Design and Analysis of a Hardware-Efficient Compressed Sensing Architecture for Data Compression in Wireless Sensors

Fred Chen, *Member, IEEE*, Anantha P. Chandrakasan, *Fellow, IEEE*, and Vladimir M. Stojanović, *Member, IEEE*

***Abstract*—This work introduces the use of compressed sensing (CS) algorithms for data compression in wireless sensors to address the energy and telemetry bandwidth constraints common to wireless sensor nodes. Circuit models of both analog and digital implementations of the CS system are presented that enable analysis of the power/performance costs associated with the design space for any potential CS application, including analog-to-information converters (AIC). Results of the analysis show that a digital implementation is significantly more energy-efficient for the wireless sensor space where signals require high gain and medium to high resolutions. The resulting circuit architecture is implemented in a 90 nm CMOS process. Measured power results correlate well with the circuit models, and the test system demonstrates continuous, *on-the-fly* data processing, resulting in more than an order of magnitude compression for electroencephalography (EEG) signals while consuming only 1.9 μW at 0.6 V for sub-20 kS/s sampling rates. The design and measurement of the proposed architecture is presented in the context of medical sensors, however the tools and insights are generally applicable to any sparse data acquisition.**

***Index Terms*—Biomedical electronics, circuit analysis, compressed sensing, electroencephalography, encoding, low power electronics, sensors, wireless sensor networks.**

I. Introduction

OVER the past two decades, advancements in microelectronics have enabled relatively cheap, distributed sensor nodes capable of moderate scale sensing, data collection, computation and communication. In turn, wireless sensor networks have emerged as a research area that spans a broad range of applications from agriculture to health care. Although the applications are diverse, many of the technical challenges facing the field are similar. From the protocol layer down to the circuit level most of the challenges are related to the stringent energy constraints of each sensor node [1]. In most applications, whether because of cost or utility, there is a need for each sensor node to have a lifetime in the 10 year range or beyond. For example, even with a sensor lifetime of 10 years, a network with 4000 nodes, such as in a large office building, requires on average a battery changed per day [2]. Similarly, for patients who require implantable medical devices, limiting the frequency of replacing batteries both reduces costly surgeries and improves the quality of life. With the energy density of modern portable batteries in the range of 1 W-hr/cc, even a 10 year device lifespan requires the sensor to consume on the order of 10 μW of average power per cubic centimeter of battery volume.

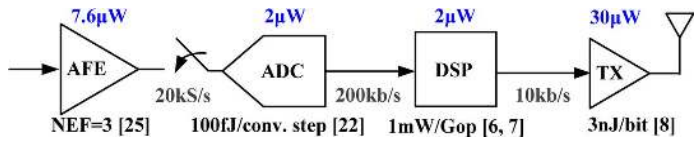


Fig. 1. Energy costs and power consumption for typical circuits in bio-sensor applications. It is assumed that the DSP filters some data and that the TX power scales with data rate.

Medical monitoring is an emerging application area that exemplifies the stringent energy constraints imposed on wireless sensor nodes and their corresponding circuits. Fig. 1 shows the typical circuit blocks used in sensors for medical monitoring and their associated energy cost and power consumption at a given sample rate. As Fig. 1 shows, the cost to wirelessly transmit data is orders of magnitude greater than for any other function. With the exception of ultra-wideband (UWB) radios, which have limited range and reliability issues, state-of-the-art radio transmitters exhibit energy-efficiencies in the nJ/bit range while every other component consumes at most only tens of pJ/bit. This cost disparity suggests that some data reduction strategy at the sensor node should be employed to minimize the energy cost of the system. In applications such as implantable neural recording arrays, the high energy cost to transmit a bit of information and the radio's limited bandwidth actually necessitate data compression or filtering at the sensor in order to reduce both energy consumption and data throughput [3].

Existing strategies for implementing integrated data compression or filtering solutions under these constraints largely revolve around detecting and extracting specific signal data [3]–[7]. However, the filtered data often contains limited information. For example, in neural recorders, the data is typically limited to just the time and amplitude of a neural spike event rather than the signal itself [3], [5]. Even when the event detection is used to trigger a full signal capture [4], 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 typically provide only macro level decisions based on the original signals [6], [7]. For these signal processing strategies, there is a tradeoff between data reduction, robustness, implementation cost, and

Manuscript received March 19, 2011; revised August 16, 2011; accepted September 18, 2011. Date of current version February 23, 2012.This paper was approved by Associate Editor Roland Thewes.

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: fredchen@mit.edu).

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/JSSC.2011.2179451

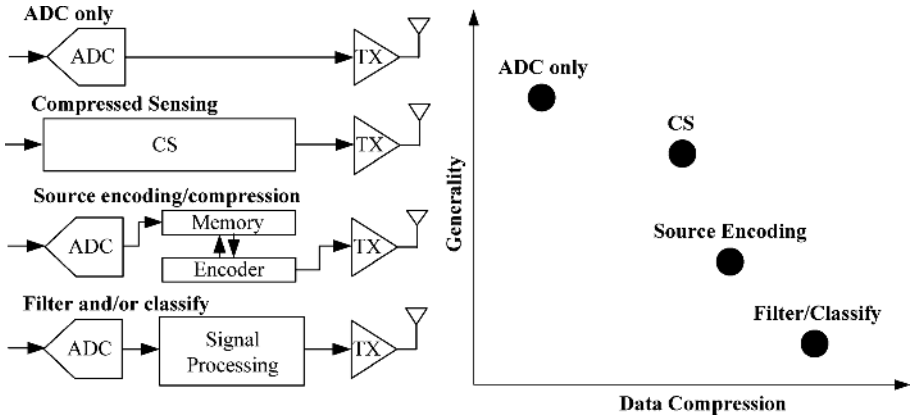


Fig. 2. Relative merits of CS compared to different data acquisition and compression approaches used in wireless sensors.

the granularity of information captured. In each case, the goal is to minimize the number of bits transmitted (to minimize the average radio power) while reliably preserving the signal information at a minimum implementation cost.

In this work, we introduce the design and implementation of a sensor architecture based on the theory of compressed sensing that offers an improved set of tradeoffs toward achieving this goal. As Fig. 2 shows, a CS based sensor system combines the positive qualities of existing data acquisition and compression systems: it provides a flexible and general interface like an analog-to-digital converter (ADC), yet still enables data compression proportional to the signal information content, which is consistent with the performance of source coding. For wireless sensor applications, this combination of characteristics is particularly attractive as it would enable a single hardware interface across many applications while simultaneously addressing the energy cost of wireless transmission. Traditional data acquisition architectures have been based on the principles of Shannon's sampling theorem which requires that the sampling rate must be greater than twice the maximum frequency of the signal being sampled. Compressed sensing is an emerging field whose theory leverages known signal *structure* to acquire sampled data at a rate proportional to the information content rather than the frequency content of a signal [9]. In theory, this would enable far fewer data samples than traditionally required when capturing signals with relatively high bandwidth, but a low information rate. As shown in Table I, many biophysical signals of interest fall into this category where their required sampling rates far exceed the information rate (frequency of event occurrences). Although these examples are in the context of medical applications, they can be generally applied to any field where the signals of interest are sparse.

To demonstrate the practicality of the proposed system, a CS encoder [10] is designed and fabricated in a 90 nm CMOS process based on circuit modeling and power analysis tradeoffs discussed in the remainder of the paper. Section II begins by providing background on CS theory and addresses its applicability to data compression. Section III specifies the hardware parameters of the CS framework that are used to compare implementation costs. Based on these parameters, Sections IV and V develop the circuit-level power/performance cost models for implementing the CS framework in the analog and digital domains.

TABLE I
CHARACTERISTICS OF COMMON MEASURED BIO-SIGNALS [4]

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	200 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

Section VI then analyzes the implementation tradeoffs and describes the actual system implementation. Section VII presents measurement results where the system demonstrates the ability to blindly encode EEG signals at a high compression factor, and shows that developed circuit cost models correlate well with the experiment. Finally, Section VIII discusses more generally the possible extensions of the power model and the CS architecture before concluding the paper.

II. COMPRESSED SENSING BACKGROUND

In this section, we provide an overview of the basic principles of CS and the relevance of each principle to the proposed sensor system.[1] CS is based on the following 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, $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \boldsymbol{\psi}_2 \ldots \boldsymbol{\psi}_{\mathrm{L}}]$ such that $\boldsymbol{f} = \boldsymbol{\Psi x}$ or equivalently:

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

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

[1]The background provided is only meant to give sufficient context under which the proposed system hardware design can be discussed. For a more theoretical and thorough background on compressed sensing, please refer to [9], [11], [15].

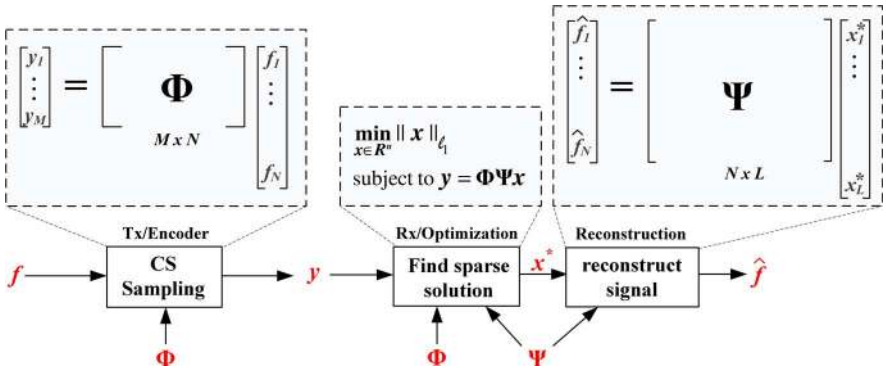


Fig. 3. CS sampling framework where $\boldsymbol{f}$ is the signal being sampled, $\boldsymbol{y}$ is the set of compressed measurements, $\boldsymbol{\Phi}$ is the measurement matrix, $\boldsymbol{\Psi}$ is the signal basis under which $\boldsymbol{f}$ is sparse, $\boldsymbol{x}^*$ is the resulting coefficient vector when finding the sparse solution and $\hat{\boldsymbol{f}}$ is the reconstructed signal.

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 captured in some other basis. So sparseness also implies compressibility and *vice versa*. A familiar example of such a signal is a sine wave which requires many coefficients in time to represent, but requires only one non-zero coefficient in the Fourier domain. Fortunately, many sensor signals such as the bio-signals from Table I have sparse representations in either the Gabor or wavelet domains [12], [13] thus making them suitable for data compression using 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. This generalized sampling framework is shown in Fig. 3 where the N-dimensional input signal, $\boldsymbol{f}$, is encoded into an M-dimensional set of measurements, $\boldsymbol{y}$ through a linear transformation by the $M \times N$ measurement matrix, $\boldsymbol{\Phi}$, where $\boldsymbol{y} = \boldsymbol{\Phi f}$. When $M < N$ such that the system is underdetermined, there are an infinite number of feasible solutions for $\boldsymbol{f}$. However, when the signal to be recovered is known to be sparse in some basis, $\boldsymbol{\Psi}$ where $\boldsymbol{f} = \boldsymbol{\Psi x}$, then the sparsest solution (fewest significant non-zero x_i) out of the infinitely possible is often the correct solution. A common and practical approach to find the sparse solution is to solve the following convex optimization problem:[2]

$$\min_{\boldsymbol{x} \in \boldsymbol{R}^n} ||\boldsymbol{x}||_{\ell_1} \text{ subject to } \boldsymbol{y} = \boldsymbol{\Phi\Psi x} \tag{2}$$

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 to (2). The problem of minimizing the ℓ_1-norm in (2) has been shown to be solved efficiently [14]. Its feasibility implies that an N-dimensional signal can be recovered from a lower order number of samples, M, provided that the signal is sparse under some basis. We rely on this result to reduce the data that the sensor must transmit; the ratio N/M is essentially the data compression factor (CF) realized by the CS system and is proportional to the radio power that would be saved.

[2]In practice there are many viable approaches to finding the "sparse" solution for the underdetermined system of equations described by $y = \boldsymbol{\Phi\Psi}x$. We simply present ℓ_1 minimization as one approach whose complexity is known to be tractable and thus demonstrates the practicality of the reconstruction process.

C. Incoherent Sampling

In addition to sparseness, CS also relies on incoherence between the sensing modality $(\boldsymbol{\Phi})$ and the signal model $(\boldsymbol{\Psi})$ to minimize the number of measurements (M) needed to recover the signal. Coherence measures the largest correlation between any row of $\boldsymbol{\Phi}$ and column of $\boldsymbol{\Psi}$ and can be defined by the operator μ as

$$\mu(\boldsymbol{\Phi\Psi}) = \max_{k,j} |\langle \phi_k, \psi_j \rangle| \tag{3}$$

where $\mu^2(\boldsymbol{\Phi\Psi})$ can range between 1 and N [15]. The less coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, the fewer the number of measurements needed to recover the signal. A lower bound on the number of measurements needed to recover the overwhelming majority of terms in an S-sparse signal (a signal with S significant non-zero terms out of N in the basis $\boldsymbol{\Psi}$) was shown to be

$$M \geq C \cdot \mu^2(\boldsymbol{\Phi\Psi}) \cdot S \cdot \log N \tag{4}$$

where C is a small known constant (empirically $\sim 2-2.5$ [16]) and N is the dimensionality of the signal to be recovered [15]. Since S is a measure of the information in the signal, this lower bound shows that the number of samples required to recover a signal in a CS framework is proportional to the information content of the signal.

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 [17]. As suggested in [17], this means that a random sensing matrix can be employed as a universal encoder and 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 a generic infrastructure for data acquisition and compression that is agnostic to the type of signals being acquired, provided that they are sparse.

III. CS Implementation Parameters

In order to improve the energy efficiency of the system, the overhead to process and compress the data cannot outweigh the energy savings gained at the transmitter from data reduction. Thus, for CS to be a practical solution for wireless sensor

nodes, an energy efficient hardware implementation of the encoder must be realized. Given the relative immaturity of the field, there have been few works that discuss the tradeoffs or costs of realizing CS in hardware [18]–[20] and even fewer measured results [20].

The CS parameters described in Section II can be translated into a set of required hardware specifications. As shown in Fig. 3, the CS encoder essentially amounts to performing a linear projection from the N-dimensional input, $\boldsymbol{f}$, to an M-dimensional set of measurements, $\boldsymbol{y}$, using the matrix, $\boldsymbol{\Phi}$. In the context of data compression, this amounts to transforming every block of N samples of $\boldsymbol{f}$ into M measurements ($\boldsymbol{y}$). We define B_f and B_y as the bits needed to represent the dynamic range of each sample in $\boldsymbol{f}$ and $\boldsymbol{y}$ respectively. Thus, the effective compression factor is $(\mathrm{N}\cdot B_f)/(\mathrm{M}\cdot B_y)$.[3] A common approach to facilitate an efficient hardware implementation of $\boldsymbol{\Phi}$ is to use a pseudo-random Bernoulli matrix where each entry, $\Phi_{m,n}$, is ± 1 [18], [19]. This minimizes the size of $\boldsymbol{\Phi}$ and subsequent matrix-multiply operations by representing each matrix entry with only a single bit. Any other choice of a full rank $M \times N$ matrix would result in additional circuit complexity, data storage, and computation requirements.

As with traditional signal processing algorithms, the CS encoding can be implemented in either the analog or digital domains. In early proposed applications of CS, the linear projection was applied in the analog domain prior to digitization either because the dominant consumer of power was the sensing mechanism [21] or to reduce the required sampling frequency of the ADC [18], [19]. However, unlike previously proposed applications for CS, wireless sensor applications are rarely limited by ADC performance. Thus, the next two sections will model the dependencies and costs of both systems. In an effort to provide a level comparison between these implementations, the analysis of circuit costs are described in terms of the required system parameters: N, M, B_f, B_y, and the signal bandwidth (BW_f) in Hertz.

IV. Analog CS Encoder Power Model

Fig. 4 shows the block diagram and example circuits for an analog implementation similar to those described in [18] and [19]. In the circuits shown, the input is amplified through an operational transconductance amplifier (OTA) while the multiplication is realized with a double-balanced passive mixer. The sample-and-hold (S/H) circuit following the mixer acts as an integrating (summing) stage as well as the S/H input to the ADC.[4] Although there are many possible alternative circuit realizations, the example provided is representative of how hardware costs in this architecture will scale.

[3]Note that in the remainder of the text, we will commonly refer to the ratio N/M as the compression factor (CF) to highlight dependencies on compression performance and resolution. In practice, the required value for B_y scales with B_f such that $(\mathrm{N}\cdot B_f)/(\mathrm{M}\cdot B_y)$ does not vary much over resolution such that N/M is representative of the compression factor.

[4]A reset switch to a common mode voltage and at least one more S/H circuit (not shown) need to be time multiplexed with the circuit shown to enable continuous integration of the input while the ADC quantizes the previously integrated sample.

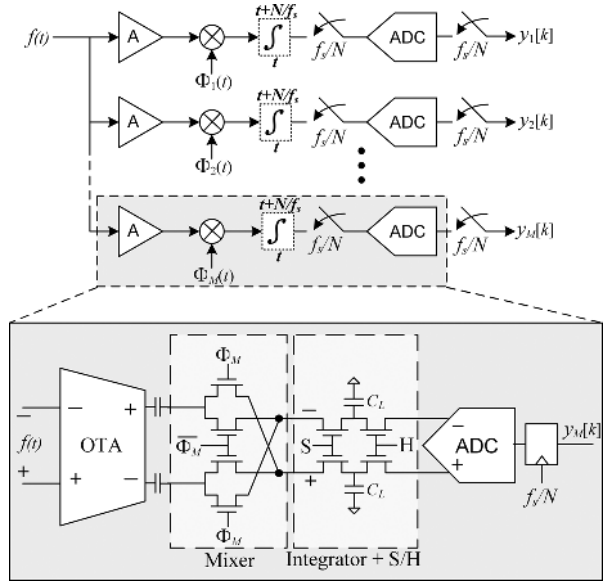


Fig. 4. Block diagram and example circuitry for an analog implementation of the CS linear transformation. The passive mixer is driven by the matrix coefficients at a rate of f_s. During the sample phase $(S = 1)$, the sample-and-hold (S/H) circuit also acts as a passive integrator.

A. Analog-to-Digital Converter

In the architecture shown in Fig. 4, the matrix coefficients, $\Phi_\mathrm{i}(t)$, need to be applied at the Nyquist frequency, f_s, of the signal or higher in order to avoid aliasing [19]. However, the sampling frequency of each ADC only needs to be f_s/N where $f_\mathrm{s} > 2\mathrm{BW}_f$ and N is the number of integration samples per compression block. The output of each ADC produces one measurement result, $y_\mathrm{i}[k]$, so the resolution of the ADC should be equal to the required measurement resolution, B_y. The resulting power of the array of M ADCs is then

$$P_\mathrm{ADC} = (M/N)\cdot \mathrm{FOM}\cdot 2^{B_y}\cdot f_s \tag{5}$$

where the figure-of-merit (FOM) of the ADC is a design specification. In subsequent analysis, the FOM used to show tradeoffs is 100 fJ/conversion step, which is consistent with the general performance of modern ADCs over a wide range of resolutions and sampling speeds [22].[5]

B. Integrator and Sample/Hold

The simplified Norton and Thevenin equivalent noise circuits in Fig. 5 show that the constraints on the sampling circuit are partially dictated by the mixer and OTA. When the sampling circuit is tracking the input, the noise bandwidth of the system is set by the sampling capacitor and the series resistance of the CMOS switch (R_sw), mixer (R_mix) and the OTA output resistance (R_o). For practical purposes, R_o should be dominant to insure that the OTA looks like a current source and so that the combined circuit acts like an integrator where the appropriate noise model more closely resembles the Norton equivalent model. As described in [19], if the S/H is assumed to be a perfect integrator, then the frequency response of the integrator is a sinc

[5]For sensor applications, it is assumed that the required resolution and bandwidth of the ADC are low enough so that the ADC efficiency is not noise limited such that the ADC power will scale 2X with resolution rather than 4X (i.e., the FOM will stay constant as the performance requirements scale).

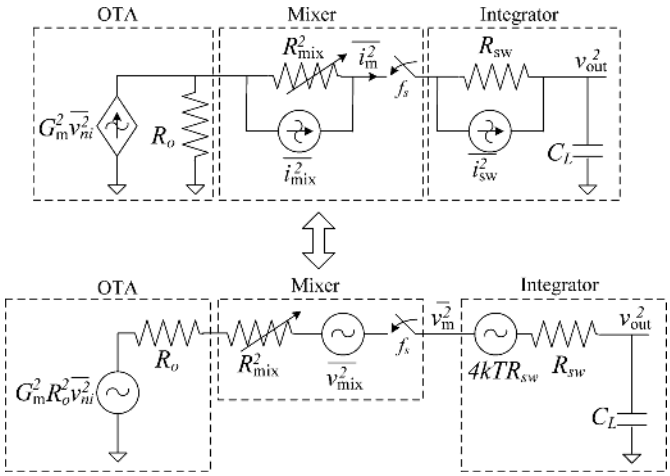


Fig. 5. Simplified Norton and Thevenin equivalent noise models for the OTA, mixer and integrator.

pulse where the gain (G_I) and noise bandwidth (BW_N) of the integrator can be expressed as

$$G_I^2\mathrm{BW}_N = \int_0^\infty |H_i(f)|^2 df = \underbrace{\left(\frac{N}{f_s \cdot C_L}\right)^2}_{\text{gain}} \cdot \underbrace{\left(\frac{f_s}{2N}\right)}_{\text{bandwidth}} \tag{6}$$

where the integration period is N/f_s. However, to the extent that R_o is not infinite then the equivalent noise model moves closer to the Thevenin equivalent model where the noise bandwidth is given by the low-pass filter response over a finite integration window:

$$\begin{aligned} G_I^2\mathrm{BW}_N &= \int_0^\infty |H_i(f)|^2 df \\ &= \underbrace{R_o^2}_{\text{gain}} \cdot \underbrace{\frac{1}{4R_oC_L}(1 - e^{-2N/R_oC_Lf_s})}_{\text{bandwidth}} \end{aligned} \tag{7}$$

where it is assumed that the equivalent resistance seen by the capacitor is dominated by R_o. The circuit properly approximates an ideal integrator for integration periods where $N/f_s < 0.1R_oC_L$. The bandwidth of the unloaded OTA, assumed to be set by a single dominant pole $1/(2\pi R_oC_p)$, should at least match the required bandwidth of the signal, $\mathrm{BW}_f = f_s/2$, so the lower bound on the size of the integrating capacitor to functionally act as an integrator can be described by[6]

$$C_L > \frac{10 \cdot N_{\max}}{R_o f_s} = 10\pi \cdot N_{\max} \cdot C_p \tag{8}$$

where C_p is the capacitance at the dominant pole, and $N_{\max}$ is the maximum number of samples to compress. The power due to switching the integrator and S/H circuits is then modeled by

$$P_i = \frac{M}{N} \cdot V_{\mathrm{DDA}}^2 \left(\frac{C_L}{16} + C_G\right) \cdot f_s \tag{9}$$

where C_G is the total gate capacitance of the switches. In (9) it is assumed that the single-ended voltage swing is between $1/4V_{\mathrm{DDA}}$ and $3/4V_{\mathrm{DDA}}$, and that the common mode reset voltage is at $1/2V_{\mathrm{DDA}}$. Even if C_p is unrealistically assumed to consist of only parasitics and wiring, a reasonably useful value of $N_{\max} = 100$ would still require C_L to be on the order of 3 pF in most modern processes. Thus, the power attributed to switching the switches themselves is negligible compared to C_L.

C. Mixer

The passive mixer shown in Fig. 4 is described in [23] where it is shown to have a theoretical voltage conversion gain (G_C) ranging between -3.92 dB and -2.1 dB and a measured noise figure (NF) of 3.8 dB. The primary impact of the mixer performance is its impact on the specifications for the OTA. For a $G_c = -3$ dB and a NF $=$3.8 dB, the current noise density at the output of the mixer, $\overline{i_m^2}$, is then

$$\overline{i_m^2} = \overline{i_{\mathrm{OTA}}^2} \cdot \mathbf{10}^{(G_C/2+\mathrm{NF})/10} = \mathbf{1.7} \cdot \overline{i_{\mathrm{OTA}}^2} \tag{10}$$

where $\overline{i_{\mathrm{OTA}}^2}$ is the noise current density out of the amplifier (into the mixer). For a pseudo-random bit sequence (PRBS) of N samples, the resulting noise accumulated during an integration window is N times the output noise of a single sample, where the output noise density is filtered by the gain and effective noise bandwidth of an integrator with $1/N$th the integration period. The total integrated output noise needs to be less than the quantization noise of the ADC leading to[7]

$$v_{\mathrm{out,rms}}^2 \simeq 1.7 \cdot \overline{i_{\mathrm{OTA}}^2} \cdot G_I^2\mathrm{BW}_N \leq \frac{V_{\mathrm{DDA}}^2}{12 \cdot 2^{2B_y}}. \tag{11}$$

D. OTA

The lower bound on power consumption in the amplifier is typically set by the input referred noise ($v_{ni,\mathrm{rms}}$) requirement. A figure of merit that captures the relationship between $v_{ni,\mathrm{rms}}$ and power consumption in the amplifier is the noise efficiency factor (NEF) which was first introduced in [24] and captures the effective number of transistors contributing noise:

$$\mathrm{NEF} = v_{ni,\mathrm{rms}}\sqrt{\frac{2I_{\mathrm{amp}}}{\pi \cdot V_T \cdot 4kT \cdot \mathrm{BW}_{\mathrm{amp}}}} \tag{12}$$

where I_{amp} is the total amplifier current, V_T is the thermal voltage (kT/q) and $\mathrm{BW}_{\mathrm{amp}}$ is the bandwidth of the amplifier. Measured NEFs in state of the art low-noise amplifiers fall in between 2 and 3 [25]–[27]. For future analysis, a NEF of 3 will be used[8] and the required power for the array of amplifiers can then be calculated by rewriting (12) as

$$\begin{aligned} P_{\mathrm{amp}} &= M \cdot V_{\mathrm{DDA}} I_{\mathrm{amp}} \\ &\geq M \cdot V_{\mathrm{DDA}} \cdot \frac{(\mathrm{NEF})^2}{v_{ni,\mathrm{rms}}^2} \cdot \frac{\pi \cdot V_T \cdot 4kT \cdot \mathrm{BW}_f}{2} \end{aligned} \tag{13}$$

[6]To the extent that the integrator is non-ideal will essentially introduce errors in the matrix entries (weights of each input sample) and require a means to back out the actual matrix applied as in [19].

[7]The noise term due to the sampling switch of the S/H circuit has been omitted since it will be insignificant for any practical values of R_{sw} and R_o. The mixer power is dominated by the clocking and logic to generate the sequence of matrix coefficients, $\Phi_{\mathrm{i}}(t)$, which is discussed later.

[8]In practice, a realistic NEF for each application must be determined in order to properly weigh the costs. For the purpose of analysis, the NEF chosen is on the low-end of what has been generally published in state-of-the-art amplifiers for bio-applications to establish a lower bound on the amplifier power.

The output noise constraint in (11) can then be rewritten in terms of $v_{ni,\mathrm{rms}}$ such that[9]

$$\underbrace{\left(0.92\sqrt{N}\cdot\frac{G_m}{f_S C_L}\right)^2}_{G_A}\cdot\underbrace{\overline{v_{ni}^2 f_s}}_{v_{ni,\mathrm{rms}}^2}\leq\frac{V_{\mathrm{DDA}}^2}{12\cdot 2^{2B_y}} \tag{14}$$

where G_A is the total voltage gain from the input of the amplifier to the input of the ADC. The required value of G_A varies by application and the expected dynamic range of the input signal, but a common specification used in previously published bio-sensor applications is 40 dB [3], [25], [27]. This constraint, however, assumes that the total gain is set such that the input range of the ADC is perfectly accommodated. Since we are integrating over N samples, the instantaneous voltage (variance) on the integrator can be expected to grow by a factor of $\sqrt{N}$ and cannot be allowed to exceed the available headroom such that $v_{\mathrm{in}}G_A\sqrt{N}\leq 2V_{\mathrm{DDA}}$. This constraint reduces the available headroom and thus the required noise floor for a given resolution. Combining this additional constraint with (13) and (14) results in the minimum amplifier power required:[10]

$$P_{\mathrm{amp}}=2\mathrm{BW}_f\cdot 3M\cdot N\cdot 2^{2B_y}\cdot\frac{G_A^2\mathrm{NEF}^2}{V_{\mathrm{DDA}}}\cdot\frac{\pi(kT)^2}{q}. \tag{15}$$

E. Analog CS Encoder Power

The total power for the analog implementation of the CS encoder, excluding the matrix generation and mixer (multiply) cost, can be summarized as

$$\begin{aligned}P_{\mathrm{CS},a}=2\mathrm{BW}_f\Bigg[&\underbrace{\frac{M}{N}\cdot\mathrm{FOM}\cdot 2^{B_y}}_{\mathrm{ADCs}}\\&+\underbrace{\frac{M\cdot V_{\mathrm{DDA}}^2\cdot 10\pi\cdot N\cdot C_p}{16}}_{\mathrm{integrators}}\\&+\underbrace{3M\cdot N\cdot 2^{2B_y}\cdot\frac{G_A^2(\mathrm{NEF})^2}{V_{\mathrm{DDA}}}\cdot\frac{\pi(kT)^2}{q}}_{\mathrm{amplifiers}}\Bigg].\end{aligned} \tag{16}$$

As expected, the costs of all components scale with the number of measurements, M, but they are also dependent on the input signal bandwidth. So even if the number of samples in the CS framework is independent of the signal bandwidth, the cost to implement the circuits is not.

[9]In (11), the noise bandwidth of the OTA is simplified from $(\pi/4)f_s$ to f_s which implies $\sim$25% margin on the required bandwidth to accommodate the signal.

[10]The instantaneous voltage can be allowed to be as large as $2V_{\mathrm{DD}}$ since we are assuming a differential system where V_{DDA} corresponds to the differential ADC input range. The resulting power gain then becomes $\mathrm{N}\cdot\mathrm{G}_\mathrm{A}^2$ instead of G_A^2.

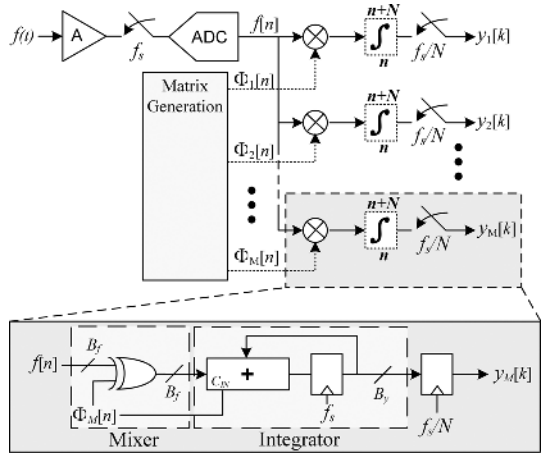


Fig. 6. Block diagram and circuitry for a digital implementation of the CS encoder.

V. Digital CS Encoder Power Model

Fig. 6 shows the block diagram and circuits for an equivalent digital implementation of the CS encoder. The input signal is first amplified and then digitized by a single ADC sampling at the Nyquist rate, f_s. The ADC output is passed to M parallel accumulators that accumulate the incoming sample based on their respective sequence of matrix coefficients, $\Phi_{\mathrm{i}}[n]$. Recall that the coefficient matrix is a Bernoulli random matrix where all elements are ± 1. Thus, the multiplication function can be simply implemented with an XOR gate and the *carry-in* input of the accumulator. The output of the accumulator is then captured every N samples at which time the accumulator is reset.

A. Accumulator and XOR

Each measurement, $y_{\mathrm{i}}[k]$, requires an accumulator with at least B_y bits of resolution which results in B_y flip-flops and XORs, and a B_y-bit adder. In order to model the delay and energy costs associated with these circuits, a logical effort (LE) [28] model is adopted to determine the sizing of each gate and the methodology for sizing the adder is similar to [29]. A slightly simplified version of the alpha-power law delay model used in [30] is used to map the normalized delay of the LE model to real delay. The LE delay of the accumulator is used to scale V_{DD} until the timing constraint is just met, resulting in the following minimum operating V_{DD}:

$$V_{\mathrm{DD,MIN}}\simeq\frac{\alpha_d V_{\mathrm{th}}}{1-2.5\cdot K_d\cdot(D_{FF}+D_{\mathrm{ADD},B_y})\cdot f_s} \tag{17}$$

where K_d and α_d are technology fitting parameters and D_{FF} and $D_{\mathrm{ADD},By}$ are the LE delay of the flip-flop and the critical path of a B_y-bit adder.[11] The dynamic power consumption

[11]The adder topology chosen is a ripple carry adder. For sensor applications, the resolution and speed of the adder are such that the circuit power will likely be dominated by leakage. Under this assumption, it is appropriate to choose the most compact adder topology which is a ripple carry adder. If the operating conditions change then the subsequent analysis can be adopted to other adder architectures.

can be calculated by accounting for all of the gate and parasitic capacitances at each node along with the switching activity at those nodes.[12] So for MB_y-bit accumulators and XORs operating at $V_{\mathrm{DD,MIN}}$ this results in a dynamic switching energy of

$$P_{\mathrm{accum,dyn}} = M \cdot [27 \cdot (B_y + \log_2 \sqrt{N}) + 2] \cdot f_s \cdot V_{\mathrm{DD,MIN}}^2 \cdot C_{\mathrm{inv}} \quad (18)$$

where C_{inv} is the capacitance of the reference inverter. The $\log_2 \sqrt{N}$ term is added to avoid overflow in the accumulators during integration and is analogous to the headroom constraint reflected in (15) of the analog model. Unlike the analog integrator, the dynamic range of the digital accumulator can be expanded by extending the headroom rather than lowering the noise floor such that it does not impact the noise or resolution requirements of the OTA and ADC.

One component of energy consumption that LE does not explicitly model is the sub-threshold leakage current. To account for leakage, an additional normalized parameter is added that captures the relative leakage current in each gate compared to the reference inverter.[13] Similar to how the normalized delay in LE was used to model delay, we use the normalized leakage parameter to arrive at a power consumption expression due to leakage:

$$P_{\mathrm{accum,leak}} = M \cdot [22.5 \cdot (B_y + \log_2 \sqrt{N}) + 4] \cdot V_{\mathrm{DD,MIN}} \cdot I_{\mathrm{leak,ref}}(V_{\mathrm{DD,MIN}}) \quad (19)$$

where $I_{\mathrm{leak,ref}}(V_{\mathrm{DD,MIN}})$ is the leakage of the reference inverter at a supply voltage of $V_{\mathrm{DD,MIN}}$.

B. ADC and Amplifier

The constraints on the ADC and amplifier for the digital CS encoder system are similar to those in the analog system discussed earlier. For the ADC, it is now dependent on signal resolution (B_f) instead of measurement resolution (B_y) and samples at the Nyquist rate such that:

$$P_{\mathrm{ADC},D} = \mathrm{FOM} \cdot 2^{B_f} \cdot f_s \quad (20)$$

Similarly for the amplifier, the noise constraint on the amplifier is now only determined by the quantization noise of the ADC such that

$$P_{\mathrm{amp},D} = G_A^2(\mathrm{NEF})^2 \cdot \frac{12 \cdot 2^{2B_f}}{V_{\mathrm{DDA}}} \cdot \frac{2 \cdot \pi(kT)^2 \cdot \mathrm{BW}_f}{q} \quad (21)$$

with the same assumptions regarding G_A and NEF as before.

[12]The wires for the accumulator bank are all local and as such are lumped in with the parasitic portion of the LE delay model. Furthermore, when the design is leakage dominated, there is relatively no impact from wire estimation errors.

[13]The simplifying assumptions made in determining the leakage parameter are that the NMOS and PMOS leakage in the reference inverter are the same, and that the leakage current scales linearly with gate width and the number of off branches. The probability of the gate being in a certain leakage state is also taken into consideration.

C. Digital CS Encoder Power

The total power for the digital implementation of the CS encoder, excluding the matrix generation cost, can be summarized in (22) where $B_y^* = B_y + \log_2 \sqrt{N}$.

$$P_{\mathrm{CS},D} = 2\mathrm{BW}_f \cdot \left[\underbrace{\mathrm{FOM} \cdot 2^{B_f}}_{\mathrm{ADC}} + \underbrace{12 \cdot 2^{2B_f} \cdot \frac{G_A^2(\mathrm{NEF})^2}{V_{\mathrm{DDA}}} \cdot \frac{\pi(kT)^2}{q}}_{\mathrm{Amplifier}} \right] + M \cdot V_{\mathrm{DD,MIN}} \left[\underbrace{(22.5B_y^* + 4) \cdot I_{\mathrm{leak,ref}}(V_{\mathrm{DD,MIN}})}_{\text{leakage current}} + \underbrace{(54B_y^* + 4) \cdot \mathrm{BW}_f \cdot V_{\mathrm{DD,MIN}} \cdot C_{\mathrm{inv}}}_{\text{switching current}} \right]. \quad (22)$$

VI. CS Implementation

For wireless sensor applications, systems are typified by low sampling frequencies, medium resolutions and small amplitude input signals. Since the purpose of the CS encoder is for data compression, a desirable target is for 10X compression. Based on (4), compression block lengths between 100 to 1000 samples require roughly 11–17 measurements per significant term to recover the signal. Minimally, the system should be designed to recover a 1-sparse signal, but a more practical choice is to build in margin. Thus, to reconstruct 3–4 significant terms per block requires the following range of specifications for the system: $M \sim 50$, $N > 500$, $\mathrm{BW}_f = 0.1{-}10$ kHz, $B_f = 8$, $B_y > 8$, and $G_A > 100$ (40 dB).

A. Analog CS Versus Digital CS

To determine which implementation is most suitable for wireless sensor applications, the power models developed in Sections IV and V are used to map the power costs based on technology parameters extracted from the 90 nm CMOS process intended for the test chip fabrication. These results are captured in Fig. 7 which plots the relative power $(P_{\mathrm{CS},a}/P_{\mathrm{CS},D})$ of the analog CS encoder versus the digital CS encoder over the range of specifications. To help visualize this multi-dimensional design space, each sub-plot [Fig. 7(a)–(c)] captures the dependencies across only two of the three most sensitive parameters: N, G_A, and B_y. The remaining parameters, when not swept, are kept at a specification of $M = 50, \mathrm{N} = 500, G_A = 40$ dB, $B_f = 8, B_y = 10$, and $\mathrm{BW}_f = 200$ Hz, which is shown on each plot as the target specification. The general conclusion that can be drawn from the plots in Fig. 7 is that the digital implementation is more efficient at higher signal gains (G_A),[14]

[14]The choice of gain in the system is not arbitrary but rather a reflection of the magnitude of the input signal relative to the full scale range of the ADC. If the gain is too low (high) then the amplified signal may underutilize (saturate) the range of the ADC, and thus achieve a lower effective resolution. The calculations shown assume that the system is appropriately designed to accommodate the maximum expected input.

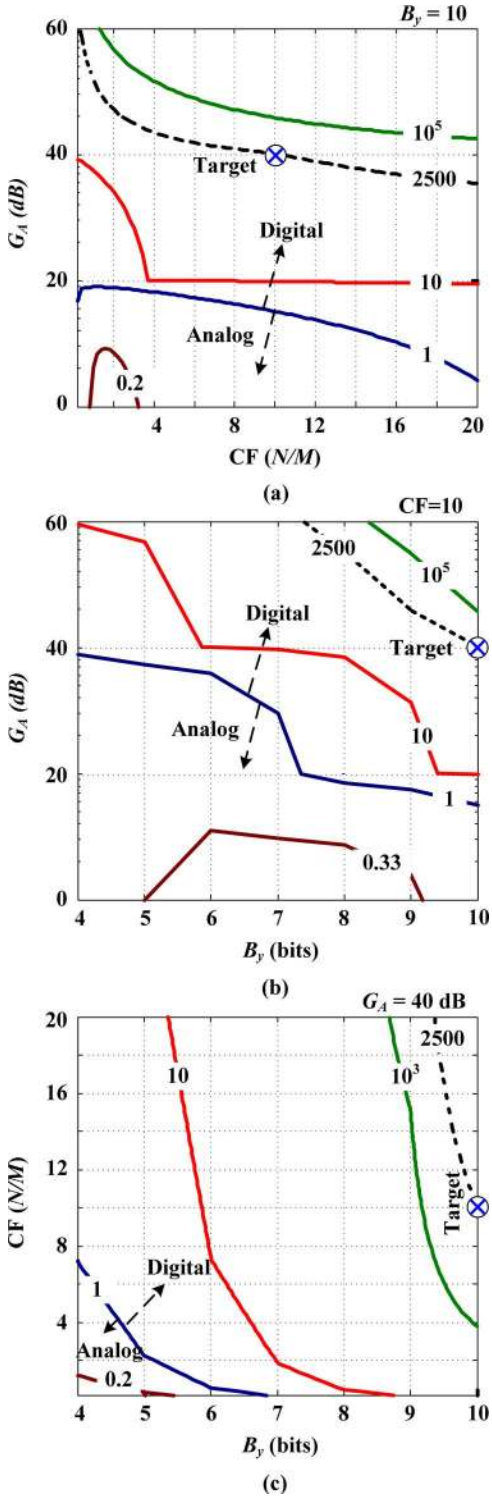


Fig. 7. Relative power cost of analog vs. digital CS encoder implementations ($P_{\mathrm{CS},a}/P_{\mathrm{CS},D}$) across the specification space for (a) the compression factor ($\mathrm{CF} = N/M$) and amplifier gain (G_A), (b) the measurement resolution (B_y) and G_A, and (c) B_y and CF. In each plot M is fixed so the CF is really a sweep of N. The targeted specification of $M = 50, N = 500, G_A = 40$ dB, $B_f = 8, B_y = 10$, and $\mathrm{BW}_f = 200$ Hz is shown on each plot along with its corresponding cost contour. All power calculations are based on the 90 nm CMOS process used for fabrication.

compression factors (N/M) and measurement resolutions (B_y). For the target specification, even potential inaccuracies

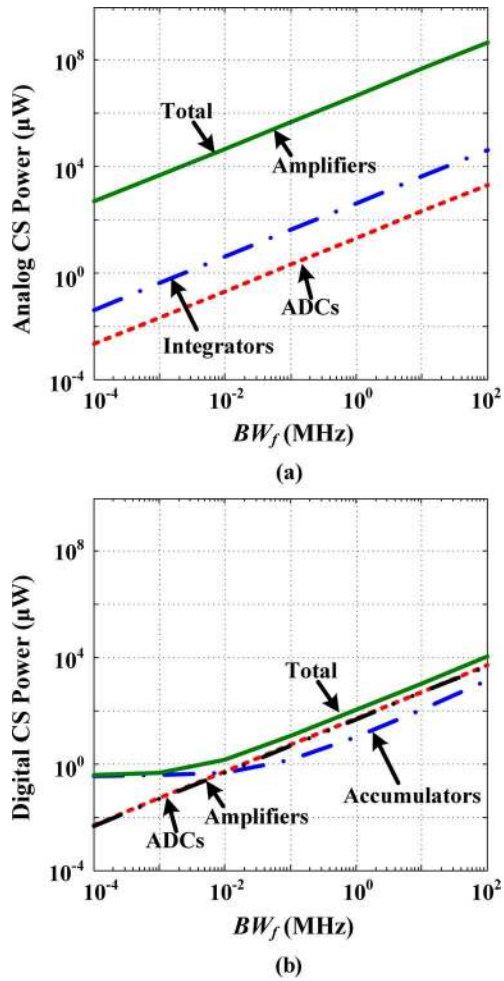


Fig. 8. Power breakdown of the (a) analog CS implementation and the (b) digital CS implementation over input signal bandwidth for a 90 nm CMOS technology where $M = 50, N = 500, B_y = 10, B_f = 8$, and $G_A = 40$ dB.

in the power models cannot account for several orders of magnitude in power difference, so a digital implementation clearly presents the more power efficient option.

The common power limitation of the analog implementation stems from the noise and headroom requirements of the amplifier. In each case, higher signal gain, compression (larger N), and resolution translates into a lower input referred noise requirement. The steep power cost for low noise in the OTA is then multiplied by the number of parallel measurements, M. The amplifier's power dominance is shown in Fig. 8 where the power breakdown of both the analog and digital CS implementations is plotted across operating frequencies (input signal bandwidth). As expected, the digital implementation is limited by leakage at low sample rates and the ADC and OTA at higher sampling rates.

B. Matrix Generation

The problem of generating the measurement matrix ($\mathbf{\Phi}$) coefficients is a common problem for both analog and digital realizations, and in many cases it can be the limiting factor for power

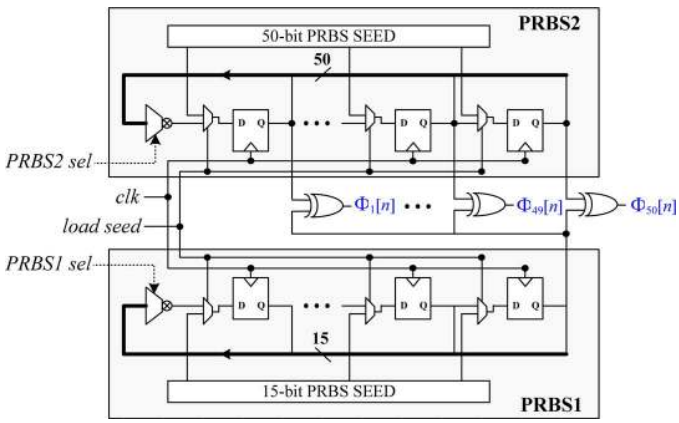


Fig. 9. Block diagram of the measurement matrix generation block. The PRBS seeds are loaded every Nth sample in conjunction with the resetting of the accumulators.

and area. Since the matrix needs to approximate a random matrix, one straightforward approach is to use a look-up table or a memory to implement the matrix. However, to get compression factors of 10X or more with an M of 50 requires an N greater than 500 which equates to at least 25,000 entries. Although this may seem like a small amount of memory, it would dwarf the area of the accumulators, ADC and AFE combined and also dominate the power consumption since it is both large (leaky) and needs to run at the Nyquist rate. Additionally, the size of the memory would limit the maximum achievable compression factor of the encoder. Another approach, which was adopted in [19], is to use an independent PRBS generator for each measurement, $y_i[k]$. While this is much more compact than the memory implementation, it still roughly doubles the size of the accumulator array when the length of the PRBS generator polynomial is close to the resolution of the measurement. For example, generating an independent 2^{15}-PRBS sequence for each measurement would require 750 (15×50) flip-flops. Even in [19], where the number of measurements is smaller, the PRBS generators and associated clocks were the largest contributor to power consumption.

Since power consumption is paramount in sensor applications, we propose an alternative realization of the matrix generation that requires only two PRBS generators. As shown in Fig. 9, the matrix generation circuit consists of the *state* of one PRBS generator XOR'd with the *output* of a second PRBS generator to create the columns of $\mathbf{\Phi}$ on a sample by sample basis. The seed and sequence length of each PRBS is programmable to enable the synthesis of a wide variety of pseudo-random matrices. It may seem that this same result could be achieved with only a single PRBS generator since shifted versions of the same sequence should appear uncorrelated with one another. However, because the input is often oversampled, the inner product of a measurement matrix that is derived from a single shifted PRBS sequence and the input will appear correlated. Not taking into account any additional overhead to seed the PRBS generators, the resulting implementation requires only 65 flip-flops for an M of 50 compared to what would have been 750 flip-flops to enable PRBS sequences with the same run length for the approach described in [19]. With these improvements, the matrix generation power is reduced to less than 10% of the digital backend (accumulator) power for the digital CS implementation.[15]

C. CS System Architecture

Fig. 10 shows the resulting block diagram of the proposed system annotated with example waveforms of the signal compression and reconstruction. Based on the analysis presented in Section VI.A, which shows the digital CS encoder to be three orders of magnitude more efficient for the targeted specifications, the architecture chosen for implementation is the digital one shown in Fig. 6. The final implementation uses 16-bit accumulators in the encoder to avoid overflow for compression block lengths up to 4000 samples for a 10-bit measurement resolution target or alternatively an 11-bit resolution for 1000 sample block lengths. For the digital system, there is a small incremental power cost to allow this additional flexibility to experimentally explore dependencies on resolution and compression factors.

VII. Measurement Results

In order to validate the predicted hardware costs and demonstrate the system, the encoder circuits shown in Figs. 6 and 9 were fabricated in a 90 nm CMOS process. The test chip consists of a low-area 8-bit SAR ADC [31] and the CS encoder block described in the previous section [10]. Fig. 11 shows the die photo of the chip with the layout superimposed along with the test infrastructure and the measured power for the CS encoder. The digital CS encoder, including control circuitry, matrix generation and clock power, consumes only 1.9 μW at 0.6 V for sampling frequencies below 20 kS/s. As expected, the measured power is largely dominated by leakage for the sampling frequencies of interest. Considering that the operating point is in the leakage limited regime, the results correlate well with the model developed in Section V which predicts $\sim$0.6 μW of power consumption for the digital backend and matrix generation (no clocks, buffers or control) under the same operating conditions.

For testing, pre-recorded sensor signals were either driven into the ADC from an external DAC or passed directly as digitized data into the CS block through an on-chip deserializer. The output of the ADC could be observed synchronously with the output of the CS encoder block to enable a comparison between the quantized and reconstructed signals. Fig. 12 shows an example of a continuous data acquisition for a CF of 20. In this example, a pre-recorded EEG signal [32] driven by the off-chip DAC is sampled, compressed and then reconstructed off-line. The input is quantized by the ADC and continuously compressed from 1000 8-bit ADC samples into 50 16-bit accumulator measurements netting an effective CF of 10.[16] As Fig. 12 shows, the reconstructed signal faithfully represents the

[15]In the case where the system is not in the leakage limited regime, dynamic power consumed in the accumulators can be roughly halved by interpreting the matrix as 1's and 0's rather than $+1$'s and -1's as described in [10]. This allows the accumulator clocks to be gated when multiplied with 0. To enable this, each accumulator needs an additional bit or two to accommodate the any DC offset in the signals.

[16]It should be noted that not all 16 bits in the accumulator are required to recover the 8-bit signal so the actual compression performance is better than 10X.

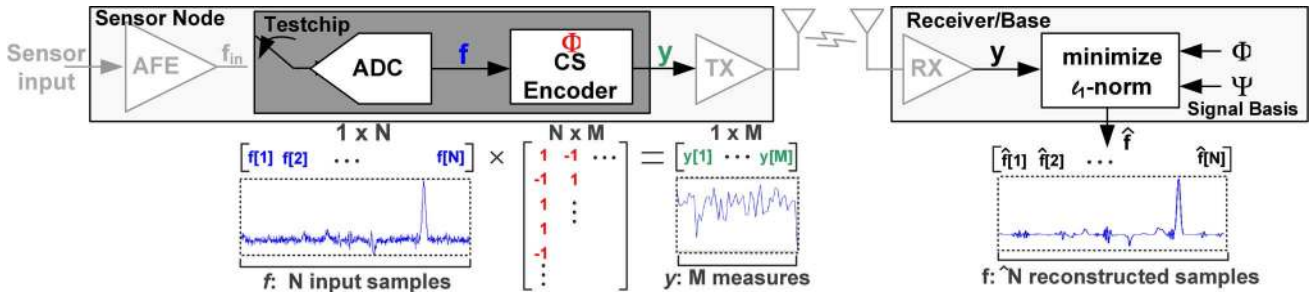


Fig. 10. Block diagram of the proposed sensor system and test chip showing the equivalent mathematical function of the CS encoder and reconstruction on an example waveform.

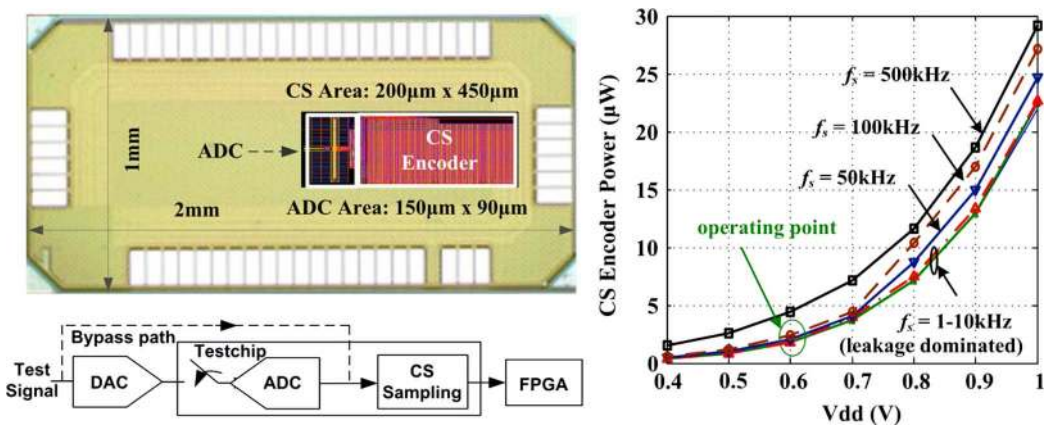


Fig. 11. Die photo, testing infrastructure and measured power for the CS encoder.

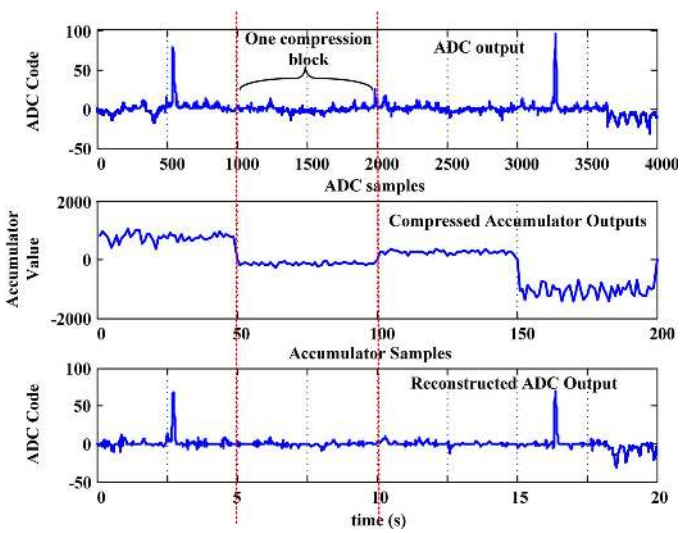


Fig. 12. Measured result showing continuous data acquisition of an EEG signal (driven by an off-chip DAC) showing the ADC output, compressed measurements, and reconstructed waveform when 1000 input samples (N) are compressed to 50 measurements (M).

distinguishing features of the original ADC output despite being somewhat lossy.

As with any lossy compression scheme, there is a question of how much loss is acceptable. From Sections II, III and VI, we know the quality of the recovered signal depends on the signal sparseness and compression factor (N/M) but it also depends on the resolutions of the ADC and CS encoder. Since both of these factors also translate into hardware cost, there is an opportunity to further reduce the power if the recovered signal quality is relatively insensitive to either parameter. To explore this space, a synthetic EEG signal with over a dozen non-zero elements is created and driven by the off-chip DAC into the test chip. The measured signal-to-noise and distortion ratio (SNDR)[17] for the reconstructed signal under varying compression factors and resolutions is plotted in Fig. 13. Since the number of non-zero elements exceeds what can be reconstructed from only 50 measurements, it is expected that the reconstruction will not be perfect. However, in each case, the large amplitude spike signal is well recovered which is indicative of the CS reconstruction process which is more robust when recovering higher energy components of the signal. The effect of having a lower resolution ADC, is emulated by masking out the ADC's LSBs in hardware while the effect of transmitting fewer bits from the CS encoder is mimicked by dropping bits during reconstruction. As the plots show, there is little perceptual difference between the reconstructed signal from an 8-bit and 5-bit ADC output. The same is true when the measurement resolution is reduced to 8-bits by dropping LSBs in the accumulator. Relaxing both resolution requirements would further lower the costs of the ADC and OTA as well as improve the compression factor (by transmitting fewer bits). Furthermore, it is interesting to note that the reconstruction error from the on-chip ADC output is lower than from an ideal ADC at lower resolutions. This is due to lower quantization error introduced by the

[17]SNDR is defined as the reference signal energy $(\mathrm{f}_{\mathrm{REF}}^2)$ divided by the error energy between the reconstructed signal and the reference.

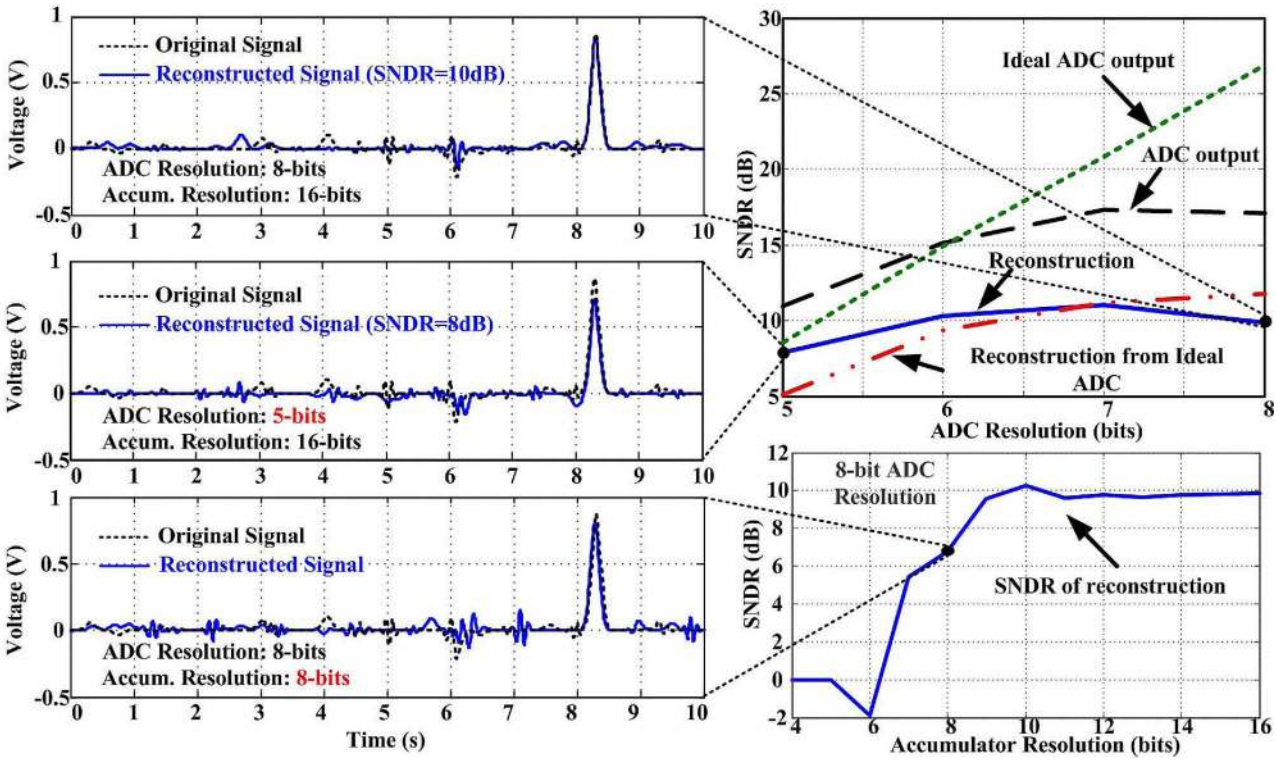


Fig. 13. SNDR of the ideally and actual quantized signals and associated reconstructed signals for each versus measurement resolution (B_y) and ADC resolution (B_f). Select accompanying waveforms provide relative points of reference for the quality of the reconstructed signals.

non-linearity (non-uniform quantization) of the on-chip ADC. This is not a wholly unexpected result as uniform quantizers are not necessarily optimal for CS signal recovery [33].

VIII. Discussion

Thus far, the work presented has focused on the design costs of a CS system for a wireless medical sensor. In this section, we discuss some general implications of the circuit analysis results and possible extensions of our modeling framework and the CS architecture.

A. Modeling Results

In the case of the digital system, the LE model is relatively mature and there are few modeling assumptions so the predicted results correlate well with the measured results. For the analog system, however, there are some built-in assumptions to the model that will generally produce optimistic power numbers. For example, it is assumed that the circuit components perform ideally such that the integrator and mixer perform perfect accumulation and multiplication like their digital counterparts. In reality this will not be true, so when comparing the digital and analog systems at the same specifications, the resulting system performance will not be identical. For the power comparison in this work, the results favored the digital implementation despite the optimistic analog power estimate, but care should be taken to analyze these assumptions when the system specifications result in similar power performance.

B. Model Applicability

The inputs to the power modeling framework presented consist only of technology parameters, circuit performance specifications and system specifications. So to the extent that these inputs are well defined, the model is applicable to any CS application. One clear extension of the model is to analyze the power tradeoffs for AIC applications. AICs, which are identical to the analog system presented, have been proposed as a way to reduce the sampling frequencies of ADCs but it has never been clear if it is generally a more power efficient approach than an ADC alone. The cost of the digital system presented is similar to a single ADC at higher frequencies so the AIC comparison would likely yield similar conclusions as those presented in Fig. 7. Similar to high-speed ADCs, whose performance is often limited by sampling jitter, AICs will see a similar limitation in the mixer block at higher frequencies as has been noted in [19].

C. Compression Performance and Cost

The measured results have shown compression performance that is on the same order of magnitude as previous feature extraction systems [3]–[6] without requiring any decision making at the sensor node, while the energy-efficiency and power cost of the system is on par with or better than a custom feature extraction ASIC [7]. However, since CS is performing data compression rather than any decision making, it is more appropriate to compare it to other compression/source coding schemes. For comparison, we limit this discussion to lossless compression alternatives since the quality of the recovered signal is known and independent of the signal type.

From Fig. 13, we can see that when the input signal ("Original Signal") is quantized to 8 bits (B_f), the quality (SNDR) of the reconstructed signal is the same for accumulator resolutions (B_y) greater than 10 bits. Thus, the CS system still preserves the peak reconstruction performance by transmitting only 10 bits per measurement resulting in a coding efficiency of 0.5 bits per sample. In other words, it takes the CS system 500 bits (50 measurements $\times$ 10 bits/measurement) to represent the 1000 sample, 8-bit input sequence.

Comparatively, the theoretical entropy of the same 8-bit input signal is significantly higher at 3.2 bits per sample. This represents the coding efficiency that one might achieve with an infinite length Huffman code [34] which is calculated as

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \tag{23}$$

where p is the probability mass function of X, and X represents the distribution of samples in the signal. This result is to be expected as the sample entropy does not take advantage of correlations between samples in the signal. Typically, the Lempel–Ziv–Welch (LZW) compression algorithm is more suitable for this purpose as it is more efficient at encoding repetitive data [35]. Again, for comparison, we pass the 8-bit test signal used in Fig. 13 into an LZW encoder. The size of the minimum encoded output from the LZW algorithm is 2950 bits (295 10-bit code words) resulting in a coding efficiency of 2.95 bits per sample. So in this example, when compared to CS, a $\sim$6X penalty in transmission energy is paid to achieve lossless compression.

For LZW to improve its coding efficiency, the block length (and input length) of the encoder must increase such that longer repetitions in the signal can be more efficiently encoded.[18] For LZW, this requires a larger code dictionary and longer code sequence to be stored before transmission, which requires greater hardware cost. As seen in our power analysis, digital circuits for low bandwidth applications, such as wireless sensors, will often be leakage limited, so more storage implies more power. Thus, for any alternative compression scheme to be competitive with CS in terms of power, the storage requirements must be on the order of 1000 flip-flops[19] or less. In the case of LZW, the example just described consumes only $\sim$3 k storage elements for the coded output, but the corresponding dictionary needed to generate that output code requires an 11 k memory[20] where the storage requirements for both the output code and dictionary increase as higher compression is desired. Even without accounting for differences in computational complexity (which favors CS), the CS compression system, though lossy, offers 6X higher compression at over 10X lower implementation (storage/power) cost.

IX. Conclusion

This work has presented an application of CS theory that addresses the energy and telemetry bandwidth constraints of wireless sensor nodes by enabling data compression without loss of generality. The circuit models have been developed to enable the power/performance analysis of analog and digital implementations of the proposed CS encoder over a range of system specifications. The analysis reveals that a digital implementation, rather than the more commonly proposed analog encoder, is a significantly more energy-efficient and suitable architecture for wireless sensor applications. Furthermore, a compact and efficient method of generating the encoding matrix *on-the-fly* is presented that enables a low-power and low-area solution to one of the design limitations of CS encoders. The fabricated test chip demonstrates the first fully integrated circuit realization of a CS encoder, validates the circuit model and choice of implementation, and demonstrates the ability to continually and blindly compress bioelectrical signals at compression factors of 10X or more without the need for any general purpose memory or processing at the sensor node. The proposed system provides a generic platform that can be adopted to compress data for any application that captures sparse signals, and measurements show that a proper choice of metrics could enable further hardware reduction.

[18]For example, the code efficiency for encoding the test input sequence repeated twice is 2.34 bits/sample instead of 2.95 bits/sample.

[19]The CS system presented uses a total of 865 flip-flops in the accumulators and PRBS generators.

[20]This calculation assumes that the codebook is initialized with all 256 (for an 8-bit input) possible single sample sequences. The resulting code book size for the test input signal has 550 20-bit codes (each code is a 10-bit sequence prefix followed by the new 10-bit character).

Acknowledgment

The authors would like to thank MIT CICS for support and R. Sredojević, M. Georgas, and F. Lim for their helpful discussions.

References

[1] A. Gaddam, S. C. Mukhopadhyay, G. S. Gupta, and H. Guesgen, "Wireless sensors networks based monitoring: Review, challenges and implementation issues," in *Proc. 2008 3rd Int. Conf. Sensing Technology*, Nov. 2008, pp. 533–538.

[2] C. Links, "Wireless sensor networks: Maintenance-free or battery-free?," *RTC Mag.*, vol. 2, pp. 18–21, 2009.

[3] 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, Jan. 2007.

[4] B. Gosselin and M. Sawan, "Circuits techniques and microsystems assembly for intracortical multichannel ENG recording," in *Proc. 2009 IEEE Custom Integrated Circuits Conf.*, 2009, pp. 97–104.

[5] R. Olsson and K. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, pp. 2796–2804, 2005.

[6] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. P. Chandrakasan, "A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 804–816, Apr. 2010.

[7] V. Karkare, S. Gibson, and D. Markovic, "A 130-μW, 64-channel spike-sorting DSP chip," in *Proc. IEEE Asian Solid-State Circuits Conf.*, 2009, vol. 1, pp. 289–292.

[8] J. L. Bohorquez, J. L. Dawson, and A. P. Chandrakasan, "A 350 μW CMOS MSK transmitter and 400 μW OOK super-regenerative receiver for medical implant communications," in *2008 Symp. VLSI Circuits Dig. Tech. Papers*, 2008, pp. 32–33.

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

[10] F. Chen, A. P. Chandrakasan, and V. Stojanović, "A signal-agnostic compressed sensing acquisition system for wireless and implantable sensors," in *Proc. 2010 IEEE Custom Integrated Circuits Conf.*, 2010, pp. 1–4.

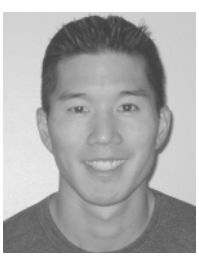
[11] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[12] S. Miaou and S. Chao, "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 539–543, 2005.

[13] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE 14th Workshop on Statistical Signal Processing*, 2007, pp. 181–184.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[15] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.
[16] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[17] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
[18] 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)*, May 2007, pp. 1959–1962.
[19] 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.
[20] R. Robucci, J. D. Gray, J. Romberg, and P. Hasler, "Compressive sensing on a CMOS separable-transform image sensor," *Proc. IEEE*, vol. 98, no. 6, pp. 1089–1101, Jun. 2010.
[21] P. K. Baheti, "An ultra low power pulse oximeter sensor based on compressed sensing," *Proc. Wearable and Implantable Body Sensor Networks, 2009*, pp. 144–148, 2009.
[22] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," in *Proc. 2008 IEEE Custom Integrated Circuits Conf.*, Sep. 2008, vol. 1, pp. 105–112.
[23] A. R. Shahani, D. K. Shaeffer, and T. H. Lee, "A 12-mW wide dynamic range CMOS front-end for a portable GPS receiver," *IEEE J. Solid-State Circuits*, vol. 32, pp. 2061–2070, 1997.
[24] 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.
[25] 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.
[26] J. Holleman and B. Otis, "A sub-microwatt low-noise amplifier for neural recording," in *Proc. 2007 IEEE Engineering in Medicine and Biology Conf.*, 2007, pp. 3930–3933.
[27] M.-Z. Li and K.-T. Tang, "A low-noise low-power amplifier for implantable device for neural signal acquisition," in *Proc. 2009 IEEE Engineering in Medicine and Biology Conf.*, 2009, pp. 3806–3809.
[28] I. Sutherland, B. Sproull, and D. Harris, *Logical Effort: Designing Fast CMOS Circuits*. San Mateo, CA: Morgan Kaufmann, 1999.
[29] D. Harris and I. Sutherland, "Logical effort of carry propagate adders," in *Proc. 37th Asilomar Conf. Signals, Systems & Computers*, 2003, vol. 1, pp. 873–878, 2003.
[30] D. Markovic, V. Stojanovic, B. Nikolic, M. A. Horowitz, and R. W. Brodersen, "Methods for true energy-performance optimization," *IEEE J. Solid-State Circuits*, vol. 39, no. 8, pp. 1282–1293, 2004.
[31] F. Chen, A. P. Chandrakasan, and V. Stojanović, "A low-power area-efficient switching scheme for charge-sharing DACs in SAR ADCs," in *Proc. 2010 IEEE Custom Integrated Circuits Conf.*, 2010, pp. 1–4.
[32] Swartz Center for Computational Neuroscience, University of California at San Diego. [Online]. Available: http://sccn.ucsd.edu
[33] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. 2009 IEEE Int. Symp. Information Theory*, Jun. 2009, vol. 3, pp. 6–10.
[34] D. A. Huffman, "Minimum-redundancy coding for the discrete noiseless channel," *IEEE Trans. Inf. Theory*, vol. 7, no. 1, pp. 27–38, Jan. 1961.
[35] T. A. Welch, "A technique for high-performance data compression," *Computer*, vol. 17, no. 6, pp. 8–19, 1984.



Fred Chen (S'00–M'11) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 2011, the M.S. degree from the University of California at Berkeley in 2000, and the B.S. degree from the University of Illinois at Urbana-Champaign in 1997, all in electrical engineering.

From 2000 to 2005, he was with Rambus Inc., Los Altos, CA, where he worked on the design of high-speed I/O and equalization circuits. He has also previously held a design position at Motorola, Inc., Libertyville, IL. His current research interests include energy-efficient circuits and systems, and circuit design in emerging technologies.

Dr. Chen was a recipient of the 2010 ISSCC Jack Raper Award for Outstanding Technology Directions Paper.

Anantha P. Chandrakasan (M'95–SM'01–F'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley in 1989, 1990, and 1994, respectively.

Since September 1994, he has been with the Massachusetts Institute of Technology, Cambridge, where he is currently the Joseph F. and Nancy P. Keithley Professor of Electrical Engineering. He was the Director of the MIT Microsystems Technology Laboratories from 2006 to 2011. Since July 2011, he has been the Head of the MIT EECS Department. His research interests include micro-power digital and mixed-signal integrated circuit design, wireless microsensor system design, portable multimedia devices, energy efficient radios and emerging technologies. He is a co-author of *Low Power Digital CMOS Design* (Kluwer Academic Publishers, 1995), *Digital Integrated Circuits* (Pearson Prentice-Hall, 2003, 2nd edition), and *Sub-threshold Design for Ultra-Low Power Systems* (Springer 2006). He is also a co-editor of *Low Power CMOS Design* (IEEE Press, 1998), *Design of High-Performance Microprocessor Circuits* (IEEE Press, 2000), and *Leakage in Nanometer CMOS Technologies* (Springer, 2005).

Dr. Chandrakasan was a co-recipient of several awards including the 1993 IEEE Communications Society's Best Tutorial Paper Award, the IEEE Electron Devices Society's 1997 Paul Rappaport Award for the Best Paper in an EDS publication during 1997, the 1999 DAC Design Contest Award, the 2004 DAC/ISSCC Student Design Contest Award, the 2007 ISSCC Beatrice Winner Award for Editorial Excellence and the ISSCC Jack Kilby Award for Outstanding Student Paper (2007, 2008, 2009). He received the 2009 Semiconductor Industry Association (SIA) University Researcher Award. He has served as a technical program co-chair for the 1997 International Symposium on Low Power Electronics and Design (ISLPED), VLSI Design'98, and the 1998 IEEE Workshop on Signal Processing Systems. He was the Signal Processing Sub-committee Chair for ISSCC 1999–2001, the Program Vice-Chair for ISSCC 2002, the Program Chair for ISSCC 2003, the Technology Directions Sub-committee Chair for ISSCC 2004–2009, and the Conference Chair for ISSCC 2010–2011. He is the Conference Chair for ISSCC 2012. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 1998 to 2001. He served on SSCS AdCom from 2000 to 2007 and he was the meetings committee chair from 2004 to 2007.

Vladimir Stojanović (S'96–M'04) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2005, and the Dipl.Ing. degree from the University of Belgrade, Serbia, in 1998.

He is the Emanuel E. Landsman Associate Professor of Electrical Engineering and Computer Science at MIT. He was with Rambus Inc., Los Altos, CA, from 2001 through 2004. His research interests include design, modeling and optimization of integrated systems, from CMOS-based VLSI blocks and interfaces to system design with emerging devices like NEM relays and silicon-photonics. He is also interested in design and implementation of energy-efficient electrical and optical networks, and digital communication techniques in high-speed interfaces and high-speed mixed-signal IC design.

Dr. Stojanovic received the 2006 IBM Faculty Partnership Award and the 2009 NSF CAREER Award, as well as the 2008 ICCAD William J. McCalla, 2008 IEEE TRANSACTIONS ON ADVANCED PACKAGING, and 2010 ISSCC Jack Raper Best Paper Awards.