Event, Bordeaux, France, Article No. 52, 12 pages. (acceptance rate: 84/311=27.0%).

MLSys 2022 **Jie Zhao**, Xiong Gao, Ruijie Xia, Zhaochuang Zhang, Deshi Chen, Lei Chen, Renwei Zhang, Zhen Geng, Bin Cheng, and Xuefeng Jin, *Apollo: Automatic Partition-based Operator Fusion through Layer by Layer Optimization*, In Proceedings of the 5th Conference on Machine Learning and Systems (MLSys 2022), 29 August - 1 September, 2022, Santa Clara, pages 1-19. (acceptance rate: 51/247=20.6%).

PLDI 2021 **Jie Zhao**, Bojie Li, Wang Nie, Zhen Geng, Renwei Zhang, Xiong Gao, Bin Cheng, Chen Wu, Yun Cheng, Zheng Li, Peng Di, Kun Zhang, and Xuefeng Jin, *AKG: Automatic Kernel Generation for Neural Processing Units using Polyhedral Transformations*, In Proceedings of the 42nd ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2021), 20-25 June, 2021, Vitural, Canada, pages 1233-1248. (acceptance rate: 87/320=27.2%).

MICRO 2020 **Jie Zhao** and Peng Di, *Optimizing the Memory Hierarchy by Compositing Automatic Transformations on Computations and Data*, In Proceedings of the 53rd IEEE/ACM International Symposium on Microarchitecture (MICRO-53), 17-21 October, 2020, Global Online Event, pages 427-441. (acceptance rate: 82/446=18.4%).

ICS 2019 HuiHui Sun, Florian Fey, **Jie Zhao**, and Sergei Gorlatch, *WCCV: Improving the Vectorization of IF-statements with Warp-coherent Conditions*, In Proceedings of the 33rd ACM International Conference on Supercomputing (ICS 2019), 26-28 June, 2019, Phoenix, Arizona, USA, pages 319-329. (acceptance rate: 45/193=23.3%).

CC 2018 **Jie Zhao**, Micheal Kurse, and Albert Cohen, *A Polyhedral Compilation Framework for Loops with Dynamic Data-dependent Bounds*, In Proceedings of the 27th International Conference on Compiler Construction (CC 2018), 24-25 February, 2018, Vienna, Austria, pages 14-24. (acceptance rate: 18/52=34.6%).

SKG 2012 Baoliang Li, **Jie Zhao**, Junhui Wang, and Wenhua Dou, *A max-plus algebra approach for network-on-chip end-to-end delay estimation*, In Proceedings of the 8th International Conference on Semantics, Knowledge and Grids (SKG 2012), 22-24 October 2012, Beijing, China, pages 217-220.

HPCC 2012 **Jie Zhao**, Rongcai Zhao, and Lin Han, *A nonlinear array subscript dependence test*, In Proceedings of the IEEE 14th International Conference on High Performance Computing and Communication (HPCC 2012), 24-25 February, 2012, Liverpool, UK, pages 764-771.

Journal Publications

TACO 2025 Jiawei Tan, Jiapeng Zhang, Zhuo Tang, Xiong Xiao, Bingting Jiang, **Jie Zhao**, and Kenli Li, *ASSG: Enhanced Workload Balancing via Adaptive State Scheduling Granularity Approach for Stateful Distributed Stream Processing*, ACM Transactions on Architecture and Code Optimization, 2025, 22(4): Article 160, 26 pages.

TACO 2025 Zhanyuan Di, Leping Wang, Zhaojia Ma, En Shao, **Jie Zhao**, Ziyi Ren, Siyuan Feng, Dingwen Tao, Guangming Tan, and Ninghui Sun, *Accelerating Parallel Structures in DNNs via Parallel Fusion and Operator Co-Optimization*, ACM Transactions on Architecture and Code Optimization, 2025, 21(3): Article 92, 26 pages.

TOCS 2023 **Jie Zhao**, Jinchen Xu, Peng Di, Wang Nie, Jiahui Hu, Yanzhi Yi, Sijia Yang, Zhen Geng, Renwei Zhang, Bojie Li, Zhiliang Gan, and Albert Cohen, *Modeling the Interplay between Loop Tiling and Fusion in Optimizing Compilers using Affine Relations*, ACM Transactions on Computer Systems, 2023, 41(1-4): Article 5, 45 pages.

TACO 2019 **Jie Zhao** and Albert Cohen, *Flextended tiles: a flexible extension of overlapped tiles for polyhedral compilation*, ACM Transactions on Architecture and Code Optimization, 2019, 16(4): Article 47, 25 pages.

Journal of Supercomputing 2018 **Jie Zhao** and Rongcai Zhao, *K-DT: A formal system for the evaluation of linear data dependence testing techniques*, The Journal of Supercomputing, 2018, 71(1): 340-368.

Science China 2017 **Jie Zhao** and Rongcai Zhao, *Identifying superword level parallelism with directed graph reachability*, Science China: Information Sciences, 2017, 60(1): 019103.

Computer Journal 2016 **Jie Zhao**, Rongcai Zhao, and Jinchen Xu, *Code generation for distributed-memory architectures*, The Computer Journal, 2016, 59(1): 119-132.

Journal of Supercomputing 2015 **Jie Zhao**, Rongcai Zhao, Xi Chen and Bo Zhao, *An improved nonlinear data dependence test*, The Journal of Supercomputing, 2015, 71(1): 340-368.

IET Software 2013 **Jie Zhao**, Rongcai Zhao, Lin Han, and Jinlong Xu, *QP test: a dependence test for quadratic array subscripts*, IET Software, 2013, 7(5): 271-282.

Workshops, Posters, Preprints, and Others

PPoPP 2025 Poster Zhanyuan Di, Leping Wang, Ziyi Ren, En Shao, **Jie Zhao** Siyuan Feng, Dingwen Tao, Guangming Tan, and Ninghui Sun, *Poster: Magneto: Accelerating Parallel Structures in DNNs via Co-Optimization of Operators*, In Proceedings of 30th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP 2025), 1–5 March, 2025, Las Vegas, NV, USA, pages 563–565. (poster acceptance rate: 49/189=25.9%).

arXiv 2021 Jinhui Yuan, Xinqi Li, Cheng Cheng, Juncheng Liu, Ran Guo, Shenghang Cai, Chi Yao, Fei Yang, Xiaodong Yi, Chuan Wu, Haoran Zhang and **Jie Zhao**, *OneFlow: Redesign the Distributed Deep Learning Framework from Scratch*, arXiv:2110.15032v6 [cs.DC], https://doi.org/10.48550/arXiv.2110.15032.

CGO 2018 SRC **Jie Zhao**, *A general purpose automatic overlapped tiling technique in polyhedral frameworks*, Student Research Competition at 16th IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2018), 24-25 February, 2018, Vienna, Austria.

IMPACT 2017 **Jie Zhao** and Albert Cohen, *A general compilation algorithm to parallelize and optimize counted loops with dynamic data-dependent bounds*, the 7th International Workshop on Polyhedral Compilation Techniques (IMPACT 2017), 23 January, 2017, Stockholm, Sweden.

Teaching Experiences

CP 2025 **Principle Instructor of Compiler Principles**, *Undergraduate-level*, 52/52 teaching hours, fall 2025.

CP 2024 **Principle Instructor of Compiler Principles**, *Undergraduate-level*, 52/52 teaching hours, fall 2024.

CA 2020 **Principle Instructor of Computer Architecture**, *Undergraduate-level*, 34/60 teaching hours, spring 2020.

PC 2022 **Guest Instructor of Principles of Compiler**, *Undergraduate-level*, 2/40 teaching hours, fall 2022, invited by the National University of Defense and Technology.

ACA 2021 **Guest Instructor of Advanced Computer Architecture**, *Graduate-level*, 2/40 teaching hours, fall 2021.

ACA 2019 **Guest Instructor of Advanced Computer Architecture**, *Graduate-level*, 2/40 teaching hours, fall 2019.

Graduated Student Mentoring

MEng 2024 **Mentoring Wei Zhao's MEng thesis in 2024**, *Zhengzhou University*, MEng thesis: Research on Operator Fusion and Optimization Techniques Based on LLVM, Defensed in June 2024.
Co-advise with Prof. Lin Han

MEng 2024 **Mentoring Zuoyan Zhang's MEng thesis in 2024**, *Information Engineering University*, MEng thesis: Research on Error Detection Methods for Floating-point Arithmetic Expressions, Defensed in June 2024.
Co-advise with Prof. Shaozhong Guo

MEng 2023 **Mentoring Xiaoyue Xu's MEng thesis in 2023**, *Zhengzhou University*, MEng thesis: Research on Automatic Generation of DCU Kernel Code Based on LLVM, Defensed in June 2023.
Co-advise with Prof. Lin Han

Mphil 2023 **Mentoring Guanghui Song's Mphil thesis in 2023**, *Information Engineering University*, Mphil thesis: Automatic Generation and Optimization of Mixed-precision Computation for Loop Nested Code, Defensed in June 2023.
Co-advise with Prof. Shaozhong Guo

PhD 2022 **Mentoring Xiaohan Tao's PhD thesis in 2022**, *Information Engineering University*, PhD thesis: High Performance Code Generation and Optimization for Sunway Heterogeneous Many-core Architecture, Defensed in June 2022.
Co-advise with Prof. Jianmin Pang

Mphil 2022 **Mentoring Boyang Wang's Mphil thesis in 2022**, *Zhengzhou University*, Mphil thesis: Vector Code Generation based on Polyhedron Model, Defensed in June 2022.
Co-advise with Prof. Jianmin Pang

PhD 2021 **Mentoring Yingying Li's PhD thesis in 2021**, *Information Engineering University*, PhD thesis: Research on Key Techniques of Polyhedral Compiler Optimizations for Heterogeneous Systems, Defensed in June 2021.
Co-advise with Prof. Jianmin Pang

Scholar Services

- Program Committee Co-Chairs for IMPACT 2024
- Program Committee for CGO 2026, CC 2026
- Program Committee for IMPACT 2025, 2023, 2022 and 2021
- Reviewers for ACM Transactions on Architecture and Code Optimization, ACM Transactions on Embedded Computing Systems, IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Elsevier Parallel Computing, , Springer Journal of Computer Science and Technology, Springer Journal of Supercomputing, IET Software
- Reviewers of National Natural Science Foundation of China (2019 and 2025)