

2024 IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2024)

**Edinburgh, United Kingdom
2-6 March 2024**



**IEEE Catalog Number: CFP24CGO-POD
ISBN: 979-8-3503-9510-5**

**Copyright © 2024 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP24CGO-POD
ISBN (Print-On-Demand):	979-8-3503-9510-5
ISBN (Online):	979-8-3503-9509-9
ISSN:	1931-0544

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Contents

Frontmatter

Welcome from the General Chairs	iii
Welcome from the Program Chairs	v
CGO 2024 Organization	vii
CGO 2024 Sponsors and Supporters	xi

Compilers for Machine Learning

A Tensor Algebra Compiler for Sparse Differentiation

Amir Shaikhha, Mathieu Huot, and Shideh Hashemian — <i>University of Edinburgh, United Kingdom; University of Oxford, United Kingdom</i>	1
--	---

Energy-Aware Tile Size Selection for Affine Programs on GPUs

Malith Jayaweera, Martin Kong, Yanzhi Wang, and David Kaeli — <i>Northeastern University, USA; Ohio State University, USA</i>	13
---	----

PolyTOPS: Reconfigurable and Flexible Polyhedral Scheduler

Gianpietro Consolaro, Zhen Zhang, Harenome Razanajato, Nelson Lossing, Nassim Tchoulak, Adilla Susungi, Artur Cesar Araujo Alves, Renwei Zhang, Denis Barthou, Corinne Ancourt, and Cédric Bastoul — <i>Huawei Technologies, France; Mines Paris-PSL, France; Huawei Technologies, China</i>	28
--	----

Machine-Learning Guided Optimizations

AskIt: Unified Programming Interface for Programming with Large Language Models

Katsumi Okuda and Saman Amarasinghe — <i>Massachusetts Institute of Technology, USA; Mitsubishi Electric Corporation, Japan</i>	41
---	----

Revealing Compiler Heuristics through Automated Discovery and Optimization

Volker Seeker, Chris Cummins, Murray Cole, Björn Franke, Kim Hazelwood, and Hugh Leather — <i>Meta AI Research, USA; University of Edinburgh, United Kingdom</i>	55
--	----

SLaDe: A Portable Small Language Model Decompiler for Optimized Assembly

Jordi Armengol-Estapé, Jackson Woodruff, Chris Cummins, and Michael F. P. O’Boyle — <i>University of Edinburgh, United Kingdom; Meta AI Research, USA</i>	67
---	----

TapeFlow: Streaming Gradient Tapes in Automatic Differentiation

Milad Hakimi and Arrvinth Shriraman — <i>Simon Fraser University, Canada</i>	81
--	----

Compilers for GPUs

A Framework for Fine-Grained Synchronization of Dependent GPU Kernels

Abhinav Jangda, Saeed Maleki, Maryam Mehri Dehnavi, Madan Musuvathi, and Olli Saarikivi — <i>Microsoft Research, USA; University of Toronto, Canada</i>	93
---	----

Enhancing Performance through Control-Flow Unmerging and Loop Unrolling on GPUs

Alnis Murtovi, Giorgis Georgakoudis, Konstantinos Parasyris, Chunhua Liao, Ignacio Laguna, and Bernhard Steffen — <i>TU Dortmund, Germany; Lawrence Livermore National Laboratory, USA</i>	106
--	-----

Retargeting and Respecializing GPU Workloads for Performance Portability

Ivan R. Ivanov, Oleksandr Zinenko, Jens Domke, Toshio Endo, and William S. Moses — <i>Tokyo Institute of Technology, Japan; RIKEN R-CCS, Japan; Google DeepMind, France; University of Illinois at Urbana-Champaign, USA; Google DeepMind, USA</i>	119
--	-----

Seer: Predictive Runtime Kernel Selection for Irregular Problems

Ryan Swann, Muhammad Osama, Karthik Sangaiah, and Jalal Mahmud — <i>AMD, USA</i>	133
--	-----

Custom Processors

AXI4MLIR: User-Driven Automatic Host Code Generation for Custom AXI-Based Accelerators

Nicolas Bohm Agostini, Jude Haris, Perry Gibson, Malith Jayaweera, Norm Rubin, Antonino Tumeo, José L. Abellán, José Cano, and David Kaeli — <i>Northeastern University, USA; Pacific Northwest National Laboratory, USA; University of Glasgow, United Kingdom; University of Murcia, Spain</i>	143
--	-----

Ecmas: Efficient Circuit Mapping and Scheduling for Surface Code

Mingzheng Zhu, Hao Fu, Jun Wu, Chi Zhang, Wei Xie, and Xiang-Yang Li — <i>University of Science and Technology of China, China</i>	158
--	-----

PresCount: Effective Register Allocation for Bank Conflict Reduction

Xiaofeng Guan, Hao Zhou, Guoqing Bao, Handong Li, Liang Zhu, and Jianguo Yao — <i>Shanghai Jiao Tong University, China; Shanghai Enflame Technology, China</i>	170
--	-----

Tackling the Matrix Multiplication Micro-kernel Generation with Exo Adrián Castelló, Julian Bellavita, Grace Dinh, Yuka Ikarashi, and Héctor Martínez — <i>Universitat Politècnica de València, Spain; Cornell University, USA; University of California at Berkeley, USA; Massachusetts Institute of Technology, USA; Universidad de Córdoba, Spain</i>	182
Compiler Construction	
One Automaton to Rule Them All: Beyond Multiple Regular Expressions Execution Luisa Cicolini, Filippo Carloni, Marco D. Santambrogio, and Davide Conficconi — <i>Politecnico di Milano, Italy</i>	193
Whose Baseline Compiler Is It Anyway? Ben L. Titzer — <i>Carnegie Mellon University, USA</i>	207
Enabling Fine-Grained Incremental Builds by Making Compiler Stateful Ruobing Han, Jisheng Zhao, and Hyesoon Kim — <i>Georgia Institute of Technology, USA</i>	221
Custom Environments	
Compile-Time Analysis of Compiler Frameworks for Query Compilation Alexis Engelke and Tobias Schwarz — <i>TU Munich, Germany</i>	233
DrPy: Pinpointing Inefficient Memory Usage in Multi-Layer Python Applications Jinku Cui, Qidong Zhao, Yueming Hao, and Xu Liu — <i>North Carolina State University, USA</i>	245
SCHEMATIC: Compile-Time Checkpoint Placement and Memory Allocation for Intermittent Systems Hugo Reymond, Jean-Luc Béchenec, Mikaël Briday, Sébastien Faucou, Isabelle Puaut, and Erven Rohou — <i>Université de Rennes - Inria - CNRS - IRISA, France; Nantes Université - École Centrale Nantes - CNRS - LS2N - UMR 6004, France</i>	258
Latent Idiom Recognition for a Minimalist Functional Array Language using Equality Saturation Jonathan Van der Cruyssen and Christophe Dubach — <i>McGill University, Canada</i>	270
Static/Dynamic Analyses	
BEC: Bit-Level Static Analysis for Reliability against Soft Errors Yousun Ko and Bernd Burgstaller — <i>Yonsei University, South Korea</i>	283
Boosting the Performance of Multi-solver IFDS Algorithms with Flow-Sensitivity Optimizations Haofeng Li, Jie Lu, Haining Meng, Liqing Cao, Lian Li, and Lin Gao — <i>Institute of Computing Technology at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Zhongguancun Laboratory, China; TianqiSoft, China</i>	296
Representing Data Collections in an SSA Form Tommy McMichen, Nathan Greiner, Peter Zhong, Federico Sossai, Atmn Patel, and Simone Campanoni — <i>Northwestern University, USA</i>	308
Revamping Sampling-Based PGO with Context-Sensitivity and Pseudo-instrumentation Wenlei He, Hongtao Yu, Lei Wang, and Taewook Oh — <i>Meta, USA</i>	322
Supporting Tools	
Compiler Testing with Relaxed Memory Models Luke Geeson and Lee Smith — <i>University College London, United Kingdom; Arm, United Kingdom</i>	334
High-Throughput, Formal-Methods-Assisted Fuzzing for LLVM Yuyou Fan and John Regehr — <i>University of Utah, USA</i>	349
EasyTracker: A Python Library for Controlling and Inspecting Program Execution Théo Barollet, Christophe Guillon, Manuel Selva, François Broquedis, Florent Bouchez-Tichadou, and Fabrice Rastello — <i>University Grenoble Alpes - Inria - CNRS - Grenoble INP - LIG, France</i>	359
OptiWISE: Combining Sampling and Instrumentation for Granular CPI Analysis Yuxin Guo, Alex W. Chadwick, Márton Erds, Utpal Bora, Ilias Vougioukas, Giacomo Gabrielli, and Timothy M. Jones — <i>University of Cambridge, United Kingdom; Arm, USA; Arm, United Kingdom</i>	373
Practice and Experience	
EasyView: Bringing Performance Profiles into Integrated Development Environments Qidong Zhao, Milind Chabbi, and Xu Liu — <i>North Carolina State University, USA; Scalable Machines Research, USA</i>	386
Experiences Building an MLIR-Based SYCL Compiler Ettore Tiotto, Víctor Pérez, Whitney Tsang, Lukas Sommer, Julian Oppermann, Victor Lomüller, Mehdi Goli, and James Brodman — <i>Intel Corporation, Canada; Codeplay Software, United Kingdom; Intel Corporation, USA</i>	399

Unveiling and Vanquishing Goroutine Leaks in Enterprise Microservices: A Dynamic Analysis Approach Georgian-Vlad Saioc, Dmitry Shirchenko, and Milind Chabbi — <i>Aarhus University, Denmark; Uber Technologies, Denmark; Uber Technologies, USA</i>	411
Acceleration Techniques	
A System-Level Dynamic Binary Translator using Automatically-Learned Translation Rules Jinhu Jiang, Chaoyi Liang, Rongchao Dong, Zhaohui Yang, Zhongjun Zhou, Wenwen Wang, Pen-Chung Yew, and Weihua Zhang — <i>Fudan University, China; University of Georgia, USA; University of Minnesota at Twin Cities, USA</i>	423
Instruction Scheduling for the GPU on the GPU Ghassan Shobaki, Pnar Muyan-Özçelik, Josh Hutton, Bruce Linck, Vladislav Malysenko, Austin Kerbow, Ronaldo Ramirez-Ortega, and Vahl Scott Gordon — <i>California State University, Sacramento, USA; Advanced Micro Devices, USA</i>	435
JITSPMM: Just-in-Time Instruction Generation for Accelerated Sparse Matrix-Matrix Multiplication Qiang Fu, Thomas B. Rolinger, and H. Howie Huang — <i>Advanced Micro Devices, USA; NVIDIA, USA; George Washington University, USA</i>	448
oneDNN Graph Compiler: A Hybrid Approach for High-Performance Deep Learning Compilation Jianhui Li, Zhennan Qin, Yijie Mei, Jingze Cui, Yunfei Song, Ciyong Chen, Yifei Zhang, Longsheng Du, Xianhang Cheng, Baihui Jin, Yan Zhang, Jason Ye, Eric Lin, and Dan Lavery — <i>Intel, USA; Intel, China</i>	460
Author Index	471