MIT Open Access Articles

Data-Driven Approaches for Computation in Intelligent Biomedical Devices: A Case Study of EEG Monitoring for Chronic Seizure Detection

Massachusetts Institute of Technology

Review

Data-Driven Approaches for Computation in Intelligent Biomedical Devices: A Case Study of EEG Monitoring for Chronic Seizure Detection

Naveen Verma [1,*], **Kyong Ho Lee** [1,*] **and Ali Shoeb** [2]

[1] Princeton University, Engineering Quadrangle B-226, Olden Street, Princeton, NJ 08544, USA

[2] Massachusetts General Hospital & MIT , 77 Massachusetts Avenue, Cambridge, MA 02139, USA;
E-Mail: ashoeb@mit.edu

[*] Authors to whom correspondence should be addressed; E-Mails: nverma@princeton.edu (N.V.);
kyonglee@princeton.edu (K.H.L.); Tel.: +1-609-258-1424.

Abstract: Intelligent biomedical devices implies systems that are able to detect specific physiological processes in patients so that particular responses can be generated. This closed-loop capability can have enormous clinical value when we consider the unprecedented modalities that are beginning to emerge for sensing and stimulating patient physiology. Both delivering therapy (e.g., deep-brain stimulation, vagus nerve stimulation, *etc.*) and treating impairments (e.g., neural prosthesis) requires computational devices that can make clinically relevant inferences, especially using minimally-intrusive patient signals. The key to such devices is algorithms that are based on data-driven signal modeling as well as hardware structures that are specialized to these. This paper discusses the primary application-domain challenges that must be overcome and analyzes the most promising methods for this that are emerging. We then look at how these methods are being incorporated in ultra-low-energy computational platforms and systems. The case study for this is a seizure-detection SoC that includes instrumentation and computation blocks in support of a system that exploits patient-specific modeling to achieve accurate performance for chronic detection. The SoC samples each EEG channel at a rate of 600 Hz and performs processing to derive signal features on every two second epoch, consuming 9 μJ/epoch/channel. Signal feature extraction reduces the data rate by a factor of over $40\times$, permitting wireless communication from the patient's head while reducing the total power on the head by $14\times$.

1. Introduction

Sensors and stimulators are emerging that are allowing biomedical devices to advance to new frontiers. Much of this has been initiated by innovations in biologically compliant materials and technologies, but its relevance has become apparent thanks to clinical work that has led to the conception of new *systems* that are beginning to take advantage of the unprecedented modalities that these technologies offer for delivering therapy and for sensing physiology. Deep-brain stimulation (DBS), for instance, involves the implantation of a stimulation electrode in targeted locations of brain as well as the use of a simple control device (similar to a pacemaker) that actuates stimulation. Devices employing DBS have shown astounding efficacy in some patient towards the treatment of a range of neurological conditions from Parkinson's disease to dystonia [1]. Vagus nerve stimulation (VNS), is an alternate approach that has been used to treat epilepsy patients; it avoids the need for brain surgery by using an implanted stimulator in the neck that is actuated either using a similar control device or through manual control by the patient [2]. Similarly, sensors are emerging that are capable of recording extremely rich physiological signals over long time periods as a part of chronic disease management (*i.e.*, as a long-term means of managing pathophysiologic patient states). Flexible materials, such as silk [3] and silicone polydimethylsiloxane (PDMS) [4], that are also histologically promising, provide a substrate both for electrodes and active electronics; this, for instance, forms the basis for high-resolution, multi-channel recording systems for bio-potentials that manifest on curvilinear biological surfaces (such as the cortex of the brain or the surface of cardiac tissue [3,5]).

In order to take advantage of such technologies in *intelligent, closed-loop* systems, however, we require the ability to detect specific physiological processes with a high level of accuracy. Identifying specific, clinically-relevant indicators in bio-potential signals that are available in implantable and wearable systems currently poses a dominating limitation. As we describe in this paper, the signals themselves are physiologically complex, and the correlations are both diverse and subject to noise. An important concern for chronic wearable and implantable systems is also that the computations must be achieved with very low power consumption; the total power budget for wearable devices, for instance, is 1–10 mW, and that for implantable devices is 10–100 μW. These power levels are derived from the energy delivery options that are viable from application to application (ranging from primary and chargeable storage sources (e.g., thin-film batteries) to wireless power transfer), as described in [6]. Ultimately, the power and/or energy constraints make it essential to carefully consider the algorithmic approach and the computations required in order to identify opportunities for reducing the complexity or improving the efficiency.

Understanding the algorithmic approaches is also important when we consider how limited current clinical decision-support systems are in terms of accuracy. Current bedside hospital monitors, for instance, employ models that are insufficient for detecting acute events with high specificity [7]. Alarm

fatigue is thus one of the most prominent problems in intensive care units [8]. Nonetheless, automatic patient monitoring remains an important aspect for managing the scale of healthcare delivery, even in the hospital setting. Beyond this, chronic out-patient monitoring (following hospital discharge) is also emerging as an important mode of healthcare [9]. In the face of limited clinical resources, it further stresses the importance for accuracy in order to manage the resulting increase in scale (*i.e.*, false-alarm rates must be low when monitoring large patient populations that can only be accessed through out-of-hospital response). It also places greater emphasis on the need for low-energy platforms due to the mobility required in the monitoring devices.

A powerful development in the domain of clinical decision support has been the emergence of *data-driven* modeling techniques. These refer to methods of modeling physiological signals based on observing and analyzing data. This is in contrast to attempts at modeling the underlying processes, which are exceedingly complex, and therefore tend to yield low-order models. In the remainder of this paper, we describe why data-driven techniques are a promising means to achieve the accuracy required in practical chronic biomedical devices. We then describe the computational approaches that can be used to take advantage of these. Finally, we consider a practical device that attempts to exploit data-driven modeling. An SoC for chronic seizure detection is described, illustrating the design considerations and methodologies involved. The SoC employs an architecture that performs local acquisition of patient electroencephalographs (EEGs), analog-to-digital conversion, and signal feature extraction. Local feature extraction results in a compressed signal representation for data-driven classification, making wireless transmission from the scalp viable.

2. The Need for Data-Driven Techniques

Currently, clinical decision support systems rely on algorithms that employ very simple models of the physiological signals in order to detect the correlations of interest. Often, for instance, the models are based on pre-determined thresholds of specific signal parameters [10]. These parameters can typically be extracted through signal processing, and thus low-energy DSP has been the predominant computational approach in current biomedical devices [11]. In this section, however, we highlight three key challenges in biomedical detection applications that motivate a change in focus from *signal processing*, which involves computations to transform signals, to *inference*, which involves computations to model and detect specific signal correlations. This leads to the emphasis on data-driven modeling techniques.

2.1. Detection Challenges in Biomedical Applications

Clinically relevant processes that must be detected in biomedical applications are generally physiologically complex. The ability to model these, and their manifestations in the signals that are available through chronic sensing, poses a primary challenge. Thus, the concern, as we try to achieve higher accuracy, is not just the ability to extract specific signal parameters with high signal-to-noise ratio, but rather the ability to discriminate the meaning of those parameters. Although the development of physiology-based models has played an important role in understanding the underlying processes, their viability in practical detection systems is limited. First, these have typically resulted in low-order models [12]; this leads to simple threshold rules that yield limited accuracy. Second, even after the

physiological processes have been modeled, the specific manifestations in the signals must also be modeled. The primarily signals of interest in low-power chronic biomedical devices are collaterally representative of the target physiological processes. As an example, neural electrophysiology appears in scalp electroencephalographs (EEG) after volume conduction through brain tissue and then signal transduction across the tissue membranes and skull that surround the brain. The physiological sources thus manifest in the signals through complex coupling mechanisms. Physiology-based models thus face severe complexities that limit the accuracy that they have been able to provide.

A second challenge is that that signals available through low-power chronic sensing typically represent the superposition of numerous physiological sources. Thus, the models must be able to reject these with very high specificity, usually higher than (linear) signal processing affords. In the case of EEG-based seizure detection, for instance, electrical activity from muscles (*i.e.*, electromyographic (EMG) activity) is a major concern. As an example, Figure 1 is from the case-study in [13], where amplitude-integrated EEG (AEEG) is used to monitor a neonatal patient. AEEG is a method of presenting the EEG over long time periods (using filtering, peak rectification, and semi-logarithmic compression). Using a threshold-margin-based detection approach, seizure activity was not detected in the initial period despite observing abnormal symptoms in the patient. The EEG trace, exhibited the artifact shown below, which was suspected to be EMG activity. In the following periods, the electrode position was changed, and then a muscle relaxant was administered, suppressing the artifact. At this point, the AEEG margins indicated electrographic seizure activity, consistent with the clinical symptoms. In addition to physiological interference sources, mechanical motion artifacts and coupling of electrical environmental noise are prominent concerns [14]. Although these degrade the performance of detection algorithms that are based on simplistic rules, data-driven algorithms that explicitly model artifact-corrupted signals can retain their accuracy and thus mitigate the need for further instrumentation and/or signal processing to suppress the artifacts [15].

Figure 1. Seizure detection based on amplitude EEG margins. The electromyographic (EMG) activity in the EEG trace (initially observed) blocked seizure detection until a muscle relaxant was administered [13].
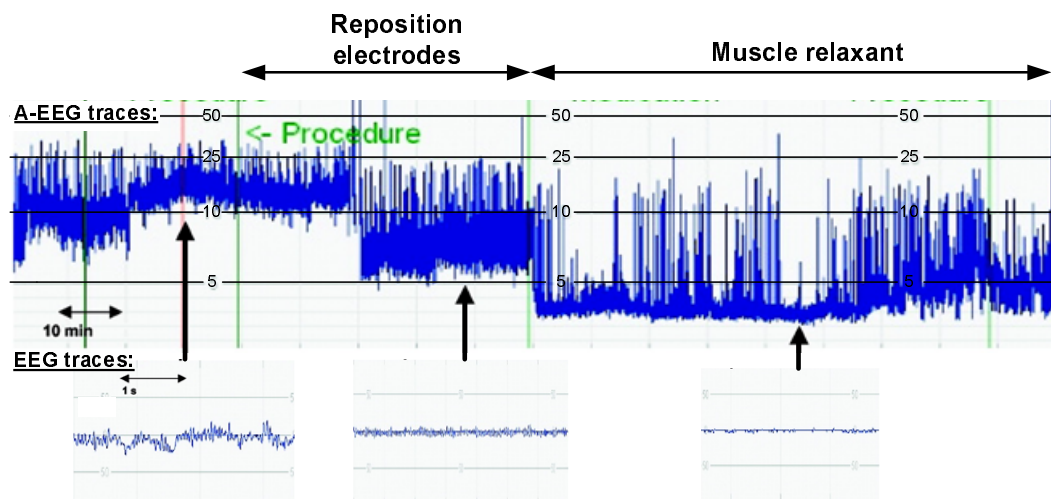
The third challenge is that efficient methods for developing and adapting the detection models are required. This can play a critical role in improving accuracy by allowing patient-specific factors to be more readily incorporated. The importance of patient-specific factors is seen across applications. In [10], even very simple adaption of model thresholds based on sparse sampling from a target patient group shows to yield substantial performance improvement for cardiac monitoring applications. Adaption has shown to be beneficial even when very rich models are possible. For instance, a great deal of clinical work has gone into characterizing the typical behavior exhibited in a patient's EEG during a seizure. Most importantly, a characteristic frequency range of spectral activity has been identified as a strong indicator [16]. Figure 2, however, shows two actual EEG bursts from a patient. Although, no clinical seizure corresponds with the burst in Figure 2a, the electrographic activity has the rhythmic behavior typically associated with a seizure. On the other hand, Figure 2b corresponds to a clinical seizure, and, although its rhythmic behavior is consistent with that typically exhibited during seizures, it is more specific, both in terms of the spectral content and in terms of the spatial EEG channels that are excited. Thus, modeling the background and seizure characteristics, and doing this on a patient-by-patient basis improves the accuracy of seizure detection [17]. The ability to efficiently and systematically develop new models is therefore a key aspect for practical detection systems.

Figure 2. EEG from an epileptic patient showing (**a**) typical background burst resembling a seizure and (**b**) burst corresponding to a clinical seizure, having specific spectral and spatial characteristics.

(a)   (b)
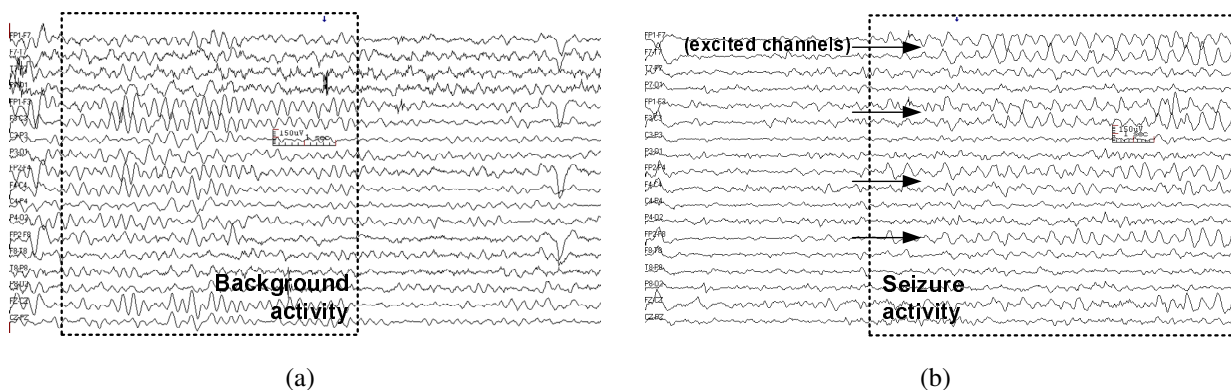
2.2. *Exploiting the Availability of Medical Data*

With respect to the challenges faced in biomedical detection applications, two important developments have recently emerged. The first is the availability of patient data in the healthcare domain, and the second the advancement of machine-learning techniques for data-analysis. Hospitals today widely employ data-bases where patient signals recorded from bedside monitors (e.g., EEG, ECG, *etc.*) are logged along with observational and/or diagnostic annotations [12,18]. As a result, a large amount of data is available, corresponding to patients with a broad range of clinical conditions. Practices for annotating the datasets are also improving, leading to high-quality, clinically-relevant labels are being directly applied to the signals [18]. A technological advancement of importance is low-power recording technologies that can provide long-term access to the same physiologically-indicative signals recorded

by hospital bedside monitors [14]. Through these, biomedical devices can directly take advantage of the annotated databases without the need for supplementary models to correlate with alternate signals.

Machine-learning techniques provide powerful methods for exploiting the large-scale availability of patient data. These have particularly advanced thanks to applied work related to physiological data analysis [19]. More specifically for chronic biomedical devices, the *data-modeling* and *classification* methods (*i.e.*, methods for applying the models) have seen substantial development [20]. The efficiency with which machine-learning techniques allow the development of quantitative models from large datasets, means that patient-specific models can be readily developed as patient-specific data becomes available, and thus a high degree of specificity can be achieved. In [17], for instance, a patient-specific seizure detector achieves a sensitivity of 93%, a specificity of 0.3 ± 0.7 false alarms per hour, and a latency of 6.77 ± 3.0 s, compared to 66%, 2.0 ± 5.3 false alarms per hour, and 30.1 ± 15 s for a patient-generic detector.

The generality of the modeling and classification methods means that they can be applied across a wide range of biomedical applications. Figure 3 shows the generic structure of machine-learning-based detection algorithms. The seizure detection application is based on [21], where a sensitivity of 96%, a specificity of 2 false detections per day, and a latency of 3.0 s is achieved across 24 patients, and the arrhythmia beat classification application is based on [22], where a sensitivity above 75% and a false positive rate below 1.5% is achieved. The two primary aspects of the algorithms are (1) biomarker extraction and (2) biomarker interpretation. Biomarkers correspond to defined parameters in the signals that are determined to have some correlation with the target physiological patient state. The precise correlation, however, is determined by constructing a model of the biomarkers by employing a training dataset. Real-time interpretation of the biomarkers is then achieved by applying that model using a classifier. The classifier thus requires a training phase; this, however, does not face real-time detection-throughput constraints and can be performed offline, either once or periodically to track slowly evolving physiology [23].

By treating the biomarkers as a feature vector, a machine-learning classifier can be used. Various options for machine-learning-based classification are available, and these offer the possibility to efficiently scale the models to handle a large number of features with diverse correlations. The following section focuses on classification approaches that can be applied in biomedical devices. Biomarkers, however, are closely coupled to the clinical factors, and therefore their precise choice varies from application to application. For instance, in Figure 3, two applications are considered. For seizure detection, information is contained in the spectral content of the patient's electroencephalograph (EEG), while for cardiac arrhythmia detection information is contained in the waveform morphology of the patient's electrocardiograph (ECG); the biomarkers are thus chose accordingly. Even within an application such as arrhythmia detection, which is associated with a range of cardiovascular conditions, various biomarkers have been used based on the physiological factors they represent [22]. Finally, the optimal biomarkers can also be patient dependant; for instance, seizures in some patients have exhibited notable correlation with cardiac beat-rate in addition to EEG activity. Thus, detection accuracy improves when beat-rate features (extracted from the ECG) are included [21]. As a result, programmable platforms that can support a wide range of biomarker extraction processing can play an important role.

Figure 3. The typical structure of data-driven biomedical detection algorithms consists of biomarker (feature) extraction followed by biomarker interpretation, which is based on a generated model.
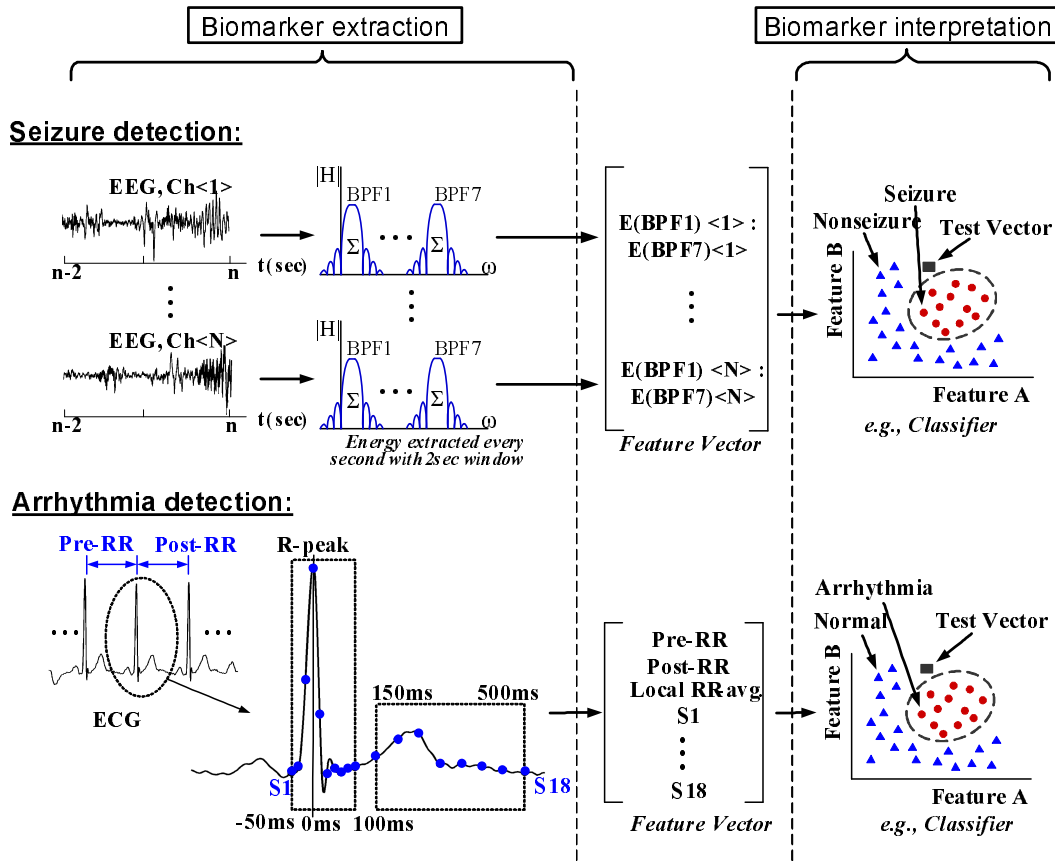

3. Data-Driven Classification Frameworks

Biomedical signals are non-stationary; as mentioned, they reflect the state of numerous dynamic physiological processes. Thus, an approach based on simply detecting changes in the signal characteristics is not sufficient. Instead, biomedical devices must overlook *physiologic* signal changes, and recognize specific *pathophysiologic* signal changes that are expressed due to a dynamic disease state that is of interest. As an example, the EEG trace (over 18 channels) in Figure 4 exhibits a burst at the 2994 s mark corresponding to the start of an epileptic seizure. This is characterized by rhythmic activity (most prominent on the FP2-F4 and T8-P8 channels). Earlier rhythmic activity, for instance that between the 2989–2992 marks, is distinct from typical background activity. Nonetheless, it in fact represents a normal characteristic of sleep EEG (known as a *spindle*), and should be differentiated by a detection device.

Biomedical signals can also exhibit a wide range of characteristics within both the physiologic and pathophysiologic states. For instance, in the physiologic state, EEG exhibits different characteristics depending on whether an individual is awake or asleep. Similarly, in the pathophysiologic state, the EEG observed at the onset, middle, and end of a seizure may be very different [24]. Depending on the application, however, a system may be required to differentiate between the physiologic and pathophysiologic states, or to discriminate among various sub-states within each of these two states.

Differentiating between the physiologic and pathophysiologic states of a signal amounts to solving a binary classification problem. A binary classifier aims to label an observation as a member of one of two classes in a manner that optimizes a chosen performance objective; for example, this could be minimization of the misclassification probability. In the seizure onset detection problem, the observation is an EEG epoch (*i.e.*, some window of time) and the two classes are seizure (Class C_1) and non-seizure (Class C_2).

Figure 4. Biomedical signals represent various physiologic (e.g., 2989–2992 s marks) and pathophysiologic states (e.g., 2994 s mark). Thus, simple detection of signal changes (*i.e.*, anomaly detection) is insufficient.
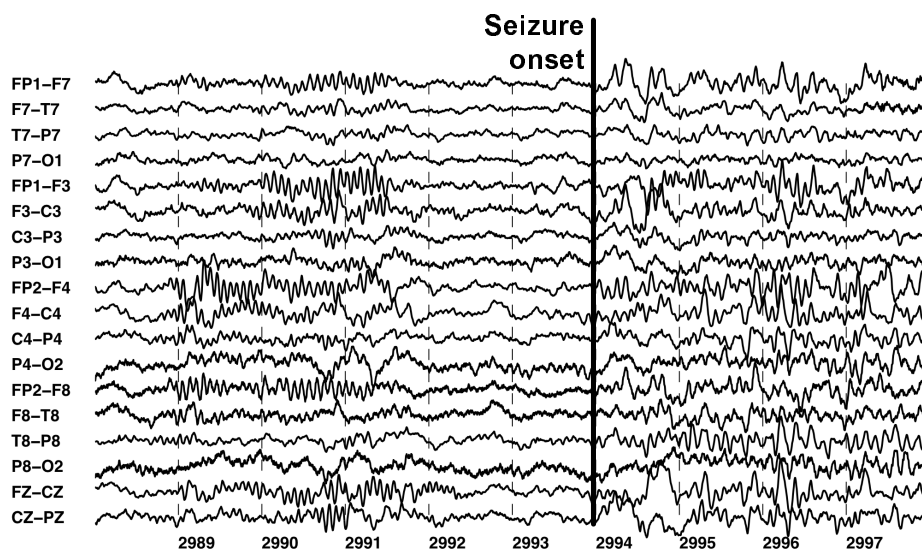
The non-seizure class, in actual fact, includes EEG gathered during awake, sleep, and ambulatory periods. Discriminating among the sub-states of the physiologic and pathophysiologic states amounts to solving a multi-class classification problem. However, since the most powerful and robust classification algorithms available are formulated within the setting of binary classification, multi-class classification problems are often solved using a collection of binary classifiers [25]. As an example, consider differentiating between K classes using binary classification. One approach, known as *one vs. all*, involves combining the outputs of K classifiers, each of which discriminates between a class (e.g., C_1) and the union of the remaining $K - 1$ classes (e.g., $C_2 \cup C_3 \cup \ldots C_K$).

For classification, rather than directly using the observed signal, the signal is represented by biomarkers that correspond to salient signal properties. This generically reduces the complexity of the observation without attempting to discriminate the data, and it creates a representation that is appropriate for the specific task of model-based classification via the formalizations described below. Namely, the observation is converted into a set of *features*. The choice of features strongly affects the success of a binary classifier. Ideally, the distributions of the feature values should be as distinct as possible for observation belonging to the different classes. Then, by letting a feature be denoted by x_i and a vector

of features by $\vec{x} = [x_1 \ x_2 \ x_3 \ x_k]$, we can represent a binary classifier as a function that maps a feature vector $\vec{x}$ to Class C_1 or C_2. More succinctly

$$F(\vec{x}) \to \{C_1, C_2\} \tag{1}$$

Several techniques exist for determining the mapping function. Below, we discuss the merits of the generative and discriminative approaches. While the generative approach may provide classification outputs that indicate class membership and *degree* of membership, the discriminative approach simply aims to provide an indication of class membership; this is both sufficient for many of the medical devices envisioned, and, as described below, it is more viable for model construction based on the data typically available [26,27].

3.1. Discriminative Approach

Discriminative approaches aim to derive a *decision boundary* or boundaries that divide the feature space into regions containing feature vectors from the same class. A powerful algorithm for determining decision boundaries is the support-vector machine (SVM) algorithm [28]. In its simplest form, the SVM algorithm determines a decision boundary in the form of a hyper-plane [29]. The hyper-plane is chosen to maximize the classification margin, which is the geometric distance between the hyper-plane and the boundary data instances for each class that are observed during training [30]. The chosen boundary cases are known as the support vectors.

As an example, Figure 5a shows the linear decision boundary separating intra-cranial EEG (IEEG) seizure feature vectors (circles) from non-seizure feature vectors (crosses). Here, the x and y-axis features represent the energy in the spectral band 8 ± 8 Hz and 25 ± 11 Hz, respectively. The seizure feature vectors are distributed into a cluster that lies away from the non-seizure data as well as several sparse points that lie in close proximity to the non-seizure data. The seizure vectors closest to the non-seizure vectors are associated with the onset period while those further away are associated with later stages of the seizure (e.g., the vectors numbered 1–5 are extracted from the first 7 s of a seizure). It can be seen that the two classes are not well separated by a linear decision boundary, as numerous non-seizures data points are classified as seizure feature vectors. SVMs, however, allow nonlinear boundaries to be evoked through a kernel transformation, improving the classifier performance (by, for instance, reducing misclassification of the non-seizure feature vectors, as shown in Figure 5b).

The mapping function that establishes whether a feature vector falls within the region of the feature space defined by seizure activity is expressed below for the case of a linear boundary. Here, the vector W (which is normal to the separating hyper-plane) and the bias term b are parameters determined by the SVM learning algorithm.

$$f(\vec{x}) = \begin{cases} C_1, & \text{if } W^T\vec{x} + b \geq 0 \\ C_2, & \text{if } W^T\vec{x} + b < 0 \end{cases} \tag{2}$$

A highly flexible nonlinear boundary can be achieved using a radial-basis function kernel, where the discriminant function takes the form shown in the equation below. The coefficient α_i, the support-vectors

$\vec{x}_i$, and the bias term b are determined by the SVM learning algorithm based on *observed* data. The parameter γ, which controls how tightly the nonlinear boundary circumscribes a class, is user-defined.

$$f(\vec{x}) = \begin{cases} C_1, & \text{if } \sum_{i=1}^{N} \alpha_i e^{-\gamma|\vec{x}-\vec{x}_i|^2} + b \geq 0 \\ C_2, & \text{if } \sum_{i=1}^{N} \alpha_i e^{-\gamma|\vec{x}-\vec{x}_i|^2} + b < 0 \end{cases} \tag{3}$$

The number of support-vectors N is strongly affected by the complexity of the classification task. As the similarity between the physiologic and pathophysiologic data increases, more support-vectors may be needed to define a more complex decision boundary, and the computational cost scales accordingly.

Nonetheless, the SVM algorithm is well suited for biomedical detection tasks because its learning mechanism focuses on determining a decision boundary by identifying boundary data cases from among those that are *observed* for each class. This readily allows the generation of the models. It has also been shown to perform well in high-dimensionality classification problems [31] and in settings where one class has a much smaller number of training samples, as is often the case for the pathophysiologic class [32].

Figure 5. Support-vector machine decision boundaries for intracranial EEG from a seizure detection application using (**a**) a linear classification kernel and (**b**) a radial-basis function classification kernel (seizure data is represented by circles and non-seizure data is represented by crosses).

(a)

(b)
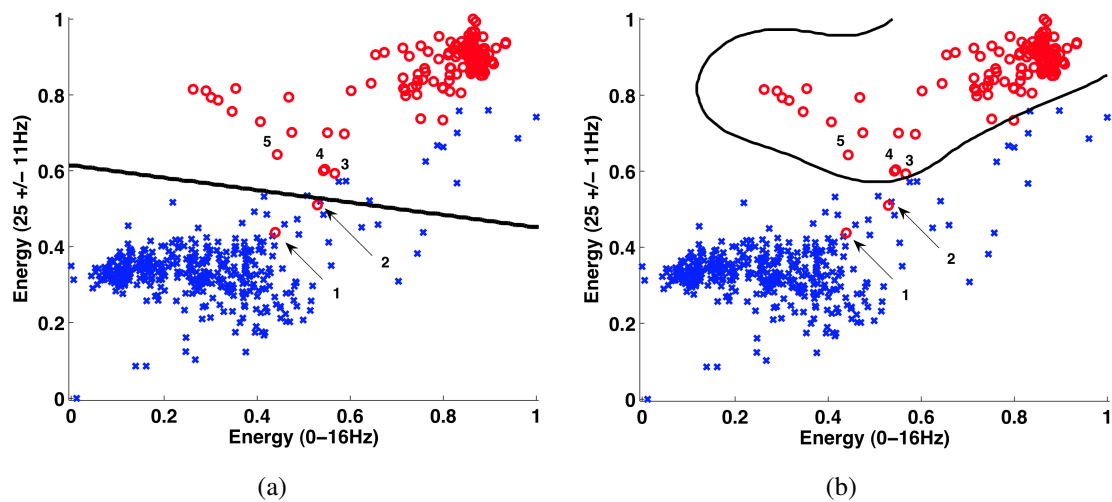
3.2. Generative Approach

In contrast to discriminative approaches (which focus on boundary data cases), generative approaches first model the probability distributions of each class and then label an observation as a member of the class whose model is most likely to have generated an observation [26]. Mathematically, this logic is represented as follows [33]:

$$f(\vec{x}) = \begin{cases} C_1, & \text{if } P(C_1|\vec{x}) \geq P(C_2|\vec{x}) \\ C_2, & \text{if } P(C_1|\vec{x}) < P(C_2|\vec{x}) \end{cases} \tag{4}$$

$P(C_i|\vec{x})$, which is known as the *a posteriori* probability, is the probability that the source of the observation is the class C_i given that we have observed the feature vector $\vec{x}$. The *a posteriori* probability of a class is typically not known. It can, however, be computed using Bayes' Rule. Bayes' Rule formulates the *a posteriori* probability in terms of probability functions and densities that can be more readily estimated from data:

$$P(C_i|\vec{x}) = \frac{f(\vec{x}|C_i)P(C_i)}{f(\vec{x})} \tag{5}$$

The *prior* $P(C_i)$ is the probability that observations originate from the class C_i. The *likelihood function* $f(\vec{x}|C_i)$ describes how likely it is to observe the feature vector $\vec{x}$ assuming that the source is the class C_i. A tractable and flexible choice for $f(\vec{x}|C_i)$ is a mixture of gaussians:

$$f(\vec{x}|C_i) = \sum_{j=1}^{K} \pi_j \mathcal{N}(\vec{x}, \mu_j, \Lambda_j) = \sum_{j=1}^{K} \frac{\pi_j}{(2\pi)^{M/2}|\Lambda_j|^{1/2}} e^{-\frac{1}{2}(x-\mu_j)^T \Lambda_j^{-1}(x-\mu_j)} \tag{6}$$

The parameters of the mixture model (number of gaussian densities K and the weight π_j, mean μ_j, and covariance matrix Λ_j of each gaussian density) can be estimated using training feature vectors derived from the class C_i by using the Expectation Maximization (EM) algorithm [34]. As an example, consider the scalp EEG seizure (circles) and non-seizure (crosses) feature vectors shown in Figure 6a. These feature vectors were originally 126 dimensional but have been projected onto a 2 dimensional space through principal components analysis for illustration. Figure 6b shows the likelihood function for the non-seizure class estimated using the EM algorithm with a mixture of 3 gaussians; notice how the gaussian mixture facilitates the approximation of the multimodal, non-gaussian distribution of the non-seizure class. In general, it is important to remember that reliably estimating the parameters of a gaussian mixture model in the original high-dimensional feature space (which may be necessary in order to model the relevant physiology) is much more challenging and requires an abundance of data; this challenge is often referred to as the *curse of dimensionality*. In the biomedical context, gathering an abundance of data can be difficult, especially for the pathophysiologic class.

Using the prior and likelihood function we can rewrite the *a posteriori* probability and mapping function as follows:

$$f(\vec{x}) = \begin{cases} C_1, & \text{if } f(\vec{x}|C_1)P(C_1) \geq f(\vec{x}|C_2)P(C_2) \\ C_2, & \text{if } f(\vec{x}|C_1)P(C_1) < f(\vec{x}|C_2)P(C_2) \end{cases} \tag{7}$$

If the prior probabilities $P(C_i)$ are not known, then the above decision rule can be reduced to comparing the ratio of the likelihood functions to a threshold λ. The value of λ is determined experimentally (or analytically if mathematically tractable forms of $f(\vec{x}|C_i)$ are chosen [33]) using training data to yield a desired probability of true and false detections:

$$f(\vec{x}) = \begin{cases} C_1, & \text{if } \frac{f(\vec{x}|C_1)}{f(\vec{x}|C_2)} \geq \lambda \\ C_2, & \text{if } \frac{f(\vec{x}|C_1)}{f(\vec{x}|C_2)} < \lambda \end{cases} \qquad (8)$$

From the above decision rule, we can derive a boundary that separates the values of $\vec{x}$ where $F(\vec{x}) = C_1$ from the values of $\vec{x}$ where $F(\vec{x}) = C_2$. As in the case of a discriminative classifier, this decision boundary can be a line, an open curve, or a closed curve depending on the form of the term $f(\vec{x}|C_i)$. Beyond this, however, we can also use the likelihood functions $f(\vec{x}|C_i)$ to assess an observation's degree membership in the class C_i. Unlike a discriminative classifier, which must assign an observation to a class, a generative classifier can thus use the degree of membership of an observation to implement a *reject option*. This can be used to state that an observation is unlikely to belong to either class whenever values of $f(\vec{x}|C_i)$ are very small. Such an option can be useful when classifying signals exhibiting many regimes peripheral to those of concern. This arises in the context of brain-machine interface applications [35].

> **Figure 6.** (**a**) 126 dimensional seizure (red) and non-seizure (blue) feature vectors extracted from Scalp EEG and projected onto a two dimensional space spanned by the first two principal components; (**b**) Three gaussian mixture model of the non-seizure likelihood function ($f(\vec{x}|C_{nonseizure})$) estimated using the EM algorithm.

(a)   (b)
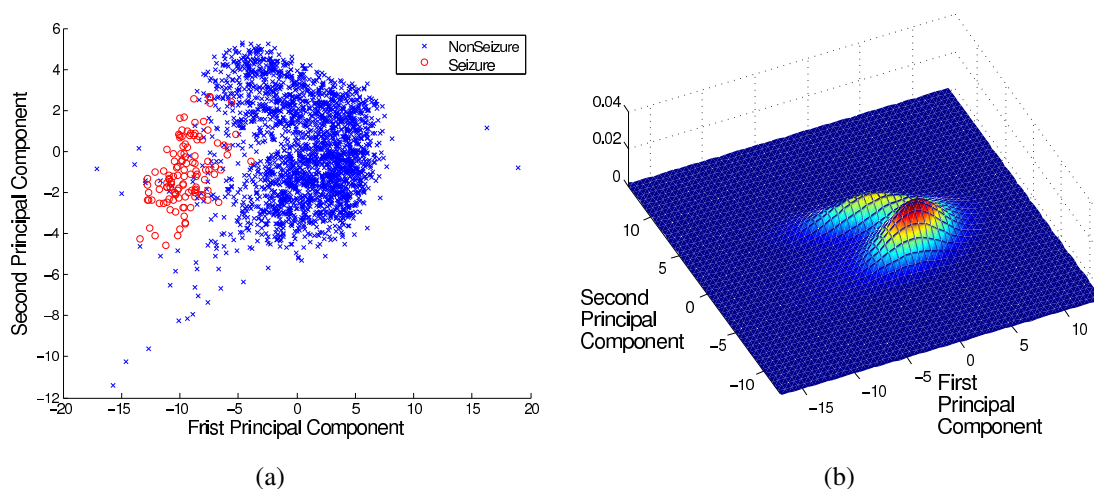
For illustration, Figure 7 shows the decision boundary resulting from applying the above (generative) decision rule with $f(\vec{x}|C_{seizure})$ modeled using a single gaussian, $f(\vec{x}|C_{nonseizure})$ modeled using a mixture of 3 gaussians, and $\lambda = 30$. A smaller λ would result in a wider boundary that includes more seizure feature vectors, and potentially, more non-seizure feature vectors. Also shown in Figure 7 is the boundary resulting from using a discriminative classifier (SVM using radial basis function with $\gamma = 3$). Note how the model assumption of a single guassian for $f(\vec{x}|C_{seizure})$ forces the decision boundary to be an ellipsoid, while the SVM results in a more curved decision boundary that encloses more seizure feature vectors. However, when there is confidence that the proposed form of $f(\vec{x}|C_i)$ describes the data well, a generative model may outperform a discriminative model [26].

In summary, the discriminative approach can be more readily applied to biomedical applications characterized by high-dimensional feature vectors and limited training data; this is thanks to the limited

characterization required of the statistics underlying the data. However, when information about the statistics is available (thanks to sufficient training data), the generative approach can provide models that produce accurate results for classification *as well as* degree of class membership.

Figure 7. Seizure (red) and Non-seizure (blue) feature vectors from Figure 6a separated using decision boundaries estimated using a generative and discriminative approach. The decision boundary derived using the generative approach (black) with $f(\vec{x}|C_{seizure})$ modeled using a single gaussian, $f(\vec{x}|C_{nonseizure})$ modeled using a mixture of 3 gaussians, and $\lambda = 30$. The decision boundary derived using the discriminative approach used a Support Vector Machine and radial basis kernel with parameter $\gamma = 3$.
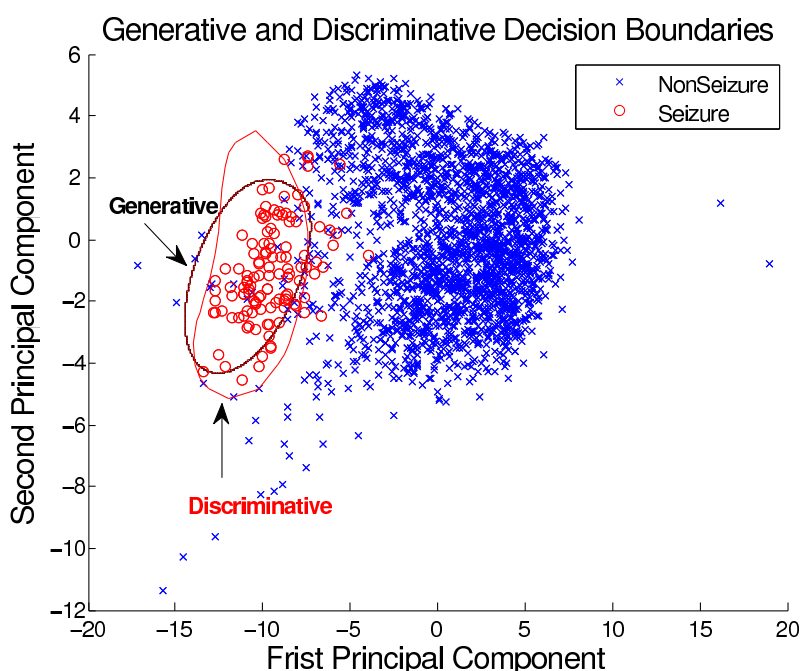
4. Case Study: Chronic Seizure Detection System

Since data-driven techniques introduce distinct benefits for addressing the challenges mentioned in Section 2.1, we are motivated to incorporate these in practical biomedical devices. In this context, system- and application-level considerations for the devices also become important. For instance, recognizing the energy limitations and obtrusiveness of instrumenting patients with many sensing channels, SVM-based approaches have been investigated that minimize the number of electrodes from which continuous recording is required. In the seizure detection system of [36], coarse detection is thus performed continuously using only 60% (on average) EEG channels; the resulting detector misses only 7 out of 143 seizures, has a detection latency of 7.0 s, and achieves a false alarm rate of 0.11 per hour (compared to 4 out of 143, 6.0 s, and 0.07 per hour, respectively, for a full 18-channel system). In addition, low-energy devices that exploit the topology of the analog instrumentation front end to also directly extract typical biomarkers have been reported for EEG-sensing applications [37] and ECG sensing applications [38]. Anticipating the need for low-energy classification, studies have also been pursued investigating simplified classifier implementations that reduce computational complexity

through the use of approximate models (e.g., employing a reduced set of support vectors for SVM classification in seizure detection [39]); although this compromises the classification model, limiting generality and accuracy, it can lead to low-power implementations.

This section describes details of an SoC designed at the circuit and system levels to take advantage of data-driven modeling for chronic seizure onset detection. The SoC incorporates a low-noise instrumentation amplifier for EEG acquisition, an ADC, and a feature extraction processor for SVM-based classification. Each SoC corresponds to one EEG channel, and the number of channels required ranges from 2–18 depending on the patient. Seizure onset detection can serve to generate signals to alert the patient and/or caregivers, control automatic data-logging and charting systems, or actuate therapeutic stimulators in a closed-loop manner (e.g., vagus-nerve stimulators [17]).

The system-level partitioning of the detector is driven by patient safety considerations. Namely, no cables originating from the patient's scalp are permitted since these can pose a strangulation hazard in case the patient convulses during a seizure. Additionally, the battery size and weight on the patient's head and neck must be minimal in order to reduce the risk of injury in case the patient loses balance. To minimize energy consumption on the scalp, the SoC performs EEG acquisition and feature-extraction processing locally, but leaves the classification computation for a device worn away from the patient's head. As described below, local feature extraction plays an important role in reducing the energy of wireless communication from the head, mitigating the need in this application for further data-rate reduction through classification; however, the form factor and/or implantation protocols in future systems can make the integration of classifiers compelling.

4.1. Algorithm and System Design

Clinical studies have determined that seizure onset information is contained in the spectral energy distribution of a patient's EEG [16]. The SoC thus extracts biomarkers corresponding to the spectral energy distribution in seven frequency bins from each EEG channel. The seizure detection algorithm in Figure 3, summarizes the overall processing employed in the system. The seven features from each channel are combined to form a total feature vector of up to 126 dimensions (since as many as 18 channels may be used). For continuous seizure detection, each feature vector is classified in real-time by a pre-trained SVM (utilizing a radial-basis function kernel). The SVM is trained based on patient-specific data, and this is verified to yield substantial improvement in sensitivity and specificity compared to patient-generic detectors (testing is based on 536 hours of patient data across 16 patients from Boston Children's Hospital [17]).

As mentioned, wireless transmission of information from the scalp sensors is essential to avoid the presence of cables. Table 1 considers the total power for a full-scale, 18-channel EEG monitoring system, illustrating the opportunity afforded by local biomarker (feature) extraction as compared to complete wireless EEG transmission. In the case of wireless EEG, the radio transmitter power dominates (assuming a low-power Zigbee radio [40]). Local feature extraction, however, reduces the transmission data-rate by a factor of over forty at the cost of negligible digital computation power. As a result, the total system power on the scalp is reduced by a factor of fourteen. The remaining radio power is no longer dominant, and thus further data-rate reduction is not essential. As a result, the classification computation is left to be performed away from the scalp where the battery size and weight constraints

are less severe. A direct implementation of the SVM classification kernel on an MSP430 processor consumes approximately 4 mW.

Table 1. Power comparison showing the benefit of local feature extraction *vs.* complete wireless EEG [41]. An 18-channel EEG recording system is assumed, and the aggregate power over all channels is reported.

	No Local Processing	**Local Feature Extraction**
Instrumentation amps	**72 μW**	**72 μW**
ADCs (12 b, 600 Hz/ch.)	**3 μW**	**3 μW**
Digital Processors	-	**2.1 μW**
Radio (CC2550)	**1733 μW**	**43 μW**
-Active: *bit-rate* $\times$ 40 nJ/bit	-Active: 43.2 kb/s [1] $\times$ 40 nJ/bit	-Active: 2 kb/2 s $\times$ 40 nJ/bit
-Start-up: 4.8 μW (avg./s)	-Start-up: 4.8 μW	-Start-up: 4.8 μW/2 s
-Idle mode: 0.46 μW	-Idle mode: 0.46 μW	-Idle mode: 0.46 μW
Total	**1808 μW**	**120 μW**

[1] Assumes EEG waveforms are downsampled to 200 Hz (as in typical EEG recording interfaces).

4.2. *Low-Noise EEG Acquisition*

As shown in Table 1, the power consumption of the instrumentation amplifier is an important concern. The amplifier employed thus aims to maximize power efficiency by operating at a low supply voltage ($<$1 V) and by using a voltage-feedback topology, where a minimum number of noise contributing devices are required. For a chronic wearable system, robustness to signal interferences (such as 60 Hz environmental coupling, motion artifacts on the electrodes, *etc.*) are also primary concerns. Since data-driven modeling affords a high detection specificity with respect to the biomarkers, some degree of signal interference is tolerated. Nonetheless, stability in the acquisition system is critical so that the learned correlations remain valid during real-time detection.

Figure 8 summarizes the key instrumentation considerations, which include $1/f$ and wideband device noise sources, 60 Hz coupling, interfering electrographic scalp activity (which may be correlated between recording and reference electrodes), and the electrode offsets. For the electrodes, low-cost Ag/AgCl electrodes, which are the most common, are presumed. In the equivalent circuit model shown, signal transduction is dominated by the capacitor (C_P) and the resistor (R_S). For on-scalp electrodes, the worst case value of these can be as low as 50 nF and as high as 2 MΩ, respectively [42]. The input impedance of the instrumentation amplifier is constrained accordingly (to have much lower capacitance and much higher resistance). In addition, transduction through C_P requires the formation of a charge double layer at the electrode surface, which causes a low-frequency electrode offset voltage (EOV). The EOV can be as high as 100 mV, posing the possibility of easily saturating a low-voltage design.

Figure 9 show the architecture and first stage of the instrumentation amplifier. Chopper stabilization [43] is used to handle low-frequency ($1/f$) noise, and input capacitors (C_{IN}) are used to passively isolate large electrode offsets from the active portion of the low-voltage amplifier. Several other implementations of chopper stabilized EEG amplifiers have recently been reported. In [44], input

capacitors are also used, but chopper modulation is performed before these. This yields very high common-mode rejection ratio by mitigating the effect of capacitor mismatch [14], but it limits the EOV tolerable and degrades the amplifier's input resistance (compromising the ability to handle large electrode impedances in on-scalp sensing applications). Alternatively, the design in [45] uses current-feedback to handle EOV while maintaining very high input impedance through the use of input transconductors. The magnitude of EOV cancelation achievable through active feedback, however, is limited in low supply voltage designs, and the use of current-feedback transconductors degrades the noise efficiency somewhat.

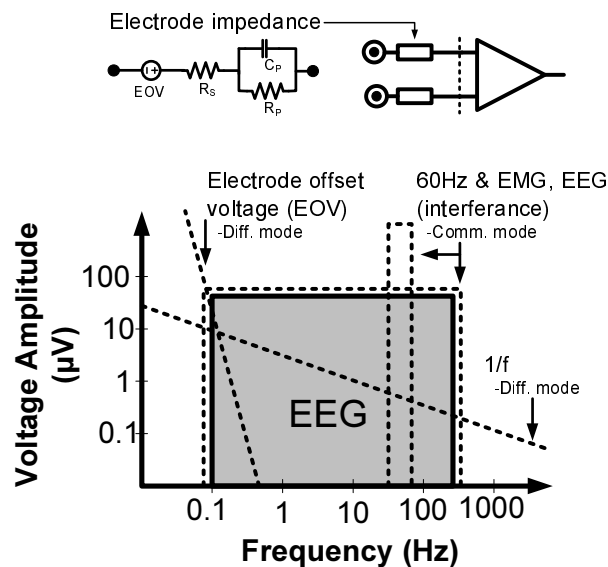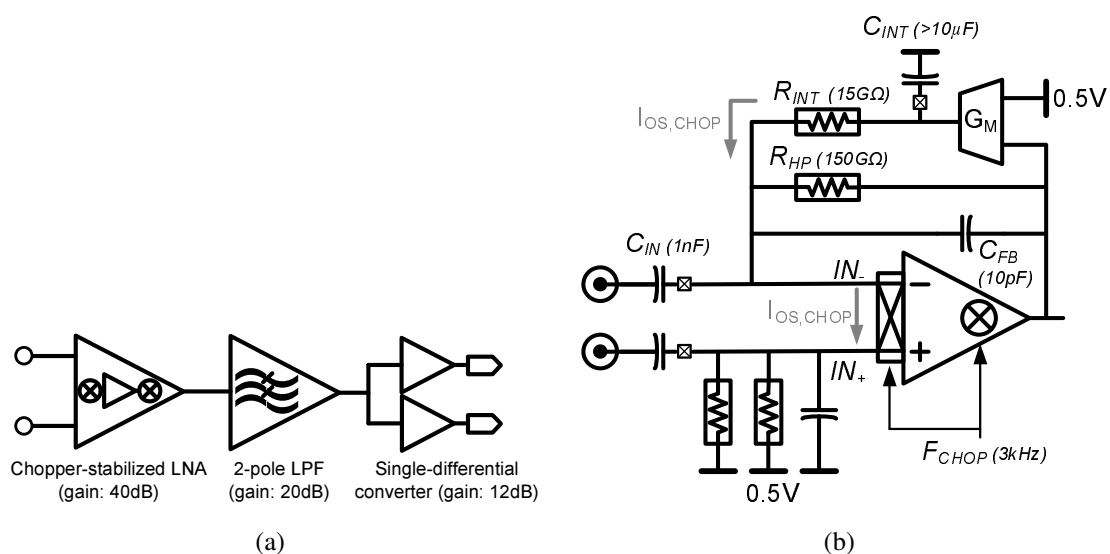**Figure 8.** Instrumentation challenges with EEG sensing.

Figure 9. The instrumentation amplifier showing (**a**) the overall architecture and (**b**) the first stage low-noise amplifier.

In the topology of Figure 9b, the voltage-feedback is retained for noise efficiency, but input chopper modulation is performed after the input capacitors, at the virtual ground node. Although this degrades the resistance between the op-amp nodes, the input resistance seen by the electrodes remains high since

feedback ensures that electrodes do not need to cause substantial voltage swing at the virtual ground node. As shown, however, input offset voltage in the op-amp does cause an offset current ($I_{OS,CHOP}$) between its input nodes. To avoid $I_{OS,CHOP}$ saturating the amplifier, the $G_M - C$ servo loop is used in feedback, providing a cancelation current. Further details of the amplifier, including noise analysis, are provided in [41].

The ADC used to digitize the instrumentation amplifier output is a differential, 12-bit SAR converter (circuit details are provided in [46]). The resolution was selected based on algorithm performance from testing (in a manner similar to the feature-extraction parameter analysis described below). It was determined that detection performance improved up to 10-bits; thus, the 12-bit ADC with 10.6 effective number of bits (ENOB) was employed. The instrumentation amplifier has a bandwidth of 200 Hz (required for generic EEG acquisition). In order to ease the anti-aliasing filtering requirements in the instrumentation front-end, the ADC samples at 600 Hz, since (as shown in Table 1) its power consumption is not a dominating concern.

4.3. Feature-extraction Processing

Although the importance of spectral energy biomarkers in the EEG is well established for seizure onset detection, the precise parameters required for their representation (assuming an SVM classifier) is not well known. In order to determine this, analysis is pursued to evaluate the performance of the algorithm over the parameter space. The parameters considered include the dynamic range of the EEG samples, the precision of the feature-extraction computation, the bin resolution of the spectral energy distribution, and the frequency response of the spectral analysis filters. Due the data-driven nature of the detection algorithm employed, its performance over the parameter space depends on the characteristics of the data. Thus, the analysis is approached by testing the performance using available patient data. As an example, Figure 10 considers the analysis pertaining to the bin resolution for the spectral energy distribution. The critical metrics for seizure onset detection include sensitivity, latency, and specificity (represented by the number of false alarms per hour). As shown, these are determined for bin resolutions ranging from 2–8, in a 0–20 Hz band (*i.e.*, the band of interest for seizure detection [16]). Based on over 1100 h of data across 30 patients, the results show that the performance improves substantially up to a resolution of 6 bins. Thus, in the implementation, a bin resolution of seven filters is chosen.

In addition to the performance metrics, Figure 10 also shows the number of support vectors required. The number of support vectors is representative of the classification model complexity. Although SVM classification is performed away from the patient's head, it is worth noting that there is a trade-off between the quality of the biomarkers and complexity of the classification task. Nonetheless, in the range of bin resolutions required for sufficient performance, the number of support vectors required does not change dramatically.

Figure 11 shows the implementation of the feature extraction processor. In order to determine the precision and frequency response required for the spectral analysis filters, analysis similar to Figure 10 is performed, evaluating the performance over various filter parameters. Although the input EEG is sampled at a rate of 600 Hz (as described above), the band of interest for seizure detection is 0–20 Hz. In order to ease the implementation of the filters, the input is first down-sampled by eight using a decimation filter of order 48. Each of the seven spectral analysis filters then require an order of

46 to achieve the chosen frequency response. The spectral analysis filters have the following frequency response: bandwidth of 3 Hz, transition band of 3 Hz, and stop-band attenuation of -25 dB. The filters are all implemented using a folded-delay line FIR structure.

Figure 10. Analysis (from 1136 h of patient data) to determine the feature-extraction parameters required in order to accurately classify the data (the analysis for spectral bin resolution is shown).
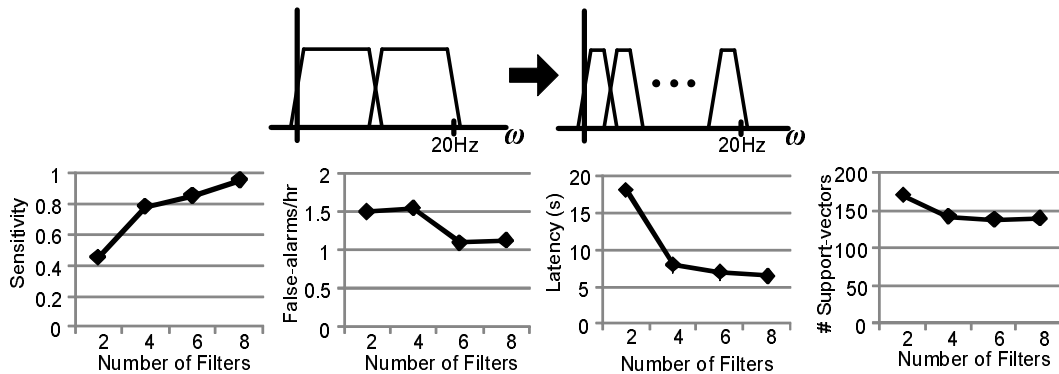

Figure 11. Implementation of the feature-extraction processor (with parameters determined from analysis).
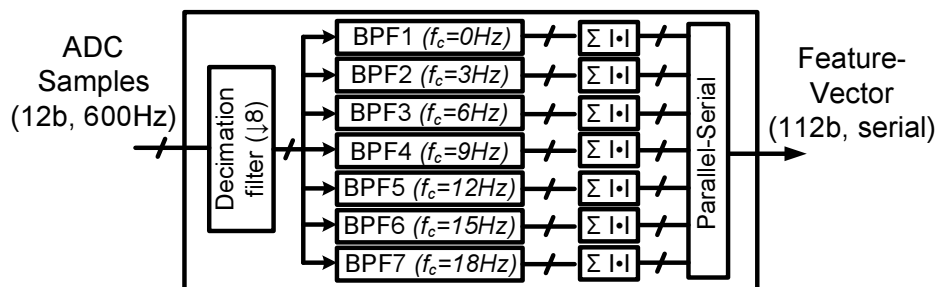

Table 2, shows the final parameters used in the feature-extraction processing, which have been chosen based on analysis of the algorithm performance with patient data.

4.4. SoC Performance

The SoC is implemented in a 5M2P 180nm CMOS process. Figure 12 shows the die photograph along with a performance summary. Although the feature extraction processor occupies a large portion of the die area, its energy is very low, making local feature extraction for communication data-rate reduction compelling from the perspective of the energy and safety concerns related to the patient's head and neck.

Figure 13 shows a demonstration of EEG acquisition, digitization, feature extraction, and classification. A recording electrode at the occipital location (O2) and a reference electrode at the scalp mid-line indicate a relaxed eyes-closed state in the subject through the appearance of an α-wave in the EEG. The α-wave corresponds to a 10 Hz rhythm. In the test shown, the subject periodically closes and opens his eyes to evoke the onset of the α-wave. An SVM has been trained (using 20 s of data) to detect the subject's relaxed eyes-closed state. The waveforms show the EEG acquired using the on-chip

instrumentation amplifier and ADC. The presence of the α-wave is thus annotated. The second waveform indicates the output of the classifier so that the accuracy and latency of the detector can be characterized. Over a five minute run, the α-wave is correctly detected with less than 2.5 s of latency.

Table 2. Parameters selected for feature-extraction processing based on analysis of algorithm performance.

Feature Extraction Parameters	Parameter Values
Decimation filter input rate	600 Hz
Decimation filter output rate	75 Hz
Decimation filter BW	150 Hz
Decimation filter input precision	12-bits
Decimation filter output precision	16-bits
Decimation filter order	48
Spectral analysis band	0–20 Hz
Spectral analysis bins	7
Spectral analysis filter BW	3 Hz
Spectral analysis filter pass band	3 Hz
Spectral analysis filter output precision	16-bits
Spectral analysis filter order	46

Figure 12. Performance summary and die photograph of seizure detection SoC.

|	
Supply voltage	1 V
I-amp LNA power	3.5 μW
I-amp input impedance	>700 MΩ
I-amp noise (input referred, 0.5 Hz–100 Hz)	1.3 μVrms
I-amp electrode offset tolerance	>1 V
I-amp CMRR	>60 dB
ADC resolution	12 b
ADC energy per conversion	250 pJ
ADC max. sampling rate	100 kS/s
ADC INL/DNL	0.68/0.66 LSB
ADC SNDR	65 dB
Digital energy per feature-vector	234 nJ
Feature vector computation rate	0.5 Hz
Total energy per feature-vector	**9 μJ**
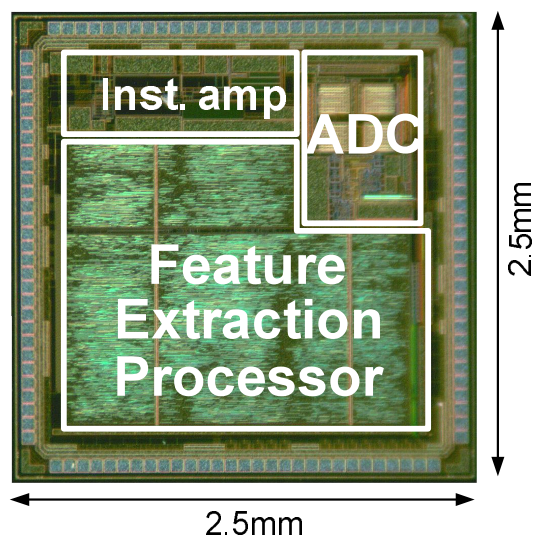
4.5. *Low-energy Programmable Processors Through Voltage Scaling*

In the SoC described, a dedicated processor is used for biomarker extraction. This leads to very low energy, motivating the computation *vs.* communication trade-off of Table 1. As described in

Section 2.2, however, the optimal choice of biomarkers depends on a wide range of considerations from application-to-application. Programmable platforms can thus help cover the range of choices pertaining to biomarker selection and processing. The challenge is that programmability incurs much higher energy. A generic approach that is valuable for reducing energy is voltage reduction. Voltage reduction quadratically lowers the active switching energy (CV_{DD}^2). Although, this degrades circuit speed, allowing leakage power sources to integrate over a longer period, energy is typically minimized at voltages near the subthreshold regime, since here the circuit speed begins degrading rapidly [47]. This is shown, for instance in the case of an MSP430 (implemented in low-power 65 nm CMOS) [48], where the energy is minimized at a V_{DD} of 0.5 V where an operating frequency of 400 kHz is achieved.

Figure 13. Operation of seizure detection system including EEG acquisition, digitization, feature extraction, and classification of α-state (wired transmission of the feature vectors is employed to simplify demonstration of the detection system).
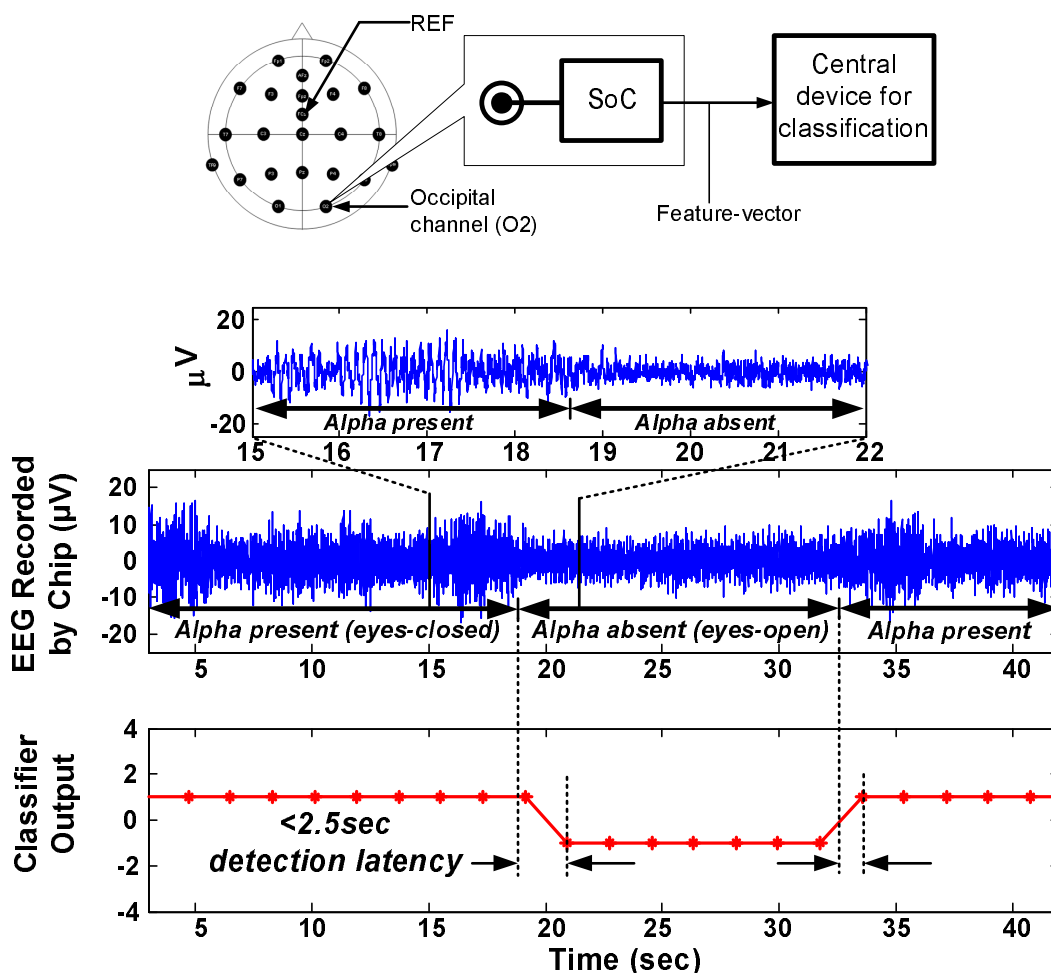
Since many physiological signals of interest have low-bandwidths (*i.e.*, below 10 kHz) [14]. The energy savings afforded by aggressive voltage reduction are compelling. The ability to operate circuits at ultra-low-voltages, however, requires specialized circuit techniques to tolerate the loss of digital noise margins. Voltage scaling dramatically exacerbates sensitivity to process variation. This is most severe in memory circuits, such as SRAMs, which are relied on heavily in programmable platforms. Due to the high number of devices used in large arrays, it is beneficial to use specialized bit-cells (e.g., 8-transistor

topologies). Although these require more devices than the standard 6-transistor topologies, they achieve better area efficiency at low voltages, since the wider noise margins enable aggressive device sizing which would otherwise exacerbate variability to an intolerable level [49]. Logic-gate upsizing to ensure noise margins in the logic circuits may also be required, and can be achieved using a Monte Carlo simulation-based modeling and validation methodology [50].

5. Conclusions and Challenges

The emergence of advanced biomedical sensing and stimulation technologies raises the opportunity for chronic biomedical devices that are clinically compelling. A critical aspect in systems that aim to take advantage of these, however, is the ability to discern specific physiological processes so that proper therapy can be automatically delivered or so that clinically relevant states can be continuously tracked. Interpreting the correlations in the patient signals that are available through chronic sensors is challenging. Namely, physiological processes, and the mechanisms through which they are exhibited in the signals, are typically too complex to model accurately. Further, the precise correlations exhibited may be variable from patient-to-patient, and these must be discriminated with high specificity since numerous physiological processes simultaneously affect the signals. The ability to efficiently and accurately model the data is thus critical. Fortunately, thanks to recent clinical practices, databases now exist where patient signals are annotated with diagnostic information. Additionally, efficient algorithms (from the domain of machine-learning) have emerged that allow us to exploit such databases to model the clinically-relevant correlations between the data and the physiological patient states. This paves the way for promising detection algorithms that can allow chronic biomedical devices to be intelligent.

Incorporating such algorithms in practical devices introduces energy, instrumentation, form-factor, and computational resource constraints. To demonstrate system design, the case study of an SoC for a chronic seizure detection system is considered. The SoC incorporates a low-noise instrumentation amplifier for EEG acquisition, an ADC, and a processor for spectral feature extraction. Analysis, based on patient data testing, is presented for determining the feature-extraction parameters and the computational precision required. This leads to a system that is viable for chronic seizure detection; however, several open challenges remain for future systems. The implementation of low-power machine-learning classifiers is opposed by the complexity of the classification task. The ability to efficiently scale the classifier model for the datasets of interest is thus important in order to avoid compromising the accuracy of a detector [39]. In the case study considered, the power consumption of classification motivated moving its implementation off the scalp, where the battery size and weight limitations are the most severe. Additionally, although data-driven modeling provides a high degree of specificity by training based on previously observed data, physiological changes in the patient can invalidate the generated model. Although, this is not observed to limit seizure detection applications (where seizures in a given patient appear to be stereotypical and stable), it can be important in applications where patient monitoring is required following an acute event to which the physiology responds by evolving over time [23]. Nonetheless, patient data is one of the most important resources in modern healthcare. Exploiting it to achieve accurate patient monitoring systems is emerging as a promising approach. Incorporating these methods in chronic biomedical devices is thus highly

compelling, since new modalities for delivering therapy and sensing patient signals could make such devices invaluable for effective healthcare delivery.

Acknowledgments

References and Notes

1. Benabid, A.L. Deep brain stimulation for Parkinson's disease. *Curr. Opin. Neurobiol.* **2003**, *13*, 696–706.
2. Schachter, S.C. Vagus nerve stimulator. *Epilepsia* **1998**, *39*, 677–686.
3. Kim, D.H.; Viventi, J.; Amsden, J.; Xiao, J.; Vigeland, L.; Kim, Y.S.; Blanco, J.; Frechette, E.; Contreras, D.; Kaplan, D.; *et al.* Dissolvable films of silk fibroin for ultrathin, conformal bio-integrated electronics. *Nat. Mater.* **2010**, *9*, 511–517.
4. Graudejus, O.; Yu, Z.; Jones, J.; Morrison, B.; Wagner, S. Characterization of an elastically stretchable microelectrode array and its application to neural field potential recordings. *J. Electrochem. Soc.* **2009**, *156*, 85–94.
5. Viventi, J.; Kim, D.H.; Moss, J.; Kim, Y.S.; Blanco, J.; Anetta, N.; Hicks, A.; Xiao, J.; Huang, Y.; Callans, D.J.; *et al.* Dissolvable films of silk fibroin for ultrathin, conformal bio-integrated electronics. *Sci. Trans. Med.* **2010**, *2*, 2694–2699.
6. Chandrakasan, A.; Verma, N.; Daly, D. Ultralow-power electronics for biomedical applications. *Annu. Rev. Biomed. Eng.* **2008**, *10*, 247–274.
7. Leong, T.; Aronsky, D.; Michael Shabot, M. Guest editorial: Computer-based decision support for critical and emergency care. *J. Biomed. Inform.* **2008**, *41*, 409–412.
8. Wallis, L. Alarm fatigue linked to patient's death. *Am. J. Nurs.* **2010**, *110*, 16.
9. Dishman, E. Inventing wellness systems for aging in place. *IEEE Comput.* **2004**, *37*, 34–41.
10. Tsien, C. TrendFinder: Automated detection of alarmable trends. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2000.
11. Tompkins, W.J. *Biomedical Digital Signal Processing: C Language Examples and Laboratory Experiments for the IBM PC*; Prentice-Hall: Bergen County, NJ, USA, 1993.
12. Hau, D.; Coiera, E. *Learning Qualitative Models from Physiological Signals*; AAAI Technical Report SS-94-01; AAAI Press: Menlo Park, CA, USA, 1994; pp. 67–71.
13. Hagmann, C.F.; Robertson, N.J.; Azzopardi, D. Artifacts on electroencephalograms may influence the amplitude-integrated EEG classification: A qualitative analysis in neonatal encephalopathy. *Pediatrics* **2006**, *118*, 2552–2554.
14. Yazicioglu, R. Biopotential Readout Circuits for Portable Acquisition Systems. Ph.D. Thesis, Katholieke Universiteit, Leuven, Belgium, 2008.

15. Shoeb, A.; Schachter, S.; Schomer, D.; Bourgeois, B.; Treves, S.T.; Guttag, J. Detecting seizure onset in the ambulatory setting: Demonstrating feasibility. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2005**, *4*, 3546–3550.

16. Gotman, J.; Ives, J.; Gloor, G. Frequency content of eeg and emg at seizure onset: Possibility of removal of EMG artefact by digital filtering. *EEG Clin. Neurophysiol.* **1981**, *52*, 626–639.

17. Shoeb, A.; Bourgeois, B.; Treves, S.; Schachter, S.; Guttag, J. Impact of patient-specificity on seizure onset detection performance. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, 2007; pp. 4110–4114.

18. Meyfroidt, G.; Guiza, F.; Ramon, J.; Bruynooghe, M. Machine learning techniques to examine large patient databases. *Best Pract. Res. Clin. Anaesthesiol.* **2009**, *23*, 127–143.

19. Sajda, P. Machine learning for diagnosis and detection of disease. *Annu. Rev. Biomed. Eng.* **2006**, *8*, 537–565.

20. Podgorelec, V.; Druzovec, T.W. Some applications of intelligent systems in medicine. In *Proceedings of the IEEE 3rd International Conference on Computational Cybernetics*, Mauritius, 2005; pp. 35–39.

21. Shoeb, A.; Guttag, J. Application of Machine Learning to Epileptic Seizure Detection. In *Proceedings of the International Conference on Machine Learning*, Haifa, Israel, 2010.

22. de Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206.

23. Sorenson, D.; Grissom, C.; Carpenter, L.; Austin, A.; Sward, K.; Napoli, L.; Warner, H.; Morris, A.; and for the Reengineering Clinical Research in Critical Care Investigators. A frame-based representation for a bedside ventilator weaning protocol. *J. Biomed. Inform.* **2008**, *41*, 461–468.

24. Jouny, C.; Bergey, G. Dynamics of Epileptic Seizures During Evolution And Propagation. In *Computational Neuroscience in Epilepsy*, 1st ed.; Soltesz, I., Staley, K., Eds.; Academic Press: San Diego, CA, USA, 2008; Volume 28, pp. 457–470.

25. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Networks* **2002**, *13*, 415–425.

26. Rubinstein, Y.D.; Hastie, T. Discriminative vs Informative Learning. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*; AAAI Press: Menlo Park, CA, USA, 1997; pp. 49–53.

27. Ng, A.Y.; Jordan, M.I. On Discriminative *vs.* Generative classifiers: A comparison of logistic regression and naive Bayes. *Adv. Neural Inf. Process. Syst.* **2001**, *14*, 841–848.

28. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.

29. If the classes cannot be well separated by a hyper-plane, the SVM can be used to determine a quadratic or circumferential decision boundary. The SVM determines a nonlinear boundary in the input feature space by solving for linear boundary in higher-dimensional feature space, which is formed by introducing a *kernel*.

30. Theodoridis, S.; Koutroumbas, K. Linear Classifiers. *Pattern Recognition*; Elsevier: Oxford, UK, 2009; Chapter 3.

31. Joachims, T. Text Categorization with Support-Vector Machines: Learning with Many Relevant Features. In *Proceedings of the European Conference on Machine Learning*; Springer: Berlin, Germany, 1998; pp. 137–142.

32. Akbani, R.; Kwek, S.; Japkowicz, N. Applying Support-Vector Machines to Imbalanced Datasets. In *Proceedings of the European Conference on Machine Learning*; Springer: Berlin, Germany, 2004.

33. Kay, S.M. Cramer-Rao Lower Bound. *Fundamentals of Statistical Signal Processing: Detection Theory*; Prentice-Hall: Bergen County, NJ, USA, 1998; Chapter 3, pp. 60–92.

34. Bishop, C.M. Probability Distributions. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006; Chapter 2, pp. 67–127.

35. Lotte, F.; Congedo, M.; Lauyer, A.; Lamarche, F.; Arnaldi, B. A Review of Classification Algorithms for EEG-Based Brain-Computer Interface. *J. Neural Eng.* **2007**, *4*, R1–R13.

36. Shih, E.I. Reducing the Computational Demands of Medical Monitoring Classifiers by Examining Less Data. Ph.D. Thesis, Massachu-setts Institute of Technology, Cambridge, MA, USA, 2010.

37. Avestruz, A.T.; Santa, W.; Carlson, D.; Jensen, R.; Stanslaski, S.; Helfenstine, A.; Denison, T. A 5 μW/channel spectral analysis IC for chronic bidirectional brain-machine interfaces. *IEEE J. Solid-State Circuits* **2008**, *43*, 3006–3024.

38. Yazicioglu, R.F.; Kim, S.; Torfs, T.; Merken, P.; Hoof, C.V. A 30 μW analog signal processor ASIC for biomedical signal monitoring. In *Proceedings of the IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, San Francisco, CA, USA, 2010; pp. 124–125.

39. Shoeb, A.; Carlson, D.; Panken, E.; Denison, T. A Micropower Support Vector Machine Based Seizure Detection Architecture for Embedded Medical Devices. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, Minneapolis, MN, USA, 2009; pp. 4202–4205.

40. Texas Instruments. Low-cost low-power 2.4 GHz RF transmitter. Available Online: http://focus.ti.com/docs/prod/folders/print/cc2550.html (accessed on 1 September 2010).

41. Verma, N.; Shoeb, A.; Bohorquez, J.; Dawson, J.; Guttag, J.; Chandrakasan, A.P. A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. *IEEE J. Solid-State Circuits* **2010**, *45*, 804–816.

42. Dozio, R.; Baba, A.; Assambo, C.; Burke, M.J. Time based measurement of the impedance of the skin-electrode interface for dry electrode ECG recording. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2007)*, Lyon, France, 2007; pp. 5001–5004.

43. Enz, C.; Temes, G. Circuit techniques for reducing the effects of op-amp imperfections: auto-zeroing, correlated double sampling, and chopper stabilization. *Proc. IEEE* **1996**, *84*, 1584–1614.

44. Denison, T.; Consoer, K.; Santa, W.; Avestruz, A.T.; Cooley, J.; Kelly, A. A 2 μW 100 nV/rtHz chopper-stabilized instrumentation amplifier for chronic measurement of neural field potentials. *IEEE J. Solid-State Circuits* **2007**, *42*, 2934–2945.

45. Yazicioglu, R.; Merken, P.; Puers, R.; Hoof, C.V. A 60 μW 60 nV/$\sqrt{Hz}$ readout front-end for portable biopotential acquisition system. *IEEE J. Solid-State Circuits* **2007**, *42*, 1100–1110.

46. Verma, N.; Chandrakasan, A.P. An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor nodes. *IEEE J. Solid-State Circuits* **2007**, *42*, 1196–1205.

47. Wang, A.; Chandrakasan, A.P. A 180 mV subthreshold FFT processor using a minimum energy design methodology. *IEEE J. Solid-State Circuits* **2005**, *40*, 310–319.

48. Kwong, J.; Ramadass, Y.K.; Verma, N.; Chandrakasan, A. A 65 nm Sub-V_t microcontroller with integrated SRAM and switched capacitor DC-DC converter. *IEEE J. Solid-State Circuits* **2009**, *44*, 115–126.

49. Verma, N. Ultra-Low-Power SRAM Design In High Variability Advanced CMOS. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.

50. Kwong, J.; Chandrakasan, A. Variation-Driven Device Sizing for Minimum Energy Sub-threshold Circuits. In *Proceedings of the International Symposium on Low Power Electronics and Design*, Tegernsee, Germany, 2006; pp. 8–13.