

# **2021 IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2021)**

**Virtual Symposium  
27 February – 3 March 2021**



**IEEE Catalog Number: CFP21CGO-POD  
ISBN: 978-1-7281-8614-6**

**Copyright © 2021 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:	CFP21CGO-POD
ISBN (Print-On-Demand):	978-1-7281-8614-6
ISBN (Online):	978-1-7281-8613-9

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# Contents

## Frontmatter

<b>Message from the General Chair</b> . . . . .	iii
<b>Message from the Program Chairs</b> . . . . .	v
<b>CGO 2021 Organization</b> . . . . .	vi
<b>Report from the Artifact Evaluation Committee</b> . . . . .	x
<b>CGO 2021 Sponsors</b> . . . . .	xii

## Keynote

<b>Data Layout and Data Representation Optimizations to Reduce Data Movement (Keynote)</b> Mary Hall — <i>University of Utah, USA</i> . . . . .	1
--	---

## Compiler Infrastructure

<b>MLIR: Scaling Compiler Infrastructure for Domain Specific Computation</b> Chris Lattner, Mehdi Amini, Uday Bondhugula, Albert Cohen, Andy Davis, Jacques Pienaar, River Riddle, Tatiana Shpeisman, Nicolas Vasilache, and Oleksandr Zinenko — <i>Google, USA; Indian Institute of Science, India; Google, France</i> . . . . .	2
<b>Progressive Raising in Multi-level IR</b> Lorenzo Chelini, Andi Drebes, Oleksandr Zinenko, Albert Cohen, Nicolas Vasilache, Tobias Grosser, and Henk Corporaal — <i>Eindhoven University of Technology, Netherlands; Inria, France; ENS Paris, France; Google, France; Google, Switzerland; University of Edinburgh, UK</i> . . . . .	15
<b>Towards a Domain-Extensible Compiler: Optimizing an Image Processing Pipeline on Mobile CPUs</b> Thomas Koehler and Michel Steuwer — <i>University of Glasgow, UK; University of Edinburgh, UK</i> . . . . .	27
<b>BuildIt: A Type-Based Multi-stage Programming Framework for Code Generation in C++</b> Ajay Brahmakshatriya and Saman Amarasinghe — <i>Massachusetts Institute of Technology, USA</i> . . . . .	39

## Dealing with Precision

<b>An Interval Compiler for Sound Floating-Point Computations</b> Joao Rivera, Franz Franchetti, and Markus Püschel — <i>ETH Zurich, Switzerland; Carnegie Mellon University, USA</i> . . . . .	52
<b>Seamless Compiler Integration of Variable Precision Floating-Point Arithmetic</b> Tiago Trevisan Jost, Yves Durand, Christian Fabre, Albert Cohen, and Frédéric Pétrot — <i>Université Grenoble Alpes, France; CEA LIST, France; Google, France; CNRS, France; Grenoble INP, France; TIMA, France</i> . . . . .	65
<b>UNIT: Unifying Tensorized Instruction Compilation</b> Jian Weng, Animesh Jain, Jie Wang, Leyuan Wang, Yida Wang, and Tony Nowatzki — <i>University of California at Los Angeles, USA; Amazon, USA</i> . . . . .	77
<b>Unleashing the Low-Precision Computation Potential of Tensor Cores on GPUs</b> Guangli Li, Jingling Xue, Lei Liu, Xueying Wang, Xiu Ma, Xiao Dong, Jiansong Li, and Xiaobing Feng — <i>Institute of Computing Technology at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; UNSW, Australia; Jilin University, China</i> . . . . .	90

## Binary Profiling, Tracing, Sampling

<b>Cinnamon: A Domain-Specific Language for Binary Profiling and Monitoring</b> Mahwish Arif, Ruoyu Zhou, Hsi-Ming Ho, and Timothy M. Jones — <i>University of Cambridge, UK; University of Sussex, UK</i> . . . . .	103
<b>GPA: A GPU Performance Advisor Based on Instruction Sampling</b> Keren Zhou, Xiaozhu Meng, Ryuichi Sai, and John Mellor-Crummey — <i>Rice University, USA</i> . . . . .	115
<b>ELFies: Executable Region Checkpoints for Performance Analysis and Simulation</b> Harish Patil, Alexander Isaev, Wim Heirman, Alen Sabu, Ali Hajiabadi, and Trevor E. Carlson — <i>Intel Corporation, USA; Intel Corporation, Belgium; National University of Singapore, Singapore</i> . . . . .	126
<b>Vulkan Vision: Ray Tracing Workload Characterization using Automatic Graphics Instrumentation</b> David Pankratz, Tyler Nowicki, Ahmed Eltantawy, and José Nelson Amaral — <i>University of Alberta, Canada; Huawei Technologies, Canada</i> . . . . .	137

## Parallelism - Optimizing, Modeling, Testing

<b>Loop Parallelization using Dynamic Commutativity Analysis</b>	
Christos Vasiladiotis, Roberto Castañeda Lozano, Murray Cole, and Björn Franke — <i>University of Edinburgh, UK</i> . . . . .	150
<b>Fine-Grained Pipeline Parallelization for Network Function Programs</b>	
Seungbin Song, Heelim Choi, and Hanjun Kim — <i>Yonsei University, South Korea</i> . . . . .	162
<b>YaskSite: Stencil Optimization Techniques Applied to Explicit ODE Methods on Modern Architectures</b>	
Christie L. Alappat, Johannes Seiferth, Georg Hager, Matthias Korch, Thomas Rauber, and Gerhard Wellein — <i>University of Erlangen-Nuremberg, Germany; University of Bayreuth, Germany</i> . . . . .	174
<b>GoBench: A Benchmark Suite of Real-World Go Concurrency Bugs</b>	
Ting Yuan, Guangwei Li, Jie Lu, Chen Liu, Lian Li, and Jingling Xue — <i>Institute of Computing Technology at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; UNSW, Australia</i> . . . . .	187

## Memory Optimization and Safeness

<b>Memory-Safe Elimination of Side Channels</b>	
Luigi Soares and Fernando Magno Quintão Pereira — <i>Federal University of Minas Gerais, Brazil</i> . . . . .	200
<b>Variable-Sized Blocks for Locality-Aware SpMV</b>	
Naveen Namashivayam, Sanyam Mehta, and Pen-Chung Yew — <i>HPE, USA; University of Minnesota at Twin Cities, USA</i> . . . . .	211
<b>Object Versioning for Flow-Sensitive Pointer Analysis</b>	
Mohamad Barbar, Yulei Sui, and Shiping Chen — <i>University of Technology Sydney, Australia; CSIRO's Data61, Australia</i> . . . . .	222
<b>Scaling Up the IFDS Algorithm with Efficient Disk-Assisted Computing</b>	
Haofeng Li, Haining Meng, Hengjie Zheng, Liqing Cao, Jie Lu, Lian Li, and Lin Gao — <i>Institute of Computing Technology at Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; TianqiSoft, China</i> . . . . .	236

## Compiling Graph Algorithms, Compiling for GPU's

<b>Compiling Graph Applications for GPUs with GraphIt</b>	
Ajay Brahmakshatriya, Yunming Zhang, Changwan Hong, Shoaib Kamil, Julian Shun, and Saman Amarasinghe — <i>Massachusetts Institute of Technology, USA; Adobe, USA</i> . . . . .	248
<b>Efficient Execution of Graph Algorithms on CPU with SIMD Extensions</b>	
Ruohuang Zheng and Sreepathi Pai — <i>University of Rochester, USA</i> . . . . .	262
<b>r3d3: Optimized Query Compilation on GPUs</b>	
Alexander Krolik, Clark Verbrugge, and Laurie Hendren — <i>McGill University, Canada</i> . . . . .	277
<b>C-for-Metal: High Performance SIMD Programming on Intel GPUs</b>	
Guei-Yuan Lueh, Kaiyu Chen, Gang Chen, Joel Fuentes, Wei-Yu Chen, Fangwen Fu, Hong Jiang, Hongzheng Li, and Daniel Rhee — <i>Intel Corporation, USA</i> . . . . .	289

## Compiling for Spatial, Quantum, and Embedded Devices

<b>Relaxed Peephole Optimization: A Novel Compiler Optimization for Quantum Circuits</b>	
Ji Liu, Luciano Bello, and Huiyang Zhou — <i>North Carolina State University, USA; IBM Research, USA</i> . . . . .	301
<b>StencilFlow: Mapping Large Stencil Programs to Distributed Spatial Computing Systems</b>	
Johannes de Fine Licht, Andreas Kuster, Tiziano De Matteis, Tal Ben-Nun, Dominic Hofer, and Torsten Hoefler — <i>ETH Zurich, Switzerland; MeteoSwiss, Switzerland</i> . . . . .	315
<b>Thread-Aware Area-Efficient High-Level Synthesis Compiler for Embedded Devices</b>	
Changsu Kim, Shinnung Jeong, Sungjun Cho, Yongwoo Lee, William Song, Youngsok Kim, and Hanjun Kim — <i>POSTECH, South Korea; Yonsei University, South Korea</i> . . . . .	327

## JIT and Binary Translation

<b>HHVM Jump-Start: Boosting Both Warmup and Steady-State Performance at Scale</b>	
Guilherme Ottoni and Bin Liu — <i>Facebook, USA</i> . . . . .	340
<b>Enhancing Atomic Instruction Emulation for Cross-ISA Dynamic Binary Translation</b>	
Ziyi Zhao, Zhang Jiang, Ying Chen, Xiaoli Gong, Wenwen Wang, and Pen-Chung Yew — <i>Nankai University, China; University of Georgia, USA; University of Minnesota at Twin Cities, USA</i> . . . . .	351
<b>An Experience with Code-Size Optimization for Production iOS Mobile Applications</b>	
Milind Chabbi, Jin Lin, and Raj Barik — <i>Uber Technologies, USA</i> . . . . .	363

**AnghaBench: A Suite with One Million Compilable C Benchmarks for Code-Size Reduction**  
Anderson Faustino da Silva, Bruno Conde Kind, José Wesley de Souza Magalhães, Jerônimo Nunes Rocha, Breno Campos Ferreira  
Guimarães, and Fernando Magno Quintão Pereira — *State University of Maringá, Brazil; Federal University of Minas Gerais, Brazil* 378

**Author Index** . . . . . 391