Guest Editors' Introduction: Hardware and Algorithms for Energy-Constrained On-Chip Machine Learning (Part 2)

JAE-SUN SEO, Arizona State University

YU CAO, Arizona State University

XIN LI, Duke University

PAUL WHATMOUGH, Arm Research

Recently, machine/deep learning algorithms have unprecedentedly improved the accuracies in practical recognition and classification tasks. However, to achieve incremental accuracy improvement, state-of-the-art deep neural network (DNN) algorithms tend to present very deep and large models, which poses significant challenges for custom hardware implementations in terms of computation, memory, and communication. This is especially true for energy-/resource-constrained hardware platforms (e.g., mobile/wearable devices, self-driving cars, IoT systems, etc.), where various constraints exist in power, performance, and area.

There is a timely need to map the latest machine-learning algorithms to application-specific hardware (e.g., ASIC, FPGA, emerging devices, etc.) in order to achieve orders of magnitude improvement in performance, energy efficiency, and density. Recent progress in emerging hardware-friendly algorithms and nanoelectronic technology will further help shed light on future hardware-software platforms for on-chip learning.

The objective of this special issue is to explore the potential of on-chip machine learning, to reveal emerging algorithms and design needs, and to promote novel applications for learning. A holistic approach of concurrent innovations in hardware and algorithms is essential to support real-time information analytics under stringent power constraints in a mobile system.

With the support of the ACM Special Interest Group in Design Automation and the IEEE Council of Electronic Design Automation, we organized the third workshop on Hardware and Algorithms for Learning On-a-chip (HALO) in 2017. Our workshop promoted close interaction between hardware design and cognitive/learning algorithms. To address the rapid progress in this field, this special issue solicits high-quality presentations from HALO 2017. In addition, a general call for papers was distributed to the research community in order to solicit a broader range of research works on energy-efficient on-chip machine-learning algorithm and hardware designs. After the peer review process, the guest editors decided to split the special issue into two parts. For the second part of this special issue, the following six papers were accepted for publication:

1. [Hardware Optimizations of Hyperdimensional Computing: Rematerialization of Hypervectors, Binarized Bundling, and Combinational Associative Memory](#)

Authors' addresses: J.-S. Seo, School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287; Y. Cao, School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287; X. Li, Electrical and Computer Engineering Department, Duke University, Durham, NC 27708; P. Whatmough, Arm Research, Boston, MA 02451.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2019 Copyright held by the owner/author(s).

1550-4832/2019/12-ART31

<https://doi.org/10.1145/3359336>

2. [Application and Thermal Reliability-Aware Reinforcement-Learning-Based Multicore Power Management](#)
3. [Comprehensive Analytic Performance Assessment and K-Means-Based Multicast Routing Algorithm and Architecture for 3D-NoC of Spiking Neurons](#)
4. [MV-Net: Towards Real-Time Deep Learning on Mobile GPGPU Systems](#)
5. [Heterogeneous Scheduling of Deep Neural Networks for Low-Power Real-Time Designs](#)
6. [QuTiBench: Benchmarking Neural Networks on Heterogeneous Hardware](#)

We are very grateful to all authors who submitted their manuscripts to this special issue, as well as the reviewers who offered high-quality reviews in a timely manner. Finally, we would like to thank Dr. Yuan Xie and Dr. Ramesh Karri, the previous and current editors-in-chief of the *ACM Journal of Emerging Technologies in Computing Systems*, and their editorial team for their consistent support on this special issue. We hope that the readers will peruse the papers selected for this special issue and that it will facilitate further research in this exciting area.