A Signal-agnostic Compressed Sensing Acquisition System for Wireless and Implantable Sensors

Fred Chen, Anantha P. Chandrakasan, and Vladimir Stojanović
Department of EECS, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

***Abstract*—A signal-agnostic compressed sensing (CS) acquisition system is presented that addresses both the energy and telemetry bandwidth constraints of wireless sensors. The CS system enables continuous data acquisition and compression that are suitable for a variety of biophysical signals. A hardware efficient realization of the CS sampling demonstrates data compression up to 40x on an EEG signal while maintaining low perceptual loss in the reconstructed signal. The proposed system also simultaneously relaxes the noise and resolution constraints of the analog front end (AFE) and ADC by nearly an order of magnitude. The CS sampling hardware is implemented in a 90 nm CMOS process and consumes 1.9 µW at 0.6 V and 20 kS/s.**

I. Introduction

The utility of a wireless sensor is constrained by its limited energy source which imposes strict energy requirements on the sensor circuits. Fig. 1 shows the energy costs of typical circuit blocks used in sensors for medical monitoring. In applications such as implantable neural recording arrays, the high energy cost to transmit a bit of information and the radio's limited bandwidth necessitate data compression or filtering at the sensor to limit energy consumption and data throughput [1].

As shown in Table I, many bio-signals of interest occur infrequently, so transmitting data only when a significant event is detected can be effective in reducing data and power [2]. Existing integrated data compression strategies for biophysical monitoring adopt this approach and rely on known signal characteristics in order to detect significant events [1-5]. However, the filtered data contains limited information. In neural recorders, the data is often limited to just the time and amplitude of a neural spike event rather than the signal itself [1, 3]. Even when the event detection is used to trigger a full signal capture [2], the system is susceptible to missing events entirely if detection thresholds are not properly set. Meanwhile, feature extraction approaches require training, are usually signal specific and still do not provide the original signal [4, 5]. All of these signal processing strategies exemplify a common goal in energy constrained wireless sensors: reduce the number of bits transmitted to minimize the average radio power (typically dominant) while maintaining the captured signal information.

TABLE I
COMMON MEASURED BIO-SIGNALS

Signal	*Sampling Rate*	*Frequency of Events*	*Event Duration*	*Duty Cycle (%)*
Extracellular APs	30 kHz	10 – 150 /s	1 – 2 ms	2 to 30
EMG	15 kHz	0 – 10 /s	0.1 – 10 s	0 to 100
EKG	250 Hz	0 – 4 /s	0.4 – 0.7 s	0 to 100
EEG, LFP	100 Hz	0 – 1 /s	0.5 – 1 s	0 to 100
O_2, Ph, Temp.	0.1 Hz	0.1 /s	N/A	Very low

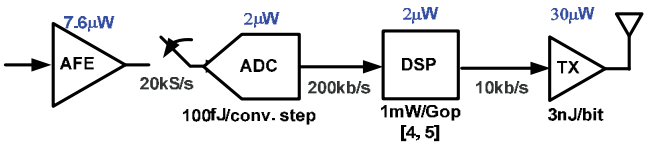

Fig. 1. Energy costs for typical circuits in bio-sensor applications. It is assumed the DSP performs some data filtering and the TX power scales with data rate.

In this work, we present the design and implementation of a new sensor system architecture based on the theory of compressed sensing that more efficiently achieves this goal. As results will show, this approach reduces the average radio power by exploiting signal sparseness to encode the data at a high compression factor. The reconstruction process also enables power reduction in the frontend circuitry by relaxing the noise and resolution requirements of the AFE and ADC. Unlike event detection based data compression, this approach enables a faithful reconstruction of the entire original signal and is applicable across a variety of signals without knowing the signal details *a priori*. While most of the results and examples presented are in the context of medical applications, they can be generally applied to other fields as well. The following sections will first give background on CS theory followed by a discussion of how the proposed system can reduce energy consumption in wireless sensors. The remaining sections will describe the hardware implementation, measured results and conclusions.

II. Compressed Sensing Background

CS is an emerging field whose theory leverages known signal structure to acquire data at a rate proportional to the information rate rather than the Nyquist rate [6]. CS is based on several key concepts which will be discussed hereafter: signal sparsity, signal reconstruction and incoherent sampling.

A. Signal Sparsity

CS theory relies first and foremost on the signal of interest, $\boldsymbol{f}$, having a sparse representation in some basis, $\boldsymbol{\Psi}=[\boldsymbol{\psi}_1\boldsymbol{\psi}_2...\boldsymbol{\psi}_L]$ such that $\boldsymbol{f}=\boldsymbol{\Psi x}$ or:

$$\boldsymbol{f} = \sum_{i=1}^{L} x_i \boldsymbol{\psi}_i \tag{1}$$

where $\boldsymbol{x}$ is the coefficient vector for $\boldsymbol{f}$ under the basis $\boldsymbol{\Psi}$. For $\boldsymbol{f}$ to be sparse in $\boldsymbol{\Psi}$, the coefficients, x_i, must be mostly zero or insignificant such that they can be discarded without any perceptual loss. If $\boldsymbol{f}$ has the most compact representation in $\boldsymbol{\Psi}$, then $\boldsymbol{f}$ should be compressible if acquired in some other basis. So sparseness also implies compressibility and *vice versa*. A familiar example is a sampled sine wave which has many coefficients in time, but requires only one non-zero coefficient in the Fourier basis. Fortunately, many of the bio-signals in Table I have sparse representations in either the Gabor or wavelet domains [7, 8] thus making them suitable for CS.

B. Signal Recovery From Incomplete Measurements

CS theory also proposes that rather than acquire the entire signal and then compress, it should be possible to capture only the useful information to begin with. The challenge then is how to recover the signal from what would traditionally seem to be an incomplete set of measurements. Again, the sparseness of the signal is relied upon to make this possible.

Consider an $M \times N$ measurement matrix, $\boldsymbol{\Phi}$, where $\boldsymbol{y}=\boldsymbol{\Phi f}$ and $\boldsymbol{f}$

is the N-dimensional signal to recover. When $M < N$ such that the system is underdetermined, there are an infinite number of feasible solutions for $\boldsymbol{f}$. However, if the signal to be recovered is known to be sparse, then the sparsest solution (most 0's) out of the infinitely possible is often the correct solution. A common and practical approach used to determine the sparse solution is to solve the convex optimization problem:

$$\underset{x}{\text{minimize}} \| \boldsymbol{x} \|_{l_1} \text{ subject to } \boldsymbol{y} = \boldsymbol{\Phi\Psi x} \tag{2}$$

where $\boldsymbol{\Psi}$ is the $N \times L$ basis matrix and $\boldsymbol{x}$ is the coefficient vector from (1). The recovered signal is then $\hat{\boldsymbol{f}} = \boldsymbol{\Psi x}^*$ where $\boldsymbol{x}^*$ is the optimal solution. The l_1-norm cost function serves as a proxy for sparseness as it heavily penalizes small coefficients so the optimization drives them to zero. The problem of minimizing the l_1-norm in (2) has been shown to be solved efficiently and require only a small set of measurements ($M<<N$) to enable perfect recovery [9]. The implication of these results is that an N-dimensional signal can be recovered from a lower order number of samples, M, provided that the signal is sparse in some basis. We rely on this result from CS theory to reduce the data that the sensor must transmit.

C. Incoherent Sampling

In addition to sparseness, CS also relies on incoherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, where coherence measures the largest correlation between any row of $\boldsymbol{\Phi}$ and column of $\boldsymbol{\Psi}$. The less coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, the fewer the measurements (M) needed to recover the signal [9]. The canonical basis (delta functions) and Fourier basis in a discrete Fourier transform is an example of a maximally incoherent $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ pair.

In terms of hardware cost and complexity, it is desirable if the signal basis, $\boldsymbol{\Psi}$, does not need to be known *a priori* in order to determine a viable sensing matrix, $\boldsymbol{\Phi}$. Fortunately, random sensing matrices with sufficient sample size exhibit low coherence with any fixed basis [9]. This means that a random sensing matrix can acquire the sufficient measurements needed to enable signal reconstruction of *any* sparse signal without knowing *a priori* what the proper basis $\boldsymbol{\Psi}$ for the signal is. We leverage this principle to build an infrastructure for data acquisition and compression that is agnostic to the type of signals being acquired, provided that they are sparse.

III. Compressed Sensing for Wireless Sensors

Fig. 2 shows a block diagram of the proposed system annotated with example waveforms of the signal compression and reconstruction along with the equivalent mathematical function of the CS block. In previous proposed practical applications of CS theory, the measurement $\boldsymbol{y}=\boldsymbol{\Phi f}$ was applied prior to digitization either because the dominant consumer of power was the sensing mechanism [10] or because the sampling frequency of the ADC was the limiting factor [11]. In most wireless biomedical sensor applications, the sampling frequency is rarely a limitation, as is evident from Table I, while the dominant consumer of energy is in the wireless telemetry circuits. Thus, the proposed system architecture implements the CS sampling on the digitized samples where the hardware and power costs benefit from CMOS scaling.

Fig. 2 illustrates how CS theory is applied to reduce the total sensor power. The CS block compresses the N ADC samples, $\boldsymbol{f}$, into M linear combinations, $\boldsymbol{y}$, by realizing the matrix-multiply operation $\boldsymbol{\Phi f}$ ($\boldsymbol{f}^{\mathbf{T}}\boldsymbol{\Phi}^{\mathbf{T}}$ as shown). Thus, the number of samples that need to be transmitted is reduced from N to M. The radio power, assuming that it scales with data rate, is then clearly reduced by N/M which significantly lowers the average sensor power and energy per sample as the radio power is dominant.

Another nice property regarding the CS framework is that the reconstruction algorithm is robust to noise. The intuition behind this is that if the noise is additive white Gaussian noise or even slightly colored, then it should not be sparse in the basis, $\boldsymbol{\Psi}$, whereas the signal will be. Thus, as results will show, the resolution of the sampled data, and hence the ADC, can be relaxed without significant detriment to the quality of the reconstruction. Since the input referred noise of the AFE should be designed to roughly track the input referred quantization noise of the ADC [3], lowering the required ADC resolution also relaxes the required input referred noise of the amplifier and reduces the power in both blocks.

IV. System and Circuit Implementation

The proposed CS system reduces energy consumption at the sensor node by shifting the system complexity (reconstruction) to the receiver which is presumably less energy constrained. What is not addressed in previous CS works [10, 11], however, is that there must also be an efficient implementation of the CS encoding at the sensor in order to fully realize these gains.

To facilitate an efficient yet flexible hardware realization capable of signal-agnostic data capture, the sensing matrix, $\boldsymbol{\Phi}$, was chosen to be a random Bernoulli matrix where each entry, $\Phi_{i,n}$, is +/-1. Since each output, y_i, of the CS block is just a linear combination of the ADC samples, $\boldsymbol{f}$, and all the entries in $\boldsymbol{\Phi}$ are +/-1, the CS transformation can be implemented with an accumulator per output where each incoming ADC sample is added or subtracted based on the current matrix entry. Thus, the number of columns, N, can be programmable by changing the number of samples to accumulate over, while the maximum number of rows, M, needs to be determined beforehand since one accumulator is needed per row. To avoid overflow, a 16-bit accumulator was used to provide a sufficient range of N values while M was chosen to be 50 to provide flexibility in exploring compression factors (N/M). The block diagram of the matrix-multiply operation along with the accumulator circuits are shown in Fig. 3. At every N^{th} sample, the pseudo-random bit sequence (PRBS) generators are re-seeded and the serializer block captures the state of the

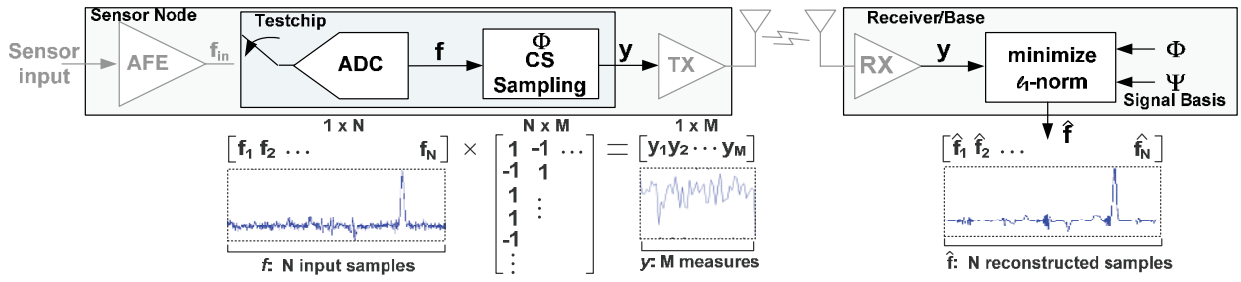

Fig. 2. Block diagram of the proposed sensor system architecture showing the equivalent mathematical function of the CS sampling matrix.

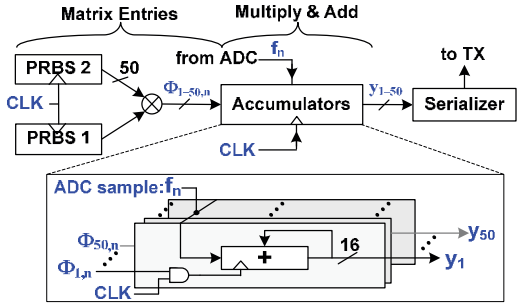

Fig. 3. Block diagram of the CS transformation and accumulator circuits.

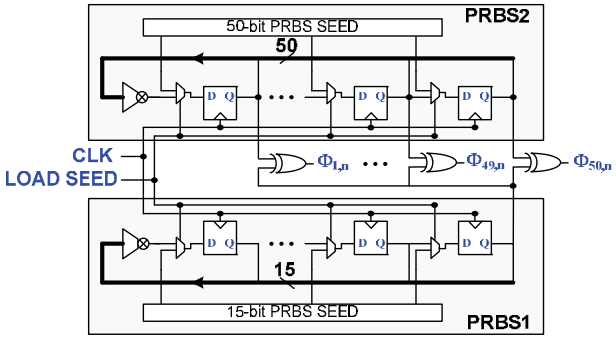

Fig. 4. Circuit for generating CS matrix entries.

accumulators so that continuous data acquisition can occur while previously captured data is either transmitted or stored.

To enable signal reconstruction, the "random" entries in Φ must be such that the rows are uncorrelated. A straightforward implementation using a separate PRBS generator per row would more than double the area and power of the CS block. Instead, the entire random matrix is generated from just two independent PRBS generators by XOR'ing the state of one PRBS with the current output of the other on a per sample basis. Fig. 4 shows the circuit used to generate the matrix entries, $\Phi_{i,n}$, where $i \in [1,M]$ and $n \in [1,N]$. The seed and sequence length of each PRBS is programmable to allow for the creation of different matrices.

We further halve the power cost by recoding the ADC output into two's complement format, so that the output is nominally zero, which allows gating of the accumulators on '-1' entries instead of performing subtractions. The resulting matrix has 1/0 entries and is functionally equivalent to its +/-1 counterpart. The only drawback is that the accumulators may saturate for a smaller N if the input has a large DC component.

V. Measurement Results

The testchip power, area and testing infrastructure are shown in Fig. 5. The testchip, which consists of an 8-bit SAR ADC and the CS block, uses a novel charge-sharing DAC scheme [12] to reduce the ADC area so that together the ADC and CS block consume only 200 x 550 µm in a 90 nm CMOS process. For the range of operating frequencies in medical applications, the full (M=50) CS block power consumes only 1.9 µW at 0.6 V and 20 kS/s, and is dominated by leakage. For testing, pre-recorded sensor signals were either driven into the ADC from an external DAC or passed directly into the CS block through an on-chip deserializer. The output of the ADC could be observed synchronously with the output of the CS to enable comparison between the sampled and reconstructed signals.

Fig. 6 shows an example of a continuous data acquisition for an N:M ratio of 1000:50; the input is continuously compressed from 1000 8-bit ADC samples into 50 16-bit accumulator

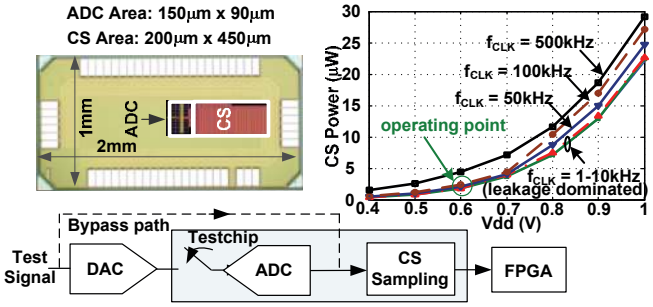

Fig. 5.CS area and power of the testchip and testing setup block diagram.

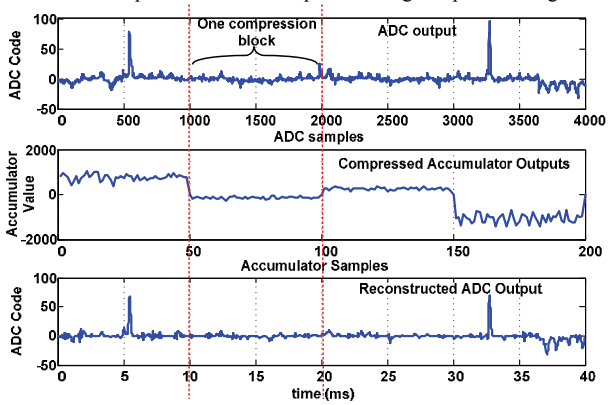

Fig. 6. Continuous data acquisition of an EEG signal showing the ADC output, compressed samples, and reconstructed waveform for N=1000, M=50.

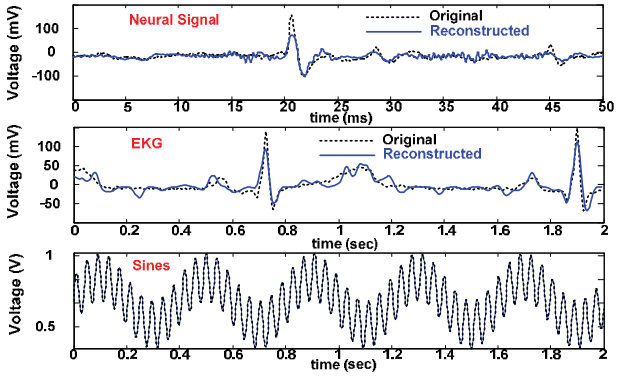

Fig.7. Original and reconstructed signals: a neural signal, EKG and sine waves.

outputs netting an effective compression factor of 10. In this example, a pre-recorded EEG signal [13] driven by the off-chip DAC is sampled, compressed and then reconstructed from an over-complete Gabor dictionary in which EEG signals have a sparse representation [8]. As Fig. 7 shows, the same procedure using different basis can be applied to other signal classes with similar results.

As Fig. 6 also shows, the reconstructed signal faithfully represents the key features of the original ADC output but can be lossy. The quality of the recovered signal depends on the signal sparseness, the compression factor N/M as well as the ADC and accumulator resolutions. In order to quantify the impact of each parameter as well as the effects of noise, an artificial noiseless reference EEG signal (f_{REF}) was created from over a half a dozen Gabor frames. The signal-to-noise-and-distortion ratio ($SNDR$) of the reconstructed waveform (f_{REC}) is then plotted for various compression factors and resolutions where $SNDR$ is defined as:

$$SNDR = \| f_{REF} \|_{l_2}^2 / \| f_{REF} - f_{REC} \|_{l_2}^2 \qquad (3)$$

To emulate input from lower resolution converters, bits from the ADC could be masked out in hardware. As Fig. 8 shows, there is little perceptual difference between the reconstructed

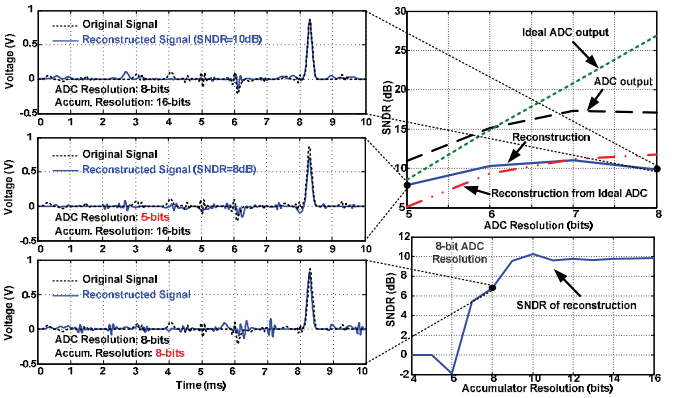

Fig. 8. SNDR of the original and reconstructed signals vs. ADC resolution and accumulator resolution with select accompanying waveforms.

signal from an 8-bit and 5-bit ADC. The same is true when the resolution of the accumulator is reduced to 8-bits by dropping the LSBs. Fig. 8 also shows the SNDR of the reconstructed waveform from an ideal ADC and from the ADC on the testchip. As the SNDR plots show, there is little difference between the reconstructed signal from the perfectly linear, idealized ADC and the non-linear ADC on the testchip. Interestingly, the on-chip ADC actually outperforms the ideal ADC at lower resolutions because of its non-uniform quantization. Collectively, these results show the robustness of the reconstruction process which relaxes the ADC resolution to 5-bits and allows the accumulator resolution to halve, thus improving the effective compression by a factor of two.

Since the maximum M directly impacts power and area costs, it is important to understand the tradeoffs as it scales as well. Fig. 9 plots the SNDR of the reconstructed signal over a range of N/M for several M. The plots show that compression factors at or below 40x enable good signal reconstruction for values of M above 20. Thus, the number of accumulators in the sampling matrix could be halved for signals with similar sparsity. It should be noted that for sparser signals, lower M, or higher compression factors (N/M) can be realized.

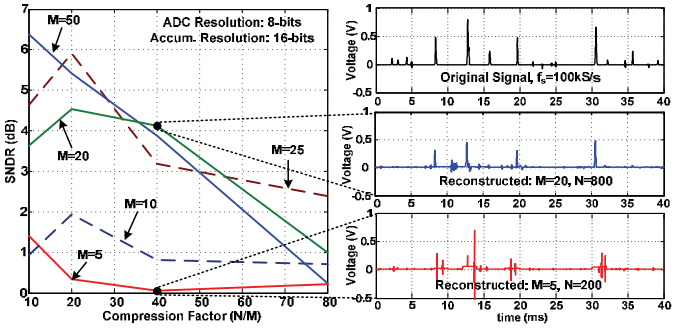

Fig. 9. *SNDR* of the reconstructed signal vs. compression factor (N/M).

TABLE II
AVERAGE SENSOR POWER PER CHANNEL WITH CS DATA COMPRESSION

	Previous Works				CS (This Work)	
	[2]	[3]	MICS [1]	UWB [14]	MICS	UWB
AFE	5.6μV_{rms}	8.9μV_{rms}	5.1μV_{rms}	4.9μV_{rms}	163μV_{rms}	78μV_{rms}
	8.6μW	68μW	42.4μW	6.6μW	< 0.1μW	< 0.1μW
ADC	10-bit	5-bit	10-bit	9-bit	5-bits	5-bits
	7.4μW	<1μW	<1μW	17.5μW£	< 0.1μW	1.1μW
Data Reduction	3-48x, 8-bit	12x, 5-bit	10x, 1-bit	0.6x, 16-bit	10-40x, 8b	10-40x, 8b
	122μW‡	75μW	24.8μW	--	0.8μW	0.8μW
TX (MICS/UWB)	5-66kb/s	8kb/s	3.3kb/s	640kb/s	**0.2-0.8kb/s**	**8-32kb/s**
	--	--	67.5μW	12.5μW	4-16μW	0.2-0.7μW
Total	138μW†	170μW†	135μW	37μW§	**5-17μW***	**2-3μW***

*The CS block is assumed to have a size of M=20, operate at 0.6V, and require 5-bit ADC and 8-bit accumulator resolution. The other circuits are assumed to have the same energy efficiency as the reference (e.g. the FOM of the ADC in MICS(CS) and MICS[1] is the same).
§Does not include power for multiplexing overhead £Includes buffers to drive ADC
†Does not include power for wireless transmitter. ‡Inferred from total power/channel

Table II demonstrates the impact on power that the proposed CS block would have on previous neural recording systems. The performance of the AFE, ADC and TX are assumed to be the same both with and without CS, only with relaxed specifications. The power of the ADCs and the input referred noise of the AFE are assumed to scale roughly as 2^B, where B is the ADC resolution, while AFE power scales roughly as 2^{2B}. Since the CS block is leakage limited, its power scales linearly with area, or M. Even at modest compression factors (10x) and for UWB systems where the radio power is not dominant [14], adoption of the CS architecture results in significant average power savings. The CS based system is even up to 10x lower power than wired systems [2, 3] and the energy cost of the CS block is comparable to feature extraction processors [4, 5].

VI. CONCLUSION

This work has presented a novel architecture and circuit implementation for data acquisition and compression in energy constrained wireless sensors. The proposed system enables recovery of the full signal while simultaneously compressing the amount of transmitted data and relaxing the noise and resolution constraints of the AFE and ADC. This circuit architecture demonstrates lower energy cost compared to existing data compression alternatives and enables continuous data capture, compression and transmission without requiring any general purpose memory or processing at the sensor. The proposed system is also generic enough to capture and compress data for any sparse signal, which is a departure from traditional compression techniques that are typically signal-specific.

ACKNOWLEDGMENTS

The authors would like to thank W. Wattanapanitch in Prof. Sarpeshkar's group for providing the recorded neural data.

REFERENCES

[1] R. Harrison, et al., "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE JSSC*, vol. 42, 2007.
[2] B. Gosselin and M. Sawan, "Circuits techniques and microsystems assembly for intracortical multichannel ENG recording," *IEEE Custom Integrated Circuits Conference*, pp. 97-104, 2009.
[3] R. Olsson and K. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 2796-2804, 2005.
[4] N. Verma, et al., "A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System," *IEEE Journal of Solid-State Circuits*, vol. 45, pp. 804-816, 2010.
[5] V. Karkare, et al., "A 130-µW, 64-Channel Spike-Sorting DSP Chip," *IEEE Asian Solid-State Circuits Conference*, vol. 1, pp. 289-292, 2009.
[6] D. Donoho, "Compressed Sensing," *Trans. on Information Theory*, 2006.
[7] S. Miaou and S. Chao, "Wavelet-Based Lossy-to-Lossless ECG Compression in a Unified Vector Quantization Framework," *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 539-543, 2005.
[8] S. Aviyente, "Compressed Sensing Framework for EEG Compression," *IEEE 14th Workshop on Statistical Signal Processing*, 2007.
[9] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, 2006.
[10] P. Baheti and H. Garudadri, "An ultra low power pulse oximeter sensor based on compressed sensing," *Body Sensor Networks*, 2009.
[11] J.N. Laska, et al., "Theory and Implementation of an Analog-to-Information Converter using Random Demodulation," *ISCAS*, 2007.
[12] F. Chen, et al., "An Area-efficient Switching Scheme for Charge-sharing DACs in SAR ADCs", (unpublished)
[13] Swartz Center for Computational Neuroscience, http://sccn.ucsd.edu
[14] M. Chae, et al., "A 128-Channel 6mW Wireless Neural Recording IC with On-the-Fly Spike Sorting and UWB Transmitter," *ISSCC*, 2008.