Real-Time Trainable Data Converters for General Purpose Applications

Loai Danial, Student Member, IEEE, and Shahar Kvatinsky, Senior Member, IEEE

The Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 3200003, ISRAEL

***Abstract*— Data converters are ubiquitous in data-abundant systems, where they are heterogeneously distributed across the analog-digital interface. Unfortunately, conventional data converters trade off speed, power, and accuracy. Furthermore, intrinsic real-time and post-silicon variations dramatically degrade their performance. In this paper, we employ novel neuro-inspired approaches to design smart data converters that could be trained in real-time for general purpose applications, using machine learning algorithms and artificial neural network architectures. Our approach integrates emerging memristor technology with CMOS. This concept will pave the way towards adaptive interfaces with the continuous varying conditions of data driven applications.**

Keywords—Analog-to-digital conversion, adaptive systems, digital-to-analog conversion, memristors, machine learning, neuromorphic computing, reconfigurable architectures.

I. Introduction

With the advent of high-speed, high-precision, and low-power mixed-signal systems, the demand for accurate, fast, and energy-efficient data converters is on the rise. These systems, which operate on a variety of real-world signals, are widely employed in medical imaging, biosensors, consumer electronics, instrumentation, and telecommunication.

Unfortunately, the intrinsic speed-power-accuracy tradeoff in analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) is pushing them out of the application band of interest [1]. Furthermore, with the continuous downscaling of technology motivated by Moore's law, this tradeoff has become a chronic bottleneck of modern systems design due to alarming deep sub-micron effects [1]. Those effects are poorly handled with particular technology-dependent design techniques that overload data converters with enormous overhead, exacerbating the tradeoff and severely degrading their performance [1]. Conventional data converters lack design standards and are customized with sophisticated design flow. Data converter architectures are optimized for special purpose applications, from high-speed, to high-resolution, to low-power applications [2]. These methods not only require exhaustive characterization and massive validation, but they are also expensive to develop, with a long time-to-market.

This paper takes a different, systematic approach, beyond Moore's law, to design general purpose data converters [3][4]. We propose that the converted data be used to train the converter in order to autonomously adapt to the exact specifications of the running application and adjust to environmental variations. This approach will reduce the time to market, efficiently scale with newer technologies, drastically reduce its cost, significantly standardize the design flow, and enable a generic architecture for general purpose applications.

The proposed trainable data converters utilize machine learning (ML) algorithms to train an artificial neural network (ANN) architecture based on the promising technology of memristors [5]. With their synapse-like behavior, memristors are becoming more and more prevalent in the design and realization of artificial neural systems [6]. Their small footprint, analog storage properties, low energy consumption, and non-volatility characteristics allow them to mimic neural synapses, where the conductance of the memristor is considered as the synapse weight [7].

II. Trainable Data Converters

A. Figure of Merit (FOM) Dynamic Optimization

The field of machine learning (ML) is dedicated to the study and implementation of systems that can learn from data to make decisions, predictions, and classifications based on past examples. Data conversion can be viewed as a special case of the classification optimization and signal restoration problem that could easily be solved using ML.

When comparing data converters with different specifications, a numerical quantity known as a figure of merit (*FOM*) is used to fairly characterize the performance of each converter relative to its alternatives. The FOM is defined as

$$FOM = \frac{P}{2^{ENOB} \cdot f_s} \left[\frac{J}{conv}\right], \quad (1)$$

and relates the converter power dissipation during conversion, *P*, to its performance in terms of sampling frequency, f_s, and effective number of resolution bits *(ENOB)*. The FOM best captures the fundamental speed-power-accuracy tradeoff [1]. It accurately reflects the merits of the converter in a certain context and for a specified purpose.

We propose trainable data converter architectures for general purpose applications [3][4], as shown in Fig. 1. A set of parameters is determined to meet the requirements of the running application. First, f_s is determined, followed by the number of resolution bits *N*, followed by the full-scale voltage V_{FS}, which specifies the converter input dynamic range. Then, the converter is trained by a supervised ML algorithm, called online stochastic gradient descent. The training is done in real-time to optimize the ENOB and power dissipation, where the correct digital labels corresponding to the analog input are supplied and compared to the actual digital output. This procedure is equivalent to a dynamic FOM optimization. The technique is not exclusive to application reconfiguration but can also be used for device mismatch self-calibration, adaptation, power optimization, and noise tolerance with generic, standard methodology [3][4].

B. Neural Network Data Converter Architectures

ANNs are distributed networks that collectively make decisions based on a successive adjustment of weights. This mechanism precisely describes ADCs in time-scale with successive binary-weighted approximation (SAR). While bit comparison is equivalent to neural activation, each reference scale during the

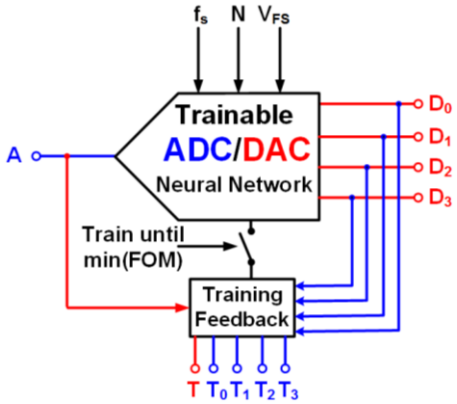

Fig. 1. Scheme of the trainable 4-bit ADC/DAC (blue/red path) neural network. The network receives f_s, V_{FS}, N, and is provided a specific teaching dataset T_i for real-time training. The training continues until the converter achieves the optimal FOM.

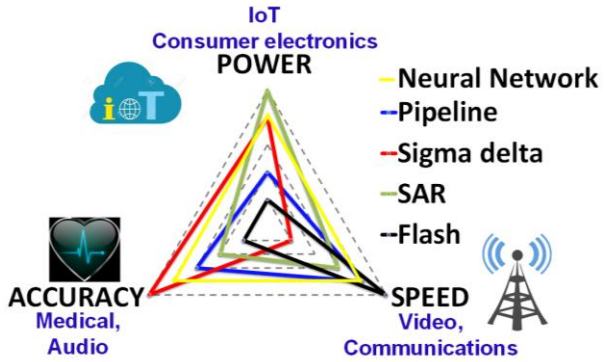

Fig. 2. Spider diagram of conventional ADC architectures (colored lines), design tradeoff, and associated applications (blue text), along with our trainable, general purpose, neural network architecture (yellow), which breaks through the tradeoff as shown by the balanced diagram.

successive binary search algorithm is equivalent to a binary-weighted synapse. The temporal binary search algorithm of a SAR is transformed to a spatial neural network with binary-weighted synapses and pipelined forward-propagated neurons. Analogously, a single layer neural network with binary-weighted synapses could be viewed as a special case of the binary-weighted DAC architecture. The proposed networks [3] are based on energy optimization, where their energy level is minimal when they are used as data converters.

In a real-time operation where non-ideal, non-linear, stochastic, and varying conditions affect the conversion accuracy, the correct weights are not distributed deterministically in binary-weighted style. In this case, the weights should be updated in real-time. We exploit the intelligent properties of the neural networks for online training by ML algorithms. The training algorithm ensures the learning capability of our networks in terms of accuracy and speed [3][4]. The synapse circuit design is composed of a single memristor, connected to a shared terminal of two MOSFET transistors (p-type and n-type) [7]. The circuit utilizes the intrinsic dynamics of the memristive crossbar, which inherently implements Ohm's and Kirchhoff's laws for ANN hardware realization [7]. The neurons are implemented by opAmps and comparators [6]. While the training feedback of the ADC is simple and realized by digital hardware, the feedback of the DAC is sophisticated and implemented by pulse-width modulator [3][4].

III. General Purpose Data Converters

Having demonstrated the dynamic mechanism of the trainable data converters [3][4], we now discuss their real-time training capability for general purpose applications. For every selected f_s, the converter is trained correspondingly by a training data-set with the same specifications, thus achieving optimal ENOB. Analogously, the power consumption is dynamically optimized for every f_s to achieve minimal power dissipation of the network.

Interestingly, the collective optimization of the proposed architecture breaks through the speed-power-accuracy tradeoff, and dynamically scales the FOM to achieve 8.25 fJ/conv.step for a 4-bit ADC in 180nm CMOS technology, two orders of magnitude less than the average FOM of state-of-the-art ADCs [2][4]. The architecture's versatility is realized by a simple and minimalistic design with a reconfigurable cost-effective single-channel. The proposed architecture utilizes the resistive parallel computing capabilities of memristors to achieve high speed. It moreover utilizes their analog non-volatility to achieve high accuracy by enabling the ML algorithm to adjust its conductance precisely and *in situ*. All these features will enable a general-purpose application architecture that is trained to fit different specifications from high-speed, to high-accuracy, to low-power, as illustrated in Fig. 2.

IV. Conclusions And Future Work

This concept paper proposes a real-time trainable ADC architecture for general purpose applications, which breaks through the speed-power-accuracy tradeoff. Motivated by the analogies between mixed-signal circuits and the neuromorphic paradigm, we exploit the intelligent properties of an ANN, and propose neural network architectures for data converters that are trained online by a supervised ML algorithm. A hybrid CMOS–memristor circuit design is proposed to realize the neural network. We believe that the proposed data converters constitute a milestone with valuable results for emerging applications with varying conditions, such as wearable devices and automotive applications. Large-scale challenges still need to be investigated by leveraging mixed-signal architectures (pipelined, time-interleaved, and oversampling) and deep neural network concepts.

V. Acknowledgements

This research was partially supported by the Israeli Planning and Budgeting Committee Fellowship, by the Viterbi Fellowship at the Technion Computer Engineering Center, by the Israel Innovation Authority KAMIN grant no. 57681, and by EU COST Action IC1401.

References

[1] P. Kinget *et al.*, "Impact of Transistor Mismatch on the Speed-Accuracy-Power Trade-Off of Analog CMOS Circuits," *CICC*, pp. 333–336, May 1996.

[2] B. Murmann, "ADC Performance Survey 1997-2017," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html.

[3] L. Danial *et al.*, "DIDACTIC: A Data-Intelligent Digital-to-Analog Converter with a Trainable Integrated Circuit using Memristors," *JETCAS*, Vol. 8, No. 1, pp. 146-158, March 2018.

[4] L. Danial *et al.*, "Breaking Through the Speed-Power-Accuracy in ADCs using a Memristive Neuromorphic Architecture," *TETCI* (in press).

[5] L. Gao *et al.*, "Digital-to-Analog and Analog-to-Digital Conversion with Metal Oxide Memristors for Ultra-Low Power Computing," *NANOARCH*, pp. 19–22, July 2013.

[6] M. Prezioso *et al.*, "Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors," *Nature*, Vol. 521, No. 7550, pp. 61–64, May 2015.

[7] D. Soudry *et al.*, "Memristor-Based Multilayer Neural Networks with Online Gradient Descent Training," *TNNLS*, Vol. 26, No. 10, pp. 2408–2421, October 2015.