

2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM 2023)

**Marina Del Rey, California, USA
8-11 May 2023**



**IEEE Catalog Number: CFP23054-POD
ISBN: 979-8-3503-1206-5**

**Copyright © 2023 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:	CFP23054-POD
ISBN (Print-On-Demand):	979-8-3503-1206-5
ISBN (Online):	979-8-3503-1205-8
ISSN:	2576-2613

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

2023 IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM) FCCM 2023

Table of Contents

Message from the FCCM 2023 General and Program Chairs	xii
Organizing Committee	xiii
Program Committee	xiv
Additional Reviewers	xv
Sponsors	xvi

Session 1 – High-Level Synthesis

LightningSim: Fast and Accurate Trace-Based Simulation for High-Level Synthesis	1
<i>Rishov Sarkar (Georgia Institute of Technology) and Cong Hao (Georgia Institute of Technology)</i>	
PASTA: Programming and Automation Support for Scalable Task-Parallel HLS Programs on Modern Multi-Die FPGAs	12
<i>Moazin Khatti (Simon Fraser University), Xingyu Tian (Simon Fraser University), Yuze Chi (University of California, Los Angeles, USA), Licheng Guo (University of California, Los Angeles, USA), Jason Cong (University of California, Los Angeles, USA), and Zhenman Fang (Simon Fraser University)</i>	
SCCL: An open-Source SystemC to RTL Translator	23
<i>Zhuanhao Wu (University of Waterloo, Canada), Maya Gokhale (Lawrence Livermore National Labs, USA), Scott Lloyd (Brigham Young University, USA), and Hiren Patel (University of Waterloo, Canada)</i>	
Lasa: Abstraction and Specialization for Productive and Performant Linear Algebra on FPGAs... ..	34
<i>Xiaochen Hao (Peking University), Mingzhe Zhang (Tsinghua University), Ce Sun (University of Science and Technology of China), Zhuofu Tao (University of California, Los Angeles), Hongbo Rong (Intel), Yu Zhang (University of Science and Technology of China), Lei He (Eastern Institute for Advanced Study, China; University of California), Eric Petit (Intel), Wenguang Chen (Tsinghua University), and Yun Liang (Peking University)</i>	

Session 2 – Architecture and CAD

Placement Optimization for NoC-Enhanced FPGAs	41
<i>Srivatsan Srinivasan (University of Toronto, Canada), Andrew Boutros (University of Toronto, Canada), Fatemehsadat Mahmoudi (University of Toronto, Canada), and Vaughn Betz (University of Toronto, Canada)</i>	
BRAMAC: Compute-in-BRAM Architectures for Multiply-Accumulate on FPGAs	52
<i>Yuzong Chen (Cornell University) and Mohamed Abdelfattah (Cornell University)</i>	
A Machine Learning Approach for Predicting the Difficulty of FPGA Routing Problems	63
<i>Andrew David Gunter (University of British Columbia) and Steven Wilton (University of British Columbia)</i>	
CXL over Ethernet: A Novel FPGA-Based Memory Disaggregation Design in Data Centers	75
<i>Chenjiu Wang (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Ke He (n/a), Ruiqi Fan (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Xiaonan Wang (Wuxi Institute of Interconnect Technology), Wei Wang (n/a), and Qinfen Hao (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences)</i>	

Session 3 – Applications/ML

Model-Platform Optimized Deep Neural Network Accelerator Generation Through Mixed-Integer Geometric Programming	83
<i>Yuhao Ding (University of Hong Kong, Hong Kong), Jiajun Wu (University of Hong Kong, Hong Kong), Yizhao Gao (University of Hong Kong, Hong Kong), Maolin Wang (AI Chip Center for Emerging Smart Systems, Hong Kong), and Hayden Kwok-Hay So (University of Hong Kong, Hong Kong)</i>	
MSD: Mixing Signed Digit Representations for Hardware-Efficient DNN Acceleration on FPGA with Heterogeneous Resources	94
<i>Jiajun Wu (University of Hong Kong, Hong Kong), Jiajun Zhou (University of Hong Kong, Hong Kong), Yizhao Gao (University of Hong Kong, Hong Kong), Yuhao Ding (University of Hong Kong, Hong Kong), Ngai Wong (University of Hong Kong, Hong Kong), and Hayden Kwok-Hay So (University of Hong Kong, Hong Kong)</i>	
Optimizing Hybrid Binary-Unary Hardware Accelerators using Self-Similarity Measures	105
<i>Alireza Khataei (University of Minnesota, USA), Gaurav Singh (University of Minnesota, USA), and Kia Bazargan (University of Minnesota, USA)</i>	
Efficient Implementation of Ring-Binary-LWE-Based Lightweight PQC Accelerator on the FPGA Platform	114
<i>Pengzhou He (Villanova University), Tianyou Bao (Villanova University), Yazheng Tu (Villanova University), and Jiafeng Xie (Villanova University)</i>	

Session 4 – Applications/ML

ATHEENA: A Toolflow for Hardware Early-Exit Network Automation	121
<i>Benjamin Biggs (Imperial College London), Christos-Savvas Bouganis (Imperial College London), and George Constantinides (Imperial College London)</i>	
Modular and Lean Architecture with Elasticity for Sparse Matrix Vector Multiplication on FPGAs	133
<i>Abhishek Kumar Jain (AMD, USA), Chirag Ravishankar (AMD, USA), Hossein Omidian (AMD, USA), Sharan Kumar (AMD, USA), Maithilee Kulkarni (AMD, USA), Aashish Tripathi (AMD, USA), and Dinesh Gaitonde (AMD, USA)</i>	
HARFLOW3D: A Latency-Oriented 3D-CNN Accelerator Toolflow for HAR on FPGA Devices	144
<i>Petros Toupas (Imperial College London; Information Technologies Institute-CERTH), Alexander Montgomerie-Corcoran (Imperial College London), Christos-Savvas Bouganis (Imperial College London), and Dimitrios Tzovaras (Information Technologies Institute-CERTH)</i>	
Power2Picture: Using Generative CNNs for Input Recovery of Neural Network Accelerators Through Power Side-Channels on FPGAs	155
<i>Lukas Huegle (Karlsruhe Institute of Technology (KIT), Germany), Martin Gotthard (Karlsruhe Institute of Technology (KIT), Germany), Vincent Meyers (Karlsruhe Institute of Technology (KIT), Germany), Jonas Krautter (Karlsruhe Institute of Technology (KIT), Germany), Dennis R. E. Gnad (Karlsruhe Institute of Technology (KIT), Germany), and Mehdi B. Tahoori (Karlsruhe Institute of Technology (KIT), Germany)</i>	

Session 5 – Applications

Computing and Compressing Electron Repulsion Integrals on FPGAs	162
<i>Xin Wu (Paderborn University, Germany), Tobias Kenter (Paderborn University, Germany), Robert Schade (Paderborn University, Germany), Thomas D. Kühne (Paderborn University, Germany), and Christian Plessl (Paderborn University, Germany)</i>	
Tensor-Product-Based Accelerator for Area-Efficient and Scalable Number Theoretic Transform	174
<i>Yuying Zhang (Hong Kong University of Science and Technology), Sarveswara Reddy Sathi (Hong Kong University of Science and Technology), Zili Kou (Hong Kong University of Science and Technology), Sharad Sinha (Indian Institute of Technology, India), and Wei Zhang (Hong Kong University of Science and Technology)</i>	
SQL2FPGA: Automatic Acceleration of SQL Query Processing on Modern CPU-FPGA Platforms	184
<i>Alec Lu (Simon Fraser University) and Zhenman Fang (Simon Fraser University)</i>	
DGNN-Booster: A Generic FPGA Accelerator Framework For Dynamic Graph Neural Network Inference	195
<i>Hanqiu Chen (Georgia Institute of Technology, USA) and Cong Hao (Georgia Institute of Technology, USA)</i>	

Posters

Designing a Configurable IEEE-Compliant FPU that Supports Variable Precision for Soft Processors	202
<i>Chris Keilbart (Simon Fraser University, Canada), Yuhui Gao (Simon Fraser University, Canada), Martin Chua (University of British Columbia, Canada), Eric Matthews (Simon Fraser University, Canada), Steven J.E. Wilton (University of British Columbia), and Lesley Shannon (Simon Fraser University, Canada)</i>	
HyBNN: Quantifying and Optimizing Hardware Efficiency of Binary Neural Networks	203
<i>Geng Yang (Xidian University, China), Jie Lei (Xidian University, China), Zhenman Fang (Simon Fraser University, Canada), Yunsong Li (Xidian University, China), Jiaqing Zhang (Xidian University, China), and Weiyang Xie (Xidian University, China)</i>	
Accelerating 128-bit Floating-Point Matrix Multiplication on FPGAs	204
<i>Fumiya Kono (Shizuoka Institute of Science and Technology, Japan), Naohito Nakasato (The University of Aizu, Japan), and Maho Nakata (Cluster for Pioneering Research, Japan)</i>	
Efficient Implementation of a Genetic Algorithm for the Capacitated Vehicle Routing Problem on a High-Performance FPGA	205
<i>Maximilian Jakob Heer (University of Rhode Island, USA), Jose Quevedo (University of Rhode Island, USA), Marwan F. Abdelatti (University of Rhode Island, USA), Resit Sendag (University of Rhode Island, USA), and Manbir Sodhi (University of Rhode Island, USA)</i>	
An Efficient Piecewise Linear Approximation of Non-Linear Operations for Transformer Inference	206
<i>Haodong Lu (Fudan University, China), Qichang Mei (Fudan University, China), and Kun Wang (Fudan University, China)</i>	
Feature Extraction Accelerator for Streaming Time Series	207
<i>Prithviraj Yuvaraj (UCR)</i>	
UPTRA: An Ultra-Parameterized Temporal CGRA Modeling and Optimization	208
<i>Yuan Dai (State Key Lab of ASIC and System, Fudan University, China), Yunhui Qiu (State Key Lab of ASIC and System, Fudan University, China), Qilong Zhu (State Key Lab of ASIC and System, Fudan University, China), Jingyuan Li (State Key Lab of ASIC and System, Fudan University, China), Wenbo Yin (State Key Lab of ASIC and System, Fudan University, China), and Lingli Wang (State Key Lab of ASIC and System, Fudan University, China)</i>	
ReLoDAQ: Resource-Efficient, Low Overhead 200 Gbit s ⁻¹ Data Acquisition System for 6G Prototyping	209
<i>Christian Karle (Karlsruhe Institut of Technology, Germany), Marc Neu (Karlsruhe Institut of Technology, Germany), Johannes Pfau (Karlsruhe Institut of Technology, Germany), Jan Sperling (Karlsruhe Institut of Technology, Germany), and Jürgen Becker (Karlsruhe Institut of Technology, Germany)</i>	

FASBM: FPGA-Specific Approximate Sum-Based Booth Multipliers for Energy Efficient Hardware Acceleration of Image Processing and Machine Learning Applications	210
<i>Zainab Aizaz (Maulana Azad National Institute of Technology, MANIT, India), Kavita Khare (Maulana Azad National Institute of Technology, MANIT, India), and Aizaz Tirmizi (Lakshmi Narain College of Technology, LNCT, India)</i>	
SpCNA: An FPGA-Based Accelerator for Point Cloud Convolutional Neural Networks	211
<i>Gong-Lang Zhou (Sun Yat-Sen University, China), Kaiyuan Guo (Novauto Co., Ltd., China), Xiang Chen (Sun Yat-Sen University, China), and Kwok Wa Leung (Sun Yat-Sen University, China)</i>	
A Flexible and Scalable Reconfigurable FPGA Overlay Architecture for Data-Flow Processing ...	212
<i>Anna Drewes (Otto-von-Guericke-Universität Magdeburg, Germany), Vitalii Burtsev (Otto-von-Guericke-Universität Magdeburg, Germany), Bala Gurumurthy (Otto-von-Guericke-Universität Magdeburg, Germany), Martin Wilhelm (Otto-von-Guericke-Universität Magdeburg, Germany), David Broneske (Otto-von-Guericke-Universität Magdeburg, Germany), Gunter Saake (Otto-von-Guericke-Universität Magdeburg, Germany), and Thilo Pionteck (Otto-von-Guericke-Universität Magdeburg, Germany)</i>	
Improving Performance of HPC Kernels on FPGAs using High-Level Resource Management	213
<i>Antonio Filgueras (Universitat Politècnica de Catalunya; Barcelona Supercomputing Center), Miquel Vidal (Barcelona Supercomputing Center), Daniel Jiménez-González (Universitat Politècnica de Catalunya; Barcelona Supercomputing Center), Carlos Álvarez (Universitat Politècnica de Catalunya; Barcelona Supercomputing Center), and Xavier Martorell (Universitat Politècnica de Catalunya; Barcelona Supercomputing Center)</i>	
Accelerating Graph Analytics with oneAPI and Intel FPGAs	214
<i>James Bickerstaff (University of Pittsburgh, USA), Luke Kljucaric (University of Pittsburgh, USA), and Alan George (University of Pittsburgh, USA)</i>	
Decision Forest Training Accelerator Based on Binary Feature Decomposition	215
<i>Thiem Van Chu (Tokyo Institute of Technology), Yu Mizutani (Tokyo Institute of Technology), Yuta Nagahara (Tokyo Institute of Technology), Shungo Kumazawa (Tokyo Institute of Technology), Kazushi Kawamura (Tokyo Institute of Technology), Jaehoon Yu (Tokyo Institute of Technology), and Masato Motomura (Tokyo Institute of Technology)</i>	
b8c: SpMV Accelerator Implementation Leveraging high Memory Bandwidth	216
<i>José Oliver (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain), Carlos Álvarez (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain), Teresa Cervero (Barcelona Supercomputing Center, Spain), Xavier Martorell (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain), John D. Davis (Barcelona Supercomputing Center, Spain), and Eduard Ayguadé (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain)</i>	
Scalable Quantum Error Correction for Surface Codes using FPGA	217
<i>Namitha Liyanage (Yale University), Yue Wu (Yale University), Alexander Deters (Yale University), and Lin Zhong (Yale University)</i>	

MSBF-LSTM: Most-Significant Bit-First LSTM Accelerators with Energy Efficiency Optimisations	218
<i>Sige Bian (Southeast University, China), He Li (Southeast University, China), Chengcheng Wang (Southeast University, China), Changjun Song (Southeast University, China), and Yongming Tang (Southeast University, China)</i>	
Accelerating FPGA-Based Wi-Fi Transceiver Design and Prototyping by High-Level Synthesis ...	219
<i>Thijs Havinga (Ghent University, Belgium), Xianjun Jiao (Ghent University, Belgium), Wei Liu (Ghent University, Belgium), and Ingrid Moerman (Ghent University, Belgium)</i>	
Dynamically Scheduled Memory Operations in Static High-Level Synthesis	220
<i>Robert Szafarczyk (University of Glasgow, UK), Syed Waqar Nabi (University of Glasgow, UK), and Wim Vanderbauwhede (University of Glasgow, UK)</i>	
Transformer-OPU: An FPGA-Based Overlay Processor for Transformer Networks	221
<i>Yueyin Bai (Fudan University, China), Hao Zhou (Fudan University, China), Keqing Zhao (Fudan University, China), Jianli Chen (Fudan University, China), Jun Yu (Fudan University, China), and Kun Wang (Fudan University, China)</i>	
Compiler-Assisted Kernel Selection for FPGA-Based Near-Memory Computing Platforms	222
<i>Veronia Iskandar (Chair of Adaptive Dynamic Systems, Technische Universität Dresden, Germany), Mohamed A. Abd El Ghany (Electronics Department/ IES Lab, German University in Cairo, Egypt/ TU Darmstadt, Germany), and Diana Goehringer (Chair of Adaptive Dynamic Systems, Technische Universität Dresden, Germany)</i>	
Clustering Classification on FPGAs for Neuromorphic Feature Extraction	223
<i>Luke Kljucaric (University of Pittsburgh, USA) and Alan D. George (University of Pittsburgh, USA)</i>	
Making BRAMs Compute: Creating Scalable Computational Memory Fabric Overlays	224
<i>Md Arafat Kabir (University of Arkansas), Joshua Hollis (University of Arkansas), Atiyehsadat Panahi (Cadence Design Systems), Jason Bakos (University of South Carolina), Miaoqing Huang (University of Arkansas), and David Andrews (University of Arkansas)</i>	
OCMGen: Extended Design Space Exploration with Efficient FPGA Memory Inference	225
<i>Sanjay Gandham (University of Central Florida, USA), Lingxiang Yin (University of Central Florida, USA), Hao Zheng (University of Central Florida, USA), and Mingjie Lin (University of Central Florida, USA)</i>	
PRAD: A Bayesian Optimization-Based DSE Framework for Parameterized Reconfigurable Architecture Design	226
<i>Bingbing Peng (State Key Laboratory of ASIC and System, Fudan University, China), Shaoyang Sun (State Key Laboratory of ASIC and System, Fudan University, China), Yuan Dai (State Key Laboratory of ASIC and System, Fudan University, China), Jingyuan Li (State Key Laboratory of ASIC and System, Fudan University, China), Yunhui Qiu (State Key Laboratory of ASIC and System, Fudan University, China), Kaihang Wang (State Key Laboratory of ASIC and System, Fudan University, China), Wenbo Yin (State Key Laboratory of ASIC and System, Fudan University, China), and Lingli Wang (State Key Laboratory of ASIC and System, Fudan University, China)</i>	

Moth: A Hardware Accelerator for Neural Radiance Field Inference on FPGA	227
<i>Yuanfang Wang (State Key Lab of ASIC & System, Fudan University, China), Yu Li (State Key Lab of ASIC & System, Fudan University, China), Haoyang Zhang (State Key Lab of ASIC & System, Fudan University, China), Jun Yu (State Key Lab of ASIC & System, Fudan University, China), and Kun Wang (State Key Lab of ASIC & System, Fudan University, China)</i>	

PhD Student Forum

FCCM 2023 PhD Student Forum Introduction	228
<i>Debjit Pal (University of Illinois Chicago)</i>	
Reformulating the FPGA Routability Prediction Problem with Machine Learning	230
<i>Andrew David Gunter (The University of British Columbia Vancouver, Canada) and Steve Wilton (The University of British Columbia Vancouver, Canada)</i>	
Hardware/Software Co-design for Machine Learning Accelerators	233
<i>Hanqiu Chen (Georgia Institute of Technology, USA) and Cong Hao (Georgia Institute of Technology, USA)</i>	
From Acceleration to Accelerating Acceleration: Modernizing the Accelerator Landscape using High-Level Synthesis	236
<i>Rishov Sarkar (Georgia Institute of Technology) and Cong Hao (Georgia Institute of Technology)</i>	
Power Side-Channel Attacks and Defenses for Neural Network Accelerators	239
<i>Vincent Meyers (Karlsruhe Institute of Technology, Germany)</i>	
Enabling Elastic Resource Management in Cloud FPGAs via A Multi-layer Collaborative Approach	242
<i>Wenbin Teng (School of Computer Science and Technology, University of Science and Technology of China), Lei Gong (School of Computer Science and Technology, University of Science and Technology of China), Chao Wang (School of Computer Science and Technology, University of Science and Technology of China), and Xuehai Zhou (School of Computer Science and Technology, University of Science and Technology of China)</i>	
A Framework for Graph Machine Learning on Heterogeneous Architecture	245
<i>Yi-Chien Lin (University of Southern California, USA)</i>	
DataMaster: A GNN-based Data Type Optimizer for Dataflow Design in FPGA	247
<i>Zheyuan Zou (University of Science and Technology of China), Lei Gong (University of Science and Technology of China), Chao Wang (University of Science and Technology of China), and Xuehai Zhou (University of Science and Technology of China)</i>	

Author Index	251
---------------------------	------------