A 311 nW Integrated Neural Amplifier and Spike Enhancement Filter Achieving 98.99% Spike Detection Sensitivity

Le Yang
Electrical and Computer Engineering Department
UC San Diego
San Diego, CA USA
ley006@ucsd.edu

Hakan Töreyin
Electrical and Computer Engineering Department
San Diego State University
San Diego, CA USA
htoreyin@sdsu.edu

***Abstract*— In implantable neural monitoring, handling increasing data volumes from numerous channels is a challenge for transmission. A viable solution is on-chip data spike detection. This study introduces a low-power circuit integrating an analog front-end, spike enhancement filter, and detector. The amplifier adopts a two-stage operational transconductance design to both perform linear filtering of the biopotential recordings and convert them into current. The spike enhancement filter is designed as a current-mode analog signal processing circuit, utilizing translinear loops to emulate the underdamped dynamics of a particle in a monostable potential well, implemented via a second-order differential equation. The filter's output, enhanced with spikes, undergoes a spike detector stage employing hard thresholding. This circuitry is designed using TSMC 65nm CMOS technology. Through simulations utilizing the Wave_clus database, the proposed system demonstrates an average spike detection sensitivity of 98.99% while consuming 311 nW when powered by a 1 V supply, with a compact footprint of 0.0348 mm^2.**

Keywords—spike detection, ultra-low power computing, integrated sensing and computing

I. Introduction

Detecting neural spikes in intracortical neural recordings is essential for various systems utilizing electrodes for neural monitoring, such as brain-machine interfaces [1] and neuroscience experiments [2]. Typically, this process involves conditioning the raw electrical recording through amplification and band-pass filtering to emphasize neural spikes while minimizing noise. Subsequently, spike detection occurs, often followed by spike sorting to categorize spikes based on their originating neurons. Following this, a decoding stage interprets brain activity to deduce intention. Alternatively, intention can be decoded directly from detected spikes, circumventing the spike sorting process [3], [4], [5]. The accuracy and efficiency of these systems, whether employing spike sorting or not, can be impacted by the quantity of detected spikes. Detecting a high volume of spikes provides a comprehensive understanding of neural activity and circuits.

Theoretically, electrodes have the capability to detect spikes from neurons within a range of approximately 140 µm, corresponding to around 1000 neurons in the rat cortex [6]. However, current methodologies are limited to detecting spike events within about 50 µm due to signals falling below the electrode's noise floor [7]. To enhance spike detectability, current amplifier designs focus on minimizing noise, although this comes at the expense of increased power consumption. Moreover, the use of high-density multielectrode arrays (HD-MEAs) is employed to overcome spatial resolution limitations [8], [9], [10], although this leads to heightened hardware complexity and data bandwidth, consequently increasing both size and power consumption [11], [12]. Typically, the transceiver serves as the bottleneck, dissipating significantly more power compared to other components within the system [11]. For instance, in a 1000-channel system with 10-bit resolution and a 20-kSps sampling rate, data transfer requirements would amount to 200-Mbps, while even the best wireless systems currently provide only up to tens of Mbps [13].

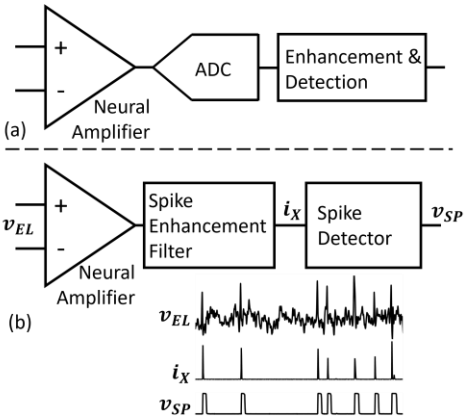

Figure 1. (a) Conventional and (b) proposed spike detection method.

A compelling strategy to reduce on-chip data volume involves implementing on-chip data spike detection. An algorithm for spike enhancement, employing a noise-enhanced filter, has demonstrated effectiveness in improving the signal-to-noise ratio (SNR) of spikes originating from distant neurons and achieving robust spike detection [14]. Notably, this filter capitalizes on stochastic resonance (SR), where optimal noise intensity maximizes SNR enhancement. Consequently, the noise constraint on the amplifier could potentially be alleviated. With this motivation, our study introduces neural signal conditioning electronics that integrate a neural amplifier with the noise-enhanced filtering technique described in [14], utilizing custom-designed analog signal processing tailored specifically for

This work was supported by the National Science Foundation under Grant Number 1916160.

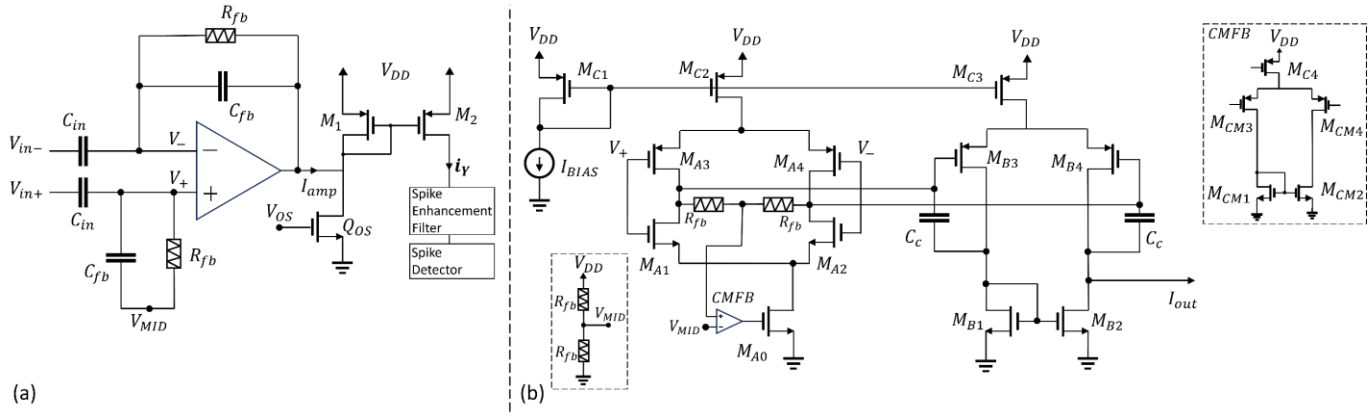

Figure 2. Neural amplifier. (a) Closed-loop circuit. (b) Operational transconductance amplifier schematics.

neural spike detection for the first time. Unlike the conventional approach to spike detection, which involves amplification followed by digitization for subsequent spike detection and inference (Figure 1(a)), the proposed method conducts spike detection through analog domain computation directly on the analog outputs of the neural amplifier (Figure 1(b)).

II. Proposed Integrated Neural Amplifier, Spike Enhancement Filter and Detection Circuit

The system, which integrates signal amplification, spike enhancement filtering, and spike detection stages, is designed using TSMC 65nm CMOS technology.

A. Neural Amplifier

The neural amplifier is designed as a two-stage differential-input single-ended-output operational transconductance amplifier (OTA) with a capacitive-resistive closed-loop feedback system (Figure 2(a)). The closed-loop voltage mid-band gain is set by the ratio C_{in}/C_{fb}. The low-cutoff frequency, f_L, is determined by C_{fb} and feedback resistors implemented as pseudo-resistors, R_{fb}, where two diode-connected PMOS transistors are connected in series. Assuming the dominant pole is at the output, the high-cutoff frequency, f_H, is controlled by the transconductance of the OTA and the load capacitance. The load capacitance is affected by the parasitic capacitances of the input device of the $Q_{M2} - Q_{M3}$ current-mirror and the offset current source device implemented using Q_{OS}. The purpose of $Q_{M2} - Q_{M3}$ and Q_{OS} is to provide the spike enhancement filter with a consistently positive current input, denoted as, i_Y.

The OTA schematic is presented in Figure 2(b). The first stage of the OTA comprises a fully-differential current-reused amplifier with common-mode feedback (CMFB). CMFB is realized using a five-transistor OTA, where the error signal between the output DC level of the first stage and a mid-rail voltage, V_{MID}, is fed back into the circuit through the bias current sink implemented using M_{A0}. This CMFB mechanism facilitates dynamic adjustment of the amplifier's bias voltage, ensuring that all transistors operate within the subthreshold saturation region. The second stage of the OTA utilizes a similar five-transistor architecture with C_C Miller compensation. This compensation network ensures OTA stability and sets the open-loop bandwidth.

TABLE I. Device Sizes and Component Values

OTA stage			
Devices	***W/L (µm)***	***Devices***	***W/L (µm)***
$M_{A0,1,2}$	6.77/0.5	$M_{A3,4}$	25.84/0.4
$M_{B1,2}$	6.28/0.61	$M_{B3,4}$	25.84/0.4
$M_{CM1,2}$	6.28/0.61	$M_{CM3,4}$	25.84/0.4
$M_{C1,4}$	1.8/0.4	$M_{C2,3}$	25.84/0.4
Other components			
Devices	***W/L (µm)***	***Devices***	***Capacitance(fF)***
PMOS pseudo-resistor	0.45/10	C_{in}	2000
		C_{fb}	10
		C_c	580

The mid-rail voltage used in the closed-loop circuit in Figure 2(a) and in the CMFB network in Figure 2(b) is determined as $V_{MID} = V_{DD}/2$, achieved through a voltage divider comprising two identical pseudo-resistors, R_{fb}, connected in series. Device sizes of the amplifier are provided in Table I.

B. Spike Enhancement Filter

The filter used for neural spike pre-emphasis is determined by the following 2nd order differential equation [14]:

$$\frac{d^2x(t)}{dt^2} + \gamma\frac{dx(t)}{dt} = -\frac{dU(x)}{dx} + y(t). \quad (1)$$

The filter governs the dynamics of particle's movement within a potential well, $U(x)$, with a damping factor of γ. Two forces act on the particle on the right-hand side of equation (1); the neural recording, $y(t)$, and a force proportional to the slope of $U(x)$. The output of the filter is represented by $x(t)$. In accordance with findings from [14], we selected a monostable well potential, $U(x) = ax^2$.

For energy-efficient implementation of (1), following [15], we devised a translinear circuit utilizing MOS devices in weak-inversion to solve two 1st order differential equations: $\frac{dz(t)}{dt} = y(t) - ax(t)$ and $\frac{dx(t)}{dt} = z(t) - \gamma x(t)$. The circuit represents the variables, x, y, and z as currents, denoted as i_X, i_Y, and i_Z.

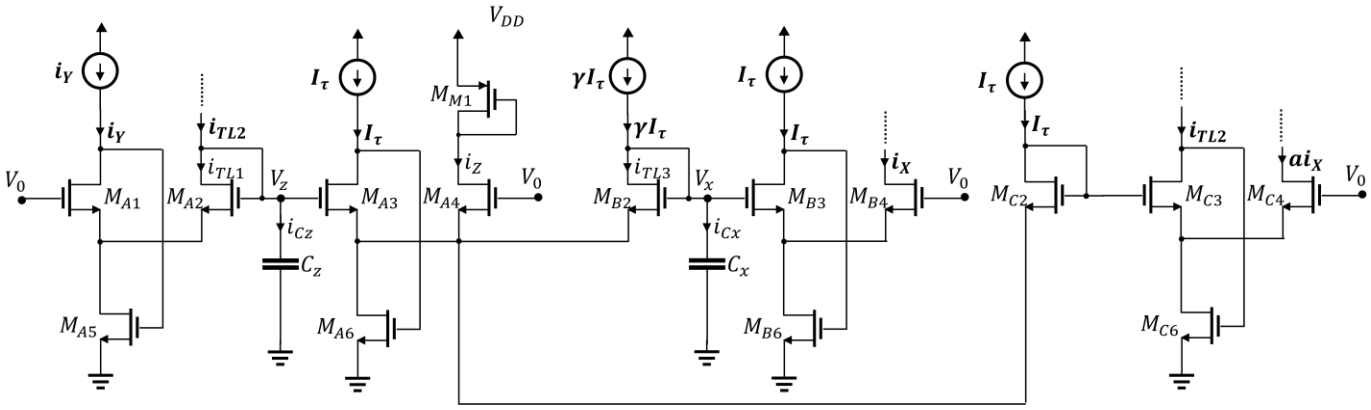

Figure 3. Spike enhancement filter schematics.

By utilizing the relations, $\frac{di_X}{dt} = \frac{di_X}{dv_X}\frac{dv_X}{dt}$ and $\frac{di_Z}{dt} = \frac{di_Z}{dv_Z}\frac{dv_Z}{dt}$, the expressions to implement are as follows:

$$C_z\frac{dv_Z}{dt} + i_{TL1} = i_{TL2}$$
$$C_x\frac{dv_X}{dt} + i_{TL3} = \gamma I_\tau \quad (2)$$

The circuit is presented in Figure 3. In (2), $i_{TL1} = \frac{i_Y I_\tau}{i_Z}$, $i_{TL2} = \frac{a i_X I_\tau}{i_Z}$, $i_{TL3} = \frac{i_Z I_\tau}{i_X}$, and $I_\tau = \frac{CU_T}{\tau\kappa}$, where U_T represents the thermal voltage, κ denotes the gate coupling coefficient in weak inversion, and τ is the time constant. The current summations in (2) occur at the capacitor voltage nodes, v_Z and v_X. The circuit comprises three translinear loops formed by gate-source voltages of four transistor groups; $Q_{A1} - Q_{A4}$ calculate $i_{TL1} = \frac{i_Y I_\tau}{i_Z}$, Q_{A4} and $Q_{B2} - Q_{B4}$ calculate $i_{TL3} = \frac{i_Z I_\tau}{i_X}$, and Q_{A4} and $Q_{C2} - Q_{C4}$ calculate $i_{TL2} = \frac{a i_X I_\tau}{i_Z}$.

The current sources in Figure 3 are realized using pMOS current mirrors. The parameter a is set to unity, and all devices of the filter, except for the current mirror output generating γI_τ, are sized as $8\,\mu m/2\,\mu m$. The value of γ is fixed at 0.25, consequently, the output device of the corresponding mirror is sized as $8\mu m/8\mu m$. The C_Z and C_X are 100 fF. The voltage V_0 is set to 400 mV to ensure the devices remain in saturation within the current range of 1 nA-100 nA.

C. Spike Detector

Spike detection is performed through hard thresholding. The spike detector comprises a current comparator followed by two inverters to refine the output signal (Figure 4). The current comparator is composed of two current mirrors copying the filter output, i_X, and a threshold current, I_{TH}, onto the same output branch. If I_{TH} exceeds (falls below) i_X, the output voltage, V_{comp}, is pulled to a low (high) voltage by M_{CC2} (M_{CC3}) [16]. The V_{comp} is then fed into two inverters connected back-to-back, resulting in spike events being identified as a pulse train.

III. Simulation Results and Discussion

The proposed neural signal conditioning system schematics is simulated using Cadence Virtuoso. The performance results

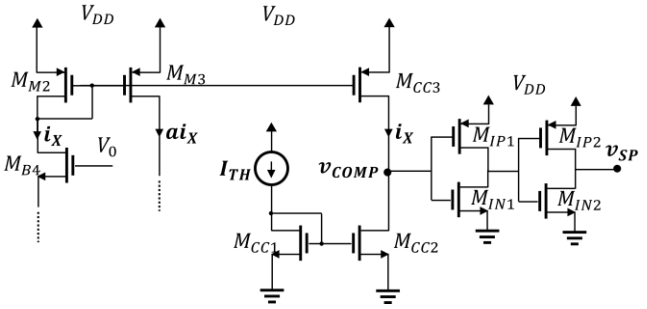

Figure 4. Spike detector circuit.

TABLE II. Amplifier Performance

Supply (V)	1.0
Voltage to current gain (dB)	-87.7
Power consumption (nW)	216.26
Bandwidth (Hz)	2.8-7.9k
Input-referred noise (μV)	13.9
Input impedance@500Hz (MΩ)	338.5
CMRR (dB)	55
PSRR (dB)	47.5

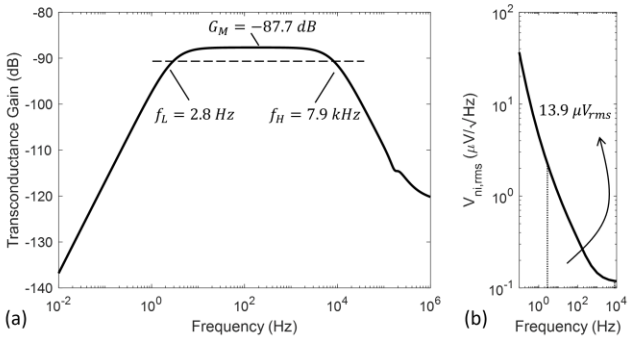

Figure 5. Closed-loop gain and noise of the neural amplifier.

of the amplifier are summarized in Table II. Both stages are biased with 100 nA, and the CMFB amplifier is biased with 10 nA, resulting in a power dissipation of 216 nW. The closed-loop amplifier achieves a bandwidth ranging from 2.8 Hz to 7.9 kHz with a transconductance gain of 41.2 nA/mV (Figure 5(a)). The simulated common-mode gain is 73.2 pA/mV. Additionally, the

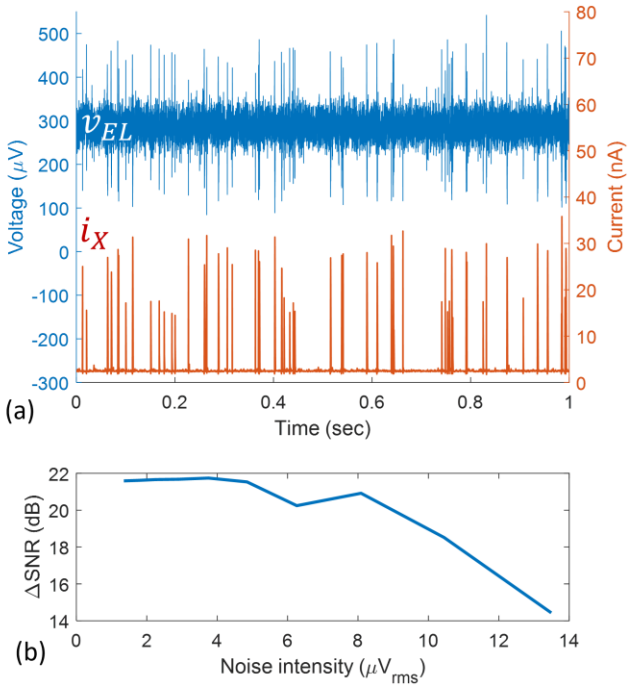

Figure 6. Spike enhancement of the proposed filter. (a) Example waveforms of the amplifier input, v_{EL}, and the filter output, i_X. (b) ΔSNR for various input-referred noise intensities.

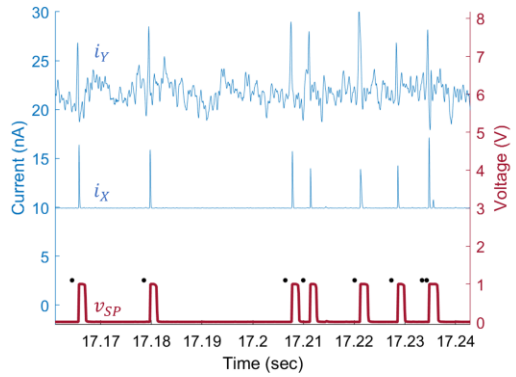

Figure 7. Example waveforms from the filter and the detector. The black dots indicate the true spike events of the recording.

integrated input-referred noise over the closed-loop bandwidth is simulated as 13.9 μV_{rms} (Figure 5(b)). The spike enhancement and the detector dissipate 95 nW of power.

To evaluate the spike detection performance, the neural amplifier is provided with neural recordings from the Wave_clus synthetic extracellular dataset [17]. Each recording within the dataset consists of 60-second samples generated at a sampling rate of 24 kHz. Actual in vivo spike waveforms from three different neurons are utilized in each recording. The noise levels in the recording names (e.g., 0.05, 0.2) indicate the noise standard deviation, calculated as the ratio of spike amplitude to the noise standard deviation. The recordings are further infused with the amplifier's integrated input-referred noise, modeled as white noise, using MATLAB. In Figure 6(a), the waveforms illustrate a one-second portion of the recording "Easy 1 – noise 0.05" with added noise, fed to the amplifier, alongside the corresponding output of the filter.

TABLE III. SPIKE DETECTION PERFORMANCE

Recording Noise		*# of Spikes*	*#FN+FP*	*Se (%)*	*Acc (%)*
Easy 1	0.05	3514	9	99.77	99.74
	0.1	3522	6	99.86	99.83
	0.15	3477	38	99.71	98.92
	0.2	3474	293	96.17	91.94
Easy 2	0.05	3410	3	99.94	99.91
	0.1	3520	2	99.97	99.94
	0.15	3411	86	98.83	97.51
	0.2	3526	349	93.85	90.46
Difficult 1	0.05	3383	1	100	99.97
	0.1	3448	1	100	99.97
	0.15	3472	2	100	99.94
	0.2	3414	37	99.36	98.91
Difficult 2	0.05	3364	1	100	99.97
	0.1	3462	3	99.94	99.91
	0.15	3440	34	99.59	99.01
	0.2	3493	209	96.91	94.18

The SNR enhancement, ΔSNR, of the filter is assessed by computing the SNR as the mean spike peak-to-peak amplitude averaged over 20 arbitrarily-selected spikes, divided by the root-mean-square (RMS) value of the noise obtained as an average RMS of 20 arbitrarily-selected noise-only segments of the recordings. Figure 6(b) presents ΔSNR for various noise intensities. Notably, for 4 μV_{rms} intensity, ΔSNR exhibits a slight improvement with the proposed filter, indicating the facilitation of stochastic resonance.

Threshold is selected as 1 nA greater than the baseline level of the filter output, I_x. The simulated threshold detector output waveforms are further processed in MATLAB to identify the true positives (TP), false positives (FP), and false negatives (FN) to calculate the sensitivity as $Se = TP/(TP + FN)$, accuracy as $Acc = TP/(TP + FP + FN)$, and the sum $FN + FP$. Figure 7 shows the input of the spike enhancement filter and the outputs of both the filter and the detector for 80-millisecond segment of the recording "Easy 1 – noise 0.05".

The detection performance for all recordings is summarized in Table III. The circuit achieves an average sensitivity of 98.99±1.79%, an average accuracy of 98.13±3.09%, and an average $FN + FP$ of 67.13±112.77. When compared with the numerical implementation of the filter in [14], the average sensitivity is slightly lower (99.97% vs. 98.99%). This discrepancy could potentially be attributed to the implementation of the damping parameter γ. Unlike the filter in [14], which incorporates an adaptive damping constant adjustment based on the signal amplitude, the proposed circuit employs a constant damping term.

The chip layout is presented in Figure 8. A comparison of the proposed circuit with previous works presenting simulation-based spike detection performances is detailed in Table IV. Regarding detection performance, the proposed method outperforms [18] and exhibits comparable performance to [19]. It's also important to note that the studies in [18], [19] utilize

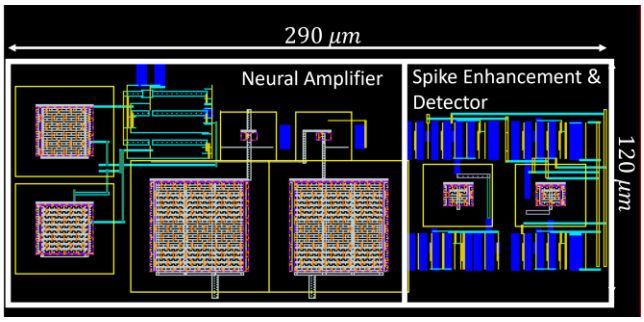

Figure 8. Chip layout.

different databases. The proposed circuit dissipates approximately one order of magnitude less power than [20]. While the study in [18] achieves lower power and area, it's worth mentioning that it does not include an analog front-end amplifier stage. In fact, excluding the amplifier, the proposed spike enhancement and detector dissipate less power (0.095 μW) within a slightly larger area (0.0077 mm^2) than [18].

IV. Conclusion

This paper introduces an integrated analog front-end amplifier and spike detection circuitry, achieving real-time spike detection with ultra-low power consumption in a compact footprint. The high efficiency of this approach is attributed to the significant SNR enhancement provided by the spike enhancement filter, which employs noise-enhanced spike enhancement. Consequently, reducing the power dissipation becomes feasible due to the lowered noise constraint on the neural amplifier. Notably, the system demonstrates robust spike detection performance, as evidenced by its high sensitivity and accuracy on a benchmark database. These simulation outcomes underscore the potential of the proposed system for integration as the analog front-end and spike detector stage in implantable neural monitoring applications, particularly those that directly infer spike events, bypassing the need for a sorting stage.

TABLE IV. Comparison with Previous Works

	This Work Simulated	***TBCAS23 Simulated [18]***	***BioCAS23 Simulated [19]***	***ISCAS22 Simulated [20]***
Technology	65 nm	180 nm	40 nm	65 nm
Integrated Amplifier	Yes	No	No	Yes
Power (μW/channel)	AFE: 0.216 SP: 0.095	AFE: N/A SP: 0.28	AFE: N/A SP: 0.95	AFE: 2.55 SP: 0.6
Area (mm^2/channel)	AFE: 0.021 SP: 0.0077	AFE: N/A SP: 0.0067	AFE: N/A SP: N/A	AFE:0.0027 SP:0.001
Detection Accuracy	98.13%	96%	90.9%-100%	N/A

AFE: Analog front-end block, SP: Spike enhancement and detector block

References

[1] J. Dai *et al.*, "Reliability of motor and sensory neural decoding by threshold crossings for intracortical brain–machine interface," *J. Neural Eng.*, vol. 16, no. 3, p. 036011, Apr. 2019, doi: 10.1088/1741-2552/ab0bfb.

[2] G. Buzsáki, "Large-scale recording of neuronal ensembles," *Nat Neurosci*, vol. 7, no. 5, Art. no. 5, May 2004, doi: 10.1038/nn1233.

[3] G. W. Fraser, S. M. Chase, A. Whitford, and A. B. Schwartz, "Control of a brain–computer interface without spike sorting," *J. Neural Eng.*, vol. 6, no. 5, p. 055004, Sep. 2009, doi: 10.1088/1741-2560/6/5/055004.

[4] E. M. Trautmann *et al.*, "Accurate Estimation of Neural Population Dynamics without Spike Sorting," *Neuron*, vol. 103, no. 2, pp. 292-308.e4, Jul. 2019, doi: 10.1016/j.neuron.2019.05.003.

[5] B. P. Christie *et al.*, "Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain–machine interface performance," *J. Neural Eng.*, vol. 12, no. 1, p. 016009, Dec. 2014, doi: 10.1088/1741-2560/12/1/016009.

[6] D. A. Henze, Z. Borhegyi, J. Csicsvari, A. Mamiya, K. D. Harris, and G. Buzsáki, "Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo," *Journal of Neurophysiology*, vol. 84, no. 1, pp. 390–400, Jul. 2000, doi: 10.1152/jn.2000.84.1.390.

[7] C. Pedreira, J. Martinez, M. J. Ison, and R. Quian Quiroga, "How many neurons can we see with current spike sorting algorithms?," *Journal of Neuroscience Methods*, vol. 211, no. 1, pp. 58–65, Oct. 2012, doi: 10.1016/j.jneumeth.2012.07.010.

[8] A. Obaid *et al.*, "Massively parallel microwire arrays integrated with CMOS chips for neural recording," *Science Advances*, vol. 6, no. 12, p. eaay2789, Mar. 2020, doi: 10.1126/sciadv.aay2789.

[9] "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings." Accessed: Jul. 22, 2022. [Online]. Available: https://www.science.org/doi/full/10.1126/science.abf4588

[10] A. Nurmikko, "Challenges for Large-Scale Cortical Interfaces," *Neuron*, vol. 108, no. 2, pp. 259–269, Oct. 2020, doi: 10.1016/j.neuron.2020.10.015.

[11] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and Analysis of a Hardware-Efficient Compressed Sensing Architecture for Data Compression in Wireless Sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012, doi: 10.1109/JSSC.2011.2179451.

[12] M. Shaeri, A. Afzal, and M. Shoaran, "Challenges and Opportunities of Edge AI for Next-Generation Implantable BMIs." arXiv, Apr. 13, 2022. doi: 10.48550/arXiv.2204.02362.

[13] A. Uran *et al.*, "A 16-Channel Neural Recording System-on-Chip With CHT Feature Extraction Processor in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, pp. 1–1, 2022, doi: 10.1109/JSSC.2022.3161296.

[14] C. B. Güngör, P. P. Mercier, and H. Töreyin, "Investigating well potential parameters on neural spike enhancement in a stochastic-resonance pre-emphasis algorithm," *J. Neural Eng.*, vol. 18, no. 4, p. 046062, May 2021, doi: 10.1088/1741-2552/abfd0f.

[15] B. A. Minch, "Synthesis of static and dynamic multiple-input translinear element networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 2, pp. 409–421, Feb. 2004, doi: 10.1109/TCSI.2003.822557.

[16] C. B. Güngör, P. P. Mercier, and H. Töreyin, "A 2.2 nW Analog Electrocardiogram Processor Based on Stochastic Resonance Achieving a 99.94% QRS Complex Detection Sensitivity," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 17, no. 1, pp. 33–44, Feb. 2023, doi: 10.1109/TBCAS.2023.3235786.

[17] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering," *Neural Computation*, vol. 16, no. 8, pp. 1661–1687, Aug. 2004, doi: 10.1162/089976604774201631. https://figshare.le.ac.uk/articles/dataset/Simulated_dataset/11897595/1

[18] Z. Zhang, P. Feng, A. Oprea, and T. G. Constandinou, "Calibration-Free and Hardware-Efficient Neural Spike Detection for Brain Machine Interfaces," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 17, no. 4, pp. 725–740, Aug. 2023, doi: 10.1109/TBCAS.2023.3278531.

[19] J. Chen *et al.*, "A Clockless Robust Bionic Spike Detector for Implantable Brain-Computer Interfaces," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2023, pp. 1–5. doi: 10.1109/BioCAS58349.2023.10388732.

[20] T. Guo, H. Liu, and Y. Liu, "An Ultra Compact Neural Front-End with CT-NEO Based Spike Detection for Implantable Applications," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2022, pp. 2826–2830. doi: 10.1109/ISCAS48785.2022.9937847.