2026 IEEE 34th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM 2026)

Atlanta, Georgia, USA
13-16 May 2026

IEEE Catalog Number: CFP26054-POD
ISBN: 979-8-3315-5816-1

IEEE
Advancing Technology for Humanity

Copyright © 2026, IEEE

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:	CFP26054-POD
ISBN (Print-On-Demand):	979-8-3315-5816-1
ISBN (Online):	979-8-3315-5815-4
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



TABLE OF CONTENTS

The Optimal, The Fast, and The Hybrid: Automatic Placement and Routing for AIE Arrays 1
Hang Yan, James Yen, Rongbo Zhang, Andrew Boutros, Vaughn Betz

Déjà Vu Packing: Optimizing FPGA Logic Clustering Runtime via Pattern Memoization 10
Milo Liebster, Amin Mohaghegh, Andrew Boutros

HGQ-LUT: Fast LUT-Aware Training and Efficient Architectures for DNN Inference 19
Chang Sun, Zhiqiang Que, Bakhtiar Zadeh, Qibin Liu, Kevin H. Alvarez, Wayne Luk, Maria Spiropulu

FHPSAC: FPGA-based High-Parallelism SAC Accelerator 29
Jiabin Xu, Wang Fan, Xuegong Zhou, Wei Cao, Jialin Chen, Fengzhe Zhang, Fan Zhang, Xinsheng Yu

GraphLeap: Decoupling Graph Construction and Convolution for Vision GNN Acceleration on FPGA 38
Anvitha Ramachandran, Dhruv Parikh, Viktor Prasanna

Δ^2-PSUM: A Low-Latency Soft-Error-Resilient Binary Neural Network Inference Processor 48
Sae-Byeok Jeong, Tae-Hwan Kim

NetCilk: Extending Task-Level Parallelism Seamlessly across FPGAs 53
Mohamed Shahawy, Canberk Sönmez, Paolo Ienne

EZCache: Easy Action-Enabled FPGA Caches for Non-Stalling Datapaths in SmartNICs and Beyond 62
Ahmed Abdelsalam, Vishal Gondaliya, Ezz Hamed, Pragati Medleri Hire Math, Marc Gepigon
Joshua Landgraf, Nadeen Gebara, Bob Groza, Dongwook Lee, Anshuman Verma, Andrew Putnam

SPAC: Automating FPGA-based Network Switches with Protocol Adaptive Customization 71
Guoyu Li, Yang Cao, Lucas H L Ng, Alexander Charlton, Qianzhou Wang, Will Punter
Philippos Papaphilippou, Ce Guo, Hongxiang Fan, Wayne Luk, Saman Amarasinghe
Ajay Brahmakshatriya

A Robotics Middleware for FPGAs Supporting Dynamic Function Exchange and Streaming Data Distribution 81
Alexander Philipp Nowosad, Christian Lienen, Marco Platzner

ViM-Q: Scalable Algorithm-Hardware Co-Design for Vision Mamba Model Inference on FPGA 90
Shengzhe Lyu, Yuhan She, Patrick S. Y. Hung, Ray C. C. Cheung, Weitao Xu

ReCoVLM: A Reconfigurable FPGA–GPU Co-Design for Edge Vision-Language Inference 100
Jingyu Wang, Yongjiang Xue, Kailai Zhuang, Mingze Sun, Qingzeng Song

LUT-LLM: Efficient Language Model Inference with Memory-based Computations on FPGAs 109
Zifan He, Shengyu Ye, Rui Ma, Yang Wang, Jason Cong

VSALUT: A Lightweight Low-Dimensional VSA Classifier for Efficient Inference on FPGA 119
Nuntipat Narkthong, Xiaolin Xu

DiffRouter: A Differentiable Routing Framework for UltraScale FPGAs 124
Xiaohan Gao, Zhili Xiong, Yusu Wang, David Z. Pan

PathSteiner: Improving PathFinder with Quasi-Optimal Steiner-Tree Initialization 133
Shashwat Shrivastava, Luka Kurešević, Alexandros Poupakis, Chirag Ravishankar, Dinesh Gaitonde
Stefan Nikolić, Mirjana Stojilović

Enabling Net-Level Global Vision in FPGA Routing via Precomputed Steiner Potential Fields 142
Yaozhang Liu

Enabling Context-Switchable Monolithic 3D FPGA Design Using Bistable Ferroelectric Inverters 151
Faaiq Waqar, Matthew Chen, Zifan He, Zishen Wan, Minji Shon, Wei-Hsing Huang, Jason Cong Shimeng Yu

ReFHE-NTT: Resource-Driven NTT FPGA Architecture for Fully Homomorphic Encryption 161
Valentino Guerrini, Giuseppe Sorrentino, Alessandro Barenghi, Davide Conficconi

TransDot: An Area-efficient Reconfigurable Floating-Point Unit for Trans-Precision Dot-Product Accumulation for FPGA AI Engines 171
Jiayi Wang, Maohua Nie, Sin-Chen Lin, C.-J. Richard Shi, Ang Li

AIE4ML: An End-to-End Framework for Compiling Neural Networks for the Next Generation of AMD AI Engines 176
Dimitrios Danopoulos, Enrico Lupi, Chang Sun, Sebastian Dittmeier, Michael Kagan, Vladimir Loncar Maurizio Pierini

When Systolic Arrays Meet AI Engines: Architectural Constraints on AMD Versal ACAP 185
Jason Kimko, Jason Cong

Adaptive AIE–PL Systems for Efficient End-to-End Pyramidal 3D Image Registration 194
Giuseppe Sorrentino, Paolo S. Galfano, Claudio Di Salvo, Eleonora D'Arnese, Davide Conficconi

LegoMap: Optimization for High-Throughput Transformer Computing on AI Engine-Based FPGAs 204
Hailiang Hu, Haodong Chang, Donghao Fang, Zhenrui Wang, Wuxi Li, Rongjian Liang, Bo Yuan, Jiang Hu

Hardware Stencil Accelerator with Periodic Boundary Conditions 213
Damien Simon, Olivier Sentieys, Sylvain Lefebvre

Trident: Efficient FPGA Acceleration of XMSS Tree in Post-Quantum Signature Scheme SLH-DSA 222
Tianyou Bao, Joshua Ennis, Kirill Morozov, Jiafeng Xie

ImageHD: Energy-Efficient On-Device Continual Learning of Visual Representations via Hyperdimensional Computing 231
Jebacyril Arockiaraj, Dhruv Parikh, Viktor Prasanna

FASTR: FPGA-based Acceleration of Hierarchical Foundation Models for SAR ATR 241
Sachini Wickramasinghe, Cauligi Raghavendra, Viktor Prasanna

AccelOrb: FPGA Acceleration of Orb v2 for Fast Molecular Dynamics 251
Sunjae Kim, Gwanhong Park, Jeawoo Lim, Faaiz Asim, Jongeun Lee

HESP-Stream: A Sensor-Direct, Instruction-Driven FPGA System for Real-Time Sparse Event Processing 256
Jihui Qi, Yongjiang Xue, Qingzeng Song

AutoINV: Automated Invariant Generation Framework for Formal Verification on High-Level Synthesis Designs 257
Xiaofeng Zhou, Linfeng Du, Guangyu Hu, Sharad Sinha, Hongce Zhang, Wei Zhang

LiveGraph: High-Performance On-FPGA Dynamic Graph Updating Framework 258
Yufeng Luo, Peikun Hong, Jing Wang, Feiyang Wu, Chao Li, Minyi Guo

Accelerating Topology Optimization on AMD Versal AIE-ML Engines 259
Kaustubh Manohar Mhatre, Vedant Tewari, Aditya Ray, Farhan Khan, Ridwan Olabiyi, Ashif Iquebal Aman Arora

LLM-Assisted Analysis of On-Chip Protocol Implementations 260
Melisande Zonta-Roudes, Nora Hinderling, Supraja Sridhara, Srinidhi Nagendra, Shweta Shinde

A Streaming FPGA Architecture for Sparse Matrix Multiplication with Sparsity-Aware Data Reordering 261
Shuxuan Li, Nikela Papadopoulou, Wim Vanderbauwhede

STEEL: Sparsity-Aware Fused Attention for Energy-Efficient Long-Sequence Inference on AMD's XDNA™ NPU 262
Victor J.B. Jung, Gagandeep Singh, Joseph Melber, Kristof Denolf, Francesco Conti, Luca Benini

Towards Compiler-Driven Dynamic Partial Reconfiguration with MLIR 263
Gabriel Rodriguez-Canal, Nick Brown, Maurice Jamieson, Nicolas Bohm Agostini, Ankur Limaye Vito Giovanni Castellana, Joseph Manzano, Antonino Tumeo

Chariot: Compiler-Aware Heterogeneous Graph Representation Learning for Automated HLS Optimization 264
Yanlong Huang, Jierui Liu, Yuhan She, Rongliang Fu, Tsung-Yi Ho, Hong Yan, Ray C.C. Cheung

DTCore: A Compiler-Directed Control-Minimal FPGA Compute Engine 265
Wei-Cheng Zeng, Kuan-Ting Lai

Utilising Pipelined Buses to Provide Resource Efficient Dynamic Thread Scheduling in Barrel Scheduled Processors 266
Thomas Bain, David Thomas, Graeme Bragg

Design Space Exploration for Layer Pipelined DNN Accelerators Based on FPGAs 267
Yiwei Wang, Chang Wu

A Hybrid Ising FPGA-COBI Architecture with Hardware-Based Problem Decomposition 268
Ruihong Yin, Yue Zheng, Chaohui Li, Ahmet Efe, Abhimanyu Kumar, Ziqing Zeng, Ulya R. Karpuzcu Sachin S. Sapatnekar, Chris H. Kim

Precision-aware Communication in CGRAs 269
Shwet Chitnis, Fergus Xu, Ayush Kulkarni, Jiayi Wang, Jingqun Zhang, Arjun Raje, Ang Li

Towards Small Language Models on FPGAs 270
Filip Wojcicki, Omar Sharif, Ebby Samson, Paul Kelly, George Constantinides, Christos Bouganis Wayne Luk

SCOPE: SCalable and Observable PEripheral-Borrowing Framework for FPGA-Based Prototyping 271
Jiarun Yan, Ao Liu, Si Zhang, Congwu Zhang, Yazhou Wang, Shiqi Liu, Ke Zhang

LENS-HLS: LLM-Enhanced Learning-based Design Space Exploration for High-Level Synthesis 272
Yujie Yan, Guanhua Chen, Keren Zhu

ForgeBench: A Machine Learning Benchmark Suite and Auto-Generation Framework for Next-Generation HLS Tools 273
Andy Wanna, Hanqiu Chen, Cong Callie Hao

An Efficient Dataflow Framework for DiT-Based Image Generation 274
Yazhe Zhang, Shouyu Du, Zhenyu Xu, Miaoxiang Yu, Dingjiang Yan, Zhiheng Ni, Qing Yang, Tao Wei

DORA: Dataflow-Instruction Orchestration Architecture for DNN Acceleration 275
Xingzhen Chen, Zhuoping Yang, Jinming Zhuang, Shixin Ji, Sarah Schultz, Zheng Dong, Weisong Shi Peipei Zhou

HAMG: A Hierarchical Automated MemTile-based GEMM Accelerator for Versal AIE-ML 276
Kai Shao, Erwei Wang, He Li

TileFuse: Fused Mixed-Precision Kernels for Quantized LLM Inference on AMD XDNA2 NPUs 277
Wesley Pang, Gregory Jun, Feiyang Liu, Deming Chen

Lightweight Lossless Compression Scheme for Video Applications on Versal ACAP Devices 278
Sharan Kumar, Sandesh Goyal, Srini Srinivasan, Aashish Tripathi

FAST-Prefill: FPGA Accelerated Sparse Attention for Long Context LLM Prefill 279
Rakshith Jayanth, Viktor Prasanna

Lightweight Queueing Abstraction for Rapid Simulation and Automated Tuning of Input-Dependent Streaming Pipelines on FPGAs 280
Shashank Obla, Bin Li, James C. Hoe

EDSSC: An Efficient FPGA-based Accelerator for Dynamic Sparse Spectral Clustering 281
Zhengyan Liu, Ce Guo, Zehuan Zhang, Qiang Liu, Wayne Luk

Combining HLS and Computation Coding for Balanced Matrix-Vector Engines on FPGA 282
Sayanti Pal, Alexander Lehnert, Marc Reichenbach

Sparse Compressed Quantized Dataflow Architecture (QUDA) for Neuromorphic Computing 283
Shadi Matinizadeh, Anup Das

Efficient Split-Posit DSP Slice Architecture for Embedded FPGA Fabrics 284
Arun M, Madhav Rao

Reconfigurable Computing Challenge: A High-Throughput FPGA Accelerator for Real-Time 3D Gaussian Splatting 285
Guan-Wei Lai, Yi-Chen Hu, Chia-Yu Kuo, Wei-Chien Cheng, Yi-Chung Wu

Reconfigurable Computing Challenge: Real-Time Graph Neural Networks for Online Event Selection in Big Science 289
Marc Neu, Frank Baptist, Thomas Lobmaier, Fabio Papagno, Torben Ferber, Jürgen Becker

Reconfigurable Computing Challenge: Deploying SmolVLA Model on an AMD XDNA NPU 294
Xiaoyu Liang, Elaine Cao, Chun-Ning Kao, David Lin, Jiajie Li, Zhiru Zhang

Reconfigurable Computing Challenge: FPGA-Gym-v2: FPGA-Based RL Environment Acceleration with LLM-Assisted Onboarding 299
Jiayi Li, Hongxiao Zhao, Wenshuo Yue, Yihan Fu, Daijing Shi, Anjunyi Fan, Bonan Yan

Reconfigurable Computing Challenge: FPGA-Based WebAssembly Stack Co-Processor 303
Qiuping Wu, Mugeng Liu, Yixuan Zhang, Hongxiao Zhao, Yihan Fu, Gang Huang, Yun Ma, Bonan Yan

Reconfigurable Computing Challenge: Transformer for Jet Tagging on Versal AI Engines 307
Gram Koski, Sean Lipps, Zhenghua Ma, G Abarajithan, Ryan Kastner

Reconfigurable Computing Challenge: RapidScan High-Throughput Parameterized HLS-based Streaming String Matching Library for FPGAs 311
Shashank Obla, Tommy Tracy, Matthew Beck, James C. Hoe, Kevin Skadron, Wajih Ul Hassan

Impact of Ultra Ethernet Transport on Co-Tenant Latency in Reconfigurable Data Center NICs: A Simulation Study 315
Vyshnavi Alladi, Ronald Sass

To Infinity and Beyond: A Reconfigurable Architecture for Space Vision Computing 317
Claudio Di Salvo, Davide Conficconi

From Acceleration to Accelerating Acceleration: Applications and Tools for Democratizing Hardware Design 319
Rishov Sarkar, Cong Callie Hao

Unleashing Heterogeneous Systems Capabilities to Enhance Compute-Intensive Workloads 321
Giuseppe Sorrentino, Davide Conficconi