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 the green transformation. These technologies are present in AI, the Internet of Things, multimedia, digital consumer electronics, mobile, graphics, encryption, robotics, automotive, networking, medical, health care, and biometrics. They are based on novel architectures and schemes for single/multiple/many cores, NoC, embedded systems, reconfigurable computing, grid, ubiquitous, dependable computing, globally asynchronous locally synchronous (GALS), 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 chip performance. Regardless of their goals, the industry has been challenged with developing optimal hardware and software solutions 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 26 in April 2023. 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 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 16 submissions for COOL Chips 26 and selected the 11 best based on technical merit and innovation.

0272-1732 © 2024 IEEE
Digital Object Identifier 10.1109/MM.2024.3353949
Date of current version 12 February 2024.

THE ARTICLES

This special issue of *IEEE Micro* captures one contribution from COOL Chips 25 in 2022 and three from COOL Chips 26 in 2023. The four articles focus on a memcapacitive, in-memory computing array for SNN compression, a 3-D rendering processor with hybrid deep neural network (DNN), a complement neural processing unit for online learning with a convolutional neural network (CNN) and SNN, and an integrity verification scheme for secure nonvolatile memory (NVM). All of them were emerging topics the COOL Chips 25 and 26.

In "A Compressed Spiking Neural Network Onto a Memcapacitive In-Memory Computing Array," Oshio et al.[A1] design a neuromorphic circuit using a low-power, memcapacitive adopting analog in-memory computing synapse. Circuit-nonlinearity-aware training is combined with network compression techniques to prevent the SNN from losing accuracy caused by the neuron circuit's nonlinearity and the synapse's low resolution. With Spice simulations, the proposed circuit performs MNIST classifications with almost no loss from ideal accuracy (97.64%) and consumes 15.7 nJ per inference.

In "A Low-Power Artificial-Intelligence-Based 3-D Rendering Processor With Hybrid Deep Neural Network Computing," Han et al.[A2] develop a low-power, AI-based 3-D rendering processor. It removes useless computations, and a hybrid neural engine, which utilizes coarse- and fine-grained sparsity exploitation, accelerates the remaining inference tasks by allocating them to the two different neural engines, which focus on zero skipping and data reusability. The proposed processor was fabricated with 28-nm CMOS technology, giving us a maximum of 118-frames/s rendering while consuming 99.95% lower power than modern GPUs.

In "COOL-NPU: Complementary Online Learning Neural Processing Unit," Kim et al.[A3] propose a complementary online learning neural processing unit, COOL-NPU, for accurate and energy-efficient online learning systems. It reduces energy consumption by combining

APPENDIX: RELATED ARTICLES

A1. R. Oshio, T. Sugahara, A. Sawada, M. Kimura, R. Zhang, and Y. Nakashima "A compressed spiking neural network onto a memcapacitive in-memory computing array," *IEEE Micro*, vol. 44, no. 1, pp. 8–16, Jan./Feb. 2024, doi: 10.1109/MM.2023.3285529.

A2. D. Han, J. Ryu, S. Kim, S. Kim, J. Park, and H.-J. Yoo, "A low-power artificial-intelligence-based 3-D rendering processor with hybrid deep neural network computing," *IEEE Micro*, vol. 44, no. 1, pp. 17–27, Jan./Feb. 2024, doi: 10.1109/MM.2023.3328965.

A3. S. Kim et al., "COOL-NPU: Complementary online learning neural processing unit," *IEEE Micro*, vol. 44, no. 1, pp. 28–37, Jan./Feb. 2024, doi: 10.1109/MM.2023.3330169.

A4. T. Kubo and S. Takamaeda-Yamazaki, "Cachet: Low-overhead integrity verification on metadata cache in secure nonvolatile memory systems," *IEEE Micro*, vol. 44, no. 1, pp. 38–48, Jan./Feb. 2024, doi: 10.1109/MM.2023.3335354.

CNN and SNN, eliminating the power overhead due to the redundant weight update. The proposed SNN core reduces the energy consumption of SNN-gradient generation by two-step encoding and reduces inference power by a hierarchical cache with lookup table mode. The proposed processor was fabricated with a Samsung 28-nm CMOS, achieving 6.94 mJ/frame and 0.73 mean average precision for object detection while reducing 47.7% energy with a slight loss of accuracy.

In "Cachet: Low-Overhead Integrity Verification on Metadata Cache in Secure Nonvolatile Memory Systems," Kubo and Takamaeda-Yamazaki[A4] propose a novel integrity verification scheme for secure NVM systems. Cachet employs set hash functions to authenticate the metadata cache, and they enable the metadata integrity update without imposing additional overhead during system recovery. Cachet reduces the execution time by 21%, NVM writes by 30%, and power consumption overheads by 22%.

Low power, low energy, power/energy efficiency, and power/energy awareness are still some of the most important factors for any kind of chip design. To cope with this subject, not only devices and circuits but also a wide variety of innovations, including architectures, algorithms, and software, are essential. The COOL Chips conference series will continue to cover all kinds of low-power and high-performance—also known as "COOL"—chips and look for future contributions.

ACKNOWLEDGMENTS

It has been a pleasure to put together this special issue on COOL chips. We thank the former Editor-in-Chief of *IEEE Micro* Lizy Kurian John for her support and guidance. We also thank Tadao Nakamura, steering committee chair of COOL Chips 25 and 26, and Makoto Ikeda, organizing committee chair, for their help in arranging this special issue. The Special Issue on COOL Chips would not have been possible without their help.

RYUSUKE EGAWA is a professor at Tokyo Denki University, Tokyo, 120-8551, Japan, and has been a program committee cochair of the IEEE COOL Chips conference series since 2022. Contact him at egawa@mail.dendai.ac.jp.

YASUTAKA WADA is a professor at the Department of Information Science, Meisei University, Tokyo, 191-8506, Japan, and has served as a program committee cochair of the IEEE COOL Chips conference series since 2022. Contact him at yasutaka.wada@meisei-u.ac.jp.