EDITORIAL

Journal of Semiconductors
(2020) 41, 020101
doi: 10.1088/1674-4926/41/2/020101

Preface to the Special Issue on Reconfigurable Computing for Energy Efficient AI Microchip Technologies

Guest Editors

Haigang Yang
Institute of Electronics, Chinese Academy of Sciences, Beijing, 100190, China
Email: yanghg@mail.ie.ac.cn

Yajun Ha
ShanghaiTech University, Shanghai, 201210, China
Email: hayj@shanghaitech.edu.cn

Lingli Wang
Fudan University, Shanghai, 200433, China
Email: llwang@fudan.edu.cn

Wei Zhang
Hong Kong University of Science and Technology, Hong Kong, China
Email: wei.zhang@ust.hk

Yingyan Lin
Rice University, Houston, TX 77005-1892, USA
Email: yingyan.lin@rice.edu

Many artificial intelligence (AI) processing tasks, especially those related to deep neural networks (DNNs), are both computation and memory intensive. Yet the traditional computing platforms such as CPU are increasingly facing difficulties in dealing with those massive processing workloads. Reconfigurable computing (RC) features the ability to perform computations in hardware to increase execution capabilities, and at the same time retain much of the flexibility of a software solution. The microchip design based on the reconfigurable computing models and principles has emerged as an effective means to ensure that the AI applications can be accelerated to not only meet the performance and throughput targets but also the power and energy efficiency requirements.

In this special issue on reconfigurable computing for energy efficient AI microchip technologies, there are contributions from nine research groups of great reputation. The first review article analyses a trend towards dynamic reconfigurable computing, while the second review article discusses some FPGA design issues emerging in the coming AI era. Further, the third review article carries out a survey on software and hardware architecture of neural network accelerator. The remaining research works cover frontier topics respectively, namely, a neural network acceleration approach with coarse-grained dataflow reconfiguration; a H-tree based reconfiguration mechanism for homogeneous processing element array design; an efficient DNN training implementation with FPGA based batch-level parallelism; a low bit-width training method for high performance DNN design; a routing algorithm for FPGAs with time-multiplexed interconnects; and a proposed way of optimizing energy efficiency of CNN-based object detection with dynamic voltage and frequency scaling.

We sincerely hope that this special issue could provide a valuable reference and perspective for the research community working in this exciting field and inspire much more researchers to explore the new ground for future innovations. We would like to thank all the authors who have contributed high-quality peer-reviewed articles to this special issue. We are also grateful to all the reviewers for their expert suggestions, as well as to the editorial and production staff of *Journal of Semiconductors* for their superb assistance.