

2021 IEEE 33rd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2021)

**Virtual Conference
26 – 29 October 2021**



**IEEE Catalog Number: CFP21307-POD
ISBN: 978-1-6654-4302-9**

**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:	CFP21307-POD
ISBN (Print-On-Demand):	978-1-6654-4302-9
ISBN (Online):	978-1-6654-4301-2
ISSN:	1550-6533

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

2021 IEEE 33rd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD) SBAC-PAD 2021

Table of Contents

Message from the General Chairs	viii
Message from the Program Chairs	ix
Organizing Committee	x
Steering Committee	xi
Program Committee	xii
Reviewers	xiv
Sponsors	xvi

Session 1: Accelerated Computing

DACHash: A Dynamic, Cache-Aware and Concurrent Hash Table on GPUs	1
<i>Hao Zhou (The University of Mississippi, USA), David Troendle (The University of Mississippi, USA), and Byunghyun Jang (The University of Mississippi, USA)</i>	
A Low-Power Hardware Accelerator for ORB Feature Extraction in Self-Driving Cars	11
<i>Raúl Taranco (Universitat Politècnica de Catalunya, Spain), José-Maria Arnau (Universitat Politècnica de Catalunya, Spain), and Antonio González (Universitat Politècnica de Catalunya, Spain)</i>	
Opening the Black Box: Performance Estimation During Code Generation for GPUs	22
<i>Dominik Ernst (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), Georg Hager (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), Matthias Knorr (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), Gerhard Wellein (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), and Markus Holzer (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)</i>	
SECD: Efficient Hardware/Software Co-design of FPGA-Based DNN Accelerators for Edge Inference	33
<i>Jude Haris (University of Glasgow, UK), Perry Gibson (University of Glasgow, UK), José Cano (University of Glasgow, UK), Nicolas Bohm Agostini (Northeastern University, USA), and David Kaeli (Northeastern University, USA)</i>	

Session 2: Memory Systems

Improving Phased Transactional Memory via Commit Throughput and Capacity Estimation	44
<i>Catalina Munoz Morales (UNICAMP, Brazil), Bruno Honorio (UNICAMP, Brazil), Alexandro Baldassin (UNESP, Brazil), and Guido Araujo (UNICAMP, Brazil)</i>	
Synchronization Strategies on Many-Core SMT Systems	54
<i>Agustín Navarro Torres (Universidad de Zaragoza, Spain), Jesús Alastruey Benedé (Universidad de Zaragoza, Spain), Pablo Ibáñez Marín (Universidad de Zaragoza, Spain), and Maria Carpen Amarie (Huawei Technologies, Switzerland)</i>	
Design and Evaluation of Associative Processing Kernels	64
<i>Jonathas Silveira (University of Campinas, Brazil) and Lucas Wanner (University of Campinas, Brazil)</i>	
A Task-Based Execution Engine for Distributed Operating Systems Tailored to Lightweight Manycores with Limited On-Chip Memory	74
<i>João Vicente Souto (Federal University of Santa Catarina (UFSC), Brazil), Márcio Castro (Federal University of Santa Catarina (UFSC), Brazil), and Pedro Henrique Penna (Pontifical Catholic University of Minas Gerais (PUC Minas), Brazil)</i>	

Session 3: Computer Architecture

Efficient Tensor Slicing for Multicore NPUs using Memory Burst Modeling	84
<i>Rafael Sousa (IC-UNICAMP, Brazil), Byungmin Jung (LGE San Jose Labs, USA), Jaehwa Kwak (LGE San Jose Labs, USA), Michael Frank (MagiCore, USA), and Guido Araujo (IC-UNICAMP, Brazil)</i>	
Sparsity-Aware Power Gating for Tensor Cores	94
<i>Ehsan Atoofian (Lakehead University, Canada)</i>	
Employing Simulation to Facilitate the Design of Dynamic Binary Translators	104
<i>Vanderson Martins do Rosario (IC-UNICAMP, Brazil), Rafael Zinsly (IC-UNICAMP, Brazil), Sandro Rigo (IC-UNICAMP, Brazil), and Edson Borin (IC-UNICAMP, Brazil)</i>	
Register Flush-Free Runahead Execution for Modern Vector Processors	114
<i>Hikaru Takayashiki (Tohoku University, Japan), Masayuki Sato (Tohoku University, Japan), Kazuhiko Komatsu (Tohoku University, Japan), and Hiroaki Kobayashi (Tohoku University, Japan)</i>	

Session 4: Scheduling and Distributed Systems

Shelf Schedules for Independent Moldable Tasks to Minimize the Energy Consumption	126
<i>Anne Benoit (LIP, ENS Lyon, France), Louis-Claude Canon (Univ. Bourgogne Franche-Comté, CNRS, France), Redouane Elghazi (Univ. Bourgogne Franche-Comté, CNRS, France), and Pierre-Cyrille Héam (Univ. Bourgogne Franche-Comté, CNRS, France)</i>	

Enabling Microservices Management for Deep Learning Applications Across the Edge-Cloud Continuum	137
<i>Zeina Houmani (Inria Avalon team, LIP Laboratory, UMR CNRS - ENS de Lyon, University of Lyon, France), Daniel Balouek-Thomert (University of Utah, USA), Eddy Caron (Inria Avalon team, LIP Laboratory, UMR CNRS - ENS de Lyon, University of Lyon, France), and Manish Parashar (University of Utah, USA)</i>	
FAIR: Fully-Adaptive Framework for Improving Resource Provisioning in Collaborative CPU-FPGA Cloud Environments	147
<i>Michael Guilherme Jordan (Federal University of Rio Grande do Sul, Brazil), Guilherme Korol (Federal University of Rio Grande do Sul, Brazil), Mateus Beck Rutzig (Federal University of Santa Maria, Brazil), and Antonio Carlos Schneider Beck (Federal University of Rio Grande do Sul, Brazil)</i>	
HPC Data Storage at a Glance: The Santos Dumont Experience	157
<i>Andre Ramos Carneiro (UFRGS and LNCC, Brazil), Jean Luca Bez (UFRGS, Brazil), Carla Osthoff (LNCC, Brazil), Lucas Mello Schnorr (UFRGS, Brazil), and Philippe O. A. Navaux (UFRGS, Brazil)</i>	
Sparbit: a new Logarithmic-Cost and Data Locality-Aware MPI Allgather Algorithm	167
<i>Wilton Jaciel Loch (Santa Catarina State University (UFES), Brazil) and Guilherme Piêgas Koslovski (Santa Catarina State University (UFES), Brazil)</i>	

Session 5: Applications

Efficient Online 4D Magnetic Resonance Imaging	177
<i>Marco Barbone (Imperial College London, UK), Andreas Wetscherek (The Royal Marsden NHS Foundation Trust, UK), Thomas Yung (Imperial College London, London, UK), Uwe Oelfke (The Royal Marsden NHS Foundation Trust, UK), Wayne Luk (Imperial College London, UK), and Georgi Gaydadjiev (Bernoulli Institute, University of Groningen, UK)</i>	
Functional Approximation and Approximate Parallelization with the ACCEPT Compiler	188
<i>Lucas Reis (University of Campinas, Brazil) and Lucas Wanner (University of Campinas, Brazil)</i>	
Sampling-Based Sparse Format Selection on GPUs	198
<i>Gangyi Zhu (Ohio State University) and Gagan Agrawal (Augusta University, USA)</i>	
FSCHOL: An OpenCL-Based HPC Framework for Accelerating Sparse Cholesky Factorization on FPGAs	209
<i>Erfan Bank Tavakoli (Arizona State University, USA), Michael Riera (Arizona State University, USA), Masudul Hassan Quraishi (Arizona State University, USA), and Fengbo Ren (Arizona State University, USA)</i>	
Author Index	221