A Low-Power Area-Efficient Compressive Sensing Approach for Multi-Channel Neural Recording

Mahsa Shoaran, Mariazel Maqueda Lopez, Vijaya Sankara Rao Pasupureddi, Yusuf Leblebici and Alexandre Schmid
Microelectronic Systems Laboratory (LSM), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne
Email: mahsa.shoaran@epfl.ch

***Abstract*—High-density wireless intracranial neural recording is a promising technology enabling the autonomous diagnosis and therapy of brain diseases. Increasing the number of recording channels is accompanied by the increased amount of data resulting in an unacceptable transmission power. A comprehensive study of possible compressed sensing methods in the context of neural signals has been done, and the compression of signals originating from different channels in the spatial domain has been implemented at the system and circuit levels. Results of the simulations in a UMC 0.18μm CMOS technology and subsequent reconstructions show the possibility of compressing with ratios as high as 2.6 with a recovery SNR of at least 10dB using extremely compact and low-power circuits. The power efficiency and limited area per channel confirm the relevance of the proposed approach for multi-channel high-density neural interfaces.**

I. Introduction

Wireless multi-channel neural recording systems are approaching technology limits due to the high overall data rates and increased transmission power which fall beyond the available bandwidth of the state-of-the-art wireless links and the acceptable range of heat generation for implantable devices. Monitoring the activity of large number of neurons is a prerequisite for understanding the cortical structures [1]. Further development of the recording interfaces is required to enable large-scale information extraction to support informative analysis as well as effective diagnosis and treatment.

Recently, the rapidly developing theory of compressed sensing (CS) has been studied to tackle the data rate issue in the context of biological signals [2]-[4]. Many biological signals such as action potentials, EEG and ECG have an information rate much smaller than the rate dictated by the Nyquist sampling theorem [2]. This property enables the recovery of the original signal from a small number of linear measurements resulting in bandwidth saving. Additional details regarding the CS theory can be found in [5].

Considering the sparse behavior of neuronal activity resulting from the low firing rate of neurons and the large difference of the voltage amplitude in the spiking and non-spiking modes, the simplest compression scheme can be realized by taking few (M) linear measurements out of N (N>>M) samples of the signal in a defined time frame. The concept can be implemented either in the analog domain prior to digitization (neural CS or NCS of Fig. 1(a)), or downstream the ADC (Fig. 1(b)). In the analog approach, the total data rate and digitization power is decreased but the area overhead due to the multi-path nature of the CS topology results in a large area per channel. Although the power analysis in [2] shows the superior performance of a digital (Fig. 1(b)) over the analog implementation, including M multiplication and accumulation blocks in each channel can result in a large area which summed up to the area of the low-noise amplifier, the Nyquist rate ADC and the random sequence generator, disqualifies this approach for a multi-channel recording interface which should include electronics for many channels in a limited die area.

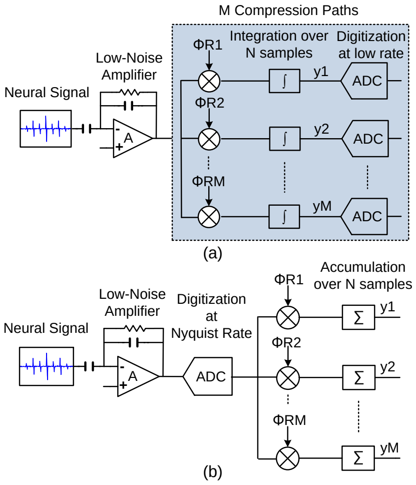

Fig. 1. (a) Analog single-channel NCS and (b) Digital single-channel NCS.

Considering the area and power inefficiency of a single-channel NCS scheme, the compression algorithm has been implemented in a more efficient and compact way for multi-channel data acquisition. The spatial sparsity of the signals generated by different channels on the electrode array is efficiently exploited in the proposed scheme.

This paper is organized as follows. Section II presents the conventional CS topologies. Spatial CS is discussed in Section III. Simulation results of the proposed architecture are presented in Section IV. Section V concludes the paper.

II. Conventional CS Topologies

Only few methods for circuit-level implementation of a CS system are proposed in the literature. It can be implemented either in the analog or digital domains. In the analog domain, the multiplication and integration can be performed in the voltage [6] or current mode [7]. In a conventional neural

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

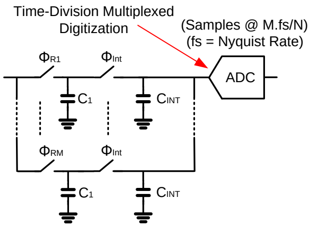

Fig. 2. Multiplication with random sequences, integration and multiplexed digitization in an analog approach.

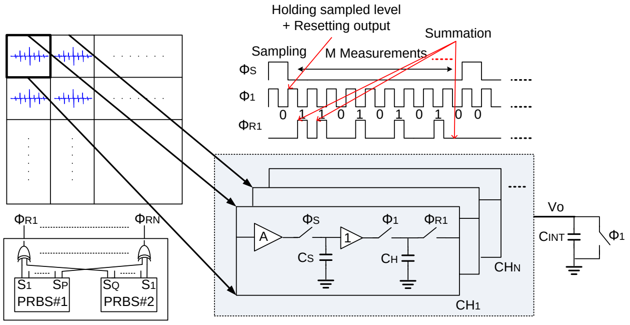

Fig. 3. Proposed multi-channel SCS.

interface, the weak neural potentials are amplified through low-noise amplifiers in the front-end of the system. Thus, performing CS in the current mode requires adding a G_m stage which is power hungry and area inefficient to guarantee a linear operation. As an additional drawback, a trans-impedance amplifier with large linear resistors and high DC-gain OTA is required to convert the current signal back to voltage prior to digitization resulting from the lack of properly designed current-mode ADCs.

The shaded block in Fig. 1(a) can be implemented as shown in Fig. 2. The amplified signal passes through M compression paths. The signal is sampled according to the level of the random sequence controlling the sampling switch in each path. The resetting switch to the ground is not shown in this figure. During the integration phase, charge sharing occurs between the sampling and integrating capacitors, which emulates a lossy integration over C_{INT}. Rather than designing M similar ADCs with a very low sampling rate of f_s/N, time multiplexing is applied to achieve a moderate sampling rate which falls in the high-performance and low-power region of operation of SAR ADCs. This implementation circumvents the need for high-power opamp-based integrators [6]-[7]. However, the integrating capacitors should be large compared to the sampling capacitors to achieve close-to-ideal integration.

Detailed analysis in [2] and [7] discuss the noise and power contributions of the mixer, integrator and sample/hold, LNA and ADC. Although the sampling rate of a single ADC is not a limiting design parameter in a digital implementation due to the relatively low frequency of neural signals [2], in a multi-channel system, on the other hand, the digitization power and area usage are multiplied by the number of channels and therefore significantly contribute to the total power dissipation and die area. Assuming a minimal area of 200μm×200μm for the LNA [8] and 200μm×550μm for the ADC and CS encoder [2], the minimum required area per channel is approximately 400μm×400μm. A novel area and power-efficient implementation is discussed in the next Section.

III. Spatial CS (SCS)

A. System Architecture

Wideband neural signals comprising high amplitude action potentials or spikes with a firing rate of 10-100Hz followed by long periods of low activity are sparse in the time domain. The lower frequency EEG signals have a sparse representation in Gabor or wavelet domains [2].

In addition to single-channel neural data which is sparse over time, the entire multi-channel array can be considered as an image sensor array consisting of pixels (here the pixels are the electrodes recording from different spatial dimensions of the brain) in which very few are active at each time instant. This spatial sparsity can be derived from the fact that each channel is active over a small percentage of time, and the signals recorded by neighbouring electrodes, depending on the spatial resolution and pitch of the electrodes, are delayed, attenuated or amplified forms of the same signal in the worst case of correlation. Nevertheless, at large scale, the full image is considered as sparse. The block diagram of the proposed spatial compression scheme is depicted in Fig. 3.

The amplified signals of the individual channels are sampled on C_S and kept constant during M measurements. The linearity of the track and hold circuit is guaranteed by using PMOS source-to-bulk connected source followers. The sampled signal charges the holding capacitor in the first half cycle of the clock. In the second half, the holding capacitors of all channels are connected to the integrating capacitor, based on the random value controlling the in-pixel switch. Thus, the signals of all channels in the array are multiplied by the instantaneous random value and summed together on C_{INT} ($C_{INT} >> C_H$). The compressed voltage $V_o(n)$ can be written as:

$$\frac{C_H\phi_{R1}(n)V_1(n-1/2)+\ldots+C_H\phi_{RN}(n)V_N(n-1/2)}{C_H\phi_{R1}(n)+\ldots+C_H\phi_{RN}(n)+C_{INT}} \tag{1}$$

where $V_i(n)$ is the tracked level of the signal originating from channel number i at time nT, with T being the period of the clock signal. $\phi_{Ri}(n)$ is the level (1 or 0) of the random sequence applied to ith channel at time nT and N is the number of channels. As a significant advantage, this design encodes the full array to one single data which is digitized using a single ADC. As a benefit of CS, the sampling rate of the latter ADC is N/M times smaller than the sampling rate of the unique ADC which is required in a non-compressed but time-multiplexed topology. In large arrays, similar to the image compression topologies [9], one ADC can be allocated to each column which receives the randomly summed value of signals generated from that column. The accumulation is performed after the ADCs. Thus, the cost of implementation in terms of in-pixel area and power is much less than previous topologies.

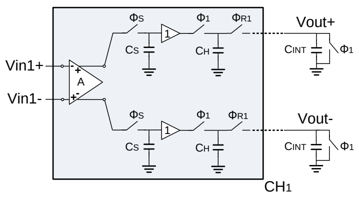

Fig. 4. Differential implementation of the proposed topology.

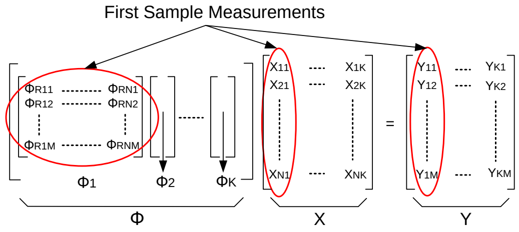

Fig. 5. Random matrix multiplication.

Using a differential topology (Fig. 4), the non-linearity and dc components caused by the source follower buffer circuit are partially removed.

B. Pseudo Random Bit Sequence (PRBS) generator

The actual implementation of CS using any of the presented topologies requires the generation of M rows of the measurement matrix at low power consumption and small area overhead. In a single-channel approach, each channel needs to be loaded with M sequences. In SCS on the other hand, each channel is only loaded with one sequence. The measurement matrix supporting the first M measurements required for recovering the first sample of each channel is created by taking the first M values of the in-channel sequences located in the columns of the matrix (Φ_1 in Fig. 5). Pseudorandom sequences which exhibit low coherence with any fixed sparsity basis [2] are a proper choice for the implementation of the measurement matrix. In this design, the sequence generation is achieved by XORing the multiple outputs of different length PRBS generators (Fig. 3). Considering a test recording array of 4×4 and a value of M equal to 7 (compression ratio of 16/7), the M×N matrices are generated as shown in Fig. 5. In this figure, N is the number of channels (16) and K is the number of samples per channel in a segment of signal. 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.

The possible correlation between the generated sequences in this design does not affect the performance of the reconstruction because the similarity periods do not occur simultaneously and they are shifted in time. Hence, each row of the matrix which consists of the individual samples of the different sequences, is uncorrelated with the next rows. The sequence generators operate at a speed which is M times faster than the Nyquist rate. True Single-Phase Clocked (TSPC) flip-flops are used resulting in very low power consumption and a compact implementation. A small number of 9 flip-flops and 16 XOR gates is sufficient to generate the required sequences for 16 channels. Unlike the approach used in [2], a PRBS generator with M flip flops is not necessary. All the N sequences are generated by XORing the different combinations of the two blocks with smaller number of flip-flops. Since the length of the sequences is not required to be very long, a small number of flip flops can be used in the PRBS generators and the XORed output random sequences are sufficiently long to support the state-of-the-art biomedical applications.

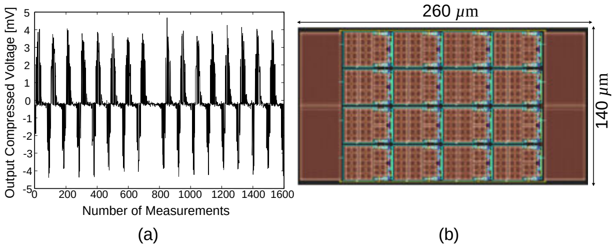

Fig. 6. (a) Output differential voltage and (b) Layout of the CS array.

IV. Simulation Results

In order to demonstrate the functionality of the proposed system, an array of 16 neural recorder channels has been designed in a UMC 0.18μm CMOS technology (Fig. 6(b)). Synthetic neural data [10] with more than 10 nonzero elements out of 200 and additive white gaussian noise with an overall SNR of 40dB (typical of a neural recording front-end) is applied as the amplified signal prior to the CS stage. The total power consumption of the array and random sequence generator (excluding the amplifiers) drawn from a 1.2-V supply is 1.07μW. The Basis Pursuit Denoising Method provided by the SPGL1 solver [11] is used for reconstruction. Fig. 6(a) shows the differential voltage which is observed at the output of the circuit in Fig. 4, and corresponding to an effective sampling rate of 20K per channel. The original and reconstructed signals of two sample channels based on the proposed circuit are shown in Fig. 7 and compared to the conventional single-channel approach with the same compression ratio (M = 87, N = 200) with an approximated area of 400μm×400μm per channel (excluding LNAs). The recovery SNR of the reconstructed signal ($\hat{x}$) with respect to the original signal (x) is calculated from the performance measure defined as:

$$SNR = -20\log_{10}\|x-\hat{x}\|_2/\|x\|_2. \tag{2}$$

The averaged SNR of 16 channels obtained using MATLAB and circuit simulations are 18.42dB and 12.94dB, respectively, and 12.87dB including the circuit transient noise analysis. The SNR is 21.57dB for the single-channel approach, averaged over 100 simulations. The quality of spike reconstruction using SCS remains sufficiently good (10.62dB) even for highly overlapped input signals (Fig. 9(d)). To validate the randomness of the generated sequences, results of CS and reconstruction with a random Gaussian matrix and the designed PRBS generator are shown in Fig. 8. Fig. 9(a) depicts the averaged SNR of channels versus compression factor. Fig. 9(b) and (c)

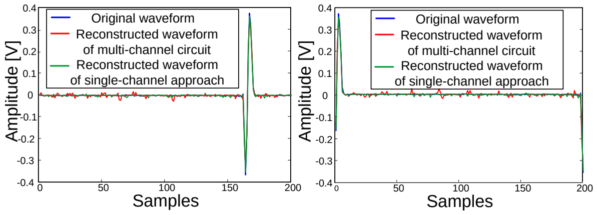

Fig. 7. Original and reconstructed signals within 10 miliseconds for two sample channels using proposed circuit simulations and conventional single-channel MATLAB simulations.

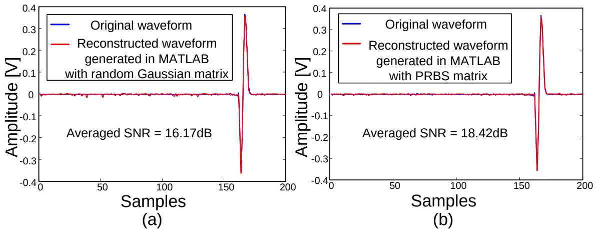

Fig. 8. Original and reconstructed signals within 10 miliseconds for a sample channel using (a) a random Gaussian matrix and (b) the proposed PRBS generator.

demonstrate the SNR versus channel numbers and SNR versus samples, with samples being the columns of matrix X in Fig. 5. Although the SNR of all sample reconstructions is not high, the channel-based SNR is sufficiently high which guarantees the precise recovery of the spiking and non-spiking periods in all channels of the array. The proposed CS block follows a very low-noise amplifier in the front-end of the system, in each channel ([8]). Table I presents the performance summary and the comparison with the only published circuit-level CS system for neural recording. Further optimization of the design in terms of circuit non-idealities (mainly caused by lossy integration), noise, distortion and the effect of performance metrics of the ADC in overall SNR are under consideration.

V. Conclusion

A multi-channel CS topology appropriate for high-density intracranial neural recording is proposed. Without applying any thresholding or signal-dependent pre-processing, the functionality is proven through system as well as circuit level simulations. An efficient method for multi-path random sequence generation is also presented. In future, combining the SCS approach with video compression algorithms [12] and spatial redundancy removal methods [3] will be considered to achieve more power saving.

Acknowledgment

The authors acknowledge the support of the Swiss NSF under project number 200021-130166.

References

[1] G. Buzsaki, “Large-scale recording of neuronal ensembles,” *Nat Neurosci*, vol. 7, pp. 446-451, 2004.

[2] 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.

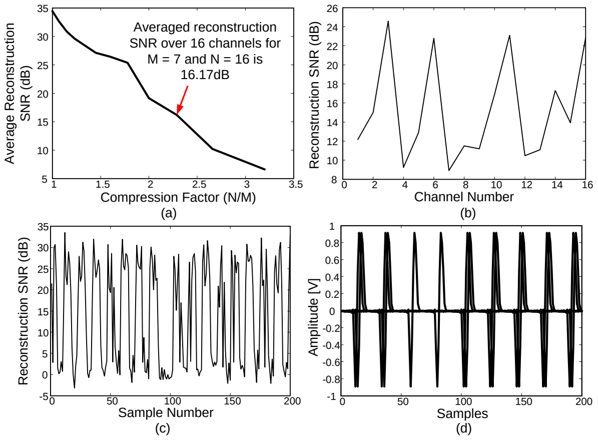

Fig. 9. (a) Averaged reconstruction SNR for 16 channels versus compression factor. (b) Reconstruction SNR versus channel number. (c) Reconstruction SNR versus number of samples. (d) Correlated multi-channel neural signals applied to the proposed compression topology.

TABLE I
PERFORMANCE SUMMARY OF THE CS CIRCUIT

Parameter	This Work	[2]
Technology (nm CMOS)	**180**	**90**
Supply Voltage (V)	**1.2**	**0.6**
Array Dimension	**4×4**	**1×1**
Sampling Rate per channel	**20k**	**20k**
Data Reduction	**2.3×**	**10-40×**
CS Area per channel	**48μm×48μm**	**200μm×450μm**
AFE Noise (μV_{rms})	**1.77**	**78**
AFE Power (μW)	**5.8**	**< 0.1**
CS power (μW)(full array)	**1.07**	**1.9**

[3] 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.

[4] A. M. R. Dixon, E. G. Allstot, D. Gangopadhyay, D. J. Allstot, “Compressed Sensing System Considerations for ECG and EMG Wireless Biosensors,” *IEEE Transactions on Biomedical Circuits and Systems*, vol.6, no.2, pp. 156-166, 2012.

[5] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, pp. 21-30, 2008.

[6] 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. 2007 IEEE Int. Symp. Circuits and Systems (ISCAS)*, pp. 1959-1962, 2007.

[7] 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*, vol. 58, no. 3, pp. 507-520, 2010.

[8] M. Shoaran, C. Pollo, Y. Leblebici and A. Schmid, “Design Techniques and Analysis of High-Resolution Neural Recording Systems Targeting Epilepsy Focus Localization,” *International Conference of the IEEE EMBS*, pp. 5150-5153, 2012.

[9] V. Majidzadeh, L. Jacques, A. Schmid, P. Vandergheynst, Y. Leblebici, “A (256×256) pixel 76.7mW CMOS imager/ compressor based on real-time in-pixel compressive sensing,” *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, no. 1, pp. 2956-2959, 2010.

[10] L. Smith, N. Mtetwa, “Manual for the noisy spike generator matlab software,” 2006. [Online]. Available: http://www.cs.stir.ac.uk/ lss/noisyspikes/

[11] E. V. Berg and M. P. Friedlander, “SPGL1: A solver for large-scale sparse reconstruction”, June 2007. [Online]. Available: http://www.cs.ubc.ca/labs/scl/spgl1

[12] C. H. Chung, L. G. Chen, Y. C. Kao, F. S. Jaw, “Multichannel evoked neural signal compression using advanced video compression algorithm,” *International IEEE Conference on Neural Engineering (NER)*, pp. 697-701, 2009.