An Energy-Efficient Multi-Sensor Compressed Sensing System Employing Time-Mode Signal Processing Techniques

Akgun, Omer Can; Mangia, Mauro; Pareschi, Fabio; Rovatti, Riccardo; Setti, Gianluca; Serdijn, Wouter A.

DOI
10.1109/ISCAS.2019.8702667

Publication date
2019

Document Version
Accepted author manuscript

Published in
2019 IEEE International Symposium on Circuits and Systems (ISCAS)

An Energy-Efficient Multi-Sensor Compressed Sensing System Employing Time-Mode Signal Processing Techniques

Omer Can Akgun*, Mauro Mangia†, Fabio Pareschi‡, Riccardo Rovatti†, Gianluca Setti‡ and Wouter A. Serdijn*
*Section Bioelectronics, Delft University of Technology, the Netherlands - Email: {o.c.akgun,w.a.serdijn}@tudelft.nl
†ARCES, University of Bologna, Italy - Email: {mauro.mangia2,riccardo.rovatti}@unibo.it
‡DET, Politecnico di Torino, Italy - Email: {fabio.pareschi,gianluca.setti}@polito.it

***Abstract*—This paper presents the design of an ultra-low energy, rakeness-based compressed sensing (CS) system that utilizes time-mode (TM) signal processing (TMSP). To realize TM CS operation, the presented implementation makes use of monostable multivibrator based analog-to-time converters, fixed-width pulse generators, basic digital gates and an asynchronous time-to-digital converter. The TM CS system was designed in a standard 0.18 μm IC process and operates from a supply voltage of 0.6V. The system is designed to accommodate data from 128 individual sensors and outputs 9-bit digital words with an average reconstruction SNR of 35.31 dB, a compression ratio of 3.2, with an energy dissipation per channel per measurement vector of 0.621 pJ at a rate of 2.23 k measurement vectors per second.**

***Index Terms*—compressed sensing, time-mode, time-mode signal processing, rakeness, energy efficiency, ultra-low energy**

I. Introduction

Today, most of the circuit design is done in CMOS. The advancement and scaling of CMOS technologies has, in fact, always been based on improving digital systems' performance. Yet, with each new technology node, the supply voltage of the process node is scaled down as well, which reduces the headroom that is available to the transistors for operating in saturation region. Without transistors operating in the saturation region, it is very hard to realize signal processing and amplification functions in the analog domain. One possible solution to this problem is using time-mode signal processing (TMSP) techniques [1]–[4].

Time-mode (TM) circuits represent an analog signal by the time difference between two binary switching events. Consequently, time-mode operation is inherently lower power when compared to standard CMOS digital operation. For example, to transfer N-bit accurate data in a standard CMOS digital circuit, the number of switchings required on the data line may change from 0 to N if the data is transmitted in parallel, whereas, in a TM circuit, transfer of the data always takes two switchings if the rising and falling edges of a pulse are used for information transmission. Based on these very simple observations, it is arguable that more low-power signal processing systems may be implemented using TMSP techniques in the future.

With today's increasing focus on both Internet of Things and biosensors/bioelectronics applications, the design of very low-energy sensing nodes is the key to enlarge these paradigms to a wider range of applications. Furthermore, a sensor is no longer a block devoted only to signal acquisition; it must also be able to acquire, digitize, compress and finally transmit sensed data with an almost negligible energy cost.

In this field, the adoption of Analog to Information Conversion (AIC) based on the Compressed Sensing paradigm (CS) [5], [6] emerges as a low-energy solution that combines data acquisition and data compression. The CS mechanism is based on the projection, in the analog domain, of input signal instances on a set of predefined *sensing sequences* that are in number less than the input signal intrinsic dimensionality. The obtained output, the *measurement* vector, is then transmitted to a decoder block to recover the original signal. Alternatively, CS can be used as a low power compression scheme after signal digitization [7], [8].

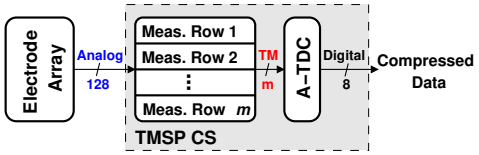


Figure 1. Proposed TMSP CS system high-level block diagram.

The research work presented here focuses on multi-channel electrode arrays that characterize a huge set of biomedical applications [7], [9], [10], where the input signal is the collection of readings. The proposed system is shown in Figure 1 and realizes the CS paradigm by using a very energy-efficient TMSP system. The analog input from 128 electrodes is sensed and compressed in time-mode and the resulting time-mode signals are converted to digital using a single asynchronous time-to-digital converter (A-TDC).

The paper is organized as follows. Sec. II recaps the CS mathematical background, in Sec. III the TMSP Compressed Sensing System is presented, while Sec. IV reports simulation results. Finally, we draw the conclusions.

II. Compressed Sensing

The CS working principle transforms the real information content of an n dimensional input signal x into a measurement vector y that is the output of m linear projections of x over the rows of a matrix A, called the *sensing matrix* [5]. This simple acquisition scheme reflects the implementation of an ideal encoder. Furthermore, in a real scenario, noise and other system non-idealities affect the final encoder output, which can be expressed as

$$y = A(x + \nu_x) + \nu_y \ , \tag{1}$$

where both ν_x and ν_y are additive disturbances modelling the mentioned non-idealities for the input signal and for the

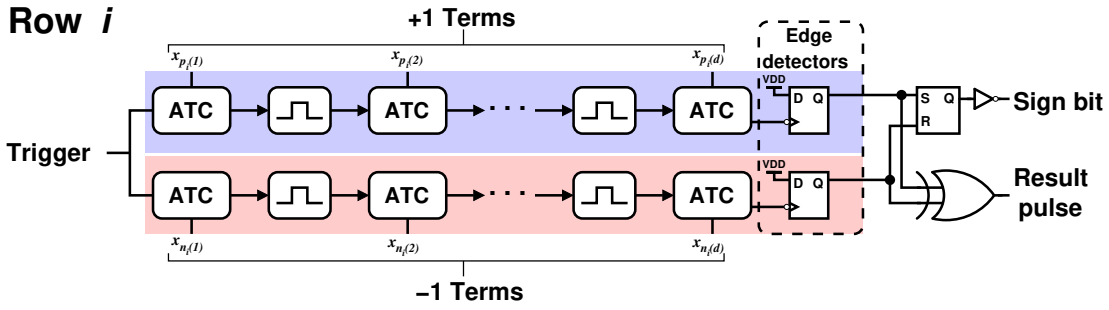


Figure 2. A compressed sensing measurement row implementation using time-mode signal processing.

measurement vector, e.g., quantization error or signal noise. In case of $m < n$, CS provides a compression ratio defined as $CR = n/m$.

On the receiver side, a decoder block must be able to recover the initial information x from the transmitted digital words composing y and the knowledge of the sensing matrix A. Since A is a rectangular matrix, retrieving x from y is an ill-posed problem, i.e., multiple solutions exist. To guarantee a correct signal reconstruction, the CS decoder requires a proper assumption on the acquired class of signal, i.e., all input instances x must be κ-sparse. This means that an $n \times n$ orthonormal matrix D exists, which defines the *sparsity basis*, so that any vector $x = D\xi$ is such that ξ, the n-dimensional vector containing projections of x over the columns of D, has no more than $\kappa \ll n$ non-zero components.

With this assumption, the decoder is designed to give the reconstructed input signal $\hat{x} = D\hat{\xi}$, where $\hat{\xi}$ is the sparsest vector ξ that matches (1) with a proper tolerance. This is mapped in the solution of the following optimization problem [6]

$$\hat{\xi} = \arg\min_{\xi} \|\xi\|_1 \quad \text{s.t. } \|AD\xi - y\|_2 < \varepsilon \ , \tag{2}$$

where $\|\cdot\|_p$ stand for the ℓ_p norm and ε balances the effects of both ν_x and ν_y.

Finally, proper decoding is ensured assuming A satisfies some constraints and a minimum number of measurements are available. In standard CS theory, this is ensured by generating entries of A as instances of independent and identically distributed random variables [6]. Along all possible CS encoders already proposed in the literature, circuit implementations that adopt either antipodal or ternary random sensing matrices are more advantageous [8], [11]–[13]. This means that the sensing matrix entries are still random but are limited to either $A_{j,k} \in \{-1, 1\}$ or $A_{j,k} \in \{-1, 0, 1\}$ where, in the latter case, an increase in the number of zeros implies a reduction in the number of operations needed to compute y.

If we assume that for the i-th row of A only d entries are non-zeros, where the number of 1 and -1 are d_i^+ and d_i^-, then the i-th measurement becomes

$$y_i = \sum_{j=1}^{d_i^+} x_{p_i(j)} - \sum_{j=1}^{d_i^-} x_{n_i(j)} \ , \tag{3}$$

where $p_i(\cdot)$ and $n_i(\cdot)$ map the positions of positive and negative entries of the i-th row of A, respectively.

The CS approach was expanded in [8], [14] where the authors proposed a soft adaptation of the second-order statistics of the sensing matrix rows to the second-order statistics of the acquired class of signals, and this method is called rakeness-based CS. Rakeness-based CS imposes the following $n \times n$ correlation matrix C_A to the rows of A,

$$C_A = \frac{n}{2}\left(\frac{C_x}{\mathrm{tr}(C_x)} + I_n\right) \ , \tag{4}$$

where $\mathrm{tr}(\cdot)$ stands for matrix trace, I_n is the $n-$dimensional identity matrix and C_x is an estimation of the input signal correlation matrix.

III. Time-Mode Signal Processing Compressed Sensing System

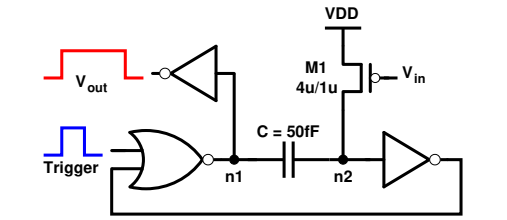
We applied time-mode operation and TMSP methods to the design of a rakeness based CS system in a standard 0.18 μm IC process. Each measurement row of the implemented CS system is defined by (3) and is mapped to a TMSP implementation, as shown in Figure 2. As in (3), positive and negative computations are separated into two time-mode processing chains, one for each summation. Each measurement row is triggered by a trigger signal and as each analog-to-time converter (ATC) converts a voltage input value into a pulse whose width is proportional to the input signal value, the signal propagates through the chain of ATCs and fixed-width pulse generators (FWPGs). FWPGs, represented by the pulse blocks in Figure 2, are required to be able to trigger the next ATC in the chain with the falling edge of the previous ATC pulse. In this specific implementation, we created two parallel chains of d ATCs and connected $d - d_i^+$ and $d - d_i^-$ ATC inputs to AC-ground, for the upper and the lower chains, respectively. The ATCs with 0 inputs represent the 0 values in the sensing matrix and these AC-ground connections are required to equalize the time offset coming from the ATC. The structures and operation principles of both the ATC and the negative-edge triggered fixed-width pulse generator are explained in the following sub-section.

At the end of each chain, negative-edge triggered flip-flops are employed to capture the final falling edge of the signal generated by the chain of ATCs. A *result pulse*, whose width value is proportional to the calculation given in (3) is generated by *XOR*-ing the signals captured by the edge detectors. This *XOR* operation effectively creates a pulse signal whose width is the absolute value of the time difference of the signals at the edge detector outputs. As the sign information of the subtraction is lost in the *XOR* operation, a sign bit is generated by using a nand-based *SR-latch* and an inverter (Figure 2).

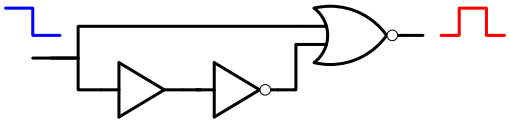
To convert the *result pulse* into a digital value for transmission and/or storing, an 8-bit A-TDC similar to the one in [15] was designed and employed. The result pulse is fed into the A-TDC for conversion, and at the same time the negative edge of the result pulse is used for triggering the next row for calculations. With such an implementation, we were able to use a single TDC

for processing the result pulse of all the measurements for an input signal bandwidth limited to 200 Hz.

A. Sub-blocks



(a) Monostable Multivibrator



(b) Negative edge triggered fixed-width pulse generator

Figure 3.

As the ATC in the system, we employ a monostable multivibrator (MSMV) as shown in Figure 3(a) [16]. In this implementation, a pMOS transistor (M1) acts as a variable resistor whose resistance is modulated by the current input signal. When the MSMV is triggered by an input pulse, nodes **n1** and **n2** are pulled to logic-low and M1 starts charging node **n2**. The gate of M1 is driven by the input signal that is to be converted into time, and sampling is realized by modulating the instantaneous resistance of M1. Thus, the RC time constant of the multivibrator is modulated as well, resulting in a pulse whose width is proportional to the amplitude of the input signal. The pulse width generated by the ATC is given in [16] by

$$T = C(R + R_{on}) \ln \left[\frac{R}{R + R_{on}} \frac{V_{DD}}{V_{DD} - V_{th}} \right], \qquad (5)$$

where R is the average resistance of the pMOS transistor during pulse generation, R_{on} the resistance of the NOR gate, and V_{th} the switching threshold of the inverter. Assuming $R_{on} << R$ and $V_{th} = V_{DD}/2$, (5) is simplified to $T = 0.69RC$. Furthermore, this ATC implementation has an inherent *timeout* feature and will always generate a pulse event at node **n1** regardless of the input signal value at V_{in}, avoiding stalling of the chain. Capacitor C and transistor M1 were made bigger than the minimum values required for correct operation to mitigate process variation effects.

A negative edge triggered FWPG, shown in Figure 3(b), is used between the ATC blocks as we require the triggering of the next ATC in the chain to occur during the falling edge of the pulse generated by the previous ATC. By triggering the next ATC with the falling edge of the previous ATC output, time addition operation is realized.

The result of a transistor-level transient simulation of a measurement row with $d = 16$ is shown in Figure 4. On the top and bottom halves of the figure, time-mode calculations for the *+1 Terms* and *-1 Terms* are shown, respectively. Only the last 2 ATC outputs in their respective chains, i.e., $d = 15$ and $d = 16$, are shown. The outputs of the ATCs are shown in red and the outputs of the FWPGs are shown in blue. The ATCs generate signals with varying pulse widths based on their input signal values, and, therefore the signal in the chain is delayed by an amount proportional to the voltage inputs applied to the

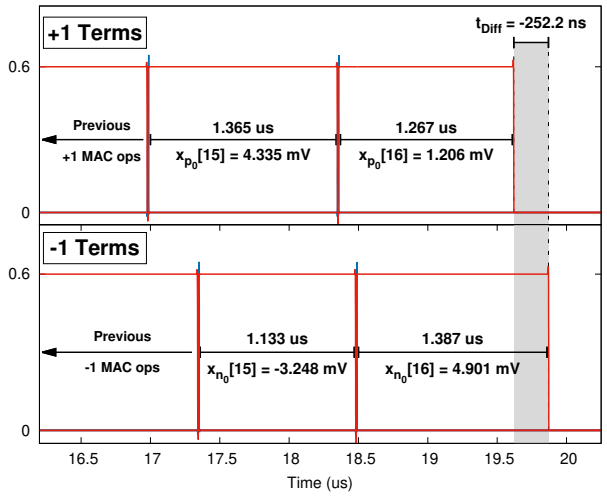


Figure 4. Transient simulation of one measurement row of the designed TMSP CS system with $d = 16$.

ATCs. The falling edges of the generated pulse signals by the last ATCs in the chain ($d_i^+ = di_i^- = d = 16$) denote the end of time-mode summation operations, and their difference gives the time-mode result of (3). For example, in the figure, the last ATC in the *+1 Terms* chain generates the final falling edge earlier, representing a summation value smaller than the one generated by the *-1 Terms* chain. Therefore, the resulting difference, marked by t_{Diff}, is negative. The sign information of the calculation is captured using the *SR-latch* and later saved together with the output of the TDC, resulting in a 9-bit digital value.

B. System Simulations

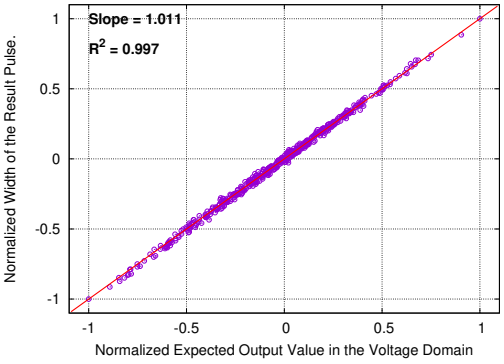


Figure 5. Comparison of the normalized generated pulse width with respect to the normalized expected value from 512 measurements for a CS system with d = 16.

To verify the correct time-mode operation of the designed system, extensive transistor level SPICE simulations were run using the HSPICE simulator. First, we verified the operation of the ATC. For input signals that vary between -5 mV and 5 mV around half VDD (0.3V), the ATC realizes the conversion function $t_{pulse} = 30.8 \cdot \frac{\mu s}{V} \cdot V_{in} + 1.235\,\mu s$, generating pulses in the range $1.081\,\mu s$ - $1.389\,\mu s$. We also simulated both a single ATC and a chain of ATCs and FWPGs for their time-mode noise (jitter) performance. The simulations show that a single ATC together with a FWPG has a random jitter of 6.49 ns for a conversion range of 309 ns, resulting in an SNR of 33.55 dB.

As the number of ATCs and FWPGs increase in the chain, for every doubling of the number of elements the increase in SNR is 3 dB. Therefore, a chain with at least 2 ATCs satisfy our input SNR requirement of 35 dB. In our verification and high-level simulations, jitter values for different d values were implemented in our models through the ν_x term in (1).

Next, we simulated and verified the implementation of a CS measurement row with $d = 16$ by varying the input signals and comparing the normalized output of the design in Figure 2 to the normalized expected value in the voltage domain. The results of simulating 512 measurements is shown in Figure 5. As can be seen from the figure, the generated time-mode signals strongly correlate with the expected value in the time domain with an R^2 value of 0.997.

Table I
PER SAMPLE ENERGY DISSIPATION OF THE MAIN SUB-BLOCKS IN THE CS SYSTEM.

Block	Energy (pJ)
ATC	0.0213
FWPG	0.0117
TDC	1482

After verifying the correct operation of our design, we characterized each element in the implementation for their energy dissipation per computation at a VDD of 0.6V. Energy dissipation results of the major circuit blocks are presented in Table I. The TDC energy dissipation value includes the energy dissipation of the edge detectors, SR latch and the XOR gate. These values were used to find an energy-optimal implementation, which is presented in the next section.

IV. System Simulation Results

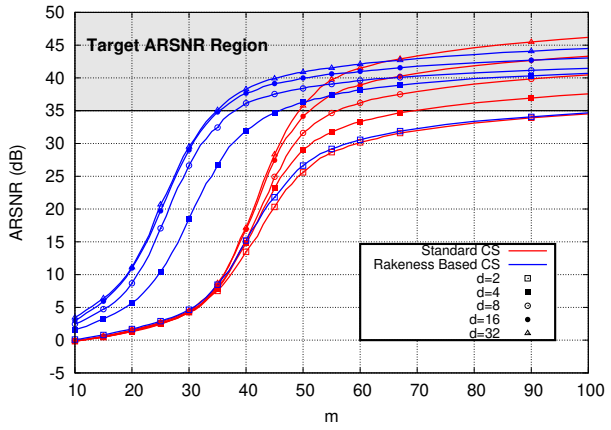


Figure 6. ARSNR simulations for varying values of m and d for time-mode, standard and rakeness-based CS.

To verify the effectiveness of the proposed architecture, we investigated an energy optimum time-mode CS implementation using the results from the previous section. First, to generate the A matrices, we performed Monte Carlo simulations on a set of synthetic low-pass signals. We used a chunk of input signals composed by $n = 128$ successive samples, such that each input vector x is sparse with respect to the discrete cosine transform and where $\kappa = 12$. The A matrices are such that only d entries in each row are non-zero. Furthermore, for the Standard CS, purely random ternary sensing matrices are employed while for the rakeness-based CS a proper correlation profile is imposed to each row of A according to (4) and by adopting generation methods described in [8], [17]. The input signal correlation matrix needed in (4) was estimated on a separated dataset.

Table II
ARSNR, ENERGY DISSIPATION, COMPRESSION RATIO AND CS CR EFFICIENCY IN TERMS OF ENERGY PER MEASUREMENT VECTOR FOR A 128 CHANNEL TMSP CS IMPLEMENTATION.

	m	d	ARSNR (dB)	E (pJ)	CR	CRE
Rakeness	40	8	35.31	79.46	3.20	0.0403
	36	16	35.41	90.52	3.56	0.0393
	41	8	35.68	81.45	3.12	0.0383
	37	16	35.95	93.03	3.46	0.0372
Standard	57	8	35.22	113.23	2.25	0.0198
	58	8	35.49	115.22	2.21	0.0192
	52	16	35.78	130.75	2.46	0.0188
	59	8	35.80	117.20	2.17	0.0185

To assess the average performance of the implementation, 500 different sets of sensing matrices, for each CS approach, were applied to 500 different input instances in order to obtain signal reconstructions by the solution of (2). After that it was possible to compute, as a figure of merit, the Average Reconstruction Signal to Noise Ratio (ARSNR)

$$\mathrm{ARSNR} = \mathrm{E}\left[\frac{\|\mathrm{x}\|_2}{\|\mathrm{x} - \hat{\mathrm{x}}\|_2}\right] \tag{6}$$

where $\mathrm{E}[\cdot]$ stands for the expectation. The results of our simulations with a targeted ARSNR value of 35 dB are given in Figure 6. We considered both standard CS and rakeness-based CS implementations. As expected, rakeness-based CS performs better than standard CS and requires less resources, i.e., smaller d and m, for satisfying the ARSNR requirement.

As all the d and m pairs that satisfy the ARSNR requirement are feasible implementation candidates, to be able to compare them in terms of energy efficiency, we define a metric *Compression Ratio per Energy (CRE)* as $\mathrm{CRE} = \frac{\mathrm{CR}}{\mathrm{E}}$, where E is the energy per measurement vector in pJ. A higher CRE means higher energy efficiency for the achieved compression ratio. The results of our simulations are given in Table II. The rows are sorted for CRE in descending order, separately for both standard and rakeness-based CS implementations. When the rakeness-based approach is compared to a standard implementation, compression ratio increases by 42% and the energy dissipation reduces by 29.8% for the most energy efficient implementations. Our implementation for the best CRE dissipates 0,621 pJ per channel per measurement vector and outputs 2.23 k measurement vectors per second.

V. Conclusions

This paper presents the design and the simulation results a of time-mode, rakeness-based compressed sensing system. Time-mode signal processing techniques have been applied for accumulating and subtracting voltage signal values in the time-domain using energy-efficient simple circuitry. The system was designed and simulated in a standard 0.18 μm process and operates from a supply voltage of 0.6 V. For an optimal 128-channel implementation, the energy dissipation per channel is 0.621 pJ per measurement vector for a compression ratio of 3.2.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 752819 and from ACRI Young Investigator Training Program.

REFERENCES

[1] F. Yuan, "CMOS time-to-digital converters for mixed-mode signal processing," *The Journal of Engineering*, January 2014. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/joe.2014.0044

[2] D. Miyashita, R. Yamaki, K. Hashiyoshi, H. Kobayashi, S. Kousai, Y. Oowaki, and Y. Unekawa, "An LDPC decoder with time-domain analog and digital mixed-signal processing," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 73–83, Jan 2014.

[3] Z. Chen and J. Gu, "Analysis and design of energy efficient time domain signal processing," in *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. ACM, 2016, pp. 100–105.

[4] V. Ravinuthula, V. Garg, J. G. Harris, and J. A. B. Fortes, "Time-mode circuits for analog computation," *International Journal of Circuit Theory and Applications*, vol. 37, no. 5, pp. 631–659, 2009. [Online]. Available: http://dx.doi.org/10.1002/cta.488

[5] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[7] J. Zhang, Y. Suo, S. Mitra, S. Chin, S. Hsiao, R. F. Yazicioglu, T. D. Tran, and R. Etienne-Cummings, "An efficient and compact compressed sensing microsystem for implantable neural recordings," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 4, pp. 485–496, Aug 2014.

[8] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, "Rakeness-based design of low-complexity compressed sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, 2017.

[9] M. Aghagolzadeh and K. Oweiss, "Compressed and distributed sensing of neuronal activity for real time spike train decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 2, pp. 116–127, April 2009.

[10] K. D. Wise, D. J. Anderson, J. F. Hetke, D. R. Kipke, and K. Najafi, "Wireless implantable microsystems: high-density electronic interfaces to the nervous system," *Proceedings of the IEEE*, vol. 92, no. 1, pp. 76–97, Jan 2004.

[11] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 1, pp. 149–162, Feb. 2016.

[12] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, Sept 2011.

[13] M. Shoaran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 6, pp. 857–870, Dec 2014.

[14] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1001–1014, May 2012.

[15] O. C. Akgun, "An asynchronous pipelined time-to-digital converter using time-domain subtraction," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–5.

[16] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 4th ed. Oxford University Press, 1998.

[17] G. Jacovitti, A. Neri, and G. Scarano, "Texture synthesis-by-analysis with hard-limited Gaussian processes," *IEEE Transactions on Image Processing*, vol. 7, no. 11, pp. 1615–1621, Nov. 1998.