(+86)156-1611-1252
Beijing, China
jameschennerd@gmail.com

Dong Chen

HomePage: dongchen-coder.github.io
Google Scholar: Dong Chen

My research centers on program analysis for correctness and performance, spanning both symbolic and neural methods. I lead teams working on program synthesis, small language model training and LLM infrastructure. I am also interested in system software, advanced computing paradigms including optical and quantum computing, and programming language theories.

Education

Ph.D in Computer Science, *University of Rochester* 2014.09-2019.05
BS/MS in Computer Science, *National University of Defense Technology* 2007.09-2013.12

Experience

Research Software Engineer / Tech Lead (Top Minds Program) **2022.05-**
Huawei *Beijing, China*
Assistant Professor **2019.06-2021.05**
National University of Defense Technology *Changsha, China*
Intern(x2) **2016.06-2016.08, 2018.06-2018.08**
Qualcomm, Graphics Compiler Team *CA, USA*
Intern **2015.06-2015.08**
FutureWei Technologies, Compiler Team *CA, USA*

Skills

Tools and Languages C++, Python, Parallel Programming, LLVM
Communication Chinese (native), English (working proficiency)

Projects

SLM Training (Team Leader) **2025.06–**
- RL, finetuning, continuous pre-training, evaluation with Qwen3 models for running performance summarization.
- Investigating structure pruning and new model architectures (rwkv, mamba, mixtrue-of-lookup-experts) for edge devices.

LLM Infra (Team Leader) **2025.01–**
- Auto-tuning framework for triton-based kernels with dynamic shapes, integrated into the vLLM framework and evaluated on LLaMA3.1 and Qwen3 models, achieving average end-to-end speedups of 1.37× and 1.42× respectively, under dynamic batching workloads.
- Implementing a tile-level distributed computation and communication overlapping inference/training framework for dense/MoE models.

Coding Agent and Program Synthesis **2024.01–**
- *GPU Kernel Optimization*: Designed a Monte Carlo Tree Search (MCTS)–based agent for automated GPU kernel optimization.
- *Automatic GitHub Issue Resolution*: Developed a task-graph–based multi-agent framework enabling precise plan execution; achieved state-of-the-art performance on SWE-bench-lite, resolving 28.33% of issues (June 4–17, 2024).
- *Program Synthesis for Locality Analysis*: Proposed and implemented an input–output–example–driven, syntax-guided synthesis framework for program locality analysis; designed a domain-specific language (DSL) and a unification-based search algorithm to efficiently explore the program space.

Static Analysis for Memory Safety **2022.05-2024.12**
- explores techniques to reason about program properties automatically (sparse-value flow analysis, abstract interpretation, etc).
- implements tools to identify memory bugs for large-scale industrial codes, such as null pointer dereference, memory leaks, etc.

Compiler Leasing **2019.01-2022.12**
- proposes a framework that enables fine-grained control of data replacements in a cache by a compiler.
- designs and implements an algorithm to derive optimal leases for each reference in a program to minimize cache misses.

Static Sampling for Locality Analysis **2018.05-2021.12**
- designs and implements an LLVM compiler pass that predicts the cache performance of loop nests. It specializes the loops to enable static profiling of reuse intervals.

Write Locality **2016.01-2016.12**
- designs and implements a linear-time algorithm to model cache writebacks from the memory access trace of a program.
- implements a scheduling algorithm to minimize writebacks by grouping co-running programs, with the writeback model.

OpenCL Performance portability **2012.01-2013.12**
- designs a source-to-source translator based on LLVM infrastructure. It automatically transforms OpenCL kernel for GPU with fine-grained parallelism to vectorized code for CPU.

Publications

[**DATE'26**] Yuhan Kang, Wenrui Zhang, *Dong Chen*, Yang Shi, Jianchao Yang, Zeyu Xue, Jing Feng and Mei Wen. "DyGen: A Constant-Time Kernel Generator for Dynamic-Shape Neural Networks". 29th Design, Automation and Test in Europe Conference.
[**ICSE-SEIP'26**] Chaofan Wang, Tingrui Yu, Chen Xie, Jie Wang, *Dong Chen*, Wenrui Zhang, Yuling Shi, Xiaodong Gu, Beijun Shen. "EVOC2RUST: A Skeleton-guided Framework for Project-Level C-to-Rust Translation". 48th International Conference on Software Engineering, Software Engineering in Practice Track.
[**LMPL'25**] Ting Yuan, Wenrui Zhang, *Dong Chen*, Jie Wang. "CG-Bench: Can Language Models Assist Call Graph Construction in the Real World?". Proceedings of the 1st ACM SIGPLAN International Workshop on Language Models and Programming Languages
[**ACL-findings'25**] Linhao Zhang, Daoguang Zan, Quanshun Yang, Zhirong Huang, Dong Chen, Bo Shen, Tianyu Liu, Yongshun Gong, Pengjie Huang, Xudong Lu, Guangtai Liang, Lizhen Cui, Qianxiang Wang. "CodeV: Issue Resolving with Visual Data" Findings of the Association for Computational Linguistics: ACL 2025, 7350–7361
[**TechReport'24**] Dong Chen, Shaoxin Lin, Muhan Zeng, Daoguang Zan, Jian-Gang Wang, Anton Cheshkov, Jun Sun et al. "CodeR: Issue Resolving with Multi-Agent and Task Graphs". arXiv preprint arXiv:2406.01304 (2024).
[**TechReport'24**] Daoguang Zan, Zhirong Huang, Ailun Yu, Shaoxin Lin, Yifan Shi, Wei Liu, *Dong Chen* et al. "SWE-bench-java: A GitHub Issue Resolving Benchmark for Java". arXiv preprint arXiv:2408.14354 (2024).
[**TechReport'24**] Wenrui Zhang, Tiehang Fu, Ting Yuan, Ge Zhang, Dong Chen, and Jie Wang. "A Lightweight Framework for Adaptive Retrieval In Code Completion With Critique Model." arXiv preprint arXiv:2406.10263 (2024).
[**TOSEM24, co-first author**] Daoguang Zan, Ailun Yu, Wei Liu, *Dong Chen*, Bo Shen, Wei Li, Yafen Yao, Yongshun Gong, Xiaolin Chen, Bei Guan, Zhiguang Yang, Yongji Wang, Qianxiang Wang, Lizhen Cui. "CodeS: Natural Language to Code Repository via Multi-Layer Sketch". https://arxiv.org/abs/2403.16443
[**MICRO24**] Jianchao Yang, Mei Wen, Dong Chen, Zhaoyun Chen, Zeyu Xue, Yuhang Li, Junzhong Shen, Yang Shi. "HyFiSS: A Hybrid Fidelity Stall-Aware Simulator for GPGPUs". 57th Annual IEEE/ACM International Symposium on Microarchitecture.
[**JSA23**] Hao Ming, Tingting Pan, *Dong Chen*, Chencheng Ye, Haikun Liu, Liting Tang, Xiaofei Liao and Hai Jin. "VIDGCN: Embracing Input Data Diversity with A Configurable Graph Convolutional Network Accelerator". Journal of Systems Architecture.
[**TACO22**] Chen Ding, *Dong Chen*, Fangzhou Liu, Benjamin Reber, Wesley Smith. "CARL: Compiler Assigned Reference Leasing". ACM Transactions on Architecture and Code Optimization.
[**LCPC21**] *Dong Chen*, Chen Ding, Dorin Patru. "CLAM: Compiler Leasing of Accelerator Memory". 32nd Workshop on Languages and Compilers for Parallel Computing.
[**ISMM21**] *Dong Chen*, Chen Ding, Fangzhou Liu, Benjamin Reber, Wesley Smith, and Pengcheng Li. "Uniform Lease vs LRU Cache: Analysis and Evaluation". The 2021 ACM SIGPLAN International Symposium on Memory Management.
[**MEMSYS20**] Ian Prechtl, Ben Reber, Chen Ding, Dorin Patru, *Dong Chen*. "CLAM: Compiler Lease of Cache Memory". The 6th International Symposium on Memory Systems.
[**PPoPP20p**] Fangzhou Liu, *Dong Chen*, Wesley Smith, and Chen Ding. "PLUM: static parallel program locality analysis under uniform multiplexing". 25th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (Poster).
[**PhD Thesis**] *Dong Chen*. "Program locality analysis based on reuse intervals". University of Rochester, 2019.
[**LCPC19**] *Dong Chen*, Chen Ding, and Dorin Patru. "CLAM: Compiler leasing of accelerator memory." Languages and Compilers for Parallel Computing: 32nd International Workshop, LCPC 2019, Atlanta, GA, USA, October 22–24, 2019, Revised Selected Papers 32, pp. 89-97. Springer International Publishing, 2021.
[**MEMSYS19**] *Dong Chen*, Fangzhou Liu, Mingyang Jiao, Chen Ding, Sreepathi Pai. "Statistical Caching for Near Memory Management". 5th International Symposium on Memory Systems.
[**PLDI18**] *Dong Chen*, Fangzhou Liu, Chen Ding, Sreepathi Pai. "Locality analysis through static parallel sampling". 39th ACM SIGPLAN Conference on Programming Language Design and Implementation. (Artifact evaluated).
[**LCPC18**] *Dong Chen*, Chunling Hu, Chucheow Lim, Sreepathi Pai, Chen Ding. "POSTER: Static Sampling for GPU Code". 31th International Workshop on Languages and Compilers for Parallel Computing.
[**LCPC17**] *Dong Chen*, Fangzhou Liu, Chen Ding, Chucheow Lim. "POSTER: Static Reuse Time Analysis Using Dependence Distance". 30th International Workshop on Languages and Compilers for Parallel Computing.
[**TACO17**] Chencheng Ye, Chen Ding, Hao Luo, Jacob Brock, *Dong Chen*, Hai Jin. "Cache Exclusivity and Sharing: Theory and Optimization". ACM Transactions on Architecture and Code Optimization.
[**TACO17**] Pengcheng Li, Xiaoyu Hu, *Dong Chen*, Jacob Brock, Hao Luo, Eddy Z Zhang, Chen Ding. "LD: Low-Overhead GPU Race Detection Without Access Monitoring". ACM Transactions on Architecture and Code Optimization.
[**MEMSYS16**] *Dong Chen*, Chencheng Ye, Chen Ding. "Write Locality and Optimization for Persistent Memory". 2nd International Symposium on Memory Systems
[**Frontiers15**] Mei Wen, Dafei Huang, Changqing Xun, *Dong Chen*. "Improving performance portability for GPU-specific OpenCL kernels on multi-core/many-core CPUs by analysis-based transformations". Frontiers of Information Technology & Electronic Engineering Vol.16 No.11 P.899-916
[**EuroPar14**] Dafei Huang, Mei Wen, Changqing Xun, *Dong Chen*, Xing Cai, Yuran Qiao, Nan Wu, Chunyuan Zhang. "Automated Transformation of GPU-Specific OpenCL Kernels Targeting Performance Portability on Multi-Core/Many-Core CPUs". 20th International European Conference on Parallel and Distributed Computing.

[**JZUS13**] Changqing Xun, *Dong Chen*, Qiang Lan, and Chunyuan Zhang. "Efficient fine-grained shared buffer management for multiple OpenCL devices". Journal of Zhejiang University Science C 14, no. 11 (2013): 859-872.
[**AMM13**] *Dong Chen*, Hua You Su, Wen Mei, Li Xuan Wang, and Chun Yuan Zhang. "Scalable parallel motion estimation on multi-GPU system". Applied Mechanics and Materials 347 (2013): 3708-3714.
[**HPCC13**] *Dong Chen*, Changqing Xun, Dafei Huang, Mei Wen, Chunyuan Zhang. “Automatic mapping single-device OpenCL program to heterogeneous multi-device platform". 15th Conference on High-Performance Computing and Communications.

Professional Activities

Professional Services: Artifact Evaluation Committee for OOPSLA26, POPL26, POPL25, OOPSLA25, PLDI25. Reviewer for DL4C@NeurIPS25, VerifAI@ICLR25, JCST. Sub-reviewer for MEMSYS19, ICS19, LCPC18, ICS17, MEMSYS17, NPC17.
Teaching Assistant: Data Structure, Programming Language Design and Implementation, Advanced Compiler.