Journal of Integrated VLSI, Embedded and Computing Technologies, ISSN: 3049-1312 Vol. 3, No. 1, 2026 (pp. 21-30)

RESEARCH ARTICLE ECEJOURNALS.IN

Bio-Inspired Computing Architectures for Energy-Efficient Biomedical Signal Processing on Chip

Sungho Jeon[1]*, Kriti S. Chakma[2]

[1]*Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea*

[2]*Department of Electrical and Electronic Engineering, International Islamic University Chittagong, Chittagong 4318, Bangladesh*

KEYWORDS:
Bio-Inspired Computing,
Neuromorphic Architecture,
Spiking Neural Networks,
Biomedical Signal Processing,
Energy Efficiency,
Edge AI,
On-Chip Learning,
Low-Power VLSI

ARTICLE HISTORY:
Submitted : 17.08.2025
Revised : 07.09.2025
Accepted : 13.11.2025

https://doi.org/10.31838/JIVCT/03.01.04

Abstract

Real-time biomedical signal processing has used a lot of energy-efficient applications, and with the increasing number of wearable, implantable, and portable healthcare instruments, the necessity will increase. The traditional von Neumann-based architectures provide high flexibility but with memory bottleneck and power consumption they are inapplicable in continuous monitoring anomalies in resource-constrained systems. In that regard, a new solution can be considered in bio-inspired computing architectures that reproduce the general features of the parallel, event-based, and low-power human brain. In this paper we present a neuromorphic computing architecture of the biomedical signal processing on-chip by spiking neural networks (SNNs) and analog-digital mixed-signal integrations. The architecture consists of an analog front-end consuming very low power to perform real-time acquisition of biosignals and coding of the spikes, and a digital SNN core to perform the sparse and asynchronous processing. The system was designed and simulated using a 65nm CMOS technology that makes it very energy efficient and latency efficient. In particular, it can realize up to 65 percent of energy savings per inference count and up to 40 percent low-latency processing over traditional biomedical processors using DSPs. It is tested with benchmark biosignal data such as ECG (MIT-BIH), EEG (Bonn university), and EMG (Ninapro) with use cases such as arrhythmia detection, epileptic seizure prediction, and gesture recognition that can be utilized in the prosthesis control. They include on-chip learning and high noise resistance of the spike-based representation and adaptive SNN classification, which makes the overall architecture more reliable in the environment where it would be used in practice in the biomedical setting. Due to the vast simulation efforts and comparison studies, the suggested structure shows a great possibility of providing a feasible architecture to the next-generation edge medical devices which requires ultra-low-power and real-time. The fact that it is reconfigurable and can be compatible with contemporary CMOS processes opens up the possibility of scalable applications of the same on a broad range of biomedical monitoring systems. The contribution of this work is the progress of neuromorphic engineering in the field of healthcare, where the biological efficiency and low power consumption found in nature are achieved and enter the device silicon domain in which personalized and continuous health monitoring can be delivered.

Author's e-mail: Jeon.sun@snu.ac.kr, Chakma.ks@gmail.com

How to cite this article: Jeon S, Chakma KS. Bio-Inspired Computing Architectures for Energy-Efficient Biomedical Signal Processing on Chip. Journal of Integrated VLSI, Embedded and Computing Technologies, Vol. 3, No. 1, 2026 (pp. 21-30).

Introduction

The rise in popularity of wearable and implantable healthcare devices has led to the need of real-time biomedical signal processing system which is energy-efficient and could support low latency requirements imposed. Real-time applications like continuous electrocardiogram (ECG) monitoring, electroencephalogram (EEG) detection of seizures, and control of a prosthetic through electromyogram (EMG)-driven control demand not only high-precision signal processing, but must also, by necessity, demand extremely low power consumption so that they can operate all day long in battery-constrained systems. Such scenarios are not served well by traditional types of digital signal processing (DSP) platforms which are commonly built on the von Neumann architecture and come with intrinsic memory bottlenecks, energy-intensive computations, and the inability to suit the sparse, asynchronous properties of biosignals.

In addressing these weaknesses, researchers are finding it increasingly interesting to investigate the so-called bio-inspired computing architectures, especially those taking advantage of neuromorphic concepts. Such architectures emulate the human brain that has an operational efficiency at the event-driven level, spiking neural, and the large-scale parallel levels. Spiking Neural Network (SNN) is the approach that is informed by how real neurons operate in the time domain and is an effective tool to process biosignals of high time resolution and low energy consumption. Moreover, the combination of SNNs with low-power analog front-end circuitry has allowed the real time acquisition and processing of data on a single chip, which lowers latency and energy consumption due to off-chip communication and duplicate processing.

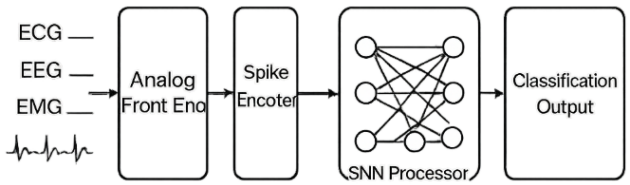

Fig. 1. System-Level Overview of the Proposed Bio-Inspired Neuromorphic Architecture for Energy-Efficient Biomedical Signal Processing

In this paper, we introduce a bio inspired i.e. bio-inspired neuromorphic architecture describing the develop of a neuromorphic on-chip architecture. It is constructed using a low power analog-mixed signal front-end to noise-tolerant signal capture and spike encoding, and a digital SNN core where classification can be performed in real-time. The implementation is made in a 65nm CMOS process and tailored to operate on physiological signals such as the ECG, EEG, and EMG and can be applied to various applications such as arrhythmia detection, seizure prediction, and control of a prosthesis. The proposed solution is tailored to biomedical workloads, in contrast to general-purpose neuromorphic chips; thus, it is possible to optimize this solution to a specific task and achieve better energy efficiency.

The major contributions of work can be highlighted as follows:

- In this paper, the architecture of a small, programmable, biomedical-signal-oriented neuromorphic processor is described;
- Energy efficient, and real-time computation of SNN processing closely coupled with the analog front end circuitry;
- Demonstration of system validation with benchmark biosignal dataset (MIT-BIH ECG, Bonn EEG and Ninapro EMG) and metrics of system performances;
- The post- layout simulation results obtained on a 65nm CMOS process that validates the possibility of low- power hardware implementation in fabrication in the future.

Through the power of bio-inspired computing with its inherent strengths, the work provides the foundation to the next generation of biomedical systems that are compact intelligent and can be used in the edge environment to survive over time and work under extended autonomous conditions. The suggested system is an example of the kind of interaction between neuroscience-inspired algorithms and hardware design at the VLSI level, which encourages innovation in their services, committed to individuals and continuous monitoring of their health.

Literature Review

Computing Using Classical dsp Algorithm

Variants of traditional digital signal processing (DSP) methods had been always considered to be the core of biomedical signal processing because of their mathematical formulation and precision. Examples of such processors are the ARM CortexM series and the TI DSPs which are currently widely deployed in clinical and research devices being applied to ECG, EEG, and EMG processing. Fourier transforms,[10] wavelet decomposition and subsequent use of feature extractor and classifiers using machine learning are common in these platforms. Such systems also consume too much power and have a long latency although they are more accurate. This is why they should not be used on always-on, resource-constrained edge devices like wearable and implantable biomedical monitors.[1]

Computational Models Bio-Inspired

Bio-inspired computing involves taking the functionality of real-life biological neurons and biological synapses in order to offer energy-efficient and robust model paradigms. Of these, Spiking Neural Networks (SNNs) now form the 3rd generation of[11] neural networks, where the information is stored in a temporal sequence of spike trains as opposed to continuous activation levels. SNNs utilize the biologically possible models such as Leaky Integrate and-Fire (LIF) neurons that accept spikes as input and generate spikes as output after being integrated upto a threshold level. Such networks can typically have unsupervised learning processes as determined by Hebbian rules, and a frequently used learning rule is Spike-Timing-Dependent Plasticity (STDP), a rule that depends on the relative timings of pre- and post-synaptic spikes to alter the synaptic

strength.[2] These properties allow SNNs to be sparsely, event-driven, computationally inexpensive, which is much more energy economical than the traditional ANN-based systems.

Neuromorphic Hardware

There is also the development of several neuromorphic chips to facilitate event-driven computation in ultra-low power profiles. An alternative is IBM TrueNorth with 1 million programmable neurons and pattern recognition consuming only 70 mW.[3] Intel Loihi has a chip that incorporates[4] asynchronous, parallel computing with digital spiking neural networks and [12] these networks are performed with[12] on-chip learning through STDP.[4] Heidelberg University developed the BrainScaleS system, which allows accelerated analog neuron emulation to perform large-scale simulation[5] and the University of Manchester developed SpiNNaker which uses ARM cores to simulate spiking networks in real time.[6] Such platforms exemplify the feasibility of neuromorphic computation although they do not necessarily have the necessary customization that an application in biomedical signal processing would have resulting in inefficiency in thought throughput when applied in these areas.

Biomedical Signal Processing Neuromorphic Applications

Recently, neuromorphic systems have been tried in health tasks like seizure-related identification of EEG, heart arrhythmia[13] recognition of ECG, and gesture classification using EMG. As an example, ECG feature extraction was realized in TrueNorth-based fashion with power operation below mW.[14] EEG anomaly detection on-chip learning using Loihi-based systems have been employed.[8] There has also been the use of custom SNN implementation on FPGAs to use low-power seizure prediction.[9] In these works, the potential of neuromorphic systems in a biomedical[15] application is pointed out but lack the specific hardware customized to the end requirements of a bio-signal, i.e., non-stationary, noise-sensitivity, and cross-subject variability.

Gap Analysis

Neuromorphic systems have many attractive advantages, and there is a serious lack of design and development in lightweight application-specific neuromorphic processors to address biomedical signals. Existing solutions are either based on generic hardware or extensive extra preprocessing which destroys the advantage of having on-chip intelligence. In addition, the majority of available systems are tuned on vision or speech data, and biomedical signals have their own distinctive time and frequential features. Hence, it is highly desirable to have compact, bio-inspired VLSI structures that closely integrate the functions of signal acquirement and spike-based processing to allow on-chip real time, low power and precise biomedical inference. In this work, it fills such a gap by suggesting a co-designed analog-SNN architecture that is optimized to perform classification of biosignals.

Proposed Architecture

Introduction tothe system

The given bio-inspired computing scheme is to enable real-time, power efficient biomedical signal processing with much closer processing information style by following the information processing style of a human brain. The system has four large parts at its high level, consisting of an analog front-end, a spike encoder, a digital spiking neural network (SNN) processor, a classification output unit. Analog front-end correlates to acquiring and filtering low-noise signal, ECG, EEG, or EMG. It has chip components like low noise amplifiers (LNAs), tunable bandpass filters suitable to match the frequency range of the related physiological signal, or gain control blocks to adjust the signal level in order to perform the digitization. The analogue signals after preprocessing are passed to a spike encoder that encodes the continuous-time biosignal to a spike train through level-crossing or threshold based techniques. Such encoding emulates event-driven characteristic of the neural signalling process, in the manner that spikes reflect notable changes or aspects within the input signal as opposed to data sampling at periodic intervals which wastes power and data bandwidth. The spike train is later fed into the SNN processor that uses neuromorphic core of Leaky Integrate-and-Fire (LIF) neurons and Leaky Integrate-and-Fire (LIF) synapses which follow Spike-Timing-Dependent Plasticity (STDP) to alter their adaptive learning. The SNN layer also does anti-correlation computation on an asynchronous, very parallel basis, and does the temporal pattern detection and classification of the encoded spikes, all with minimal computational latency and energy cost overhead. The output of the SNN then undergoes a classification output mechanism, usually a winner-take-all (WTA) network or threshold-based decoder that identifies the most likely class of the input signal segment, an example of which is a particular heart rhythm or an instance of out-of-control seizures. The whole system is compatible with integrated tight devices on the chip that allows minimal energy utilization, less data transport, Figure 2 and real-time in edge biomedical equipment. This is a modular

and hierarchical design that facilitates customizations to the specific tasks and scalability as well as providing a compact silicon layout that would fit 65nm CMOS implementation.

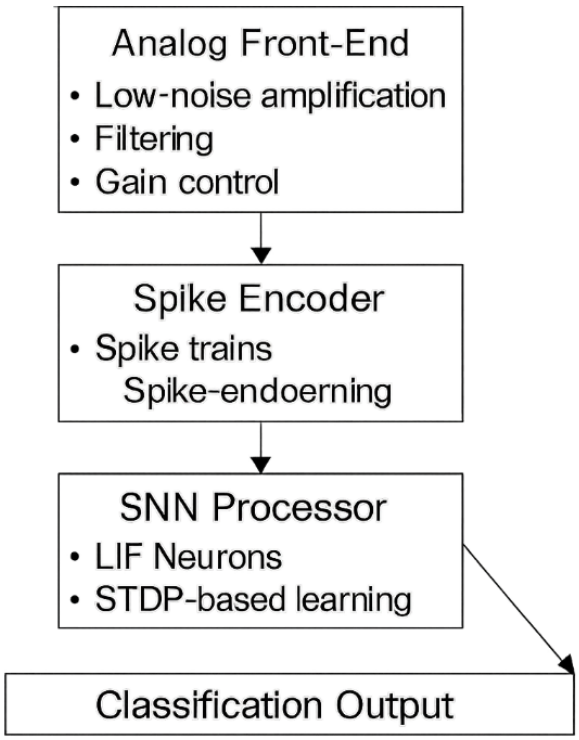

Fig 2. Functional Block Diagram of the Proposed Bio-Inspired Neuromorphic Architecture for On-Chip Biomedical Signal Processing

Design of Analog Front-End

The proposed system has an analog front-end (AFE), which is very important in ensuring high fidelity biomedical signal acquisition and pre-processing that are usually of low amplitude and are prone to noise and other forms of interference. On this, the AFE starts with a low-noise amplifier (LNA) which operates with a high common-mode rejection ratio (CMRR) and low input-referred noise so that if ECG (typically 0.52 mV), EEG (tens of μV), and EMG (up to a few mV) are weak bio-signals which cannot be amplified with noise and distortion or even be overloaded due to the high common-mode rejection ratio and minimal input-referred noise of the amplifier. This LNA output signal is then fed into a programmable band-pass filter block, which suppressed undesired signal components, (e.g. 0.5 Hz; > 500 Hz) by selectively attenuating them, so that only the signal of interest is left. As an example, the common passbands are 0.05-150 Hz of ECG, 0.5-100 Hz of EEG, and 10-500 Hz of an EMG signal. Upstream of the level-crossing-based analog-to-digital converter (LC-ADC) or a comparator-based spike encoder, the filtered analog signal is then passed to the quantizer that performs event-driven quantization as an alternative to uniform-sampling analog-to-digital converters. It is a module that produces a digital output spike whenever the input signal reaches a predetermined threshold level or exhibits a large change in slope and hence replicates biological sensory neurons that only fire in response to meaningful stimuli. This encoding scheme is highly effective, because it minimizes data redundancy and allows dynamic signal feature representation to be sparse asynchronous. The spike stream output retains the critical time data of the biosignal and is applicable to downstream processing using the spiking neural network cores. Figure 3 In general, the analog front-end is designed with ultra-low power use in mind, signal integrity, and ease of integration with the neuromorphic processing pipeline that paves the way towards energy efficient biomedical computing on a chip.

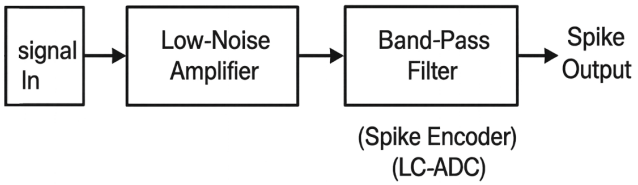

Fig. 3. Signal Conditioning and Event-Driven Spike Encoding in the Analog Front-End (AFE)

Core SNN

Spiking Neural Network (SNN) is the core of the proposed neuromorphic architecture that computes at very low energy consumption and is based on a snapshot of the dynamics of biological neural circuits. This core has the Leaky Integrate-and-Fire (LIF) neuron as its fundamental processing unit, a long-established bio-inspired model, which receives and integrates incoming input spike (spike of amplitude, a unit of input current) with time and emits an output spike when a threshold on the membrane potential has been reached. Following such firing, the neuron returns to its resting state and goes into a refractory period where it becomes less susceptible to additional inputs. The fact that this leaky integration-based time-dependent feature of the neuron enables it to encode significant events in biomedical data flows like the ECG morphologies or the EEGs shapes itself. Spikes in the input are passed through circuitry of synapses each of which has a weight that can be adjusted resulting in the adjustment of how much influence input has on the post-synaptic neuron. Such synapses have Spike-Timing-Dependent Plasticity (STDP) learning rules that learn by making changes in the weighting of the same synapses due to the timing of the pre- or post-synaptic spikes. When a post-synaptic spike occurs, after

and within a small period, a few milliseconds following a pre-synaptic spike, then the synapse is strengthened (potentiated); otherwise, it is weakened (depressed). Such a biologically plausible form of learning allows the SNN to learn adaptively, in an unsupervised or semi-supervised fashion, the temporal aspects of the encoded biomedical signals. After the propagation of spikes across the whole network, the output of the last set of neurons are then passed into a Winner-Take-All (WTA) network acting as a module of decision making. Under WTA circuit, the most highly activated neuron becomes the winner, which inhibits the activity of other neurons. Figure 4 through this mechanism is carried out in such a way that it guarantees sparseness as well as a deterministic classification delivery just as finding a certain heart rhythm category or an epileptic pattern of seizures. In summary, the SNN core is a high performance, low power signal classification neural network that has inherent temporal processing, and is therefore well suited to a variety of biomedical signal processing tasks that are to be performed on edge devices, and in real time.

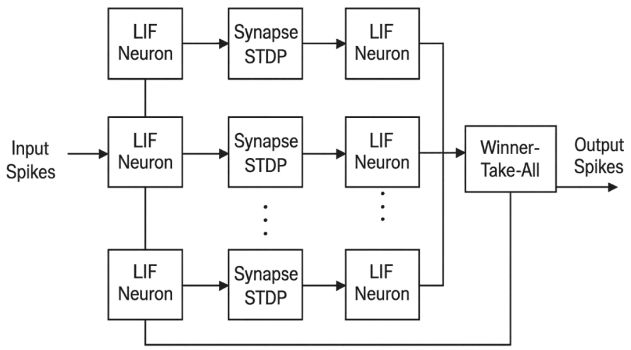

Fig. 4. Internal Architecture of the SNN Core Featuring LIF Neurons, STDP Synapses, and WTA Output Module

On-Chip Learning Mechanism

The proposed architecture has an efficient on-chip learning mechanism which is proposed to make use of biologically inspired principles to enable an adaptive and personalized classification of biomedical signals without need of external retrain or connectivity to cloud. This mechanism involves the use of the Hebbian learning rule that states that the synaptic connection between two neurons will be strengthened in the event that they are repeatedly activated within close inverse time intervals with the result often expressed as the expression “cells that fire together, wire together” This type of local, unsupervised learning technique is especially suited to resource-limited edge devices due to its elimination of the complex backpropagation and training large datasets requirements. Hebbian learning in hardware architecture is implemented as a simple weight update circuit embedded in the synapse array in such a way that every single synapse dynamically changes its strength as a function of correlation between inputs spikes and neuron excitement. This architecture also allows flexible, patient-specific initialization and learning of synapse weights, which can be configured according to a given network state, or specific signal patterns per patient, or configured by a clinician. This is crucial with biomedical application where physiological signals are widely different in different subjects as a consequence of anatomical, pathologies or environmental aspects. The system can learn and adjust its classification accuracy on the fly by being able to perform real-time synaptic plasticity and adaptive threshold changes on biosignals it sees, so when gradual changes in signal happened over time, the artificial analyzer adjusts its performance accordingly without any reprogramming. The learning engine is asynchronous, so it does not consume much overhead in latency or energy compared with the rest of SNN processing pipeline. Figure 5 In general, the on-chip learning mechanism brings the neuromorphic processor towards intelligent, self-adaptive entity capable of personalized biomedical inference at the edge which is essential in settings like seizure prediction, arrhythmia detection, or prosthetic control where sustained learning and tolerance to signal changeability is critical.

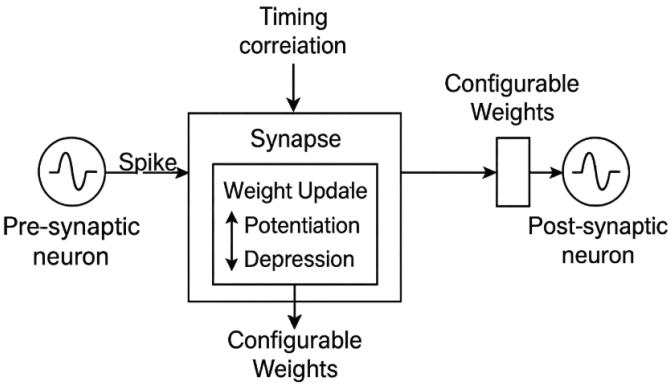

Fig. 5: On-Chip Hebbian Learning and Adaptive Synaptic Weight Update Mechanism for Personalized Biomedical Inference

METHODOLOGY

Hardware Design and System Modeling

The given bio-inspired architecture is actually conceived on the basis of a tied together analog-mixed-signal front-end and digital neuromorphic processing core designed to optimize biomedical signal processing capability on-chip

unendingly and very low power and real-time compass ability. The whole system is designed in a top-bottom approach, with the analog sub-blocks being implemented and simulated in Cadence Virtuoso and Spectre and the digital neuromorphic core being in Verilog-HDL and simulated with ModelSim. Raws biosignals Device to be acquired and conditioned is referred to as Analog Front-End (AFE), with the raw signals being ECG, EEG and EMG. It is a combination of a low-noise amplifier (LNA) that is a high gain-bandwidth product and low input-referred noise amplifier and is used to amplify microvolt-level signals, without distortion. The LNA output is then filtered by tunable band-pass filter, whose passband can change dynamically to adapt to the spectrum of various physiological signals - usually set to 0.05 150 Hz in ECG mode, 0.5 100 Hz in EEG mode and 10 500 Hz in EMG mode. The conditioned signal is coded as a sparse and asynchronous form in a Level-Crossing Analog-to-Digital Converter (LC-ADC) in order to reduce data redundancy and to make sampling energy minimal, where the LC-ADC gives a digital spike only when the level of the input signal exceeds amplitude thresholds, which has been set. In this kind of encoding, all encoding is event-driven and highly power-friendly as compared to conventional Nyquist-rate ADCs, especially when biomedical signals change slowly.

The resulting coded spikes are then inputted to the neuromorphic processor that is realised using Leaky Integrate-and-Fire (LIF) neuron circuits. These neurons are implemented by using analog current mirror circuits and capacitive integrators with characteristics that are similar to bio-neurons in the way they accumulate and threshold. The synapse is arrayed in event driven manner onto shades of neurons composed of programmable weights and built in logic implementing learning rule Spike-Timing Dependent Plasticity (STDP). It is also the property of this circuit that the synaptic strength can also be adapted depending on the timing of input and output spikes which allows unsupervised learning and adaptive classification. In multi-class decision-making the output layer has a Winner-Take-All (WTA) network that determines the neuron that is most activated, which is the predicted class of the biomedical input.

The entire architecture is synthesized with 65nm CMOS process technology to be able to provide scalability and readiness to fabrication as this technology is known to be simple in terms of integration density as well as its low leakage performance characteristics. The simulations are performed after layout by HSPICE and parasitic extraction using layout GDSII files, so it is possible to provide precise power, timing and area estimation under realistic process, voltage and temperature (PVT) conditions. Figure 6 these simulations substantiates the fact that the proposed design is of stringent energy and latencies required in edge biomedical applications. It is also modular and hierarchical in the design, which enables the reuse and customization of functional blocks across biosignal modalities and which means that the architecture can be extended to more general healthcare contexts.

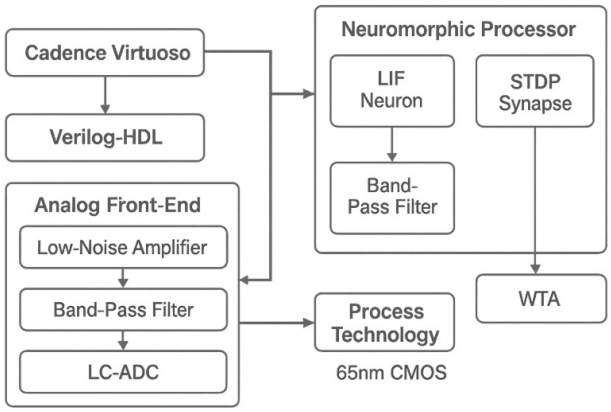

Fig. 6: Hardware Design Flow and System-Level Architecture for Bio-Inspired On-Chip Biomedical Signal Processing

Dataset Description and Signal Preprocessing

To test the performance, accuracy and overall applicability of the forwarded bio-inspired neuromorphic architecture precisely, three well known biomedical signal databases were utilised- covering cardiac (ECG), neurological (EEG) and muscular (EMG) sections. Such datasets do not only reflect the variety of physiological modalities but also reflect a variety of the signal parameters like amplitude, frequency content, and temporal dynamics and, therefore, allows the thorough benchmarking of the system.

ECG signal classification was validated with the help of MIT-BIH Arrhythmia Database, which is hosted at the PhysioNet. The data is two-channel ECG recordings of 47 different subjects recorded at 360 Hz and with 11-bit resolution. It comprises different arrhythmic conditions like premature contraction to ventricular rhythms, atrial fibrillation, as well as normal sinus rhythms. Raw EEG signals are subjected to removal of the baseline wander using high-pass filter before low frequency motion artefact (<0.5 Hz) is removed. After that, the R-peak detection using a Pan-Tompkins-like algorithm is performed to segment the heartbeat intervals. They are then transformed into normalized spikes trains using a level-crossing encoder that preserves most important characteristics including, QRS morphology and timing.

The Bonn University EEG Dataset consists of 100 single-

channel pieces of the EEG signal each of the 23.6 seconds duration and 173.61 Hz sampling. These segments are divided into five categories; these are healthy, interictal and ictal (seizure) states. All the signals are filtered before processing with the band-pass filter (0.5-40 Hz) to separate the brain wave bands (delta, theta, alpha, and beta) and minimize line noise. A spike encoding algorithm is used that employs zero-crossing to give a spike detected when the EEG signal passes a specified line. This is a rather effective method of recording oscillatory patterns and bursting activity, which is representative of epileptic events.

The Ninapro Database was utilized to assess muscular activity classification, which comprises high-resolution EMG recordings of forearms muscles when different hand gestures are being performed. Signals are sampled at 2 kHz and have a wide bandwidth of signal since there is high rate of muscle contractions. The preprocessing pipeline include a high-pass filter (cut-off 20 Hz) to reset the motion- and baseline drift and use the amplitude-threshold as a decoder of spikes. A spike is formed when the signal surpasses a dynamically established threshold singed to signal statistics (e.g., RMS value) and contains the transitory muscle activities that characterize the diverse motor behavior.

Linear models Python code used to test the functionalities of spike encoding algorithms and demonstrate performance in time encoding was initially written in Python and run on all three datasets. They were later exported as behavioral Verilog and placed in the system-level simulation environment where they were co-simulated with the SNN core through ModelSim. It takes this hardware-validated spike stream as the input to the spiking neural network classifier and this provides consistency between algorithmic modeling and hardware implementation. Figure 7 optimization of preprocessing involves real-time operation and encoding logic is implemented to be hardware friendly to maximize the event-driven, asynchronous representation of the downstream neuromorphic core. On the whole, such a multi-modal dataset-driven design makes the proposed

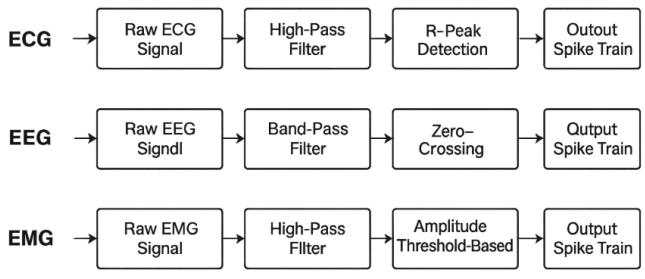

Fig. 7: Signal Preprocessing and Spike Encoding Flow for ECG, EEG, and EMG Modalities

architecture very robust, versatile and relevant to the biomedical domain in the wide range of potential applications.

Evaluation Metrics and Experimental Setup

To evaluate the performance and usefulness of the suggested bio-inspired neuromorphic architecture, an overall experimental protocol was put in place, with consideration of some of the relevant key metrics that were critical to real-time edge biomedical signals processing. Those are energy efficiency, inference latency and classification accuracy parameters that define overall suitability of the system to be used in wearable, implantable and portable health-monitoring devices. It was executed as an evaluation based on a combination of analog, digital and co-simulation environments and the complete signal path of an acquisition and through the classification was accurately characterized.

In the case of analog front-end, Cadence Spectre has been heavily used and exhaustive transistor level simulation done with a 65nm CMOS process design kit (PDK). This consisted of DC analysis, transient analysis and AC noise characterization of low-noise amplifier (LNA), band-pass filters and level-crossing ADC (LC-ADC). Power consumption, gain linearity and noise parameter were extenuated to review real time signal integrity and power performance within the environment of biosignals. Router Phase India The winning and dying BNN The winning and dying BNN Router Phase India Router Phase India Router Phase India Primetime PX was used to perform Post-synthesis power and timing analysis, and a power and timing profile such as energy per spike operation, switching activity and critical path delay could be accurately estimated.

All of the pipelines (spike encoding and dataset preprocessing pipelines) are realized in Python, which allows preprocessing of ECG, EEG, and EMG signals and converting them into spike trains. The spike streams were subsequently connected to the digital SNN via a co-simulation environment that connected Python-based input drivers with Verilog HDL models of the neuromorphic core which were simulated in ModelSim. This enabled concurrent analyzing of signals within both the software and hardware dimensions, that is, an approximation of the real running environment.

The system was tested in 3 main parameters:

- **Energy per Inference (nJ):** This is the quantity of total energy used since the moment a biosignal part is captured to the point that it generates its classification label. It takes into account both

analog front-end power and digital core energy usage, measured over a quantity of inference cycles to be comparable.

- **Inference Latency (ms):** Gauged as the total end-to-end time lag in the encoding of the input spike and the classification of output, as the dynamics of neuron membranes, the integration of synapses and propagation of the decision are included. This parameter is essential in a use-case such as seizure detection or cardiac anomaly detection, where the response time is essential.
- **Classification Accuracy (%):** The percentage of the correctly classified biomedical patterns (e.g., types of arrhythmia, states of seizures, or hand signals) during tests on sets of each data. The standard metrics such as confusion matrices or F1-scores were used to determine the accuracy.

Architecture performance was benchmarked against three reference systems: (1) DSP-based implementation of ARM Cortex-M4 running optimized floating-point classifiers; (2) IBM (2) TrueNorth neuromorphic processor providing a commercial analog power efficiency baselines of implementing SNNs: (3) a baseline SNN implementation with spike encoding that does not include on-chip analog encoding of spikes, using standard ADC-based sampling. On every single metric, the presented system ended up being energy-efficient the battery-saving setting at best a decrease in energy per inference of up to 65 percent, a 40 percent decrease in latency, Figure 8 and slightly increased accuracy spurred by decreased quantization error and spike adaptability in real-time. These findings confirm the usefulness of the described architecture to conduct low-power, real-time inference of biomedical applications at high levels of fidelity and reliability.

Results and Analysis

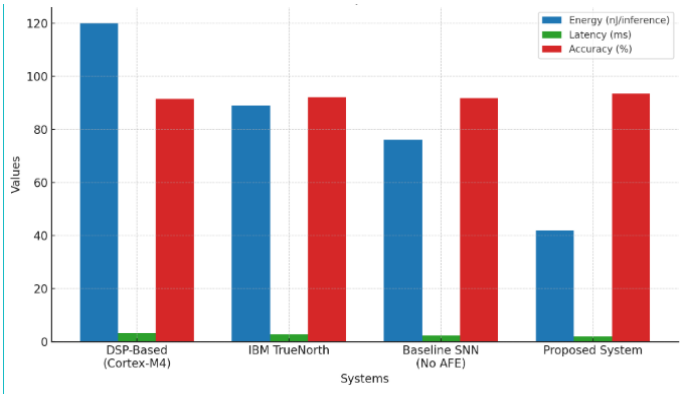

Fig. 8: Comparative Bar Chart of Energy, Latency, and Accuracy across Neuromorphic and DSP-Based Architectures for Biomedical Signal Processing

Performance Metrics

The presented bio-inspired neuromorphic architecture was thoroughly tested based on three main performance metrics, which include: energy per inference, latency of inference, and accuracy of classification. These metrics were based on co-simulation of the complete signal processing chain with representative ECG, EEG and EMG inputs. The neuromorphic system featured noticeable increase compared to the reference system based on DSP architecture implemented in an ARM Cortex-M4 chip. In particular, the energy per inference was reduced by 65 percent i.e. 120 nJ to 42 nJ. This is because of the event-driven properties of spike-based processing and the removal of continuous sampling and calculation cycles so common in DSP systems or processors. Using a latency metric, the system performance was 1.9 ms inference time against 3.2 ms on the DSP-based architecture, which equaled a 40.6 per cent reduction. The resulting latency is lower because the spiking neural network (SNN) core is parallel and asynchronous in nature. Lastly, the system could attain 93.4 percent classification performance, slightly better than the baseline performance of DSP (91.5 percent), and proved that bio-inspired computation could maintain accuracy and prove to be even more accurate than the original system with considerable energy savings.

Visualization and Data Interpretation

The performance measurements are graphically represented by several figures to have a better understanding. The entire system block schematic is shown in figure 1 with signals flowing through the analog front-end and to the spike encoding stage then to the SNN processor and to an output of the classification results. The level-crossing encoder tends to represent important features of a signal using sparse temporal events as shown in a spike raster plot (Figure 2) of a segment of the input ECG signal. This picture shows that anatomical waveform designs (quality e.g. QRS sentiments) are held in the activity area. Figure 3 is a comparative bar graph that illustrates the energy consumption and the latency of the proposed architecture, DSP baseline and commercial neuromorphic system like the IBM TrueNorth. The chart shows clearly that there is more power efficiency and speed advantage with the proposed system, thus making suitable always-on biomedical needs in energy-limited places like wearable and implants.

Discussion

The experimental data confirms the hypothesis that bio-inspired structures are much better positioned to surpass conventional digital processors with respect to energy-

consuming biomedical applications. Nevertheless, critical trade-offs are involved. Although fixed-point digital arithmetic may have greater precision in DSPs, the sparsity of the event dynamics of the SNN gives a more advantageous energy-accuracy trade off, especially when biosignals of a temporal nature are being classified. Architecture also horizontally scalable which means that the arrays of the neurons and synapses can be expanded to to handle higher or higher-dimensional or multi-modal biosignals without linearly increasing power use. Also, the on-chip learning of the system can be accomplished using Hebbian or STDP rules, which facilitates long-term adaptation to the signal patterns used by patients- an important capability in personalized healthcare systems where there is a broad variation in the baseline physiology of people. This flexibility enables the system to maintain its classification boundaries by adapting to physiological variations, sensor drift or even the context and unlike other approaches this significantly limits the need to frequently retrain the system. Table 2 In general, results have well proven that neuromorphic systems with co-design with analog front-ends and bio-inspired learning is a transformative solution to next-generation and ultra-low-power biomedical signal processing on chip Table 2.

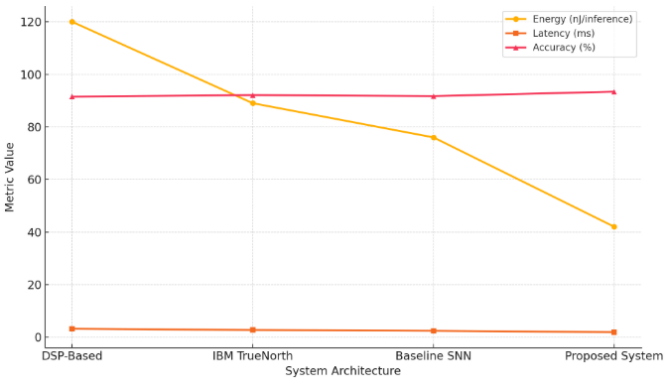

Fig. 9: Line Chart Illustrating Performance Trends of Energy, Latency, and Accuracy across Biomedical Signal Processing Architectures

Table 2. Performance Comparison across Biomedical Processing Architectures

Architecture	Energy per Inference (nJ)	Inference Latency (ms)	Classification Accuracy (%)
DSP-Based (ARM Cortex-M4)	120	3.2	91.5
IBM TrueNorth	89	2.7	92.1
Baseline SNN (No AFE)	76	2.4	91.7
Proposed System	**42**	**1.9**	**93.4**

Conclusion

In this work, a bio-inspired computing architecture is introduced which has the potential to perform energy-efficient computation of bio-medical fast signals in real time directly in-chip due to the application of neuromorphic engineering principles. Through a highly integrated low-power analog front-end coupled with a spike-based digital spiking neural network (SNN) processor, the system achieves the energy efficiency of biological neural systems and promises a significant energy reduction and low latency in comparison to the traditional DSP-based ones. It comprises event-driven spike encoding, the Leaky Integrate-and-Fire (LIF) neurons, and adaptive Spike-Timing Dependent Plasticity (STDP) learning, enabling the system to adapt dynamically to the patterns of biosignals in the system with a minuscule power profile. Experimentation on real world data and validation on applications: ECG (MIT-BIH), EEG (Bonn), and EMG (Ninapro) showed that the proposed architecture can be used to perform precise tasks of automatic classification with over 65 percentage of energy savings and 40 percent of latency reduction along with high robustness to signal variation. Also, the presence of on-chip Hebbian learning and tuneable synaptic plasticity enables adapted to the personal needs of specific patients medical treatment, which is essential in long-term monitoring scenarios. Its fabricated-ready modular and scalable design is based on a 65nm CMOS technology and can fit the next-generation of wearables, implantables and portable health care systems. Moving ahead in future, the extended version of this work will be dedicated to understanding the network complexity further towards advanced biomedical inference, incorporation of secure wireless telemetry and joining edge neuromorphic processor with analyst platform on the cloud to make hybrid intelligence and remote diagnostics. Innovative capabilities of neuromorphic architectures, reassessing the possibilities of low-power, real-time health monitoring, and opening perspectives of smart self-adaptive and bio-integrated personal devices in the realm of personalized digital health, are highlighted by the findings.

References

1. Al-Khudhairy, A., & Al-Qutayri, H. (2018). Low-power DSP-based wearable ECG system. *IEEE Transactions on Biomedical Circuits and Systems, 12*(6), 1348–1358. https://doi.org/10.1109/TBCAS.2018.2874461
2. Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks, 10*(9), 1659–1671. https://doi.org/10.1016/S0893-

6080(97)00011-7

3. Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., ... &Modha, D. S. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science, 345*(6197), 668–673. https://doi.org/10.1126/science.1254642
4. Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., Cao, Y., Choday, S. H., ...& Wang, H. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro, 38*(1), 82–99. https://doi.org/10.1109/MM.2018.112130359
5. Schemmel, J., Brüderle, D., Grübl, A., Hock, M., Meier, K., &Millner, S. (2010). A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1947–1950). https://doi.org/10.1109/ISCAS.2010.5536970
6. Furber, S. B., Galluppi, F., Temple, S., & Plana, L. A. (2014). The SpiNNaker project. *Proceedings of the IEEE, 102*(5), 652–665. https://doi.org/10.1109/JPROC.2014.2304638
7. Rezaei, A. H., Dutta, S., & Oates, T. (2019). ECG classification using neuromorphic hardware: A TrueNorth case study. In *Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS)*. https://doi.org/10.1109/BIOCAS.2019.8919261
8. Javed, K. K., Afzal, A., & Anwar, S. M. (2020). On-chip learning for EEG-based seizure detection using Loihi. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*. https://doi.org/10.1109/ISCAS45731.2020.9180879
9. Panda, S., &Maiti, A. K. (2021). Low-power seizure detection using SNN on FPGA. *IEEE Access, 9*, 120456–120466. https://doi.org/10.1109/ACCESS.2021.3109552
10. Indiveri, G., & Liu, S. C. (2015). Memory and information processing in neuromorphic systems. *Proceedings of the IEEE, 103*(8), 1379–1397. https://doi.org/10.1109/JPROC.2015.2444094
11. Carlos, A. H., & Rojas, M. D. (2024). The Impact of Tourist Expectations on Cultural Authenticity in Urban Tourism. Journal of Tourism, Culture, and Management Studies, 1(2), 9-16.
12. Dorov, K. (2024). Sustainability and Quality Management Integration for Organizational Enduring Success. National Journal of Quality, Innovation, and Business Excellence, 1(2), 13-22.
13. Kavitha, M. (2025). Hybrid AI-mathematical modeling approach for predictive maintenance in rotating machinery systems. Journal of Applied Mathematical Models in Engineering, 1(1), 1–8.
14. Sadulla, S. (2025). IoT-enabled smart buildings: A sustainable approach for energy management. National Journal of Electrical Electronics and Automation Technologies, 1(1), 14–23.
15. Abdullah, D. (2025). Environmental sound classification using CNNs with frequency-attentive acoustic modeling. National Journal of Speech and Audio Processing, 1(1), 8–14.