A Seizure-detection IC Employing Machine Learning to Overcome Data-conversion and Analog-processing Non-idealities

Jintao Zhang, Liechao Huang, Zhuo Wang, and Naveen Verma

Princeton University, Princeton, NJ 08544, USA

***Abstract* — This paper presents a seizure-detection system wherein the accuracy required of the analog frontend is substantially relaxed. Typically, readout of electroencephalogram (EEG) signals would dominate the energy of such a system, due to the precision (noise, linearity) requirements. The presented system performs data conversion and analog multiplication for EEG feature extraction via simple circuits to demonstrate that feature errors can be overcome by appropriate retraining of a classification model, using a machine-learning algorithm. This precludes the need to design a high-precision frontend. The prototype, in 32nm CMOS, results in features whose RMS error normalized to their ideal values is 1.16 (i.e. errors are larger than ideal values). An ideal implementation of the seizure detector exhibits sensitivity, latency, false alarms of 5/5, 2.0 sec., 8, respectively. The feature errors degrade this to 5/5, 3.6 sec., 443, causing high false alarms; but retraining of the classification model restores this to 5/5, 3.4 sec., 4.**

***Index Terms* — Machine learning, analog processing circuits, electroencephalography, epilepsy, system-on-chip, error compensation.**

I. Introduction

Increasingly, embedded sensing systems aim to perform local analysis of sensor signals to address communication constraints and/or enable increasingly high-value functions within the device. This poses two challenges. The first challenge is that the sensor signals are often derived from complex physical processes, which typically cannot be modeled analytically. Machine-learning algorithms have thus become invaluable in sensing applications, because they enable data-driven methods for constructing high-order models for signal analysis.

The second challenge is that system resources (energy, area, etc.) are highly constrained. Often the analog frontend used for sensor readout poses the critical bottleneck. As an example, in the seizure-detection system presented in [1], the instrumentation and data-conversion blocks dominate system performance and power, consuming over 60% of the total power. The reason for this is that the instrumentation amplifiers and ADCs must be designed for adequately low noise (device, quantization), nonlinearity (gain-compression, INL), etc. Relaxing these requirements has the potential to substantially mitigate the performance and energy bottlenecks, but will degrade the quality of sensor readout.

This paper explores how machine-learning algorithms, employed to enable analysis over physically complex sensor signals, can also be exploited to overcome non-idealities in the analog frontend. The concept of Data-driven Hardware Resilience (DDHR) was previously presented in the context of overcoming digital faults, leading to computational errors [2]. DDHR leverages data-driven training using data derived from non-ideal system hardware to construct a model for inference (e.g., classification). The resulting model is referred to as an *error-aware model*, and is shown to enable high classification performance even in the presence of severe computational errors arising due to highly non-ideal hardware. Previous work on DDHR has led to two major outcomes: (1) the system performance achieved corresponds to the fundamental mutual information between the error-affected data and its class membership, regardless of the severity of errors [2]; and (2) the error-aware model can be trained entirely within a low-power device via a simple machine-learning algorithm [3].

This paper extends the DDHR concept, demonstrating its application for overcoming non-idealities in the analog frontend due to data conversion and analog feature extraction. A seizure detector is implemented and characterized by replaying analog EEG data acquired from epileptic patients.

II. System Overview

Fig. 1 shows the algorithmic block diagram of the EEG-based seizure detector [4]. Samples from a channel of patient EEG at 64Hz are fed to a feature extractor, which computes the spectral energy distribution. This is done via a bank of 8 band-pass FIR filters, each followed by an absolute-value accumulator, representing the energy over a 2 sec. epoch. The 8 features are concatenated with the energies computed from two previous epochs, giving a

total of 24 features per EEG channel. In the system demonstration, four channels are supported, giving a 96-dimensional feature vector in total.

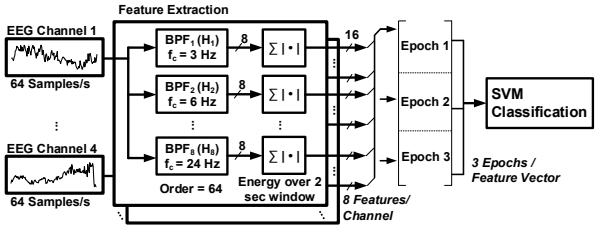


Fig. 1. Algorithmic block diagram of seizure detector.

Fig. 2 shows the system block diagram. Specialized integrating ADCs are used for both data conversion and multiplication of the analog EEG by programmable digital multipliers. FIR filtering can be formulated as multiplication between time samples of the signal and a matrix, whose elements are shifted versions of the filter coefficients (thereby implementing convolution for each output sample). Thus, analog multiplication in this way eliminates the need for explicit multiplication later, leaving only digital additions in order to implement the filtering and absolute-value accumulation needed for feature extraction. Finally, a digital support-vector machine (SVM), with programmable kernel function, is used for

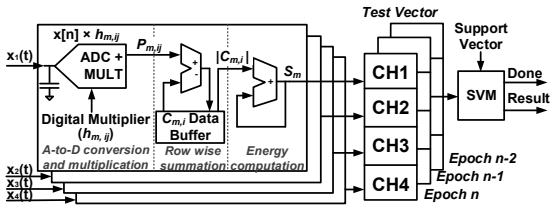
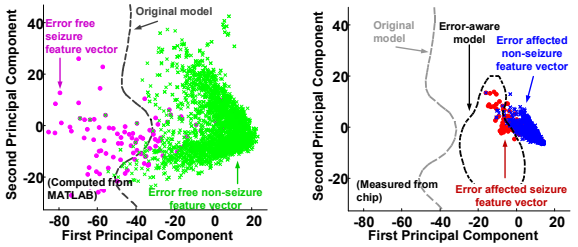


Fig. 2. System block diagram.

classification of the computed feature vectors.

Formulating FIR filtering as matrix multiplication implies that the analog samples *x[n]* must be converted and multiplied by different filter coefficients, $h_{m,ij}$ according to the convolution sum for each filter output sample (m denotes the m-th filter and i, j denotes the i-th row and j-th column of the matrix). Taking analog EEG samples in real time, this requires that each multiplication output from the ADC, $P_{m,ij}$, be added to the running value of the corresponding convolution sum, $C_{m,i}$. Thus, each $C_{m,i}$ must be buffered and then added to the $P_{m,ij}$ values computed by the ADC. Finally, once all filter output samples from a 2 sec. epoch are computed, absolute-value accumulation is performed over the $C_{m,j}$ output samples from each filter by a second 22-b digital adder, giving the spectral-energy features, S_m, from each epoch of each EEG channel. These features are then fed to a 3-stage digital shift register to form the final feature vector by concatenating the features from 3 epochs.

Data conversion and analog multiplication via the simple integrating ADCs suffers from considerable error due to circuit non-idealities. However, the demonstrated system tolerates these errors by appropriately training an error-aware model for SVM classification. Using data from the demonstrated system as an illustration, Fig. 3a shows the distribution of error-free feature vectors (derived from ideal data conversion and digital computation), along with the original SVM classification model. Fig. 3b shows the distribution of feature vectors computed using the prototype IC, along with both the original and error-aware SVM classification models. Previous work has suggested methods for constructing such a model within the sensing device itself [3]; however, for this demonstration, the error-aware model is constructed by acquiring the error-affected feature vectors and performing training offline.



(a) Error-free data with original model. *(b) Error-affected data with error-aware model.*

Fig. 3. Distribution of feature vectors and classification models (to aid visualization, feature vectors are projected to two dimensions via principal component analysis).

III. Circuit-level Implementation

A. ADC and Mixed-signal Feature Extractor

As shown in Fig. 2, the system performs feature extraction over four analog EEG channels. To do this, the circuit shown in Fig. 4 is used. Each EEG signal is sampled using a passive sample-and-hold. Then, data conversion is performed by first resetting an integration capacitor C_{INT} to a nominally mid-rail voltage (V_{MID}), and then driving it with one of the MOSFET current sources. This generates a constant ramp, such that using a comparator, the number of clock cycles taken to reach the analog sample is determined, giving the digital code.

Changing the value of C_{INT} scales the time taken (slope of the ramp), effectively implementing multiplication within the data conversion process. Thus, implementing C_{INT} as a switchable 5-b binary-weighted capacitor bank (with unit capacitance of 67fF) enables application of a digital multiplier. Employing both a charging and discharging current source (PMOS and NMOS) allows for signed multipliers and input samples (with respect to V_{MID}). The current and capacitance values are set for a nominal ADC resolution of 8 bits. The multiplied output from the ADC is then added to the running value of the corresponding convolution sum $C_{m,i}$ implicitly, by loading the buffered value of $C_{m,i}$ into the counter as the starting point of the integration process. This necessitates a 16-b counter.

The motivation to use an integrating ADC topology to

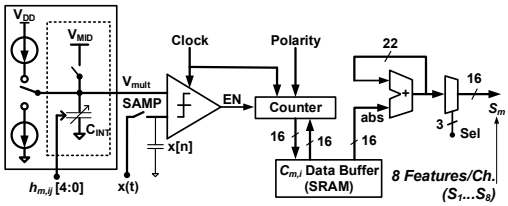


Fig. 4. ADC and mixed-signal feature extractor.

perform analog multiplication is driven by dynamic range considerations. Generally, multiplication substantially increases the dynamic range required, which has a severe impact on voltage-based analog circuits. On the other hand, the integrating ADC uses C_{INT} and the current source to effectively implement time-based multiplication, wherein the primary source of energy consumption is CV^2 loss in the digital counter. Thus, the impact of multiplication can be substantially less than in typical voltage-based circuits.

The analog computation performed is affected by numerous non-idealities, leading to computational errors. The major non-idealities include the following: non-linearity of the MOSFET current sources; bottom-plate parasitics and mismatch of the switchable unit capacitors implementing C_{INT}; comparator delay and offset dependence on the voltage of the analog input sample; variable charge-injection error in the input sample and hold; etc. Prototype measurements in Section IV characterize the errors in the feature vectors. Rather than correcting or calibrating the non-idealities on the circuit level, these errors are overcome on an algorithmic level via the error-aware model loaded in the SVM engine.

B. Support Vector Machine (SVM) Engine

The SVM engine performs classification over the computed feature vector by using a model that is preloaded from training. Classification involves computing the following function:

$$\sum_{i=1}^{N} \alpha_i y_i K(\vec{x},\ \overrightarrow{sv_i}) + b, \qquad (1)$$

where $\vec{x}$ is the 96-dimensional feature vector, $\overrightarrow{sv_i}$ is one of N support vectors derived from training (representing the model), and K is a kernel function employed to enhance the flexibility of the classification model (and may introduce additional training parameters). Following this, thresholding is applied for binary classification.

Fig. 5 shows a block diagram of the SVM engine. The primary operation involving the elements of $\vec{x}$ and $\overrightarrow{sv_i}$ is multiplication-accumulation (MAC), while K involves various non-linear operations. In our implementation, the kernel function is programmable; but most notably, for greatest flexibility, a radial-basis function (RBF) kernel is used, which involves computation of the exponential function. Thus, the SVM engine's microarchitecture, which is based on the design in [5], involves a specialized MAC unit and a CORDIC unit.

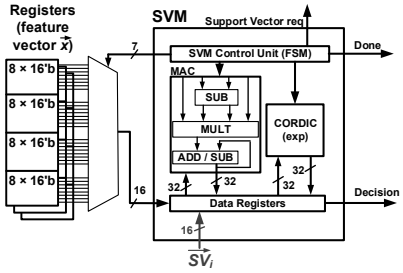
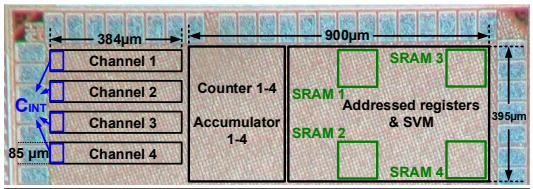


Fig. 5. Block diagram of SVM engine.

IV. Prototype Measurements and System Demo

The seizure detection system is prototyped in a 32nm CMOS SOI process from IBM. The die photo and operation summary is provided in Fig. 6. Further testing details are presented below.

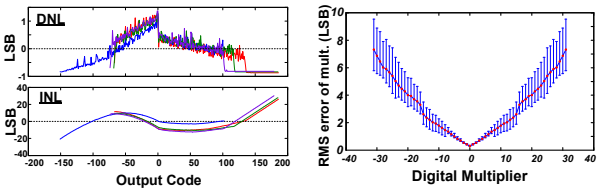


Technology	32nm CMOS SOI	Sample Rate	Up to 5kHz (typ. 64Hz)
Supply Voltage	Analog: 0.9 V Digital: 0.9 V SRAM: 0.9 V	ADC ENOB	2.8-3.6 (f_{in} = 100 Hz)/ 2.1-2.6 (f_{in} = 1k Hz)
		ADC INL/DNL	20.8-30.1/1.2-1.4 (LSB)
		Energy / Feature Vector	12.04 uJ / Feature Vector
ADC Resolution	8 bits		
Input Range	200 mV - 700 mV	Energy / class.	1.35 uJ / class.

Fig. 6. Die photo of prototype system and operation summary.

A. ADC and Multiplication Errors

The ADC performs both data conversion and multiplication. The errors are characterized first using standard ADC metrics and then using more general metrics for computation. Fig. 7a shows the INL/DNL measured for each of the four channels (digital multiplier of 31 is used for characterization); the max. INL is 30 LSB at the 8-b level. Fig. 7b shows the root-mean-squared (RMS) error of the multiplication output for each multiplier value, when the ADC is driven by an input having uniform distribution over the nominal analog range. As shown, both the ADC nonlinearity and multiplication errors are severe. Nonetheless, system measurements presented next show high seizure-detection performance is maintained.



(a) ADC INL/DNL (shown for 4 channels)

(b) Multiplication RMS error (with error bars for 4 channels).

Fig. 7. ADC and analog multiplication non-idealities.

B. Seizure-detection System

The seizure-detection system is tested by replaying patient data from the CHB-MIT seizure dataset [6] using a 14-b arbitrary waveform generator. The baseline model for SVM classification is constructed by computing feature vectors offline based on the algorithm shown in Fig. 1. The error-aware model for SVM classification is constructed by acquiring feature vectors from the system itself during a training phase. In both cases, the SVM model is trained offline and loaded to the chip.

Non-idealities in data-conversion and analog-multiplication cause severe errors in the feature vectors computed by the chip. Specifically, the RMS error of the features normalized to the ideal values of the features is measured to be 1.16, indicating errors larger than the feature values themselves. Proceeding with system characterization, Table I shows the ideal performance achieved (i.e., by implementing the algorithm in Fig. 1 in MATLAB) alongside the performance achieved using the chip, both with a baseline model and with the error-aware model. The performance metrics are sensitivity (number of seizures correctly declared), latency (delay of seizure declaration with respect to expert-identified electrographic onset) and specificity (number of seizures incorrectly declared). As shown, feature errors result in poor performance using the baseline model (high false alarms); however, despite the feature errors, the error-aware model restores the system performance to nearly the ideal level.

Table I
Seizure-detection system performance

	Ideal Performance	Chip w/ baseline model	Chip w/ error-aware model
Sensitivity	5/5	5/5	5/5
Latency	2.0 sec.	3.6 sec.	3.4 sec.
Specificity*	8	443	4

V. Conclusions

This paper explores the possibility of greatly relaxing the precision requirements of analog frontend circuits by taking advantage of algorithmic capabilities to model how information is represented in the resulting outputs. Namely, by using a machine-learning algorithm to train a classification model based on the resulting outputs, errors due to circuit non-idealities are compensated by the classifier. Using simple analog circuits for data conversion and feature extraction, the prototype seizure-detection system shows that performance is effectively restored to ideal-system levels, even in the presence of severe errors.

Acknowledgements

Funding provided by NSF (grant CCF-1253670), AFOSR (grant FA9550-14-1-0293), and Systems on Nanoscale Information fabriCs (SONIC), one of six SRC STARnet Centers, sponsored by MARCO & DARPA.

References

[1] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. P. Chandrakasan, "A micropower 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.

[2] Z. Wang, K. H. Lee, and N. Verma, "Overcoming computational errors in sensing platforms through embedded machine-learning kernels," *IEEE Trans. VLSI Syst.*, DOI: 10.1109/TVLSI.2014.2342153.

[3] Z. Wang, R. Schapire, and N. Verma, "Error adaptive classifier boosting (EACB): leveraging data-driven training towards hardware resilience for signal inference," *IEEE Trans. Circuit and Syst. I*, vol. 62, no. 4, pp. 1136-1145, Apr. 2015.

[4] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 975–982.

[5] K. H, Lee and N. Verma, "A low-power processor with configurable machine-learning accelerators for high-order and adaptive analysis of medical sensor signals," *IEEE J. Solid-State Circuits*, vol. 48, no. 7, pp. 1625–1637, Jul. 2013.

[6] [Online]. Available: http://www.physionet.org, Physionet