

赵捷

Curriculum Vitae

计算机科学系
湖南大学信息科学与工程学院
长沙市岳麓区麓山南路麓山门
410082 中华人民共和国
✉ jiezhaoh@hnu.edu.cn
🌐 <https://yaozhujia.github.io/>



个人简介

赵捷长期从事编译优化相关的研究，致力于通过编译器优化实现不同应用在各种平台上的高效实现，建立上层应用与底层体系结构之间的桥梁软件。更具体地，赵捷的研究兴趣包括：

- 代码生成与优化
- 深度学习编译系统
- 数值程序分析

工作经历

- 2024.01–今 **教授**，计算机科学系，湖南大学信息科学与工程学院，410082 长沙.
- 2023.09–
2023.12 **副教授**，云计算与大数据系统实验室，中国人民大学信息学院，100872 北京.
- 2019.01–
2023.07 **讲师**，高性能计算教研室，中国人民解放军战略支援部队信息工程大学网络空间安全学院，450001 郑州.

交流经历

- 2023.08–今 **学术合作**，华为技术有限公司北京研究所，北京.
- 2021.04–
2023.04 **学术合作**，希姆计算科技有限公司，杭州.
- 2019.02–
2021.12 **学术合作**，华为技术有限公司杭州研究所，杭州.
- 2020.04–
2021.04 **学术合作**，阿里巴巴集团，北京.
- 2018.05–
2018.09 **学术合作**，华为技术有限公司法国研究所，巴黎.

教育背景

- 2015.03–
2018.12 **数学专业博士**，巴黎高等师范学校/INRIA, 75005 巴黎，法国，导师：Professor Albert Cohen.
博士论文题目：A Combined Language and Polyhedral Approach for Heterogeneous Parallelism
- 2012.09–
2016.06 **计算机专业博士**，中国人民解放军信息工程大学网络空间安全学院，450001 郑州，导师：赵荣彩教授.
博士论文题目：并行化编译中的依赖关系分析
- 2009.09–
2012.06 **计算机专业硕士**，中国人民解放军信息工程大学信息工程学院，450001 郑州，导师：赵荣彩教授.
硕士论文题目：面向分布存储的编译优化技术研究
- 2005.08–
2009.07 **计算机专业学士**，清华大学计算机科学与技术系，100084 北京.

荣誉获奖

- 2023.10 ACM SIGHPC China 新星奖
- 2023.09 中国人民大学杰出青年学者 A 岗
- 2021.05 中原信息化“专家智库贡献奖”二等奖
- 2020.10 **IEEE/ACM MICRO-53 最佳论文提名 (国内第三次)**
- 2019.12 信息工程大学四院新教员教学能力竞赛优秀奖
- 2019.10 信息工程大学学术论文成果三等奖
- 2019.08 HPC China 2019 优秀论文奖
- 2017.01 信息工程大学优秀博士学位论文奖
- 2016.04 ACM SIGPLAN PLMW Scholarship for OOPSLA 2016
- 2014.11 **军队科技进步二等奖, 序 7, XX 计算机基础数学库系统**
- 2014.11 军队公派出国留学资助奖学金, 赴法国攻读博士学位
- 2013.07 **军队优秀硕士学位论文奖**
- 2013.01 信息工程大学优秀硕士学位论文奖

参与课题

- 某重大工程项目 2024-2027 **某重大工程, XXX 编译与代码翻译, 2024.03-2027.07, 子课题负责人, 总经费 1200 万 (子课题部分 300 万), 在研.**
- 区域创新发展联合项目 2021-2024 **国家自然科学基金, 基于多面体模型的深度学习张量编译器, 2021.01-2024.12, 信息工程大学部分负责人, 总经费 260 万 (信息工程大学部分 80 万), 在研.**
清华大学牵头, 信息工程大学、北京一流科技有限公司参与 (该项目按照重点项目标准资助)
- 青年基金 2019-2021 **国家自然科学基金, 国产处理器数学函数精度分析与优化, 2019.01-2021.12, 一般参与人, 25 万, 在研.**
- 青年基金 2018-2020 **国家自然科学基金, 面向异构系统的多面体编译优化技术, 2018.01-2020.12, 项目负责人, 24 万, 结题.**
- 开放课题 2013-2015 **数学工程与先进计算国家重点实验室开放课题, 面向多核/众核架构的新型编程模型, 2013.11-2015.10, 主要完成人, 8 万, 结题.**
- 专项子课题 2008-2013 **国家重大专项工程子课题, 并行识别与自动并行化系统, 2008.07-2013.06, 主要完成人, 480 万, 结题.**
- 专项子课题 2008-2013 **国家重大专项工程子课题, XX 计算机基础数学库系统, 2008.07-2013.06, 一般参与人, 160 万, 结题.**
依托该项目的成果获得 2014 年军队科技进步二等奖

科研成果

出版专著

- TUP Book **赵捷, 李宝亮, 多面体编译理论与深度学习实践, 清华大学出版社, 2022 年 11 月 1 日出版, 293 页.**
共计 439 千字, 负责约 350 千字

会议论文 (英文)

- 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, 16-20 September, 2024, Vienna, Austria, to appear. (acceptance rate: 143/694=20.6%).
CCF A 类会议, 软件工程领域顶级会议

- 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%).
CCF A 类会议, 计算机系统领域顶级会议
- 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%).
CCF A 类会议, 高性能计算与并行编程领域顶级会议
- 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%).
CCF A 类会议, 软件工程领域顶级会议
- OSDI 2023 **Jie Zhao**, Siyuan Feng, Xiaoqiang Dan, Fei Liu, Chengke Wang, Sheng Yuan, Wen Yuan 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%).
CCF A 类会议, 计算机系统领域顶级会议
- 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%).
机器学习与系统软件领域顶级会议
- 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%).
CCF B 类会议, 体系结构与编译器领域顶级会议
- 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, Bordeaux, France, Article No. 52, pages 1-12. (acceptance rate: 84/311=27.0%).
CCF B 类会议, 分布式与并行处理高水平会议

- 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, California, USA, pages 1-19. (acceptance rate: 51/247=20.6%).
机器学习与系统软件领域顶级会议
- 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, Vitual, Canada, pages 1233-1248. (acceptance rate: 87/320=27.2%).
CCF A 类会议，程序设计语言与编译器领域顶级会议
- 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%).
CCF A 类会议，体系结构与编译器领域顶级会议，最佳论文提名（该会议 53 年历史上国内科研单位第三次获此提名）
- 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'19), 26-28 June, 2019, Phoenix, Arizona, USA, pages 319-329. (acceptance rate: 45/193=23.3%).
CCF B 类会议，高性能计算与编译器领域高水平会议
- 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'18), 24-25 February, 2018, Vienna, Austria, pages 14-24. (acceptance rate: 18/52=34.6%).
CCF B 类会议，编译器领域高水平会议（注：该会议原来是 CCF B 类会议群 ETAPS 下的一个会议，现已独立召开）
- 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), 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), 24-25 February, 2012, Liverpool, UK, pages 764-771.
CCF C 类会议
- AHPCN 2012 **Jie Zhao**, Rongcai Zhao, and Yuan Yao, *A nested loop fusion algorithm based on cost analysis*, In Proceedings of 5th International Symposium on Advances of High Performance Computing and Networking (AHPCN), 24-25 February, 2012, Liverpool, UK, 1096-1101.
[会议论文 \(中文\)](#)
- HPC China 2019 李颖颖, **赵捷**, 庞建民, 多面体模型中分裂分块算法的设计与实现, 2019 全国高性能计算学术年会大会论文集, 2019 年 8 月 22 日-24 日, 130-141, 中国, 内蒙古, 呼和浩特, 130-141 页.
大会优秀论文奖

期刊论文 (英文)

- 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 Xuefeng Jin, *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.
CCF A 类期刊, SCI、EI 检索
- TACO 2019 **Jie Zhao** and Albert Cohen, *Flexextended tiles: a flexible extension of overlapped tiles for polyhedral compilation*, ACM Transactions on Architecture and Code Optimization, 2019, 16(4): Article 47, 25 pages.
CCF A 类期刊, SCI、EI 检索
- 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.
CCF C 类期刊, SCI、EI 检索
- Science China 2017 **Jie Zhao** and Rongcai Zhao, *Identifying superword level parallelism with directed graph reachability*, Science China: Information Sciences, 2017, 60(1): 019103.
CCF A 类期刊, SCI、EI 检索
- Computer Journal 2016 **Jie Zhao**, Rongcai Zhao, and Jinchen Xu, *Code generation for distributed-memory architectures*, The Computer Journal, 2016, 59(1): 119-132.
CCF B 类期刊, SCI、EI 检索
- 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.
CCF C 类期刊, SCI、EI 检索
- 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.
CCF B 类期刊, SCI、EI 检索

期刊论文 (中文)

- 软件学报 2023 宋广辉, 郭绍忠, **赵捷**, 陶小涵, 李飞, 许瑾晨, 面向 Stencil 计算的自动混合精度优化, 软件学报, 2023, 34(12): 5704-5723.
- 软件学报 2023 陶小涵, 朱雨, 庞建民, **赵捷**, 徐金龙, 面向申威异构架构的并行代码自动生成, 软件学报, 2023, 34(4): 1570-1593.
- 计算机学报 2020 李颖颖, **赵捷**, 庞建民, 多面体模型中分裂分块算法的设计与实现, 计算机学报, 2020, 43(6): 1038-1051.
- 软件学报 2019 李雁冰, 赵荣彩, 韩林, **赵捷**, 徐金龙, 李颖颖, 一种面向异构众核处理器的并行编译框架, 软件学报, 2019, 30(4): 981-1001.
- 软件学报 2018 **赵捷**, 李颖颖, 赵荣彩, 基于多面体模型的编译“黑魔法”, 软件学报, 2018, 29(8): 2371-2396.
- 中国科学: 信息科学 2017 **赵捷**, 赵荣彩, 基于有向图可达性的 SLP 向量化识别方法, 中国科学: 信息科学, 2017, 47(3): 310-325.
- 软件学报 2015 许瑾晨, 黄永忠, 郭绍忠, 周蓓, **赵捷**, 一个浮点数学函数库测试平台, 软件学报, 2015, 26(6): 1306-1321.
- 计算机学报 2014 **赵捷**, 赵荣彩, 韩林, 李宝亮, 面向 MPI 代码生成的 Open64 编译器后端, 计算机学报, 2014, 37(7): 1620-1632.
- 软件学报 2014 刘晓嫻, 赵荣彩, **赵捷**, 徐金龙, 面向 DOACROSS 循环的流水并行代码自动生成, 软件学报, 2014, 25(6): 1154-1168.
- 软件学报 2013 李雁冰, 赵荣彩, 刘晓嫻, **赵捷**, 面向 OpenMP 自动并行化的代价模型, 软件学报 (增刊), 2014, 25(S2): 101-110.

软件学报 **赵捷**, 赵荣彩, 丁锐, 黄品丰, 基于嵌套循环分类的并行识别技术, 软件学报, 2012, 23(10): 2695-2704.

Workshops 及 Preprint

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>, 13 pages.

CGO 2018 **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), 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), 23 January, 2017, Stockholm, Sweden, 10 pages.

面向多面体编译模型的研讨会

专利发明

国家发明专利 2015 单征, 庞建民, 姚远, **赵捷**, 丁锐, 刘晓娟, 基于二次规划模型的非线性下标表达式的依赖测试方法, 2015.12.09, 中国, ZL201310147595.9.

教学经验

CA 2020 **计算机体系结构主讲教师**, 本科生学位课, 共 60 学时, 授课 34 学时, 战略支援部队信息工程大学计算机专业 2020 年春季学期.

PC 2022 **编译原理客座讲师**, 本科生学位课, 共 40 学时, 授课 2 学时, 国防科技大学计算机专业 2022 年秋季学期.

ACA 2021 **高级计算机体系结构客座讲师**, 研究生学位课, 共 40 学时, 授课 2 学时, 战略支援部队信息工程大学计算机专业 2021 年秋季学期.

ACA 2019 **高级计算机体系结构客座讲师**, 研究生学位课, 共 40 学时, 授课 2 学时, 战略支援部队信息工程大学计算机专业 2019 年秋季学期.

研究生指导

硕士论文 2024 **指导赵薇硕士生论文**, 郑州大学, 硕士论文题目: 基于 TVM 的算子融合与优化技术研究, 已于 2024 年 6 月答辩.

指导教师: 赵捷博士

硕士论文 2024 **协助指导张作言硕士生论文**, 战略支援部队信息工程大学, 硕士论文题目: 浮点算术表达式的误差检测方法研究, 已于 2024 年 6 月答辩.

指导教师: 郭绍忠教授

硕士论文 2023 **指导许啸跃硕士生论文**, 郑州大学, 硕士论文题目: 基于 LLVM 的 DCU 内核代码自动生成技术研究, 已于 2023 年 6 月答辩.

指导教师: 赵捷博士

硕士论文 2023 **协助指导宋广辉硕士生论文**, 战略支援部队信息工程大学, 硕士论文题目: 面向循环嵌套的混合精度计算自动生成与优化, 已于 2023 年 6 月答辩.

指导教师: 郭绍忠教授

博士论文 2022 **协助指导陶小涵博士论文**, 战略支援部队信息工程大学, 博士论文题目: 面向申威异构众核平台的高性能代码生成与优化, 已于 2022 年 6 月答辩.

指导教师: 庞建民教授

- 硕士论文 **协助指导王博漾硕士论文**, 郑州大学, 硕士论文题目: 基于多面体模型的 SIMD 向量化编译技术研究, 已于 2022 年 6 月答辩.
指导教师: 庞建民教授
- 博士论文 **协助指导李颖颖博士论文**, 战略支援部队信息工程大学, 博士论文题目: 面向异构系统的多面体编译优化关键技术研究, 已于 2021 年 6 月答辩.
指导教师: 庞建民教授

学术活动

- 开源社区贡献
 - 华为 MindSpore 框架社区专家委员会成员
- 国际期刊审稿
 - «IEEE Transactions on Computers»
 - «IEEE Transactions on Parallel and Distributed Systems»
 - «Parallel Computing»
 - «ACM Transactions on Architecture and Code Optimization»
 - «Journal of Computer Science and Technology»
 - «IET Software»
 - «Journal of Supercomputing»
- 国内期刊审稿
 - « 计算机学报 »
 - « 计算机研究与发展 »
 - « 电子学报 »
- 自然科学基金项目评审专家
 - 2019 信息科学面上项目