Journal of Signal Processing Systems (2022) 94:913–915
https://doi.org/10.1007/s11265-022-01807-x

Guest Editorial: Special Issue on Advances in Signal Processing Systems

Gabriel Falcao[1] · Joseph R. Cavallaro[2] · Farhana Sheikh[3] · Leonel Sousa[4]

Published online: 18 August 2022

We are pleased to present this special issue of the *Journal of Signal Processing Systems* on Advances in Signal Processing Systems. The thirteen papers in this special issue represent new research results in several areas in the design and implementation of signal processing systems. The papers can be grouped into three major themes related to recent emerging areas: (1) computer-aided-design tools and architectures for machine learning; (2) advanced algorithms and architectures for 5G and 6G wireless communication systems; and (3) efficient edge computation. Relating to the first theme, in the article, "*Efficient Neuromorphic Signal Processing with Resonator Neurons*," Intel's new neuromorphic research processor, Loihi 2, is presented. Several example algorithms important for machine learning are described and evaluated on the new architecture that supports an extended range of stateful spiking neuron models with programmable dynamics. The examples showing the speedups achieved with the new architecture include short time Fourier transform, optical flow estimation, audio classification tasks, and an efficient spike-based spectrogram encoder.

✉ Gabriel Falcao
gff@deec.uc.pt

Joseph R. Cavallaro
cavallar@rice.edu

Farhana Sheikh
farhana.sheikh@intel.com

Leonel Sousa
las@inesc-id.pt

1 University of Coimbra, Instituto de Telecomunicações, 3030-290 Coimbra, Portugal

2 Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA

3 Intel Corporation, Hillsboro, OR 97124, USA

4 INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, 1000-029 Lisbon, Portugal

The paper, titled "*Optimization of General Matrix Multiply Library for Ternary Weight for Fast DNN Inference*," proposes a library that supports SIMD vectorized instructions for matrix multiplication of numbers with precision under 8 bits. These are implemented using simple logic operations that replace the long-latency multiply-and-add operations. The proposed solution accumulates partial sums to exploit parallelism in 8-bit and 16-bit SIMD instructions. The performance was tested in quantized deep neural networks (DNNs) with ternary weights and sub-8-bit activations and run on ARM-based CPUs.

In "*Generating Efficient FPGA-based CNN Accelerators from High-Level Descriptions*," authors address a problem that arises with the evolution of Convolutional Neural Networks (CNNs) and their hardware implementations. The paper proposes to exploit reconfigurable systems and High-Level Synthesis (HLS) to address the time-consuming RTL-based development effort and the design space exploration of these systems. The hardware accelerator generation framework characterizes the input CNN and produces hardware-aware metrics, that through HLS tools are used to generate a synthesizable RTL representation of the inference accelerator.

Learning in IoT devices is covered by the article "*Energy and Loss-aware Selective Updating for SplitFed Learning with Energy Harvesting-Powered Devices*", which proposes to reduce the communication overheads in data-privacy preserving decentralized learning frameworks, for IoT devices. Their selective update model analyzes both training loss and energy available on the energy-constrained client-side to communicate a limited number of activations and gradients. In this paper the authors report significant energy savings and minimal losses in accuracy.

Intelligence in speech processing is addressed by "*A New Neural Beamformer for Multi-channel Speech Separation*" where the authors exploit deep learning technology together with cross-attention mechanisms, to produce multi-channel

speech separation models under reverberant environments. The experimental results report superior performance for the proposed method, compared to current state-of-the-art approaches in terms of scale-invariant signal-to-noise ratio, perceptual evaluation of speech quality and short-time objective intelligibility measure.

There are multiple excellent articles addressing the second theme related 5G and 6G networks. The article titled "*An Adaptable and Scalable Generator of Distributed Massive MIMO Baseband Processing Systems*," concerns the emerging area of configurable hardware and software for wireless base stations for 5G and 6G systems. The authors propose a generator that can produce hardware realizations for systems based on parameters including baseband algorithms, MIMO system configuration, and distributed hardware processing elements. The performance and adaptability are analyzed and compared.

Conventional phased array architectures above 100 GHz require hundreds of milli-Watts per antenna, which constrains the scaling to massive arrays. The emergence of simplified, low-resolution architectures as an answer to power-efficient scaling in millimeter wave phased array receivers has been addressed in an article entitled "*Low-Resolution Architectures for Power-Efficient Scaling of mmWave Phased Array Receivers*". The power savings are quantified, and tradeoffs are compared against the penalties introduced in beamforming quality, array elements under-utilization and die area.

In keeping with the mmWave networks theme, "*Millimeter-Wave Wideband Channel Estimation using Analog True-Time-Delay Array under Hardware Impairments*" addresses the overhead of fast and accurate channel estimation in the establishment of a high-rate link between base station and the user. The authors propose a frequency-domain compressive sensing-based algorithm that leverages frequency-dependent beams of a true-time-delay (TTD) array installed on the user's equipment, to reduce the estimation overhead. The proposed algorithm is then optimized and compared with the state-of-the-art in terms of the number of training frames, estimation accuracy, computational complexity, and sensitivity to hardware impairments in antenna arrays.

High-throughput forward error correction is an essential component of high-frequency wireless systems and the article "*Ultra-high-throughput EMS NB-LDPC decoder with full-parallel node processing*" presents a very high-throughput decoder architecture for Non-Binary Low Density Parity Check codes based on the Hybrid Extended Min-Sum algorithm. They propose a fully pipelined hardware for the check node processor that computes one row of the parity check matrix per clock cycle. The paper discusses results for a specific code implemented on a 28 nm process node design, and compares throughput and BER performance.

Low-latency systems are a key feature of next generation wireless networks and the article titled "*Hardware Architecture for Guessing Random Additive Noise Decoding Markov Order (GRAND-MO)",* proposes the use of a hard-input variant of the Guessing Random Additive Noise Decoding (GRAND) technique — the GRAND Markov Order, — to eliminate the need of interleavers / de-interleavers, used for mitigating burst noise in communication channels with memory, thus reducing the overall latency in these systems. The proposed solutions are compared for short-length, high-rate codes in the paper.

Three articles address efforts to improve computation at the edge where area resources are limited, and reduced latency is required. The first one, "*Efficient Hardware Implementation Architectures for Long Integer Modular Multiplication over General Solinas Prime*," addresses the complexity involved in the multiplication of long integers, used in cryptography and homomorphic encryption. This article proposes a new method where the intermediate, partial results become shorter and can be added in parallel using Wallace-tree-based adders with a small area overhead. Several optimization schemes are proposed to reduce latency and area requirements.

In the second edge-compute related article, fast reconstruction of depth images from sparse data is addressed by presenting an efficient accelerator for autonomous systems. In "*Efficient Reconfigurable Mixed Precision $\ell 1$ Solver for Compressive Depth Reconstruction*," the authors exploit variable precision approximate computing techniques with fine-grain tuning in the accelerator for reaching a better tradeoff between algorithmic performance and implementation efficiency. The solution enables compressive depth reconstruction by varying the precision scaling in single bit granularity during the iterative optimization process. The mixed precision $\ell 1$ solvers implemented are compared in terms of area and power improvements, and also for depth image quality loss, compared to floating-point implementations.

Finally, vision processing at the edge is addressed by the article "*Increased Leverage of Transprecision Computing for Machine Vision Applications at the Edge*," which describes a new approach for monitoring workloads on resource-constrained edge devices in runtime and leveraging accuracy-throughput tradeoffs. The optimization techniques employed identify the necessary configurations for each task in terms of optimal accuracy, energy and memory on a graphics processing unit (GPU). Experimental results show that for small accuracy losses, the GPU performance gains can be significant.

This special issue was preceded by an international workshop on Signal Processing Systems, planned for Coimbra, Portugal, but held virtually due to the pandemic in October 2021, where forty-seven papers were presented. The thirteen articles in this special issue have undergone rigorous peer-review according to the journal's high standards.

We would like to thank all the authors for contributing their research results to this special issue. The anonymous

reviewers provided valuable comments and suggestions to improve the quality and thoroughness of detail in the papers and we thank them for their important efforts. We hope that you enjoy reading this special issue, and that it will inspire the community to build on the results presented here.

Gabriel Falcao received the Ph.D. degree in Electrical and Computer Engineering from the University of Coimbra in 2010. In 2011, 2012, and 2017, he was a Visiting Professor with EPFL, and in the summer of 2018, he was a Visiting Academic with ETHZ, Switzerland. He is currently a Tenured Assistant Professor with the Department of Electrical and Computer Engineering, University of Coimbra and a Researcher with Instituto de Telecomunicações. He published more than 100 peer-reviewed articles in highly prestigious international journals and conferences. His research interests include parallel computer architectures, energy-efficient processing, GPU- and FPGA-based accelerators, and compute-intensive signal processing applications, namely those related with communications and imaging. He is currently working on building energy-efficient architectures for AI, with a particular focus on processing-in-memory hardware. He is a Senior member of IEEE, member of the IEEE Signal Processing Society, the Applied Signal Processing Systems Technical Committee (ASPS TC) (previously IEEE Technical Committee on Design and Implementation of Signal Processing Systems, DISPS TC), and a Full Member of the HiPEAC Network of Excellence. He was a recipient of a Google Faculty Research Award in 2013/14 together with J. Barreto. He was General Co-Chair of the IEEE SiPS in 2020 and 2021, and Local Chair of Euro-Par 2021.

Joseph R. Cavallaro received the B.S. degree from the University of Pennsylvania, Philadelphia, Pa, in 1981, the M.S. degree from Princeton University, Princeton, NJ, in 1982, and the Ph.D. degree from Cornell University, Ithaca, NY, in 1988, all in electrical engineering. In 1988, he joined the faculty of Rice University, Houston, TX, where he is currently a professor of electrical and computer engineering. His research interests include computer arithmetic, and DSP, GPU, FPGA, and VLSI architectures for applications in wireless communications. During the 1996–1997 academic year, he served at the National Science Foundation as Director of the Prototyping Tools and Methodology Program. He was a Nokia Foundation Fellow and a Visiting Professor at the University of Oulu, Finland in 2005 and continues his affiliation there as an Adjunct Professor. He is currently the Director of the Center for Multimedia Communication at Rice University. He was an advisory board member of the IEEE SPS TC on Design and Implementation of Signal Processing Systems and a past Chair of the IEEE CAS TC on Circuits and Systems for Communications. He is a Senior Area Editor of the IEEE Transactions on Signal Processing and currently serves as an AE for the Journal of Signal Processing Systems. He was General/Program Co-chair of the 2003, 2004, and 2011 IEEE International Conference on Application-Specific Systems, Architectures and Processors (ASAP), General/Program Cochair for the 2012 and 2014 ACM/IEEE GLSVLSI, General Co-Chair of the 2020 and 2021 IEEE SiPS workshops, and General Chair of the 2020 IEEE Asilomar Conference on Signals, Systems, and Computers. He served on the IEEE CAS Society Board of Governors during 2014 and is a Fellow of the IEEE.

Dr. Farhana Sheikh is a Principal Engineer at Intel's Programmable Solution Group's CTO & Strategy Office. She has over 15 years of experience in ASIC and DSP/communications research including adaptive DSP, crypto, graphics, quantum wireless control, and 5G+ wireless. Since joining PSG, after 10+ years in Intel Labs, Farhana's research focuses on 2-D and 3-D chiplet + FPGA integration research, with a focus on 3D heterogeneous integration for next generation wireless and sensing applications. Farhana has published over 50 papers and filed 22 patents, has initiated the AIB-3D open-source specification for 3D chiplet heterogeneous integration. Farhana was instrumental in enabling Intel 16 for Intel's IDM2.0 and is the co-creator of Intel's University Shuttle Program. Outside of Intel she volunteers for IEEE Solid-State Circuits Society (SSCS) and is the SSCS Women in Circuits Committee Chair. Farhana is a co-recipient of 2020, 2019, and 2012 IEEE ISSCC Outstanding Paper Awards. In 2021, Farhana was recognized for her mentorship work with students and faculty by the Semiconductor Research Corporation (SRC) that awarded her the 2021 Mahboob Khan Outstanding Industry Liaison Award. Recently, she was elected IEEE SSCS Member-at-Large for 2022-2024.

Leonel Sousa received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade de Lisboa (UL), Lisbon, Portugal, in 1996. He is currently a Full Professor with the IST, UL. He is also a Senior Researcher with the Research and Development Instituto de Engenharia de Sistemas e Computadores (INESC-ID). His research interests include computer arithmetic, VLSI architectures, and parallel computing. He has contributed to more than 300 papers in journals and international conferences, for which he got several awards, including the DASIP13 Best Paper Award, the SAMOS11 Stamatis Vassiliadis Best Paper Award, the DASIP10 Best Poster Award, and the Award UL/Caixa Geral de Depósitos for the quality of the publications in 2022. He has contributed to the organization of several international conferences, namely as program chair and general chair and has given keynotes in some of them. He has edited five special issues of international journals, he is currently Associate Editor of the IEEE Transactions on Computers, and Senior Editor of the Journal on Emerging and Selected Topics in Circuits and Systems. He is Fellow of the IET, Distinguished Scientist of ACM, and Distinguished Contributor of the IEEE Computer Society.