Compact Low-Power Cortical Recording Architecture for Compressive Multichannel Data Acquisition

Mahsa Shoaran, *Student Member, IEEE*, Mahdad Hosseini Kamal, *Student Member, IEEE*, Claudio Pollo, Pierre Vandergheynst, and Alexandre Schmid, *Member, IEEE*

***Abstract*—This paper introduces an area- and power-efficient approach for compressive recording of cortical signals used in an implantable system prior to transmission. Recent research on compressive sensing has shown promising results for sub-Nyquist sampling of sparse biological signals. Still, any large-scale implementation of this technique faces critical issues caused by the increased hardware intensity. The cost of implementing compressive sensing in a multichannel system in terms of area usage can be significantly higher than a conventional data acquisition system without compression. To tackle this issue, a new multichannel compressive sensing scheme which exploits the spatial sparsity of the signals recorded from the electrodes of the sensor array is proposed. The analysis shows that using this method, the power efficiency is preserved to a great extent while the area overhead is significantly reduced resulting in an improved power-area product. The proposed circuit architecture is implemented in a UMC 0.18 μm CMOS technology. Extensive performance analysis and design optimization has been done resulting in a low-noise, compact and power-efficient implementation. The results of simulations and subsequent reconstructions show the possibility of recovering fourfold compressed intracranial EEG signals with an SNR as high as 21.8 dB, while consuming 10.5 μW of power within an effective area of 250 μm $\times$ 250 μm per channel.**

***Index Terms*—Compressive sensing, cortical signals, multichannel recording, reconstruction.**

I. Introduction

WIRELESS monitoring of brain activity through implantable devices is a promising technology enabling advanced and cost-effective diagnosis and treatment of brain disorders such as stroke, Parkinson's disease, depression and epilepsy [1]–[3]. Recording from multiple sites, however, introduces a major technological bottleneck as the large bandwidth requirement for data telemetry which is not easily achievable by state-of-the-art wireless technology. The increased power consumption of transmission for large recording arrays can cause major safety and biocompatibility concerns regarding the applicability of such devices. Thus, some type of data reduction prior to telemetry is needed to meet the requirements of an implantable device.

Compressive sensing (CS) [4], [5] is an emerging compression method with interesting advantages over traditional methods thanks to its low encoder complexity and universality with respect to the signal model. A compressive sensing system samples a high-dimensional signal into a smaller number of linear measurements than dictated by the Nyquist sampling theorem. CS has been recently studied in the context of biological signals (e.g., ECG [6], [7], EEG [8] and iEEG [9]) to tackle the data rate issue. When compared to thresholding and activity-dependent recording, CS has the advantage of preserving the temporal information and morphology of the signal for the entire recording period. It is also possible to apply CS along with other methods (such as interpacket redundancy removal, Huffman coding [6] or dynamic power management of the front-end LNA) in order to further relax the stringent energy and bandwidth requirements of implantable system.

While the majority of research presented in literature focus on power minimization of the implantable system, there is also a stringent need to minimize the circuit area in order to include the highest number of recording units into the available die area. Large-scale recording of cortical activity is particularly important in the case of diseases like epilepsy which spread over wide regions of cortical area. The state-of-the-art research targeting such applications progresses toward minimally invasive flexible and dense recording arrays with high-resolution recording capability of intracranial EEG (iEEG) signals [10]–[12]. The high resolution (i.e., small spacing of recording sites) provides the capability of capturing higher frequency activity than traditionally recordable by large widely-spaced electrodes, giving a profound insight into the fundamental mechanisms underlying such abnormalities. Electrocorticographic signals recorded from human cortex can be used as an alternative to invasive spike recordings through penetrating electrodes, in order to control prosthetic limbs in BMIs as shown in [13].

The common microelectronic approach to CS ([8], [14], [15]) consists of on-the-fly compression of consecutive samples of each recording unit over time, either in analog [Fig. 1(a)] or digital [Fig. 1(b)] domain. Even though this approach results in a significant energy efficiency, its large area usage disqualifies the concept for a multichannel recording interface which should include the circuits supporting many channels in a limited die area. To overcome this issue, a new multichannel measurement scheme [Fig. 1(c)] along with an appropriate recovery scheme

Manuscript received July 10, 2013; revised October 26, 2013 and January 13, 2014; accepted January 25, 2014. Date of publication April 09, 2014; date of current version January 15, 2015. This work was supported by the Swiss NSF under Grants 200021-130166 and 200020-149742. This paper was recommended by Associate Editor R. Genov.

M. Shoaran and A. Schmid are with the Microelectronic Systems Laboratory, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland (e-mail: mahsa.shoaran@epfl.ch).

M. Hosseini Kamal and P. Vandergheynst are with the Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland.

C. Pollo is with the Neurosurgery Department, University Hospital, 3010 Bern, Switzerland.

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/TBCAS.2014.2304582

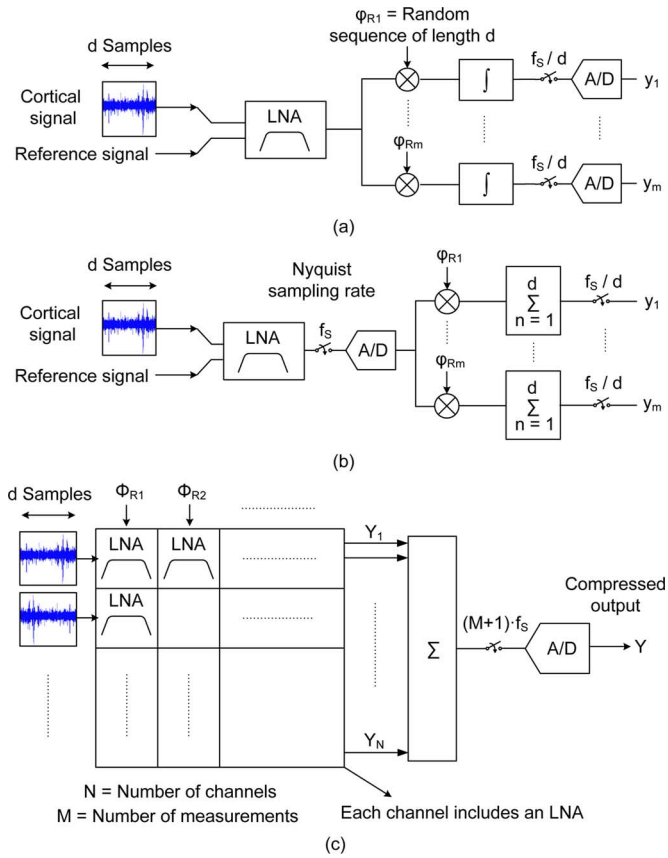

Fig. 1. Block diagram of (a) the analog single-channel CS, (b) the digital single-channel CS, and (c) the proposed multichannel CS architectures. In this design, $d = 1024$, $N = 16$ and $M = 4$.

are proposed, which encode the whole array to a single compressed data stream. In the proposed approach, the compression is carried out in the analog domain, and in a multichannel fashion. This technique circumvents the need to place one ADC per channel and results in a significant area saving. Based on this approach, a wireless monitoring system consisting of several recording/compressing units is proposed (Fig. 2). Taking benefit of the area-efficient implementation of CS, the number of recording units implantable on the cortex which satisfy the energy constraints of the system is scaled up by a factor equal to the compression ratio.

This paper is organized as follows. Section II presents a background on CS and describes the multichannel acquisition model. The circuit-level implementation of the proposed method is discussed in Section III. Experimental results of the proposed architecture are presented in Section IV and Section V concludes the paper.

II. Theoretical Model

Compressive sensing exploits the known structures in the signals to lower the required sampling ratio below the Nyquist rate, while providing a signal recovery of acceptable quality [17]. The basic concepts of CS are provided in the following discussion, emphasizing the benefit of applying CS in a neural recorder array. A detailed explanation of CS can be found in [4], [5], [18].

A. Compressive Sampling

Compressive signal acquisition prescribes sampling a signal $x \in \mathbb{R}^d$ with a significantly smaller number of samples than the signal sampled at the Nyquist rate. The linear measurements $y \in \mathbb{R}^m$, $m \ll d$ are formed by computational projection of the signal onto the measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times d}$. Compressive measurements over a defined time interval are obtained by

$$y = \mathbf{\Phi} x. \tag{1}$$

The corresponding block diagram is shown in Fig. 1(a) and (b). The acquisition system is under-determined, i.e., the number of solutions for x is infinite. However, when the signal x is known to be sparse in a basis $\mathbf{\Psi} \in \mathbb{R}^{d \times d}$ in which $x = \mathbf{\Psi}\alpha$ can be sparsely represented, the sparsest representation of x is the solution to the problem. A signal is said to be sparse in a basis if it can be represented by S non-zero coefficients, i.e., $S \ll d$.

To ensure a successful recovery of x, the linear map $\mathbf{\Phi}$ should satisfy the Restricted Isometry Property (RIP) [19]. The measurement matrix satisfies the RIP for an S-sparse signal x (S non-zero coefficients) and a small restricted isometry constant $\delta_S < 1$ if

$$(1 - \delta_S) \|x\|_2^2 \leq \|\mathbf{\Phi} x\|_2^2 \leq (1 + \delta_S) \|x\|_2^2 . \tag{2}$$

$\|x\|_2$ represents the ℓ_2 norm of x, which is defined as $\|x\|_2 = (\sum_i |x_i|^2)^{1/2}$. Sampling matrices with elements generated independently and according to a subgaussian distribution, such as the independent and identically distributed (i.i.d) Gaussian or the Bernoulli/Rademacher distributions (random 0 and 1), satisfy the RIP.

B. Multichannel Acquisition Model

In the case of multichannel neural recording, measurement limitations such as die area and power consumption suggest the use of a multichannel compression technique rather than acquiring each channel separately. Therefore, it is important to consider a measurement scheme which fulfills the physical constraints of the system. Let $\mathbf{X} \in \mathbb{R}^{d \times N}$ represent the multichannel iEEG signal where d is the dimension of the signal in each channel and in a defined time-window called *compression block* and N is the number of channels. We define a reshaping operator $\mathcal{P} : \mathbb{R}^{d \times N} \rightarrow \mathbb{R}^{d \cdot N}$ which transposes the input matrix and vectorizes the resulting matrix by concatenating its columns after each other. The linear compressive measurements are obtained by acquiring $M = p/d$ measurements from columns of $\mathbf{X}$, i.e., M measurement at each time-sample from all channels, where $p \ll d \times N$ is the total number of measurements. Hence, the multichannel linear map can be represented in matrix form $\mathbf{\Phi_{MC}} \in \mathbb{R}^{(Md) \times (dN)}$ as follows:

$$\mathbf{\Phi}_{\mathrm{MC}} = \begin{bmatrix} \mathbf{\Phi}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{\Phi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{\Phi}_d \end{bmatrix},$$

$$\mathbf{\Phi}_i = \begin{bmatrix} \phi_i^{1,1} & \cdots & \phi_i^{1,N} \\ \phi_i^{2,1} & \cdots & \phi_i^{2,N} \\ \vdots & \ddots & \vdots \\ \phi_i^{M,1} & \cdots & \phi_i^{M,N} \end{bmatrix} \tag{3}$$

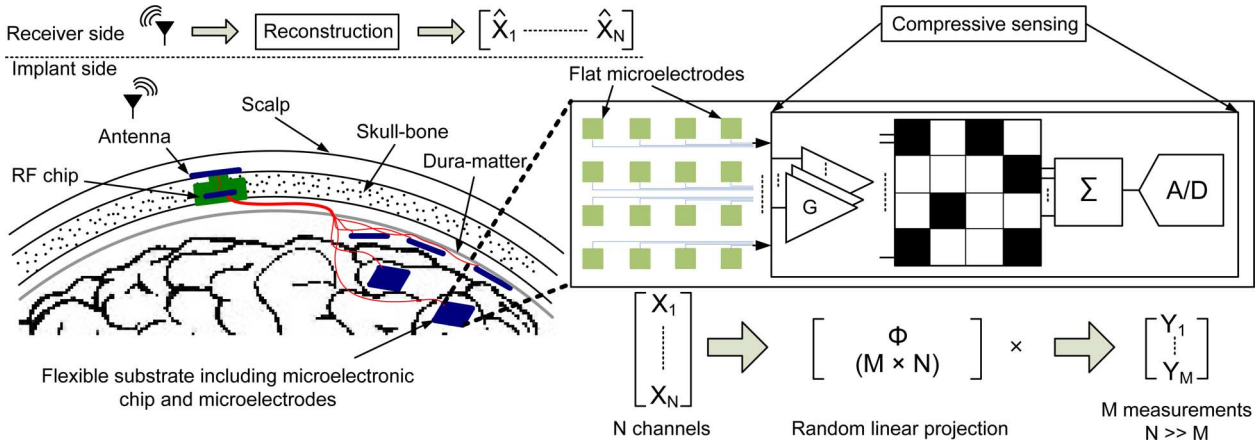

Fig. 2. System-level view of the proposed multichannel compressive sensing method used in an implantable neural recording interface [16]. The neural signals sensed at the electrode sites are amplified, randomly projected, summed up and digitized through a single on-chip ADC. An RF unit placed within a burr hole in the skull transmits the compressed and digitized data originating from several recording chips to an external receiver and powers the implanted system.

where $\mathbf{\Phi}_i \in \mathbb{R}^{M\times N}$ and $\phi_i^{k,l}$ is uniformly selected from $\{0,1\}$ to approximate a measurement matrix similar to the Bernoulli matrix. The multichannel measurement vector $Y \in \mathbb{R}^p$ is defined as

$$Y = \mathbf{\Phi}_{\mathrm{MC}}\, \mathcal{P}(\mathbf{X}). \tag{4}$$

C. Sparse Recovery

The recovery of the multichannel iEEG signal from the compressive measurements explicitly employs a multichannel sparsity domain $\mathbf{\Psi}_{\mathrm{MC}}$. The underlying neural signals in channels share similar structures. Hence, the transform domain of the multichannel signal is a block-diagonal matrix which presents the sparsity domain of channels along the diagonal and is defined as

$$\mathbf{\Psi}_{\mathrm{MC}} = \mathbf{I}_N \otimes \mathbf{\Psi}. \tag{5}$$

$\mathbf{\Psi} \in \mathbb{R}^{k\times d}$ is the sparsity domain of each channel, $\mathbf{I}_N$ is the identity matrix of size $N \times N$, and $\otimes$ represents the Kronecker product. The compressive sensing scheme [4], [5] attempts to recover the sparsest solution to $\mathbf{X}$ from the measurements Y using the following convex minimization:

$$\underset{\mathbf{X}\in\mathbb{R}^{\mathbf{d}\times\mathbf{N}}}{\operatorname{argmin}} \left\|\mathbf{\Psi}_{\mathrm{MC}}^T \mathbf{X}_{\mathrm{vec}}\right\|_1 \text{ s.t. } Y = \mathbf{\Phi}_{\mathrm{MC}}\, \mathcal{P}(\mathbf{X}) \tag{6}$$

where $\mathbf{X}_{\mathrm{vec}}$ is the vector form of $\mathbf{X}$ which includes the concatenation of its columns. The ℓ_1 norm of a vector ξ is defined as $\|\xi\|_1 = \sum_i |\xi_i|$. The measurement consistency in the recovery algorithm ($Y = \mathbf{\Phi}_{\mathrm{MC}}\, \mathcal{P}(\mathbf{X})$) pertains to obtaining a signal $\mathbf{X}$ which is in accordance with the measurements $\mathbf{Y}$ through the measurement matrix $\mathbf{\Phi}_{\mathrm{MC}}$, i.e., (6) recovers a signal ($\mathbf{X}$) which is sparse in the sparsity domain ($\mathbf{\Psi}_{\mathrm{MC}}$) and satisfies the measurement consistency constraint.

D. Mixed Norm Recovery

The multichannel neural signals have high inter-channel dependency, as the signals recorded by the adjacent channels are delayed or scaled version of each other, depending on the spatial resolution and pitch of the electrodes which indicates the propagation of neural activity within the brain. The dependent structure of multichannel neural signals suggests the design of a recovery model which exploits the similarity of neural signals. The Gabor coefficients [20] of iEEG signals recorded by adjacent channels in a Gaussian window are shown in Fig. 3(a). The Gabor coefficients are observed to follow similar activity among neural channels for each frequency.

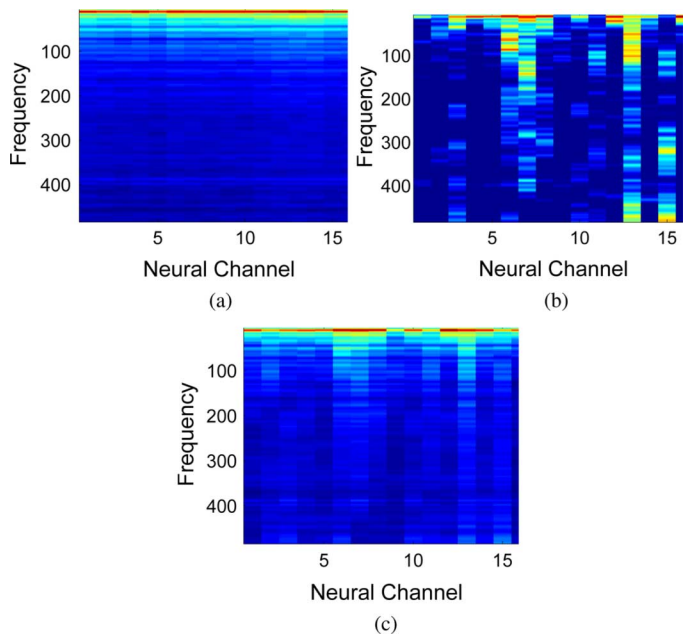

Fig. 3. The structure of Gabor coefficients of multichannel neural signals in a Gabor window. (a) Original neural signals: the Gabor coefficients represent group structure along frequencies. (b) ℓ_1 norm: the Gabor coefficients are scattered and recovery does not respect the group structure. (c) $\ell_{1,2}$ mixed norm: the joint recovery preserves the group structure of Gabor coefficients.

Sparse recovery induces coefficient-wise sparsity without considering the inter-channel correlation of the neural signals. The neural recovery model should employ the cross correlations of iEEG signals to improve the reconstruction quality. An appropriate model for multichannel neural signals should highlight the group structure of Gabor coefficients, i.e., the model should lead to a sparsity on the number of active frequencies

and promote similar activity on the neural channels for the selected frequency.

We model the dependency of neural signals using the $\ell_{1,2}$ mixed norm [21]. The $\ell_{1,2}$ mixed norm of a given vector ξ is defined as

$$\|\xi\|_{1,2} = \sum_{g} \left(\sum_{m} |\xi_{g,m}|^2 \right)^{1/2} \tag{7}$$

where ξ is divided into a set of non-overlapping groups. The elements of ξ are indexed by a pair (g, m), where g defines the group number and m represents the corresponding index inside the group. It is important to differentiate between the $\ell_{1,2}$ mixed norm and the sparsity inducing ℓ_1 norm. In the former, a group of coefficients are discarded or retained together, since the same threshold is applied to the ℓ_2 norm of each group. In the latter, each coefficient is shrunk independently from other elements and the same threshold is applied to each element. Consequently, the ℓ_1 recovery does not model the block structure in the neural signals. However, the $\ell_{1,2}$ norm respects the group structure of neural signals and imposes sparsity on the group of coefficients rather than each coefficient independently. Thus, the $\ell_{1,2}$ scheme results in recovery of dense blocks for sparse number of frequencies. Fig. 3 compares the recovered Gabor coefficients of multichannel neural signal employing sparse and joint recovery in a Gabor window, with the Gabor coefficients of the original neural signal. We observe that the recovered Gabor coefficients using $\ell_{1,2}$ [Fig. 3(c)] yield the same structure as of the original multichannel neural signals [Fig. 3(a)]. Furthermore, the sparse recovery [Fig. 3(b)] does not respect the group structure of neural signals and results in recovery of Gabor coefficients which are sparse and independently spread along different frequencies for each neural channel. This behaviour is explained by the fact that the sparse recovery does not consider the group structure of neural signals. The solution to the multichannel neural recovery using the $\ell_{1,2}$ mixed norm is obtained by replacing the ℓ_1 norm by the mixed norm as

$$\mathbf{\Psi}_{\mathrm{MC}} \underset{\alpha_{\mathrm{MC}} \in \mathbb{R}^{\mathbf{d}\cdot\mathbf{N}}}{\operatorname{argmin}} \|\alpha_{\mathrm{MC}}\|_{1,2} \text{ s.t. } Y = \mathbf{\Phi}_{\mathrm{MC}}\, \mathcal{P}(\mathbf{\Psi}_{\mathrm{MC}}\, \alpha_{\mathrm{MC}}). \tag{8}$$

III. Circuit Implementation

In order to realize the CS acquisition scheme proposed in Section II, the architecture shown in Fig. 4 is designed and implemented in a 1P6M 0.18 $\boldsymbol{\mu}$m CMOS technology. The main focus of this design consists of accommodating a large number of recording units into the available die area, while preserving sufficiently low-noise and low-power performance. Low power consumption and compact area are crucial in high-density implantable recording systems.

A. Recording Channels

The fabricated integrated circuit includes 16 recording channels. Each channel consists of a low-noise amplifier with a band-pass transfer function, an additional low-pass filter to limit the high cut-off frequency, a second gain stage, and a buffered sample-and-hold circuit. The differential outputs of all channels individually connect to the summing stage and randomly accumulate at the output of this stage. The result is then digitized through a single ADC.

A three-stage configuration is used in each channel [22], in order to minimize the total die area and provide the desired amplification and filtering of the input signal.

1) Low-Noise Amplifier: An area-efficient T network-based capacitive feedback amplifier [23] with moderate-sized input capacitors is used as the front-end gain stage G_1, which provides a mid-band gain of 29.8 dB [Fig. 5(a)]. The mid-band gain of this stage is realized by multiplying two capacitive ratios. This potentially provides a high closed-loop gain in a single stage. Using this topology, the total size of the capacitors is smaller than the size of capacitors in a conventional capacitive feedback topology [24]. Four back-to-back MOS devices biased in subthreshold region are used to implement the high-value feedback resistors with sufficient linearity for creating a low-frequency high-pass pole [Fig. 4(a)]. The variation on the high-value resistance over the voltage swing across the resistor is 1.5% at 200 mV and remains smaller than 2% up to a voltage swing of 600 mV across the resistor. Owing to the series connection of the high-value resistors in a symmetric combination, the equivalent resistance exhibits symmetric variations around the quiescent point in all process corners. The simulated high-pass cut-off frequency varies in the range of 14–118 Hz, under the worst-case process corner, supply voltage and temperature variations.

The differential mid-band gain of this stage is calculated as

$$A_M = -\left(\frac{C_1}{C_2}\right)\left(\frac{2C_4 + C_2 + C_3}{C_3}\right). \tag{9}$$

In general, the smaller the input capacitance in the feedback loop is, the larger the ratio of the OTA noise referred to the input of the LNA will be, since the OTA's input parasitics become comparable to C_1. To tackle this issue, the additional gain provided by the second capacitive ratio is selected small compared to the first term in (9) to satisfy the low-noise operation as well. In this design, $C_2 = C_3 = 200$ fF, $C_1 = 1.4$ pF and $C_4 = 400$ fF, resulting in $C_{\mathrm{tot}} = 4$ pF. The conventional topology [24] requires a total capacitance of 64×200 fF ≈ 12.8 pF to achieve a similar mid-band gain.

A second benefit of replacing the very large input capacitance by a moderate one is that the larger input impedance provided by a moderate input capacitor reduces the effect of attenuation of cortical signals in the electrode-channel interface.

A folded-cascode OTA with a continuous-time common-mode feedback (CMFB) circuit is implemented in the LNA [Fig. 4(a)]. Through the dc path provided by the large-value feedback resistors, the same CM voltage as V_{CM} is generated at the input gates of the OTAs used in G_1 and G_2. The total bias current of the OTA is 7 μA. The simulated passband of the LNA ranges from 28 Hz to 144 kHz. The linearity performance of the LNA is described by the THD metrics which is 0.24% for a 2 mV_{p-p} input signal. This LNA consumes 8.4 μW of power, corresponding to 89% of the total power consumed by the channel.

The ac-coupled architecture provided by the capacitive feedback topology enables the amplifier to reject the large dc offsets (as large as 1–2 V [24]) which are commonly generated in

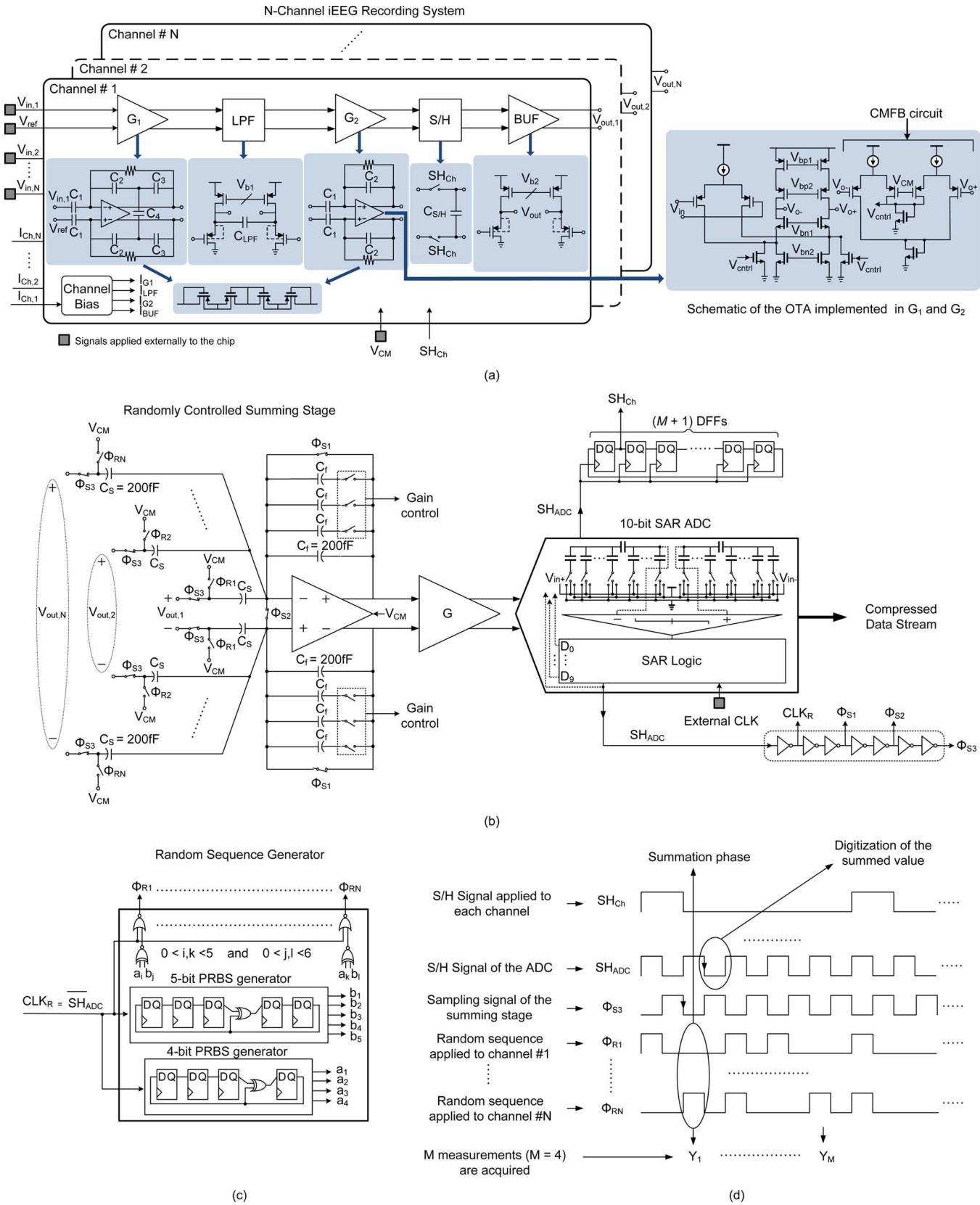

Fig. 4. Circuit implementation of the proposed multichannel CS architecture. (a) Individual channels. (b) Variable-gain summing stage, gain stage and 10-bit SAR ADC with attenuated binary-weighted DAC. (c) Multi-output random sequence generator. (d) Timing diagram of the proposed system.

an electrode-tissue interface. The common centroid layout of the input differential pair in the LNA and the passive components such as input and feedback capacitors results in improved matching and offset performances, which are mainly limited by

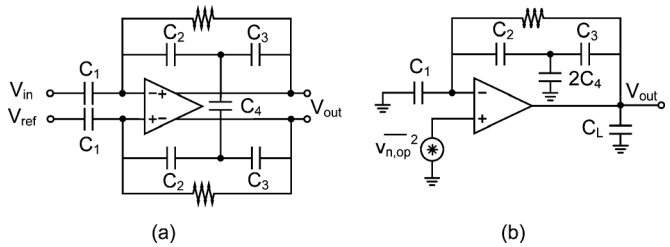

Fig. 5. (a) Compact low-noise differential amplifier. (b) Half-circuit equivalent used for noise analysis.

the front-end LNA. The measured input-referred offset of the stand-alone channel is 0.85 μV.

2) Filter, Amplifier and S/H: The commonly used low bandwidth of interest (less than 2 kHz) for capturing iEEG signals (e.g., for epileptic activity detection) requires additional low-pass filtering after the LNA. Limiting the bandwidth using a low-pass filter stage with minimal power and noise costs is more beneficial to limit the die area per channel than aggressively increasing the C_L of the LNA. A first-order source-follower based low-pass filter (adapted from [25]) is used which achieves sufficient linearity at a small bias current [Fig. 4(a)]. The DC-gain loss due to the bulk transconductance which is inherent to the source-follower architecture is reduced by a source-to-bulk connection. The internal feedback of the source-follower circuit and processing the signal in voltage domain increases the linearity range of the filter.

The presented filter exhibits a THD of 0.08% (equal to 10–11 bits of linearity) for a filter input signal of 60 mV_{p-p} (or 2 mV_{p-p} at the input of a channel) consuming a bias current of 5.4 nA. The cut-off frequency is 1.9 kHz and is tunable by the current.

A second gain stage [G_2 in Fig. 4(a)] provides an additional gain of $A_{M2} = 13.64$ dB using a conventional capacitive feedback architecture. The power consumption of the second gain stage is negligible (0.72 μW) compared to the front-end LNA, thanks to the relaxed noise requirements. Using a folded-cascode OTA architecture, an output voltage swing as large as 200 mV_{p-p} is achieved in the second gain stage. The size of the transistors in this OTA are selected smaller compared to the OTA used in the front-end LNA, thanks to the relaxed noise and matching requirements.

The amplified and band-pass filtered signal of each channel goes through a sample-and-hold (S/H) stage followed by a source-follower buffer to keep the signal constant during the consecutive M measurements ($M = N/\text{CR}$ in this case). The S/H circuit is controlled by the signal SH_{Ch} which is $M+1$ times slower than the SH signal of the ADC and is generated using a single $(M+1)$-bit shift register [Fig. 4(b)]. The output common-mode voltages of the LPF and BUF are defined by the gate-source voltage of the PMOS transistors which are located at their input, knowing that the output bias voltage of the preceding stages are defined through the CMFB of G_1 and G_2. The source-to-bulk connected source-follower buffers at the output of each channel exhibit a THD of 0.12%, which equals to 9–10 bits of linearity for a 200 mV_{p-p} input signal. The total power consumption of the LPF, buffer and in-channel biasing circuitry is less than 0.3 μW.

The required settling time to guarantee B-bit performance imposes the following condition at the output of the second gain stage, prior to the S/H:

$$e^{-(\frac{1/2f_S}{\tau})} < 0.5 \times 2^{-B} \tag{10}$$

for $B = 8$ and $f_S = 4$ kS/s this condition results in

$$f_{-3\,\text{dB}} = 1/(2\pi\tau) > \frac{(B+1)\ln 2}{\pi} f_s \approx 1.98 f_s \approx 8\,\text{kHz} \tag{11}$$

in which $f_{-3\,\text{dB}}$ represents the bandwidth of the circuit when the sampling capacitor is connected. Assuming a sampling capacitor of 0.4 pF, the achieved bandwidth is 55 kHz, consuming a low bias current of 0.6 μA.

B. Randomly Controlled Summing Stage

The sampled signals of N channels of the array connect to the summing stage at the sampling phase (ϕ_{S3}) which follows the two in-phase events ϕ_{S1} and ϕ_{S2}. By applying a proper timing strategy (ϕ_{S1} comes before ϕ_{S2} and ϕ_{S2} before ϕ_{S3}) as presented in Fig. 4(b) and using bottom-plate sampling, it is possible to substantially reduce the effect of channel charge injection of the switches. The effect of clock feedthrough is reduced by the differential implementation. A variable voltage gain is achieved through the controlling switches in series with the feedback capacitors. Due to the randomized summation at this stage, the programmability of the gain is crucial. The gain stage preceding the ADC provides an additional gain of three to perfectly accommodate the full-scale input range of the ADC. A hybrid two-stage class A/AB topology [28] is used as the OTA in this stage, which provides the desired rail-to-rail output swing.

The operation of the summing stage is as follows. In sampling mode, ϕ_{S1}, ϕ_{S2} and ϕ_{S3} are on, allowing the differential voltage across the two sampling capacitors (C_S) at the output of each channel to track the differential output voltage of that channel. In summation mode (ϕ_{S1}, ϕ_{S2} and ϕ_{S3} are off), the charge stored on the sampling capacitors of those channels with random value equal to one are transferred to the capacitors in the feedback path (C_f). This enables the summation of the sampled values of channels, based on the value of the corresponding random sequence.

A reasonable lower bound for the bandwidth of OTA used in the summing stage can be calculated from (10) which equals 50 kHz at a sampling rate of $(M+1)f_S = 20$ kS/s. The stability and speed of the feedback loop is guaranteed for the worst case where all random values are equal to one, while a single feedback capacitor is connected which corresponds to $A_{M,\text{INT}} = N$, for equal unit capacitors and equal inputs.

1) System Resolution: Assuming the worst-case situation in which all N channels of the array are summed together while all controlling random values are equal to one, the following constraints on the signal headroom and noise prior to the ADC will exist:

$$NV_{\text{in,sig}}A_M \leq 2VDD \tag{12}$$

$$NV_{n,\text{rms}}^2 A_M^2 \leq (2VDD)^2/12 \cdot 2^{2By} \tag{13}$$

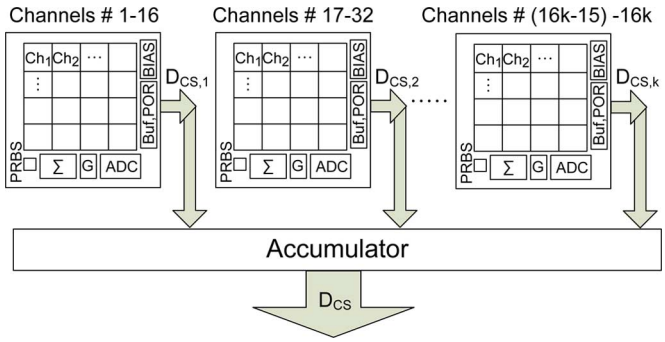

Fig. 6. Proposed architecture for realizing a high-density recording system. Each block of this figure corresponds to one of the implantable chip-electrode combinations of Fig. 2 (excluding the electrodes).

where A_M is the total mid-band gain from the input of the channels to the input of the ADC, $V_{\text{in,sig}}$ is the peak-to-peak amplitude of the input signal, $V_{n,\text{rms}}$ is the rms value of the input-referred noise of each channel and N is the number of channels.

Having $V_{\text{in,rms}} = V_{\text{in,sig}}/2\sqrt{2}$ for a sinusoidal waveform, results in $B_y \approx B_{\text{sig}} + \log_2 \sqrt{N} - 0.58$. Generalizing this equation to any arbitrary waveform with $V_{\text{in,rms}} = V_{\text{in,sig}}/K$ yields

$$B_y \approx \log_2 \frac{V_{\text{in,sig}}}{V_{n,\text{rms}}} + \log_2 \sqrt{N} - 1.78 \tag{14}$$

which is independent of the parameter K. Considering a background neuronal noise of 5–10 μV_{rms} [26] and a typical amplitude of surface cortical signals of up to 1 mV and adding some extra margin, B_{sig} and B_y, are set to 8 and 10 for 16 channels.[1] The effective compression ratio of the proposed CS topology is then approximated by

$$\text{CR}_{\text{eff}} \approx \frac{NB_{\text{sig}}}{(M+1)(B_{\text{sig}} + \log_2 \sqrt{N})}. \tag{15}$$

2) Large-Scale Implementation: The imposed requirement on the ADC resolution degrades the power efficiency of the system at large number of channels and sets an upper limit for the maximum number of channels connected to the summing stage prior to the ADC. Thus, the model shown in Fig. 6 is adapted for large arrays. The digital outputs of 16-channel blocks are summed together by a digital accumulator with sufficiently large number of bits. The power and area cost of adding the required number of bits to the resolution of the accumulator is significantly less compared to the cost of increasing the resolution of the ADC. Designing an ADC for target resolutions higher than 10 bits requires additional topological modifications which are commonly achievable by consuming more power. Using this method, the proposed CS concept is still more area- and power-efficient for large arrays than single-channel approaches.

[1]This calculation excludes the rms thermal noise of the electrodes which can be in the order of 10–20 μV [27], depending on the size, material and operating temperature of the electrodes. In reality, this noise further degrades the dynamic range of the input signal and the effective number of bits achievable at the system level.

In this model, the limit on the compressed array size (N) is imposed by the acceptable power efficiency of the summing stage and ADC which are sampled at a rate of $(1 + N/\text{CR})f_S$ with f_S being the Nyquist sampling rate. The multichannel system is then divided into sub-blocks of N-channels, with N being the optimum number of channels to be compressed into a single data stream. Each sub-block is encoded to a digital data stream which after reconstruction, generates the signals originated from the channels constituting that sub-block.

3) Random Matrix Generator: In order to guarantee an efficient implementation of compressive sensing, the power and area cost of CS circuits including the measurement matrix generator must be negligible compared to the rest of the integrated circuit. In a single-channel approach, each channel is loaded with m sequences (building the rows of the measurement matrix explained in (1)), whereas in the proposed model, each channel needs to be driven by only one sequence. As shown in (3), the measurement matrix supporting the consecutive M measurements for recovering each sample of individual channels is filled by the corresponding M values of the in-channel sequences. This approach circumvents the need to place a memory to store the elements of the measurement matrix.

As shown in Fig. 4(c), multi-output sequence generation is achieved by XORing the multiple outputs of maximal-length pseudo-random binary sequence (PRBS) generators [29]. For a recording array of 4×4 and a value of M equal to 4 (resulting in $\text{CR} = 16/4$), the 16 sequences driving the individual channels are generated by XORing the states of a 4-bit PRBS generator with another 5-bit PRBS generator. As opposed to the single-channel compressing system which has to be physically designed for a specific predefined m and redesigned by varying the compression ratio, the proposed scheme is easily adaptable for different values of M by adjusting the clock frequency.

4) Noise Analysis: Power, area and noise are the three major performance metrics of an implantable system. Ideally, the noise performance of the system is limited by the background extracellular noise and by the noise of the electrodes. In addition to the low-noise amplifier at the front-end of each channel, the noise induced by the switched capacitor summing stage should be minimized. While the effect of flicker noise can be reduced by using large input devices (generally PMOS), the thermal noise induced by the switching circuits can affect the total noise of the system, prior to the ADC. A comprehensive sampled noise analysis based on an approach similar to [30] is presented to give an insight into the appropriate values of the design parameters of the summing stage. As a general noise reduction technique, the input differential pair of the folded-cascode OTA used in the front-end LNA is biased in subthreshold region [31]. This results in a large g_m/I_D of the input transistors and reduces the input-referred noise of the OTA. The mean-square value of the total thermal noise at the output of each channel can be approximated by integrating the PSD of OTA noise at the input [31], shaped by the low-pass transfer function of the loop [see the half-circuit model shown in Fig. 5(b)], and fed into the low-pass filter and second gain stage, resulting in

$$\overline{v_{n,\text{out}}^2} \approx S_{n,\text{op}} \cdot \frac{A_{M,\text{tot}}^2}{4\tau} = \frac{4kT}{\kappa g_m} \cdot \frac{A_{M,\text{tot}}^2}{4\tau} \tag{16}$$

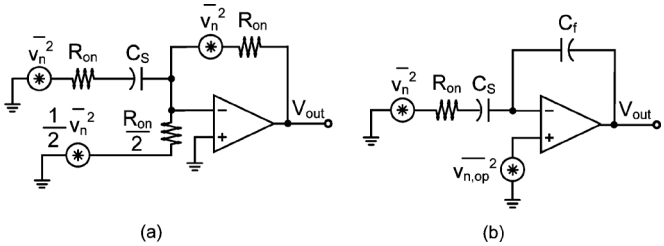

Fig. 7. Noise analysis models of the summing stage of Fig. 4 in (a) sampling phase and (b) integrating phase.

where k is Boltzmann's constant, T is the absolute temperature, κ is the reciprocal of the subthreshold slope factor, $A_{M,\mathrm{tot}} \approx 1/\beta_1\beta_2$, $1/\beta_1$ is obtained from (9) and $1/\beta_2$ is the mid-band gain of the second gain stage. Assuming $\tau_1 = C_L/\beta_1 g_m$ the settling time constant at the output of the LNA with a load capacitor of C_L and an input transconductance of g_m, and $\tau_2 = C_{\mathrm{LPF}}/g_{m,\mathrm{LPF}}$ the time constant at the output of the LPF, $\tau_2 > \tau_1$ and (16) yields

$$\overline{v_{n,\mathrm{out}}^2} \approx \frac{kT.A_{M,\mathrm{tot}}^2.g_{m,\mathrm{LPF}}}{\kappa g_m C_{\mathrm{LPF}}}. \tag{17}$$

The equivalent half-circuits of the summing stage during the sampling and integrating phases are shown in Fig. 7. The PSD of the thermal noise generated on the sampling capacitor (C_S) which is due to the switch placed in-series with C_S and during ϕ_{S3} can be written as [30]

$$S_v(f) = \frac{4kTR_{\mathrm{on}}}{1 + (2\pi f\tau_0)^2}. \tag{18}$$

Here, R_{on} is the on resistance of the switch and τ_0 is the time constant of the C_S branch during sampling (ϕ_{S3}), which is expressed as

$$\tau_0 \approx C_S \cdot \left(R_{\mathrm{on}} + \frac{R_{\mathrm{on}}}{2 + g_m R_{\mathrm{on}}}\right). \tag{19}$$

The corresponding noise integrated from 0 to infinity is equal to $kT/[C_S(1 + 1/(2 + g_m R_{\mathrm{on}})]$ which summed up to the non-significant noise terms generated by the two switches ϕ_{S1} and ϕ_{S2} can be estimated as kT/C_S.

During the second phase (ϕ_{S3} is open), the equivalent time constant of this circuit can be approximated by $\tau = C_S(R_{\mathrm{on}} + 1/g_m)$. The mean-square value of the switching noise assuming that in the worst case, ϕ_{Ri} is equal to one for $i = 1, \ldots, N$, is calculated as [Fig. 7(b)]

$$\overline{v_{n1}^2} \approx \frac{4kTR_{\mathrm{on}}}{4\tau} = \frac{kT/C_S}{1 + 1/x}. \tag{20}$$

The parameter x is introduced as $x = R_{\mathrm{on}}g_m$. The noise power in C_S due to the op-amp when ϕ_{S3} is open is found as

$$\overline{v_{n2}^2} \approx \frac{4kT/\kappa g_m}{4\tau} = \frac{kT}{\kappa C_S(1 + x)}. \tag{21}$$

Thus, the total input-referred noise of the differential summing stage can be expressed as

$$\begin{aligned}\overline{v_{n,\mathrm{in}}^2} &\approx \frac{2kT}{C_S} + \frac{2kT/C_S}{1 + 1/x} + \frac{2kT}{\kappa C_S(1 + x)} \\ &= \frac{kT}{C_S}\left(\frac{4x + 2 + 2/\kappa}{x + 1}\right) = \frac{kT}{\tau g_m}(4x + 2 + 2/\kappa)\end{aligned} \tag{22}$$

referred to the input of each summing branch (or equivalently to the output of each channel). Based on this equation, the noise is minimized for $x \ll 1$. The total noise power calculated in (22) should be smaller than the output noise power of channels calculated by (17) which satisfies the noise requirement prior to the ADC. There is a clear trade-off between the summing stage's noise and the circuit area which is proportional to C_S. However, considering the extra margin assumed in calculating the system resolution based on (14), and accounting for the electrode and background noise, enables letting the noise due to this stage be as large as the noise of channels, keeping in mind that additional noise reduction does not improve the effective number of bits but further degrades the area efficiency of the implantable system.[2] With this assumption, a minimum value of C_S equal to 200 fF is required to guarantee the proper performance of the summing stage. For an optimal design, the maximum permissible value of τ based on (10) can be substituted into $\tau = C_S(1 + x)/g_m$ resulting in the lowest g_m which can fulfill the noise and bandwidth requirements of this stage. The assumed constraint on x (i.e., $x \ll 1$) determines the minimum size of the switches in this case. For equally-sized capacitors, these switches are selected $(M + 1)$ times larger than the in-channel S/H switches which operate at a lower frequency.

C. Analog-to-Digital Converter

The stringent area and power constraints of the implantable system motivate the compact and energy-efficient implementation of the ADC. Based on system-level requirements, an ADC with 10 bits of resolution and a sampling rate of 20 kS/s translates into a data recorded in each channel with 8 bits of resolution and a sampling rate of 4 kS/s. The SAR ADC is a popular architecture which enables low-power data conversion for medium resolution/speed applications. A binary-weighted capacitive array with attenuation capacitor [Fig. 4(b)] is used which enables the compact implementation of the ADC. The potentially small input capacitance of this array[3] results in relaxed specifications of the stage preceding the ADC, in terms of bandwidth and power consumption.

As shown in the timing diagram of Fig. 4(d), the data sampled from a channel is kept constant during M randomized summations and consecutive digitizations through the ADC. Thanks to the low bandwidth of iEEG signals, the PRBS generator and ADC are clocked at a rate faster than the sampling frequency of the individual channels. This allows the computations to be done

[2]The typical value of the in-band input-referred noise power due to the LNA is more than 10 times smaller than the electrode noise power for this application.

[3]The total required capacitance of the attenuated array is $2^{(B/2)-1}$ times smaller than a conventional binary-weighted array where B is the number of bits.

TABLE I
COMPARISON OF THE AFE AND ADC PERFORMANCE WITH PUBLISHED LITERATURE

Parameter	[38]	[39]	[37]	[40]	[41]	[42]	[23]	[43]	[44]	This Work
Technology [μm CMOS]	0.35	0.13	0.13	0.18	0.35	0.35	0.35	0.13	0.13	0.18
Bandwidth [Hz]	10-5k	0.023-11.5k	167-6.9k	10-7.2k	1-5k	217-7.8k	1-8.5k	1-5k	0.1-5k	39-1.9k
Gain [dB]	33	38.3	47.5	39.4	34	45.7	38.1	54-60	54-60	43.8
Input-Referred Noise [μV_{rms}]	6.08	1.95	3.8	3.5	7	4.43	13.3	5.1	6.5	4.2
NEF	5.55	2.48	2.16	3.35	4.6	2.16	7.87	4.4	7.2	4.12
AFE Area [mm^2]	0.020	-	< 0.080	0.062	0.020	-	0.056	0.045	< 0.090	0.040
CMRR	60	63	83	70.1	-	58	74	75	75	59.8
PSRR	-	63	-	63.8	-	40	55	-	-	-
AFE Power Consumption [μW]	8.4	12.5	1.92	7.92	4.2	1.26	6	8.5	4.5	9.4
ADC Sampling Fequency [kS/s]	111	10-100	22.5	-	-	256	-	$\leq$100	57	20
ENOB	6	6.2	7.65	-	-	8.3	-	7.6	7.8	9.2
INL [LSBs]	-	-	-0.45	-	-	1.52	-	0.7	-	-0.76
DNL [LSBs]	-	-	-0.71	-	-	0.89	-	0.6	-	-0.37
ADC Power Consumption [μW]	2.77	< 1	0.5	-	-	31.3	-	1.5	1.8	2.6

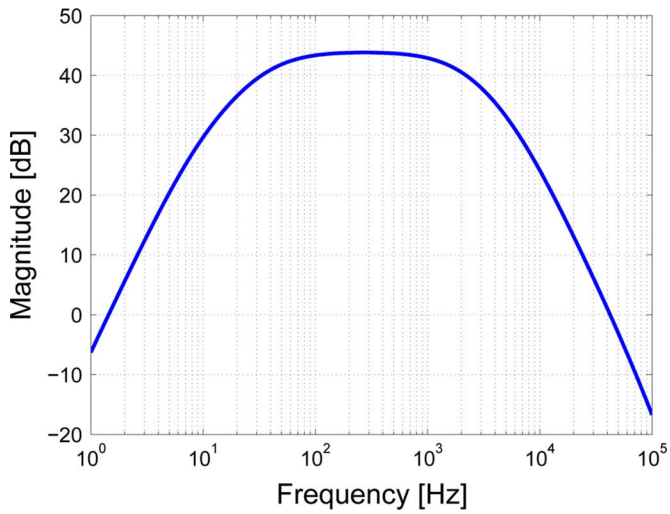

Fig. 8. Measured frequency response of the stand-alone channel.

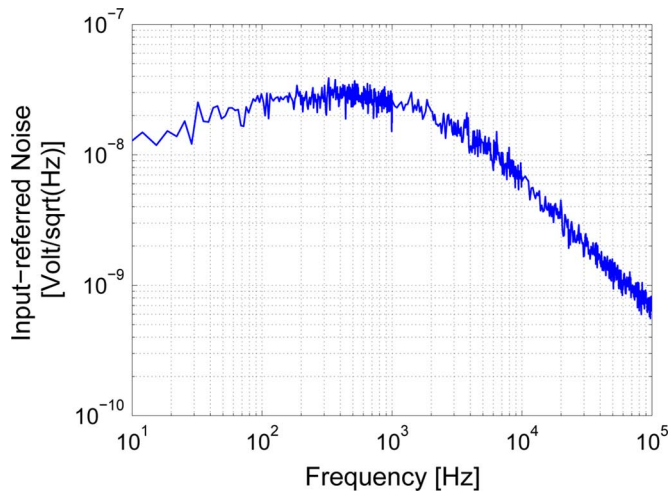

Fig. 9. Measured input-referred noise of the stand-alone channel.

within the input sampling period. All the required signals are generated using a single external clock of 400 kHz, as shown in the timing diagram of Fig. 4(d). The sample-and-hold signal of the ADC is generated at a rate equal to 1/20 of the external clock, which enables 10-bit conversion of the summed value. Digitization occurs during the half-cycle in which random sequences are zero [shown in Fig. 4(d)], thus converting the previous accumulated value. At the same time, the sampling signals of the summing stage [Φ_{S3} and also the in-phase signals Φ_{S1} and Φ_{S2} in Fig. 4(b)] are equal to one, which resets the output of the summing stage and stores the corresponding channel's output for the following accumulation phase. During the next half-cycle, the sampled values at the input branches of the integrator which are stored on the capacitors C_S, are summed together based on the value of the random sequence controlling that branch. Thus, the output of the summing stage at the rising edge of the Φ_{S3} represents the randomized summation of the channels' samples which must be converted to the digital bit stream, using the ADC.

IV. EXPERIMENTAL RESULTS

The measured frequency response at the output of G_2 and input-referred noise of the stand-alone channel are shown in Figs. 8 and 9. The high-pass pole is measured at 39 Hz, while the simulated cut-off frequency is 28 Hz. Thus, the HP-pole differs by 40% with respect to the desired value. The discrepancy between the simulated and measured high-pass pole is due to the inaccurate model of the transistors in weak inversion region, and the extra parasitics which affect the total capacitance of the feedback path. A more precise control on the high-pass cut-off frequency can be achieved by adjusting the feedback capacitor and resistor using digital words as shown in [37].

The input-referred noise density at the channel input is 25 nV/$\sqrt{\text{Hz}}$ at 1 kHz. The input-referred noise integrated over the signal bandwidth is 3.2 μV_{rms}, while it increases to 4.2 μV_{rms} when integrated from 1 Hz to 100 kHz. The Noise Efficiency Factor (NEF) of the analog front-end (including G_1, LPF and G_2) is defined as [32]

$$\text{NEF} = V_{n,\text{rms}} \cdot \sqrt{\frac{2I_{\text{tot}}}{\pi \cdot U_T \cdot 4kT \cdot \text{BW}}}. \tag{23}$$

The measured NEF is equal to 4.12 and 3.23 for the noise integration bandwidths of 1 Hz-100 kHz and passband of the signal, respectively.

Table I presents the performance summary of the analog front-end (AFE) and ADC which are compared with published works.

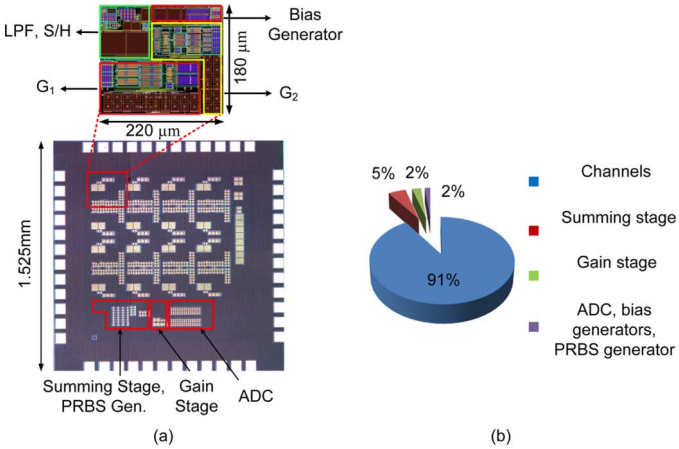

Fig. 10. (a) Microphotograph of the chip and an individual channel's layout. (b) Power breakdown of a 16-channel compressive sensing array.

TABLE II
POWER CONSUMPTION AND AREA USAGE OF INDIVIDUAL BLOCKS

Parameter	Value
$P_{Channel}$	$9.4\mu W$
$P_{SummingStage}$	$8.4\mu W$
$P_{GainStage}$	$4.2\mu W$
P_{ADC}	$2.6\mu W$
P_{PRBS}	$0.03\mu W$
$A_{Channel}$	$220\mu m \times 180\mu m$
$A_{SummingStage}$	$290\mu m \times 150\mu m$
$A_{GainStage}$	$150\mu m \times 100\mu m$
A_{ADC}	$290\mu m \times 150\mu m$
A_{PRBS}	$90\mu m \times 80\mu m$

The microphotograph of the fabricated chip is shown in Fig. 10(a). The total current consumption of the chip including the buffers and bias generators is 140 μA drawn from a 1.2 V power supply, corresponding to effective current of 8.75 μA per channel. The achieved power density of the system is 7.2 mW/cm^2, significantly below the safety limit of 80 mW/cm^2 [33] for an implantable system. The contribution of different blocks of the system to the total power consumption is shown in Fig. 10(b). The area occupation and power consumption of each block of the system are shown in Table II.

In order to demonstrate the effectiveness of the proposed acquisition model, a long segment of multichannel iEEG signal recorded from subdural strip and greed electrodes implanted on the left temporal lobe of a patient with medically refractory epilepsy have been used as the input. The signals are recorded during an invasive pre-surgical evaluation phase to pinpoint the areas of the brain involved in seizure generation and to study the feasibility of a resection surgery. This data includes minutes of pre-ictal, ictal and post-ictal activities sampled at 32 kS/s, using Neuralynx. The signals recorded by 16 adjacent channels of a greed of the electrodes are applied into the proposed CS system.

A. Recovery Performance

Many biological signals can be sparsely represented in either Gabor or wavelet domains [8]. We employ Gabor transform as the sparsity domain of neural signals for multichannel neural recovery based on sparse and mixed norm methods. The recovery

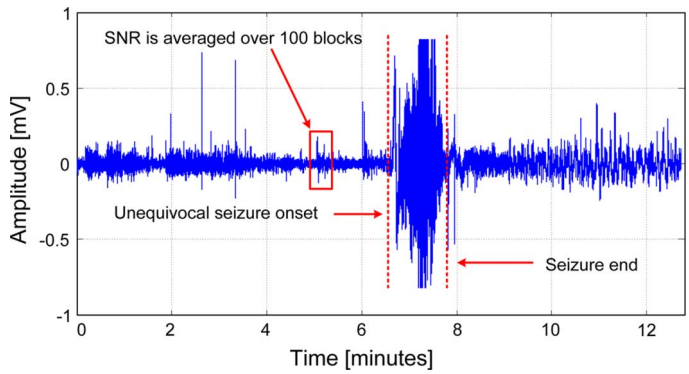

Fig. 11. One channel of human intracranial EEG recording using strip and greed electrodes implanted on the left temporal lobe. The recovery SNR is calculated by averaging over 100 blocks of signal in the low-voltage fast activity region.

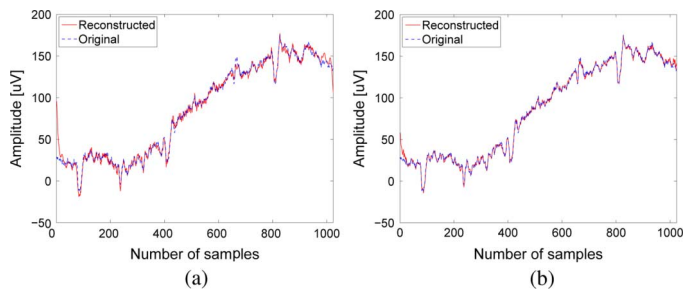

Fig. 12. Comparison of recovery performance using different reconstruction methods for a block of length 1024 and compression ratio of 4. (a) ℓ_1 recovery $\mathrm{SNR}_{\mathrm{CH1}} = 21.3$ dB. (b) $\ell_{1,2}$ recovery $\mathrm{SNR}_{\mathrm{CH1}} = 28.04$ dB.

SNR of the reconstructed signal $(\hat{x})$ with respect to the original signal (x) is calculated from

$$\mathrm{SNR} = -20\log_{10}\|x-\hat{x}\|_2/\|x\|_2 \tag{24}$$

for each recording channel (e.g., $\mathrm{SNR}_{\mathrm{CH1}}$ represents the recovery SNR of first channel). The mean SNR of 16 channels are averaged over 100 blocks of the signal, as shown in Fig. 11. The performance of the circuit is validated for low-voltage fast activities which are shown to be associated with seizure onset. The reconstructed signals versus the original signals corresponding to one block of a single channel data using ℓ_1 and $\ell_{1,2}$ recovery are shown in Fig. 12(a) and (b). The length of each compression block (d) is equal to 1024 samples and is equivalent to 256 msec at a 4 kHz sampling frequency. The digitized data after ADC is used for recovery. As shown in these figures, applying the $\ell_{1,2}$ recovery on the compressed data produced by the adjacent channels results in an improved recovery performance, compared to the sparse recovery. The averaged SNRs using the ℓ_1 and $\ell_{1,2}$ recovery are 16.64 and 21.80 dB, respectively. Based on the statistical analysis reported in [34], a minimum SNR of 10.45 dB (corresponding to a PRD of 30%) is acceptable to maintain the diagnostically important data in the recovered signal, e.g., for successful seizure detection. Reducing the number of measurements to $M = 1$, i.e., CR $= 16$ results in average SNR $= 13.72$ dB, using $\ell_{1,2}$ recovery. Thus, the system is able to successfully recover low-voltage iEEG signals compressed by a ratio as high as 16. Fig. 13 presents the average reconstruction SNR for sparse and joint recovery

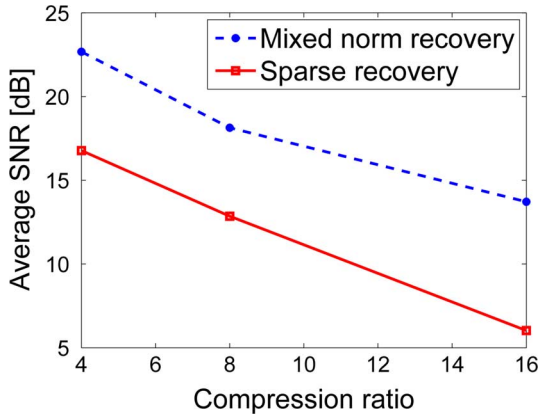

Fig. 13. Comparison of mixed norm and sparse recovery performance for different compression ratios. SNRs are averaged over 20 compression blocks.

for different compression ratios. The achieved compression ratios vary within the range of 16/15 up to 16, corresponding to $\mathrm{CR} = 16/\mathrm{M}$, where M is an integer number between 1 and 15.

As confirmed by the average SNR, the system is potentially able to recover the signal over the entire recording period, with some tolerable loss (i.e., $\mathrm{SNR} > 10.45$ dB). However, the amplitude and frequency of the recorded signals can significantly vary, depending on the distance between the microelectrodes and their surface area [35]. The amplitude of the very high frequency oscillations (250–500) recorded using macroelectrodes [36] is much smaller (5–30 versus 100–1250 μV) than human fast ripples recorded from smaller microelectrodes [35]. The amplitude of the fast ripples recorded by the high-density arrays will expectedly fall within the range of tens-hundreds of microvolts, which is efficiently recovered at the output of the system, as shown in Fig. 12.

The required time for the reconstruction of the 1024-length epochs of the multichannel signal is 1.12 second per channel, using a 2.66 GHz processor with 4 GB of RAM. To achieve a real-time performance, the speed of the reconstruction could be improved by hardware implementation of the algorithm on FPGA and using custom acceleration techniques.

B. Effect of Circuit Non-Idealities and Non-Adjacent Channels

A second database consisting of multichannel intracranially recorded signals of the slices of a rat somatosensory cortex under bicuculline, which blocks the synaptic inhibition and consequently mimics the epilepsy, is applied to the CS recording system. This signal includes epileptiform burst activity and extracellularly detected spikes. However, the verifiable signals of this database are associated with electrodes randomly located on the array. Consequently, these channels exhibit limited synchronous activity during the seizure, compared to the previous database. This effect is reflected in relatively lower recovery SNR of neural signals presented in Table III. SNR is evaluated for each channel and for the multichannel signal, by comparing $\mathbf{X}_{\mathrm{vec}}$ with the reconstructed multichannel data stream (defined as $\mathrm{SNR}_{\mathrm{T}}$).

In order to study the effect of circuit non-idealities (such as quantization and thermal noise) on the recovery performance, a comparison of SNR is also presented in Table III. The reconstruction results are compared by including and excluding different noise sources in simulations. The results are compared to the recovery performance of the compressed signal generated by matrix multiplication in MATLAB, using the same matrix as the output of the on-chip PRBS generator. CS can improve the attainable signal fidelity in the presence of sensor noise as shown in [45]. Although the reconstruction performance of a CS system is not as good as a simple quantizer for noiseless inputs [46], for more practical noisy signals recorded by the sensors, CS achieves a better performance, i.e., lower energy and improved PRD (Percentage Root-Mean Squared Difference) [45]. The CS system filters some of the input noise during reconstruction [45], and consequently is a correct choice for noisy environments such as a neural interface. Otherwise expressed, the recovery algorithm discards the coefficients below a certain threshold in the sparse representation of the signal. The discarded coefficients can be interpreted as filtered noise.

TABLE III
RECOVERY QUALITY IN THE PRESENCE OF NOISE (SNRS ARE IN DECIBELS)

Performance Metric	Gabor Transform, ℓ_1 recovery, CR = 4				
	×QN [a] ×cktN[b]	✓QN ×cktN	×QN ✓cktN	✓QN ✓cktN	*MAT LAB*
$\mathrm{SNR}_{\mathrm{CH1}}$	14.83	14.64	14.75	14.32	15.12
$\mathrm{SNR}_{\mathrm{T}}$	10.82	10.52	10.76	10.23	10.97

[a]Excluding the quantization noise.
[b]Excluding the noise of the circuit.

The recovery performance of the proposed system is marginally affected by the quantization noise of the ADC, as confirmed by the results of simulations presented in Table III. Excluding the circuit noise in simulations (thermal and flicker) results in a negligible improvement of the recovery performance which confirms the robustness of the CS system against non-idealities induced by the circuit. Consequently, the specifications related to the resolution of the ADC, the required noise performance of the analog front-end and the summing stage preceding the ADC and therefore the total power consumption and area of the chip can be further relaxed without jeopardizing the recovery performance.

C. Comparison and Future Work

Table IV summarizes the performance of the system and presents a comparison with published works. In this table, compression power and area refer to the extra power consumption and area usage of the signal digitization, compression and thresholding blocks which are commonly added to the total power consumption and area of the analog front-end. The authors in [8] apply compressive sensing on a single-channel pre-recorded EEG data by acquiring measurements in the digital domain. The power saving is significant while the area overhead is not addressed. Due to the youthfulness of the field and the lack of similar electronic architectures that use CS in brain implants, we have compared our results with a Discrete Wavelet Transform (DWT)-based design [49] for intra-cortical implants and several additional systems based on spike/AP detection [37], [47], [48], [50] for implantable neural recording

TABLE IV
COMPARISON OF SYSTEM PERFORMANCE WITH PUBLISHED LITERATURE

Parameter	[8]	[37]	[47]	[48]	[49]	[50]	This Work
Technology [μm CMOS]	0.09	0.13	0.5	0.18	0.5	0.5	0.18
Power supply [V]	0.6	1.2	3.3	1.8	-	3	1.2
Compression method	DCS	PWL Spike det.	Spike det.	AP det.	DWT Spike det.	Spike det.	MCS
Number of channels	1	1	100	16	32	32	16
Compression area per channel [mm^2]	0.103	0.080	< 0.160	> 0.0475	0.18	0.12	0.008
Compression power per channel [μW]	1.9	1.18	27	> 96	95	75	0.95
Sampling rate per channel [kS/s]	$\leq$20	90	15	30	25	20	4
Compression ratio	$\leq$10	125	150	48	$\leq$20	12.5	$\leq$16

applications. While the design in [49] mainly addresses the area-efficiency of the implantable system and proposes an architecture that sequentially evaluates the DWT of the multi-channel data in the digital domain, our results outperform this approach in terms of area and power efficiency. In addition, high compression ratios are achieved by means of the following thresholding and redundancy removal stages, while the DWT by itself does not result in any data compression. Thresholding, however, results in a significant loss of the signal in non-spiking regions while a more precise recovery is achieved at much lower compression ratios (e.g., at $\mathrm{CR} = 2$ in [49]). The chip includes several memory registers containing threshold values of different channels and additional blocks such as controllers, address generator and buffer units which degrade the power and area efficiency of the system. Some of the reported spike detector systems achieve significant data reduction [37], [47] with negligible overhead in terms of compression power and area [37]. However, the patient-specific threshold setting in such systems can result in design complexity in a real neural interface in addition to the loss of signal in non-spiking regions. Furthermore, the transmitted signal may not be acceptable to the clinicians who usually prefer to have access to the entire iEEG data, even though somewhat lossy, for a thorough neurological examination.

As a final remark, the method presented in this paper can also be implemented in a digital framework, similar to the concept presented in [8], but in a multichannel fashion. By using an ADC per channel and applying the random controlling signals to the digital outputs of channels, which pass through a single digital accumulator block, the full array will be encoded into a digital compressed data. Consequently, any possible advantage of applying CS in the digital domain as expressed in [8] can be exploited in a more area-efficient approach, compared to applying CS per channel. However, the system implemented in this paper benefits from using a single ADC per array which results in a small effective area per channel.

As an alternative approach to the proposed method, implementing an intricate ADC which performs the randomized summation of the channels' outputs will eliminate the summing stage and improve the area efficiency of the multichannel compressive sensing system. Further comparison and characterization of the proposed approaches is required to achieve an optimal model for multichannel signal compression.

V. CONCLUSION

A new multichannel architecture for compressive recording of cortical signals at the surface of the cortex is proposed. In addition to the area efficiency, the proposed method is easily adaptable to different compression ratios, depending on the sparsity of the input signals. The power efficiency resulting from the compressive sensing methodology in addition to the minimal area cost, make this approach highly relevant for power- and area-constrained multichannel sparse signal acquisition. This approach can be investigated in other applications than neural recording, which require data recording from multiple nodes. Extensive system-level analysis and simulations confirm the relevance and efficiency of the system for high-density recording applications, compared to alternative compression methods.

ACKNOWLEDGMENT

The entire signal processing analysis of this work is devised by the signal processing laboratory. The authors wish to thank Dr. K. Schindler and Mr. M. Fuchs for providing the intracranial EEG signals.

REFERENCES

[1] M. Mollazadeh, K. Murari, G. Cauwenberghs, and N. V. Thakor, "Wireless micropower instrumentation for multimodal acquisition of electrical and chemical neural activity," *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 6, pp. 388–397, 2009.

[2] M. J. Cook *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study," *Lancet Neurol.*, vol. 12, pp. 563–571, 2013.

[3] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran, "Brain-controlled interfaces: Movement restoration with neural prosthetics," *Neuron*, vol. 52, pp. 205–220, 2006.

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

[5] E. J. Candès, "Compressive sampling," in *Proc. Int. Congress of Mathematicians*, 2006, pp. 1433–1452.

[6] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2456–2466, 2011.

[7] A. M. R. Dixon, E. G. Allstot, D. Gangopadhyay, and D. J. Allstot, "Compressed sensing system considerations for ECG and EMG wireless biosensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 2, pp. 156–166, 2012.

[8] F. Chen, A. P. Chandrakasan, and V. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, pp. 744–756, 2012.

[9] M. H. Kamal, M. Shoaran, Y. Leblebici, A. Schmid, and P. Vandergheynst, "Compressive multichannel cortical signal recording," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 4305–4309.

[10] G. A. Worrell *et al.*, "High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings," *Brain*, vol. 131, pp. 928–937, 2008.

[11] J. J. V. Gompel *et al.*, "Phase I trial: Safety and feasibility of intracranial electroencephalography using hybrid subdural electrodes containing macro- and microelectrode arrays," *Neurosurg. Focus*, vol. 25, 2008.
[12] M. Stead *et al.*, "Microseizures and the spatiotemporal scales of human partial epilepsy," *Brain*, vol. 133, pp. 2789–2797, 2010.
[13] M. S. Fifer, S. Acharya, H. L. Benz, M. Mollazadeh, N. E. Crone, and N. V. Thakor, "Toward electrocorticographic control of a dexterous upper limb prosthesis: Building brain-machine interfaces," *IEEE Pulse*, vol. 3, pp. 38–42, 2012.
[14] X. Chen, Z. Yu, S. Hoyos, B. M. Sadler, and J. Silva-Martinez, "A sub-Nyquist rate sampling receiver exploiting compressive sensing," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 3, pp. 507–520, 2010.
[15] 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 and Systems*, 2007, pp. 1959–1962.
[16] M. Shoaran, C. Pollo, Y. Leblebici, and A. Schmid, "Design techniques and analysis of high-resolution neural recording systems targeting epilepsy focus localization," in *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Soc.*, 2012, pp. 5150–5153.
[17] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, 2007.
[18] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
[19] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
[20] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2429–2438, 1993.
[21] M. Kowalski and B. Torrésani, "Sparsity and persistence: Mixed norms provide simple signal models with dependent coefficients," *Signal, Image, Video Process.*, vol. 3, no. 3, pp. 251–264, 2009.
[22] H. Rezaee-Dehsorkh, N. Ravanshad, R. Lofti, K. Mafinezhad, and A. M. Sodagar, "Analysis and design of tunable amplifiers for implantable neural recording applications," *IEEE J. Emerg. Select. Topics Circuits Syst.*, vol. 1, no. 4, pp. 546–556, 2011.
[23] K. A. Ng and Y. P. Xu, "A compact, low input capacitance neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 5, pp. 610–620, 2013.
[24] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, 2003.
[25] S. Amico, M. Conta, and A. Baschirotto, "A 4.1-mW 10-MHz fourth-order source follower-based continuous-time filter with 79-dB DR," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2713–2719, 2006.
[26] K. S. Guillory and R. A. Normann, "A 100-channel system for real time detection and storage of extracellular spike waveforms," *J. Neurosci. Methods*, vol. 91, pp. 21–29, 1999.
[27] M. S. Chae, W. Liu, and M. Sivaprakasam, "Design optimization for integrated neural recording systems," *IEEE J. Solid-State Circuits*, vol. 43, no. 3, pp. 1931–1939, 2008.
[28] S. Rabii and B. A. Wooley, "A 1.8-V digital-audio sigma-delta modulator in 0.8-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 32, no. 6, pp. 783–796, 1997.
[29] M. Shoaran, M. M. Lopez, V. S. R. Pasupureddi, Y. Leblebici, and A. Schmid, "A low-power area-efficient compressive sensing approach for multi-channel neural recording," in *Proc. IEEE Int. Symp. Circuits and Systems*, 2013, pp. 2191–2194.
[30] R. Schreier, J. Silva, J. Steensgaard, and G. C. Temes, "Design-oriented estimation of thermal noise in switched-capacitor circuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 11, pp. 2358–2368, 2005.
[31] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 136–147, 2007.
[32] M. Steyaert, W. Sansen, and C. Zhongyuan, "A micropower low-noise monolithic instrumentation amplifier for medical purposes," *IEEE J. Solid-State Circuits*, vol. 22, no. 6, pp. 1163–1168, 1987.
[33] T. M. Seese, H. Harasaki, G. M. Saidel, and C. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue in chronic heating," *Lab. Invest.*, vol. 78, pp. 1553–1562, 1998.
[34] G. Higgins, S. Faul, R. P. McEvoy, B. McGinley, M. Glavin, W. P. Marnane, and E. Jones, "EEG compression using JPEG2000: How much loss is too much?," in *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Soc.*, 2010, pp. 614–617.
[35] A. Bragin, I. Mody, C. L. Wilson, and J. Engel, "Local generation of fast ripples in epileptic brain," *J. Neurosci.*, pp. 2012–2021, 2002.
[36] J. D. Jirsch, E. Urrestarazu, P. LeVan, A. Olivier, F. Dubeau, and J. Gotman, "High-frequency oscillations during human focal seizures," *Brain*, vol. 129, pp. 1593–1608, 2006.
[37] A. Rodriguez-Perez, J. Ruiz-Amaya, M. Delgado-Restituto, and A. Rodriguez-Vazquez, "A low-power programmable neural spike detection channel with embedded calibration and data compression," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 2, pp. 87–100, 2012.
[38] F. Shahrokhi, K. Abdelhalim, D. Serletis, P. L. Carlen, and R. Genov, "The 128-channel fully differential digital integrated neural recording and stimulation interface," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 3, pp. 149–161, 2010.
[39] S. Rai, J. Holleman, J. N. Pandey, F. Zhang, and B. Otis, "A 500 W neural tag with 2 V AFE and frequency-multiplying MICS/ISM FSK transmitter," in *Proc. IEEE Int. Solid-State Circuits Conf., Dig. Tech. Papers*, 2009, pp. 212–213, 213a.
[40] V. Majidzadeh, A. Schmid, and Y. Leblebici, "Energy efficient low-loise neural recording amplifier with enhanced noise efficiency factor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 3, pp. 262–271, 2011.
[41] J. Aziz, K. Abdelhalim, R. Shulyzki, R. Genov, B. Bardakjian, M. Derchansky, D. Serletis, and P. Carlen, "256-channel neural recording and delta compression microsystem with 3D electrodes," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, 2009.
[42] L. Wen-Sin, Z. Xiaodan, Y. Libin, and L. Yong, "A 1-v 60-uw 16-channel interface chip for implantable neural recording," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2009, pp. 507–510.
[43] K. Abdelhalim, H. Jafari, L. Kokarovtseva, J. L. P. Velazquez, and R. Genov, "64-channel UWB wireless neural vector analyzer SoC with a closed-loop phase synchrony-triggered neurostimulator," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2494–2510, 2013.
[44] K. Abdelhalim, L. Kokarovtseva, J. L. Perez Velazquez, and R. Genov, "915-MHz FSK/OOK wireless neural recording SoC with 64 mixed-signal FIR filters," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2478–2493, 2013.
[45] F. Chen, F. Lim, O. Abari, A. Chandrakasan, and V. Stojanovic, "Energy-aware design of compressed sensing systems for wireless sensors under performance and reliability constraints," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 3, pp. 650–661, 2013.
[46] V. K. Goyal, A. K. Fletcher, S. Rangan, A. Chandrakasan, and V. Stojanovic, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48–56, 2008.
[47] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.
[48] B. Gosselin, A. E. Ayoub, J. F. Roy, M. Sawan, F. Lepore, A. Chaudhuri, and D. Guitton, "A mixed-signal multichip neural recording interface with bandwidth reduction," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 3, no. 3, pp. 129–141, 2009.
[49] A. M. Kamboh, K. G. Oweiss, and A. J. Mason, "Resource constrained VLSI architecture for implantable neural data compression systems," in *Proc. IEEE Int. Symp. Circuits and Systems*, 2009, pp. 1481–1484.
[50] R. H. Olsson and K. D. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2796–2804, 2005.

Mahsa Shoaran (S'11) received the B.Sc. degree in electrical engineering, electronics, and the M.Sc. degree in microelectronics, IC design, from the Sharif University of Technology, Tehran, Iran, in 2008 and 2010, respectively.

In 2011, she joined the Microelectronic Systems Laboratory of the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, as a Ph.D. student. She has worked on system and circuit design for biomedical applications, the implantable devices for cardiovascular diseases, and brain machine interfaces. Her main research interest is low-power IC design with an emphasis on biomedical applications, innovative system design towards diagnosis, and treatment of neurological disorders and technology-oriented neuroscience.

Mahdad Hosseini Kamal (S'10) received the B.Sc. degree in electrical engineering, communications, from the K. N. Toosi University of Technology, Tehran, Iran, and the M.Sc. degree in communication systems from Ecole Polytechnique Fedérale de Lausanne (EPFL), Lausanne, Switzerland, in 2007 and 2010, respectively.

In 2010, he joined the Signal Processing Laboratory at EPFL as a Doctoral Assistant, where he is currently working toward the Ph.D. degree. His main research interests include inverse problems, compressed sensing, and biomedical signal processing.

Claudio Pollo is a member of the senior staff (Leitenderarzt) and a board certified neurosurgeon and leader of the stereotactic, functional and epilepsy surgery program in the Neurosurgical Department of University Hospital, Inselspital, Bern (Head: Prof. A. Raabe).

He previously was trained predominantly in the Neurosurgery Department of the Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne (Chair: Prof. J. G. Villemure), Switzerland. He also performed a research fellowship dedicated to deep brain stimulation (DBS) in Grenoble, France (Chair: Prof. A. L. Benabid). In 2004, he performed the postgraduate diploma in biomedical engineering delivered by the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He, as well as other members of the interdisciplinary team of the epilepsy surgery and DBS surgery program in Bern, and previously in the Lausanne-Geneva area, have presented and authored clinical and experimental studies focused on epilepsy surgery and clinical results in international congresses and journals. He is involved in different scientific projects financially sustained by the Swiss National Foundation (SNF) or Swiss Committee for Technology and Innovation (CTI), and dedicated to epilepsy surgery as well as DBS. He is author or coauthor of more than 50 original papers or book chapters. He has cofounded two startups involved in technology transfer and holds several patents.

Pierre Vandergheynst received the M.S. degree in physics and the Ph.D. degree in mathematical physics from the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1995 and 1998, respectively.

From 1998 to 2001, he was a Postdoctoral Researcher with the Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He was Assistant Professor at EPFL (2002–2007), where he is now an Associate Professor. His research focuses on harmonic analysis, sparse approximations and mathematical image processing with applications to higher dimensional, complex data processing. He was co-Editor-in-Chief of *Signal Processing* (2002–2006) and Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2007–2011), the flagship journal of the signal processing community. He is the author or coauthor of more than 70 journal papers, one monograph and several book chapters. He has received two IEEE best paper awards.

Dr. Vandergheynst is a laureate of the Apple 2007 ARTS award and of the 2009–2010 De Boelpaepe prize of the Royal Academy of Sciences of Belgium. He has been on the Technical Committee of various conferences, serves on the steering committee of the SPARS workshop, and was co-General Chairman of the EUSIPCO 2008 conference.

Alexandre Schmid (S'98–M'04) received the M.Sc. degree in microengineering and the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1994 and 2000, respectively.

He has been with EPFL since 1994, working in the Integrated Systems Laboratory as a Research and Teaching Assistant, and at Electronics Laboratories as a Postdoctoral Fellow. In 2002, he joined Microelectronic Systems Laboratory as a Senior Research Associate, where he has been conducting research in the fields of bioelectronic interfaces, non-conventional signal processing and neuromorphic hardware, and reliability of nanoelectronic devices. He also teaches in the Microengineering and Electrical Engineering Departments at EPFL. Since 2011, he has been a "Maître d'Enseignement et de Recherche" (MER) faculty member at EPFL. He is coauthor and coeditor of three books, and more than 100 articles published in journals and conferences.

Dr. Schmid served as the General Chair of the Fourth International Conference on Nano-Networks in 2009, and has served as an Associate Editor of the *IEICE Electronics Express* since 2009.