Guest Editorial
Energy-Quality Scalable Circuits and Systems for Sensing and Computing: From Approximate to Communication-Inspired and Learning-Based

Massimo Alioto, *Fellow, IEEE*, Vivek De, *Fellow, IEEE*, and Andrea Marongiu, *Member, IEEE*

I. Motivation and Technology Scaling Perspective

SUSTAINED energy efficiency improvements have been instrumental for the vertiginous evolution of electronic systems with computing, sensing, communication and storage capabilities. Energy efficiency improvements are indeed crucial for continued increase in the performance under a limited power budget, reduced operating cost, as well as for untethering traditionally wired systems. This is indeed true for high-performance systems subject to heat removal limitations (e.g., server blades), as well as for operational cost considerations when the cost of electricity is a major fraction of the total cost, as in the case of datacenters [1], or the more recent crypto-currency mining endeavors [2]. Energy reductions are also critical in portable electronics, due to the limited thermal budget and battery energy availability. Similarly, energy reductions are essential in miniaturized energy-autonomous systems such as sensor nodes, hearables, wearables and others, due to their tightly constrained energy source [3]. Overall, energy efficiency improvements have historically permitted the continuous size down-scaling and lifetime extension of electronic systems (see, [4]).

As a representative example with a relatively long-lasting trend, computers have benefitted tremendously from exponential energy improvements, as quantified by Koomey's law [5]. Koomey's empirical observation transcends Moore's law, as it has been shown to hold even when vacuum tubes and discrete transistors were used, well before the advent of the integrated circuit [5]. As shown in Fig. 1, a consistent two-order of magnitude energy reduction per decade has been observed since the inception of microprocessors. A very similar trend can be observed for many other classes of systems, such as Digital Signal Processors, according to the Gene's law in the same figure [6]. However, 90% of this energy reduction has been historically provided by advances in the integrated circuit manufacturing process and transistor miniaturization [7]. More recently, technology scaling has offered limited energy gains of only about 15% per generation. Considering the limited number of CMOS generations ahead, energy gains of a few units seem difficult to exceed. Even if CMOS is being complemented with beyond-CMOS technologies [8], CMOS is not expected to be replaced any time soon, even long after the prospective end of its scaling [9]. Even from optimistic and now outdated projections such as the ITRS roadmap, energy reductions by 3-4X are expected from CMOS technology scaling in the next decade [7]. At the same time, the several techniques adopted after the end of Dennard's scaling to continue the historical energy scaling are running out of steam. This is for example the case of low-voltage operation (as limited by the CMOS threshold voltage), parallelism (whose degree is limited by the inherent application parallelism), and application-specific acceleration.

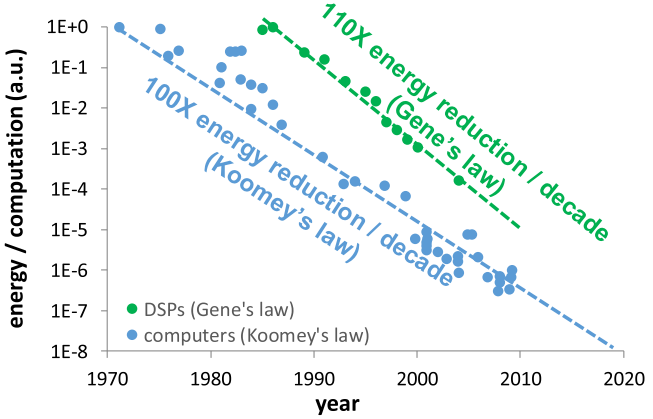


Fig. 1. Energy scaling of computers (Koomey's law [5]) since the advent of the integrated circuit, and of Digital Signal Processors (Gene's law [6]).

II. Energy-Quality Scalable Circuits and Systems

According to the above considerations, major shifts from traditional sensing/processing/communication design paradigms are now extremely important to enable sustained exponential energy reductions, in spite of the slowdown in technology scaling. In other words, new design dimensions need to be explored and traded off with energy. Interestingly, significant room for energy reduction is allowed by energy-quality (EQ) scalable circuits and systems [12], as a design dimension that is orthogonal to technology scaling. EQ-scalable design aims to dynamically and explicitly trade off energy and quality at any level, from circuit to architecture, algorithm, and system [12]. Similarly, EQ scaling is applicable to any sub-system, ranging from sensors, to analog interfaces, processing, storage, and wireless communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JETCAS.2018.2865783

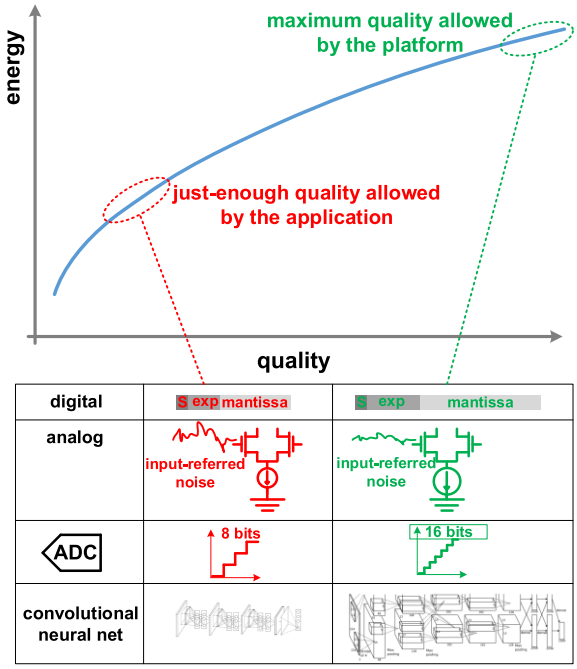


Fig. 2. Energy vs quality in a generic sub-system of an integrated circuit (e.g., digital, analog, analog-to-digital converters, convolutional neural networks).

As an illustrative example in Fig. 2, intuition suggests that greater quality can be generally achieved in arbitrary digital circuits by spending more energy, for example by increasing the arithmetic precision, and hence the accuracy of the number representation. Similar considerations hold for analog circuits, where quality can be for example raised by increasing bias currents and hence energy per sample, while reducing the input-referred noise. As another example, the same considerations hold for the resolution in analog-to-digital converters, where the quantization noise can be reduced at the cost of higher energy per conversion. At a higher level of abstraction, higher classification quality (i.e., accuracy) is achieved with convolutional neural networks having larger model size and training dataset, although at larger energy per classification. In EQ-scalable circuits and systems, the above relationship between energy and quality is exploited by adjusting the quality to the minimum allowed by the application, and hence operate at the minimum possible energy and avoid the energy waste that is traditionally paid for when operating at the fixed quality offered by the specific hardware platform.

Quality is generally defined by the output fidelity to a reference, desirable, exact or correct output. Quality is hence application-dependent, and is quantified with several well-established metrics. A few examples are accuracy, sensitivity, specificity or misclassification rate in machine learning systems for classification. PSNR or other perceptive metrics (e.g., SSIM) are used in video processing. False alarm rate is used to quantify the ability of IoT sensors to capture correct events. Mean Error Distance (MED) is used to characterize quality in approximate digital circuits.

Several approaches that can save energy at degraded quality have been proposed, although they have been mostly focused on processing. All these approaches fall under the general

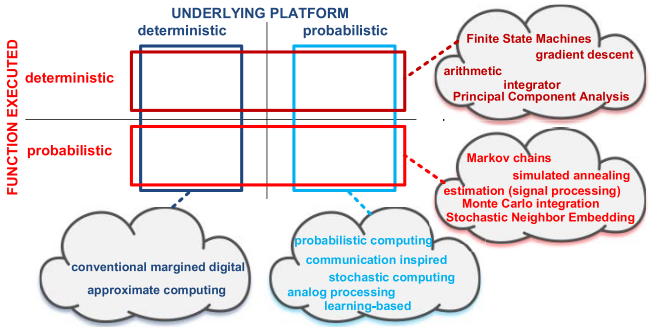


Fig. 3. Taxonomy of EQ-scalable functions and underlying platforms, and examples.

approach of energy-quality scalable circuits and systems, although they have initially developed as independent research areas. These systems can be categorized based on the operation of the underlying platform, and the type of function that is executed onto or by it. Platform and function can be either deterministic or probabilistic, as shown in Fig. 3. Deterministic platforms have perfectly repeatable and predictable behavior, as they are not affected by random phenomena, by construction or by design margining (see below). Probabilistic platforms are not perfectly repeatable, as they are affected by random phenomena that cannot be known or measured. Similar considerations hold for the functions that are implemented on a given platform. EQ scaling can be independently applied to either the platform, or the considered function, or both.

As in Fig. 3, examples of deterministic platforms are:

- MAINSTREAM DIGITAL: conventional CMOS digital circuits with clock cycle margined for the worst-case process/voltage/temperature variations hide the statistical behavior of transistors (i.e., all margined silicon dice have equivalent behavior)
- APPROXIMATE COMPUTING: quality is adjusted by setting the arithmetic precision, or introducing other quality-driven simplifications (e.g., gate pruning). These approaches are generally based on deterministic platforms and execute deterministic functions.

On the other hand, examples of probabilistic platforms that have been explored in terms of EQ scaling are:

- COMMUNICATION-INSPIRED: digital sub-systems are treated like a noisy channel, the noise being caused by inaccuracies or occasional faults.
- STOCHASTIC COMPUTING: computations are based on streams of random bits encoding signals with 0/1 density. The advantage is in the simplification of complex computations into bit-wise operations on the streams.
- PROBABILISTIC COMPUTING: building blocks down to logic gates are treated as probabilistic systems, whose output is governed by probabilities rather than deterministic values.
- LEARNING-BASED: learning-based approaches learn how to discriminate noise, circuit non-idealities, and irrelevant information from data. Learning is affected by the inevitable noise in the training dataset, thus making such systems probabilistic when learning occurs online (i.e., dataset is noisy and not repeatable).

Analogous considerations can be made on the type of function that is executed on the underlying platform. As a few examples in Fig. 3, Finite State Machines are defined by deterministic transitions between states, whereas Markov models are defined by their probabilities. In the field of optimization algorithms, deterministic algorithms such as gradient descent have probabilistic counterparts, such as simulated annealing and stochastic gradient descent. The evaluation of parameters through deterministic arithmetic computations has a probabilistic counterpart, as in the case of estimation in the area of signal processing. Numerical deterministic methods for the evaluation of integrals has Monte Carlo integration as its probabilistic counterpart. Dimensionality reduction through deterministic Principal Component Analysis (PCA) has a popular probabilistic counterpart in the Stochastic Neighbor Embedding. Beyond these few examples, similarities can be found in a wide range of existing deterministic and probabilistic functions [1], and the common kernels across applications [11].

A wide range of demonstrations of EQ-scalable System on Chip (SoC) and related building blocks with substantial energy reductions have been presented in recent years. As a few of the many available examples (including those from the same authors), demonstrations have been presented for EQ-scalable memories [13], analog-to-digital converters [14], microprocessors [15], accelerators for vision [16], and engines for machine learning [17]. A deeper and broader analysis of the properties and requirements of EQ-scalable systems is provided in the overview paper by the same authors in this special issue [18].

III. The Challenges Ahead

Several challenges need to be addressed to extract the potential energy gains of EQ scaling, while allowing their systematic and efficient design. At the sub-system level, new methods and techniques to inexpensively insert EQ knobs in all SoC components are needed, from analog and sensor interfaces to processing, power management, algorithms, and software.

Among other challenges, quality needs to be inexpensively sensed at low area and energy/power penalty in order to retain the potential energy advantages offered by EQ scaling. Methods and techniques are also needed to make the quality degradation graceful, extending the energy savings when operating at a given (lower) quality. Novel design paradigms and optimization methodologies are also needed to supervise the energy-quality control, based on an underlying application-to-hardware framework.

Novel frameworks are also needed to make EQ scaling techniques more generally applicable (e.g., general-purpose platforms), spanning multiple levels of abstraction (e.g., circuit, architecture, algorithm, software) and sub-systems. In addition, innovative frameworks and techniques are needed to minimize the overall energy via true adaptation (e.g., on-chip learning), while keeping quality within bounds in a real-time and context-aware fashion.

The above challenges require a highly inter-disciplinary collective effort, as they lie at the intersection of circuits and systems, solid-state circuits, CAD, architectures, machine learning, signal processing (e.g., computer vision, audio), and the related communities. This justifies this special issue, and explains the significant diversity of the authors' contributions. In particular, this special issue is focused on scientific and technology advances in the field of energy-quality scalable circuits, architectures, algorithms and design methodologies. Given the very large number of submissions, the abundant contributions are organized into two journal issues, i.e. Part I and Part II. The detailed content of the two parts is summarized in the following two sections.

IV. Content of Part I of This Special Issue: From Circuits to Architectures

The first part of the special issue is organized six thematic sections, starting from energy-quality scaling at circuit level. Moving to a higher level of abstraction, the next section focuses on gate- and bitcell-level approximate computing and approximate architectures. Then, energy-quality scalable architectures for neural networks and machine intelligence are considered. Several case studies on EQ-scalable hardware systems are then presented. The journal issue ends with circuit/architectural approaches that leverage beyond-CMOS technologies.

Invited Paper: The paper by Naresh Shanbhag from the University of Illinois at Urbana Champaign has been invited to bridge the traditionally separate areas of on-chip neural networks, mixed-signal circuit techniques, as well as increasingly popular architectures performing computations within the memory itself. In particular, mixed-signal computation is executed at the periphery of the memory storing the data, thus reducing energy and performance penalty due to data movement. Given its nature, the paper has been placed at the beginning of the section on architectures for neural networks and machine intelligence.

A. EQ-Scalable Circuits

The contribution by Chang *et al*. from Purdue University is about an energy-quality scalable radio-frequency front-end for integrated IoT sensor nodes. The front-end enables a wide tradeoff between input referred noise and power consumption, which can be adjusted from few tens of μWs up to the mW range.

The paper by Behroozi *et al*. from University of Wisconsin and Purdue University proposes an EQ-scalable serial bus (AxSerBus) exploiting the locality of sensory data and the intrinsic error resiliency of sensing, encoding the differences of sensory data in various modes depending on the magnitude of the differences. Since small differences are more frequent than large differences, graceful quality degradation is achieved.

The paper by Seo *et al*. from Yonsei University introduces a low-power scan test method using statistics-based scan chain reordering. Based on a new scan partition, the scan chain reordering method also relieves the routing overhead.

B. Gate- and Bitcell-Level Approximate Computing

The work by Ansari *et al*. from the University of Alberta propose approximate compressors for energy-quality scalable multiplication, leveraging properly encoding compressor's

inputs. The concept is validated in vision tasks such as image sharpening, and MIMO antenna communication systems with error control coding.

Qiqieh *et al*. from Newcastle University propose a novel approximate multiplier design based on significance-driven logic compression, i.e. lossy compression of the partial products based on their progressive bit significance. Energy-quality scaling is achieved by configuring the degree of compression and logic clustering.

Kim *et al*. from Seoul National University and Kyung Hee University present a novel approach and design methodology for SRAMs with heterogeneous bitcell size. The latter is optimized to maximize the quality of multimedia signals in the presence of variations that limit the bitcell yield. Power is saved at iso-area, thanks to the more favorable energy-quality tradeoff under voltage scaling.

C. EQ-Scalable Approximate Architectures

Onizawa *et al*. from Tohoku University and McGill University propose an energy-quality scalable architecture for 2D Gabor filtering. Power gating and voltage scaling are adaptively applied to reduce energy, while adjusting the quality through the length of the stochastic bit streams used for computation.

The paper by Liu *et al*. from the University of Alberta introduces a stochastic architecture for Deep Belief Networks based on 1) a reconfigurable structure to implement fast greedy learning algorithm, 2) an adaptive moment estimation circuit to improve the speed of training, 3) an approximate stochastic computing activation unit to implement different types of activation functions in the neurons. While keeping the accuracy close to floating-point design, significantly lower energy is achieved for both online learning and inference processes.

The contribution by Ariyarathna *et al*. from the University of Akron, Case Western and University of Calgary explores a baseband multi-beamforming method based on the spatial Fourier transform and an analog circuit approach leveraging sparse factorization.

The work by Afzali-Kusha *et al*. from the University of Southern California introduces an energy-quality scalable coarse grain reconfigurable architecture allowing voltage over-scaling. Operating voltages of processing elements are tuned and minimized subject to an output quality constraint.

D. Architectures for Neural Networks and Machine Intelligence

The above mentioned invited paper by Shanbhag *et al*. from the University of Illinois at Urbana Champaign presents a deep in-memory architecture for convolutional neural networks. The architecture reduces the energy and performance penalty due to data movement, improving the energy-delay product by an order of magnitude.

The paper by Heo *et al*. from Washington State University presents a time-domain matrix multiplier interface, based on multi-bit back-gate-driven delay elements enabling energy-quality scaling. A single-chip solution is demonstrated that includes a high-throughput digitally-driven accelerator, and a low-energy analog front-end mode. The system is validated on image recognition of handwritten digits.

In the contribution by Chen *et al*. from University of Washington and Fudan University, an accelerator for gated-recurrent-unit (GRU) inference is described, which includes on-chip incremental learning for sequential modeling. Each GRU accelerator is optimized and enhanced for efficient gradient computation. The design is validated in applications such as hand-written digit recognition, semantic natural language processing, and biomedical waveform based seizure detection

E. Case Studies of EQ-Scalable Systems

The contribution by Schabel *et al*. from North Carolina State University presents multifunctional data-centric accelerators (MDCAs) with dynamic datapath configuration for artificial neural network processors. Energy and area are traded off with reconfigure-ability as quality metrics of hardware flexibility.

Rovere *et al*. from ETH Zurich and Miromico AG propose a wake-up circuit with adaptive activity-sampling rate scaling. Being triggered upon event occurrence, the digital classifier is asynchronous and trained to minimize false positives to make the energy-quality tradeoff more favorable. The concept is applied to hand gesture recognition, and pathologic ECG beat detection.

The paper by Huang *et al*. from University of California at Berkeley introduces a neuro-inspired spike pattern classifier for an artificial olfactory system. The classifier is based on an analog feature extraction front-end converting the sensor signal into spike patterns, which are represented as high-dimensional sparse vectors. The resulting system is shown to be highly robust against parametric variations and operation failures.

The contribution by Chiang from National Chiayi University describes an integrated system comprising a MEMS and the read-out circuitry to detect seismic activity. A feedback architecture dynamically applies a negative electrostatic force that nulls the position of the physical structure implementing the accelerometer. The system was validated through vibration measurements in a real bridge.

Wells *et al*. from GeorgiaTech introduce a minimal-power object tracking system that dynamically adjusts the image sampling density, based on the content and the saliency of the regions within the scene. Also, voltage/frequency scaling is adopted to further reduce the consumption.

The paper by Ko *et al*. from GeorgiaTech is about an ROI-based algorithm for intelligent surveillance systems based on wireless image sensors, which also includes on-line rate controller to further reduce energy. Energy-quality scalability of various ROI-based processing approaches is also analyzed and compared.

In the contribution by Mohammed *et al*. from University College London, a fully on-chip interface processor that can perform online processing of local field potentials at the point of recording is proposed. The processor is the core of a system that acquires, identifies and wirelessly transmits Parkinsonian states.

Xu *et al*. from IMEC and Technische Universiteit Eindhoven introduce a flexible integrated system to be used across different applications, and with a widely adjustable energy-

quality tradeoff. In terms of applications, the same integrated circuit can be used for bio-impedance spectroscopy, respiration and ECG signal monitoring. The energy-quality tradeoff is managed through various EQ tuning knobs, such as the ADC resolution and sample rate, the selection of frequency points, bandwidth, bias currents, oversampling ratio and MLS averaging.

F. EQ Scaling Under Beyond-CMOS Technologies

The paper by Farkhani *et al*. from Aarhus Universitet presents an MTJ real-time reading method in a spintronic system for neuromorphic computing, which improves the responsiveness of MTJ-based neurons. Performance and energy improvements are achieved thanks to faster neuron operation, thanks to the early read termination.

The contribution by Tang *et al*. from University of Utah and EPFL is focused on RRAMs and their comparison with SRAMs in the context of on-chip memories and FPGAs. RRAMs are shown to be more area and energy efficient.

V. Content of Part II of This Special Issue: From Algorithms to Design Methodologies

The second part of this special issue is focused on higher levels of abstraction, compared to the first part. In particular, it is centered around algorithms and design methodologies for EQ-scalable systems, including neural network design, inference and training. The issue is opened by an overview paper by the same authors, which presents the state of the art, the related research sub-areas, as well as the open challenges and possible directions. The subsequent papers are organized according to the following sub-areas.

A. Energy-Quality Scalable Frameworks for Acquisition and Classification

The work by Salehi *et al*. from the University of Central Florida presents an adaptive framework for energy-aware acquisition of spectrally-sparse signals, along with a spin-based Adaptive Intermittent Quantizer. The sampling rate and resolution are dynamically tuned based on the incoming signal and the design target.

The contribution by Shoaran *et al*. is about an architecture for asynchronous tree operation and sequential feature extraction to improve state-of-the-art gradient-boosted-based solutions in terms of energy-area-latency product. From the point of view of dynamic EQ scaling, the architecture offers the flexibility to accommodate variable tree counts specific to each patient, to trade the predictive accuracy with energy.

Mo *et al*. from INRIA and IRISA introduces a Mixed-Integer Linear Programming (MILP) model, with two algorithmic methods (basic and accelerated version) to find the optimal solution. A problem decomposition is discussed to provide a controllable way to trade-off the quality of the solution and the computational complexity.

B. Energy-Quality Scalable Algorithms Leveraging Approximate Computing

The paper by Lee *et al*. from Queen's University Belfast explores transprecision techniques based on the knowledge of algorithm numerical properties to aggressively reduce precision. Several EQ knobs are explored, such as precision in loops, dynamic algorithm restructuring based on runtime numerical behavior, accuracy checks removal.

The work by Najafi *et al*. from Universität Bremen and Universität Hannover introduces a systematic design framework for approximate adders, including hybrid and non-equally segmented architectures. The framework permits to identify optimal configurations and predict the energy benefits of approximations.

Camus *et al*. from EPFL introduce a methodology to design approximate arithmetic circuits by artificially inserting and exploiting false paths. This technique is demonstrated for the design of an approximate adder trading off arithmetic precision for energy efficiency. High-significance carry stages are monitored to cut the carry propagation chain at lower-significance positions, and an input-induced cut mechanism is introduced to improve the output quality.

The contribution by Yantir *et al*. from the University of California at Irvine proposes a design methodology for approximate in-memory computing that combines both voltage and precision scaling in convolutional neural network acceleration. The approach is applied to an architecture based on an associative memory.

C. Energy-Quality Scalable Training and Inference Algorithms for Neural Networks

Kim *et al*. from Korea University introduce an efficient algorithm to predict and skip convolutions generating zero outputs for convolutional neural networks. A two-step zero prediction approach is developed to trade off accuracy for energy, based on the analysis of the spatial surroundings of output feature maps. A most significant bits-only computation scheme is also proposed, keeping least significant bits constant.

The work by Shrestha *et al*. from Syracuse University and Lawrence Livermore National Laboratory is focused on a modular approach to convert a standard LSTM to a spike-based LSTM, and map it into a spike-based platform. The energy-quality tradeoff is analyzed for several applications on the IBM TrueNorth platform, such as parity check, Extended Reber Grammar and Question classification.

Sarvar *et al*. from Purdue University study the impact of cross-layer approximations from neural network size, to weight pruning, approximate MAC and approximate memory. A cross-layer framework to co-optimize these techniques is also introduced.

The contribution by He *et al*. from the Washington University in St. Louis and the University of the Chinese Academy of Sciences proposes a training framework for energy-accuracy scalable neural networks that are made more resilient to make the energy-quality tradeoff more graceful and favorable. The main idea is to set weights that sets the loss function to a nearly-global minimum with flat valleys around it, thus

making the effect of weight approximations on accuracy less pronounced.

Koteshwara *et al*. from the University of Minnesota propose machine learning classifiers based on an architecture that allows on-demand incremental-precision adjustment. The latter starts from a low precision baseline, dynamically trading off quality with energy and throughput. Energy-quality scaling is also enabled by adjustable thresholds for multi-level classification. The concept is applied to seizure detection.

The work by Zhang *et al*. from the University of Wisconsin Madison introduces a training framework for model compression and energy-accuracy scalable neural networks, enabling an order of magnitude lower complexity in both processing and memory. Simplifications are proposed to eliminate redundant neurons across different layers.

The contribution by Truong *et al*. from RMIT University, the University of Sydney and others proposes hardware-friendly convolutional neural networks for real-time inference, based on integer computation in the training phase, as opposed to quantized network that are trained in floating-point representation. The approach is validated with multiple time-series datasets, and is shown to be robust against precision downscaling.

The work by Galindez *et al*. from KU Leuven is about a circuit-aware machine learning scheme that dynamically tunes the quality of sensor front-ends to trade off energy/power and quality (e.g., through circuit-level trade off noise tolerance and power). This tradeoff is studied analytically, and a run-time quality control framework is proposed. Tuning also makes the system robust against sensor failure contexts, as shown with various datasets.

D. Energy-Quality Scalable Design Methodologies for Systems Performing Pattern Recognition and Data Sensemaking

Kang *et al*. from KAIST introduce a 3-D face frontalization processor for mobile devices. The proposed processor is EQ-scalable in terms of image resolution and frontalization accuracy, thanks to a scalable processing engine architecture with workload adaptation, an accuracy-scalable weight quantization scheme with K-means clustering, and a zero skipping technique to prevent fetching of unnecessary data from off-chip memory.

The contribution by Goetschalckx *et al*. from KU Leuven explores architectures based on a hierarchy of increasingly complex classifiers, each trained for a specific sub-task. Such architectures are shown to be substantially more energy efficient than wake-up architectures. The concept is applied to speech recognition and visual object detection. A design framework is introduced to minimize the energy for a given quality.

Mangia *et al*. from Universitá degli Studi di Bologna, Politecnico di Torino and others explore compressed sensing in the context of a low-cost compression stage in a sensor network architecture. Local hubs collect sensor data and relay their compressed version to a remote concentrator. Paired with a strategy promoting diversity between the set of readings collected by different hubs, the framework can substantially reduce the energy requirements, while preserving robustness against local communication failures.

Acknowledgements and Perspectives

This special issue would not have been possible without the contribution and dedication of many people from our community. First, we are indebted with the many hundreds of reviewers who helped highlight the most impactful scientific contributions, and make this special issue a reality. We are also grateful to the JETCAS Editor-in-Chief Prof. Eduard Alarcon for his unconditioned support and valuable insights, and the JETCAS admin Ms Desiree Noel for her precious help with managing the submissions and the review process. We also thank Prof. Jan Rabaey from the University of California at Berkeley for interesting discussion and feedback on some aspects of this editorial.

Last but not least, we would like to thank the hundreds of authors who submitted their fine work to this special issue, with an unprecedented and very pleasing number of submissions. The abundant high-quality work split in these two journal issues corroborates the liveliness of our community, and its recent effort to push for remarkable advances in this field.

Overall, we trust that this special issue is a valuable landmark for the area of energy-quality scalable circuits and systems, and will help initiate and expand many exciting research directions that will keep driving energy down as we have been used to for a long time.

Massimo Alioto
Guest Editor
Department of Electrical and Computer Engineering
National University of Singapore
Singapore

Vivek De
Guest Editor
Intel Corporation
Portland (OR)
USA

Andrea Marongiu
Guest Editor
Department of Information Technology and Electrical Engineering
ETH Zurich
8006 Zürich, Switzerland

References

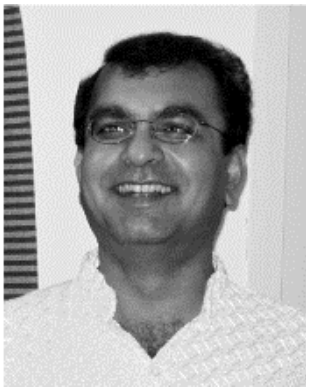
[1] Y. Guo and Y. Fang, "Electricity cost saving strategy in data centers by using energy storage," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1149–1160, Jun. 2013.
[2] I. Eyal, "The miner's dilemma," in *Proc. IEEE Symp. Secur. Privacy*, San Jose, CA, USA, May 2015, pp. 89–103.
[3] M. Alioto, Ed., *Enabling the Internet of Things: From Integrated Circuits to Integrated Systems*. Cham, Switzerland: Springer, 2017.
[4] G. Bell, "Bell's law for the birth and death of computer classes," *Commun. ACM*, vol. 51, no. 1, pp. 86–94, Jan. 2008.
[5] J. Koomey, S. Berard, M. Sanchez, and H. Wong, "Implications of historical trends in the electrical efficiency of computing," *IEEE Ann. Hist. Comput.*, vol. 33, no. 3, pp. 46–54, Mar. 2011.

[6] G. Frantz, "Digital signal processor trends," *IEEE Micro*, vol. 20, no. 6, pp. 52–59, Nov./Dec. 2000.
[7] ITRS. (2015). *International Technology Roadmap for Semiconductors: 2015 Edition*. [Online]. Available: http://www.itrs2.net/
[8] *Heterogeneous Integration Roadmap*. [Online]. Available: http://cpmt.ieee.org/technology/heterogeneous-integration-roadmap.html
[9] M. D. Hill. *Out Brief of DARPA/ISAT Workshop: Advancing Computer Systems Without Technology Progress*. [Online]. Available: http://sigarch.hosting.acm.org/2012/07/28/outbrief-of-darpaisat-workshop-advancing-computer-systems-without-technology-progress/
[10] R. L. Rivest, C. Stein, C. E. Leiserson, and T. H. Cormen, *Introduction to Algorithms*, 3rd ed. Cambridge, MA, USA: MIT Press, 2009.
[11] K. Asanovic *et al.*, "The landscape of parallel computing research: A view from Berkeley," Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2006-183, 2006. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html
[12] M. Alioto, "Energy-quality scalable adaptive VLSI circuits and systems beyond approximate computing," in *Proc. IEEE DATE*, Lausanne, Switzerland, Mar. 2017, pp. 127–132.
[13] F. Frustaci, M. Khayatzadeh, D. Blaauw, D. Sylvester, and M. Alioto, "SRAM for error-tolerant applications with dynamic energy-quality management in 28 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 50, no. 3, pp. 1310–1323, Mar. 2015.
[14] L. Freyman, D. Fick, M. Alioto, D. Blaauw, and D. Sylvester, "A 346 μm 2 VCO-based, reference-free, self-timed sensor interface for cubic-millimeter sensor nodes in 28 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 49, no. 11, pp. 2462–2473, Nov. 2014.
[15] S. Narayanan, J. Sartori, R. Kumar, and D. L. Jones, "Scalable stochastic processors," in *Proc. DATE*, Dresden, Germany, Mar. 2010, pp. 335–338.
[16] A. Alvarez, G. Ponnusamy, and M. Alioto, "EQSCALE: Energy-quality scalable feature extraction engine for sub-mW real-time video processing with 0.55 mm^2 area in 40 nm CMOS," in *Proc. ASSCC*, Seoul, South Korea, Nov. 2017, pp. 241–244.
[17] B. Moons and M. Velherst, "ENVISION: A 0.26-to-10TOPS/W subword-parallel dynamic-voltage-accuracy-frequency-scalable convolutional neural network processor in 28 nm FDSOI," in *ISSCC Dig Tech. Papers*, Feb. 2017, pp. 246–247.
[18] M. Alioto, V. De, and A. Marongiu, "Energy-quality scalable integrated circuits and systems: Continuing energy scaling in the twilight of Moore's law," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, to be published.

Massimo Alioto (M'01–SM'07–F'16) was with the University of Siena, the University of Michigan at Ann Arbor, the University of California at Berkeley, and EPFL. He was a Visiting Professor at Intel Labs–CRL. He is with the National University of Singapore, where he is currently the Director of the Integrated Circuits and Embedded Systems Area, and leads the Green IC Research Group.

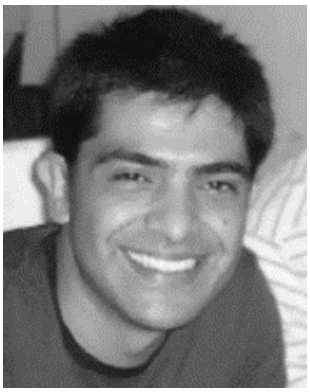
He has authored or co-authored over 250 publications on journals and conference proceedings. He has co-authored three books from Springer, including *Enabling the Internet of Things—From Integrated Circuits to Integrated Systems* (Springer, 2017). His primary research interests include ultra-low power VLSI circuits, self-powered and wireless nodes, near-threshold circuits for green computing, energy-quality scalable VLSI circuits, circuit techniques for emerging technologies, and hardware-level security, among the others.

Dr. Alioto was a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2009 to 2010, for which he is/was also member of the Board of Governors (2015–2020) and the Chair of the VLSI Systems and Applications Technical Committee from 2010 to 2012. He also served as a guest editor of numerous IEEE journal special issues and an associate editor for a number of other journals. He is/was the conference Technical Program Chair (ISCAS 2022, ICECS, NEWCAS, SOCC, PRIME, VARI, and ICM), the Track Chair (ICCD, ISCAS, ICECS, VLSI-SoC, APCCAS, and ICM), and TPC member (ISSCC and ASSCC). He is the Editor-in-Chief of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS for the 2019–2020 term and the Deputy Editor-in-Chief of the IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS (2018).



Vivek De (F'11) received the B.Tech. degree in electrical engineering from IIT Madras, Chennai, India, the M.S. degree in electrical engineering from Duke University, Durham, NC, USA, and the Ph.D. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA.

He is currently an Intel Fellow and the Director of Circuit Technology Research, Intel Labs. He is responsible for providing strategic technical directions for long-term research in the future circuit technologies and leading energy efficiency research across the hardware stack. He has 273 publications in refereed international conferences and journals with a citation H-index of 73 and 221 patents issued with over 29 patents filed (pending). He received the Intel Achievement Award for his contributions to an integrated voltage regulator technology. He received the Best Paper Award from the 1996 IEEE International ASIC Conference and nominations for the Best Paper Award at the 2007 IEEE/ACM Design Automation Conference (DAC) and the 2008 IEEE/ACM International Conference on Computer-Aided Design. He also co-authored a paper nominated for the Best Student Paper Award at the 2017 IEEE International Electron Devices Meeting. One of his publications was recognized in the 2013 IEEE/ACM DAC as one of the Top 10 Cited Papers in 50 Years of DAC. Another one of his publications received the Most Frequently Cited Paper Award in the IEEE Symposium on VLSI Circuits at its 30th Anniversary in 2017. He was recognized as a Prolific Contributor to the IEEE International Solid-State Circuits Conference at its 60th Anniversary in 2013 and a Top 10 Contributor to the IEEE Symposium on VLSI Circuits at its 30th Anniversary in 2017. He served as an IEEE/EDS Distinguished Lecturer in 2011 and an IEEE/SSCS Distinguished Lecturer from 2017 to 2018. He received the 2017 Distinguished Alumnus Award from IIT Madras.



Andrea Marongiu (M'12) received the M.Sc. degree in electronic engineering from the University of Cagliari, Italy, and the Ph.D. degree in electronic engineering from the University of Bologna, Italy.

He has been a Post-Doctoral Research Fellow with ETH Zurich, Switzerland. He currently holds an assistant professor position at the Department of Computer Science and Engineering, University of Bologna. His main research interests focus on programming models and architectures in the domain of heterogeneous multi- and many-core systems on a chip. This includes language, compiler, runtime and architecture support to efficiently address performance, predictability, energy and reliability issues in parallel, embedded systems, and HW-SW co-design of accelerator-based MPSoCs. In this field, he has authored over 100 papers in international peer-reviewed conferences and journals, books, and book chapters.

He has collaborated and collaborates with several international research institutes and companies and serves or has served as a TPC member for several international conferences and workshops in his field (DATE, SCOPES, MCSoC, EUC, FPL, and DASIP).