BrainForest: Neuromorphic Multiplier-Less Bit-Serial Weight-Memory-Optimized 1024-Tree Brain-State Classification Processor

Gerard O'Leary, *Student Member, IEEE*, Jamie Koerner, *Student Member, IEEE*, Mustafa Kanchwala, *Student Member, IEEE*, Jose Sales Filho, *Student Member, IEEE*, Jianxiong Xu, *Student Member, IEEE*, Taufik A. Valiante, and Roman Genov, *Senior Member, IEEE*

***Abstract*—Personalized brain implants have the potential to revolutionize the treatment of neurological disorders and augment cognition. Medical implants that deliver therapeutic stimulation in response to detected seizures have already been deployed for the treatment of epilepsy. These devices require low-power integrated circuits for life-long operation. This constraint impedes the integration of machine-learning driven classifiers that could improve treatment outcomes. This paper introduces BrainForest, a neuromorphic multiplier-less bit-serial weight-memory-optimized brain-state classification processor. The architecture achieves state-of-the-art energy efficiency using two layers of neuron models to implement the spectral and temporal functions needed for classification: 1) resonate-and-fire neurons are used to extract physiological signal band energy EEG biomarkers 2) leaky integrator neurons are used to build multi-timescale representations for classification. Sparse neural model firing activity is used to clock-gate device logic, thereby decreasing power consumption by 93%. An energy-optimized 1024-tree boosted decision forest performs the classification used to trigger stimulation in response to detected pathological brain states. The IC is implemented in 65nm CMOS with state-of-the-art power consumption (best case: 9.6µW, typical: 118µW), achieving a seizure sensitivity of 97.5% with a false detection rate of 2.08 per hour.**

***Index Terms*—Neural Interface processor, exponentially decaying memory, signal energy, neuromodulation, AdaBoost, bit-serial processing, neuromorphic computing.**

Received 20 April 2024; revised 30 August 2024; accepted 25 September 2024. Date of publication 16 October 2024; date of current version 5 February 2025. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by Canadian Institutes of Health Research (CIHR). This paper was recommended by Associate Editor B. Zhao. *(Corresponding author: Gerard O'Leary.)*

Gerard O'Leary, Jamie Koerner, Mustafa Kanchwala, Jose Sales Filho, Jianxiong Xu, and Roman Genov are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: gerard.oleary@mail.utoronto.ca).

Taufik A. Valiante is with the Department of Surgery (Neurosurgery), University of Toronto, Toronto, ON M5T 2S8, Canada.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TBCAS.2024.3481160.

Digital Object Identifier 10.1109/TBCAS.2024.3481160

I. Introduction

THE synchronized firing of local neural populations within the brain gives rise to oscillations known as local field potentials, or LFPs. These oscillations are now understood to reflect the underlying activity of the brain and categorizing their patterns has led to fundamental insights into physiological and pathological neural mechanisms. For example, the beta wave, which falls between 12-30 Hz is prominent in EEG when a subject is awake and focused. Recent studies have shown that abstract states can be characterized. For example, a distinct electrical signature can be detected during the recall of an image during a visual recognition task [1].

Along with interpreting brain signals, activity can be artificially induced using electrical stimulation. It has been shown that neural firing can be directly modulated during certain phases of electrical stimulus waveforms [2]. Given that we have developed technologies to interact with the brain by reading and inducing activity, the same method can be applied to the treatment of neurological disorders.

Epilepsy affects 65 million people globally, with 30% experiencing adverse drug side effects and another 30% unable to control their seizures by pharmacotherapy [3]. One emerging treatment option is the use of an implanted stimulator which disrupts the propagation of seizures as they are detected in the brain. The key targets that closed-loop stimulators aim to address include:

1) *Maximize seizure prevention.* People who achieve seizure freedom have significantly improved quality of life outcome indicators compared to those with a 75-99% seizure reduction [4]. This requires the detection of *all* seizures.
2) *Minimize side-effects.* The false detection of seizures leads to erroneous stimulation and is associated with memory impairment and depression [5].
3) *Maximize battery life.* Battery replacement surgeries increase the risk and expense of treatment [6]. Procedures can be reduced by minimizing the power consumed by the device's internal electronics.
4) *Minimize clinician burden.* Existing FDA approved devices require neurologists to manually calibrate devices over time, hindering deployment [7].

5) *Support interpretable autonomous stimulation.* Recent advances in machine learning (e.g. deep learning) could be used to greatly enhance the capabilities of devices, but many approaches do not provide interpretable outputs for their classification. This is a key concern for regulatory approval [8].

There are two closed-loop devices for the treatment of epilepsy with FDA approval: the Neuropace RNS [9] and the Liva Nova Aspire SR. The RNS is implanted within patient's skull with a depth electrode placed into a deep brain region and a strip electrode placed on the cortical surface. While this device is effective at reducing seizure frequency, only 9% of people with epilepsy achieve seizure freedom [9]. Furthermore, there are up to 2000 stimulator activations per day, resulting in avoidable side-effects [10]. The LivaNova Aspire SR is implanted in the chest, and electrodes induce activity in the vagus nerve, which in turn modulates brain activity. In this case, only 6.9% of patients are seizure free and with $>$ 200 stimulator activations per day [10]. This limited efficacy is attributed to the use of threshold-based classifiers to identify seizures [11]. The use of more advanced closed-loop computation on the device's internal electronics is constrained by power consumption, as the implant must ideally operate for the lifetime of a patient without replacement.

There exists a tradeoff in neural signal processing between on-device computation, and wireless transmission for remote processing. Devices must operate on a low power budget to maximize their battery life and reduce the number of replacement surgeries which result in a risk of infection and an additional clinical burden. A comparative study found that power dissipation for on-chip signal processing can be an order of magnitude lower than for the equivalent wireless data transmission [12]. Furthermore, it is critical that closed-loop devices should continue to operate autonomously in the event that wireless connectivity is lost. For these reasons, critical processing should be conducted on-device.

The future of ASICs for closed-loop computation demands energy-efficient, low-latency processors for responsive, safe, personalized neuromodulation. Recent developments include the use of neural co-processors as an attempt to bring the capability and flexibility of a microprocessor to a prosthesis embedded within the nervous system [13]. Such devices typically involve three stages: biomarker extraction, brain-state classification, and responsive electrical stimulation.

A broad range of biomarkers have been proposed for brain state classification [14]. For seizure detection, time-domain features such as line length and area-based methods have been described. Spectral features are among the most prevalent. As EEG can be decomposed into physiological frequency bands, the ability to efficiently extract these bands on-device is desirable [15].

Classifiers are typically trained offline, and inference is performed in a continuous manner. As each classification uses energy from the device's battery, power can be reduced by minimizing inference computations. As local field potentials (LFPs) in the brain tend to be digitally sampled every millisecond (or between 0.25-1kHz), a naïve solution would involve performing classification at the same rate. This sampling frequency is required to capture signals with high fidelity, but cognitive timescales are typically on the order of 10's to 100's of milliseconds [16]. For sudden state transitions such as those found at the onset of seizures in epilepsy, dynamic classification rates may be more appropriate. This could be achieved by finding models that can be updated asynchronously, or by using a multi-stage classification process where simple wake-up classifiers are used initially, and more complex models are computed only when necessary.

Existing work has aimed to improve on existing clinical devices by integrating information rich biomarkers to enable better identification of pathological activity [17]. Rather than using manual thresholds, academic solutions have integrated machine learning accelerators such as support vector machines [18]. However the power consumption of these implementations prevents their practical use in chronically implanted devices. Furthermore, proposed approaches typically use black-box models which do not allow for clinical interpretability.

Towards addressing these limitations, this paper is an extended description of the system-on-chip introduced in [19] which focuses on the digital brain-state classification processor. Details for the associated analog circuits can be found in [20]. The proposed solution makes use of a computationally efficient feature extraction method based on a signal energy biomarker, which has widely demonstrated clinical utility. The inspiration for this approach comes from the brain itself as will be discussed in Section IV. Neural inspired computing is also leveraged in the development of a novel classification technique (Section V), which achieves high accuracy with state-of-the-art power efficiency. Furthermore, the training process results in a white-box model which allows for the interpretation of why stimulation decisions were made. These features are integrated on a single IC, and are presented here as BrainForest.

The following section will contain a review of the state of the art, before detailing the key innovations in this work:

- Feature extraction using -fire (RAF) neuron-models. A low-power, low-area alternative to conventional filter-based methods (Section IV).
- Feature temporal representation using bit-serial exponentially decaying memories. Higher precision at greater efficiency compared to alternatives (Section V).
- Classification using bit-serial decision trees with event-driven parallel sub-computation (Section VI). The implementation is multiplier-less, SRAM-less and supports interpretable machine learning models.
- Decision function with on-chip model weight regeneration and classification biasing. The architecture removes the need for model weight memory storage to reduce power consumption and area (Section VII).

The device performance is characterized with publicly available EEG seizure datasets in Section VIII, and the benchtop characterization of the IC is described in Section IX.

TABLE I
ML-ASICS FOR BRAIN STATE CLASSIFICATION

	Lee JSSC 2013 [18]	Chen JSSC 2014 [21]	Altaf JSSC 2015 [22]	O'Leary JSSC 2019 [17]	Huang JSSC 2019 [23]	Wang ISSCC 2020 [24]	Shin JSSC 2022 [25]	Tsai TBioCAS 2023 [26]	Hou ISSCC 2024 [27]	**This work**
Technology (nm)	130	180	180	130	40	180	65	40	180	**65**
Supply Voltage	0.55-1.2	1.8	1	1.2	0.58	1.5	1.2	0.9	1.8-3.3-10	**1**
SRAM (kb)	64	-	64	96	35.8	64	3.17	73.13	-	**0**
Feature extraction	CPU	FFT, Entropy	SE	PLV, CFC, SE	FFT	MODWT-KDE	PLV, PAC, SE, HFO, LL, LMP, Hjorth	-	CFC, PAC, Spectral	**RAF model**
Classifier	SVM	LLS	D2A-LSVM	EDM-SVM	NL-SVM	LS-SVM	NeuralTree	SciCNN	EBM	**BrainForest**
White-box model	No	No	No	No	No	No	No	No	Yes	**Yes**
Class/Hr	-	-	1800	14400	-	266	-	-	-	**Activity-dependent**
Energy/Class. (μJ)	273	77.91	2.73	168.6	170.9	14.2	0.227	28.33	0.19	**0.036**
Sample Memory (s)	6	0.375	3	Exp. Decaying	2	-	0.25-2	1	-	**Exp. Decaying**
Win. Latency (s)	2	0.8	1	0	1	1.2	0.25-2	2	-	**0**
Sensitivity (%)	100†	92‖	95.7†	97.7*	96.6†	97.8†	95.6†	90.4*, 90.3†	98.5‡	**97.15***, **97.5**†
Specificity or FPR	0.05†	-	0.27†	0.185*	0.28†	99.7%†	96.8%†	95.7%*, 93.6%†	99.3%	**0.46***, **2.08**†

* EU Epilepsy Dataset. † CHB-MIT Dataset. ‖ Experimental Data. ‡ Memory Dataset.

II. RELATED WORK

An overview of published neural interface processors is first presented. This review focuses on ICs which target LFP applications. Other neural interface processors have been proposed for spiking activity from microscale recordings (e.g. [28]), but are not considered in this comparison.

One of the earliest examples of neural signal-specific digital processors is detailed in [29]. The general-purpose biomedical signal processing platform accelerates fast Fourier transform (FFT), CORDIC computation and signal energy extraction for EEG and EKG applications. The device uses a low-voltage SRAM to reduce power consumption and introduced a novel method to circumvent glitches which are associated with low-voltage powering approaches. However, the SoC relies on CPU-based classifier computation (which represents the bulk of processing), and is orders of magnitude less energy efficient than dedicated hardware as demonstrated in [17].

The IC introduced in [18] implements an SVM accelerator which supports linear, quadratic, and Gaussian kernels to allow for a trade-off between accuracy and energy efficiency. The device includes an Active-Learning Data Selection (ALDS) accelerator to select the optimal data instances in a training pool in order to reduce the analysis and labeling effort required by an expert during supervised model construction. Although this IC is one of the most significant developments in low-power brain-state classification, it relies on an MSP430 CPU for feature computation and does not include stimulation capabilities such as waveform generation.

The neural-prosthetic SoC in [21] integrates an 8-channel AFE with digital processing including FFT acceleration, approximate entropy (ApEn) features, along with a linear least squares (LLS) classifier. While specialized hardware has the advantage of reducing redundant logic and hence power consumption, limiting devices users to ApEn and LLS greatly reduces the range of applications. Furthermore, this device, along with those mentioned earlier, relies on bi-phasic pulses for symptom control.

The highly-integrated SoC presented in [22] targets epilepsy, and uses signal-band energy (SE) with a combination of SVM classifiers to tradeoff between seizure detection sensitivity and specificity. However, the device uses a windowing approach with a limited ability to capture complex temporal EEG dynamics as outlined in [17]. Furthermore, the device does not implement any general-purpose processing or neuromodulation waveform generation capabilities.

The SoC introduced in [30] is one of the few devices which integrates online learning capabilities in the form of the alternating direction method of multipliers (ADMM) method for parallel computing for SVM training [23]. This training is demonstrated to improve sensitivity by 1.2% and false alarm rates by 36% at the cost of a power dissipation of up to 7mW. However, challenges faced in practice may include "concept drift" [31] and the need for human expert epileptologists to define data labels before retraining.

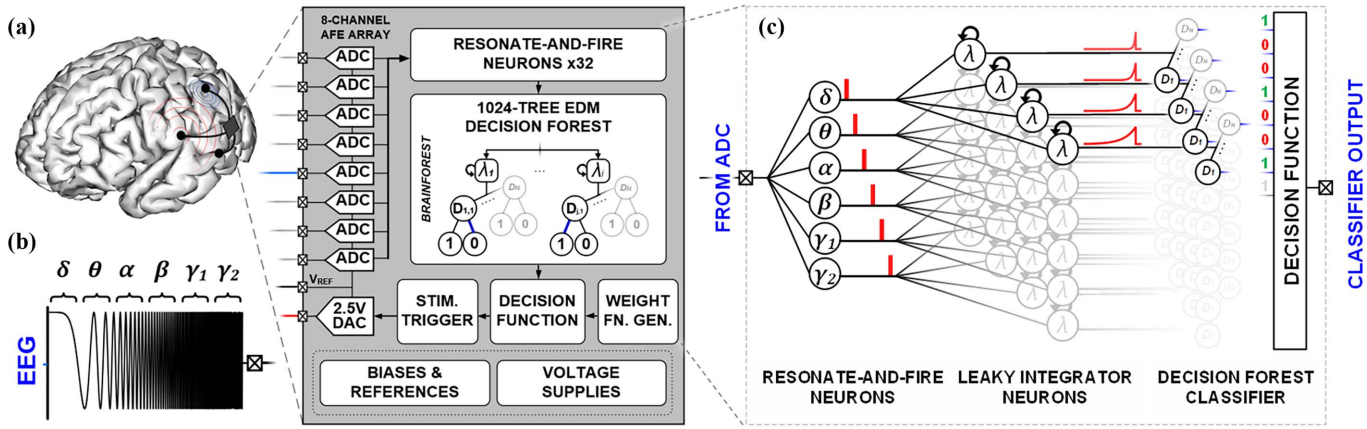

Fig. 1. BrainForest system architecture for responsive neuromodulation. (a) The SoC contains an array of 8 successive approximation register ADCs and a neurostimulation DAC [19] connected to the digital processing blocks detailed in this work. (b) Physiological signal bands present in the electroencephalogram that are used for classification. (c) The processing pipeline used to detect pathological brain states includes 1) resonate-and-fire neurons used to extract physiological signal band features, 2) leaky-integrator neurons used to build feature temporal representations and 3) a 1024-tree boosted decision forest classifier.

In 2020, Wang et al. introduced an SoC with a 2-level brain-state classifier combining a coarse-grain anomalous EEG classifier with a fine-grained least-squares SVM [24]. This is one of the first devices to introduce the concept of coarse and fine classification to reduce the effective classification rate, thus reducing power dissipation. The SoC extracts a feature vector consisting of maximum-modulus discrete wavelet transform (MODWT) and kernel density estimation (KDE) elements. Initial results are reported using synthetic minority over-sampling (SMOTE) [32] which limits the interpretability of the sensitivity and specificity compared to other work. The digital processing IC is demonstrated alongside a separate device for neural signal acquisition and stimulation.

Since the initial publication of this work in [19] a significant development included the introduction of a patient-independent SoC that utilizes a Seizure-Cluster-Inception CNN (SciCNN) for real-time, on-chip classification without the need for patient-specific retraining [26]. This SoC demonstrates how deep learning can be adapted to biomedical hardware, offering high sensitivity and specificity across various patient datasets with minimal power draw.

Recent developments in neural interface processors also include the NeuralTree SoC, a platform offering a 256-channel interface with highly energy-efficient classification and neuromodulation capabilities [25]. This system is notable for its broad range of feature extractors and tree-structured neural network (NeuralTree) classifier, which optimizes the energy-accuracy tradeoff through a patient- and disease-specific approach to symptom detection in neurological disorders such as epilepsy and Parkinson's disease. This SoC achieves a 0.227 μJ per class energy efficiency.

Further recent advances include a multi-loop neuromodulation chipset network with frequency-interleaving front-end and explainable AI designed for memory studies in freely behaving monkeys, which presents a novel approach in handling Alzheimer's related memory issues using neural interfaces [27]. This chipset implements advanced frequency interleaving to cover a range of EEG bands and focuses on the use of explainable AI.

III. Overview

The BrainForest architecture is illustrated in Fig. 1. The core consists of a 1024-tree temporal boosted decision forest. This temporal classification approach addresses two key challenges: 1) the representation of long-short-term temporal activity dependencies, 2) classification of temporal feature vectors with an optimal tradeoff between accuracy and power-consumption. The SoC contains an array of 8 successive approximation register ADCs and a neurostimulation DAC [19]. This paper will focus on the blocks to the right which perform feature extraction, classification, and stimulation functionality.

Feature extraction is implemented using resonate-and-fire neurons which isolate the physiological EEG bands which is conventionally performed using band filtering techniques such as FIR filters. It has been found in neuroscience that some cells exhibit resonant properties in these bands, where firing rates are tuned to the frequency characteristics of somatic potentials [33]. A model of these resonate-and-fire (RAF) neurons that mimics their fundamental mechanisms can be done highly efficiently with digital logic to provide a spiking response to activity in EEG input streams. This is demonstrated as a practical approximation of conventional filter-based methods with a much lower power consumption and without a significant performance difference.

This spiking activity from the RAF neurons must be converted into a temporal representation for classification. Taking inspiration from the brain once again, it has been uncovered that experience unfolding in time is encoded via a set of leaky integrator cells, each of which accepts input from the same stimulus decay at different rates [34]. BrainForest implements leaky integrate-and-fire (LIF) neurons to build a temporal representation of activity from the preceding RAF cells. As demonstrated in Section VIII a larger array of cells with different decay

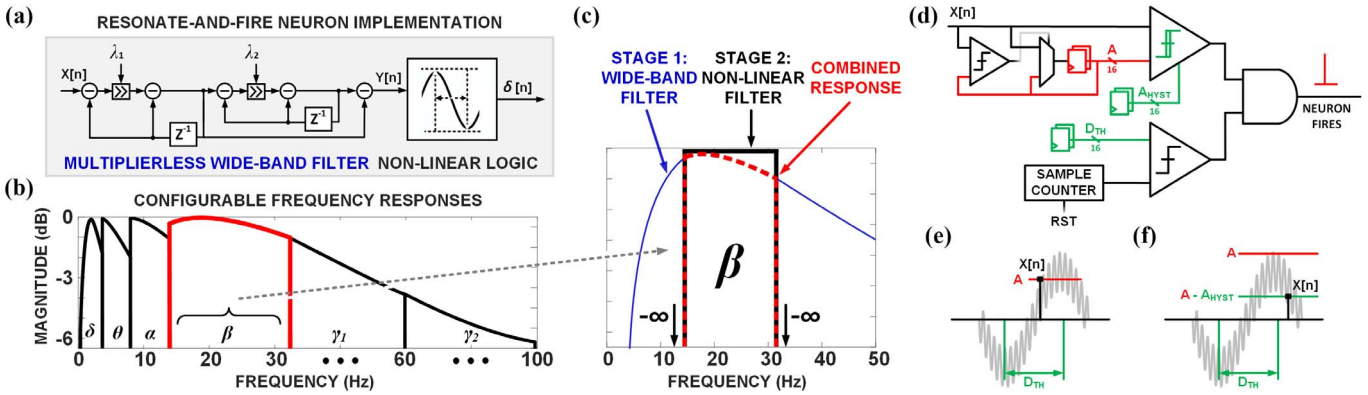

Fig. 2. The Resonate-and-fire (RAF) neural model is used to generate physiological signal band energy features. (a) The two-stage architecture includes coarse-grained multiplier-less bandpass filtering and non-linear firing output logic. (b) The ideal bands extracted using the RAF model approach. (c) The combination of coarse-grain filtering and non-linear logic results in a response approximating a brick-wall bandpass filter with full stop-band attenuation. (d) The non-linear filter logic acts as a half-wave detector. The detected output is comprised of a magnitude and event firing output which can be used for event-driven classification. (e) Conceptual operation of peak detector which follows the input samples to a maximum value. A hysteresis mechanism is included to reduce the influence of noise.

rates can encode representations with higher fidelity, a low-area implementation for each cell can enable larger arrays. This method has previously been demonstrated in [18]. In this work, the LIF mechanism is implemented using a bit-serial approach to reduce area requirements by over 90%.

The LIF temporal representation is classified using an ensemble of 1024 decision trees which are trained using existing additive machine learning models such as Adaptive Boosting (AdaBoost) and Gradient Boosted Machines (GBMs). As the classifier only operates in response to resonant activity in the brain, power is significantly reduced compared to a naïve implementation. All the trees within the model ensemble are independent, allowing each tree to be computed independently in response to feature extraction updates.

Finally the outputs from the decision forest must be aggregated in a final decision function. Additive models require the storage of weights which are associated with individual decision tree results in the ensemble. These weights are typically stored in a global on-chip SRAM. In this work, this requirement is relieved through a combination of model pre-processing and on-chip weight regeneration that exploits inherent statistical distributions. Power efficiency is further increased by leveraging domain-specific knowledge to bias the classifier and reduce the required computations during the inference of more probable classes. In the case of epilepsy less than 1% of brain activity is seizure activity, thus dynamic power can be saved during 99% of the time of the device's operation. Upon the detection of a target brain state, closed-loop intervention is delivered using the on-chip neurostimulator.

These design features are implemented without the use of SRAM or multiply operations to improve area and power efficiency. The feature extraction and classification architectures completely avoid the use of power-intensive multiply operations and global memory accesses. The architecture for the RAF and LIF models is implemented solely using comparisons, shift and add operations with local memory accesses. The decision trees are implemented using comparison and local buffers for comparison points.

IV. Resonate and Fire Neural Model

Brain states can be classified by observing activity in physiological signal bands (δ [0-4], θ [4-8], α [8-14], $\gamma 1$ [32-59], $\gamma 2$ [61-100]). The conventional approach to extracting these bands involves using FIR or IIR-based bandpass filtering [35]. Such filters require power-intensive multiply operations and have finite stop-band attenuation. It has been discovered that neurons within the nervous system exhibit resonant properties in these physiological bands, where firing rates are tuned to the frequency characteristics of somatic potentials [33]. A model of these resonate-and-fire (RAF) neurons is implemented in this work that mimics their fundamental mechanisms. This section describes the digital logic implementation to provide a spiking response to activity in distinct frequency bands.

The underlying architecture of the RAF model is shown in Fig. 2. It is composed of a two-stage model: coarse-grained multiplier-less bandpass filtering, and non-linear firing output logic. The first stage performs initial coarse bandpass filtering of the EEG signal. This is based on concatenated low-pass and high-pass filtering using a multiplier-less, resource-efficient but low-performance IIR filter, which is constrained to power-of-two coefficients to simplify computations to shift-and-add operations. Fig. 2(b) shows the ideal bands which would be extracted using this RAF model. The coarse-grained filtering does not sufficiently allow for the precise differentiation of bands of interest, so further processing is necessary. The second stage is comprised of non-linear filtering logic, which indicates the occurrence of peaks in bands of interest and stores the magnitude as a proxy for signal energy. The resulting response approximates a brick-wall bandpass filter with full stop-band attenuation as shown in Fig. 2(c).

The simplified non-linear filter-based firing output logic is demonstrated in Fig. 2(d). It fires upon detecting a half-wave in the band-filtered signal and stores the amplitude. Two thresholds must be met to trigger an output pulse: an amplitude threshold, $A - A_{\mathrm{HYST}}$, and a wave duration threshold, D_{TH}. The top branch tracks the maximum amplitude (A) until the

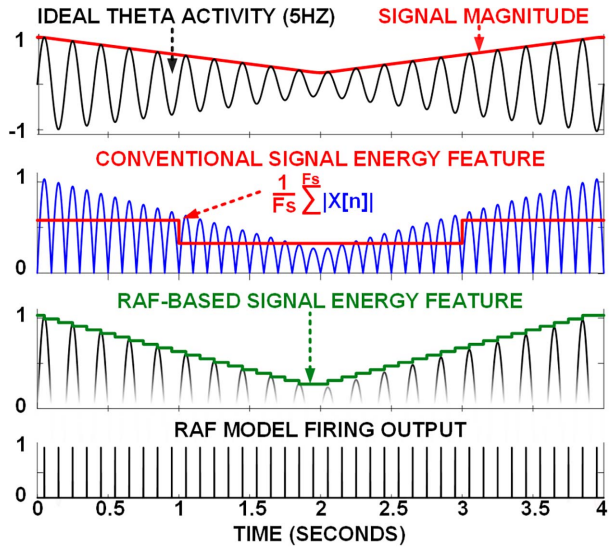

Fig. 3. Resonate-and-fire (RAF) neural model concept compared to conventional methods. Operational example using a synthetic theta wave. E) EEG band sensitivity for band extraction methods.

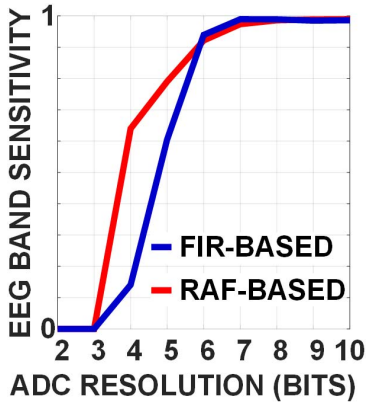

Fig. 4. Resonate-and-fire (RAF) neural model EEG band sensitivity for both FIR and RAF-based approaches. The RAF model shows greater sensitivity to targeted frequency bands at lower ADC resolutions.

signal peaks as shown in Fig. 2(e). A downward slope after the peak indicates a half-wave, triggering the output. Hysteresis prevents noise-induced detections (Fig. 2(f)). The bottom branch counts samples since the last half-wave and compares it to D_{TH}, reflecting the minimum period in the band of interest. When both thresholds are met as shown in Fig. 2(f), the neuron fires, allowing subsequent inference with the new band amplitude, A.

Fig. 3 demonstrates the use of RAF model-based computation of signal band energy versus a conventionally used FIR filter-based approach. An example θ oscillation is overlaid with the signal magnitude. The conventional approach uses the absolute value of the output of an FIR filter and takes the average over a moving window. It is illustrated that the RAF-based approach closely matches the ideal measure. Furthermore, this approach provides a firing output which can later be used to clock gate the activity of proceeding system blocks (Section VI).

The analysis in Fig. 4 compares the effectiveness of FIR-filter-based feature extraction and RAF-based feature extraction under varying signal quantization and noise conditions. The FIR-based method isolates specific frequency components using bandpass filters and evaluates the filtered signal's quality by calculating the power ratio in the target frequency band. This power ratio is the proportion of the signal's power within the desired frequency band relative to the total power across all bands. In contrast, the RAF-based method uses morphological filtering to detect waveform features (half-waves) and assesses their distribution across frequency bands. This calculation is repeated for different quantization levels by quantizing the input signal to various bit depths, simulating the effect of varying ADC resolutions. The analysis demonstrates that the RAF-based approach exhibits greater sensitivity to the energy in EEG bands at lower ADC resolutions. For classification purposes, the key requirement is that the difference in resonant sensitivity between neurons is maximized.

V. Leaky Integrate-and-Fire Neurons

This spiking activity from the RAF neurons must be converted into a temporal representation for classification. An analogous mechanism exists in the brain to represent environmental stimuli. It has been uncovered that experience unfolding in time is encoded in the brain via a set of leaky integrator cells, each of which accepts input from the same stimulus decay at different rates [34]. By comparing the state of each cell, we can infer when this stimulus occurred in the past. This allows us to build a temporal representation of firing activity from the preceding RAF neurons.

The leaky integrator model is based on the use of exponentially decaying memories (EDMs). As in the brain, this has been demonstrated to represent time-series feature data using varying-length moving averages. Multiple timescales are captured using multiple EDMs in parallel with differing decay rates, λ. This enables the learning of feature fluctuations in time e.g. comparing recent activity to average activity over long timescales. As sample rates in EEG recordings are on the order of multiple KHz, it is infeasible to store data on over years of operation. The EDM mechanism thus enables a compressed representation.

It has been shown in the case of epilepsy that a combination of long and short-term temporal information contains vital information for seizure prediction [36]. The selection of an appropriate range of λ values can be patient-specific and thus impractical to optimize manually. As is demonstrated in Fig. 9, the classification model used in the subsequent stage can be used to select the most informative values.

The implementation of LIF cells using conventional logic can result in reduced accuracy for long-term feature representation (Fig. 5). This figure also illustrates a parallel computation implementation. Note that the timescale being encoded, λ, is directly related to the number of shift operations, where the cell half-life is described as:

$$t_{1/2} = \frac{\log(0.5)}{\log(1-\lambda)} \tag{1}$$

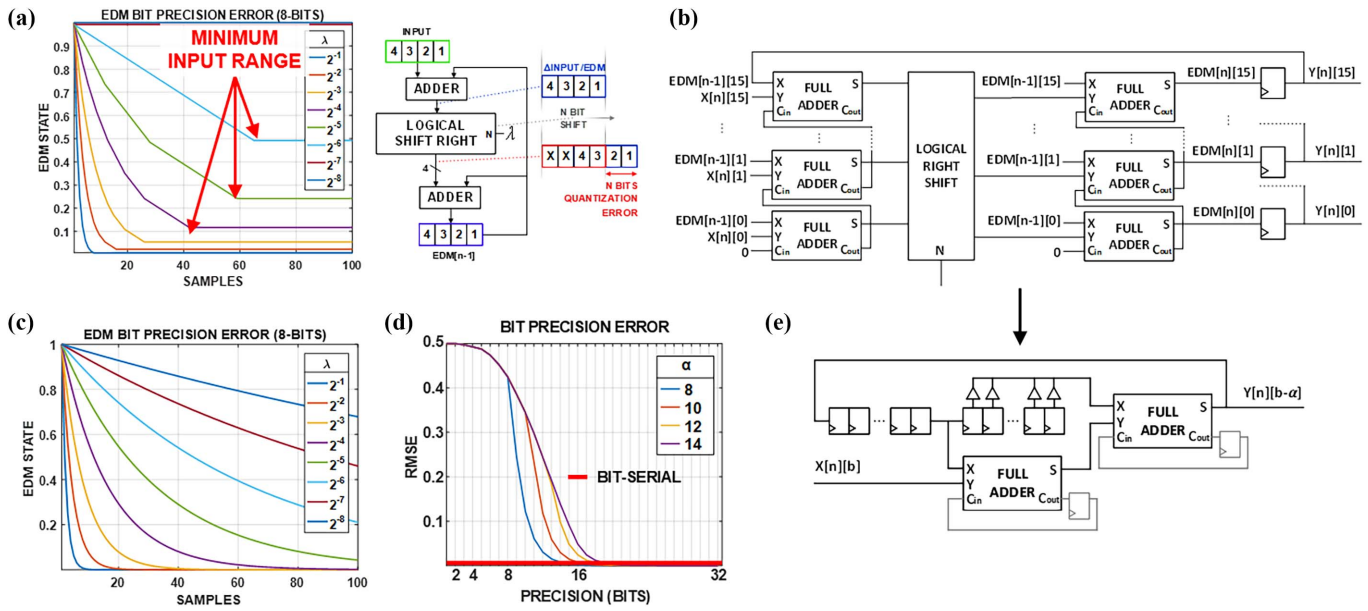

Fig. 5. Bit-Serial exponentially decaying memory (EDM) optimization. (a) Quantization errors limit the minimum input range that can be encoded and causes longer time scales to experience a greater error in their temporal representation. (b) Parallel computation implementation. (c) Migrating to a bit-serial implementation allows for the efficient implementation of dynamic bit precision to prevent quantization errors with longer-term memories. (d) The root-mean-squared error of bit-parallel vs. bit-serial EDMs. (e) Serial computation implementation.

It is shown that quantization errors limit the minimum input range that can be encoded and causes longer time scales to experience a greater error in their temporal representation (Fig. 5(a)). The source of this quantization is shown when N least significant bits are discarded during N right shift operations.

A. Bit-Serial LIF Implementation

There are three implementation challenges for the LIF model:

1) Temporal information can be encoded with an array of EDMs. The more EDMs we have the more accurate the temporal representation [34].
2) Minimize power consumption as the number of cells scale to minimize implantable battery usage.
3) Find a solution that avoids the quantization error illustrated in Fig. 5.

By moving to a bit-serial approach, these limitations can be overcome through the dynamic adjustment of bit-precision to compensate for quantization errors at long-term timescales Fig. 5(c)–5(d). The reformulation of the underlying LIF equation from parallel to serial is detailed as follows. The discrete EDM function is:

$$y[n] = y[n-1] - \lambda(y[n-1] - x[n]) \tag{2}$$

An N-bit digital representation can be described as:

$$y[n] = \sum_{b=1}^{N} 2^{y[n][b]} \tag{3}$$

Each bit, b, can be independently processed as:

$$\begin{aligned} y[n] &= y[n-1] - \lambda(y[n-1] - x[n]) \\ y[n][b] &= y[n-1][b] - \lambda(y[n-1][b] - x[n][b]) \\ y[n][b] &= y[n-1][b] - 2^{-\alpha}(y[n-1][b] - x[n][b]) \\ y[n][b] &= y[n-1][b] - \{(y[n-1][b] - x[n][b]) \gg \alpha\} \\ y[n][b] &= y[n-1][b] - y[n-1][b+\alpha] - x[n][b+\alpha] \end{aligned}$$

Note that the λ term in uses $2^{-\alpha}$ which is a right shift operator. If $x[n]$ is processed as a 1-bit LSB-first serial stream, $x[n] >> \alpha$ can be implemented as:

$$\begin{aligned} y[n] &= y[n-1] - \lambda(y[n-1] - x[n]) \\ \lambda^{-1}y[n] &= \lambda^{-1}y[n-1] - \lambda^{-1}\lambda(y[n-1] - x[n]) \\ \lambda^{-1}y[n] &= \lambda^{-1}y[n-1] - y[n-1] - x[n] \\ 2^{\alpha}y[n] &= 2^{\alpha}y[n-1] - y[n-1] - x[n] \\ (y[n] \ll \alpha) &= (y[n-1] \ll \alpha) - y[n-1] - x[n] \\ y[n][b-\alpha] &= y[n-1][b-\alpha] - y[n-1][b] - x[n][b] \end{aligned}$$

The logic which implements the serial transformation is illustrated in Fig. 5(e). The word shift operation simply requires a delay in serial computation. A cascade of flip-flops with a tri-state buffer-based gate configuration can achieve a configurable number of shifts. By introducing registers to support the delay operation needed for logical shifting, the number of flip flops in the EDM state representation also increases. In effect, the quantization error issue seen in the parallel configuration no longer exists.

The bit-serial approach has the added advantage of a significant logic reduction versus the parallel EDM. While a bit-parallel architecture could be optimized with the use of

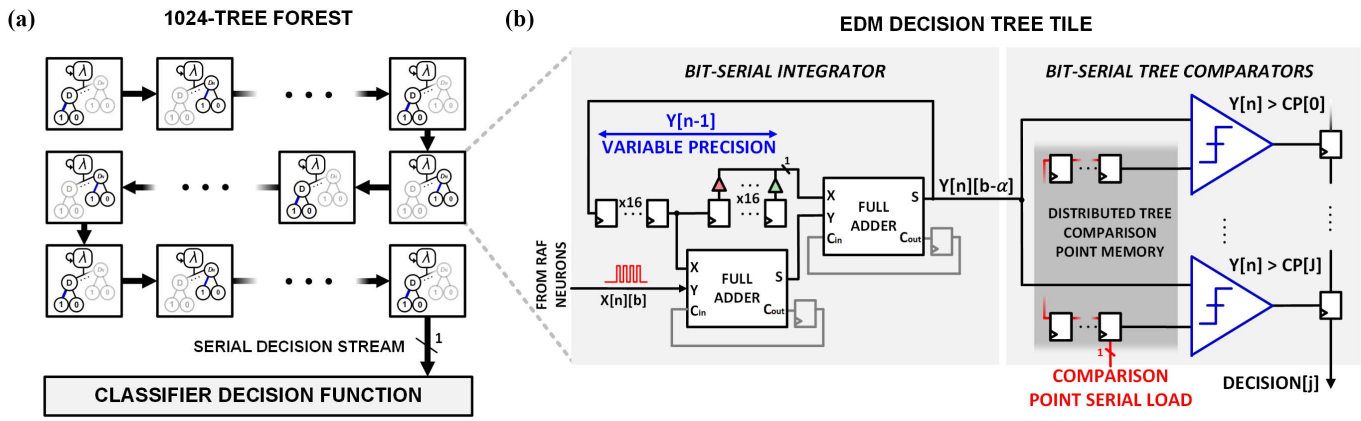

Fig. 6. BrainForest decision forest processing architecture. The EDM memory and comparator operations are tightly coupled for energy efficient processing. The bit-serial tiled architecture of the BrainForest brain-state classifier enables scalability for larger tree ensembles.

compressors, a bit-serial approach reuses the same addition logic sequentially over multiple clock cycles. A diverse range of cells with different decay rates can encode representations with higher fidelity, so a lower-area implementation for each cell enables larger arrays. The area utilization of the parallel approach can be approximated as:

$$A_{\mathrm{EDM-P}} \approx (N \log_2 N \times A_{\mathrm{MUX2}}) + N A_{\mathrm{FA}} \tag{4}$$

and in the case of a serial EDM implementation:

$$A_{\mathrm{EDM-S}} \approx N \times A_{\mathrm{DFF}} + A_{\mathrm{FA}} + A_{\mathrm{DFF}} \tag{5}$$

Synthesis results with a 65 nm process for a 16-bit parallel EDM implementation indicate an area of 495.4 μm^2, and a power consumption of 4.41 μW. A functionally equivalent 16-bit serial EDM implementation requires an area of only 67.6 μm^2, and a power consumption of 0.62 μW.

While the serial EDM implementation incurs an N-cycle processing latency with an associated static power consumption penalty under normal conditions, this is offset by a reduced dynamic power consumption and short logic paths which allow for lower voltage operation. Note that this improvement scales linearly with the number of bits. For longer time scales beyond those implemented in this work, this architecture can be efficiently scaled.

VI. Boosted Trees

The encoded temporal representation of RAF model features can be used with a classifier to infer pathological brain states. Conventional algorithms typically require the use of multipliers for inner product-based computation, and SRAMs to store the large models that these algorithms require. While machine learning techniques based on deep learning have attracted significant attention due to compelling performance in many classification tasks, their efficacy comes at a prohibitive hardware requirement for implantable applications [17]. Prior to the proliferation of deep learning, additive models were highly dominant in tasks such as image recognition [37], and underlying computations include properties that can be leveraged to create highly efficient, implant-compatible brain-state classifiers.

In brief, boosting is a supervised learning meta-algorithm that creates an ensemble weak learners which results in a strong model [38]. We start by training a simple base classifier and label the training examples with its errors. The next simple classifier is trained with weighted attention to the previous model's mistakes. By repeating this process and combining the weak linear classifiers into an ensemble, a highly accurate non linear classifier results.

BrainForest is designed for inference using most additive machine learning algorithms based on depth-of-one decision trees (Fig. 6). The two most widely used include Adaptive boosting (AdaBoost), and gradient boosting machines (GBM) [39], [40]. The tree computation in the BrainForest architecture is fundamentally based on comparison operations between temporally encoded RAF model features, and model weights which have been trained offline. The inference computations and the storage of model weights are key considerations that have been the focus of optimization with machine learning architectures.

Two design features overcome the conventional requirement for on-chip SRAMs when using decision trees: 1) By constraining the decision tree depth, D=1, memory access are fully deterministic, removing the random-access requirement. Deeper trees require non-deterministic memory access based on the outcome of each splitting point. With D=1, memory access is deterministic and there is no capacity redundancy. 2) The compactness of the required model means that parameters can be integrated locally within the model's distributed compute elements, effectively enabling in-memory computation. A drawback of this architecture is that it limits the ability to dynamically change the model size, or store multiple models. However, this application assumes that a single patient-specific seizure classification model is stored on the device with a fixed size of less than 1024 trees.

Significantly, no multiply operations are needed in this model as the computations are comparison based. Furthermore, comparison operations are particularly amenable to bit-serialization as the main computation can be implemented using simple subtraction logic. As the feature data stream is also bit-serial, the computation can be serialized with little support logic overhead (Fig. 6).

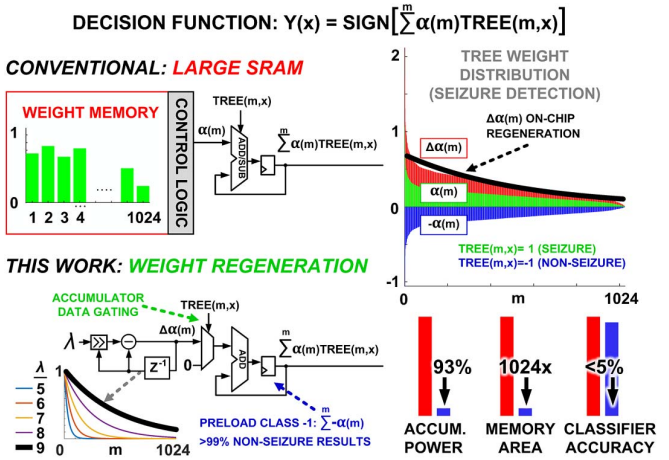

Fig. 7. Classification decision function implementation. The on-chip model weight reconstruction logic removes the memory requirement for the classifier decision function. Tree tiles are configured on the device sorted in order of weight magnitude. The weight regeneration logic approximates the tree weight based on the order of computation.

Each tree in the ensemble is independent, meaning that computation can be performed asynchronously for each tree. These sub-computations are triggered using event signaling from the RAF models as new energy magnitude values are available. Switching activity and dynamic power consumption is therefore directly related to the brain activity which is encoded by the RAF models. This results in a 93% power reduction when compared to a naïve implementation in which all trees are updated synchronously (see Section IX).

VII. Classification Decision Function

Typically, the classifier decision function is calculated by adding or subtracting the weight for each tree. These weights are typically stored in a global on-chip SRAM. In this work, a combination of model pre-processing and on-chip weight regeneration removes this requirement as shown in Fig. 7.

The pre-processing step exploits the typical distribution of values for tree coefficients with two classes. As these weights follow an exponential distribution, they can be reconstructed on-chip using function approximation logic which is pre-loaded with the max term, and decays at a rate which approximately fits the model weight distribution. The model's trees are then mapped to the tile array sorted based on the magnitude of their weights. It should be considered that while the approximation approach removes the need for a dedicated weight memory for this portion of inference, it does come at a cost of degraded accuracy. Furthermore, this degradation can be greater when all 1024 trees are in use when compared to smaller models. The impact of these constraints should be assessed for each application. When deploying the SoC to detect seizures in epilepsy, the accumulator is pre-loaded with weights to represent the interictal class. Due to the rarity of seizures, the decision tree accumulation is only active for $< 1\%$ of the device's operation, greatly reducing dynamic power consumption. Upon the detection of a desired state, neurostimulation is triggered using a 2.5 V on-chip current-mode stimulator.

TABLE II
CHB-MIT Dataset Performance

EU Patient	Sex	Age	Hours Analyzed	Clinical Seizures	Seizures Detected	Sensitivity (%)	FDR
CHB01	F	11	42	6	6	100.00	0.26
CHB07	F	1.5	19	3	3	100.00	0.10
CHB08	M	3.5	20	5	5	100.00	0.18
CHB09	F	10	19	4	3	75.00	0.36
CHB10	M	3	25	7	7	100.00	0.16
CHB18	F	18	35	5	5	100.00	12.20
CHB19	F	19	29	3	3	100.00	0.19
CHB21	F	13	33	8	8	100.00	1.88
CHB22	F	9	31	3	3	100.00	0.12
CHB23	F	6	9	7	7	100.00	5.38
Sum	-	-	**262**	**51**	**50**	-	-
Mean (STD)	-	**9.4 (5.76)**	**26.2 (9.18)**	**5.1 (1.76)**	**5.0 (1.84)**	**97.5 (7.5)**	**2.08 (3.72)**

VIII. BrainForest Model Characterization

The BrainForest classifier performance was first characterized using the CHB-MIT Scalp EEG Database [41]. The dataset consists of 22-channel EEG recordings sampled at 256 Hz from pediatric subjects with refractory seizures collected at the Boston Children's Hospital. The electrodes were positioned on the scalp according to the International 10-20 system. Each record is annotated with the number of seizures the patient experience. If a record contains one or more seizures, the annotation also includes information about the start and end of each seizure for delineation. For the purpose of validation, ten patients were selected randomly as outlined in Table II. Data was first downsampled to 256 Hz and 8 electrodes were chosen on a per patient basis based on their activity during the patient's respective seizures. For the 1024-tree model supported by this architecture, a seizure sensitivity of 97.5% with a false detection rate of 2.08 per hour have been obtained.

BrainForest was further validated using the EU Epilepsy database [42]. This database contains intracranial recordings from 30 epilepsy patients with an average continuous recording time of 150 hours per patient. A review of the results from early seizure detection approaches using this dataset can be found in [43]. For this analysis, data was first downsampled to 256 Hz and 8 electrodes were chosen on a per patient basis based on their proximity to the seizure onset zone. For a 1024-tree BrainForest model, a seizure sensitivity of 97.15% with a false detection rate of 0.46 per hour have been obtained (Table III). The BrainForest functionality is illustrated with this dataset in Fig. 8, showing raw intracranial EEG recordings for a single channel, the RAF model output, and example FIR filter-based output for comparison, and the classifier output alongside expert annotations

The white box characteristics of the GBM is illustrated in Fig. 9. The features used with the highest frequency can be directly mapped to the corresponding brain region for potential seizure onset zone identification. The model transparency also enables the analysis of feature importance's which can be used to inform future architectures and configurations. For example, it is seen that increasing the number of trees used by the model is correlated with more varied use of LIF timescales.

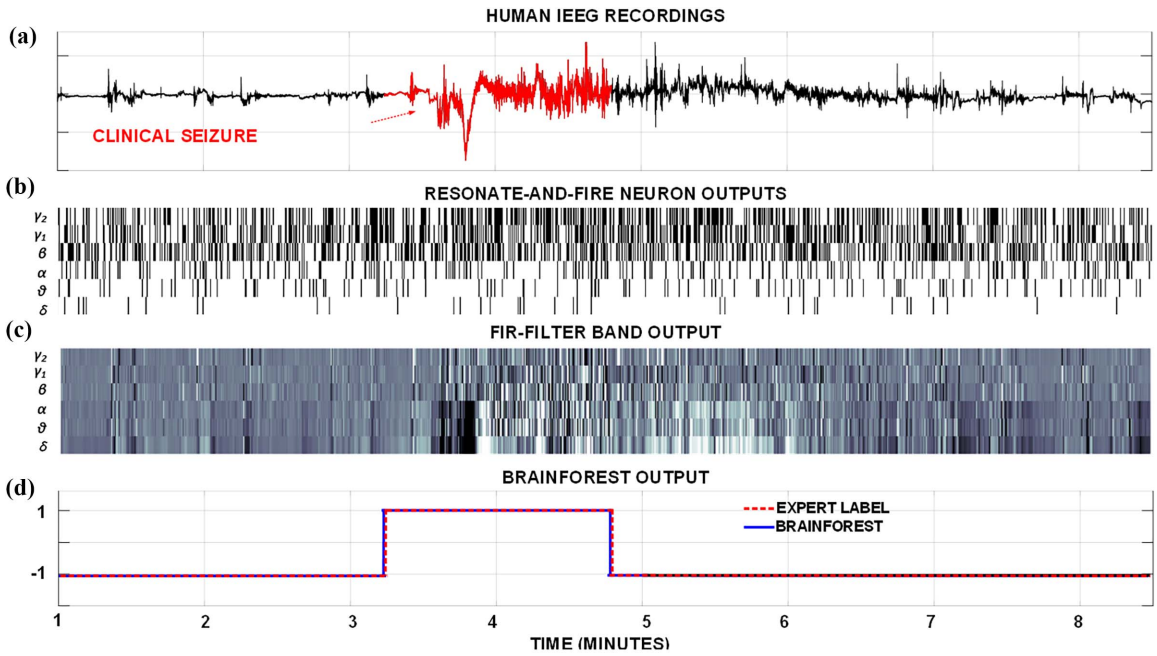

Fig. 8. BrainForest output illustrating (a) raw intracranial EEG recordings, (b) RAF model output, (c) FIR filter-based output, and (d) the classifier output alongside expert annotations.

TABLE III
EU EPILEPSY DATABASE PERFORMANCE

EU Patient	Sex	Age	Hours Analyzed	Clinical Seizures	Seizures Detected	Sensitivity (%)	FDR
EU442	M	21	118	21	21	100.00	0.33
EU548	M	17	129	27	26	96.29	1.30
EU1096	F	32	147	8	8	100.00	0.07
EU1125	F	11	108	13	12	92.31	0.16
Sum	-	-	502	69	67	-	-
Mean (STD)	-	20.25 (7.66)	125.5 (14.47)	17.25 (7.29)	16.75 (7.12)	97.15 (3.18)	0.46 (0.49)

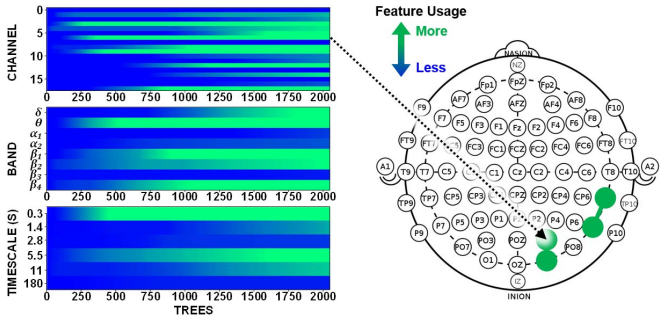

Fig. 9. BrainForest-driven seizure activity insight discovery. The white-box nature of the boosted decision tree approach means that tree weights are associated with relevant biomarkers for seizure onsets. For example, the most highly weighted channels are related to the seizure onset zone. As the number of trees in the model increases, a more varied use of channel channels, bands and leaky integrate-and-fire timescales occurs.

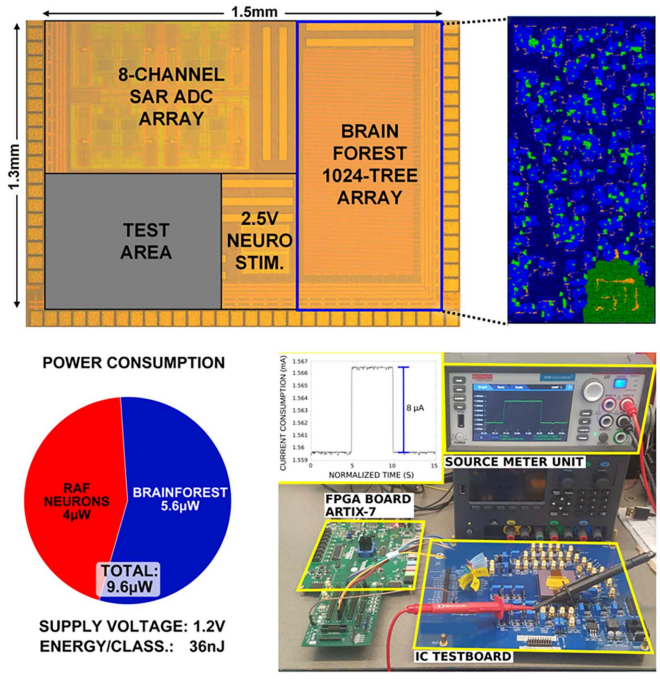

Fig. 10. BrainForest micrograph and experimental setup. A measurement of 8μA at 1.2 V is taken during the simultaneous operation of the RAF feature extractors and BrainForest classifier.

IX. BRAINFOREST BENCHTOP CHARACTERIZATION

The SoC micrograph and comparison table are shown in Fig. 10. The device was fabricated using a TSMC 65nm CMOS process with a silicon area of 1.5mm x 1.3mm (1.95mm^2) with a logic size of 160k NAND2 equivalent gates using an ARM standard cell library. The BrainForest power consumption is measured as 118μW (6μW RAF neurons, 112 μW classification logic) using a nominal voltage of 1.2 V at an operational frequency of 1 MHz. Due to the asynchronous clock-gating of

device logic, the energy use is brain-state dependent. Inference during interictal activity is 9.6μW (99% of time in operation), and 118μW during inference of seizure activity (1% of time in operation). The system performance enables real-time neural signal processing and state classification an average energy per classification of 36nJ. Given a battery capacity of 705 mAh, at a 1 V supply during full operation, the battery would last 2.53 years (118 μW). However, the RAF neuron firing rate greatly reduces during periods of low activity (e.g. interictal periods), and the power consumption drops to 9.6 μW. Assuming seizure activity accounts for 1% of a patient's brain states, the effective power consumption would instead be 10.684 μW, resulting in a battery life of 27.89 years. Note that this excludes the DAC and ADC power consumption.

X. Conclusion

Due to the low power consumption and classification performance, BrainForest could be the key to pioneering the exploration of new applications for neural interfaces in ameliorating and augmenting the brain. As high channel-count devices scale, this architecture is particularly suitable for the increased classification dimensionality requirements. Towards the five key features that closed-loop stimulators must aim to address, BrainForest achieves the following:

1) *Achieves seizure freedom.* Brainforest demonstrates the capabilities to operate with close to 100% seizure sensitivity.
2) *Minimizes side-effects.* False positives are minimized to reduce stimulation-associated side-effects.
3) *Maximizes battery life.* This work demonstrates power efficiency capable of achieving multi-year operation using existing battery technology.
4) *Minimizes clinician burden.* Automated closed-loop stimulation is supported based using machine learning to relieve the clinical burden device configuration.
5) *Supports interpretable autonomous stimulation.* BrainForest's white-box model approach allows for the interpretability of disease biomarkers and generating clinical insights into the neurological basis of brain disorders.

BrainForest could be the key to pioneering the exploration of new applications for neural interfaces in ameliorating and augmenting the brain.

Acknowledgment

The BrainForest architecture was designed and implemented by G. O'Leary. The ADCs were designed by J. Xu, the DAC was designed by L. Long, and system support blocks were designed by J. S. Filho. The authors thank CMC Microsystems for software and fabrication support.

References

[1] M. T. Kucewicz et al., "Combined single neuron unit activity and local field potential oscillations in a human visual recognition memory task," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 67–75, Jan. 2016.

[2] D. Reato, A. Rahman, M. Bikson, L.C. Parra, "Low-intensity electrical stimulation affects network dynamics by modulating population rate and spike timing," *J. Neurosci.*, vol. 30, no. 45, pp. 15067–15079, Nov. 2010.

[3] I. o. M. U. C. o. t. P. H. D. o. t. "Epilepsies," in *Epilepsy Across the Spectrum: Promoting Health and Understanding* M. J. England, C. T. Liverman, A. M. Schultz, and L. M. Strawbridge, Eds. Washington, DC, USA: National Academies Press, 2012.

[4] G. L. Birbeck, R. D. Hays, X. Cui, and B. G. Vickrey, "Seizure reduction and quality of life improvements in people with epilepsy," *Epilepsia*, vol. 43, no. 5, pp. 535–538, 2002, doi: 10.1046/j.1528-1157.2002.32201.x.

[5] A. I. Tröster, K. J. Meador, C. P. Irwin, R. S. Fisher, and SANTE Study Group, "Memory and mood outcomes after anterior thalamic stimulation for refractory partial epilepsy," *Seizure*, vol. 45, pp. 133–141, Feb. 2017.

[6] H. Kahlow and M. Olivecrona, "Complications of vagal nerve stimulation for drug-resistant epilepsy: A single center longitudinal study of 143 patients," *Seizure*, vol. 22, no. 10, pp. 827–833, Dec. 2013.

[7] G. P. Thomas and B. C. Jobst, "Critical review of the responsive neurostimulator system for epilepsy," *Med. Devices*, vol. 8, 2015, Art. no. 405.

[8] B. Patel, "ARTIFICIAL Intell.-Mach. Learn.: VALIDATION," p. 17.

[9] C. N. Heck et al., "Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the RNS System Pivotal trial,"*Epilepsia*, vol. 55, no. 3, pp. 432–441, 2014, doi: 10.1111/epi.12534.

[10] G. K. Bergey et al., "Long-term treatment with responsive brain stimulation in adults with refractory partial seizures," *Neurology*, vol. 84, no. 8, pp. 810–817, Feb. 2015.

[11] H. Cagnan, T. Denison, C. McIntyre, and P. Brown, "Emerging technologies for improved deep brain stimulation," *Nature Biotechnol.*, vol. 37, no. 9, pp. 1024–1033, Sep. 2019.

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

[13] S. Stanslaski et al., "A chronically implantable neural coprocessor for investigating the treatment of neurological disorders," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 6, pp. 1230–1245, Dec. 2018.

[14] S. Ramgopal et al., "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & Behav.*, vol. 37, pp. 291–307, Aug. 2014.

[15] D. M. Groppe et al., "Dominant frequencies of resting human brain activity as measured by the electrocorticogram," *NeuroImage*, vol. 79, pp. 223–233, Oct. 2013.

[16] D. Papo, "Time scales in cognitive neuroscience," *Frontiers Physiol.*, vol. 4, no. 86, Apr. 2013.

[17] G. O'Leary, D. M. Groppe, T. A. Valiante, N. Verma, and R. Genov, "NURIP: Neural interface processor for brain-state classification and programmable-waveform neurostimulation," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3150–3162, Nov. 2018.

[18] K. H. Lee and N. Verma, "A low-power processor with configurable embedded 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.

[19] G. O'Leary et al., "26.2 a neuromorphic multiplier-less bit-serial weight-memory-optimized 1024-tree brain-state classifier and neuromodulation SoC with an 8-channel noise-shaping SAR ADC array," in *Proc. IEEE Int. Solid-State Circuits Conf. - (ISSCC)*, 2020, pp. 402–404, [Online]. Available: https://ieeexplore.ieee.org/document/9062962

[20] J. Xu, S. Wu, H. You, J. De Sales Filho, M. Kanchwala, and R. Genov, "Advanced noise-shaping SAR ADCs utilizing single-capacitor arbitrary-resolution DACs for miniaturized neural interfaces," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2023, pp. 1–5, [Online]. Available: https://ieeexplore-ieee-org.myaccess.library.utoronto.ca/document/10388439/?arnumber=10388439

[21] W.-M. Chen et al., "A fully integrated 8-channel closed-loop neural-prosthetic CMOS SoC for real-time epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, Jan. 2014.

[22] M. A. Bin Altaf, C. Zhang, and J. Yoo, "A 16-channel patient-specific seizure onset and termination detection SoC with impedance-adaptive transcranial electrical stimulator," *IEEE J. Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, Nov. 2015.

[23] S.-A. Huang and C.-H. Yang, "A hardware-efficient ADMM-based SVM training algorithm for edge computing," 2019, *arXiv:1907.09916*.

[24] Y. Wang, Q. Sun, H. Luo, X. Chen, X. Wang, and H. Zhang, "26.3 a closed-loop neuromodulation chipset with 2-level classification achieving 1.5Vpp CM interference tolerance, 35dB stimulation artifact rejection in 0.5ms and 97.8% sensitivity seizure detection," in *Proc. IEEE Int. Solid- State Circuits Conf. - (ISSCC),* 2020, pp. 406–408.
[25] U. Shin et al., "NeuralTree: A 256-channel 0.227-μJ/Class versatile neural activity classification and closed-loop neuromodulation SoC," *IEEE J. Solid-State Circuits*, vol. 57, no. 11, pp. 3243–3257, Nov. 2022.
[26] C.-W. Tsai, R. Jiang, L. Zhang, M. Zhang, and J. Yoo, "Seizure-cluster-inception CNN (SciCNN): A patient-independent epilepsy tracking SoC with 0-shot-retraining," *IEEE Trans. Biomed. Circuits Syst.*, vol. 17, no. 6, pp. 1202–1213, Dec. 2023.
[27] Y. Hou et al., "33.4 A multi-loop neuromodulation chipset network with frequency-interleaving front-end and explainable ai for memory studies in freely behaving monkeys," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, vol. 67, Feb. 2024, pp. 548–550.
[28] T. C. Chen, K. Chen, Z. Yang, K. Cockerham, and W. Liu, "A biomedical multiprocessor SoC for closed-loop neuroprosthetic applications," in *Proc. IEEE Int. Solid-State Circuits Conf. - Dig. Tech. Papers*, 2009, pp. 434–435,435a.
[29] J. Kwong and A. P. Chandrakasan, "An energy-efficient biomedical signal processing platform," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1742–1753, Jul. 2011.
[30] S.-A. Huang, K.-C. Chang, H.-H. Liou, and C.-H. Yang, "A 1.9-mW SVM processor with on-chip active learning for epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 452–464, Feb. 2020.
[31] S. C. H. Hoi, D. Sahoo, J. Lu, and P. Zhao, "Online learning: A comprehensive survey," 2018, *arXiv:1802.02871*.
[32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, Jan. 2002.
[33] F. G. Pike, R. S. Goddard, J. M. Suckling, P. Ganter, N. Kasthuri, and O. Paulsen, "Distinct frequency preferences of different types of rat hippocampal neurones in response to oscillatory input currents," *J. Physiol.*, vol. 529, no. Pt 1, Nov. 2000, Art. no. 205.
[34] M. W. Howard et al., "A unified mathematical framework for coding time, space, and sequences in the hippocampal region," vol. 34, no. 13, pp. 4692–4707.
[35] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data–A practical approach," *J. Neurosci. Methods*, vol. 250, pp. 34–46, Jul. 2015.
[36] M. O. Baud et al., "Multi-day rhythms modulate seizure risk in epilepsy," *Nature Commun.*, vol. 9, no. 1, 2018, Art. no. 88.
[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. New York, NY, USA: Springer Science & Business Media, Aug. 2009.
[38] R. E. Schapire, "The strength of weak learnability," in *30th Annu. Symp. Found. Comput. Sci.*, Research Triangle Park, NC, USA, 1989, pp. 28–33.
[39] Y. Freund and R. E. Schapire, "A short introduction to boosting," *J. Jap. Soc. Artif. Intell.*, vol. 14, no. 5, pp. 771–780, 1999.
[40] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[41] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," PhD Thesis, Cambridge, MA, USA: MIT Press, 2009.
[42] M. Ihle et al., "EPILEPSIAE – A European epilepsy database," *Comput. Methods Programs Biomedicine*, vol. 106, no. 3, pp. 127–138, Jun. 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0169260710002221
[43] M. Bandarabadi, J. Rasekhi, C. A. Teixeira, T. I. Netoff, K. K. Parhi, and A. Dourado, "Early seizure detection using neuronal potential similarity: A generalized low-complexity and robust measure," *Int. J. Neural Syst.*, vol. 25, no. 5, Aug. 2015, Art. no. 1550019, doi: 10.1142/S0129065715500197.

Gerard O'Leary (Student Member, IEEE) received the B.Eng. degree from the National University of Ireland, Galway, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Toronto. He has held Engineering positions at Analog Devices, The European Space Agency, and ARM. His research is at the interface between neural and electronic systems, creating low-power integrated circuits which utilize signal processing and machine learning techniques to infer brain states and control activity using responsive neuromodulation. He is the recipient of the IEEE Solid-State Circuits Society (SSCS) Predoctoral Achievement Award.

Jamie Koerner (Student Member, IEEE) received the B.A. and M.A. degrees in economics, the B.C.S degree in computer science from the University of British Columbia, in 2015, 2016, and 2018, respectively, and the M.A.Sc. degree in electrical and computer engineering from the University of Toronto, in 2020. He is currently working toward the Ph.D. degree with Massachusetts Institute of Technology. His research is primarily concerned with the use of consumer electronics and deep learning-based eye tracking algorithms for neurological disease monitoring.

Mustafa Kanchwala (Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from Sir Padampat Singhania University, Udaipur, India, in 2013, and the M.S. degree in biomedical engineering from the Case Western Reserve University, Cleveland, OH, USA, in 2018. He is currently working toward the Ph.D. degree in electrical and computer engineering with the University of Toronto. His research interests include designing low power digital integrated circuits and hardware accelerators for machine learning focusing on biomedical applications.

Jose Sales Filho (Student Member, IEEE) received the B.Eng. and M.A.Sc. degrees in electrical engineering from the Universidade Federal do Rio Grande do Norte, Brazil, in 2014 and 2016, respectively. He is currently working toward the Ph.D. degree with the University of Toronto, Toronto, ON, Canada. His research interests include ultra lowpower wireless circuits for neural interfaces, readout circuits for multimodal sensory systems and wireless power transfer.

Jianxiong Xu (Student Member, IEEE) received the bachelor's degree in optical communication from the University of Electronics Science and Technology of China, Sichuan, China, in 2015, and the master's degree in electronics from the University of Toronto, Toronto, ON, Canada, in 2018. He is currently working toward the Ph.D. degree with the University of Toronto. His research interest includes designing low power circuits for biomedical applications.

Taufik A. Valiante received the MD/Ph.D. degree from the University of Toronto, in 1997, with doctoral studies completed under the supervision of Dr. Peter Carlen at Toronto Western Research Institute. In that year he began his residency in the Neurosurgery Training Program with the University of Toronto, and he became a fellow of the Royal College of Physicians and Surgeons of Canada in 2003. He pursued an epilepsy neurosurgery fellowship with the University of Washington, Seattle, under the mentorship of Dr. George Ojemann. There, he learned Dr. Ojemann's technique of language mapping for patient's undergoing epilepsy surgery. In 2003, he was recruited back to the University of Toronto as an Assistant Professor and, as a CIHR Clinician-Scientist, he launched his research program in the field of electrophysiologal characterization of neuronal networks in epilepsy. His neurosurgical staff appointment is at the Toronto Western Hospital with a practice in adult epilepsy surgery. He currently serves as the Co-Director of the Epilepsy Program with Krembil Neuroscience Centre, University Health Network and is appointed as an Associate Member of the Institute of Medical Science, University of Toronto.

Roman Genov (Senior Member, IEEE) received the B.S. degree in electrical engineering from Rochester Institute of Technology, Rochester, NY, USA, in 1996, and the M.S.E. and Ph.D. degrees in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, USA, in 1998 and 2003, respectively. Currently, he is a Professor with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he is a member of the Electronics Group and the Biomedical Engineering Group and the Director of the Intelligent Sensory Microsystems Laboratory. His research interests include analog integrated circuits and systems for energy-constrained biological, medical, and consumer sensory applications. He was a co-recipient of the Best Paper Award at the IEEE Biomedical Circuits and Systems Conference, the Best Student Paper Award at the IEEE International Symposium on Circuits and Systems, the Best Paper Award of the IEEE Circuits and Systems Society Sensory Systems Technical Committee, the Brian L. Barge Award for Excellence in Microsystems Integration, the MEMSCAP Microsystems Design Award, the DALSA Corporation Award for Excellence in Microsystems Innovation, and the Canadian Institutes of Health Research Next Generation Award. He was a Technical Program Co-Chair at the IEEE Biomedical Circuits and Systems Conference. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II: EXPRESS BRIEFS and the IEEE SIGNAL PROCESSING LETTERS, and a member of the IEEE International Solid-State Circuits Conference International Program Committee. Currently, he is an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.