

# **2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2019)**

**Madison, Wisconsin, USA  
24-26 March 2019**



**IEEE Catalog Number: CFP19PER-POD  
ISBN: 978-1-7281-0747-9**

**Copyright © 2019 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:	CFP19PER-POD
ISBN (Print-On-Demand):	978-1-7281-0747-9
ISBN (Online):	978-1-7281-0746-2

**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

# 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) **ISPASS 2019**

## Table of Contents

Message from the ISPASS 2019 General Chair	x
Message from the ISPASS 2019 Program Chair	xi
Committees	xii
Program Committee	xiii

### Paper Session I: Best Paper Nominees

GeST: An Automatic Framework For Generating CPU Stress-Tests	1
<i>Zacharias Hadjilambrou (University of Cyprus), Shidhartha Das (Arm), Paul N Whatmough (Arm / Harvard University), David Bull (Arm), and Yiannakis Sazeides (University of Cyprus)</i>	
Characterization of Unnecessary Computations in Web Applications	11
<i>Hossein Golestani (University of Michigan), Scott Mahlke (University of Michigan), and Satish Narayanasamy (University of Michigan)</i>	
Demystifying Crypto-Mining: Analysis and Optimizations of Memory-Hard PoW Algorithms	22
<i>Runchao Han (The University of Manchester), Nikos Foutris (The University of Manchester), and Christos Kotselidis (The University of Manchester)</i>	
One Size Does Not Fit All: Quantifying and Exposing the Accuracy-Latency Trade-Off in Machine Learning Cloud Service APIs Via Tolerance Tiers	34
<i>Matthew Halpern (The University of Texas at Austin), Behzad Boroujerdiad (Harvard University), Todd Mummert (IBM Research), Evelyn Duesterwald (IBM Research), and Vijay Janapa Reddi (Harvard University)</i>	

### Paper Session II: Analysis Tools

The POP Detector: A Lightweight Online Program Phase Detection Framework	48
<i>Karl Taht (University of Utah), James Greensky (Intel), and Rajeev Balasubramonian (University of Utah)</i>	

Racing to Hardware-Validated Simulation .58.....	<i>Almutaz Adileh (Ghent University), Cecilia González-Álvarez (Nokia Bell Labs), Juan Miguel de Haro Ruiz (Barcelona Supercomputing Center), and Lieven Eeckhout (Ghent University)</i>
Full-System Simulation of Mobile CPU/GPU Platforms .68.....	<i>Kuba Kaszyk (The University of Edinburgh), Harry Wagstaff (The University of Edinburgh), Tom Spink (The University of Edinburgh), Björn Franke (The University of Edinburgh), Michael O'Boyle (The University of Edinburgh), Bruno Bodin (Yale-NUS College, National University of Singapore), and Henrik Uhrenholt (Arm Sweden)</i>
Modeling Deep Learning Accelerator Enabled GPUs .79.....	<i>Md Aamir Raihan (University of British Columbia), Negar Goli (University of British Columbia), and Tor M. Aamodt (University of British Columbia)</i>
Emulating and Evaluating Hybrid Memory for Managed Languages on NUMA Hardware .93.....	<i>Shoaib Akram (Ghent University), Jennifer B. Sartor (Ghent University), Kathryn S. McKinley (Google), and Lieven Eeckhout (Ghent University)</i>

## **Paper Session III: System Characterization**

On the Impact of Instruction Address Translation Overhead .106.....	<i>Yufeng Zhou (Rice University), Xiaowan Dong (University of Rochester), Alan L. Cox (Rice University), and Sandhya Dwarkadas (University of Rochester)</i>
Quantifying Process Variations and Its Impacts on Smartphones .117.....	<i>Guru Prasad Srinivasa (University at Buffalo), Scott Haseley (University at Illinois), Mark Hempstead (Tufts University), and Geoffrey Challen (University at Illinois)</i>
Assessing the Effects of Low Voltage in Branch Prediction Units .127.....	<i>Athanasios Chatzidimitriou (University of Athens), George Papadimitriou (University of Athens), Dimitris Gizopoulos (University of Athens), Shrikanth Ganapathy (Advanced Micro Devices, Inc.), and John Kalamatianos (Advanced Micro Devices, Inc.)</i>

## **Poster Session**

Tango: A Deep Neural Network Benchmark Suite for Various Accelerators .137.....	<i>Aajna Karki (San Jose State University), Chethan Palangotu Keshava (San Jose State University), Spoorthi Mysore Shivakumar (San Jose State University), Joshua Skow (San Jose State University), Goutam Madhukeshwar Hegde (San Jose State University), and Hyeran Jeon (San Jose State University)</i>
---	--

PARADISE - Post-Moore Architecture and Accelerator Design Space Exploration Using Device Level Simulation and Experiment .139.....	
	<i>Dilip Vasudevan (Lawrence Berkeley National Lab), George Micheliannakis (Lawrence Berkeley National Lab), David Donofrio (Lawrence Berkeley National Lab), and John Shalf (Lawrence Berkeley National Lab)</i>
A Detailed Model for Contemporary GPU Memory Systems .141.....	
	<i>Mahmoud Khairy (Purdue University), Akshay Jain (Purdue University), Tor M. Aamodt (University of British Columbia), and Tim Rogers (Purdue University)</i>
DSMM: A Dynamic Setting for Memory Management in Apache Spark .143.....	
	<i>Suk-Joo Chae (Ajou University) and Tae-Sun Chung (Ajou University)</i>
Fast Modeling of the L2 Cache Reuse Distance Histograms from Software Traces .145.....	
	<i>Jiancong Ge (National ASIC Systems Engineering Technology Research Center, Southeast University, Nanjing, China) and Ming Ling (National ASIC Systems Engineering Technology Research Center, Southeast University, Nanjing, China)</i>
FlexCPU: A Configurable Out-of-Order CPU Abstraction .147.....	
	<i>Bradley Wang (UC Davis), Ayaz Akram (UC Davis), and Jason Lowe-Power (UC Davis)</i>
Hierarchical Page Eviction Policy for Unified Memory in GPUs .149.....	
	<i>Qi Yu (National University of Defense Technology), Bruce Childers (University of Pittsburgh), Libo Huang (College of Computer, National University of Defense Technology), Cheng Qian (National University of Defense Technology), and Zhiying Wang (National University of Defense Technology)</i>
Analyzing Machine Learning Workloads Using a Detailed GPU Simulator .151.....	
	<i>Jonathan Lew (University of British Columbia), Deval A. Shah (University of British Columbia), Suchita Pati (University of Wisconsin-Madison), Shaylin Cattell (University of British Columbia), Mengchi Zhang (Purdue University), Amruth Sandhupatla (University of British Columbia), Christopher Ng (University of British Columbia), Negar Goli (University of British Columbia), Matthew D. Sinclair (University of Wisconsin-Madison), Timothy G. Rogers (Purdue University), and Tor M. Aamodt (University of British Columbia)</i>

## **Paper Session IV: Workload Characterization**

Empirical Investigation of Stale Value Tolerance on Parallel RNN Training .153.....	
	<i>Joo Hwan Lee (Samsung Semiconductor, Inc.) and Hyesoon Kim (Georgia Institute of Technolog)</i>
Characterizing Sources of Ineffectual Computations in Deep Learning Networks .165.....	
	<i>Milos Nikolic (University of Toronto), Mostafa Mahmoud (University of Toronto), Yiren Zhao (University of Cambridge), Robert Mullins (University of Cambridge), and Andreas Moshovos (University of Toronto)</i>

Demystifying Bayesian Inference Workloads .177.....	
	<i>Yu Emma Wang (Harvard University), Yuhao Zhu (University of Rochester), Glenn G. Ko (Harvard University), Brandon Reagen (Facebook), Gu-Yeon Wei (Harvard University), and David Brooks (Harvard University)</i>
Workload Characterization of Nondeterministic Programs Parallelized by STATS .190.....	
	<i>Enrico Armenio Deiana (Northwestern University) and Simone Campanoni (Northwestern University)</i>
Parallelism Analysis of Prominent Desktop Applications: An 18-Year Perspective .202.....	
	<i>Siyang Feng (University of Michigan), Subhankar Pal (University of Michigan), Yichen Yang (University of Michigan), and Ronald G. Dreslinski (University of Michigan)</i>

## Paper Session V: Data Centers and Cloud Computing

µqSim: Enabling Accurate and Scalable Simulation for Interactive Microservices .212.....	
	<i>Yanqi Zhang (Cornell), Yu Gan (Cornell), and Christina Delimitrou (Cornell)</i>
Distributed Software Defined Networking Controller Failure Mode and Availability Analysis .223.....	
	<i>Paul Reeser (AT&amp;T, Inc.), Guilhem Tesseyre (Juniper Networks, Inc.), and Marcus Callaway (AT&amp;T, Inc.)</i>
A Model Driven Approach Towards Improving the Performance of Apache Spark Applications .233.....	
	<i>Kewen Wang (University of Connecticut), Mohammad Maifi Hasan Khan (University of Connecticut), Nhan Nguyen (University of Connecticut), and Swapna Gokhale (University of Connecticut)</i>
An Improved Dynamic Vertical Partitioning Technique for Semi-Structured Data .243.....	
	<i>Sahel Sharify (University of Toronto), Alan Lu (University of Toronto), Jin Chen (University of Toronto), Arnamoy Bhattacharyya (University of Toronto), Ali Hashemi (University of Toronto), Nick Koudas (University of Toronto), and Cristiana Amza (University of Toronto)</i>

## Paper Session VI: Performance Modeling and Prediction

RPPM: Rapid Performance Prediction of Multithreaded Workloads on Multicore Processors .257.....	
	<i>Sander De Pester (Ghent University), Sam Van den Steen (Ghent University), Shoaib Akram (Ghent University), and Lieven Eeckhout (Ghent University)</i>
HeteroMap: A Runtime Performance Predictor for Efficient Processing of Graph Analytics on Heterogeneous Multi-Accelerators .268.....	
	<i>Masab Ahmad (University of Connecticut), Halit Dogan (University of Connecticut), Christopher J. Michael (Naval Research Laboratory (NRL)), and Omer Khan (University of Connecticut)</i>

mRNA: Enabling Efficient Mapping Space Exploration for a Reconfigurable Neural Accelerator .282 .....	
<i>Zhongyuan Zhao (Shanghai Jiao Tong University), Hyoukjun Kwon (Georgia Institute of Technology), Sachit Kuhar (Indian Institute of Technology, Guwahati), Weiguang Sheng (Shanghai Jiao Tong University), Zhigang Mao (Shanghai Jiao Tong University), and Tushar Krishna (Georgia Institute of Technology)</i>	
DeLTA: GPU Performance Model for Deep Learning Applications with In-Depth Memory System Traffic Analysis .293 .....	
<i>Sangkug Lym (The University of Texas at Austin), Donghyuk Lee (NVIDIA), Mike O'Connor (NVIDIA), Niladrish Chatterjee (NVIDIA), and Mattan Erez (The University of Texas at Austin)</i>	
Timeloop: A Systematic Approach to DNN Accelerator Evaluation .304 .....	
<i>Angshuman Parashar (NVIDIA), Priyanka Raina (Stanford University), Yakun Sophia Shao (NVIDIA), Yu-Hsin Chen (NVIDIA), Victor A. Ying (Massachusetts Institute of Technology), Anurag Mukkara (Massachusetts Institute of Technology), Rangharajan Venkatesan (NVIDIA), Brucek Khailany (NVIDIA), Stephen W. Keckler (NVIDIA, University of Texas at Austin), and Joel Emer (NVIDIA, Massachusetts Institute of Technology)</i>	
<b>Author Index 317</b> .....	