DOI: 10.4218/etr2.12316

GUEST EDITORIAL

ETRI Journal WILEY

Special issue on SoC and AI processors

Artificial Intelligence (AI) has evolved into a general technology for a wide range of purposes and has been applied in all aspects of economy and society. It has already been extensively used in various fields, including medical services, finance, security, education, transportation, and logistics, and had led to the emergence of new commercial activities, business models, and game-changing product applications. AI is a driving force to economic and social development at the forefront of the technological revolution and industrial transformation. Additionally, System-on-a-Chip (SoC) plays a vital role in post-PC era products like smartphones, tablets, and various wearable devices where form-factor, cost, and energy-efficiency, are critical drivers. It contains multiple processing parts such as the central processing unit (CPU), graphics processing unit (GPU), image processing unit (IPU), digital signal processor (DSP), video encoder/decoder, modems, and neural processing unit (NPU). Specifically, AI processors, another name for NPU, are specially optimized for mathematics and algorithms commonly used by neural networks. They can run neural networks and machine learning tasks faster and more efficiently than CPUs.

In this special issue, we have selected papers that represent the current state-of-the-art in AI processors as well as in essential SoC blocks used in radar, RF/analog, hardware security, and design methodology.

The first paper "40TFLOPS Artificial Intelligence Processor with Function-safe Programmable Many-Cores for ISO26262 ASIL-D" by Jinho Han et al. presents AI processor architecture that has high throughput for accelerating the neural network and reducing the required external memory bandwidth for processing the neural network. For high throughput, the proposed super thread core (STC) includes 128 × 128 nano cores operating at 1.2 GHz clock frequency and the general-purpose processor (GPP) core is integrated for the control of the STC and processing AI algorithm. For the functional safety that becomes very important in automotive systems, various microarchitectural techniques are adopted, including the self-recovering cache and dynamic lockstep (DLS) function, to achieve ASIL-D of ISO26262 standard fault tolerance levels. The entire AI processor fabricated in the 28-nm CMOS process yields peak performance up to 40TFLOPS at 1.2 GHz operating frequency and 1.1 V supply voltage, with a measured energy efficiency of 1.3TOPS/W and ISO26262 ASIL-D compliant, single-point fault tolerance rate equal to 99.64%.

The next paper titled "An impulse radio (IR) radar SoC for through-the-wall human-detection applications" by Piljae Park et al. proposes through-the-wall radar (TTWR) SoC and its architecture with the test standards and methods, which can be used at disaster scenes in limited visibility conditions owing to smoke, walls, and collapse debris. Additive reception based on the coherent clocks and reconfigurability can fulfill the demands for the TTWR and a clock-based single-chip IR radar transceiver is implemented in 130-nm CMOS technology. By utilizing the repetitive coherent clock schemes, the proposed SoC can achieve signal-to-noise-ratio (SNR) enhancements. Furthermore, this paper shows the test results in various pseudo-disaster conditions of the hand-held prototype radar with the proposed TTWR SoC operating in real-time.

The third paper "AB9: a Neural Processor for Inference Acceleration" by Yong Cheol Peter Cho et al. presents a neural processor for interference acceleration with the systolic tensor core (STC) by exploiting data-reuse and parallelism characteristics inherent in neural networks, while also providing fast access to large on-chip memory. AB9 shows a superior performance and power efficiency to those of a general-purpose GPU (GPGPU) for YOLOv2, and has been fabricated with a 28-nm CMOS process along with a 40 TFLOP STC that includes 32 k arithmetic units and over 36 MB of on-chip SRAM.

To alleviate the high-computational and memory-intensive burdens in deep neural networks, the following paper "Automated Optimization for Memory-efficient High Performance Deep Neural Network Accelerators" by Hyun Mi Kim et al. investigates the efficient memory structure and operating scheme, which can provide an intuitive solution for high-performance accelerators along with dataflow control. The authors propose an efficient architecture with flexibility, while operating at high frequency despite the large memory size and PE array. They demonstrate an improvement in the efficiency and usability of the architecture by presenting

an automation algorithm for optimization. The experiments show that the proposed architecture with the increased data reuse, such as a diagonal write path, improves the performance by 1.44× on average across a wide range of neural networks. The automated optimizations dramatically improve the performance from 3.8× to 14.79× that enhances usability even further.

Recently, the importance of security has emerged in various computing fields, such as mobile, biomedical, and automotive systems. The paper "An Analysis and Efficient Hardware Implementation of True Random Number Generator Based-on Beta Source" by Seongmo Park et al. proposes an efficient hardware random number generator based on a beta source. The proposed generator generates the values of "0" and "1" and provides a method to distinguish between pseudo-random and true random numbers by comparing them with simple cumulative operations. The random-number generator produces labeled data, thus indicating whether the count value is a true random number based on the bit values of the binary count value and on the comparison of the generated labeling data that are used as reference data. The generated random numbers pass the test procedures outlined in the standards SP800-22 and SP800-90B issued by the National Institute of Standards (NIST).

To improve design productivity in SoC, high-level synthesis (HLS) has become popular, and has been used to automatically synthesize a register-transfer level (RTL) circuit from a behavioral description written in a high-level programing language such as C/C++. However, HLS tools often generate the design with the larger area overhead owing to unnecessary redundant instances. In the paper entitled "Function-Level Module Sharing Techniques in High-Level Synthesis" by Hiroki Nishigawa et al., the authors present two HLS techniques for module sharing at function level and show the effectiveness with the experimental results.

The paper "Field Programmable Analog Arrays for Implementation of Generalized Balanced OTA-C Odd/Even-nth-Order Elliptic Filters" by Maha Diab and Soliman Mahmoud presents an architecture for a field-programmable analog array based on operational transconductance amplifier (OTA) as the building block that can be used in analog signal processing units operating at low frequencies such as biopotential signals. The architecture eliminates the need for switches in the signal path and has a flexible structure. Moreover, this work presents a simplified direct circuit realization method for the synthesis of OTA-C even/odd-nth-order elliptic filters. The proposed method results in an OTA-C symmetric balanced structure with a minimum number of components and grounded capacitors for a balanced design.

The final paper "W-band MMIC Chipset in 0.1 um mHEMT Technology" by Jong Min Lee et al. developed 0.1-μm metamorphic high electron mobility transistor (mHEMT) and fabricated W-band monolithic microwave integrated circuit chipset with in-house technology to verify the performance and usability of the developed technology. The direct current (DC) characteristics of mHEMT include a drain current density equal to 747 mA/mm and a maximum transconductance of 1.354 S/mm. In addition, the RF characteristics include a cut-off frequency of 210 GHz and maximum oscillation frequency of 252 GHz. To increase the frequency of an input signal, a frequency multiplier is developed that consists of three common source doublers connected in cascade. The authors also present in this paper a low-noise amplifier with a four-stage, single-ended architecture with a common source stage and a W-band IR module with an external off-chip coupler.

The Guest Editors thank all the authors, reviewers, and the editorial staff members of the ETRI Journal for making this special issue a success. We are most pleased to have been part of this effort, and for the timely publication of these high-quality technical articles.

Ji-Hoon Kim[1]
Minjae Lee[2]
Jongsun Park[3]
Ho-Young Cha[4]

[1]*Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul, Republic of Korea*
[2]*School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea*
[3]*School of Electrical Engineering, Korea University, Seoul, Republic of Korea*
[4]*School of Electronic and Electrical Engineering, Hongik University, Seoul, Republic of Korea*

Correspondence
Ji-Hoon Kim, Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul, Republic of Korea.
Email: jihoonkim@ewha.ac.kr

AUTHOR BIOGRAPHIES

Ji-Hoon Kim received the B.S. (*summa cum laude*) and Ph.D. degrees in electrical engineering and computer science from KAIST, Daejeon, South Korea, in 2004 and 2009, respectively. In 2009, he joined Samsung Electronics. In 2018, he joined the faculty of the Department of Electronic and Electrical Engineering, Ewha Womans University, where he is currently an Associate Professor. His current interests include CPU/DSP, communication modems, and low-power SoC design for security/biomedical systems. Dr. Kim is a technical committee member of the circuits and systems for communications and VLSI systems and applications in the IEEE Circuits and Systems Society. He was the recipient of the best design award at the Dongbu HiTek IP Design Contest in 2007, and the first- place award recipient at the International SoC Design Conference Chip Design Contest in 2008.

Minjae Lee received his B.S. and M.S. degrees from Seoul University, Seoul, Rep. of Korea, in 1998 and 2000, respectively, and his Ph.D. degree from the University of California at Los Angeles, Los Angeles, CA, USA, in 2008, in electrical engineering. In 2000, he was a Consultant with GCT Semiconductor Inc., San Jose, CA, USA, and Silicon Image Inc., Sunnyvale, CA, USA, where he designed analog circuits for wireless communications and digital signal processing blocks for gigabit Ethernet. He joined Silicon Image Inc., in 2001, where he developed Serial ATA products. In 2008, he joined Agilent Technologies, Santa Clara, CA, where he was involved in the development of the next-generation high-speed ADCs and DACs. Since 2012, he has been with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, where he is currently an Associate Professor. Dr. Lee was a recipient of the 2007 Best Student Paper Award at the VLSI Circuits Symposium in Kyoto, Japan, and the 2015 Distinguished Lecture Award from GIST.

Jongsun Park received his B.S. degree in electronics engineering from Korea University, Seoul, Rep. of Korea, in 1998 and his M.S. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2000 and 2005, respectively. He joined the faculty of the Department of Electrical Engineering of Korea University, Seoul, Rep. of Korea, in 2008. He was with the Digital Radio Processor System Design Group, Texas Instruments, Dallas, TX, USA, in 2002. From 2005 to 2008, he was with the Signal Processing Technology Group, Marvell Semiconductor Inc., Santa Clara, CA, USA. His research interests focus on variation-tolerant, low-power, high-performance VLSI architectures and circuit designs for digital signal processing and digital communications. Dr. Park is a member of the Circuits and Systems for Communications Technical Committee of the IEEE Circuits and Systems Society. He served as a Guest Editor of the IEEE Transactions on Multi-Scale Computing Systems. He has also served in the technical program committees of various IEEE/ACM conferences, including ICCAD, ISLPED, ISCAS, ASP-DAC, HOST, VLSI-SoC, and APCCAS. He is an Associate Editor of the IEEE Transactions on Circuits and systems II: Express Briefs.

Ho-Young Cha received his B.S. and M.S. degrees in electrical engineering from the Seoul National University in Seoul, Rep. of Korea, in 1996 and 1999, and his Ph.D. in electrical and computer engineering from Cornell University in Ithaca, NY, in 2004. He was a Postdoctoral Research Associate at Cornell University until 2005, where he focused on the design and fabrication of wide-bandgap semiconductor devices. He was with the General Electric Global Research Center in Niskayuna, NY, from 2005 to 2007 where he developed wide-bandgap semiconductor sensors and high-power devices. Since 2007, he has been a Professor in the School of Electronic and Electrical Engineering, Hongik University, Seoul, Rep. of Korea. His research interests include wide-bandgap semiconductor devices. He has authored over 110 publications in his research area.