Guest Editorial: Alternative Computing and Machine Learning for Internet of Things

THE impending Internet of Things (IoT) wave is promising to affect every aspect of our daily lives, ranging from smart things to smart buildings, smart cities, and smart environments. A lot of attention has been devoted to the tsunami of data produced by IoT, and the related means of extracting useful actionable information from it, spawning efforts in Big Data processing and machine learning. Yet, all of this does little to address the need for IoT to capture, interpret, and act on this wall of (noisy) information at the right time, at the right place, and in the right form. Conventional computing systems are a poor match to the needs of this emerging massively distributed real-time system. Hence, alternative computing techniques present an attractive alternative, trading off computational resolution for significant gains in quality-of-service energy efficiency and robustness. This observation is based on the conjecture that most applications related to IoT have an inherent error resilience and are evolutionary (that is, learning-based). Alternative computing strategies may be conceived at every level of the design hierarchy, starting from the device level with novel 3-D nonvolatile memory/logic combinations, or at the architectural level by shifting away from the traditional von Neumann architecture to different computing paradigms such as neuromorphic and/or stochastic computation all the way up to the algorithmic and data representation levels.

To support and enable this transformation, this special issue invited the hardware and software communities to come together to present their integrative research in various topics, including: 1) IoT-aware nano-CMOS and beyond-CMOS devices, sensors, and circuits; 2) specialized and modern memory systems for IoT (e.g., memristor, STT-RAM, FeRAM, and so on); 3) reconfigurable embedded sensing and actuating, enabling runtime selection of quality, operation mode, and parameter settings of IoT devices; 4) alternative architectures for IoT-specific Big Data search, predictive analytics, deep learning, high-dimensional data, feature selection, and feature transformation; 5) IoT-specific approximate design, exploration, and optimization; 6) brain-inspired and neuromorphic components, circuits, and systems for IoT; 7) accelerators for IoT (e.g., machine learning, neuromorphic, and cognitive computing); and 8) case studies for alternative computing in the IoT era. In this special issue, we received 37 submissions and a total of six papers were accepted. The high level of competition has led to the selection of top-level contributions covering a wide spectrum of topics in the domain of alternative computing for IoT.

In "VLSI implementation of deep neural network using integral stochastic computing," Ardakani *et al.* [item 1) in the Appendix] present an efficient FPGA implementation of a deep neural network using a Virtex7 FPGA. The results show 45% and 62% reduction in area and latency on average compared with the state-of-the-art architectures. Moreover, the authors present a quasi-synchronous implementation resulting in 35% reduction in energy consumption compared with the binary radix implementations.

Wang *et al.* [item 2) in the Appendix] in "Resilience-aware frequency tuning for neural-network-based approximate computing chips" study the impact of aggressive frequency scaling as well as the resilience feature of neural network (NN) accelerators in the presence of timing errors induced by process variations, noises, and so on using a novel timing analysis method. Based on this study, the authors propose a holistic frequency scaling for NN-based approximate computing circuits based on an in-field retraining mechanism. Thanks to the proposed dynamic approach, the proposed NN can operate with much higher operating frequencies compared with the existing static NNs.

In "A 17.5-fJ/bit energy-efficient analog SRAM for mixed-signal processing," Lee *et al.* [item 3) in the Appendix] propose a novel SRAM to cope with unnecessary redundant analog-to-digital and digital-to-analog conversion in neuromorphic chips using approximate operations such as approximate write operation in inout blocks. The proposed SRAM shows promising results in a way that it achieves 64% and 1.3× better energy-efficiency and throughput compared with the conventional SRAM architectures, respectively.

Onizawa *et al.* [item 4) in the Appendix] in "Area/energy-efficient gammatone filters based on stochastic computation" present a novel stochastic computation approach to be able to replace power-and-area hungry multiplier used in digital filters resulting in an area efficient hardware. Moreover, the authors describe a gain balancing technique to

Digital Object Identifier 10.1109/TVLSI.2017.2742098

significantly improve the computation accuracy of the stochastic hardware. In addition, an efficient dynamic scaling technique is proposed to reduce the number of required stochastic bits.

In "Near-threshold RISC-V core with DSP extensions for scalable IoT endpoint devices," Gautschi *et al.* [item 5) in the Appendix] bring the idea of near-threshold (NT) computing to the IoT endpoint devices in order to achieve higher energy efficiency, and the performance scalability. The authors present a novel NT RISC-V processor core based on instruction extension and microarchitecture optimization in particular in shared memory hierarchy. According to the results, the proposed core is 3.5× faster and 3.2× more energy-efficient on average compared with the traditional cores.

Kanduri *et al.* [item 6) in the Appendix] in "Accuracy-aware power management for many-core systems running error-resilient applications" present a novel approximate dynamic voltage and frequency scaling (DVFS) technique by leveraging accuracy-performance tradeoffs in IoT and machine learning applications. The authors propose to consider the sensitivity of applications to timing errors to adjust the power-performance of the system by DVFS. The results show that the proposed approximate DVFS leads to 5× better performance per energy and 1.9× higher throughput than the traditional DVFS systems.

ACKNOWLEDGMENT

We sincerely hope that the reader finds this special issue useful and that it will inspire further research in these very important areas of IoT, machine learning, and Big Data. We would like to thank all authors who submitted papers to this special issue. Special thanks go to the referees for their time and diligence during the review process and for providing us with high-quality reviews. Finally, we would also like to thank Prof. K. Chakrabarty, Editor-in-Chief of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, for offering us the opportunity to edit this special issue.

FARSHAD FIROUZI
MSG Systems AG
Ismaning, Germany

BAHAR FARAHANI
University of Tehran
Tehran, Iran

ANDREW B. KAHNG
University of California
at San Diego
La Jolla, CA USA

JAN M. RABAEY
University of California
at Berkeley
Berkeley, CA USA

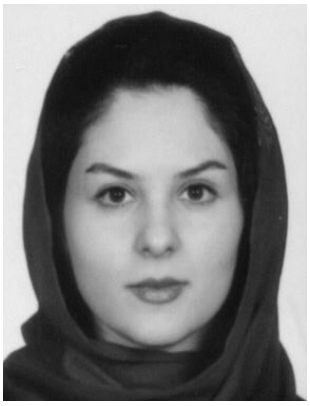
NATASHA BALAC
University of California
at San Diego
La Jolla, CA USA

APPENDIX

1) A. Ardakani, F. Leduc-Primeau, N. Onizawa, T. Hanyu, and W. J. Gross, "VLSI implementation of deep neural network using integral stochastic computing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2688–2699, Oct. 2017, doi: 10.1109/TVLSI.2017.2654298.
2) Y. Wang, J. Deng, Y. Fang, H. Li, and X. Li, "Resilience-aware frequency tuning for neural-network-based approximate computing chips," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2736–2748, Oct. 2017, doi: 10.1109/TVLSI.2017.2682885.
3) J. Lee, D. Shin, Y. Kim, and H.-J. Yoo, "A 17.5-fJ/bit energy-efficient analog SRAM for mixed-signal processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2714–2723, Oct. 2017, doi: 10.1109/TVLSI.2017.2664069.
4) N. Onizawa, S. Koshita, S. Sakamoto, M. Abe, M. Kawamata, and T. Hanyu, "Area/energy-efficient gammatone filters based on stochastic computation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2724–2735, Oct. 2017, doi: 10.1109/TVLSI.2017.2687404.
5) M. Gautschi *et al.*, "Near-threshold RISC-V core with DSP extensions for scalable IoT endpoint devices," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2700–2713, Oct. 2017, doi: 10.1109/TVLSI.2017.2654506.
6) A. Kanduri *et al.*, "Accuracy-aware power management for many-core systems running error-resilient applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2749–2762, Oct. 2017, doi: 10.1109/TVLSI.2017.2694388.

Farshad Firouzi was a Senior Scientist with imec, Belgium; KIT, Germany; SRC, USA; and SI LAB, Iran. He is currently a Senior Researcher and a Program Manager with MSG Systems AG, Ismaning, Germany, where he is involved in statistics, machine learning, neuromorphic computing, Internet of Things, Big Data, and eHealth Projects. He has authored over 40 journal and conference papers.

Dr. Firouzi has served as a Reviewer of the IEEE TRANSACTIONS ON VLSI, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, the *ACM Transactions on Design Automation of Electronic Systems*, *Microelectronics Reliability* (Elsevier), *Microelectronics* (Elsevier), ICCAD, DAC, ATS, and DATE, a Managing Guest Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, *Future Generation Computer Systems*, and the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS, a Program Chair of Complexis and IoTBDS, and a Program Committee Member in several conferences, including ICCAD and ATS.

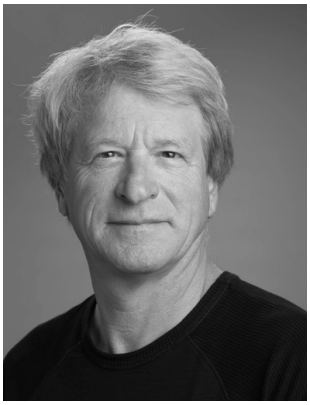


Bahar Farahani has served as an Advisor and a Consultant to various companies/organizations. She is currently the CEO and a Co-Founder of the Pirouzan Group, Iran, a top-leading company that delivers Internet of Things-aware eHealth products and services—ranging from innovative connected medical devices, to cloud-based software for medical record-keeping, smart screening, disease prediction, e-prescription, and big data analysis, and to top-quality hospitals and clinics.

Dr. Farahani has served as a Guest Editor of various journals, such as the *Future Generation Computer Systems* (Elsevier), the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, and the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS.

Andrew B. Kahng (F'10) received the Ph.D. degree in computer science from the University of California at San Diego (UCSD), La Jolla, CA, USA, in 1989.

He was with the Computer Science Department, University of California at Los Angeles, Los Angeles, CA, USA, from 1989 to 2000. Since 2001, he has been with the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering, UCSD, where he holds the Endowed Chair in high-performance computing. He has authored or co-authored over 350 journal and conference papers and three books. He holds 22 U.S. patents. His current research interests include IC physical design, the design-manufacturing interface, and combinatorial algorithms and optimization.



Jan M. Rabaey (F'95) holds the Donald O. Pederson Distinguished Professorship with the University of California at Berkeley, Berkeley, CA, USA. He is currently a Founding Director of the Berkeley Wireless Research Center, University of California at Berkeley, and the Berkeley Ubiquitous SwarmLab, University of California at Berkeley, and has served as the Electrical Engineering Division Chair at Berkeley twice. He has made high-impact contributions to a number of fields, including advanced wireless systems, low-power integrated circuits, sensor networks, and ubiquitous computing. His current interests include the conception of the next-generation integrated wireless systems, as well as the exploration of the interaction between the cyber and the biological world.

Dr. Rabaey is a member of the Royal Flemish Academy of Sciences and Arts of Belgium. He was a recipient of major awards, amongst which the IEEE Mac Van Valkenburg Award, the European Design Automation Association Lifetime Achievement Award, and the Semiconductor Industry Association University Researcher Award. He has received honorary doctorates from Lund (Sweden), Antwerp (Belgium), and Tampere (Finland). He has been involved in a broad variety of start-up ventures.

Natasha Balac is currently the Founder, the CEO, and the President of Data Insight Discovery, Solana Beach, CA, USA. Concurrently, she is directing the Interdisciplinary Center for Data Science (ICData), Calit2/Qualcomm Institute, University of California at San Diego (UCSD), La Jolla, CA, USA, where she is also a Lecturer with the Computer Science Department, Rady School of Management and the UCSD Extension. She has over 20 of experience in the field of predictive analytics and data mining that spans multiple application domains including fraud detection, risk management, predictive maintenance, marketing mix models, logistics optimization, and data visualization. She has provided over 100 courses, presentations and training workshops in the field as an experienced educator and instructor.