Special Issue on Cool Chips

Ryusuke Egawa, *Tokyo Denki University, Tokyo, 120-8551, Japan*
Yasutaka Wada, *Meisei University, Tokyo, 191-8506, Japan*

Low-power, high-speed chips (Cool Chips) encompass a broad range of architectures, applications, methodologies, and usage models and are essential fundamental techniques to realize Green Transformation (GreenX). These technologies are present in artificial intelligence, Internet of Things, multimedia, digital consumer electronics, mobile, graphics, encryption, robotics, automotive, networking, medical, healthcare, and biometrics. They are based on novel architectures and schemes for single/multi/many-cores, network on chip, embedded systems, reconfigurable computing, grid, ubiquitous, dependable computing, globally asynchronous locally synchronous, and 3-D integration. Cool software, which includes parallel schedulers, embedded real-time operating systems, binary translations and compiler issues, and low-power application techniques, is also emerging.

These technologies all aim to reduce power consumption and enhance the chip performance. Regardless of their goals, all of industry has been challenged with developing optimal solutions—both hardware and software—for power optimization according to the required performance. In general, to migrate decades' worth of legacy approaches to low-power technology, researchers approach these optimal solutions from the perspective of starting from scratch.

With this in mind, we have been organizing annual Cool Chips conferences since 1998. We celebrated Cool Chips 25 in April 2022. Cool Chips, a sister conference to Hot Chips, focuses on all aspects of cool technologies. Approximately 150 individuals attend the conference each year. In addition to regular paper presentations, the conference includes keynotes and invited talks, special topic presentations, posters, and panel discussions. To attract submissions from engineers and researchers in the industry and academia, the program committee bases acceptance on a three-page extended abstract and a six-page paper. The conference proceedings include the final presentation slides with the abstract or the paper. Program committee members reviewed each of the 18 submissions for Cool Chips 25 and selected the 12 bests based on technical merit and innovation.

0272-1732 © 2023 IEEE
Digital Object Identifier 10.1109/MM.2022.3229244
Date of current version 13 January 2023.

THE ARTICLES

This special issue of *IEEE Micro* captures two contributions from among the 12 regular presentations of Cool Chips 25. The two articles are focused on a system-on-chip (SoC) with ternary neural network (TNN)/temporal convolutional neural network (TCN) accelerator, and body bias optimization for coarse-grained reconfigurable architectures (CGRAs), respectively. All of them are the emerging topics at Cool Chips 25.

In [A1], Scherer et al. described a flexible, fully digital ternary neural network accelerator in a RISC-V-based SoC. Besides support for convolutional neural networks, they introduced extensions to the accelerator design that enable processing of time-dilated TCNs. The design achieves 5.5 μJ/inference, 12.2 mW, 8,000 inferences/s at 0.5 V for a dynamic-vision-sensor-based TCN and an accuracy of 94.5%, and 2.72 μJ/inference, 12.2 mW, 3,200 inferences/s at 0.5 V for a nontrivial 9-layer, 96 channels-per-layer network with CIFAR-10 accuracy of 86%. The peak energy efficiency is 1,036 TOp/s/W, outperforming the state-of-the-art silicon-proven TinyML accelerators by 1.67×.

In [A2], Kojima et al. reformulated the optimization problem to obtain the best body bias voltage for each domain in CGRA, and introduce continuous relaxation to solve it faster than previous work based on an integer linear program. Experimental result shows the proposed method can solve the problem within 0.5 s for all benchmarks in any conditions. For a middle-class problem, up to 5.65× speedup and a geometric mean of 2.06× speedup are demonstrated compared to the previous method with negligible loss of accuracy. Besides, they explore finer body bias control considering the power- and area-overhead of an on-chip body bias generator and suggest the most reasonable design saves 66% of energy consumption.

Low-power, low-energy, power/energy-efficiency, power/energy awareness are still one of the most important factors for any kinds of chip design. To

cope with this subject, not only devices and circuits but also wide variety of innovations, including architecture, algorithm and software, are essential. Cool Chips conference series will continue to cover all kinds of low-power and high-performance, also known as "Cool" chips, and looking for future contributions.

LOW-POWER, LOW-ENERGY, POWER/ENERGY-EFFICIENCY, POWER/ENERGY AWARENESS ARE STILL ONE OF THE MOST IMPORTANT FACTORS FOR ANY KINDS OF CHIP DESIGN.

ACKNOWLEDGMENTS

It has been a pleasure to put together this Special Issue on Cool Chips. The authors would like to thank Editor-in-Chief Lizy Kurian John for her support and guidance. They also thank Tadao Nakamura, steering committee chair of Cool Chips 25, and Makoto Ikeda, organizing committee chair, for their help in arranging this special section. The Cool Chips Special Issue would not have been possible without their help.

APPENDIX: RELATED ARTICLES

[A1] Scherer et al., "TCN-CUTIE: A 1,036-TOp/s/W, 2.72-mJ/inference, 12.2-mW all-digital ternary accelerator in 22-nm FDX technology," *IEEE Micro*, vol. 43, no. 1, pp. 42–48, Jan./Feb. 2023, doi: 10.1109/MM.2022.3226630.

[A2] Kojima et al., "A scalable body bias optimization method toward low-power CGRAs," *IEEE Micro*, vol. 43, no. 1, pp. 49–57, Jan./Feb. 2023, doi: 10.1109/MM.2022.3226739.

RYUSUKE EGAWA is a professor of Tokyo Denki University, Tokyo, 120-8551, Japan. His research interests include computer architecture and high-performance computing. Egawa received a Ph.D. degree in information science from Tohoku University. He is a Member of IEEE Computer Society, the Institute of Electronics, Information and Communication Engineers (IEICE), and Information Processing Society of Japan (IPSJ). Contact him at egawa@mail.dendai.ac.jp.

YASUTAKA WADA is a professor with Meisei University, Tokyo, 191-8506, Japan. His research interests include parallel processing and applications, green computing, heterogeneous computing, automatic parallelizing compilers, and multi/many core processor architecture. His research and development activities cover various systems, from embedded systems to high-performance computing environments, to realize high-performance, energy-efficient, and easy-to-use computer systems. Wada received a Ph.D. degree in computer science and engineering from Waseda University. He is a Member of ACM, IEEE Computer Society, the Institute of Electronics, Information and Communication Engineers (IEICE), and Information Processing Society of Japan (IPSJ). Contact him at yasutaka.wada@meisei-u.ac.jp.