This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Adaptive Learning-Based Compressive Sampling for Low-power Wireless Implants

Cosimo Aprile, *Student Member, IEEE*, Kerim Ture, *Student Member, IEEE*, Luca Baldassarre, Mahsa Shoaran, *Member, IEEE*, Gürkan Yilmaz, Franco Maloberti, *Life Fellow, IEEE*, Catherine Dehollain, *Member, IEEE*, Yusuf Leblebici, *Fellow, IEEE*, and Volkan Cevher, *Senior Member, IEEE*

***Abstract*— Implantable systems are nowadays being used to interface the human brain with external devices, in order to understand and potentially treat neurological disorders. The most predominant design constraints are the system's area and power. In this paper, we implement and combine advanced compressive sampling algorithms to reduce the power requirements of wireless telemetry. Moreover, we apply variable compression, to dynamically modify the device performance, based on the actual signal need. This paper presents an area-efficient adaptive system for wireless implantable devices, which dynamically reduces the power requirements yielding compression rates from 8× to 64×, with a high reconstruction performance, as qualitatively demonstrated on a human data set. Two different versions of the encoder have been designed and tested, one with and the second without the adaptive compression, requiring an area of 230×235 μm and 200 × 190 μm, respectively, while consuming only 0.47 μW at 0.8 V. The system is powered by a 4-coil inductive link with measured power transmission efficiency of 36%, while the distance between the external and internal coils is 10 mm. Wireless data communication is established by an OOK modulated narrowband and an IR-UWB transmitter, while consuming 124.2 pJ/bit and 45.2 pJ/pulse, respectively.**

***Index Terms*— Implantable integrated circuit, area-efficient, low-power, compressive sensing, neural signals, learning-based digital signal processing, signal recovery.**

I. Introduction

IN MOBILE applications, the power budget is defined by the battery limits, which, unfortunately, does not improve from one node to the following one, as the amount of logic gates does in the IC, as defined by the well-known Moor's law [1]. Table I gives an overview of battery power budget in some of the current electronic devices used for general daily life applications.

Among all the autonomous sensing applications, one of the most critical and challenging field is medical monitoring, in which various biological signals have to be processed with a relatively high accuracy, in order to extract reliable medical information for disease diagnosis or therapy.

For many decades, scientists have tried to understand the brain activity. Since the 1990s, clinicians have been able to implant devices capable of monitoring the neuronal activity [2]. *Micro/Nano fabrication of electromechanical systems* (M(N)EMS) industry is currently improving the capability to interface with the human brain. A multitude of applications are related to these systems, from research experiments to personal health monitoring and in-house treatments. In particular, electrodes and micro fabricated electrodes have enabled efficient electrical or optical links, enhancing the functionality of the neuronal interfaces. Since 1997, the usage of prostheses has been approved as an alternative treatment for some brain diseases, such as Parkinson and Epilepsy, and more recently, for depression [3]. Over 5% of the population worldwide experience at least one epileptic seizure during lifetime and around 50 million people are diagnosed with epilepsy [4], [5]. Moreover, in 30% of the cases, patients suffer from pharmaco-resistant epilepsy, where medications are not sufficient to treat seizures. Currently, the only available solution (when applicable) requires a long term hospitalization in order to record and localize the source of epileptic seizures, using a bulky system connected with cables trough the skull, and placed over the cortex. After localization of the epileptic foci an invasive surgery procedure is required, with the aim of physically removing the brain tissue where the stroke seizures originate. This would suggest the design of autonomous monitoring devices with minimal invasiveness. According to the vision of *Body Area Network* (BAN), such bio-electrical devices attached to the human body can either serve to carry out information to a medical host or to provide some feedback as first aid treatment.

The goal of this work is to optimize the information extraction from neural signals, merging new mathematical theory and computational methods in hardware, to reduce the power and area, outperforming the current state-of-the-art compression techniques. Our contention is to reduce the

Manuscript received April 16, 2018; revised June 26, 2018; accepted June 29, 2018. This work was supported by the European Research Council through the European Union's Horizon 2020 Research and Innovation Programme under Grant 725594 - time-data. This paper was recommended by Associate Editor X. Zhang. *(Corresponding author: Cosimo Aprile.)*

C. Aprile, K. Ture, G. Yilmaz, C. Dehollain, Y. Leblebici, and V. Cevher are with the Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland (e-mail: cosimo.aprile@epfl.ch).

L. Baldassarre is with Data Science, Gamaya, EPFL Innovation Park, 1015 Lausanne, Switzerland.

M. Shoaran is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA.

F. Maloberti is with the Department of Electrical, Computer, and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy.

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/TCSI.2018.2853983

1549-8328 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
POWER BUDGET FOR DIFFERENT APPLICATIONS

Application	Sensors	Wireless Interfaces	Power Consumption	Battery Lifetime
Pacemaker	Pacing leads	Inductive link	10 μW	Several Years
Human body monitoring	ECG, heart rate, Temperature	900 MHz ISM	1-8 mW	Several Hours
Smartphone	Multiple sensors	Bluetooth, WiFi, GSM, HSDPA, LTE	1 W	Few Hours

amount of required data, still allowing high signal reconstruction quality, leveraging both theory and practice through learning. As a result, power and time required for edge-data computation can be drastically reduced. This paper extends our previous work [6], by analyzing and implementing the fully integrated system, from the analog to digital conversion of neural signal to the wireless transmission of compressed data. Furthermore, we provide two different implementations of the data compression encoding system.

The paper is organized as follows. In Section II the system level choices are described. Compressive Sensing and Learning Based Compressive Subsampling are introduced in Section III. In Section IV the system architecture and circuits implementation are described, followed by numerical experiments. Section V presents the electrical measurements, while Section VI provides a discussion on results and concludes the paper.

II. SYSTEM LEVEL ANALYSIS

A high level view of the *System-on-Chip* (SoC) integrated on the implanted chip is depicted on the top left side of Fig. 1. The SoC is wirelessly connected to an external base station (on the right side of Fig. 1), where the compressed data is reconstructed for medical monitoring and storage.

The implanted SoC is composed by of neural amplifier, which collects the neural signals recorded by the electrodes, placed in contact with the brain surface. An *Analog to Digital Converter* (ADC), samples and digitises the amplified neural signals; the ADC output is processed by the *Digital Signal Processor* (DSP), aiming to reduce the amount of information sent by the wireless RF transmitter. Indeed, the transmitter power budget in typical wireless monitoring systems is usually one order of magnitude higher than any other system on the chip [7], [8]. In this Section, we discuss the system level design aspects and details of each block in the proposed SoC.

A. Macro and Micro-Electrodes for iEEG Recording

Bio-compatible electrodes are employed to collect the neural signal and act as an interface between the silicon microelectronics and the neurons. The electrode geometry is typically set according to the application; e.g., for measuring the single neuron activity, the size of electrodes is in the order of micrometers, while for studying the behaviour of population of neurons, the size may be larger. The micro electrodes and

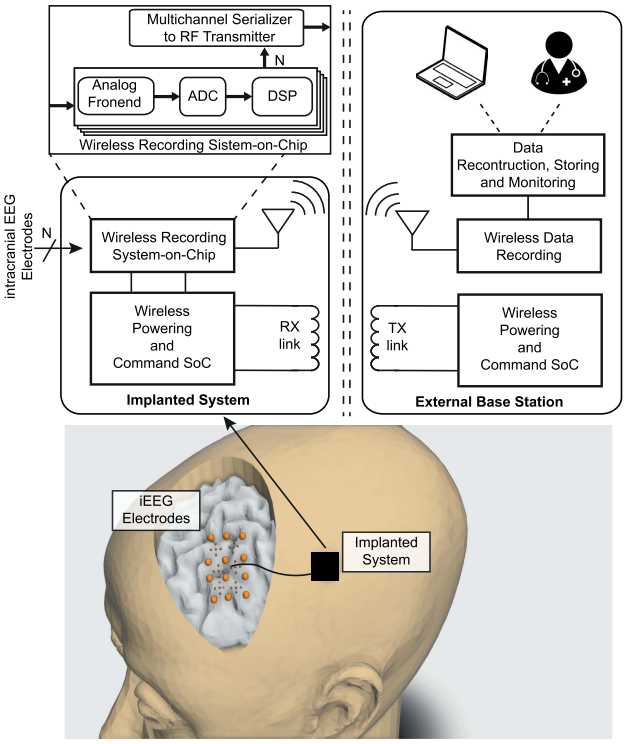

Fig. 1. Block diagram of the implantable integrated system (on the left side), wirelessly linked with an external base station (on the right), where the data is reconstructed for medical monitoring and stored. No battery is used in the implanted system.

their electronic read-out circuits have evolved according to the implant positioning and target neural activity. Needle-shaped micro electrodes are preferred for high-precision *Brain-Computer Interfaces* (BCI) [9], while flat electrodes are employed for cortical surface recordings [10]. Moreover, active electrodes are preferred to increase the quality neural data, and, consequently, power supply for the neural probes is provided. Wireless monitoring of brain activity would allow patients with implanted cortical system to safely leave the hospital during a monitoring period, that could potentially extend over several months.

Recordings from micro-electrodes of diameter less than 100 μm in the epileptic human hippocampus and neocortex have enabled the identification of several classes of electrographic activity localized to sub-millimeter-scale

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

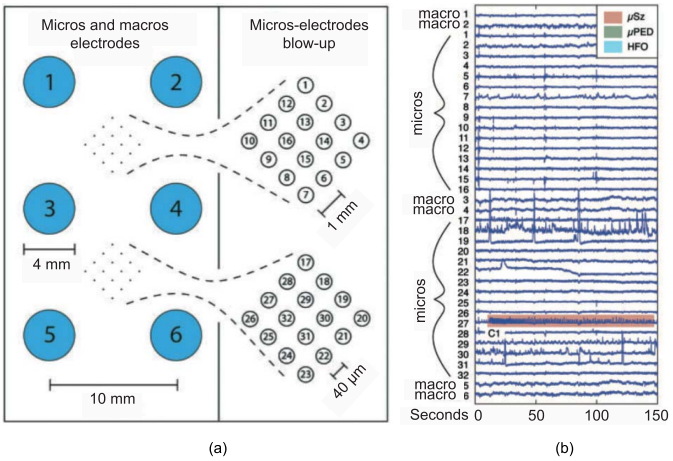

Fig. 2. Hybrid electrodes grid containing macro and microelectrode arrays (a) for iEEG signal recordings, reprinted from [12]. Signals recorded from micro and macro electrodes in (b), with an highlight on micro electrode 27 that records a seizure onset seconds before the macros.

tissue volumes, inaccessible to standard iEEG technology with macro-electrodes [11]. Moreover, Stead and colleagues [12] have observed that epileptic seizures identified on the macro-electrodes are often preceded by seizure-like activity on the micro-electrodes, depicted in Fig. 2. In particular, some of the micro-electrodes record an ongoing microperiodic epileptiform discharge, which starts minutes before the onset of the seizure itself [12], as highlighted on micro-electrode 27 shown in Fig. 2(b). Furthermore, the same researchers have also found that the signals recorded by adjacent micro-electrodes can be uncorrelated, despite their spatial vicinity. Furthermore, the sub-millimeter scale of high frequency oscillations involved in seizure generation motivate the wide-band recording of iEEG using micro-electrodes for precise monitoring of epileptic patients.

In this work, we consider neural signals collected and processed from every micro-electrode node, in order to accurately estimate the seizure onset using an implantable monitoring device. The focus of this work is on compressive sampling and wireless telemetry, while discussions on seizure detection algorithm are beyond the scope of this paper.

B. Data Processing

For each sampling electrode, the recorded signal is boosted by a *Low-Noise Amplifier* (LNA) (not described in this work). Then, the ADC, samples and digitises the analog neural signal. To meet the stringent area and power constraints of the proposed SoC, we have designed and implemented a *Successive Approximation Register* (SAR) ADC, which yields medium resolution and low-power data conversion.

Before data transmission, the digitized data is processed in order to reduce the power requirements of wireless TX. In many recently proposed implantable systems (e.g., [7], [13]–[15] and references therein), *Compressive Sampling* (CS) [16], [17] has been exploited to drastically reduce the amount of transmitted data, while still allowing robust, but complex, off-line reconstruction of original signal. CS stems from the fact that often, the information content of natural signals is much lower than the raw data content.

Given a training set of fully sampled signals, a novel Learning Based Compressive Subsampling (LBCS) [18] algorithm selects, from a representation basis like Wavelet or Hadamard, a fixed set of coefficients that capture, on average, most of the signals' energy. Only these coefficients will then be processed for new signals. Moreover, LBCS allows for very efficient linear encoders and decoders, reducing the time and power costs both on sampling and reconstruction, thus improving the conventional CS, where non-linear decoding (e.g. basis pursuit) is required for reliable signal reconstruction.

In the proposed work, we implement a fully digital encoder to compress the neural signal, which adaptively chooses the coefficients to sample, depending on the required signal quality during sampling. This process is enabled by a dynamic on-chip generation of the transformation coefficients, avoiding the large memory required to store all the transformation matrix entries [6]. Subsection IV-A addresses the analog to compressed data conversion, discussing the design and trade-offs in detail.

C. Wireless Data Telemetry

In addition to the neural data acquisition and processing, a communication channel from the implant to an external base station, namely *uplink communication*, is required to transmit digitized neural data to an external device. A *downlink communication* is needed for data transfer from external base station to the implant in order to configure sensor and processing parameters, such as sampling coefficient selections.

The proposed epilepsy monitoring system in this project implements both uplink and downlink communications. Since the downlink communication is only used for setting the system parameters, there is no need for a high data rate communication. Thus, it is sufficient a downlink receiver at the implanted SoC, which communicates at a data rate of 10 kbps. However, for the uplink communication, very high data rate communication is required, since the number of monitoring channels and their sampling rate is high. For the neural monitoring application with tens of electrodes, uplink communication should at least provide a data rate in the order of 10 Mbps. Accordingly, design of an uplink transmitter is challenging in such applications. The minimum distance for both communication types is the average human skull thickness of about 10 mm. The wireless data telemetry is further addressed in Subsection IV-C.

D. Wireless Powering

While in several biomedical applications such as hearing aids and pacemakers, batteries can occupy a significant volume, the area allocated to a neural implant is very small. Moreover, the neural recording systems consume higher amount of power and this would potentially reduce the duration of operation on battery. Considering the power demand of a neural implant aiming for continuous data transmission and the estimated power budget, current ambient energy harvesters are insufficient to fulfill this task. Wireless power transfer

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

based on inductive coupling is a proper choice, since the distance between the implant and external unit can be in the order of millimeters (human scalp thickness $\approx$ 10 mm), and sending the required power to the implant is accessible with current inductive coupling technology.

In this work, we propose a near field remote powering method, composed of four coils, an active half-wave rectifier, and a low drop-out voltage regulator, as further discussed in Subsection IV-D.

III. Learning-Based Signal Sampling

In this section, we first introduce the basics of Compressive Sensing, reviewing three recent approaches applied to neural signals. We then discuss non-linear structured recovery, before discussing Learning-Based Compressive Subsampling.

A. Compressive Sensing

Given an input signal $\mathbf{x} \in \mathbb{R}^N$ which has K non-zero coefficients, Compressive Sensing (CS) states that $\mathbf{x}$ can be robustly recovered from a signal $\mathbf{y} \in \mathbb{R}^M$ containing fewer samples than dictated by the Shannon-Nyquist theorem, with $M = \mathcal{O}(K \log \frac{N}{K})$. The compressed version of the input signal $\mathbf{x}$ can be expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{A}$ is a linear operator that either satisfies the Restricted Isometry Property (RIP) or is incoherent [19], and $\mathbf{w}$ represents the measurement noise. If the input signal $\mathbf{x}$ is not sparse in the given domain, an ortho-normal basis $\mathbf{\Phi}$ has to be used to get a sparser representation of the original signal $\mathbf{x}$. Natural signals are often characterized by sparse and structured representations in time-frequency (or space-frequency) domains, such as wavelets [20]. On the theoretical point of view, the matrix $\mathbf{A}$ can be generated with random coefficients, since i.i.d. sub-Gaussian matrices are incoherent and also satisfy the RIP condition. Moreover, they are universal, i.e., the RIP or the incoherence of $\mathbf{A}\mathbf{\Phi}$ is the same as of the original $\mathbf{A}$ [19], where matrix $\mathbf{\Phi}$ is used to form a sparser representation of the signal $\mathbf{x}$. However, sub-Gaussian matrices are prohibitively expensive to use in practice, since they require $\mathcal{O}(MN)$ space and time. Transmitting the fewer compressed samples $\mathbf{y}$ allows to save on-chip storage and telemetry power. However, the reconstruction process needed to recover $\mathbf{x}$ from $\mathbf{y}$ requires to solve non-linear optimization problems that increase both time and power requirements on the recovery node.

Bernoulli (BERN) described in [7], Multi-Channel Sampling (MCS) [14] and Structured Hadamard Sampling (SHS) presented in [21] are randomized sampling approaches recently proposed for the compression of neural signals. These three architectures are very efficient on the sampling side, but require solving non-linear optimization problems to reconstruct the original signals.

As described in [22] and references therein, a reduced number of samples is required for stable recovery, considering additional structures in the signal $\mathbf{x}$, such as interdependencies between its non-zero coefficients or constraints on its support during the recovery process. As discussed in [21], the Hierarchical Group Lasso (HGL) approach achieves the best performance over three different structured-sparsity recovery methods. This approach has been used to compare the reconstructed iEEG signals sampled through BERN, MCS and SHS methods.

B. Learning-Based Compressive Subsampling

The compression method used in this work is based on the LBCS approach [18], which requires both linear encoding and decoding with respect to a given orthonormal basis. Such method allows to simplify both the sampling and signal restoring steps, compared to standard CS approaches. In a nutshell, LBCS can be summarized considering the following compression model

$$\mathbf{y} = \mathbf{P}_\Omega \mathbf{\Psi} \mathbf{x}, \tag{2}$$

where $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$ is an orthonormal basis and $\mathbf{P}_\Omega \in \mathbb{R}^{M \times N}$ is a subsampling matrix, whose rows are canonical basis vectors. The effect of applying $\mathbf{P}_\Omega$ to $\mathbf{\Psi}\mathbf{x}$ is to return a M-dimensional vector containing only the components of $\mathbf{\Psi}\mathbf{x}$ indexed by the set $\mathbf{\Omega}$, also known as the *subsampling map*. The vector $\mathbf{y} \in \mathbb{R}^M$ is the compressed version of $\mathbf{x}$, with a nominal compression rate (CR) of $\frac{N}{M}$. The signal $\mathbf{x}$ is then approximately recovered via the fast linear decoder

$$\hat{\mathbf{x}} = \mathbf{\Psi}^* \mathbf{P}_\Omega^T \mathbf{y}. \tag{3}$$

where $\mathbf{\Psi}^*$ is the conjugate-transpose of Ψ and $\mathbf{P}_\Omega^T \mathbf{y}$ constructs a N-dimensional vector of zeros, placing the components of $\mathbf{y}$ in the positions indexed by $\mathbf{\Omega}$.

The learning process is dictated by a training set $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ of m fully sampled signals of unit norm. The optimal subsampling map $\mathbf{\Omega}$ is learnt by choosing the indices that capture most of the average energy in the transform domain:

$$\hat{\mathbf{\Omega}} = \underset{\mathbf{\Omega}, |\Omega| = M}{\arg\max} \ \frac{1}{m} \sum_{j=1}^{m} \sum_{i \in \Omega} |\langle \boldsymbol{\psi}_i, \mathbf{x}_j \rangle|^2, \tag{4}$$

where $\boldsymbol{\psi}_i$ is the i-th row of $\mathbf{\Psi}$. $\hat{\mathbf{\Omega}}$ can be exactly found by selecting the M indices whose values of $\frac{1}{m} \sum_{j=1}^{m} |\langle \boldsymbol{\psi}_i, \mathbf{x}_j \rangle|^2$ are the largest [18]. The learnt sampling scheme is then used to directly sample only those transform coefficients indexed by $\hat{\mathbf{\Omega}}$ for all signals $\mathbf{x}$.

Walsh-Hadamard based transformation has been used in recent publications [6], [23] because of its hardware friendly implementation, since each transformation coefficient requires one bit resolution, resulting in simple computations. In particular, Hosseini-Nejad *et al.* [23] propose a threshold-based Walsh-Hadamard compression, to sample the Action Potentials (AP) for brain machine interfaces. The authors apply a butterfly scheme to transform the input signal samples into the Hadamard domain. However, such butterfly-based method can be performed on very few number of consecutive samples (8 samples in [23]), limiting any kind of learning approach due to the low signal statistic. Therefore, it is used for AP signal detection, with limited application in continuous medical monitoring for diseases like epilepsy, where the whole

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

signal behavior is required by clinicians. Majidzadeh *et al.* [24] propose the generation of the full Hadamard matrix $\mathbf{\Psi} \in \mathbb{R}^{16\times16}$ for a parallel neural recording system. However, such implementation does not apply any compression mechanism, resulting in a high power consumption. The circuit implementation of LBCS technique with DCT-based transform has been proposed in [25]. Even though its implementation shows a great signal reconstruction performance, the actual hardware implementation, which requires relatively larger area and power consumption with respect to its LBCS-Hadamard counterpart, makes it more suitable for different application, such as image processing. In [6], LBCS is exploited using the Hadamard transformation matrix, where the whole Hadamard matrix is stored in static memories requiring more than 2/3 of the actual encoding area.

In this work, we propose an LBCS based compression algorithm, which performs the transformation from temporal to Hadamard domain, through on-the-fly generated Hadamard coefficients. In this implementation, only the selected rows of the Hadamard matrix (defined by $\hat{\Omega}$) are generated and used for the embedded compression, resulting in a dynamic generation of the coefficients that are used to apply the LBCS approach. Such technique drastically reduces the encoder memory requirements needed by previous LBCS-Hadamard implementation, while the signal reconstruction quality is preserved within a low power chip implementation.

C. Walsh-Hadamard Transformation

The Hadamard transform is particularly suited for hardware implementation since each coefficient can be computed by performing only simple additions or subtractions.

The reduction of hardware area in the Had-based LBCS described in [6] is possible by replacing the SRAM dedicated to store the Hadamard coefficients, with a direct computation of each matrix entry [24]. Such computation is feasible due to the intrinsic structure of the Hadamard matrix, which is summarized as follows. The non-normalized Hadamard transformation matrix $\hat{H}_n \in (-1, 1)^{N\times N}$ of size n, with $N = 2^n$ is expressed as a recursive Kronecker product of two matrices

$$\hat{H}_n = \hat{H}_1 \otimes \hat{H}_{n-1}, \quad \text{where } \hat{H}_1 \triangleq \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{5}$$

Each matrix coefficient indexes k and j, can be expressed in binary representation

$$k = \sum_{i=0}^{n-1} k_i 2^i, \quad j = \sum_{i=0}^{n-1} j_i 2^i \text{ with, } k_i, j_i \in (0, 1). \tag{6}$$

Each Hadamard entry $h_{k,j}$ can then be expressed as

$$h_{k,j} = (-1)^{\sum_{i=0}^{n-1} k_i j_i} \equiv (-1)^{mod_2(\sum_{i=0}^{n-1} l_i j_i)}. \tag{7}$$

In particular, mapping the (1, -1) to (0, 1), each Hadamard entry can be derived by

$$h_{k,j} = mod_2(\sum_{i=0}^{n-1} l_i j_i). \tag{8}$$

Such expression can be efficiently implemented in hardware, through logic AND gates to perform $l_i j_i$, while the module-2 sum is derived by a logic XOR. Thus, the circuit implementation takes the row and column indexes k and j and computes the Hadamard coefficient in the binary map (0, 1).

D. Dataset Details and Experimental Protocol

The *iEEG.org* portal contains several datasets of intracranial EEG data which are manually annotated by expert clinicians. The dataset *I001-P034-D01* has been used for the development of this research [6], [21]. It consists of approximately 1 day, 8 hours and 10 minutes of recordings at 5 kHz, or approximately 6×10^8 samples of intracranial EEG data. In order to reduce the dataset size, we use samples from the 12th and 13th seizures, and an equal number of samples before the seizure onset, for training and testing, respectively. More in details, we have used 207k samples before and after the seizure for the training signal, while for the test set we have used 153k samples.

The training set of the dataset is used to learn the sampling pattern for the LBCS approach and also to tune the variable density parameters for the SHS method. Once the sampling pattern is fixed, LBCS uses it to compress all the signal windows in the test set. The reconstruction is then performed with the linear decoder (3). For the randomized methods of MCS, BERN and SHS, we draw 20 different sampling patterns from the relative distributions for each signal window (with length N = 256) in the test and reconstruct using the tree-based HGL norm, which yields the best results [21].

IV. Implantable Architecture

The implantable chip architecture is described in this Section. The SoC designed in this work consists of the analog to digital converter, followed by the encoder which compresses the sampled data, implementing the Learning-based CS algorithm described in Section III-B. The compressed bit stream is then serialized and wirelessly sent out by the RF transmitter. The circuit can be powered wirelessly through an inductive link between the implant and a power delivery unit.

A. Analog to Compressed Data Stream

The neural signal digitization is realized by a Successive Approximation Analog to Digital Converter (SAR ADC). In such ADC topology, just one comparator is required and the design is based on a charge redistribution DAC, thus this implementation results to be energy efficient. However, the SAR ADC will require N + 1 comparison periods to prepare the final decision. Hence, the SAR ADC is expected to allow the lowest power dissipation, but also it is defined by a moderate sampling rate. The ADC design results in a compact and low-power implementation, which matches the stringent area and power constraints of our implantable SoC. The SAR ADC has 8 bit resolution and a sampling rate of 45 kHz, in order to match the 5 kS/s rate of the input signal from iEEG dataset. A compact ADC implementation is achieved by a binary-weighted capacitive array, with attenuation capacitor [14]. Since the neural signal bandwidth is relatively low, the compression computations are completed at the DSP,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

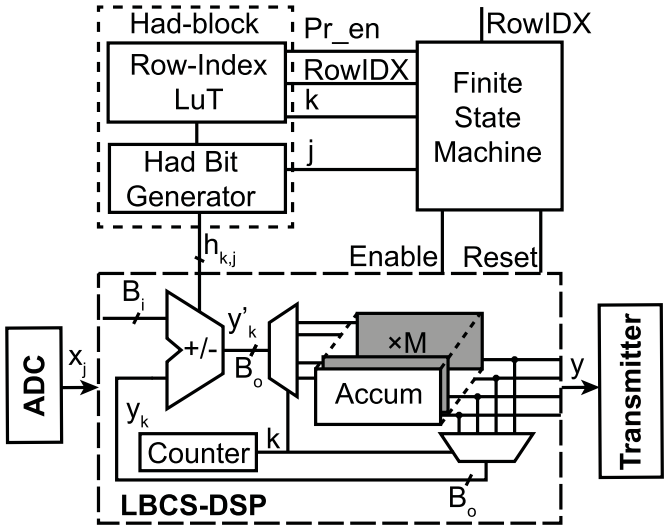

Fig. 3. One channel block diagram showing the LBCS encoder and the matrix sequence generation logic.

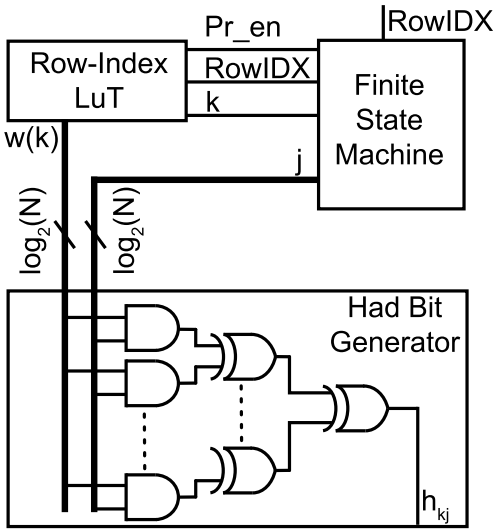

Fig. 4. Hadamard bit generator block diagram.

with the same frequency defined by the ADC. In particular, the ADC requires 9 cycles to complete the digitization of the input signal (at 5 kHz), thus running at 45 kHz. The DSP core frequency runs at the same speed, performing the data compression.

The Hadamard-based LBCS encoder block diagram is depicted in Fig. 3, where is shown the input data path from the Analog to Digital Converter (ADC), through the LBCS Digital Signal Processor (DSP) to the encoded data transmitter. The *Finite State Machine* (FSM) of the DSP drives the *Had-block* and the main DSP core, where the encoding process is executed. The Had-block generates the Hadamard bit streams and replaces the SRAM used in previous implementation [6], reducing the encoder area. The Had-block is mainly composed by the Row-Index Look up Table (LuT), and the Hadamard bit generator. The Row-Index LuT is meant to store the learnt indices of the sub-sampling matrix $\mathbf{P}_\Omega$, described in subsection III-B. Assuming that only M rows of the full Hadamard matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$ have to be used to apply the LBCS-based compression, then we can define a mapping function $w(k) =\in [0\ N-1]$, where $k \in [0\ M-1]$ is the index of the output value, and we define $h_{k,j} = h_{w(k),j}$. Then, the LuT implements such mapping function $w(k)$.

The LuT coefficients, driven by the FSM, are sent to the Hadamard-bit generator, which produces the transformation entries $h_{k,j}$, following the description in subsection III-C. Fig. 4 shows the block diagram of the Hadamard bit generator, highlighting the logic gates used to generate the $h_{k,j}$ entries [24]. During a calibration phase, the learnt Hadamard row indices, defined by the RowIDX input ($log(N)$ bit wide, to code all the possible Hadamard matrix indexes) are loaded in the LuT. As soon as the program enable (Pr_en) is active, the initialization starts and the FSM programs the M indexes into the LuT, following the RowIDX and the k signals used to correctly address the register. The FSM also generates and programs the enable and reset commands sent to the DSP, to correctly synchronize the encoding procedure, and to reset the accumulator registers (*Accum* in Fig. 3) at the end of each encoding window.

The encoder input signal x_j, digitized by the ADC with B_i bit resolution, is summed or subtracted from the previous accumulator register values, at each sampling instant j in the sampling window of length N. The LBCS-DSP block performs the embedded compression, defined as

$$y_k = \sum_{j=1}^{N} h_{k,j} x_j, \quad k \in \{1, \ldots, M\}, \tag{9}$$

where $h_{k,j}$ is the (k, j)-entry of $\mathbf{H}_\Omega = \mathbf{P}_\Omega \mathbf{H}$; the Hadamard matrix $\mathbf{H}$ (=$\mathbf{\Psi}$ described in subsection III-B), requires a single bit per entry, minimizing the computation costs in the transformation process. The encoder processing frequency is M times faster than the input signal frequency, in order to update each of the accumulator registers, where the transformation coefficients are stored. As analysed in [6], a DSP that for each sampling window performs the full-compression (with no CS or learning-based CS) would require higher power and area requirements. In particular, the area and power overhead would be higher than CR× [6].

The previous Hadamard based LBCS implementation shown in [6], has been designed for sampling window of 256 samples (N = 256), with a fixed CR of 16×. In this work, we propose the hardware implementation with an on-the-fly Hadamard generation, with a sampling window length of N = 64 and compression rate of CR = 8. The same dataset as in [6] has been taken into account, to validate the proposed hardware implementation. The N = 64 and CR = 8 combination allows to get similar average reconstruction quality, while the LBCS encoder frequency f_s is halved, resulting in a lower power consumption. Indeed, since M is defined as N/CR, the larger is the number of the Hadamard rows M, the higher is the core LBCS clock frequency, which might become a limiting factor. On the other hand, a further reduction on the number of samples N, would degrade the signal statistics over which the learning approach is based on.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

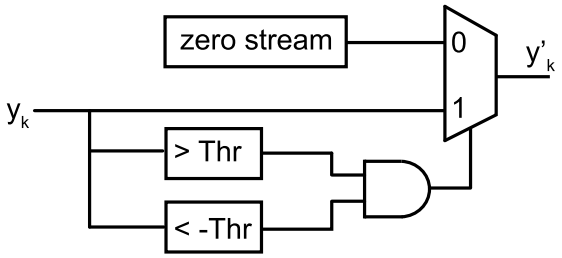

Fig. 5. Variable CR block diagram, defined by the threshold level (Thr).

B. Variable Hadamard Compression

The simulation results shown in Fig. 6-(a), depict the energy content of the N samples in the Hadamard domain, for a particular sampling window. As described in Sec. III, the Learning-based algorithm allows to define the coefficients that, in average, have the most energy contribution. However, depending on the signal evolution in the sampling window, the coefficients defined by the learning process might have a low energy content. This analysis is useful to define the system's trade-off and a variable compression rate, which adapts from window to window, depending on the energy levels defined by the neural signal evolution in time. On the system level implementation, for a window length of N = 64, a maximum compression rate of 8 has been defined, in order to allow relatively high SNR after the signal reconstruction. Since in the $\frac{N}{CR} = 8$ Hadamard coefficients the energy might be below a certain level, a threshold is also defined during the learning process, in order to transmit only the most relevant coefficients, enabling a dynamic compression. The dynamic detection of the Hadamard coefficients results in an easy hardware implementation, and allows a variable CR from window to window. Fig. 5 shows the block diagram of the variable CR implementation, depicting how, the energy content of the coefficient value y_K is transmitted or substituted with a B_O bit stream by means of a multiplexer, mathematically resumed as:

$$y'_k = \begin{cases} 0, & |y_k| < \text{Threshold} \\ y_k, & \text{otherwise.} \end{cases} \tag{10}$$

In such a design implementation, the SoC features a compression which varies from CR = 8 to CR = 64, and allows the TX to transmit fewer coefficients, thus drastically reducing its power consumption. Fig. 6-(b) shows the trade-off between the mean signal reconstruction SNR and the mean CR over the whole dataset, as the threshold varies. In particular, Fig. 6-(c) and Fig. 6-(d), show respectively the mean signal recovery quality and the mean window compression rates, with respect to the threshold levels. In particular, it is worth highlighting how a relatively small threshold (e.g., below 100) allows to reduce the number of coefficients transmitted (thus, higher CR level), while the SNR is still relatively high (above 28 dB). In particular, it is worth noticing that such SNR value is around 18 dB higher than the minimum SNR value that is acceptable to successfully allow the seizure detection [14].

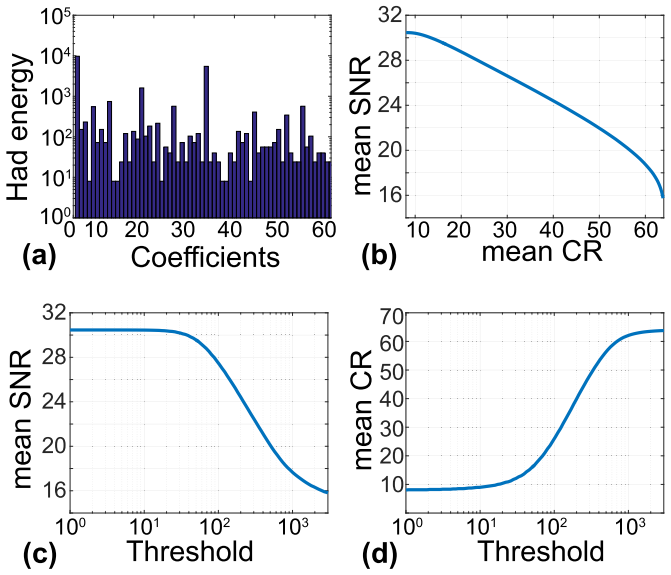

Fig. 6. SNR analysis for adaptive approach.

C. Wireless Data Transmitter

Two different wireless transmitters are designed and implemented with different data rates, operating frequency, and transmission distance, in order to cover different applications. The narrowband transmitter which operates in the MedRadio band at 416 MHz is designed for low data rate and indoor communication. The other transmitter is based on impulse-radio ultra-wideband (IR-UWB) in the 3.1-10.6 GHz frequency range and utilized for high data rate and very short distance transmission. The two transmitters provide the flexibility of sending compressed or raw data.

1) Narrowband Transmitter: The proposed on-off keying (OOK) modulated narrowband transmitter is based on the turning on and off a voltage controlled oscillator (VCO). The VCO which is shown in Fig. 7 is composed of NMOS and PMOS cross-coupled pairs and data is applied to the bias current for modulation. Reuse of the current by PMOS and NMOS pairs provides higher transconductance and higher voltage swing on the inductor. For setting the resonance frequency of the VCO, a bank of three capacitors are utilized for discrete tuning and varactors are used for fine tuning. An off-chip loop antenna is connected to the differential output of the VCO to transmit the signal and create the required inductance for LC tank [26].

2) Ultra-Wideband Transmitter: IR-UWB is a promising technique based on transmission of short pulses and it is very efficient for low range applications which requires high data-rate. In 2002, the Federal Communications Commission (FCC) approved and limited the maximum effective isotropic radiated power (EIRP) to -41.3 dbm/MHz for bandwidth between 3.1 and 10.6 GHz [27].

In this work, in addition to the narrowband transmitter, we present a high data-rate, energy and area efficient, and low complexity IR-UWB transmitter. Fig. 8 shows the schematic block diagram of the IR-UWB transmitter. The small number of circuit elements make the design simpler and area

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

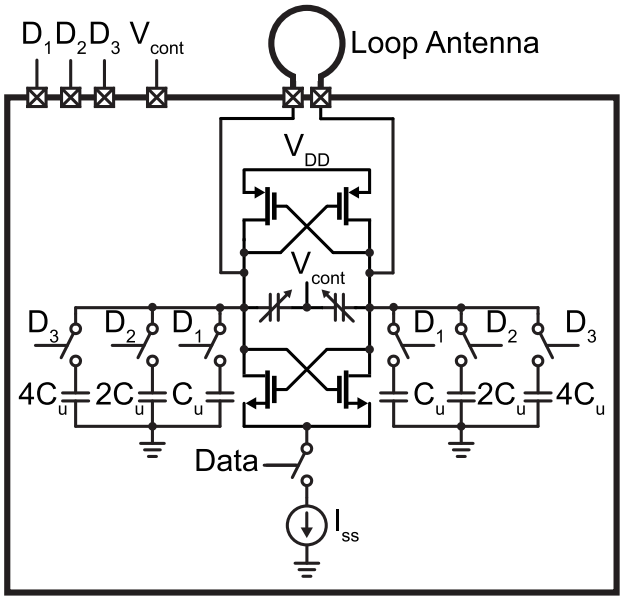

Fig. 7. Schematic of the LC cross-coupled voltage controlled oscillator.

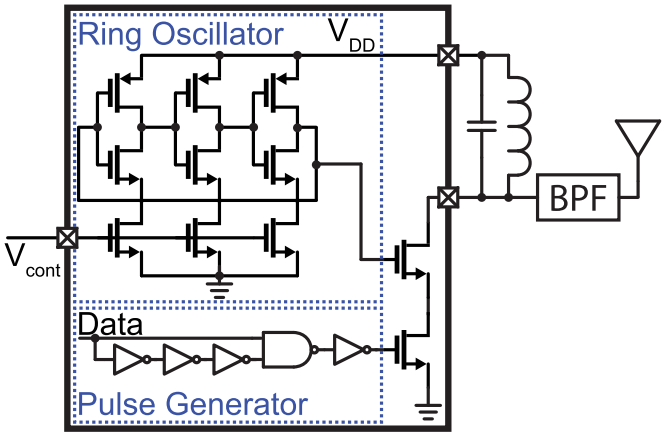

Fig. 8. Schematic of the IR-UWB transmitter.

occupation minimal. The core of the transmitter is based on the current starved ring oscillator (RO) which generates output in the range of 3.5-4.5 GHz frequency. The control voltage provides flexibility in selecting the oscillation frequency of the ring oscillator by adjusting the bias current. The pulse generator (PG) block creates short pulses at the rising edges of the data signal. The output of the RO and PG is mixed with cascode connected transistors. The drain of the transistor driven by the RO is connected to external resonator circuit formed by an inductor and a capacitor. Before the 50 Ω UWB antenna, a band-pass filter (BPF) centered at 4 GHz is used in order to satisfy the FCC regulation. For the transmission of the generated IR-UWB pulses, miniaturized, flexible and polarization-diverse UWB antenna presented in [28] can be adopted.

D. Wireless Power Transfer

To design an implantable system, wireless power transfer (WPT) method is chosen since batteries increase the total weight and dimensions of device. Considering the required power of the implant and the power transmission distance, which is in the order of millimeters, an inductive link is selected for power transmission. The losses due to remote powering are a critical concern that can cause a temperature elevation, which may damage the tissue. Hence, a power efficient transmission link composed of 4-coils, an active half-wave rectifier, and a low drop-out voltage regulator is designed and represented in Fig. 9.

TABLE II
DESIGN PARAMETERS OF INDUCTIVE LINK COILS

	External Coils		Implanted Coils	
	Outer	Inner	Outer	Inner
Inner diameter (mm)	29	13	6.9	1.9
Outer diameter (mm)	43	24	13.5	5.5
Width & Spacing (mm)	1	0.5	0.3	0.2
Number of turns	4	6	6	5

Different approaches are used for various applications, but the average power consumptions of the implants are considered nearly constant in system parameters. However, in some applications such as neural monitoring with a variable number of active electrodes, the power consumption of the implant is not always the same. Hence, the power efficiency of WPT and the dimensions of the implanted coil become the major limitations in designing the coils for remote powering. In the fundamental approach with two coupled coils, there is a direct relation between the delivered power to the load and the efficiency. The variation in the load power requires an additional approach for keeping power transfer efficiency (PTE) maximum for different activity rates. A modified version of inductive link with 4-coil instead of 2-coil has been introduced for 2 meters remote powering [29], and the structure was adapted for implant powering applications [30]. The results show a significant improvement in the efficiency. The low coupling coefficient and the low quality factor of the coils in 2-coil link are compensated by the introduced two high quality factor coils between them [31]. Moreover, the introduced coils transform different load impedances to the optimal impedance at the input of the inductive link and efficiency does not significantly change with load power. Therefore, a 4-coil inductive link is implemented to take the advantage of high PTE and tolerance for variable load power. The selected geometrical parameters for 4-coil inductive link are represented in Table II.

The operation frequency of chosen 4-coil inductive link has a significant impact on the PTE and safety of the implanted system. Absorbed power by the tissue decreases the PTE and creates a temperature increase in the surrounding. The maximum temperature elevation is limited to 1 ^{o}C by the regulations for body implants [32]. To comply with the regulations, low MHz range (1-20 MHz) operation frequency is usually chosen since the absorption of the cortical tissue is minimum in this range [33]. In addition to the absorbed energy, the power consumption of the implanted circuitry causes a temperature

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

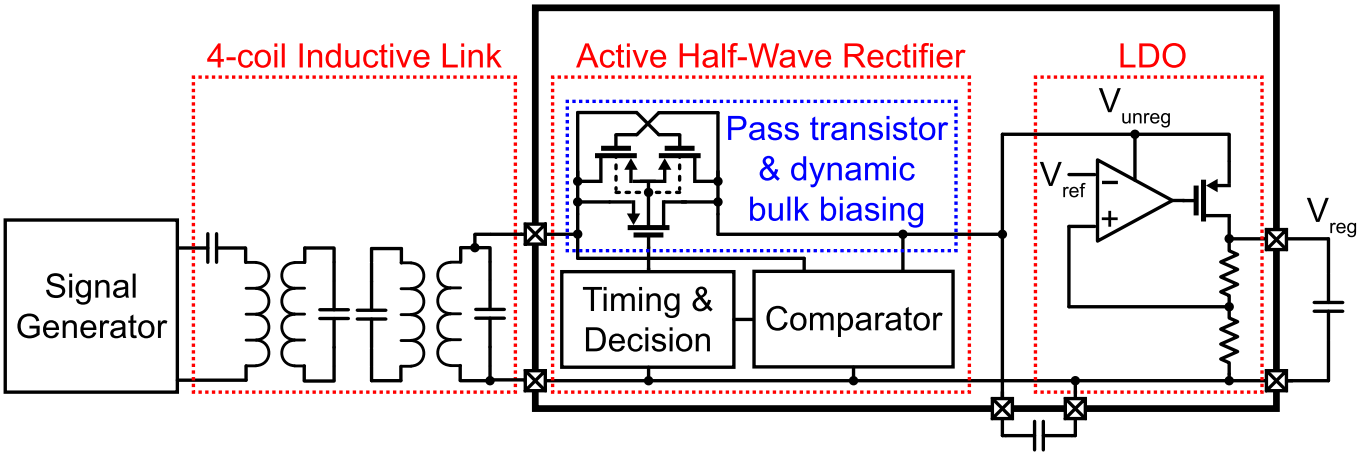

Fig. 9. Block diagram of the proposed implanted electronics for wireless power transmission.

increase in the tissue. Another study shows the limits of the maximum allowable power dissipations depending on the power dissipation level, chip size, and location of the implant [34]. In this work, the operation frequency of 8 MHz is chosen to minimize the tissue absorption and implanted system is designed for the low power consumption to limit the temperature elevation.

The induced AC voltage by the 4-coil inductive link requires to be rectified to a DC voltage. To achieve high conversion efficiency, an active half-wave rectifier is selected at the price of the losses in the comparison and decision blocks in Fig. 9. In this study, the half-wave rectifier is designed based on the work published in [35]. Pass transistor with dynamic bulk biasing constitutes the core of the rectification. To prevent the leakage from the capacitor to the input, the n-well of the two PMOS transistors are dynamically biased. Hence, the transistor conducts current only when the input voltage is higher than the voltage at the accesses of the capacitor. The comparator decides the condition of the PMOS pass transistor by comparing the input voltage and the charged voltage on the capacitance. Timing and control block applies the decision given in the comparator with an optimum switching time such that it is fast enough compared to operation frequency and minimizes the switching power losses. The low drop-out voltage regulator eliminates the ripples at the output of the rectifier and generated clean voltage supply for the other circuits in the implant. The capacitors at the output of the rectifier and regulator are implemented externally.

V. Measurement Results

The chip, fabricated in UMC 180 nm 1P6M MM/RF process technology, has been packaged and bonded to a dedicated PCB. A Xilinx development board, providing a Virtex 5 FPGA [36], is linked to the PCB trough rigid headers, as shown in Fig. 11. The board is used to set and program the SoC blocks with a PC station.

A. Sampling and Data Compression

Each block of the SoC has been independently connected to dedicated pads on the chip, in order to validate each design.

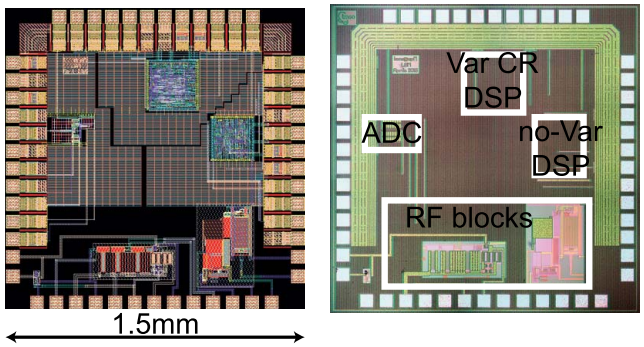

Fig. 10. Layout (on the left) and micrograph (on the right) of the tested chip.

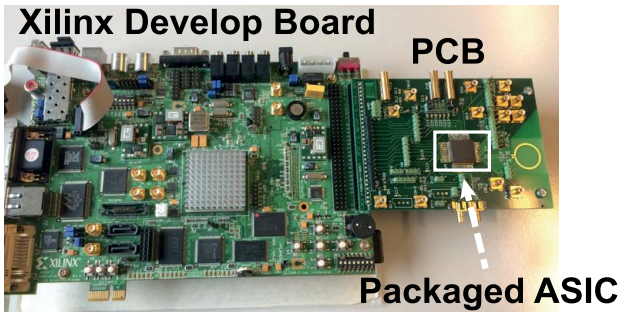

Fig. 11. Measurement setup, highlighting the FPGA and PCB link.

The analog input of ADC and the DSP digital bit streams are connected to ESD protection circuits, to reduce any possible damage due to electrostatic discharges during the measurements.

As shown in Fig. 10-left, the ADC and the two encoder versions (the variable CR on top and the non-variable version on the right side of the ASIC) do not share the power-grids, in order to separate the analog and digital domains. The power-grid has been designed in a very dense manner, with capacitors that surround the SoC blocks, stabilizing the VDD to ground fluctuations. Fig. 10-right shows the micrograph of the tested chip.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

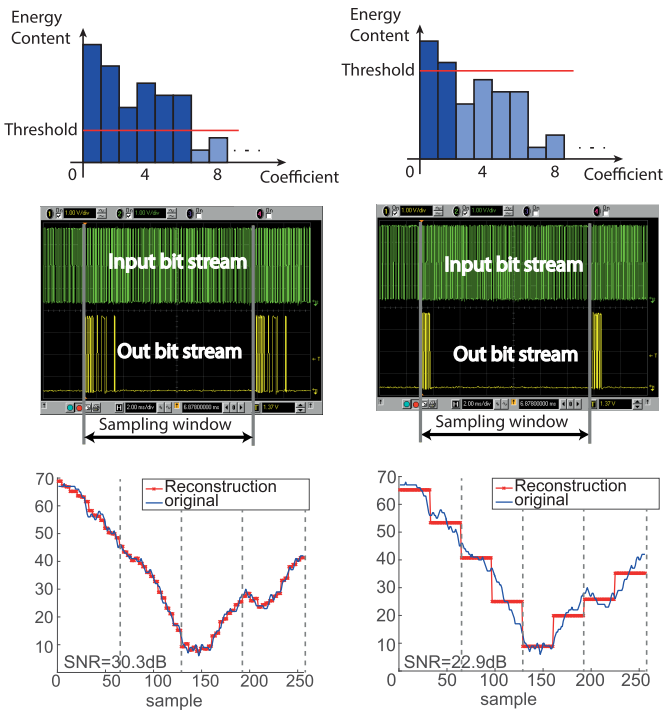

Fig. 12. Measured compressed values with low threshold (on the left) and high threshold (on the right).

The 8 bit resolution SAR-ADC with a sampling rate of 45 kHz requires an area of $230\mu\text{m} \times 150\mu\text{m}$, with a power consumption of 0.46 μW. The low power requirements of the ADC is mainly dictated by the medium resolution of 8 bits, and the low sampling frequency of the neural signals.

A Verilog code, implemented on Xilinx ISE tool, has been developed to program the encoder registers, to provide the clock at 45 kHz to the SoC, and to send the input bit stream to the encoders through the FPGA. The compressed data sequences at the output of the DSPs are collected as input to the FPGA, and analyzed with Xilinx ChipScope tool. The measurement setup is shown in Fig. 11.

The measured compressed bit streams have been plotted by an oscilloscope and are highlighted in Fig. 12. Both plots have been generated with the variable CR encoder version, in order to show, on the same plot, the dynamic generation of the transformation coefficients, and the different outputs due to low threshold (on Fig. 12 top-left) and high threshold (on Fig. 12 top-right) settings. The reconstructed signal versus the original data is plotted for 4 sampling windows, at the bottom of Fig. 12.

Table III reports the numerical results of the recovered signal, for the different compression methods discussed in this work, with fixed compression rates. In particular, this table shows how the LBCS-based signal recovery (the only method which applies a learning-based compression technique) performs better than Bernoulli [7], Multi-channel [14] or Structured Hadamard Sampling [21]. The comparison of reconstruction performance has been done considering $N = 256$ and an ADC resolution $B_i = 10$, for the iEEG dataset described in Subsection III-D. Furthermore, the LBCS

TABLE III

RECOVERY PERFORMANCE COMPARISON WITH PUBLISHED WORK ($N = 256$, $B_i = 10$)

Method	Compression rate			
	8	16	32	64
LBCS	**33.27**	**28.48**	**23.27**	**18.06**
SHS HGL	23.89	20.26	18.53	14.49
BERN HGL	20.49	16.87	13.53	11.15
MCS HGL	20.92	17.48	n.a.	n.a.

Only LBCS approach applies a learning-based compression scheme.

TABLE IV

RECOVERY PERFORMANCE SUMMARY FOR THIS WORK ($N = 64$, $B_i = 8$)

Method	Compression rate[a]			
	8	16	32	64
LBCS	**30.4**	**29.5**	**26.1**	**15.7**

[a] Average compression rate over the whole dataset.

TABLE V

COMPRESSION HARDWARE COMPARISON WITH PUBLISHED WORK

Parameter	[7]	[14]	This Work
Compression Method	BERN	MCS	LBCS
Compression Rate	10	16	Variable CR from 64[b] to 8[b]
Technology [μm CMOS]	0.09	0.18	0.18
Compression Power [μW]	1.9 at 0.6 V	17.83[a] at 1.2 V	0.47 at 0.8 V
Compression Area [mm^2]	0.090	0.090	0.059

[a] Compression power cost over 16 channels.
[b] Average compression rate over the whole dataset.

signal recovery requires the linear decoder (3), which yields the reconstructions at a fraction of the computational cost of the other methods [6].

Since the actual hardware implementation of this work has been developed with $N = 64$ and $B_i = 8$, Table IV summarizes the recovery performances for the variable encoder design, for different fixed energy thresholds (the reported CR are in average over the whole dataset). For this reason, Table IV gives an *energy content based* comparison, while Table III reports a *CR-based* comparison.

The Learning-based compression algorithm with dynamic generation of the transformation coefficients requires an area of $230\mu\text{m} \times 330\mu\text{m}$. A comparable area of $230\mu\text{m} \times 365\mu\text{m}$ is required for the adaptive DSP design, which only consumes 0.47 μW at 0.8 V. Table V reports the hardware comparison with respect to other published works.

During the measurement, each subsystem has been tested independently. The output data stream from the DSP has been serialized in a shift register of size $M \times B_O$, in a MSB first order. The serialized data is then directly transmitted by the RF block.

B. Wireless Power Transfer

The resonance frequency of each LC tank in the 4-coil inductive link is fixed at 8 MHz. Power transfer efficiency of 55% is obtained for the inductive link when the separation between the coils and the load is 10 mm and 10 mW, respectively. The performance of the rectifier and the regulator

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

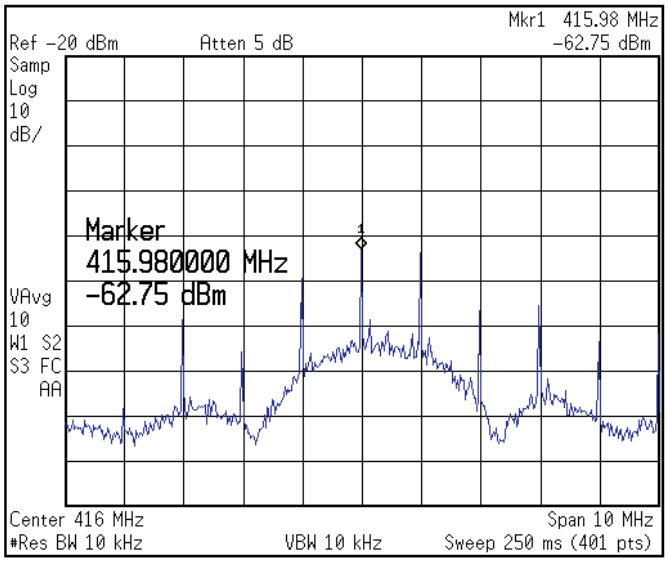

Fig. 13. Spectrum of the LC cross-coupled voltage controlled oscillator.

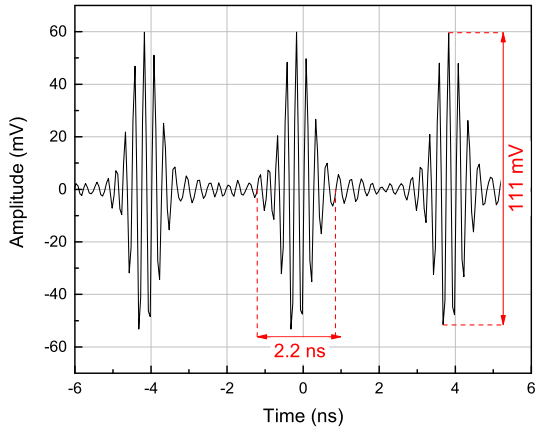

Fig. 14. Transient pulses of the IR-UWB transmitter at 250 Mpps.

is also characterized for 10 mW load and their efficiency reach to 82% and 78%, respectively. As a result, wireless power transmission beginning from the signal generator to implant load is achieved at 36% efficiency.

C. Narrowband Transmitter

The VCO is supplied with internally generated 1.8 V and the measured average power consumption during operation is 248.4 μW. Thanks to the discrete and fine tuning capacitors, VCO covers the two MedRadio bands (401-406 MHz and 413-419 MHz). Fig. 13 shows the frequency spectrum of the OOK transmitter with the highest data rate of 2 Mbps. During the measurement of the spectrum, the distance between the transmitter antenna and the receiver antenna (Taoglas Limited-TI.10.0112), which was directly connected to the spectrum analyzer, is fixed to 60 cm. A custom made OOK receiver board based on discrete components is used to demodulate the transmitted data.

D. Ultra-Wideband Transmitter

The proposed IR-UWB transmitter is fabricated and it occupies a 60 μm $\times$ 30 μm area. Fig. 14 shows the measured

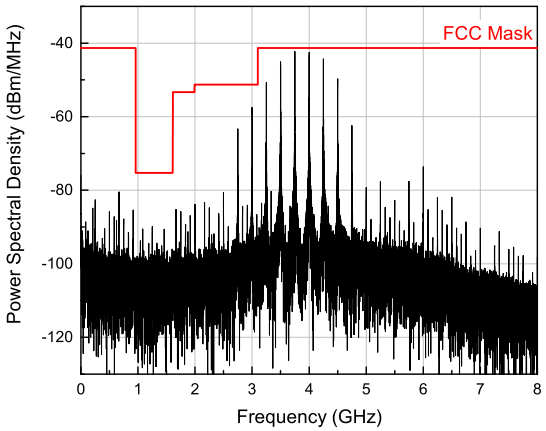

Fig. 15. Power spectral density of the IR-UWB transmitter.

output waveform of the implemented IR-UWB transmitter with 250 MHz pulse repetition rate. The maximum peak-to-peak amplitude of the measured pulse is 111 mV while its duration is 2.2 ns. Fig. 15 depicts the measured power spectral density of the transmitter and FCC regulation. The triangular envelope of the output waveform suppress the side-lobes and measured spectrum fully meets the FCC mask. When the pulse repetition frequency is 250 Mpps, the complete IR-UWB transmitter consumes 11.3 mW power which corresponds to 45.2 pJ/pulse. High throughput of the IR-UWB transmitter makes it possible to buffer the raw data and transmit it in several bursts.

VI. Conclusion

This work proposes a novel LBCS-based SoC for recording neural signals in implantable devices. The proposed encoding solution enables dynamic generation of the transformation coefficients, allowing on-the-fly compression with faster and improved off-line signal recovery than Random Bernoulli [7], Multi-channel [14] or Structured Hadamard Sampling [21]. Moreover, a variable compression rate is achieved by energy based threshold method. The proposed data compression reduces the amount of bit stream transmitted wirelessly, thus lowers the TX and implantable system's power requirements.

In the proposed design, the threshold is set during the off-line learning process. A further development of the current chip implementation can include an on-chip calibration, which sets the threshold level of the encoder in the implanted device. Moreover, in the multichannel implementation, the design should take into account the synchronization of the data from the different electrodes.

References

[1] R. R. Schaller, "Moore's law: Past, present and future," *IEEE Spectr.*, vol. 34, no. 6, pp. 52–59, Jun. 1997.
[2] A. C. Hoogerwerf and K. D. Wise, "A three-dimensional microelectrode array for chronic neural recording," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 12, pp. 1136–1146, Dec. 1994.
[3] C. B. Nemeroff *et al.*, "VNS therapy in treatment-resistant depression: Clinical evidence and putative neurobiological mechanisms," *Neuropsychopharmacology*, vol. 31, no. 7, pp. 1345–1355, 2006.
[4] M. Leonardi and T. B. Ustun, "The global burden of epilepsy," *Epilepsia*, vol. 43, no. s6, pp. 21–25, 2002.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[5] "Epilepsy-fact sheet," World Health Org., Geneva, Switzerland, Tech. Rep., Feb. 2017.

[6] C. Aprile *et al.*, "Learning-based near-optimal area-power trade-offs in hardware design for neural signal acquisition," in *Proc. Int. Great Lakes Symp. VLSI (GLSVLSI)*, 2016, pp. 433–438.

[7] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.

[8] S. Ha *et al.*, "Silicon-integrated high-density electrocortical interfaces," *Proc. IEEE*, vol. 105, no. 1, pp. 11–33, Jan. 2017.

[9] E. M. Maynard, C. T. Nordhausen, and R. A. Normann, "The utah intracortical electrode array: A recording structure for potential brain-computer interfaces," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, no. 3, pp. 228–239, Mar. 1997.

[10] D. Yoshor, W. H. Bosking, G. M. Ghose, and J. H. R. Maunsell, "Receptive fields in human visual cortex mapped with surface electrodes," *Cerebral Cortex*, vol. 17, no. 10, pp. 2293–2302, 2006.

[11] G. A. Worrell *et al.*, "High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings," *Brain*, vol. 131, no. 4, pp. 928–937, 2008.

[12] M. Stead *et al.*, "Microseizures and the spatiotemporal scales of human partial epilepsy," *Brain*, vol. 133, no. 9, pp. 2789–2797, 2010.

[13] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 1959–1962.

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

[15] M. Shoaran *et al.*, "A 16-channel 1.1mm^2 implantable seizure control SoC with sub-μW/channel consumption and closed-loop stimulation in 0.18μm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2016, pp. 1–2.

[16] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, Aug. 2006, pp. 1433–1452.

[17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[18] L. Baldassarre, Y.-H. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher. (2015). "Learning-based compressive subsampling." [Online]. Available: https://arxiv.org/abs/1510.06188

[19] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Basel, Switzerland: Birkhäuser, 2013.

[20] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 1999.

[21] L. Baldassarre, C. Aprile, M. Shoaran, Y. Leblebici, and V. Cevher, "Structured sampling and recovery of iEEG signals," in *Proc. 6th IEEE Int. Workshop Comput. Adv. Multi-Sensor Process. (CAMSAP)*, Dec. 2015, pp. 269–272.

[22] A. Kyrillidis, L. Baldassarre, M. El Halabi, Q. Tran-Dinh, and V. Cevher, "Structured sparsity: Discrete and convex approaches," in *Compressed Sensing and Its Applications*. Cham, Switzerland: Birkhäuser, 2015. pp. 341–387.

[23] H. Hosseini-Nejad, A. Jannesari, and A. M. Sodagar, "Data compression in brain-machine/computer interfaces based on the Walsh–Hadamard transform," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 1, pp. 129–137, Feb. 2014.

[24] V. Majidzadeh, A. Schmid, and Y. Leblebici, "A 16-channel, 359 μW, parallel neural recording system using Walsh-Hadamard coding," in *Proc. Custom Integr. Circuits Conf. (CICC)*, Sep. 2013, pp. 1–4.

[25] C. Aprile, J. Wüthrich, L. Baldassarre, Y. Leblebici, and V. Cevher, "DCT learning-based hardware design for neural signal acquisition systems," in *Proc. Comput. Frontiers Conf.*, 2017, pp. 391–394.

[26] J. L. Bohorquez, A. P. Chandrakasan, and J. L. Dawson, "A 350 μW CMOS MSK transmitter and 400 μW OOK super-regenerative receiver for medical implant communications," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1248–1259, Apr. 2009.

[27] *Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems*, Federal Commun. Commission, Washington, DC, USA, Apr. 2002.

[28] H. Bahrami, S. A. Mirbozorgi, R. Ameli, L. A. Rusch, and B. Gosselin, "Flexible, polarization-diverse UWB antennas for implantable neural recording systems," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 38–48, Feb. 2016.

[29] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljačić, "Wireless power transfer via strongly coupled magnetic resonances," *Science*, vol. 317, no. 5834, pp. 83–86, 2007.

[30] G. Yilmaz, O. Atasoy, and C. Dehollain, "Wireless energy and data transfer for *in-vivo* epileptic focus localization," *IEEE Sensors J.*, vol. 13, no. 11, pp. 4172–4179, Nov. 2013.

[31] A. K. Ram Rakhyani, S. Mirabbasi, and M. Chiao, "Design and optimization of resonance-based efficient wireless power delivery systems for biomedical implants," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 1, pp. 48–63, Feb. 2011.

[32] *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 KHz to 300 GHz*, IEEE Standard C95.1-2005 (Revision IEEE Standard C95.1-1991), Apr. 2006, pp. 1–238.

[33] P. Vaillancourt, A. Djemouai, J. F. Harvey, and M. Sawan, "EM radiation behavior upon biological tissues in a radio-frequency power transfer link for a cortical visual implant," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 6, Oct./Nov. 1997, pp. 2499–2502.

[34] K. M. Silay, C. Dehollain, and M. Declercq, "Numerical analysis of temperature elevation in the head due to power dissipation in a cortical implant," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2008, pp. 951–956.

[35] K. M. Silay, C. Dehollain, and M. Declercq, "Inductive power link for a wireless cortical implant with two-body packaging," *IEEE Sensors J.*, vol. 11, no. 11, pp. 2825–2833, Nov. 2011.

[36] S. J. Xilinx, "Virtex-5 LX FPGA ML501 evaluation platform," Tech. Rep. UG226 (v1.4), Aug. 2009.

Cosimo Aprile (S'16) received the B.S. degree in electronic engineering from the Politecnico di Torino, Italy, in 2010, and the joint M.S. degree in micro and nanotechnologies for ICTs from the Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland, the Institut National Polytechnique de Grenoble, and the Politecnico di Torino in 2012. He is currently pursuing the Ph.D. degree in electrical engineering with EPFL. Since 2013, he has been with the Laboratory for Information and Inference Systems and the Microelectronic Systems Laboratory, EPFL. His research interests include high speed analog circuit design, signal processing, and analog/mixed-signal integrated circuits design for optimized information extraction from signals or data volumes.

Kerim Ture received the B.S. degrees in electrical and electronics engineering and physics from Bogazici University, Istanbul, Turkey, in 2012, and the M.Sc. degree in electrical and electronics engineering from the Swiss Federal Institute of Technology, Lausanne (EPFL), Lausanne, Switzerland, in 2014, where he is currently pursuing the Ph.D. degree in wireless power transmission and radio frequency communication with the RFIC Research Group. His research interests include low-power analog and RF CMOS integrated circuit design for wireless sensor systems and biomedical applications.

Luca Baldassarre received the master's degree in physics and the Ph.D. degree in machine learning from the University of Genoa, Italy. He was with the Swiss Federal Institute of Technology under the supervision of Prof. V. Cevher on new models for compressive sensing and signal processing, and the development of new optimization algorithms. Here, he filed a patent on a machine learning-based approach for improving the scan-time and reconstruction quality of several imaging methods. He was also with the Department of Computer Science, University College London, where he conducted a Post-Doctoral Research under the supervision of Prof. M. Pontil. He has been leading data science efforts at Gamaya since 2016, a precision agriculture start-up that leverages hyper-spectral imagery, drones and satellites, to help farmers better understand their crops. He published papers in machine learning conferences and journals, with theoretical and computational analyses of new algorithms and a novel application of brain decoding based on fMRI neuroimaging.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Mahsa Shoaran received the B.Sc. and M.Sc. degrees from the Sharif University of Technology in 2008 and 2010, respectively, and the Ph.D. degree from the Swiss Federal Institute of Technology (EPFL) in 2015. She was a Post-Doctoral Fellow with the California Institute of Technology. She is currently an Assistant Professor with the School of Electrical and Computer Engineering, Cornell University, and the Director of the Cornell Neuroengineering Laboratory. Her main research interests include low-power circuit and system design for biomedical applications, brain–computer interfaces, compressive sensing, embedded classification and machine learning, and neuromodulation therapies for neurological disorders. She is a recipient of both Early and Advanced Swiss National Science Foundation Postdoctoral Fellowships. She was named a Rising Star in EE/CS by MIT in 2015.

Gürkan Yilmaz received the B.Sc. and M.Sc. degrees from the Electrical and Electronics Engineering Department, Middle East Technical University, Ankara, Turkey, in 2008 and 2010, respectively. He is currently pursuing the Ph.D. degree in microelectronics and microsystems from the Swiss Federal Institute of Technology, Lausanne (EPFL), Lausanne, Switzerland. He continued his research activities as a Post-Doctoral Researcher with the RFIC Group, EPFL, until 2017. He is also a Senior R&D Engineer with Medical Device Technologies, CSEM SA, Switzerland.

Franco Maloberti (A'84–SM'87–F'96–LF'15) is currently an Emeritus Professor with the University of Pavia, Pavia, Italy. He is also the Chairman of the Academic Committee of the Microelectronics Key-Lab of Macau. He has authored or co-authored over 550 published papers in journals or conference proceedings and four books. He holds 34 patents. He was a recipient of the XII Pedriali Prize for his technical and scientific contributions to national industrial production in 1992. He was a co-recipient of the 1996 Institute of Electrical Engineers Fleming Premium, the Best Paper Award at ESSCIRC-2007, and the Best Paper Award at the IEEJ Analog Workshop in 2007 and 2010. He received the 1999 IEEE Circuits and Systems (CAS) Society Meritorious Service Award, the 2000 CAS Society Golden Jubilee Medal, the 2000 IEEE Millennium Medal, and the IEEE CAS Society 2013 Mac Van Valkenburg Award. In 2009, he received the title of Honorary Professor of the University of Macau. He was the President of the IEEE Sensor Council from 2002 to 2003, and the Vice President, Region 8, of the IEEE CAS Society from 1995 to 1997. He is the past President of the IEEE Circuits and Systems Society. He was an Associate Editor of the IEEE TCAS-II. He was serving on the VP Publications of the IEEE CAS Society from 2007 to 2008. He was a Distinguished Lecturer of the IEEE Solid State Circuits Society from 2009 to 2010 and a Distinguished Lecturer of the Circuits and Systems Society from 2012 to 2013.

Catherine Dehollain (M'93) received the master's degree in electrical engineering and the Ph.D. degree from the Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland, in 1982 and 1995, respectively. From 1982 to 1984, she was a Research Assistant with the Electronics Laboratories, EPFL. In 1984, she joined the Motorola European Center for Research and Development, Geneva, Switzerland, where she designed integrated circuits applied to telecommunications. In 1990, she joined EPFL as a Senior Assistant with the Chaire des Circuits et Systemes, where she was involved in impedance broadband matching. Since 1995, she has been responsible for the EPFL-RFIC Group for RF activities. She has been the technical project manager of the European projects, Swiss CTI projects, and the Swiss National Science Foundation projects dedicated to mobile phones, RF wireless micropower sensor networks, and biomedical applications. Since 1998, she has been a Lecturer at EPFL in the area of RF circuits, electric filters, and CMOS analog circuits. From 2006 to 2014, she was a Maitre d'Enseignement et de Recherche, EPFL. Since 2014, she has been Adjunct Professor with EPFL. She has authored or co-authored five scientific books and 160 scientific publications. Her research interests include low-power analog circuits, biomedical remotely powered sensors, and electric filters.

Yusuf Leblebici (M'90–SM'98–F'10) received the B.Sc. and M.Sc. degrees in electrical engineering from Istanbul Technical University, Istanbul, Turkey, in 1984 and 1986, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1990. Since 2002, he has been a Chair Professor with the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, and the Director of the Microelectronic Systems Laboratory, Lausanne. He has co-authored six textbooks, and over 300 articles published in various journals and conferences. His research interests include the design of high-speed CMOS digital and mixed-signal integrated circuits, computer-aided design of VLSI systems, intelligent sensor interfaces, modeling and simulation of semiconductor devices, and VLSI reliability analysis. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATED SYSTEMS. He has been elected as a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2010 to 2011.

Volkan Cevher received the B.Sc. (valedictorian) degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1999, and the Ph.D. in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2005. He was a Research Scientist with the University of Maryland at College Park, College Park, from 2006 to 2007, and also with Rice University, Houston, TX, USA, from 2008 to 2009. He is currently an Associate Professor with the Swiss Federal Institute of Technology, Lausanne, and a Faculty Fellow with the Electrical and Computer Engineering Department, Rice University. His research interests include signal processing theory, machine learning, convex optimization, and information theory. He was a recipient of the IEEE Signal Processing Society Best Paper Award in 2016, the Best Paper Award at CAMSAP in 2015, the Best Paper Award at SPARS in 2009, and an ERC CG in 2016 and an ERC StG in 2011.