A 0.6–107 μW Energy-Scalable Processor for Directly Analyzing Compressively-Sensed EEG

Mohammed Shoaib, *Member, IEEE*, Kyong Ho Lee, *Student Member, IEEE*, Niraj K. Jha, *Fellow, IEEE*, and Naveen Verma, *Member, IEEE*

***Abstract*—Compressive sensing has been used to overcome communication constraints (energy and bandwidth) in low-power sensors. In this work, we present a seizure-detection processor that directly uses compressively-sensed electroencephalograms (EEGs) for embedded signal analysis. In addition to addressing communication, this has two advantages for local computation. First, with compressive sensing, reconstruction costs are typically severe, precluding embedded analysis; directly analyzing the compressed signals circumvents reconstruction costs, enabling embedded analysis within applications. Second, compared to Nyquist-sampled signals, the use of compressed representations reduces the computational energy of signal analysis due to the reduced number of signal samples. We describe an algorithmic formulation as well as a hardware architecture that enables two strong power-management knobs, wherein application-level performance can scale with computational energy. The two knobs are parameterized as follows: 1) ξ, which quantifies the amount of data compression, and 2) ν, which determines the approximation error within the proposed compressed-domain processing algorithm. For ξ and ν in the range 2-24×, the energy to extract signal features (over 18 channels) is 70.8-1.3 nJ, and the detector's performance for sensitivity, latency, and specificity is 96-91%, 4.7-5.3 sec., and 0.17-0.30 false-alarms/hr., respectively (compared to a baseline performance of 96%, 4.6 sec., and 0.15 false-alarms/hr.).**

***Index Terms*—Circuits, compressive sensing, energy efficiency, machine learning, seizure detection, signal processing.**

I. Introduction

COMPRESSIVE sensing is a potentially compelling technique to reduce data in sensing systems. It is based on the principles of sparsity. It states that if an N-sample signal is sparse in a secondary basis $\mathbf{\Psi}$, we can use an $(N/\xi) \times N$ projection matrix $\mathbf{\Phi}$, which is required to be incoherent with $\mathbf{\Psi}$ to create an N/ξ-sample signal [1], [2], where $\xi(\gg 1)$ represents the achieved compression factor. For most signals that are sparse in some basis $\mathbf{\Psi}$, a $\mathbf{\Phi}$ whose elements are set to ± 1 randomly with a uniform probability satisfies the incoherence property with high probability [2]. Such a choice for $\mathbf{\Phi}$ has the benefit of enabling low-energy compression, applicable to a broad range of signals; this has recently been exploited in biomedical sensors [3], [4].

The typical way in which compressive sensing has been used in such sensors is illustrated in Fig. 1(a). Embedded signals acquired from distributed sources are compressed with very low energy to alleviate the communication-bandwidth and -energy constraints between the sensors and a base station (possibly through a gateway device). At the base station, the signals can then be reconstructed for analysis. The limitation of such an approach is that it does not address the need to perform local signal analysis, because generally this has required signal reconstruction, which is computationally intensive [5] and thus impractical to perform on either the sensors or the gateway device. The need for local analysis is gaining importance in advanced sensing systems, particularly for medical applications [6] where local detection can enable closed-loop monitoring and therapeutic devices while also identifying the critical signal segments to enable transmission to centralized human experts for reconstruction and further analysis. The aim of this work is thus to enable the system model shown in Fig. 1(b), wherein compressive sensing is exploited for low-energy signal compression and subsequent reconstruction, but where the compressed representations can also be used directly to perform signal-processing operations, thereby enabling local signal analysis on either the sensors or the gateway device. In fact, the approach of performing analysis directly on compressed representations can have broad and valuable applications beyond systems where the aim is simply to move such functions from a base station to the local nodes. In this paper, we describe how analysis on compressed representations can enable a generalizable approach to substantially reduce computational energy for signal-processing operations.

Embedded medical sensors that perform analysis typically work by extracting signal features based on physiological biomarkers and then feeding these features to high-performance classifiers to detect targeted physiological states [6], [7]. However, since compressive sensing involves multiplication by the random projection matrix $\mathbf{\Phi}$, the biomarkers get obscured and thus present a challenge for signal analysis.

In [8], we proposed a methodology based on a least-squares approximation that allows us to directly analyze compressively-sensed signals. In [9], we also presented an architecture and prototype integrated circuit (IC) that exploits this methodology in an EEG-based seizure-detector. We showed in [9] that the energy of a compressed-domain processor in this case scales quadratically with the level of data compression.

Manuscript received April 25, 2013; revised July 01, 2013 and August 20, 2013; accepted September 07, 2013. Date of publication January 02, 2014; date of current version March 25, 2014. This work was supported in part by the National Science Foundation Grant No. CCF-1253670, Harold W. Dodds Honorific Fellowship from Princeton University, and Systems on Nanoscale Information fabriCs (SONIC), one of the six SRC STARnet Centers, sponsored by MARCO and DARPA. IC fabrication was provided by MOSIS. This paper was recommended by Associate Editor C.-T. Chiu.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: mshoaib@princeton.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/TCSI.2013.2285912

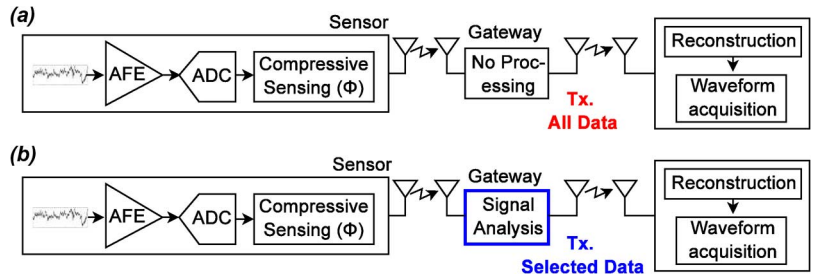


Fig. 1. Proposal to directly analyze compressively-sensed data: (a) previous approaches, (b) proposed approach.

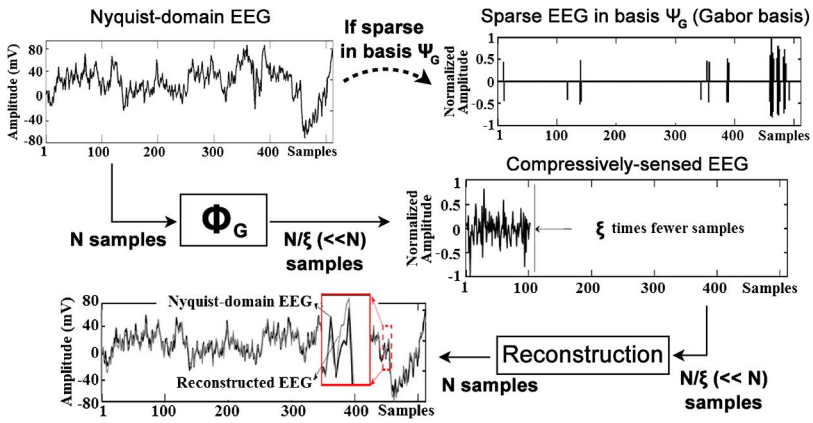


Fig. 2. EEG is sparse in the Gabor basis, enabling substantial compression; although accurate reconstruction is possible, it is computationally intensive, motivating signal analysis directly using the compressed signal.

In [10], we provided an improved approach wherein the compressed-domain equations can be solved more accurately; this is achieved by introducing a second random projection for the output signal, thereby increasing the degrees of freedom available for solving the equations. We applied the approach to a second spike-sorting application. The motivation for this application comes from the fact that in sensing systems, communication bandwidth, not just communication energy, may be of concern. A passive implant functioning as a spike-acquisition transponder is a typical example of such a case. This is a case where communication poses a bandwidth limitation, not an energy limitation, since the implant transmits with no power. In this case, the implant is severely energy constrained, and thus unable to accommodate extensive local processing. On the other hand, for communication, it can take advantage of a passive transmitter based on (inductive) backscattering. The data rate of such a transmitter, however, is limited (due to practical inductors that can be formed). The objective within the implant is thus to reduce the data rate to a level that can be supported by the passive transmitter while consuming no more energy than that required to achieve this level. We show in [8] that the new approach substantially improves the accuracy of the signal-processing system and enables two knobs for trading algorithmic performance in exchange for reduced computational complexity. The energy savings are linear with respect to each of these knobs. In this paper, we use the prototype IC from [9], extending its use to the new approach to enable the two resulting power-management knobs within an energy-scalable EEG-based seizure detector. The resulting algorithm for compressed-domain analysis increases the number of signal-transformation coefficients that need to be stored compared with a traditional Nyquist-domain algorithm. A key attribute of the IC is thus a scalable SRAM. We describe the algorithm and present detailed analysis and measurements from the IC implementation.

Fig. 2 illustrates the concept. EEG is known to be sparse in the Gabor basis $\mathbf{\Psi_G}$ [11]. This enables low-energy compression using a random projection matrix $\mathbf{\Phi_G}$. The resulting EEG-signal representation is compressed, but also substantially altered. Nonetheless, a corresponding representation of the desired signal features can be obtained by *transforming* the feature-extraction computations based on $\mathbf{\Phi_G}$. This results in a *compressed-domain seizure detector*. The resulting transformation not only overcomes the need for signal reconstruction, but also leads to computational energy savings due to a reduction in the number of input samples that need to be processed. The IC thus provides a previously unexplored approach for reducing the computational energy of signal analysis within the processor.

II. System Overview

In this section, we describe the proposed approach for transforming signal analysis from the Nyquist-sampled time domain (henceforth called the Nyquist domain) to the compressed domain. We provide simulation results for seizure detection that compare the performance of the compressed- and Nyquist-domain analyses. We also describe the architecture of the compressed-domain processor, highlighting the new knobs for computational power management that it introduces.

A. Algorithmic Formulation

Signal-classification algorithms typically base their decision rules on key features extracted from the signals *via*

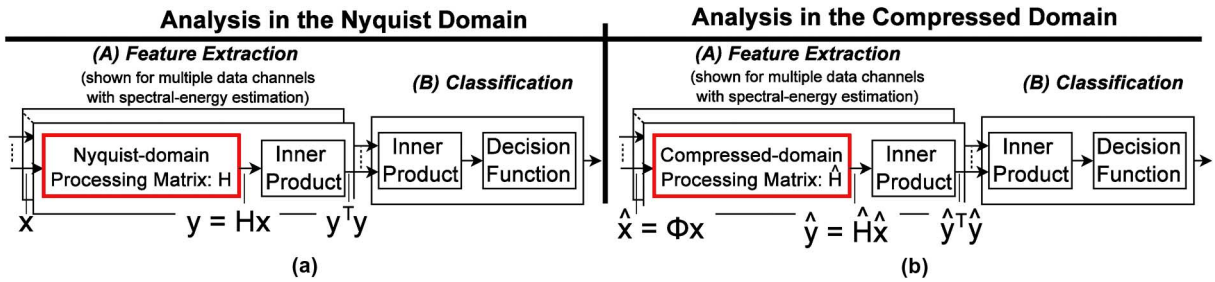


Fig. 3. Transforming any linear signal-processing function, which can be represented as a matrix $\mathbf{H}$, into an equivalent operator $\hat{\mathbf{H}}$ in the compressed domain. This needs a solution for a projection of $\mathbf{y}$, which preserves the inner product of vectors.

signal-processing functions; this is particularly true for medical detectors, where the features often correspond to physiological biomarkers. These algorithms then use a classifier to perform modeling and inference over the extracted features. Powerful classification frameworks exist in the domain of machine learning that can construct high-order and flexible models through data-driven training. In many such frameworks, the classification step utilizes a distance metric (e.g., 2-norm or inner product) between feature vectors (FVs) [12]. In certain cases, the distance metric may also be invoked within the feature extraction step, for instance, to extract spectral energies, which form generic biomarkers for neural field potentials (e.g., brain-machine interfaces [6], sleep disorders [13], etc.). In this paper, we demonstrate our results for the latter case using a seizure-detection application, where clinical studies have shown that EEG spectral energy [derived using the inner product between FVs after linear finite impulse response (FIR) filtering] can serve as a biomarker that indicates the onset of a seizure [14].

Connecting the above concepts with the multi-channel system shown in Fig. 3, in the Nyquist domain [Fig. 3(a)], an N-dimensional signal $\mathbf{x}$ from each data channel is multiplied with an $N \times N$ matrix operator $\mathbf{H}$ to achieve arbitrary linear signal-processing functions in order to derive an FV $\mathbf{y}$ (e.g., $\mathbf{H}$ could be a Toeplitz matrix to implement convolution). The inner product between FVs can then be used to derive the spectral-energy features from each data channel [as shown in Fig. 3(a), spectral-energy features can be denoted by the inner product $\mathbf{y^T y}$]. Extending this to an analysis in the compressed domain [Fig. 3(b)], we now aim to process compressed representations of the input signal, namely $\hat{\mathbf{x}} = \mathbf{\Phi x}$, where $\mathbf{\Phi}$ represents the $(N/\xi) \times N$ random-projection matrix used for compressive sensing. We thus seek to find a matrix transformation $\hat{\mathbf{H}}$ that leads to a representation of a signal which has been suitably processed as intended, but which is derived *by directly using* $\hat{\mathbf{x}}$. In this case, we also show that any other $\hat{\mathbf{H}}\mathbf{\Phi}$ combination cannot perform better than the $\hat{\mathbf{H}}\mathbf{\Phi}$ combination obtained from the presented solution (for all choices of $\mathbf{\Phi}$).

In the proposed approach, we aim to derive compressed-domain FVs $\hat{\mathbf{y}}$ that are projections of the originally intended Nyquist-domain FVs $\mathbf{y}$. With this approach, the chosen projection needs to preserve the inner product between the FVs, since the inner product is what is used by the classifier (either implicitly or for spectral-energy feature extraction). In [8] and [9], we showed that such a projection can be the $(N/\xi) \times N$ compressive sensing matrix $\mathbf{\Phi}$ itself (i.e., $\hat{\mathbf{y}}$ can be equal to $\mathbf{\Phi y}$). However, since $\mathbf{\Phi}$ is a non-square matrix of dimensionality $(N/\xi) \times N$ (determined by the compression factor), this approach allowed us to derive, at best, only a least-squares solution for $\hat{\mathbf{y}}$. Next, we showed how an $(N/\nu) \times N$ auxiliary matrix $\mathbf{\Theta}$ can be applied to $\hat{\mathbf{y}}$, where ν represents the projection factor, can be used instead of $\mathbf{\Phi}$ to introduce additional degrees of freedom that allow us to solve for $\hat{\mathbf{H}}$ exactly [10]. Additionally, by scaling ν to larger values, this approach also permitted us to obtain *approximate solutions to* $\hat{\mathbf{H}}$ *of smaller size*. The resulting solution reduced the number of operations required in the processor and thus enabled an additional knob to scale the energy of compressed-domain analysis based on the required accuracy. Below, we summarize our approach of deriving $\hat{\mathbf{H}}$ using $\mathbf{\Theta}$.

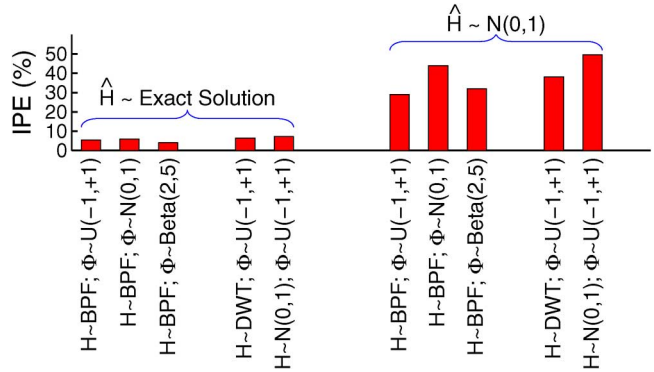


Fig. 4. IPE is lower when we use an exact solution for $\hat{\mathbf{H}}$. The low IPE value remains consistent for different choices of $\mathbf{\Phi}$ and $\mathbf{H}$.

Since $\hat{\mathbf{y}} = \mathbf{\Theta y}$, the inner product between any two FVs $\hat{\mathbf{y_1}}$ and $\hat{\mathbf{y_2}}$ in the compressed domain is given by:

$$\hat{\mathbf{y}}_\mathbf{1}^\mathbf{T}\hat{\mathbf{y}}_\mathbf{2} \Rightarrow (\mathbf{\Theta y_1})^\mathbf{T}(\mathbf{\Theta y_2}) \Rightarrow \mathbf{y_1^T}(\mathbf{\Theta^T\Theta})\mathbf{y_2}. \quad (1)$$

The right hand side will be equal to the inner product $\mathbf{y_1^T y_2}$ in the Nyquist domain if $\mathbf{\Theta^T\Theta}$ is equal to the $N \times N$ identity matrix $\mathbf{I}$. Thus, to simultaneously solve for $\mathbf{\Theta}$ and $\hat{\mathbf{H}}$, we have to solve the following constrained optimization problem:

$$\arg\min_{\mathbf{\Theta}} \|\mathbf{\Theta^T\Theta} - \mathbf{I}\|_2^2 \text{ such that } \mathbf{\Theta H} = \hat{\mathbf{H}}\mathbf{\Phi}. \quad (2)$$

1) Exact Solution for $\hat{\mathbf{H}}$ *(for Highest Accuracy):* Assuming $\mathbf{H}$ is a square matrix, we can obtain the singular value decomposition (SVD) of $\mathbf{\Phi H^{-1}}$ as $\mathbf{VSU^T}$, where $\mathbf{V}$ and $\mathbf{U}$ are orthogonal matrices (i.e., $\mathbf{U^TU} = \mathbf{V^TV} = \mathbf{I}$) and $\mathbf{S}$ is an

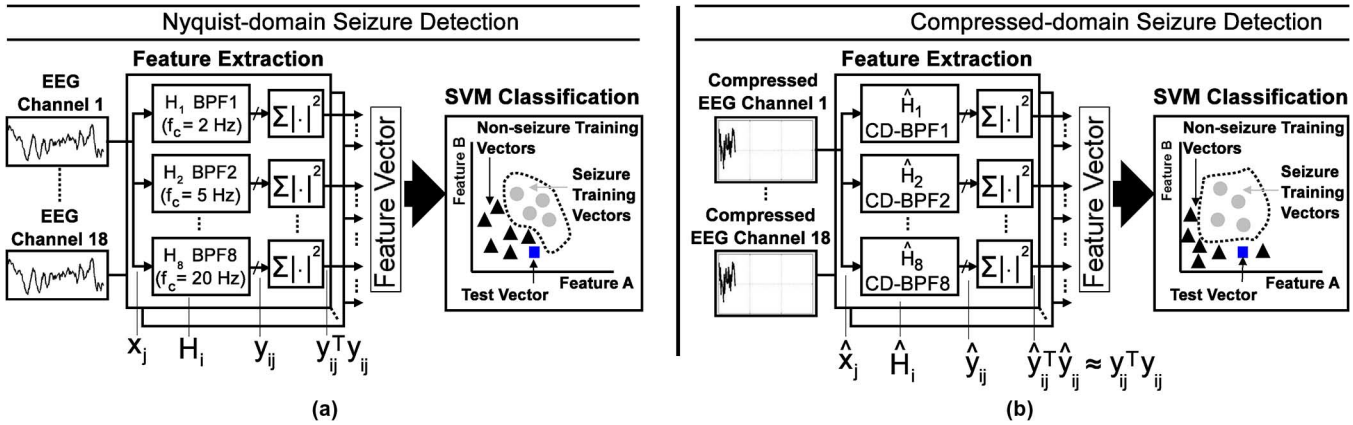


Fig. 5. The seizure detection algorithm involves feature extraction and classification using an SVM. In the compressed domain, the Nyquist-domain BPFs $\mathbf{H_i}$ are transformed to the compressed-domain BPFs (CD-BPFs) $\hat{\mathbf{H}}_\mathbf{i}$.

$(N/\xi) \times (N/\xi)$ diagonal matrix formed by the singular values of $\mathbf{\Phi H^{-1}}$. We thus have the following relationship for $\mathbf{\Theta^T \Theta}$:

$$\mathbf{\Theta^T \Theta} = (\hat{\mathbf{H}}\mathbf{\Phi H^{-1}})^\mathbf{T} \hat{\mathbf{H}}\mathbf{\Phi H^{-1}} = \mathbf{U}(\mathbf{S V^T} \hat{\mathbf{H}}^\mathbf{T} \hat{\mathbf{H}} \mathbf{V S}) \mathbf{U^T}. \tag{3}$$

The distance from the above matrix to the identity will be at least the rank deficiency of $\mathbf{U}$ [15]. The lower bound in (2) will thus be achieved if we set $\nu = \xi$,

$$\hat{\mathbf{H}} = \mathbf{S^{-1} V^T}, \quad \text{and} \quad \mathbf{\Theta} = \hat{\mathbf{H}} \mathbf{\Phi H^{-1}}. \tag{4}$$

Note that, in this case, the dimensionality of the compressed-domain processing matrix $\hat{\mathbf{H}}$ is $(N/\xi) \times (N/\xi)$ and is controllable only through one knob: the amount of data compression (i.e., ξ).

2) Approximate Solution for $\hat{\mathbf{H}}$ (for Energy Savings): In order to solve (2) approximately, we invoke the Johnson-Lindenstrauss (JL) Lemma [16], which states that $\mathbf{\Theta^T \Theta} \approx \mathbf{I}$ when each row of $\mathbf{\Theta}$ is drawn from the normal distribution $N(0, 1)$. Also, the error in the above approximation increases when we have fewer rows in $\mathbf{\Theta}$. Since $\mathbf{H}$ and $\mathbf{\Phi}$ are constants in (2), to derive rows of $\mathbf{\Theta}$ from $N(0, 1)$, we can draw each row of $\hat{\mathbf{H}}$ from $N(0, \mathbf{V S^{-2} V^T})$ [10]. This provides an approximate solution for $\hat{\mathbf{H}}$ in (2). The approximate solution is applicable in the following two cases: (1) when $\mathbf{\Theta}$ has fewer rows than columns (i.e., when $\nu > \xi$) or (2) when $\mathbf{\Theta}$ has more rows than columns (i.e., when $\nu < \xi$). As compared to the exact solution ($\nu = \xi$), the former case results in a smaller-sized $\hat{\mathbf{H}}$ and thus potentially lower energy for signal processing. The latter case, however, results in a $\hat{\mathbf{H}}$, which is bigger in size than the $\hat{\mathbf{H}}$ obtained from the exact solution. Thus, the design space of $\nu < \xi$ results in a higher computational complexity and reduced accuracy, compared to the exact solution. It is thus not a very useful design space for low-energy operation. We block out this regime in all subsequent figures that show the performance/energy of the approximate solution.

Note that when we use the approximate solution, the dimensionality of $\hat{\mathbf{H}}$ is $(N/\nu) \times (N/\xi)$ and is controllable through two independent knobs: (1) the amount of data compression (i.e., ξ) and (2) the required accuracy (i.e., ν) for solving the compressed-domain system of equations [(2)]. As an aside, we also note that given $\mathbf{\Phi}$ and $\mathbf{H}$, we can constrain the sparsity of $\hat{\mathbf{H}}\mathbf{\Phi}$ within the optimization setup [(2)] to derive an approximate solution for $\hat{\mathbf{H}}$. This is an interesting direction for future work.

In deriving $\mathbf{\Theta}$ and $\hat{\mathbf{H}}$, our aim is to preserve the inner product between the processed FVs in the compressed domain. To illustrate the benefit of the proposed transformations, we evaluate the inner-product error (IPE) in the compressed domain, which is defined as $|\hat{\mathbf{y}}_\mathbf{i}^\mathbf{T} \hat{\mathbf{y}}_\mathbf{i} - \mathbf{y}_\mathbf{i}^\mathbf{T} \mathbf{y}_\mathbf{i}| / (\mathbf{y}_\mathbf{i}^\mathbf{T} \mathbf{y}_\mathbf{i})$. We use a set of 100 EEG vectors, each comprising $N = 512$ samples and represented by vector $\mathbf{x_i}$. To illustrate the benefit of our approach, we consider three different Nyquist-domain matrices, $\mathbf{H}$, representing various signal processing functions: (1) a 64-order band-pass filter (BPF), (2) a discrete wavelet transform (DWT), and (3) a random transfer function obtained by using coefficients for $\mathbf{H}$ drawn from $N(0, 1)$. Further, we fix $\xi = 6\times$ and compute $\hat{\mathbf{y}}_\mathbf{i}$ after processing the compressed representation $\mathbf{\Phi x_i}$ with a matrix $\hat{\mathbf{H}}$, derived using (1) the exact solution presented above and (2) an $N/\xi \times N/\xi$ matrix whose entries are derived from $N(0, 1)$. Fig. 4 shows the results generated from MATLAB. We observe that the IPE is substantially lower when we use an $\hat{\mathbf{H}}$ obtained from the proposed transformation. In fact, using the exact solution, the IPE remains below 6% even for different choices of $\mathbf{\Phi}$. The key question, however, is how much IPE is acceptable for retaining the classifier performance. To provide a quantitative answer, we next study an application for epileptic seizure detection and provide an analysis based both on the end-to-end performance and the mutual information of the derived FVs with respect to the true class labels.

B. Compressed-domain Seizure Detection

We apply the above methodology to a seizure detection algorithm. Fig. 5(a) shows a block diagram of the baseline Nyquist-domain detector [14]. A two-second epoch of each EEG channel is processed using eight BPFs with passbands of 0-3 Hz, 3-6 Hz, ..., 21-24 Hz. The energy from each filter is then represented by the inner product of the output samples to form an FV, which is then used for classification by a support vector machine (SVM) classifier. The baseline detector is validated on 558 hours of EEG data in the CHB-MIT database [17]. The Nyquist-domain detector has been demonstrated to achieve an average latency, sensitivity, and specificity of 4.49 sec., 96.03%, and 0.1471 false alarms per hour, respectively [14].

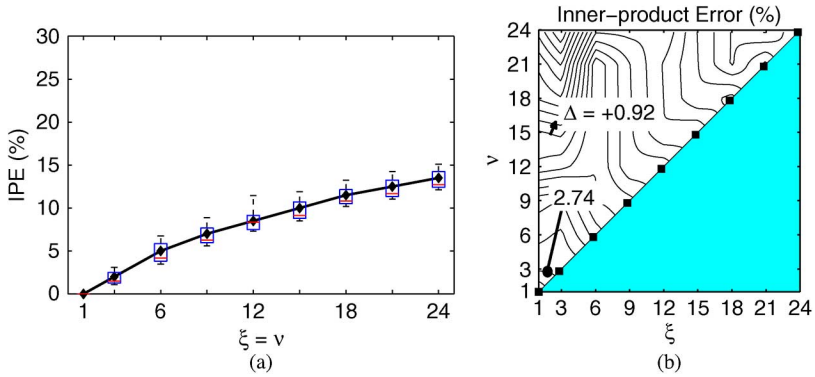


Fig. 6. IPE in (a) the exact solution is below 15% (out to compression factors of 24×) and (b) begins to degrade in the approximate solution with increasing values of ν. The exact solution is also shown in (b) with dark boxes. The cases of $\nu < \xi$, which are not favorable for low-energy operation, are blocked out. Δ shows the separation between adjacent contours.

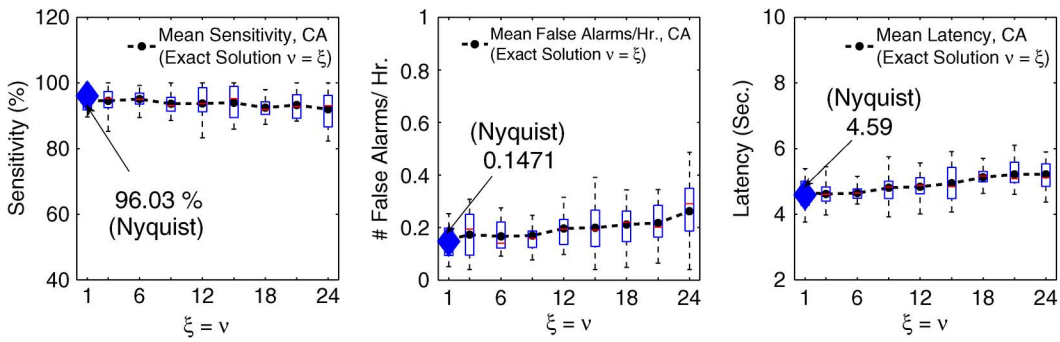


Fig. 7. Performance of the compressed-domain seizure detection algorithm using the exact solution (shown over 558 hrs. of EEG data from 21 patients) is maintained up to large ξ.

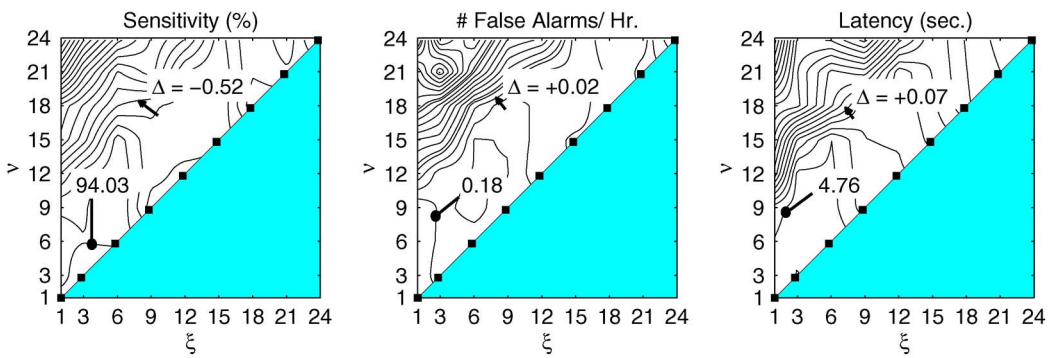


Fig. 8. Detector performance for the approximate solution measured by (a) sensitivity, (b) false alarms/hr., and (c) latency. The performance begins to degrade gradually due to the JL-approximation at higher values of ν.

To enable a transformation for the compressed domain, the ith BPF for the jth EEG channel can be formulated as matrix multiplication, namely of an input signal $\mathbf{x_j}$ by a matrix $\mathbf{H_i}$ to compute the filtered signal $\mathbf{y_{ij}}$. This formulation can lead to the compressed-domain system shown in Fig. 5(b), which is based on compressed-domain band-pass filters (CD-BPFs). In this system, we derive the corresponding matrix $\hat{\mathbf{H}}_\mathbf{i}$ using (1) an exact solution and (2) an approximate solution, as described in the previous section.

Fig. 6 shows the IPE in the FVs using the two approaches. For the results of the approximate solution, Δ is the distance between adjacent contours with units shown above the figure. Using the exact solution, the IPE is below 15% even at $\xi = 24\times$. Further, Figs. 7 and 8 show the simulated performance of the two approaches, respectively. Fig. 7 shows that despite a non-zero IPE value, performance very close to the Nyquist-domain seizure detector is retained up to large values of the compression factor ξ. For a given amount of data compression (i.e, at fixed ξ), Fig. 8 shows that using an approximate solution, the performance begins to degrade gradually with increasing values of ν. Although a low IPE helps retain performance, the information content [8], based on the mutual information of the FVs, provides a more direct relationship to the performance trend. Fig. 9 shows the mutual information values for the exact and approximate solutions. The information trends indeed correlate well with the detector performance. Thus, the proposed transformations allow us to have a small IPE up to substantial amounts of data compression, thereby preserving the information content of the FVs for classification. However, as the IPE increases at higher compression factors, the information degradation becomes substantial, affecting the end-to-end detector performance.

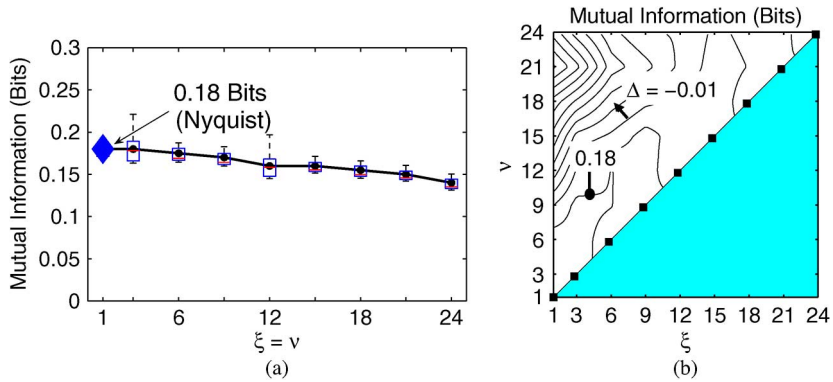


Fig. 9. Mutual information in (a) the exact solution and (b) the approximate solution follows the performance trends shown in Figs. 7 and 8, respectively.

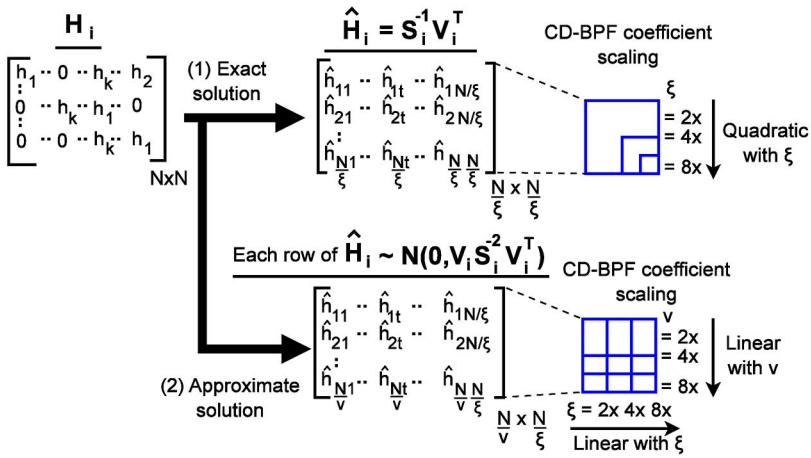


Fig. 10. CD-BPF matrices $\hat{\mathbf{H}}_\mathbf{i}$, derived using $\mathbf{H_i}$ and $\mathbf{\Phi}$, disrupt the regularity and zeros in $\mathbf{H_i}$. The complexity of the CD-BPFs scales (a) quadratically with ξ for the exact solution and (b) linearly with ξ and ν for the approximate solution.

Note that, through the results presented in this section, we aim to show general trends; showing precise correspondence between performance, IPE, and information metrics is not our intention. Based on the analysis presented in this section, we conclude that ξ and ν provide us with two knobs to control the end-to-end performance of the compressed-domain detector. Next, we study the impact of these two knobs on the size of the compressed-domain processing matrices $\hat{\mathbf{H}}_\mathbf{i}$ and thus on the processor implementation and energy.

C. Processor Architecture With Power Management

We exploit the scalability of ξ and ν as knobs for system power management. An important consequence of the algorithmic construction proposed is that the CD-BPF matrices $\hat{\mathbf{H}}_\mathbf{i}$ (which are of dimensionality $(N/\xi) \times (N/\xi)$ for the exact solution and $(N/\nu) \times (N/\xi)$ for the approximate solution) do not retain the regularity of $\mathbf{H_i}$. Even though $\mathbf{H_i}$ are of dimensionality $N \times N$, as shown in Fig. 10, the rows of $\mathbf{H_i}$ are simply selected to implement convolution, and thus are shifted versions of the impulse response of the same FIR filter. As a result, very few unique filter coefficients are required, and many of the coefficients are zero, as determined by the FIR-filter order k. However, in deriving $\hat{\mathbf{H}}_\mathbf{i}$, the shifted impulse responses and zero entries are disrupted. As shown in Fig. 10, the number of multiplications required thus no longer depends on the filter order, but rather (1) *quadratically* on the compression factor ξ for the exact solution and (2) *linearly* on both ξ and ν for the approximate solution. This scaling can potentially reduce the number of multiplications required.

To exploit this attribute, we propose the energy-scalable processor architecture whose block diagram is shown in Fig. 11. The processor consists of two computational stages: compressed-domain feature extraction followed by SVM classification. The compressed-domain feature extractor (CD-FE) includes a CD-BPF and energy-accumulator block. The coefficients for the CD-BPF are pulled from a scalable SRAM bank. Due to the disruption in regularity, the $\hat{\mathbf{H}}_\mathbf{i}$ matrices need a larger number of distinct coefficients to be stored, potentially increasing the memory requirements. Scalability in the SRAM bank is thus an important aspect of power management. We achieve this through the use of multiple subarrays, which enable fine-grained power-gating as well as reduced bit-line and word-line access energy. The total bank size in our implementation is 32 kB, which is partitioned into four subarrays. An SVM classifier (which comprises an inner-product core followed by a kernel transform) is also integrated to perform real-time seizure detection using the derived FVs. Compressively-sensed EEG signals are input directly to the processor for seizure detection. However, for the case of Nyquist inputs, a compressive-projection front-end (CPF) is also included to explicitly multiply inputs by a random projection matrix $\mathbf{\Phi}$; thus the energy savings derived from a reduced number of samples can be exploited even if the original input signal is not compressively sensed.

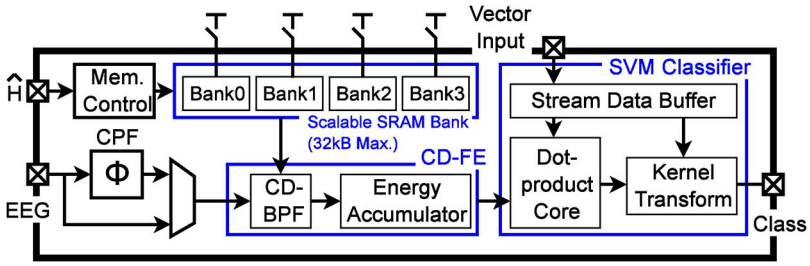


Fig. 11. Architecture block diagram of energy-scalable, compressed-domain seizure detector.

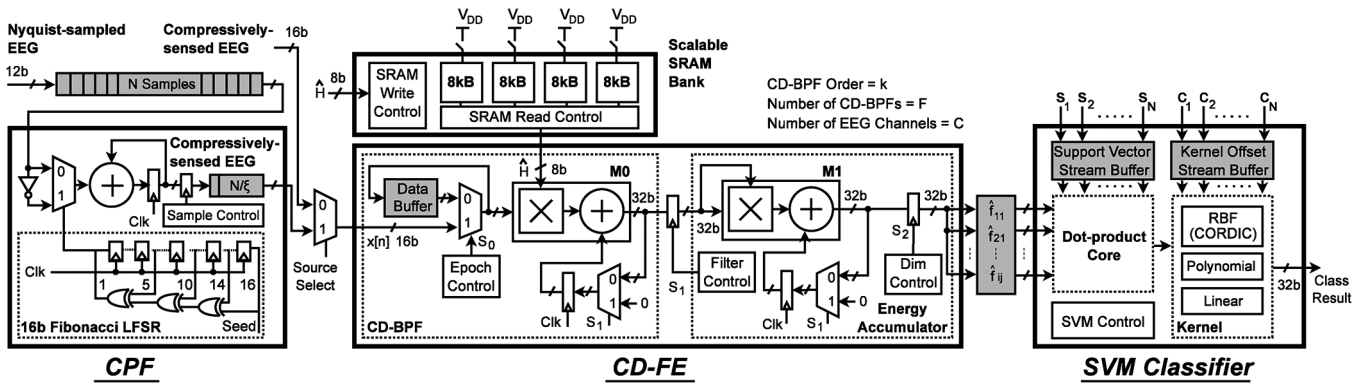


Fig. 12. Circuits used in the compressed-domain processor for seizure detection. Figure adapted from [9].

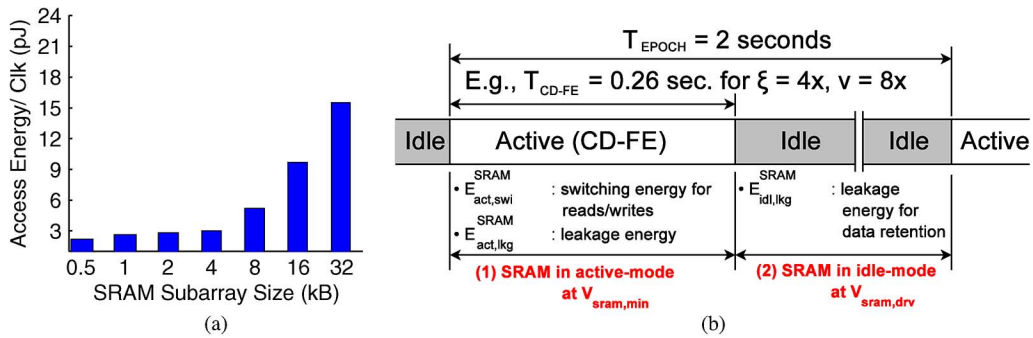


Fig. 13. (a) SRAM access energy (from NanoSim [18] at 0.7 V) is lower for smaller-sized arrays and (b) summary of energy components contributing to total SRAM energy (the $\xi = 4\times, \nu = 8\times$ case is shown for illustration).

III. Low-Energy Compressed-Domain Processor

In this section, we describe the circuits used in the compressed-domain processor. We also present an analysis of SRAM energy, which will help us understand the impact of energy scalability in the processor with respect to ξ and ν.

A. Circuits

Fig. 12 shows the circuits used in the compressed-domain processor. The CPF is selectable for front-end signal compression. It uses a 16 b linear feedback shift register (LFSR) to implement multiplication with a random projection matrix $\mathbf{\Phi}$, as shown. Since the processor operates on an EEG epoch of 2 seconds, FVs are derived at the rate of 0.5 Hz. At this low rate, the CD-FE can compute each feature dimension sequentially and store the intermediate results in a data buffer. The CD-FE can be configured to compute up to eight spectral features $(i = 0, \ldots, 7)$ for each EEG channel (j) over as many as 18 channels, yielding a maximum FV dimensionality of 144. Within the CD-FE, the control pulse S_0 initiates CD-BPF computations. A multiply-accumulate (MAC) unit (M0) is used to perform the matrix multiplications required for compressed-domain band-pass filtering using $\hat{\mathbf{H}}_\mathbf{i}$. Each filtered EEG epoch is then registered by the control pulse S_1. Energy accumulation over the output vector is then performed by a second MAC unit (M1). After the feature-extraction process [which requires $(N/\xi)(N/\xi + 1)$ MAC operations], each FV dimension $(\hat{f}_{ij})$ is stored in an intermediate FV buffer based on the control pulse S_2.

B. SRAM Energy Analysis

In our implementation, we represent filter coefficients using 8 bits of precision. Thus, to support CD-FE computations, the processor requires a maximum of 32 kB accesses per second from the memory bank. Fig. 13(a) shows that the SRAM energy per access (E_{acc}^{sram}) is reduced by choosing smaller-sized subarrays [19]. Since the ξ and ν knobs scale the memory required, designing a single 32 kB array would be sub-optimal for many of the parameter points. Instead, we design four subarrays (each of size 8 kB) to balance savings in energy per access with the overhead of further partitioning. With sub-array partitioning, leakage-energy saving can be achieved by independently power-gating each sub-array (from off-chip).

After the CD-FE computations, each FV is processed by the SVM block within the epoch duration of two seconds. The SVM can apply linear, polynomial, or radial-basis function (RBF) kernel transformations (*via* an embedded CORDIC engine). The support vectors are derived from offline training of the classifier and are provided through a dedicated interface. The classification result is encoded in the most significant bit (MSB) of the SVM output (MSB $= 1$ for seizure detected, MSB $= 0$ for no seizure detected).

The CD-FE energy comprises the logic and SRAM energy subcomponents. The SRAM consumes a substantial portion of the total CD-FE energy. Its optimization to exploit scalability with respect to ξ and ν is thus a key factor. The detector processes an EEG epoch every $T_{EPOCH} = 2$ sec. However, the optimal operating frequency (and supply voltage $V_{dd,opt}$) for the CD-FE logic is determined by minimizing the overall CD-FE energy, while ensuring a minimum throughput that allows the active CD-FE computations to be completed in $T_{CD\text{-}FE}$ (< 2) seconds for each value of ξ and ν. For the remainder of the epoch (i.e., $T_{EPOCH} - T_{CD\text{-}FE}$), the logic and SRAMs can be placed in low-energy idle modes.

Fig. 13(b) summarizes the SRAM operating modes and energies [19]. The total SRAM energy is the sum of the active-mode (E_{act}^{SRAM}) and idle-mode (E_{idl}^{SRAM}) energies for each subarray (numbering N_{sub}) that is enabled; under the assumption that the SRAMs cannot by fully power-gated in order to ensure data retention, E_{idl}^{SRAM} is not zero. During the active mode, the SRAM operates at the minimum operational supply voltage ($V_{sram,min}$) of 0.7 V for reads and writes; at this voltage, it operates at 920 kHz; this is sufficient performance for all design points (ξ, ν) of the CD-FE, thus allowing the SRAM voltage to remain at 0.7 V. During the idle mode, the SRAM operates at its minimum data-retention voltage ($V_{sram,drv}$) of 0.42 V.

In the active mode, while set to a supply voltage of $V_{sram,min}$, E_{act}^{SRAM} comprises active-switching ($E_{act,swi}^{SRAM}$) and leakage ($E_{act,lkg}^{SRAM}$) energies for a period of $T_{CD\text{-}FE}$. In the idle mode, while set to a supply voltage of $V_{sram,drv}$, E_{idl}^{SRAM} comprises only the leakage energy ($E_{idl,lkg}^{SRAM}$) for the duration ($T_{EPOCH} - T_{CD\text{-}FE}$). Thus, we can represent the SRAM energy components as follows:

$$\begin{aligned} E_{lkg}^{SRAM} &= E_{act,lkg}^{SRAM} + E_{idl,lkg}^{SRAM} \\ &= N_{sub} T_{CD\text{-}FE} \left\{ I_{V_{sram,min}} V_{sram,min} \right\} \\ &\quad + N_{sub} (T_{EPOCH} - T_{CD\text{-}FE}) \\ &\quad \times \left\{ I_{V_{sram,drv}} V_{sram,drv} \right\} \qquad (5) \\ E_{act,swi}^{SRAM} &= E_{acc}^{sram} \times \#\text{accesses} \qquad (6) \end{aligned}$$

The duration of the active mode ($T_{CD\text{-}FE}$) in (5) depends on ξ, ν, and the optimum logic voltage $V_{dd,opt}$. For smaller (larger) values of ξ and ν, there are more (fewer) coefficients in $\hat{\mathbf{H}}_\mathbf{i}$ and $T_{CD\text{-}FE}$ (the active CD-FE time) is higher (lower). For instance, $T_{CD\text{-}FE}$ is 0.26 sec. for $\xi = 4\times$ and $\nu = 8\times$, as shown in Fig. 13(b). It increases to 0.52 sec. at $\xi = \nu = 4\times$ and reduces to 0.13 sec. at $\xi = 4\times$ and $\nu = 16\times$. Further, the number of active subarrays (N_{sub}) is also a function of ξ and ν; Fig. 14 shows this dependence. Eqs. (5) and (6) also show that although $E_{act,swi}^{SRAM}$ remains invariant to changing values of V_{dd}, it is impacted by ξ and ν (since #accesses changes with ξ and ν). Note that in (6), E_{acc}^{sram} denotes the active-switching energy per access, which remains invariant to changing values of V_{dd}, ξ, and ν. Similar to $E_{act,swi}^{SRAM}$, the SRAM leakage energy E_{lkg}^{SRAM} also scales substantially with ξ and ν. Consequently, the optimal logic voltage $V_{dd,opt}$, which minimizes the SRAM and logic CD-FE energy, changes with respect to ξ and ν. We study the variation of $V_{dd,opt}$ in the next section.

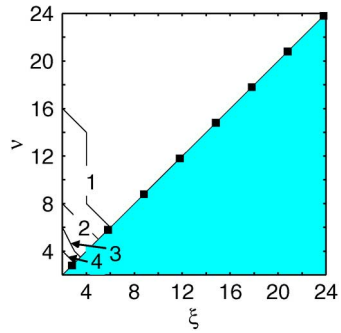


Fig. 14. N_{sub} scales substantially with ξ and ν, affecting the SRAM leakage energy.

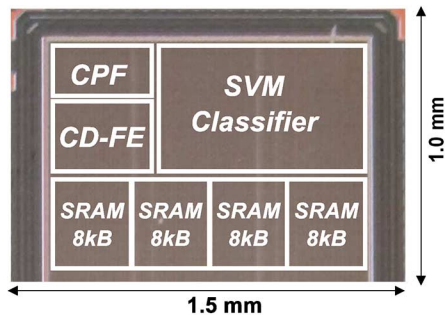


Fig. 15. Die photo of IC.

IV. Measurement Results

The IC was prototyped in a 0.13 μm CMOS process from IBM. The die photograph and performance summary are shown in Fig. 15 and Table I, respectively. 18 channels of Nyquist EEG signals are sampled at a rate of 256 Hz, and eight CD-BPFs are derived corresponding to eight Nyquist-domain BPFs, each of order $k = 64$ (based on the filter specifications required for seizure detection [7]). This leads to a total FV dimensionality of 144. Table I shows that the CPF permits EEG compression by a factor of $\xi =$ 2-24×, consuming 85-7.3 pJ of energy. In the CPF, a 16-bit LFSR generates a sequence of ± 1 values. Based on these values, each compressively-sensed signal sample is computed serially as $\hat{\mathbf{x}}_\mathbf{i} = \mathbf{x_1} \pm \mathbf{x_2} \pm \ldots \pm \mathbf{x_N}$. This process is repeated N/ξ times to provide the compressively-sensed signal $\hat{\mathbf{x}}$.

The total processor energy is in the range 0.3-2.2 μJ (for linear SVMs), 12.6-38.5 μJ [for nonlinear SVMs using a fourth-order polynomial kernel (poly4)], and 18.1-53.3 μJ (for SVMs with an RBF kernel). Since classification results are produced every two seconds (i.e., FVs are processed at a rate of 0.5 Hz), the total processor power lies in the range 0.6-107 μW for all SVM kernels. Fig. 16 shows a scatter plot of the first two principal components of the measured FVs from the IC. The FVs are derived using all data records for patient #1 in the CHB-MIT dataset. We observe from the figure that there is a good separation between the seizure and non-seizure

TABLE I
PERFORMANCE SUMMARY: ENERGY-SCALABLE, COMPRESSED-DOMAIN PROCESSOR IC

Technology	IBM 130nm LP CMOS	**SUBBLOCK ENERGY MEASUREMENTS**		
Supply voltage	CD-FE: 1.2-0.44 V		per FV	per Clock
	SRAM: 0.7/0.42 V	CPF (at 0.48 V)	85.0-7.3 pJ	10.6 fJ
	CPF/ SVM: 0.48 V	CD-FE logic (at $V_{dd,opt}$)	70.8-1.3 nJ	1.3 pJ
EEG sampling rate	256 Hz	SRAM subarray (at 0.7 V)	2.1-0.1 μJ	5.0 pJ
Clock frequency	10.2–0.3 MHz	**Total Feature Extraction**	**2.1 μJ-93.2 nJ**	**6.3 pJ**
CPF compression factor ξ	2-24×	SVM RBF	16.0-53.2 μJ	6.0 pJ
Projection factor ν	2-24×	SVM Poly4	10.5-38.4 μJ	4.8 pJ
Feature computation rate	0.5 Hz	SVM Linear	62.9-209.0 nJ	2.0 pJ
CD-BPF memory size	0.44-32 kB	**Total Processor (linear SVM)**	**2.2-0.3 μJ**	**8.3 pJ**

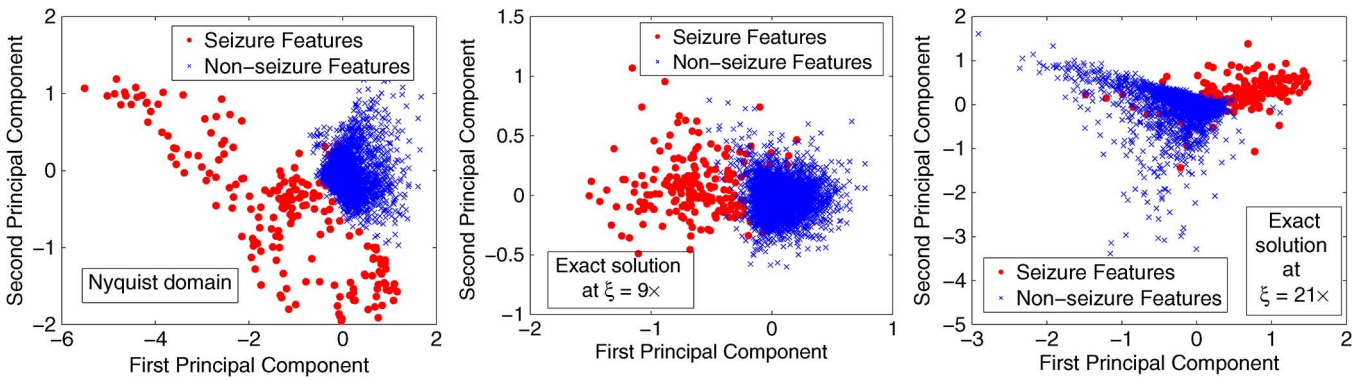


Fig. 16. Distribution of FVs derived from the IC using data from Patient # 01 in the CHB-MIT database. Although full feature data dimensionality is used for classification, projection to two dimensions *via* PCA is shown to aid visualization. Results are shown for (a) Nyquist-domain processing, (b) exact solution at $\xi = 9\times$, and (c) exact solution at $\xi = 21\times$.

FVs in the compressed domain even at $\xi = 21\times$. The results presented next consider the impact of ξ and ν scaling on the feature-extractor, classifier, and overall processor energies.

A. Determining the Optimal Voltage for the CD-FE Logic

As described in the previous section, the SRAM leakage energy changes with both ξ and ν. Thus, the optimal voltage ($V_{dd,opt}$) for the CD-FE logic changes with both ξ and ν. In order to determine $V_{dd,opt}$, we minimize the total CD-FE energy comprising the logic and SRAM energies.

Fig. 17 shows the measured subcomponents of the CD-FE energy with respect to V_{dd} when N_{sub} ranges from 1 to 4 (corresponding to four different values of ξ and ν). For all values of N_{sub}, the active energy (E_{swi}^{logic}) of the CD-FE logic increases and the leakage energy (E_{lkg}^{logic}) decreases with increasing values of V_{dd}, leading to the minimum-energy point of 0.46 V [20]. However, this is not $V_{dd,opt}$ since we need to also consider the SRAM energy. The SRAM operates at 0.7 V in the active mode. We see from Fig. 17 that the SRAM active-mode switching energy $E_{act,swi}^{SRAM}$ does not change with V_{dd} [consistent with (6)]. Further, the leakage energies in the active ($E_{act,lkg}^{SRAM}$) and idle modes ($E_{idl,lkg}^{SRAM}$) increase as N_{sub} increases. This is also expected since from (5), E_{lkg}^{SRAM} depends on N_{sub}. However, since both $E_{act,lkg}^{SRAM}$ and $E_{idl,lkg}^{SRAM}$ also depend on $V_{dd,opt}$, the increase in the leakage energies is not proportional to the increase in N_{sub} [(5)]. The total CD-FE energy is thus a nonlinear function of ξ and ν, which necessitates $V_{dd,opt}$ to be determined numerically.

Fig. 18 shows the measured CD-FE energy at different voltage values for the cases considered in Fig. 17. For these four instances, we see from the figure that $V_{dd,opt}$ for the CD-FE logic is either 0.48 V or 0.5 V. The corresponding frequencies are determined to be 380 or 400 kHz, respectively, from the frequency vs. V_{dd} plot in Fig. 19. With more measurements, we determine $V_{dd,opt}$, frequency, and active time ($T_{CD\text{-}FE}$) for the CD-FE logic when ξ and ν vary in the 2-24× range. The results are shown in Figs. 20(a), (b), and (c), respectively. For all values of ξ and ν, $T_{CD\text{-}FE}$ varies in the 0.9-0.02 sec. range and is below the epoch time of 2 sec., which allows sufficient time (T_{EPOCH}–$T_{CD\text{-}FE}$) for the SVM classifier to finish computing.

B. Feature-extractor Energy

As mentioned previously, the CD-FE energy comprises the logic and SRAM energies. In this section, we provide measurement results for these energy subcomponents using both the exact and approximate solutions for $\hat{\mathbf{H}}_\mathbf{i}$.

CD-FE Logic Energy: Recall from Section II-C that for the exact solution (i.e., when $\xi = \nu$), the CD-FE complexity scales quadratically with ξ; for the approximate solution, it scales linearly with both ξ and ν. Figs. 21(a) and (b) show the CD-FE energy for the exact and approximate solutions, respectively. For each value of ξ and ν, the energy is reported for $V_{dd,opt}$, which minimizes CD-FE's active-switching plus leakage energies as well as the SRAM energy; the $V_{dd,opt}$ values are also annotated in Fig. 21(a).

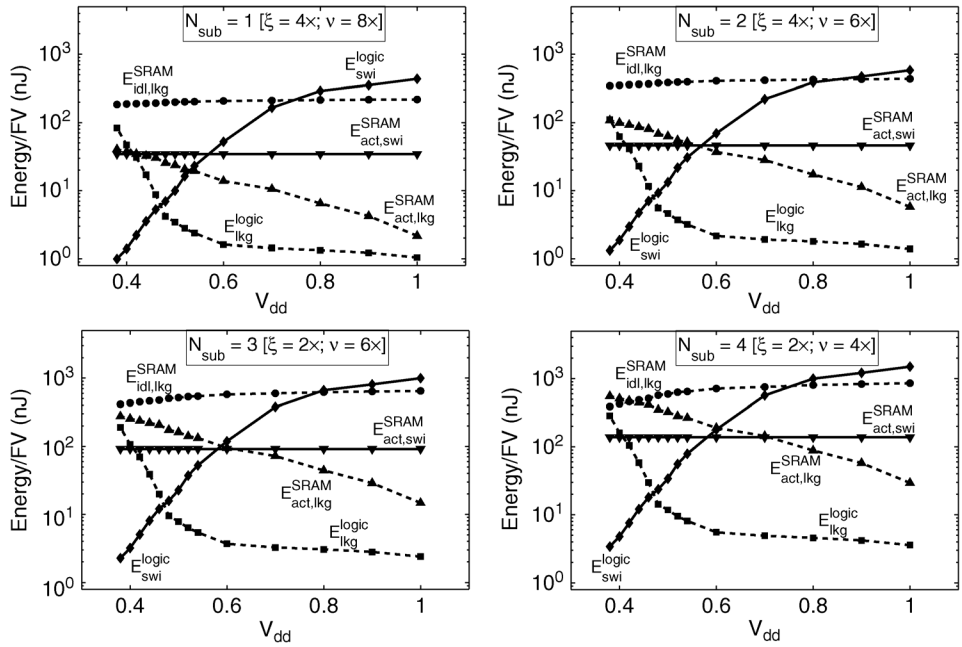


Fig. 17. The CD-FE energy subcomponents have nonlinear dependence on ξ and ν. Primarily, the SRAM leakage energy in the active ($E^{SRAM}_{act,lkg}$) and idle mode ($E^{SRAM}_{idl,lkg}$) is substantially impacted by N_{sub} and $T_{CD\text{-}FE}$. The active-mode SRAM switching energy ($E^{SRAM}_{act,swi}$) changes with ξ and ν, but not with V_{dd}.

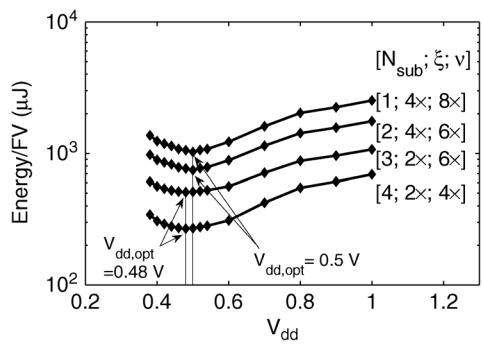


Fig. 18. For the four cases considered in Fig. 17, the optimal voltage for the CD-FE logic ($V_{dd,opt}$) is either 0.48 V or 0.5 V.

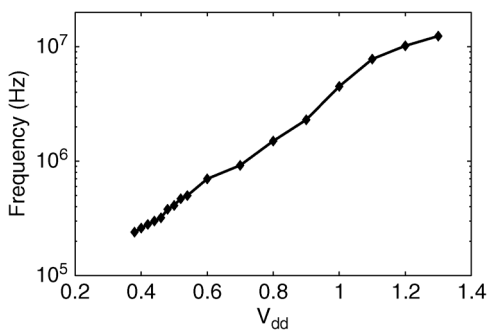


Fig. 19. The operating frequency vs. V_{dd} for the CD-FE logic.

SRAM Energy: Figs. 22(a) and (b) show the SRAM leakage energies in the idle and active modes and Fig. 22(c) shows the SRAM switching energy in the active mode, versus ξ and ν. We can see from the figures that for smaller values of ξ and ν, since the size of $\hat{\mathbf{H}}_\mathbf{i}$ is larger, $T_{CD\text{-}FE}$ is higher and the SRAM active energy dominates the idle-mode energy. This is also consistent with a higher value of $V_{dd,opt}$ at these values of ξ and ν, which enables the CD-FE computations to finish sooner. In contrast, at larger values of ξ and ν, however, there are fewer coefficients in $\hat{\mathbf{H}}_\mathbf{i}$ and the SRAM spends most of the time in the idle mode. This behavior is clear from Figs. 23(a) and (b), which show the total SRAM energy for the exact and approximate solutions. The figures show that the total SRAM energy is nearly equal to the SRAM idle-mode energy at higher values of ξ and ν. Further, the figures also show substantial scaling in the total SRAM energy and in its constituents with respect to ξ and ν. This scaling occurs due to the variation in N_{sub} and $T_{CD\text{-}FE}$ due to ξ and ν [see (5) and (6) and Figs. 14 and 20(c)]. The SRAM energy thus eventually begins to saturate due to the granularity limit of the four subarrays; a finer granularity would enhance scaling at the cost of additional hardware overhead.

Total Feature-extraction Energy: From the above results, we see that the SRAM energy can significantly dominate the CD-FE logic energy at all values of ξ and ν. This behavior validates the focus on optimizing the SRAM energy in Section III-B; for instance, at $\xi = 4\times$ and $\nu = 2\times$, the total SRAM energy is 2.1 μJ and the CD-FE logic energy is 70.8 nJ. The contribution of the energy subcomponents is also apparent in the total CD-FE energy plots shown for the exact and approximate solutions shown in Figs. 24(a) and (b), respectively (results are for 18 EEG channels with eight CD-BPFs). These plots show that the CD-FE energy profile is similar to the SRAM energy profile presented in the previous section.

Comparison With Nyquist-domain Processing: Since $\mathbf{H}_\mathbf{i}$ are Toeplitz matrices implementing convolution, the filter order determines the number of non-zero coefficients in $\mathbf{H}_\mathbf{i}$ (see Fig. 10), which in turn determine the feature-extraction energy in the Nyquist domain. However, in the compressed domain, due to the loss of regularity in $\hat{\mathbf{H}}_\mathbf{i}$, the feature-extraction energy does not depend on the filter order in the same way. Thus, in the compressed domain, the energy can initially increase due to loss of regularity in $\hat{\mathbf{H}}_\mathbf{i}$, but then can eventually decrease owing to scaling in the size of $\hat{\mathbf{H}}_\mathbf{i}$ due to both ξ and ν. Further, at a given value of ξ and ν, we can scale the total CD-FE energy by the ratio of the number of non-zero coefficients in $\mathbf{H}_\mathbf{i}$ to the

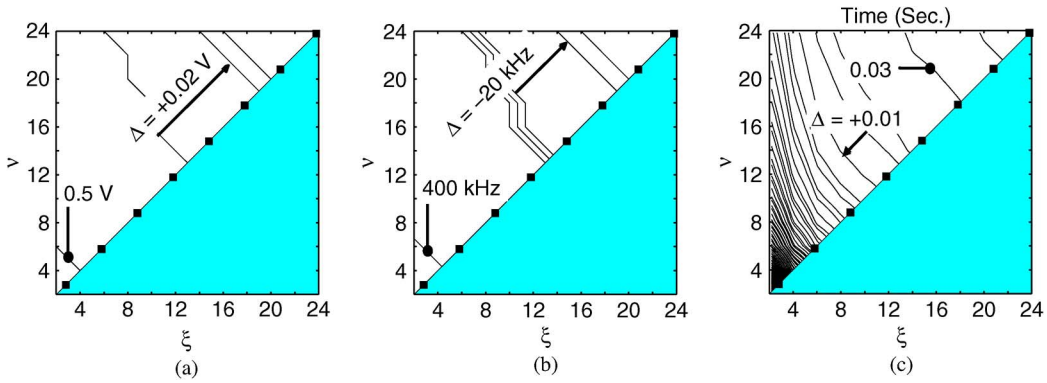


Fig. 20. As ξ and ν scale in the 2-24× range, (a) the optimal voltage for the CD-FE logic ($V_{dd,opt}$) varies in the 0.5-0.44 V range, (b) the corresponding operating frequency varies in the 400-300 kHz range, and (c) the CD-FE active time ($T_{CD\text{-}FE}$) varies in the 0.9-0.02 sec. range.

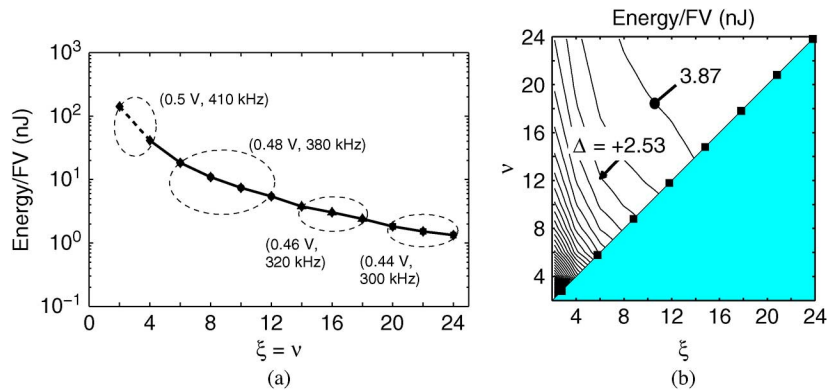


Fig. 21. The CD-FE logic energy for (a) the exact solution and (b) the approximate solution measured at $V_{dd,opt}$. The energy scales substantially with ξ and ν.

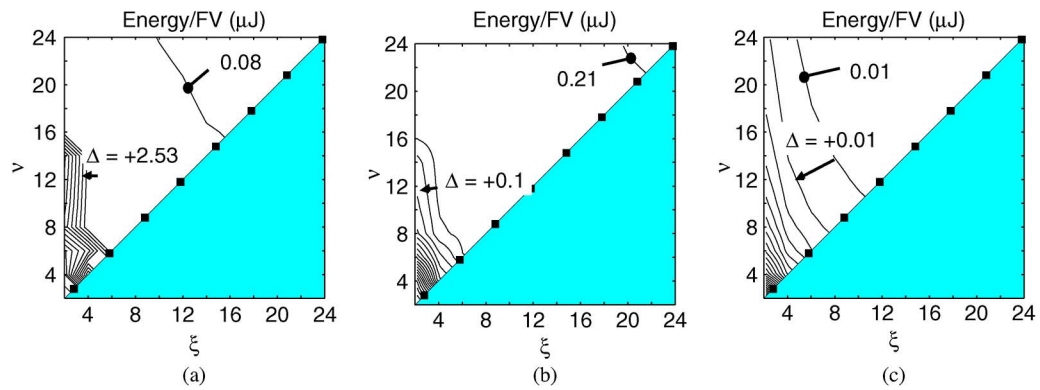


Fig. 22. Each of the SRAM energy subcomponents, i.e., (a) idle-mode leakage ($E^{SRAM}_{idl,lkg}$), (b) active-mode leakage ($E^{SRAM}_{act,lkg}$), and (c) active-mode switching ($E^{SRAM}_{act,swi}$) scales with both ξ and ν. $E^{SRAM}_{act,lkg}$ tends to dominate at smaller values of ξ and ν.

number in $\hat{\mathbf{H}}_\mathbf{i}$ to derive an estimate for the Nyquist-domain feature-extraction energy. Fig. 24 shows that for the exact solution, at $\xi > 4\times$, the total energy of compressed-domain processing is less than that projected for Nyquist-domain processing (for a 64-order FIR filter).

C. Classifier Energy

One downside of directly processing compressively-sensed EEG is that the SVM model for classification can become somewhat more complex at higher values of ξ and ν. Intuitively, this happens due to the additional error introduced in the FVs when we solve the compressed-domain equations [(2)], which necessitates complex decision boundaries in the classifier. Figs. 25(a), (b), and (c) show the classifier energy for the approximate solution using three kernels: RBF, 4th-order polynomial (poly4), and linear, respectively. Fig. 26 shows the SVM energy for the exact solution using the same three kernels. In each of these cases, the SVM operates at its minimum-energy point of 0.48 V. From Fig. 25, we can see that the increase in classifier energy opposes the reduction in CD-FE energy. We can also see that the SVM energy increase becomes worse when ν is significantly higher than ξ, which reflects the extra error introduced at the algorithmic level due to a degradation in the JL-approximation.

D. Total Processor Energy

Fig. 27 shows the effect of ξ scaling on the total processor energy for the exact solution. Figs. 28(a), (b), and (c) show the effect of ξ and ν scaling on the total processor energy for the approximate solution using the RBF, poly4, and linear classification kernels, respectively. The SVM operates at 0.48 V, the CD-FE operates at $V_{dd,opt}$ [specified in Fig. 20(a)], and the

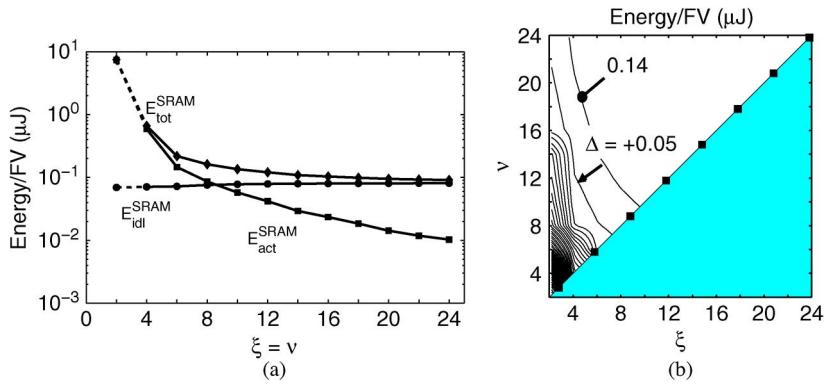


Fig. 23. The total SRAM energy for (a) the exact solution and (b) the approximate solution scales substantially at smaller values of ξ and ν. At higher values of ξ and ν, CD-BPF matrices $\hat{\mathbf{H}}_\mathbf{i}$ are smaller, which makes $E_{idl,lkg}^{SRAM}$ dominate the total SRAM energy.

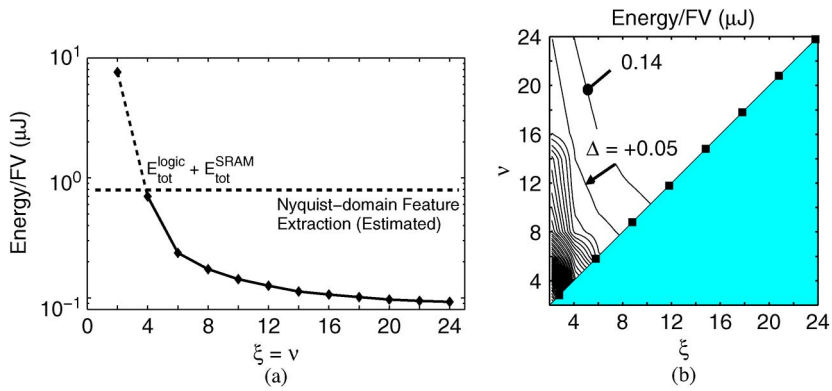


Fig. 24. The total CD-FE energy (logic + SRAM) for (a) the exact solution and (b) the approximate solution. The SRAM energy tends to dominate and thus provides scalability with both ξ and ν.

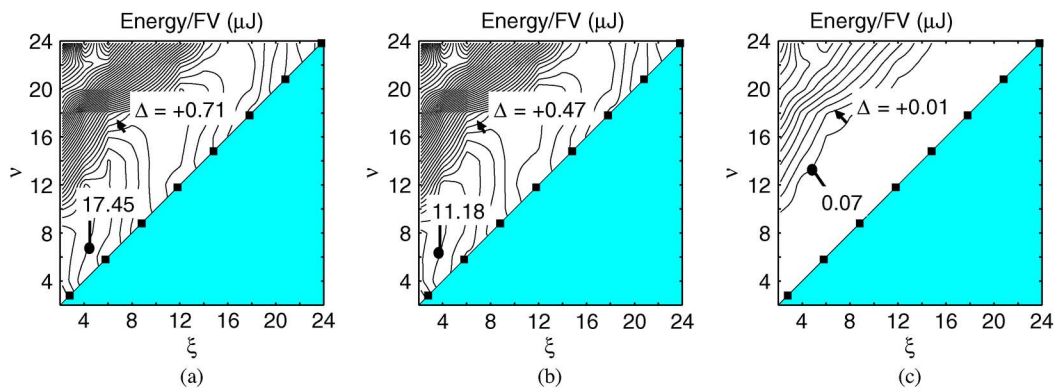


Fig. 25. The SVM classifier energy measured at the minimum-energy point of 0.48 V for the approximate solution using (a) RBF, (b) poly4, and (c) linear kernel. At a given values of ξ, the classification energy increases with increasing values of ν.

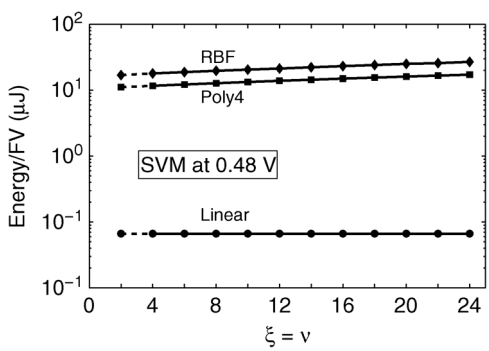


Fig. 26. SVM classifier energy for the exact solution.

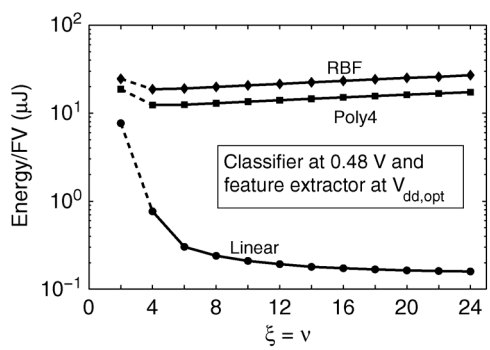


Fig. 27. Total processor energy for the exact solution.

SRAMs operate at 0.7/0.42 V during the active/idle modes. The figures show that nonlinear SVM kernels (i.e., RBF and poly4) consume significant energy, while SVMs with a linear kernel incur minimal energy, causing the energy-scaling characteristics to be dominated by CD-FE at all values of ξ and ν. For the nonlinear cases, the SVM energy actually leads to optimal ξ and

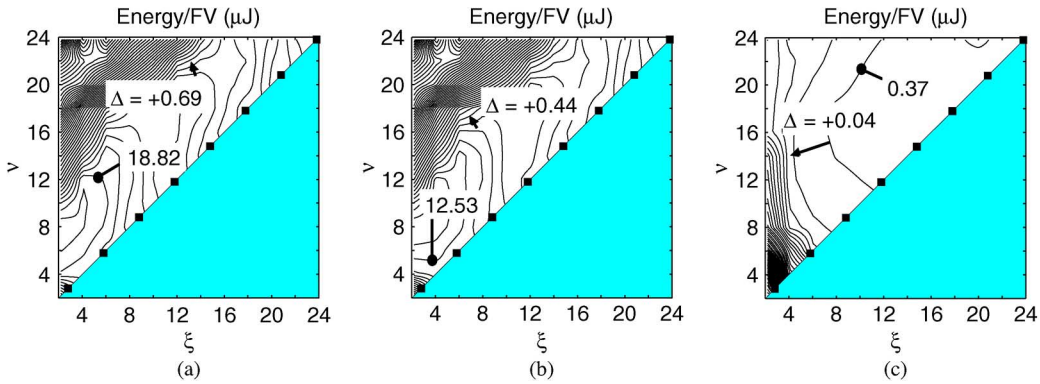


Fig. 28. Total processor energy for the approximate solution using (a) RBF, (b) poly4, and (c) linear kernel. For nonlinear SVMs, classifier energy dominates due to the extra modeling complexity in the compressed domain, while for linear kernels, CD-FE energy is higher and permits substantial scalability with ξ and ν.

ν values (e.g., for the exact solution, from Fig. 27, an optimal ξ of approximately 5$\times$ minimizes the total processor energy).

E. Summary of Measurement Results

The following is a summary of measurement results from the IC.

- In Section IV-A, we presented the energy measurements from the CD-FE block (logic and SRAM) at different values of the logic supply voltage. Since the total CD-FE energy exhibits a nonlinear relationship with respect to ξ and ν, we empirically determined the optimal CD-FE logic voltage, $V_{dd,opt}$, such that it minimizes the total CD-FE energy at a given value of ξ and ν.
- In Section IV-B, we presented energy measurements from the CD-FE block vs. ξ and ν. The CD-FE SRAM energy comprises the active- and idle-mode energies. We observed that at smaller values of ξ and ν, active-mode SRAM leakage, $E^{SRAM}_{act,lkg}$, tends to be the dominant component while at higher values of ξ and ν, the idle-mode SRAM leakage, $E^{SRAM}_{idl,lkg}$, is dominant. Further, the CD-FE logic and SRAM energy measurements showed that for values of $\xi > 4\times$, the total feature-extraction energy in the compressed domain is lower than that in the Nyquist domain.
- In Section IV-C, we presented measurement results for the SVM classifier. We observed that the classification energy can dominate the feature-extraction energy when we use compressed-domain processing with nonlinear SVM kernels. However, for linear kernels, feature-extraction energy dominates and compressed-domain processing can provide substantial energy scalability with respect to ξ and ν. Further, in Section IV-D, we also provided energy measurements from the processor (feature extraction + classification), which showed a similar trend as the classifier for the linear and nonlinear SVM kernels.

V. Conclusions

Sparsity of signals provides an opportunity to efficiently represent sensor data. Compressive sensing is one technique that exploits signal sparsity in a secondary basis to achieve very low-energy compression on the sensing node. The random projections in compressive sensing, however, affect the sensed signals, preventing the use of Nyquist-domain algorithms for signal analysis. Moreover, signal reconstruction is energy-intensive and is not desirable on low-power sensor nodes. We presented an approach to overcome these limitations in systems that use compressive sensing. We transform computations from the Nyquist domain to the compressed domain, enabling us to perform computations directly on compressively-sensed data. In particular, we presented the design of a processor that enables on-node signal analysis to detect epileptic seizures directly using compressively-sensed EEG. By using an exact solution for the compressed-domain filtering matrices, we showed that the performance of the compressed-domain detector is retained up to high compression factors. Additionally, by using an approximate solution, we derived smaller-sized compressed-domain filtering matrices, saving us more energy in the compressed domain. We showed that these methods provide us with two strong knobs to control the energy of the compressed-domain seizure-detection processor. Thus, in addition to communication energy savings, through end-to-end data reduction in a system, our methodology enables a mode of power management where the computational energy scales due to both a reduction in the number of input samples that need to be processed and due to approximations introduced at the algorithmic level.

Acknowledgment

The authors thank Ali H. Shoeb for providing the code for the baseline Nyquist-domain seizure-detection algorithm. They also thank Xiuyuan Cheng and Amit Singer from the Applied and Computational Mathematics Department at Princeton University for valuable discussions on low-dimensional error bounds for the compressed-domain equations.

References

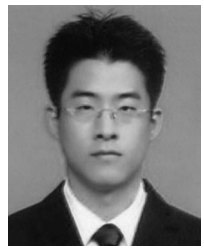
[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
[2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.
[3] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.
[4] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2456–2466, Sep. 2011.
[5] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

[6] A. Csavoy, G. Molnar, and T. Denison, "Creating support circuits for the nervous system: Considerations for brain-machine interfacing," in *Proc. Int. Symp. VLSI Circuits*, Jun. 2009, pp. 4–7.
[7] N. Verma, A. Shoeb, J. Guttag, and A. Chandrakasan, "A micro-power EEG acquisition SoC with integrated seizure detection processor for continuous patient monitoring," in *Proc. Symp. VLSI Circuits*, Jun. 2009, pp. 62–63.
[8] M. Shoaib, N. K. Jha, and N. Verma, "Enabling advanced inference on sensor nodes through the direct use of compressively-sensed signals," in *Proc. IEEE Design, Automation, and Test in Europe Conf.*, Mar. 2012, pp. 437–443.
[9] M. Shoaib, N. K. Jha, and N. Verma, "A compressed-domain processor for seizure detection to simultaneously reduce computation and communication energy," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2012, pp. 1–4.
[10] M. Shoaib, N. K. Jha, and N. Verma, "Signal processing with direct computations on compressively-sensed data," *IEEE Trans. VLSI Syst.*, under review.
[11] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE Int. Wkshp. Statistical Signal Process.*, Aug. 2007, pp. 181–184.
[12] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, Jun. 2008.
[13] W. B. Mendelson, D. A. Sack, S. P. James, J. V. Martin, R. Wagner, D. Garnett, J. Milton, and T. A. Wehr, "Frequency analysis of the sleep EEG in depression," *Psychiatry Research*, vol. 21, no. 2, pp. 89–94, 2007.
[14] A. Shoeb and J. Guttag, "Application of machine learning to seizure detection," in *Proc. Int. Conf. Machine Learning*, Jun. 2010.
[15] M. Rudelson and R. Vershynin, Non-Asymptotic Theory of Random Matrices: Extreme Singular Values, arXiv preprint arXiv:1003.2990, Apr. 2010.
[16] S. Dasgupta and A. Gupta, "An elementary proof of the Johnson-Lindenstrauss lemma," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60–65, 2002.
[17] Physionet, CHB-MIT Physionet Database [Online]. Available: http://www.physionet.org/physiobank/database
[18] Synopsys Inc., NanoSim: A Next-Generation Solution for SoC Integration Verification, Apr. 2013 [Online]. Available: http://www.synopsys.com/Tools/Verification/Pages/socintegration.aspx
[19] N. Verma, "Analysis towards minimization of total SRAM energy over active and idle operating modes," *IEEE Trans. VLSI Syst.*, vol. 19, pp. 1695–1703, Sep. 2011.
[20] A. Wang and A. P. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.

Mohammed Shoaib (S'08) received the B.Tech. and M.Tech. degrees in electrical engineering with a specialization in microelectronics and VLSI design from the Indian Institute of Technology (IIT), Madras, India, in 2007 and 2008, respectively. He received the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 2010 and 2013, respectively.

His research focuses on data-driven platforms for intelligent biomedical systems, emphasizing low-energy circuits and signal processing. He has authored or coauthored more than 20 technical papers and holds 4 U.S. patents.

Dr. Shoaib was a recipient of the IIT Madras MCM scholarship from 2004 to 2007. He also received a Ph.D. fellowship and an excellence in electrical engineering award from Princeton University in 2008 and 2011, respectively. He was a corecipient of the Qualcomm Innovation Ph.D. Fellowship 2011, a nominee for the best paper award at Healthcomm 2011, and a recipient of the Roberto-Padovani scholarship 2011, Harold W. Dodds (Princeton University Honorific) Fellowship 2012, and Princeton University Gordon Wu Prize for Excellence 2012. He is a fellow of the McGraw Center for Teaching and Learning and has had industrial experience through internships with research groups at IBM Zurich Research Laboratory, Ricoh California Research Center, and Qualcomm Corporate Research and Development, CA.



Kyong Ho Lee received the B.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 2004 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2009. He is currently working towards the Ph.D. degree in electrical engineering at Princeton University, Princeton, NJ, USA.

His research interest includes energy-efficient signal processing for biomedical applications, and ultralow energy biomedical circuit design employing machine learning techniques.

Mr. Lee was a co-recipient of the 2011 Qualcomm Innovation Fellowship.

Niraj K. Jha (S'85–M'85–SM'93–F'98) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1981, the M.S. degree in electrical engineering from S.U.N.Y. Stony Brook, NY, USA, in 1982, and the Ph.D. degree in electrical engineering from the University of Illinois, Urbana, IL, USA, in 1985.

He is a Professor of electrical engineering at Princeton University, Princeton, NJ, USA. He has co-authored or co-edited five books, 12 book chapters, and more than 390 technical papers. He has co-authored 14 papers, which have won various awards, and another six papers that have been nominated for best paper awards. He has received 14 U.S. patents. His research interests include FinFETs, low power hardware/software design, computer-aided design of integrated circuits and systems, quantum computing, and secure computing. He has given several keynote speeches in the area of nanoelectronic design and test.

Prof. Jha is a Fellow of IEEE and ACM. He has served as the Editor-in-Chief of IEEE TRANSACTIONS ON VLSI SYSTEMS and an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I AND II, IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN, IEEE TRANSACTIONS ON VLSI SYSTEMS, and *Journal of Electronic Testing: Theory and Applications*. He is currently serving as an Associate Editor of IEEE TRANSACTIONS ON COMPUTERS, *Journal of Low Power Electronics* and *Journal of Nanotechnology*. He has also served as the Program Chairman of the 1992 Workshop on Fault-Tolerant Parallel and Distributed Systems, the 2004 International Conference on Embedded and Ubiquitous Computing, and the 2010 International Conference on VLSI Design. He has served as the Director of the Center for Embedded System-on-a-chip Design funded by New Jersey Commission on Science and Technology. He is the recipient of the AT&T Foundation Award and NEC Preceptorship Award for research excellence, NCR Award for teaching excellence, and Princeton University Graduate Mentoring Award.

Naveen Verma (M'09) received the B.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2005 and 2009, respectively.

Since July 2009, he has been an Assistant Professor of electrical engineering at Princeton University, Princeton, NJ, USA. His research focuses on ultra-low-power integrated circuits and systems with an emphasis on sensing applications. Of particular importance is the use of emerging devices for the creation of functionally-diverse systems and the use of advanced signal-analysis frameworks for low-power inference over embedded signals. At the circuit level, his focus spans low-voltage digital logic and SRAMs, low-noise analog instrumentation and data-conversion, and integrated power management.

Dr. Verma was a co-recipient of 2008 ISSCC Jack Kilby Award for Outstanding Student Paper, and 2006 DAC/ISSCC Student Design Contest Award. He was a recipient of the Alfred Rheinstein Junior Faculty award at Princeton, the 2013 NSF CAREER award, and the 2013 Intel Early Career Award.