A Brain-Inspired Spike Encoder and SNN SoC for Continuous Cognitive State Monitoring

Kaushik Lakshmiramanan, Duncan Millar, Sierra Raspa, Yihui Wang, Jeremy Yun
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD, USA.

***Abstract*—Sleep quality, attentional state, and real-time cognitive assessment are increasingly central to both clinical research and consumer health technologies. Conventional methods for monitoring these parameters often lack the temporal resolution, accuracy, or adaptability required for real-time applications. Eye movements, specifically blink duration and intensity, provide a promising, non-invasive biosignal for continuous tracking of cognitive states. In this work, we present a neuromorphic System-on-Chip (SoC) designed for real-time, low-power decoding of blink-based electroencephalogram (EEG) biosignals. Designed using the open-source 1.8V SkyWater 130nm CMOS process, the system integrates analog front-end circuitry with delta-modulation-based spike encoding and on-chip classification via a Spiking Neural Network (SNN). Our design supports applications in sleep and attention monitoring, cognitive workload analysis, and potential neuromodulation therapies for neurodegenerative and mental health disorders. The platform operates with high computational efficiency, making it ideal for future scaling into wearable and embedded systems. System-level validation was conducted using full-custom layout, simulation with real EEG blink data, and verification via the Cadence Design Suite. The final chip occupies 0.742 mm^2 of a 1.6 mm $\times$ 1.6 mm die, consisting of 276,542 transistors, and consumes 63.5 μW in the analog domain and 43 μW in the digital domain.**

I. Introduction

Real-time assessment of sleep and attentional states is critical across a range of domains, including clinical diagnostics, transportation safety, education, and human-computer interaction. Traditional monitoring techniques, such as actigraphy, behavioral observation, or standard EEG systems, face limitations in spatial and temporal resolution, long-term viability, power efficiency, or portability. To address these challenges, recent research has explored noninvasive biosignal processing techniques that leverage ocular indicators, such as blink duration and strength, as proxies for cognitive state [1]. These physiological signals offer a rich, underutilized source of information for understanding arousal, fatigue, and attentional engagement [2].

This work presents a neuromorphic SoC that combines biologically-inspired signal processing with efficient, embedded computation for cognitive state decoding. The proposed system captures EEG signals noninvasively through scalp electrodes and employs an on-chip mixed-signal architecture to condition and convert these signals into sparse, event-driven spike trains. These spikes are processed in real time by a dedicated Spiking Neural Network (SNN), enabling classification of blink-related features with minimal energy and area overhead.

Fabricated using the open-source SkyWater 130nm CMOS process, the custom chip includes a low-power analog front-end, delta-modulation spike encoders, and configurable digital SNN blocks. The system extends the functionality of existing open-source neural interfacing platforms, such as OpenBCI, by offering full hardware integration optimized for wearable deployment. Furthermore, the modular architecture of the platform makes it adaptable for future extensions, including neuromodulation through biological stimulation and closed-loop neurofeedback training.

The remainder of this work details the system-level overview, technical design analysis, SNN training methodology, and experimental validation, culminating in a comprehensive performance analysis of a proof-of-concept SoC suitable for next-generation real-time brain-machine interfaces.

II. System Overview

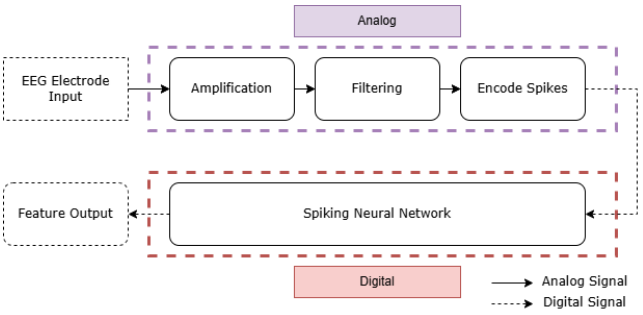


Fig. 1: A simple flowchart of the overall system and signal flow.

The project consists of the following key components: signal acquisition via electrodes, analog front-end processing, event-driven spike encoding, and SNN classification.

A. Signal Acquisition Electrodes

Our system is designed to detect eyeblink-related scalp potentials by utilizing either the FP1 or FP2 electrode placements, as defined by the international 10-20 system (see Fig. 2). The FP1 and FP2 electrodes are situated on the frontal region of the scalp, just above the left and right eyebrows, respectively. These positions are well-suited for capturing ocular artifacts, as the natural dipole formed by the cornea (positive) and retina (negative) produces a distinct potential shift during eye blinks. Specifically, blink events manifest as sharp negative amplitude deflections at these frontal sites

[3]. While FP7 and FP8 are more appropriate for monitoring lateral eye movements, the focus of this work is limited to the extraction of blink features from FP1 and FP2 signals. To enhance signal fidelity, the A1 electrode is employed as a reference, which aids in suppressing common-mode noise and improving overall noise immunity. This electrode configuration is conducive to integration into compact, wearable platforms such as smart glasses or headbands.

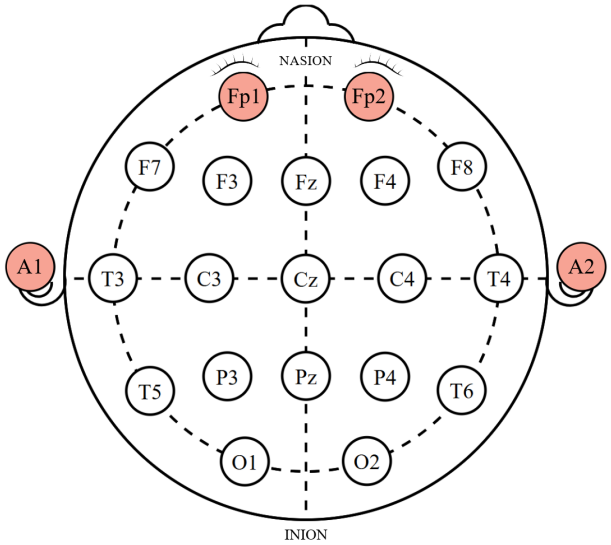


Fig. 2: Electrode configuration used for blink signal acquisition, capturing potentials between Fp1 and A1 or Fp2 and A2 according to the standard 10–20 EEG placement system.

Although blink signals originate primarily from muscular activity and are thus more accurately characterized by electromyography (EMG), they are used here more as a proof-of-concept for our system. Importantly, the circuit architecture is designed with flexibility in mind, allowing it to be repurposed for broader EEG applications, particularly those targeting neuronal local field potentials (LFPs) in the 33-125 Hz frequency band.

Signal acquisition is optimized using dry electrodes embedded in a spring-loaded guideway to minimize the formation of air gaps, which can act as antennas and introduce signal artifacts. Inevitably, electromagnetic interference from power lines (50/60 Hz) poses a challenge due to capacitive coupling between the skin and ambient electric fields from the grid. While wet electrodes with conductive gel offer superior noise suppression, they compromise user convenience and reusability, both of which are key requirements for consumer-grade wearable devices. By contrast, dry electrodes combined with effective analog signal conditioning present a balanced solution, enabling improved signal quality while preserving usability and integration feasibility into consumer wearables [4], [5].

B. Analog Front-End (AFE) Processing

Depending on the electrodes used, eyeblink signals typically exhibit amplitudes on the order of 500 μV peak-to-peak, requiring a high-gain neural preamplifier to amplify these low-level signals while suppressing extraneous noise. To ensure that only the relevant spectral components are preserved, the amplified signal is bandlimited through carefully selected cutoff frequencies. Since blink-related activity predominantly resides in the 1 Hz to 13 Hz range, a high-pass filter with a cutoff slightly below 1 Hz is implemented for AC coupling at the input stage. This effectively mitigates baseline drift resulting from slow-changing subcutaneous scalp potentials and motion artifacts. A complementary low-pass filter is incorporated to suppress high-frequency noise sources above 400 Hz, including electrocorticographic (ECoG) interference, post-amplification artifacts, static skin potentials, and ambient electromagnetic disturbances.

To address power line interference, which overlaps with the target signal bandwidth, a second-order notch filter centered at 60 Hz is applied. This is essential for attenuating noise originating from capacitive coupling with the power grid, thereby preserving the integrity of the signal within the passband and maintaining linear gain across the operating frequency range. Collectively, these analog signal conditioning techniques significantly enhance the signal-to-noise ratio (SNR), enabling accurate and robust feature extraction through spike-based encoding.

C. Event-Driven Spike Encoding

Conventional bio-potential readout circuits rely on synchronous analog-to-digital conversion, wherein continuous analog signals are quantized according to the resolution of the Analog-to-Digital Converter (ADC). However, the growing demand for higher data bandwidth, driven by the proliferation and ever-rapid scaling of multichannel electrode arrays, has introduced significant challenges in the efficient compression and transmission of neural data. Traditional ADC architectures are often inefficient when processing sparse and low-periodic signals, such as eyeblinks, where a substantial portion of samples represent idle or redundant states. This results in suboptimal latent power and computational efficiency.

To address these limitations, event-driven encoding techniques, such as absolute thresholding or peak & trough detection, have been employed in neural electrocorticography (ECoG) systems [6], [7]. These methods allow asynchronous detection of discrete events, thereby reducing unnecessary sampling. However, as illustrated in Fig. 3, absolute thresholding fails to retain sufficient signal detail for low-frequency signals, such as EEG or EMG blink events. Moreover, such schemes are more sensitive to input noise and baseline drift, which further limits their applicability in low-amplitude, non-stationary signal environments.

An alternative and more information-rich approach is event-based delta modulation, which performs asynchronous analog-to-digital conversion through level-crossing rate encoding. This method preserves relevant signal features while achieving significant data compression, drawing inspiration from biological systems. A prominent example of this technique is the Dynamic Vision Sensor (DVS) [8], which has demonstrated

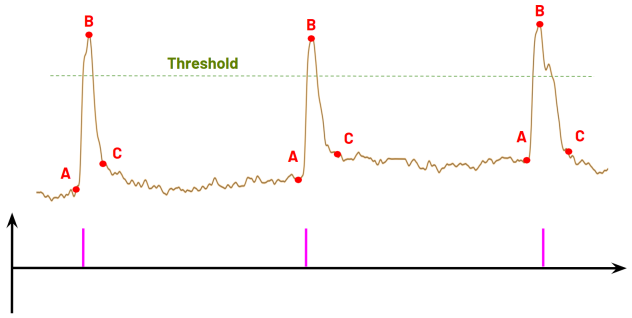


Fig. 3: EEG eyeblink waveform (inverted) illustrating the limitations of absolute thresholding in preserving signal features necessary for reliable blink classification.

its utility in encoding sparse analog signals. As shown in Fig. 4, the delta modulation circuit produces two independent spike trains representing ON and OFF events. These spikes are generated when the input signal crosses a predefined threshold in the positive (ON) or negative (OFF) direction.

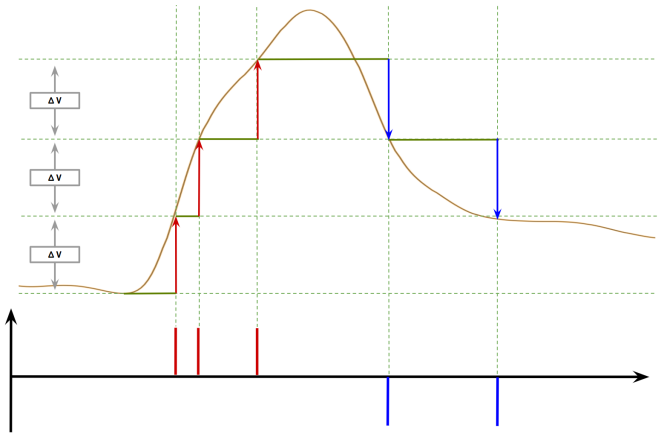


Fig. 4: Delta modulation applied to a single eyeblink waveform. The proposed SoC employs this continuous-time, level-crossing sampling technique to enable efficient, event-driven encoding into ON (red) and OFF (blue) spiking outputs.

Leveraging these advantages, the proposed system converts preconditioned analog signals into spike trains using an asynchronous delta modulation encoder. These spike-encoded signals are then input directly into a SNN for downstream processing and classification.

D. Spiking Neural Network (SNN) Classification

Spiking Neural Networks (SNNs) differ fundamentally from conventional Artificial Neural Networks (ANNs) by encoding information as discrete spike events rather than continuous activations. This temporal, event-driven representation affords SNNs several advantages, including sparse computation, inherent temporal dynamics, and energy efficiency. These properties make SNNs well-suited for low-power, real-time neuromorphic applications such as wearable or implantable systems.

In this system, the SNN module is implemented as a fully digital circuit tasked with detecting and classifying eye blinks based on spike patterns generated by the asynchronous spike encoder. The primary goal is to distinguish between soft and hard blinks so that these features can be used off-chip for estimating the user's cognitive load. A digital SNN was chosen over its analog counterpart due to its ease of scalability, which allows for expansion in the number of input channels or network depth in future work, as well as its compatibility with standard digital design flows.

The SNN architecture used is a 2–4–4–2 topology: two input neurons, two hidden layers with four neurons each, and two output neurons. Neuron states are represented using 16-bit signed fixed-point values to balance computational efficiency with numerical precision. The design must balance accuracy, hardware cost, and real-time processing capability, leading to careful consideration of the neuron model to be used.

1) Hodgkin–Huxley (HH) Model: The Hodgkin–Huxley model is a biophysically detailed model that describes how action potentials are initiated and propagated through voltage-gated sodium and potassium channels in the neuron membrane [9]. It uses a system of nonlinear differential equations to model ion channel dynamics, yielding highly accurate representations of neuronal behavior. However, this accuracy comes at a high computational cost, requiring several differential equations and parameters per neuron. Implementing HH models in hardware typically involves a large number of transistors per channel, making it infeasible for real-time, large-scale deployment on resource-constrained digital systems. Given the power and latency constraints of the SoC, the HH model is not suitable for this application.

2) Leaky Integrate-and-Fire (LIF) Model: Discovered in the early 20th century, the LIF model comprises one of the simplest spiking neuron models, treating the neuron as a passive RC circuit that integrates incoming current over time [10]. Once the membrane potential reaches a predefined threshold, a spike is emitted and the potential is reset. This model captures the basic behavior of neuronal spiking using a simple threshold mechanism and is computationally efficient, requiring minimal hardware resources. However, it lacks the ability to represent more complex dynamics such as spike-frequency adaptation, bursting, or different firing regimes. While attractive for its simplicity, the LIF model's limited expressiveness makes it less suitable for applications that demand richer temporal representations or adaptability to varied input dynamics.

3) Izhikevich (IZ) Model: To bridge the gap between biological realism and computational efficiency, the Izhikevich model was introduced in 2003 [11]. Unlike the Hodgkin–Huxley model, which is highly accurate but computationally intensive, the Izhikevich model uses a pair of coupled differential equations and a threshold-based reset mechanism to emulate a wide range of spiking and bursting behaviors with significantly reduced complexity.

Its simplicity enables efficient hardware implementation on both digital and analog neuromorphic platforms, allowing for the simulation of thousands of neurons and synapses in

parallel. In contrast to the HH model, which may require hundreds of transistors per neuron, the Izhikevich model is highly hardware-friendly and has therefore been adopted by several large-scale neuromorphic systems. Another major advantage is its adaptability: by varying only a few parameters, the model can replicate different types of neurons within a single network, such as excitatory vs. inhibitory, or bursting vs. regular-spiking.

This flexibility makes it particularly attractive for spiking neural networks intended to decode dynamic biological signals. For instance, in applications like blink classification or neural signal reconstruction, Izhikevich neurons can be tuned to capture behaviors such as adaptation and spike-frequency modulation, which simpler models like LIF are unable to represent. While it is less biophysically interpretable than the HH model, it captures enough of the essential neuronal dynamics to be functionally useful in practice. As a result, the Izhikevich model has become a widely adopted standard in computational neuroscience and neuromorphic engineering. In this work, it is selected for its ability to support biologically inspired behavior while maintaining real-time performance and scalability, key design requirements for this SNN-based blink classification system.

TABLE I: PUGH MATRIX – SNN NEURON MODEL COMPARISON

Criteria	Weight	IZ	LIF	HH
Biological Precision	20	0	-1	1
Speed	3	0	1	-1
Computational Complexity	5	0	1	-1
Power Consumption	4	0	1	-1
Total Weighted Score		**0**	**-8**	**-1**

The SNN is trained offline using supervised learning with a surrogate gradient-based backpropagation algorithm adapted for spiking dynamics. Once trained, the network is deployed on-chip to classify spike trains in real time. The output consists of two digital signals indicating whether a soft or hard blink has been detected, which external systems can further process to infer cognitive states.

III. DESIGN CONSTRAINTS

The development of a neural-recording SoC involves navigating several critical design constraints to ensure safety, manufacturability, usability, and regulatory compliance.

Health and safety considerations are of paramount importance in the design of any biomedical system. As the proposed device is intended to function as a standalone, non-invasive neural interface, it must meet stringent safety standards to avoid adverse biological effects. The system must incorporate safe, battery-powered operation with appropriate circuit isolation and protection to ensure reliable performance during prolonged use without posing electrical risks to the user.

Manufacturability is also a key constraint, particularly given the goal of integrating the system into a wearable form factor. The circuit must be compatible with standard semiconductor fabrication processes to enable scalable production. The system is designed for head-mounted use, and thus the integrated circuit will be packaged using a Quad Flat No-lead (QFN) package to facilitate compact board-level integration in wearable accessories such as smart headbands or glasses.

From a sustainability perspective, the system must prioritize energy efficiency to minimize long-term power consumption and reduce environmental impact. Low-power operation not only extends battery life but also supports the use of smaller batteries, contributing to a more compact and lightweight wearable. These optimizations support both environmental goals and user comfort.

Social and cultural factors must also be addressed. Since the system involves continuous user monitoring, ensuring data privacy is essential. Onboard processing significantly reduces the need for wireless data transmission, thereby mitigating exposure to privacy risks and side-channel attacks. However, the system's analog front-end includes a 60 Hz notch filter to suppress local power line interference, limiting its immediate applicability to regions operating on a 60 Hz electrical grid, such as North America. For deployment in countries with 50 Hz power infrastructure, the filter configuration must be adjusted accordingly. Additionally, cultural considerations, such as religious head coverings, may limit the practical wearability of head-mounted devices, and such factors must be acknowledged during product deployment.

The system must also meet regulatory requirements governing EEG-based medical devices. Compliance with guidelines such as the IEEE Recommended Practice for Neurofeedback Systems (2010–2012) is essential for safety and performance validation. Electrodes used for data acquisition must be both biocompatible and corrosion-resistant to support safe, long-term skin contact. If the device is intended for diagnostic use, further regulatory approval, such as an FDA clearance, would be required before clinical deployment.

In support of these constraints, adherence to recognized engineering standards is critical. Signal acquisition protocols are guided by the IEEE Recommended Practice for Neurofeedback Systems, ensuring robust and accurate EEG data collection. Proper electrode placement and performance are governed by the IEEE 1058-2003 Standard for Scalp Electroencephalography, which addresses issues of signal integrity and artifact rejection. Electrical safety is maintained by complying with IEC 60601-1, which stipulates requirements for electrostatic discharge (ESD) protection, voltage limiting, and overcurrent protection. These safety features are implemented at both the chip and board levels to safeguard users. Finally, material selection for components in direct contact with skin must follow ISO 10993 standards for biocompatibility, ensuring that electrodes do not cause adverse biological reactions and meet criteria for reuse.

Together, these constraints shape the system's architecture and guide design decisions to ensure that the resulting neuromorphic EEG platform is not only functionally effective but also safe, manufacturable, culturally adaptable, and compliant with regulatory expectations.

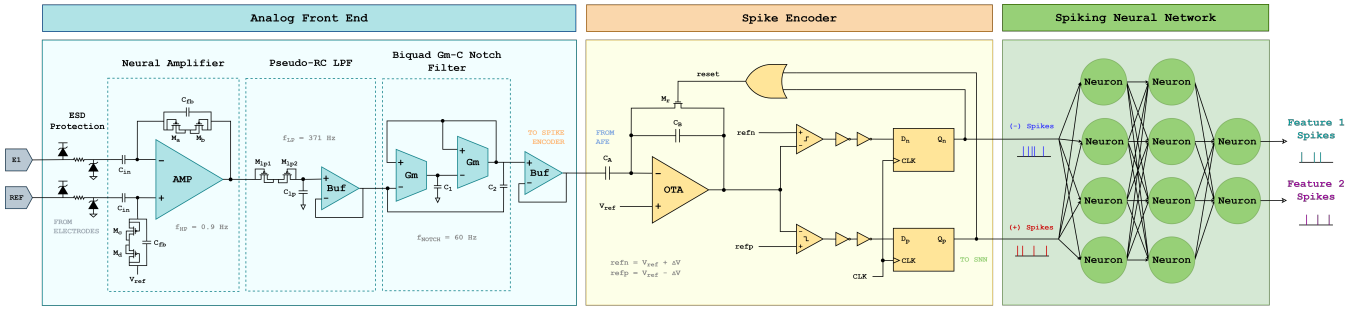


Fig. 5: Block diagram of the complete system architecture. The figure outlines the system stages from electrode input through spike-based processing to post-classification output.

IV. Design Analysis & Validation

A detailed block diagram of the complete SoC architecture is displayed in Fig. 5, illustrating the components of the analog front-end (AFE), spike encoder, and spiking neural network (SNN). The full chip layout is shown in Fig. 14.

A. AFE

1) Neural Amplifier: To optimize utilization of the limited voltage headroom, the analog front-end was designed to produce an output signal with a target peak amplitude of approximately 0.7 V_{pp}, centered around a mid-rail voltage of 0.9 V. Based on a typical input blink signal amplitude of 500 μV, this specification necessitates a voltage gain of approximately 1400 V/V. The AFE uses only a single amplification stage to minimize power consumption and area usage.

The neural amplifier employs a two-stage current-mirror op-amp – labeled in Fig. 5 as AMP – configured as a differential amplifier. Also shown in Fig. 5, capacitive feedback (C_{in} & C_{fb}) in conjunction with pseudo-resistive elements (M_a, M_b, M_c, M_d) establish the amplifier's closed-loop gain and dominant pole frequency response. This topology was first outlined in [12]. The total closed-loop gain of the neural amplifier was set by the transfer function

$$H_{cl}(s) = C_{in}/C_{fb} \tag{1}$$

Due to the gain's dependency on this capacitor sizing ratio, a relatively small 9.52 fF C_{fb} was used to maintain a high gain, as compared to the 15 pF C_{in}. Consequently, a very large equivalent resistance was required to set a sufficiently low high-pass corner frequency. This design choice effectively suppresses baseline drift and motion artifacts while preserving a stable passband for the low-frequency blink-related signals. The proposed pseudo-resistors used for M_a & M_b and M_c & M_d consist of two diode-connected PMOS transistors with local source and body connections, providing the PN-junction leakage and channel leakage currents for the high pseudo-resistance.

As shown in Fig. 6, the first stage (M_1–M_8) implements a three-current mirror OTA topology. The input differential pair (M_1, M_2), together with transistors M_3 and M_5, form a basic differential amplifier. The bias current for the entire amplifier was provided by M_{11}, which supplies a nominal current of 10 μA. The biasing network for M_{11} is depicted in Fig. 8.

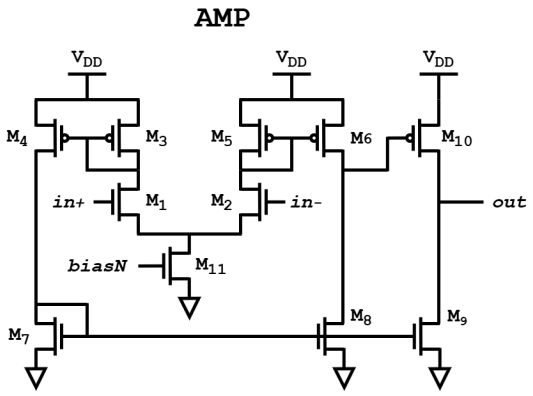


Fig. 6: The two-stage current-mirror op-amp circuit used in the neural amplifier.

TABLE II: Operating Point of AMP Transistors

Devices	W/L (μm)	I_D (μA)
M_1, M_2, M_3, M_5	0.5 / 5.0	5.0
M_4, M_6, M_7, M_8, M_9	5.0 / 5.0	25.0
M_{10}	7.0 / 5.0	25.0
M_{11}	6.0 / 0.6	10.0

Three current mirrors are integrated within the OTA: M_3/M_4, M_5/M_6, and M_7/M_8. To ensure symmetry and matching, the width-to-length (W/L) ratios of transistors in each mirrored pair are carefully maintained. In particular, the boost factor (B) for the current mirrors M_3/M_4 and M_5/M_6 was set to 5, which was achieved by sizing transistors M_4, M_6, M_7, and M_8 to have five times the width of transistors M_1, M_2, M_3, and M_5. The open-loop gain of the first-stage OTA can be approximated as

$$A_{1,ol} = B \cdot g_{m1} \cdot (r_{o6} \parallel r_{o8}) \tag{2}$$

Given the constraints imposed by the low supply voltage and the need to maintain adequate output voltage swing, a cascode configuration was not adopted in the first-stage OTA, as is usually expected. However, an OTA without a buffer can only drive capacitive loads; a resistive load – such as the preceding pseudoresistive low-pass filter – will kill the gain of the OTA [13]. As such, a Class-A gain stage (M_{10}) was

incorporated to increase the load-driving capabilities of the op-amp and add an additional gain of

$$A_{2,ol} = g_{m10} \cdot (r_{o9} \parallel r_{o10}) \tag{3}$$

2) Filters: While the high-pass cutoff frequency was defined by the neural amplifier, an additional single-pole low-pass filter was incorporated to complete the bandpass response, establishing an upper 3 dB cutoff frequency of approximately 371 Hz. To minimize area, a passive RC filter topology was formed using another high resistance pseudo-resistor element (M_{lp1} & M_{lp2}) and a 107.6 fF capacitor (C_{lp}).

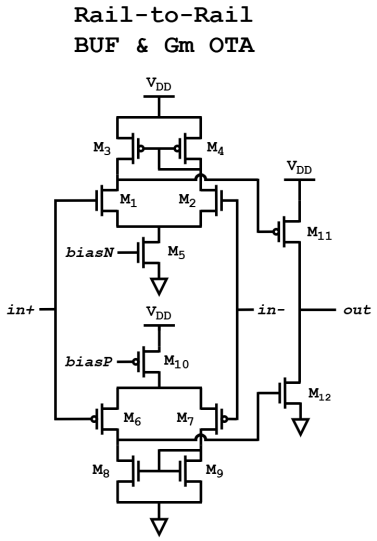


Fig. 7: Schematic of the wide-input-range OTA with output buffer. This topology was utilized for both voltage buffering and g_m stages within the filter circuits, with modifications to biasing and transistor sizing to suit each function.

TABLE III: OPERATING POINT OF BUF TRANSISTORS

Devices	W/L **(μm)**	I_D **(μA)**
$M_1, M_2, M_3, M_4,$ M_6, M_7, M_8, M_9	0.6 / 0.6	0.5
M_{11}, M_{12}	1.0 / 1.0	1.0
M_5, M_{10}	0.6 / 0.6	1.0

The output was buffered with a rail-to-rail OTA utilizing parallel PMOS and NMOS differential amplifiers, and a class-A output stage to mitigate loading effects. The buffer uses the topology shown in Fig. 7, in a negative feedback configuration. Negative feedback significantly reduces the output resistance of the common-source stage by a factor proportional to the loop gain. Accordingly, the output resistance of the buffer can be approximated as

$$R_{\text{buff}} \approx \frac{1}{A_{\text{diff}} \cdot (g_{m11} + g_{m12})} \tag{4}$$

where A_{diff} denotes the gain of the input differential pair.

TABLE IV: OPERATING POINT OF g_m OTA TRANSISTORS

Devices	W/L **(μm)**	I_D **(μA)**
M_1, M_2, M_6, M_7	0.42 / 8.0	0.005
M_3, M_4, M_8, M_9	7.0 / 0.15	0.005
M_{11}, M_{12}	0.42 / 0.3	25.0
M_5, M_{10}	0.42 / 0.6	0.01

To selectively suppress 60 Hz line noise while preserving the passband integrity, a sharp band-reject filter was implemented; a biquadratic transconductance–C (g_m–C) notch topology was employed for its low area usage and second-order response, as shown in Eqn. 5:

$$H_{notch}(s) = \frac{s^2 C_1 C_2 + g_m^2}{s^2 C_1 C_2 + s C_1 g_m + g_m^2} \tag{5}$$

where both C_1 and C_2 are 15 pF. To accommodate the low target frequency, the g_m OTA was sized accordingly and biased at 10 nA to reduce transconductance. Rail-to-rail operation was also required to maintain the output range of the neural amplifier.

3) Bias Circuits: Most OTAs in both the analog front-end and spike encoder utilized a shared, simple NMOS current mirror for biasing. For the rail-to-rail output buffers, both 10 μA NMOS and additional 10 μA PMOS bias circuits were employed. The low-current g_m OTAs, biased at 10 nA, required dedicated NMOS and PMOS biasing circuits.

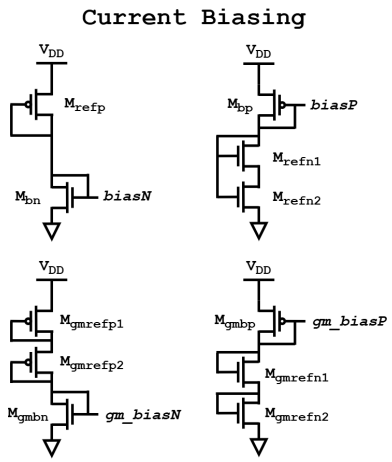


Fig. 8: Current mirror bias circuits used for the analog subsystem.

Reference currents were generated using transistors in a diode configuration; all described bias circuits are shown in Fig. 8.

The overall frequency response of the entire AFE was characterized with an AC sweep shown in Fig. 9, showing a 62.7 dB passband between 891 mHz to 371 Hz. The neural amplifier exhibited resonant peaking at about 2.3 MHz, which was attenuated by the single-pole low-pass filter.

TABLE V: OPERATING POINT OF THE BIAS TRANSISTORS

Devices	W/L (μm)	I_D (μA)
M_{bn}	0.6 / 0.6	1.0
M_{refp}	0.6 / 1.1	1.0
M_{bp}	0.6 / 0.6	1.0
M_{refn1} , M_{refn2}	0.42 / 0.15	1.0
M_{gmbn}	7.0 / 0.6	0.01
M_{gmrefp1}	0.42 / 0.35	0.01
M_{gmrefp2}	0.42 / 0.6	0.01
M_{gmbp}	7.0 / 0.6	0.01
M_{gmrefp1} , M_{gmrefp2}	0.42 / 0.15	0.01

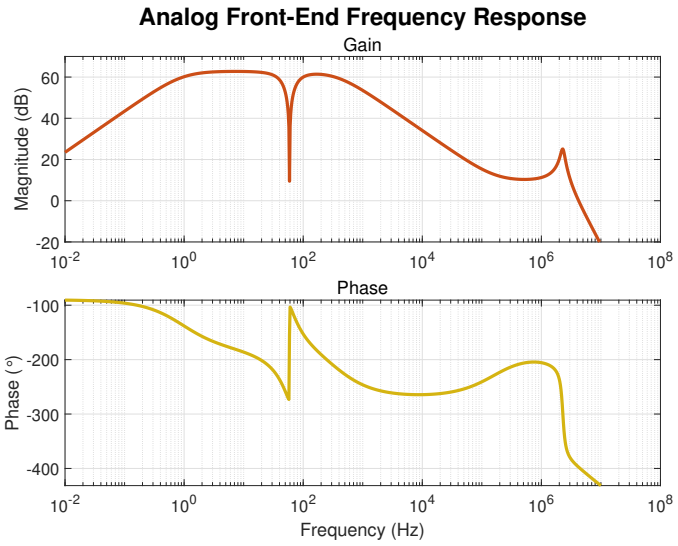


Fig. 9: Simulated frequency response of the AFE, demonstrating a gain of 62.7 dB, bandlimited characteristics, and effective notch attenuation at 60 Hz.

B. Spike Encoder

Delta modulation spike encoding was implemented via a capacitive divider differencing stage, UP and DOWN comparators, clock synchronizing delay flip-flops (D-FFs), and an internal reset mechanism. The architecture is shown in Fig. 5.

The operating principle of the spike encoder, as visualized with a transient simulation in Fig. 10, can be summarized as follows:

1) The input signal from the analog front-end (AFE) begins to deviate upward or downward.
2) In a unity-gain capacitive divider configuration (C_A = C_B = 15 pF), the inverting OTA output crosses either the positive ($refp$) or the negative ($refn$) delta threshold.
3) The corresponding UP or DOWN comparator toggles high, causing the respective D flip-flop (D-FF) to latch a logic high on the next rising clock edge.
4) This logic high propagates through an OR gate, activating the reset transistor M_r. Through negative feedback, the OTA output is pulled back to the reference voltage V_{ref}.
5) The reset action causes the comparator output to return low, clearing the D-FF input before the next clock edge.
6) The D-FF then transitions low, producing a clean, clock-synchronized, one-bit-wide UP or DOWN "spike" signal for the spiking neural network.

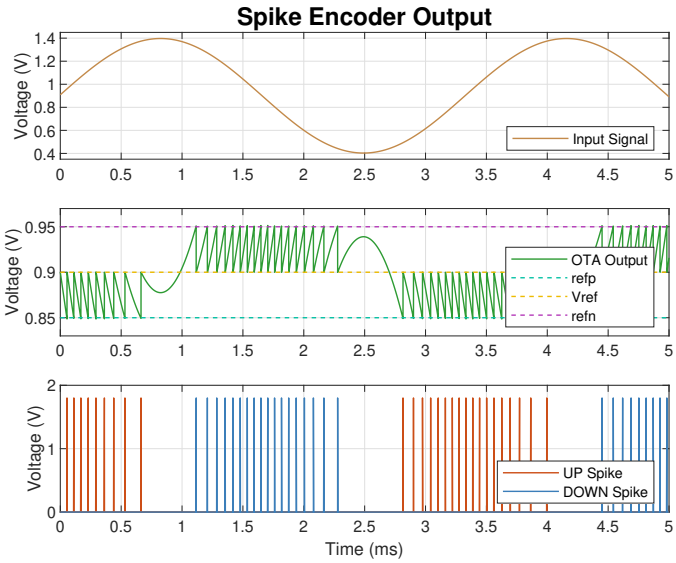


Fig. 10: Spike encoder principle of operation.

Schematics for the OTA and comparators are shown in Fig. 11. Standard cell topologies were used for the positive-edge triggered D-FFs, inverters, and OR gate.

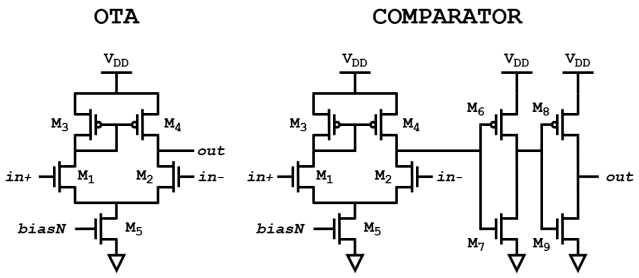


Fig. 11: OTA and comparator schematics used in the spike encoder.

TABLE VI: OPERATING POINT OF OTA TRANSISTORS

Devices	W/L (μm)	I_D (μA)
M_1, M_2, M_3, M_4, M_5	0.42 / 0.3	0.5
M_5	0.6 / 0.6	1.0

A simple five-transistor NMOS differential amplifier was used to implement the OTA. The unbuffered output was made feasible by the predominantly capacitive nature of the output load. Rail-to-rail output was unnecessary, as the output was constrained to V_{ref} (0.9 V) $\pm\Delta V_{\text{margin}}$ (50 mV), minimizing concerns related to saturation.

The comparators employed a similar differential input stage and drove a chain of inverters to produce rail-to-rail, squared output signals. Common issues such as switching noise, typically caused by input signals lingering near the threshold,

TABLE VII: OPERATING POINT OF COMPARATOR TRANSISTORS

Devices	W/L (μm)	I_D (μA)
M_1, M_2, M_3, M_4, M_5	0.42 / 0.3	0.5
M_5	0.6 / 0.6	1.0
M_6, M_8	1.0 / 0.15	–
M_7, M_9	0.65 / 0.15	–

are mitigated in this design, as the input is reset immediately after the comparator toggles. Consequently, internal positive feedback or hysteresis circuitry was not required. In this application, precise switching at the delta thresholds was prioritized to ensure consistent behavior between UP and DOWN spikes.

To generate these precisely matched reference voltages of ± 50 mV around a central reference of 0.9 V, a dedicated reference voltage circuit was designed (Fig. 12). The implementation consists of two diode-connected PMOS transistors and two series resistors. While the use of resistors increases chip area, this approach was preferred over an all-diode configuration (e.g., stacked NMOS and PMOS diodes) due to superior matching characteristics during layout. Given sufficient layout area, matching was prioritized over area efficiency.

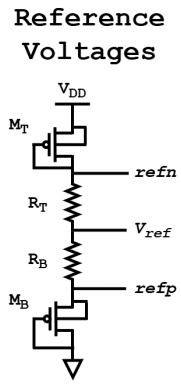


Fig. 12: The voltage reference circuit; designed with layout size-matching considerations. Locally isolated N-well body connections were utilized to limit body effect mismatch.

The reference voltage levels are governed by the PMOS saturation current equation and Ohm's Law:

$$I_D = \frac{\Delta V_{\text{margin}}}{R} = \frac{K_p}{2}\frac{W}{L}(V_{SG} - |V_{THP}|)^2 \tag{6}$$

where $\Delta V_{\text{margin}} = 50$ mV, K_p is the PMOS process transconductance parameter, and $V_{SG} = (V_{DD}/2) - \Delta V_{\text{margin}} = 850$ mV. The design targeted a drain current $I_D < 10\,\mu$A while also optimizing for area.

Resistor values R_T and R_B were each set to 9.64 kΩ. Using Eqn. 6 and performing DC analysis to extract device parameters, M_T & M_B PMOS transistors were sized with a width-to-length ratio of 5.3 μm / 0.3 μm to achieve the desired voltage drops. The R_T and R_B resistors were laid out using an interdigitated structure to enhance matching precision.

A transient simulation of the complete analog subsystem operating in real time with real recorded EEG blink signals is presented in Fig. 13. The simulation compares responses to a pronounced "hard" blink and a subtler "soft" blink, demonstrating the corresponding variation in UP and DOWN spike rates generated by the delta-modulation-based spike encoder. The delta threshold was selected to balance sensitivity and noise robustness; it was sufficiently large to suppress false-positive spikes induced by background noise, yet small enough to ensure that the downstream spiking neural network (SNN) can reliably discriminate blink events.

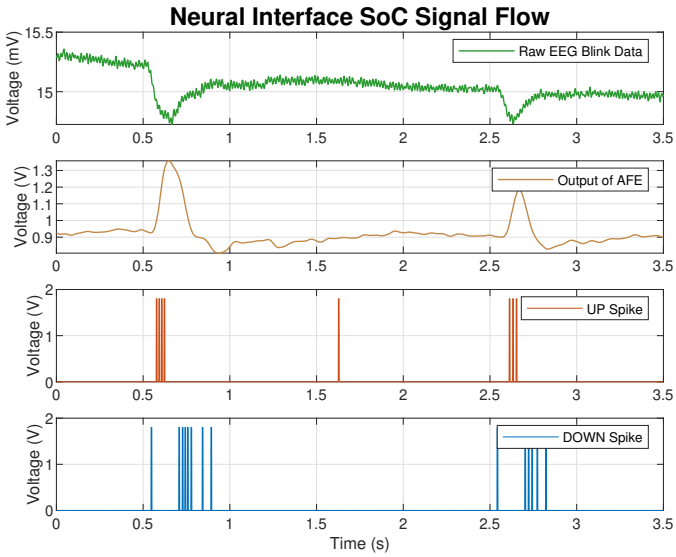


Fig. 13: Simulation results using raw EEG blink data, showing the signal conditioning stages within the analog subsystem (amplification, filtering, and spike encoding) prior to the SNN.

C. SNN

1) Design: The SNN was implemented using the Izhikevich (IZ) neuron model along with an exponential decay synapse model. The entire network was described in Verilog, synthesized into a standard cell netlist, and finalized in layout through placement and routing using Cadence design tools.

The Izhikevich model combines biological plausibility with computational efficiency and is governed by the following system of ordinary differential equations (ODEs) [14]:

$$C\frac{dv}{dt} = k(v - v_r)(v - v_t) - u + I \tag{7}$$

$$\frac{du}{dt} = a\,(b(v - v_r) - u) \tag{8}$$

subject to the after-spike resetting rule:

$$\text{if } v \geq v_{\text{peak}}, \quad \begin{cases} v & \leftarrow c \\ u & \leftarrow u + d \end{cases} \tag{9}$$

The parameters used in the model are defined as follows:

- C – Membrane capacitance
- k – Gain factor
- v_r – Resting membrane potential

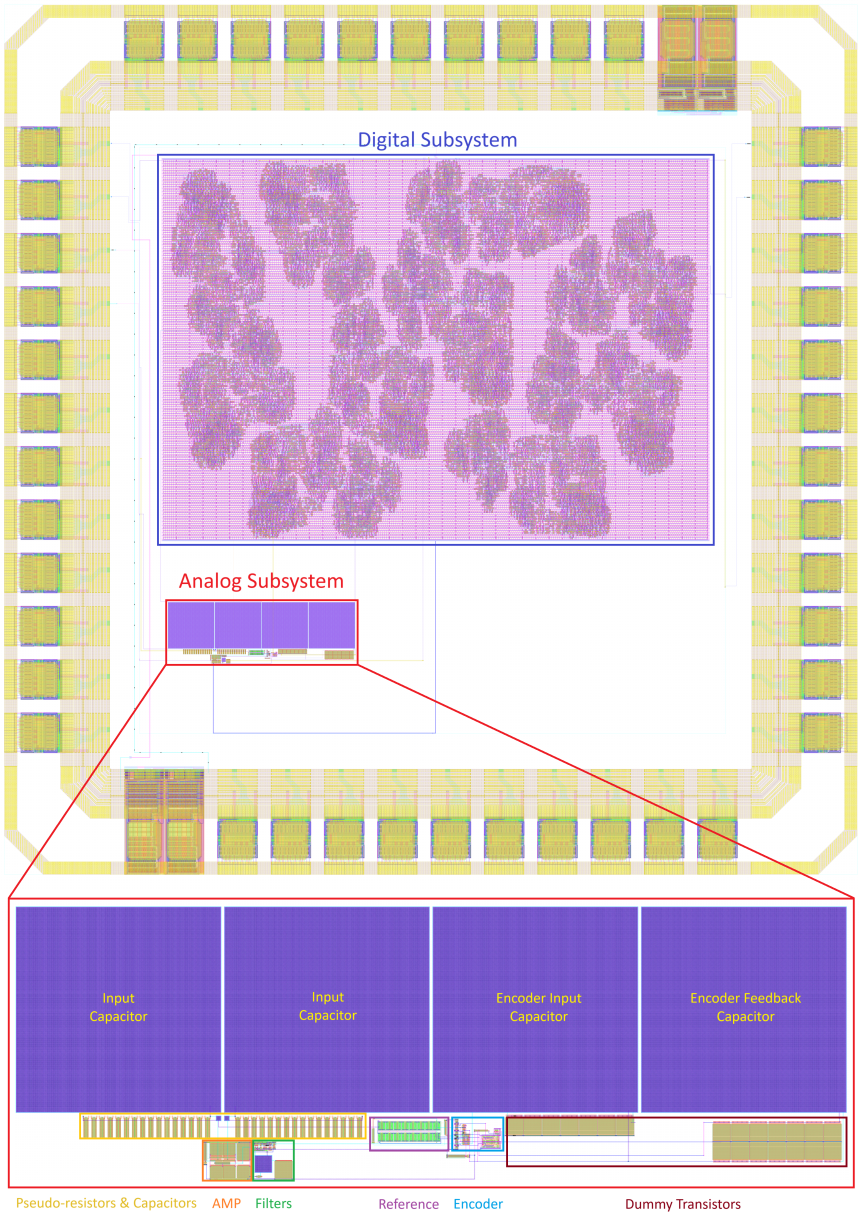


Fig. 14: The complete SoC on a 1.6 mm × 1.6 mm forty-pin padframe.

- v_t – Instantaneous threshold potential
- v_{peak} – Peak voltage threshold for spike reset
- a – Time scale of the recovery variable u
- b – Sensitivity of u to subthreshold fluctuations of v
- c – Reset value of v after a spike
- d – Increment of u after a spike

Although a simplified 4-parameter version of the IZ model exists [11], the 9-parameter model employed here enables a richer set of neuronal dynamics, supporting a broader range of spiking behaviors.

Because the Izhikevich model is described by coupled nonlinear ODEs, a Digital Differential Analyzer (DDA) is required to solve these equations iteratively over discrete time steps. In this design, the Euler method [15] was selected for numerical integration due to its simplicity, deterministic behavior, low hardware overhead, and compatibility with fixed-point arithmetic, making it well suited for real-time embedded systems [16].

The Euler method approximates solutions of first-order ODEs of the form:

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0 \tag{10}$$

by discretizing time into steps of size h and applying the iterative update rule:

$$y_{n+1} = y_n + h \cdot f(t_n, y_n) \tag{11}$$

where:

- $t_n = t_0 + nh$ is the n-th time step,
- y_n is the approximation of $y(t_n)$,
- h is the time step size.

This approach yields a piecewise linear approximation to the continuous solution of the ODE.

Applying the Euler method to the Izhikevich neuron model yields the following update equations for the membrane potential v and recovery variable u:

$$v_{\text{new}} = v + \frac{h}{C}\left[k(v - v_r)(v - v_t) - u + I\right] \tag{12}$$

$$u_{\text{new}} = u + h \cdot a\left[b(v - v_r) - u\right] \tag{13}$$

along with the spike-reset condition:

$$\text{if } v_{\text{new}} \geq v_{\text{peak}}, \quad \begin{cases} v_{\text{new}} & \leftarrow c \\ u_{\text{new}} & \leftarrow u_{\text{new}} + d \end{cases} \tag{14}$$

To optimize resource efficiency, the time step h is chosen to be a power of two, enabling the multiplication by h to be implemented as a simple bit-shift operation. In this design, h is set to 0.125, which provides a suitable balance between computational precision and performance.

Aside from lowering the system's clock frequency, reducing the number of digital bits is one of the most effective strategies to minimize power consumption since it directly decreases the number of transistors required in circuit implementations. However, due to the dynamic behavior of neuronal signals, using too few bits may result in loss of resolution between consecutive time steps, impairing the accuracy of neuron state updates. To achieve a balance between power efficiency and computational fidelity, a 16-bit fixed-point representation is adopted. This bit-width represents the minimum precision necessary to maintain the dynamic range of synaptic currents accurately.

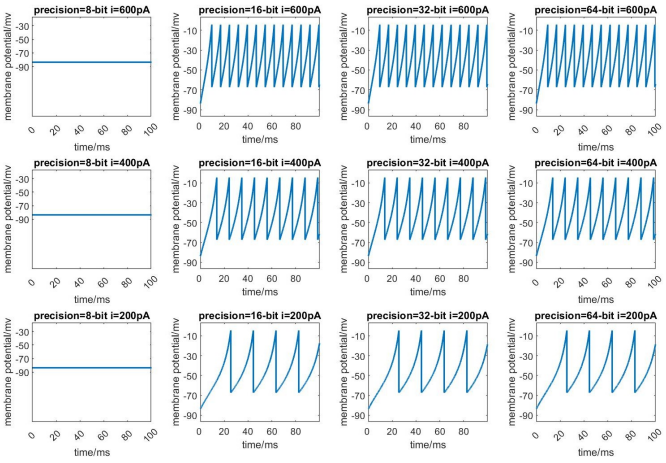


Fig. 15: Comparison of different bit-length fixed-point representations in simulating various synaptic current inputs.

A 16-bit signed fixed-point representation is utilized, as shown in Fig. 16. This format provides sufficient dynamic range while enabling efficient arithmetic operations in hardware.

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
+/-	Fractional part														

Fig. 16: Structure of the 16-bit signed fixed-point digital representation.

Synapses play a critical role in SNNs, acting as the primary channels for communication and signal integration. In this design, exponential decay is employed as the synaptic model due to its effectiveness in mimicking biological synaptic behavior and its ability to accumulate temporal information [17]. For hardware efficiency, the exponential decay is approximated using a scalar decay coefficient α, where $\alpha < 1$. The synaptic current is updated at each discrete time step using the following equation:

$$I_{\text{new}} = \alpha \cdot I + w \cdot s, \tag{15}$$

where:

- I is the synaptic current from the previous time step,
- α is the decay coefficient, satisfying $0 < \alpha < 1$,
- w is the synaptic weight,
- s is the spike input at the current time step (typically 0 or 1).

As with the Euler time step h used in neuron state updates, α is chosen to be a power of two to allow the decay operation to be implemented efficiently via bit shifting. Fig. 17 illustrates the simulated behavior of an exponential decay synapse connecting two neurons.

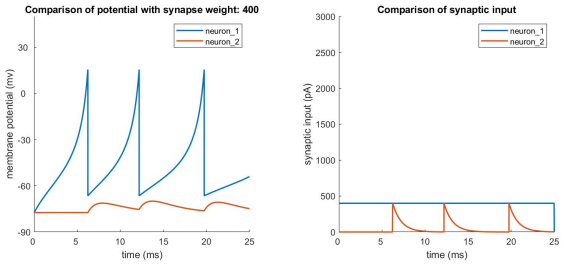


Fig. 17: Simulation of two neurons connected by an exponential decay synapse. The synapse delivers current input to the second neuron upon receiving spikes from the first neuron.

The Verilog implementation adheres to the update rules described above. However, due to the fixed-point arithmetic, precision limitations can lead to overflow or quantization errors. To mitigate these effects, a custom 16-bit fixed-point multiplier is implemented with overflow handling. The implemented SNN adopts a 2-4-4-2 architecture, where all synaptic weights are defined as parameters within the Verilog module. Fig. 18 presents the simulation results of the full network, demonstrating the behavior of neurons across each layer. For clarity, all synaptic weights are unified in the illustration.

Fixed-point 16-bit Multiplier in Verilog

```
function [15:0] fixed_16_mul;
    input [15:0] a;
    input [15:0] b;

    reg [29:0] temp;
    reg [14:0] a_temp, b_temp;

    begin
        if (a[15]) begin
            a_temp = ~(a[14:0] - 1);
        end else begin
            a_temp = a[14:0];
        end

        if (b[15]) begin
            b_temp = ~(b[14:0] - 1);
        end else begin
            b_temp = b[14:0];
        end

        temp = a_temp[14:0] * b_temp[14:0];
        if (a[15] ^ b[15]) begin
            fixed_16_mul = ~{1'b0, temp[29:15]} + 1;
        end else begin
            fixed_16_mul = {1'b0, temp[29:15]};
        end
    end
endfunction
```

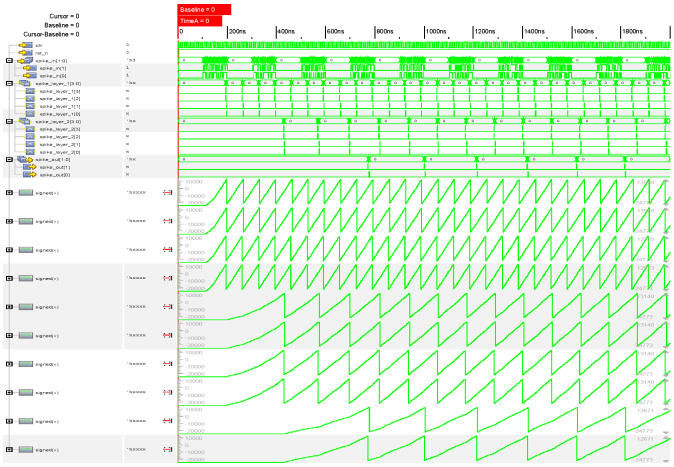


Fig. 18: Verilog simulation of the 2-4-4-2 SNN architecture. Unified synaptic weights are used for visualization clarity.

The digital SNN was synthesized using Cadence Genus, operating under a timing constraint of 20 MHz. While this frequency exceeds the application's requirements, it ensures scalability for more complex tasks involving additional neurons and higher-dimensional feature extraction. This is elaborated on more in *Section V*. Final placement and routing were completed using Cadence Innovus, and the SNN layout was globally integrated with the subsystem within the provided pad frame. Additional implementation details, including Verilog source code and scripts, are provided in *Appendix B: Code*.

2) Training: For hyperparameter tuning, an open-source raw EEG blink dataset collected from live subjects using OpenBCI headsets was utilized [18]. The dataset included labeled instances of "soft" and "hard" blinks, which represent the ground-truths used for training the neural network. Raw blink waveforms were applied to the analog subsystem within the Cadence Virtuoso Analog Design Environment (ADE) Explorer, as illustrated in Fig. 13. The resulting UP and DOWN spiking outputs were then used for training.

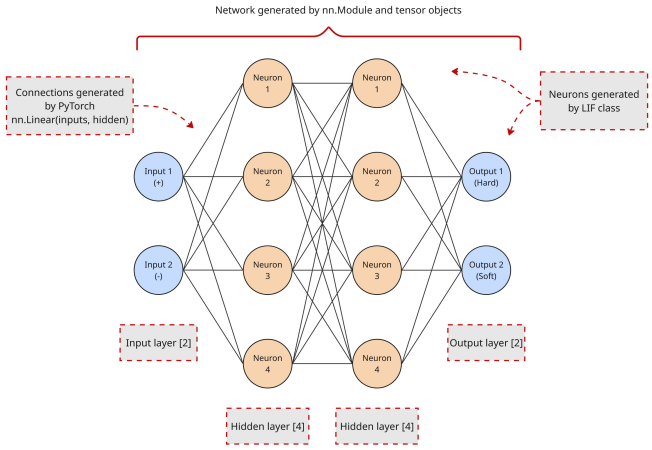


Fig. 19: The Python code setup for SNN training.

Training the SNN was performed in two phases. In the first phase, the network was trained using the pre-established Leaky Integrate-and-Fire (LIF) neuron model provided by the `snnTorch` library. At present, `snnTorch` does not include a built-in Izhikevich neuron class suitable for training. Therefore, the LIF neuron class was employed as a proxy during initial training. This two-phase training approach allowed for initial parameter optimization using LIF dynamics, while preserving the ability to later substitute Izhikevich dynamics for more biologically realistic inference in the hardware implementation.

The LIF training phase utilized a simplified 2-4-2 network configuration, which is structurally similar to the architecture illustrated in Fig. 19, but with one fewer hidden layer. The training data were passed into the two input nodes and linearly transformed to map onto the four-neuron hidden layer. A similar linear transformation was then applied to connect the hidden layer to the output neurons. The training model employed rate encoding to interpret spiking activity and determine correct classifications. The loss function used was cross-entropy, a standard choice for classification problems. To optimize the loss gradient during training, the Adaptive Moment Estimation (Adam) optimization algorithm was applied.

The labeled ground truths in the raw dataset are illustrated in Fig. 20. For training purposes, the data was modified by converting spike trains into spike windows, thereby increasing the representation of the spiking class relative to the predominantly non-spiking data. Initially, spike events constituted less than 2% of the dataset, resulting in extreme sparsity. Without this conversion, the spiking neural network (SNN) tends to achieve high accuracy by predicting the dominant non-spiking class, leading to biased learning.

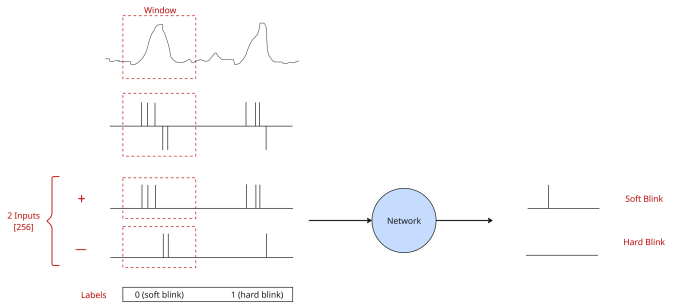


Fig. 20: Parsing the data by creating windows that contained only blinks

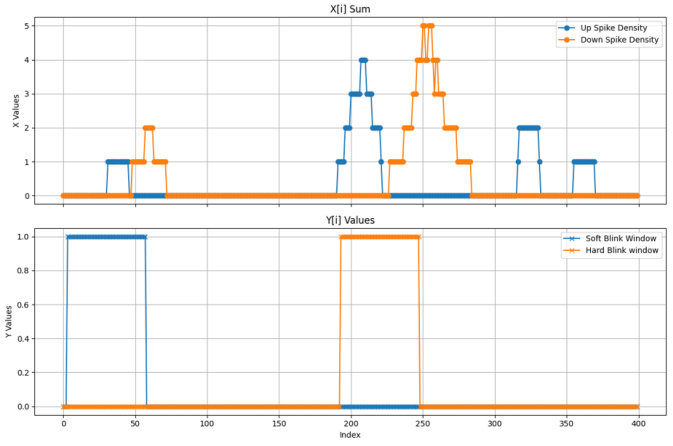


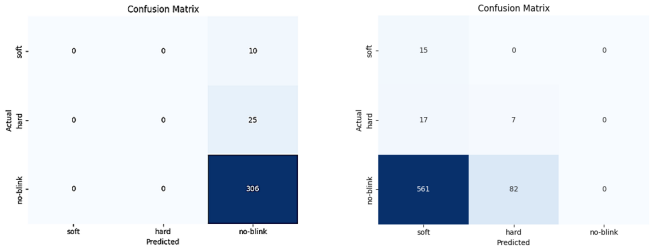
Fig. 21: Spike density for the classes and their respective label windows.

To address this class imbalance, the spiking (blinking) class was oversampled, and the non-spiking class was undersampled, as shown in Fig. 21. This windowed representation of blinking activity was employed during training of the leaky integrate-and-fire (LIF) model, achieving a peak accuracy of 86.67%. Further training epochs would be required for an increased accuracy.

Subsequently, training was extended to a 2-4-4-2 Izhikevich SNN model, as depicted in Fig. 5. The dataset was split with an 80-20 train-test ratio. The architecture comprises 2 input neurons, two hidden layers with 4 neurons each, and 2 output neurons. Synaptic weights were tuned to facilitate the detection of blink-related spike patterns.

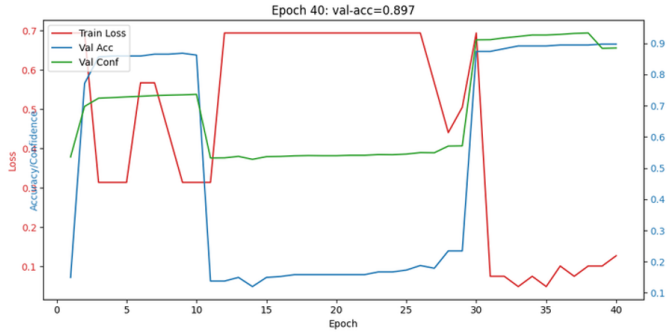
The output neurons encode the following classes: [0,0] — no blink, [0,1] — hard blink, and [1,0] — soft blink. In the output vector $[a, b]$, a and b denote spikes generated by the respective output neurons.

Model performance for unbalanced and balanced datasets is presented in Fig. 22. Corresponding training metrics (accuracy, loss, and confidence) are shown in Fig. 23. Trained synaptic weights for all inter-neuronal connections in the balanced model are visualized in Fig. 24. Training over 40 epochs yielded a minimum accuracy of approximately 25% when the network predominantly predicted a single blinking class, and a

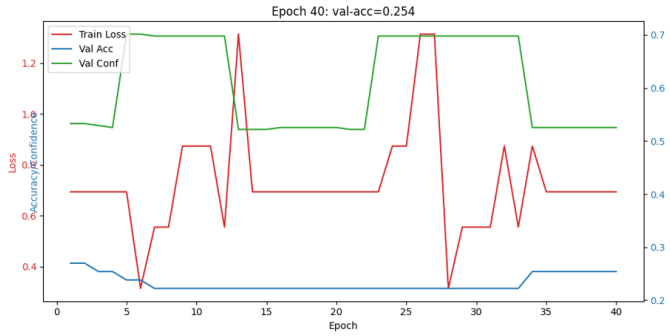


(a) (b)

Fig. 22: Confusion matrices for (a) unbalanced class distribution and (b) balanced class distribution.



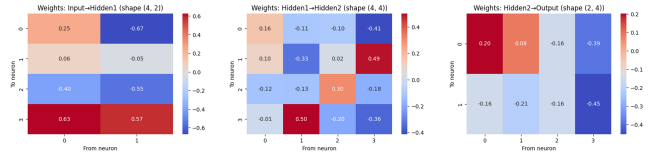
(a)



(b)

Fig. 23: Training results (accuracy, loss, and confidence) for (a) unbalanced and (b) balanced class distributions.

maximum of approximately 85% when the non-blinking class was dominant.



(a) (b) (c)

Fig. 24: Synaptic weights between neuron layers after training with a balanced dataset.

These results demonstrate the necessity of class balancing and proper data representation for effective SNN training. While the current network remains in an early stage of training, further improvements in data volume, labeling accuracy,

class distribution, and loss function design could significantly enhance overall performance.

V. Performance Optimization

A key performance consideration for the SoC is power consumption. For a single-processor system operating at a clock frequency f_{CLK}, the total power consumption $p(f_{\text{CLK}})$ comprises dynamic power $p_d(f_{\text{CLK}})$ and static power $p_s(f_{\text{CLK}})$. The dynamic component, primarily due to gate charging, capacitive discharge, and short-circuit currents during switching, is modeled as:

$$p_d(f_{\text{CLK}}) = \alpha(f_{\text{CLK}})^3 \tag{16}$$

where the power increases proportionally to the third power of frequency. The static power, resulting mainly from leakage currents, is modeled as a constant:

$$p_s(f_{\text{CLK}}) = \beta \tag{17}$$

Thus, the total digital power consumption is given by:

$$p(f_{\text{CLK}}) = \alpha(f_{\text{CLK}})^3 + \beta \tag{18}$$

where $\alpha > 0$ and $\beta \geq 0$ [19]. In the absence of performance constraints, minimum energy consumption is achieved at the lowest permissible clock frequency.

The lower bound on f_{CLK} is determined by the maximum spike density at the encoder output, which is influenced by the characteristics of the transient signal from the analog front-end (AFE). These characteristics, in turn, depend on the rate of change of the electrode input blink waveform, $V_{\text{blink}}(t)$. The minimum time between encoder spikes, or refractory period τ_{rfr}, is constrained by the clock speed and is calculated as:

$$(\tau_{\text{rfr}})_{\text{min}} = \frac{\Delta V_{\text{margin}}}{\left|\frac{d}{dt}\left[\text{Gain} \cdot V_{\text{blink}}(t)\right]\right|_{\text{max}}} \tag{19}$$

This leads to a minimum required clock frequency:

$$f_{\text{CLK}} \geq \frac{1}{(\tau_{\text{rfr}})_{\text{min}}} \tag{20}$$

For the system described, this constraint yields a minimum operational frequency of 140 Hz.

Although the system supports operation at low frequencies due to the gradual rise time of the blink waveform, it is also designed to accommodate a wide clock frequency range (150 Hz to 49 MHz) to enable power optimization and user application configurability. Clock input may be supplied externally via a crystal-based phase-locked loop (PLL) oscillator, ensuring frequency and phase stability.

Digital power estimates were obtained using Cadence Genus synthesis tools. At frequencies below 5 kHz, dynamic power was negligible relative to static power, consistent with (18). Under these conditions, total digital power was measured at approximately 43 μW. At the upper frequency limit of 49 MHz, dynamic power dominates, reaching an estimated 49 mW. Analog power consumption was characterized via transient simulations using Cadence Virtuoso ADE Explorer. The average current draw during operation was 35.3 μA, corresponding to 63.5 μW with a 1.8 V supply rail. At a clock frequency of 300 Hz, the total power consumption was 106.5 μW, allowing the SoC to operate for over five months when powered by an externally regulated 225 mAh CR2032 coin cell battery.

The full performance summary of the SoC is listed in Table VIII.

TABLE VIII:
Performance Summary of the Neuromorphic SoC

Performance Metric	**Value**
Supply Voltage	1.8 V
Analog Front-End:	
Gain	62.7 dB
Bandwidth	891 mHz–371 Hz
CMRR	46.28 dB
Area	0.02 mm^2
Spike Encoder:	
Delta Threshold (ΔV_{margin})	$\pm$0.05 V
Refractory Period (τ_{rfr})	$1/f_{\text{CLK}}$
Spike Width	1 bit
Area	0.02 mm^2
Spiking Neural Network:	
Structure	2-4-4-2
Area	0.738 mm^2
[a]CLK Speed (f_{CLK})	150 Hz – 49 MHz
Average Analog Power	63.5 μW
[b]Average Digital Power	43 μW–49 mW
Total Power Consumption	106.5 μW–49.06 mW
Total Area	0.742 mm^2
Total Transistor Count	276,542

[a] CLK speed is user adjustable within these bounds.
[b] Power consumption scales with f_{CLK} following Eqn. 18.

VI. Conclusion

This work presented the design and validation of a low-power neuromorphic System-on-Chip (SoC) for real-time EEG-based cognitive state monitoring using eyeblink signals. Leveraging event-driven delta modulation and a digital Spiking Neural Network (SNN), the system achieves compute-efficient, on-chip signal processing tailored for wearable applications. The integration of analog front-end, asynchronous spike encoding, and biologically inspired neural classification demonstrates the feasibility of deploying the SoC on compact, non-invasive EEG interfaces for attentional and sleep-state tracking, including consumer electronics such as smart glasses or headbands.

Through iterative analog and digital design, simulation, and layout in the SkyWater 130nm CMOS process, key trade-offs between power, accuracy, and scalability were identified. Event-driven delta modulation proved superior to absolute thresholding for preserving low-frequency EEG blink features in sparse, compressible spike trains. Selecting the Izhikevich neuron model for the SNN struck a practical balance between biological fidelity and computational complexity. Future steps for this work include:

1) More training, more data, and more epochs for more accurate weights.
2) Further integration and optimization between the analog front-end and spike encoder stages, such as integrating

amplification and filtering directly into the spike encoder input stage.

3) Implement a pipelined architecture for an SNN with better accuracy to reuse computational resources to save power and space.
4) Tapeout and fabrication; physical verification tests and measurements on silicon.
5) System design, packaging, and integration to integrate the SoC with physical systems.
6) In-field testing and analysis of the final packaged device.

REFERENCES

[1] E. Alyan, S. Arnau, J. E. Reiser, S. Getzmann, M. Karthaus, and E. Wascher, "Blink-related EEG activity measures cognitive load during proactive and reactive driving," *Scientific Reports*, vol. 13, no. 1, p. 19379, Nov. 2023, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-023-46738-0

[2] C. C. Liu, S. Ghosh Hajra, G. Pawlowski, S. D. Fickling, X. Song, and R. C. N. D'Arcy, "Differential neural processing of spontaneous blinking under visual and auditory sensory environments: An EEG investigation of blink-related oscillations," *NeuroImage*, vol. 218, p. 116879, Sep. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811920303657

[3] S. M. Willett, S. K. Maenner, and J. P. Mayo, "The perceptual consequences and neurophysiology of eye blinks," *Frontiers in Systems Neuroscience*, vol. 17, Aug. 2023, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2023.1242654/full

[4] H. Hinrichs, M. Scholz, A. K. Baum, J. W. Y. Kam, R. T. Knight, and H.-J. Heinze, "Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications," *Scientific Reports*, vol. 10, no. 1, p. 5218, Mar. 2020, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-020-62154-0

[5] K. E. Mathewson, T. J. L. Harrison, and S. A. D. Kizuk, "High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes," *Psychophysiology*, vol. 54, no. 1, pp. 74–82, 2017, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/psyp.12536. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.12536

[6] S. E. Paraskevopoulou and T. G. Constandinou, "A sub-1μW neural spike-peak detection and spike-count rate encoding circuit," in *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Nov. 2011, pp. 29–32, iSSN: 2163-4025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6107719

[7] T. Horiuchi, T. Swindell, D. Sander, and P. Abshire, "A low-power CMOS neural amplifier with amplitude measurements for spike sorting," in *2004 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, May 2004, pp. IV–29. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1328932

[8] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128\times 128 120 dB 15 s Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008. [Online]. Available: https://ieeexplore.ieee.org/document/4444573

[9] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1952.sp004764. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1952.sp004764

[10] L. Lapicque, "Quantitative investigations of electrical nerve excitation treated as polarization," *Biological Cybernetics*, vol. 97, no. 5, pp. 341–349, 1907. [Online]. Available: https://doi.org/10.1007/s00422-007-0189-6

[11] E. Izhikevich, "Simple model of spiking neurons," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1569–1572, Nov. 2003. [Online]. Available: https://ieeexplore.ieee.org/document/1257420

[12] R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, Jun. 2003. [Online]. Available: https://ieeexplore.ieee.org/document/1201998

[13] J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 4th ed. John Wiley & Sons, Jul. 2019.

[14] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Jul. 2006. [Online]. Available: https://direct.mit.edu/books/monograph/2589/Dynamical-Systems-in-NeuroscienceThe-Geometry-of

[15] L. Euler, *Institutiones Calculi differentialis*, 1755, google-Books-ID: sYE_AAAAcAAJ.

[16] R. Guerrero-Rivera, A. Morrison, M. Diesmann, and T. C. Pearce, "Programmable Logic Construction Kits for Hyper-Real-Time Neuronal Modeling," *Neural Computation*, vol. 18, no. 11, pp. 2651–2679, Nov. 2006. [Online]. Available: https://doi.org/10.1162/neco.2006.18.11.2651

[17] Z. Shi and T. Horiuchi, "A Summating, Exponentially-Decaying CMOS Synapse for Spiking Neural Systems," in *Advances in Neural Information Processing Systems*, vol. 16. MIT Press, 2003. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2003/hash/e3ca0449fa2ea7701a7ac53fb719c51a-Abstract.html

[18] M. Agarwal and R. Sivakumar, "Blink: A Fully Automated Unsupervised Algorithm for Eye-Blink Detection in EEG Signals," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2019, pp. 1113–1121. [Online]. Available: https://ieeexplore.ieee.org/document/8919795

[19] Z. Peng and G. Wang, "An optimal energy-saving real-time task-scheduling algorithm for mobile terminals," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, p. 1550147717707891, May 2017, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/1550147717707891

VII. APPENDIX

A. Development

TABLE IX: TEAM ROLES AND RESPONSIBILITIES

Team Member	Role and Contributions
Sierra Raspa	• Preliminary steps to SNN training • Voltage reference circuit and layout • Spike encoder layout and system-level layout
Jeremy Yun (Team Lead)	• Analog design and simulation (spike encoder, AFE, reference & bias circuits) • Signal processing and SNN training data preparation • AFE and system-level layout
Yihui Wang	• SNN design, verification, and training methods • SNN implementation and mixed-signal layout
Kaushik Lakshmiramanan	• Encoding circuit for spikes • Neural data preparation, analysis, and SNN training • System layout integration with padframe
Duncan Millar	• Initial AFE circuit design • Voltage reference layout

B. Code

The GitHub repository containing the EEG blink dataset and SNN training files can be found at https://github.com/JermYeWorm/Blink-Twice-For-Help.

C. Acknowledgement

Special thanks to Dr. Pamela Abshire, David Nchekwube, Utku Noyan, and all the students of the Spring 2025 ENEE408D VLSI Capstone course at the University of Maryland, College Park.