

Communications in Computer and Information Science

908

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Joaquim Filipe

Polytechnic Institute of Setúbal, Setúbal, Portugal

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Krishna M. Sivalingam

Indian Institute of Technology Madras, Chennai, India

Takashi Washio

Osaka University, Osaka, Japan

Junsong Yuan

University at Buffalo, The State University of New York, Buffalo, USA

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>

Chao Li · Junjie Wu (Eds.)

Advanced Computer Architecture

12th Conference, ACA 2018
Yingkou, China, August 10–11, 2018
Proceedings

Editors

Chao Li
Shanghai Jiao Tong University
Shanghai
China

Junjie Wu
National University of Defense Technology
Changsha
China

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-981-13-2422-2 ISBN 978-981-13-2423-9 (eBook)
<https://doi.org/10.1007/978-981-13-2423-9>

Library of Congress Control Number: 2018954068

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

It is a great pleasure and honor to present the proceedings of ACA 2018, the 12th Conference on Advanced Computer Architecture. ACA is sponsored by the China Computer Federation (CCF) and it is the flagship conference of the CCF Technical Committee on Computer Architecture (TCArch). It has been one of the most important academic conferences in the field of computer architecture in China since 1995.

The 2018 edition of ACA was held in the scenic area of Yingkou, a port city of the Bohai Sea. The theme this year was “Intelligent Architecture: From the Cloud to the Edge.” ACA 2018 created a forum for academic researchers and industry practitioners in China to share their insights on the next-generation computing systems. We continued the trend of making ACA an inclusive and interactive event that features invited keynotes, top paper presentation, poster showcase, and design competition, etc.

This year, we received over 120 paper registrations. Finally, there were 80 successful submissions. Each submission was reviewed by three Program Committee (PC) members on average. In all, 13 papers were rejected immediately in the first round of review and 67 papers were sent out for a second round of review. Only the papers with an average score of ≥ 3 (borderline) were considered for final inclusion, and almost all accepted papers had positive reviews or at least one review with a score of 5 (accept) or higher. Finally, the PC decided to accept 47 submissions, including 17 papers in English and 30 in Chinese. We asked the authors of all the accepted papers to submit a revised version based on the review reports.

This program would have not been possible without the efforts of the PC, the external reviewers, and the authors. We would like to express our gratitude to all the authors who submitted their papers. We would like to convey our deepest and sincerest appreciation for all the hard work and dedication of our PC members and external reviewers. We also gratefully acknowledge the kind support from our general chair, Prof. Yong Dou, organization chair, Prof. Kuanjiu Zhou, and our Steering Committee. Our thanks also go to the China Computer Federation (CCF), Technical Committee on Computer Architecture of CCF, Dalian University of Technology, the City of Yinkou, Xilinx, Baidu, and all the other institutes that kindly helped us. Finally, we greatly appreciate the steady support provided by Springer.

August 2018

Chao Li
Junjie Wu

Organization

General Chair

Yong Dou National University of Defense Technology, China

Organization Chair

Kuanjiu Zhou Dalian University of Technology, China

Program Chair

Chao Li Shanghai Jiao Tong University, China

Steering Committee

Zhenzhou Ji	Harbin Institute of Technology, China
Chenggang Wu	Institute of Computing Technology, CAS, China
Dongsheng Wang	Tsinghua University, China
Junjie Wu	National University of Defense Technology, China
Xingwei Wang	Northeastern University, China
Gongxuan Zhang	Nanjing University of Science and Technology, China

Program Committee

Quan Chen	Shanghai Jiao Tong University, China
Zidong Du	Institute of Computing Technology, CAS, China
Binzhang Fu	Huawei
Yu Hua	Huazhong University of Science and Technology, China
Weixing Ji	Beijing Institute of Technology, China
Jingwen Leng	Shanghai Jiao Tong University, China
Dongsheng Li	National University of Defense Technology, China
Duo Liu	Chongqing University, China
Yuhang Liu	Institute of Computing Technology, CAS, China
Youyou Lu	Tsinghua University, China
Guojie Luo	Beijing University, China
Bo Mao	Xiamen University, China
Songwen Pei	University of Shanghai for Science and Technology, China
Minghua Shen	Sun Yat-sen University, China
Wei Song	Institute of Information Engineering, CAS, China
Guangyu Sun	Beijing University, China
Jing Wang	Capital Normal University, China
Lei Wang	National University of Defense Technology, China

Ying Wang	Institute of Computing Technology, CAS, China
Junjie Wu	National University of Defense Technology, China
Yubing Xia	Shanghai Jiao Tong University, China
Zichen Xu	Nanchang University, China
Fengyuan Xu	Nanjing University, China
Hailong Yang	Beihang University, China
Zhibin Yu	Shenzhen Institute of Advanced Technology, China
Jingling Yuan	Wuhan University of Technology, China
Fengkai Yuan	Institute of Information Technology, CAS, China
Jidong Zhai	Tsinghua University, China
Weihua Zhang	Fudan University, China
Long Zheng	Huazhong University of Technology, China
Wenli Zheng	Shanghai Jiao Tong University, China
Junlong Zhou	Nanjing University of Science and Technology, China
Bo Wu	Colorado School of Mines, USA
Hongwen Dai	Apple Inc., USA
Lizhong Chen	Oregon State University, USA
Ruijin Zhou	VMware, USA
Shaolei Ren	University of California, Riverside, USA
Yakun Shao	NVIDIA Research, USA
Xiaoyi Lu	Ohio State University, USA
Xuehai Qian	University of Southern California, USA
Yang Hu	University of Texas at Dallas, USA
Yanqi Zhou	Baidu Silicon Valley AI Lab, USA

Additional Reviewers

Qiang Cao	Huazhong University of Technology, China
Li Jiang	Shanghai Jiao Tong University, China
Naifeng Jing	Shanghai Jiao Tong University, China
Cheng Li	University of Science and Technology of China
Tao Li	Nankai University, China
Yao Shen	Shanghai Jiao Tong University, China
Shuang Song	University of Texas at Austin, USA
Rui Wang	Beihang University, China
Chentao Wu	Shanghai Jiao Tong University, China
Qiaosha Zhou	Zhejiang Sci-Tech University, China

Contents

Accelerators

A Scalable FPGA Accelerator for Convolutional Neural Networks	3
<i>Ke Xu, Xiaoyun Wang, Shihang Fu, and Dong Wang</i>	
Memory Bandwidth and Energy Efficiency Optimization of Deep Convolutional Neural Network Accelerators	15
<i>Zikai Nie, Zhisheng Li, Lei Wang, Shasha Guo, and Qiang Dou</i>	
Research on Parallel Acceleration for Deep Learning Inference Based on Many-Core ARM Platform	30
<i>Keqian Zhu and Jingfei Jiang</i>	
Research on Acceleration Method of Speech Recognition Training	42
<i>Liang Bai, Jingfei Jiang, and Yong Dou</i>	

New Design Explorations

A Post-link Prefetching Based on Event Sampling.	53
<i>Hongmei Wei, Fei Wang, and Zhongsheng Li</i>	
The Design of Reconfigurable Instruction Set Processor Based on ARM Architecture	66
<i>Jinyong Yin, Zhenpeng Xu, Xinmo Fang, and Xihao Zhou</i>	
Stateful Forward-Edge CFI Enforcement with Intel MPX	79
<i>Jun Zhang, Rui Hou, Wei Song, Zhiyuan Zhan, Boyan Zhao, Mingyu Chen, and Dan Meng</i>	
Analytical Two-Level Near Threshold Cache Exploration for Low Power Biomedical Applications	95
<i>Yun Liang, Shuo Wang, Tulika Mitra, and Yajun Ha</i>	
DearDRAM: Discard Weak Rows for Reducing DRAM's Refresh Overhead	109
<i>Xusheng Zhan, Yungang Bao, and Ninghui Sun</i>	

Towards Efficient ML/AI

EffectFace: A Fast and Efficient Deep Neural Network Model for Face Recognition	127
<i>Weicheng Li, Dan Jia, Jia Zhai, Jihong Cai, Han Zhang, Lianyi Zhang, Hailong Yang, Depei Qian, and Rui Wang</i>	

A Power Efficient Hardware Implementation of the IF Neuron Model	140
<i>Shuquan Wang, Shasha Guo, Lei Wang, Nan Li, Zikai Nie, Yu Deng, Qiang Dou, and Weixia Xu</i>	
paraSNF: An Parallel Approach for Large-Scale Similarity Network Fusion . . .	155
<i>Xiaolong Shen, Song He, Minquan Fang, Yuqi Wen, Xiaochen Bo, and Yong Dou</i>	
An Experimental Perspective for Computation-Efficient Neural Networks Training	168
<i>Lujia Yin, Xiaotao Chen, Zheng Qin, Zhaoning Zhang, Jinghua Feng, and Dongsheng Li</i>	
Parallel Computing System	
Distributed Data Load Balancing for Scalable Key-Value Cache Systems. . . .	181
<i>Shanshan Chen, Xudong Zhou, Guiping Zhou, and Richard O. Sinnott</i>	
Performance Analysis and Optimization of Cyro-EM Structure Determination in RELION-2.	195
<i>Xin You, Hailong Yang, Zhongzhi Luan, and Depei Qian</i>	
The Checkpoint-Timing for Backward Fault-Tolerant Schemes	210
<i>Min Zhang</i>	
Quota-constrained Job Submission Behavior at Commercial Supercomputer . . .	219
<i>Jinghua Feng, Guangming Liu, Zhiwei Zhang, Tao Li, Yuqi Li, and Fuxing Sun</i>	
Author Index	233