15.5 Event-Based Spatially Zooming Neural Interface IC with 10nW/Input Reconfigurable-Inverter Fabric and Input-Adaptive Quantization

Jianxiong Xu[1], Mustafa Kanchwala[1], Mohammad Abdolrazzaghi[1], Hanfeng Cai[1], Yu Huang[1], Junyu Ma[1], Chae Lim[1], Lingyun Xu[1], Shucheng Gong[1], Weian Deng[1], Qiaosong Deng[1], Jin Che[1], Sudip Nag[1], Joshua Olorocisimo[1], Rhianna Singh[1], Yanze Wang[1], Jose Sales Filho[1], Mandana Mohaved[2], Homeira Moradi[2], George Eleftheriades[1], Taufik Valiante[2,3], Roman Genov[1]

[1]University of Toronto, Toronto, Canada
[2]Krembil Neuroscience Center, Toronto, Canada
[3]Toronto Western Hospital, Toronto, Canada

Large-scale neural interface ICs have tens of thousands of electrodes [1-3], enabling a wide range of applications including neural prostheses and therapeutic neuromodulation. However, the human brain contains 86 billion neurons and new frontiers in brain interfacing, such as understanding memory and cognition, will benefit from concurrent access to a million or more of implanted electrodes [4]. Modern microfabrication technologies, including silicon wafer thinning [1], allow for dense co-integration of electrodes and transistors on the same flexible substrate [1,4-6] and overcome the issue of the mega-scale electrode interconnect bottleneck [7]. The key remaining challenges are the low energy efficiency of neural ADCs and the high output data rate [4]. For example, one million inputs require 10nW/input ADC power - for a tissue-safe 10mW ADC total power budget [8], and a 200Gb/s wireless link - for 8b conversion at 25kHz [4]. However, the power dissipation of neural ADCs, either dedicated [9-12] or time-multiplexed [1-3,13], is over 50× higher, and implantable radios are at least 100× slower [14-15]. To address these challenges, neural spiking sparsity has been exploited in both off-line [16-17] and on-line [8,18] methods of optimum electrode selection, but this leads to significant losses in recorded information [17] and requires near-μW/input power due to static circuit biasing [8,18], respectively.

We present a 64-input event-based implantable neural interface illustrated in Fig. 15.5.1 that takes advantage of neural spikes sparsity in space, time and amplitude probability to perform: (1) on-line low-SNR event detection that triggers high-SNR event recording by reconfiguring the same ADC circuits, which we refer to as 'spatial zooming', (2) non-uniform (NU) continuous-time (CT) sampling, and (3) NU CT quantization, respectively. The inverter-based dynamically biased analog front-end (AFE) dissipates 10nW/input in the spike-detection mode (mode 1), in the spike-recording mode (mode 2), or in the artifact-tolerance mode (mode 3), and inherently reduces the output data rate by over 100×. This enables wireless powering by a glasses-mounted steerable-beam antenna array, for the application of responsive neuromodulation for memory enhancement/restoration.

Figure 15.5.1 introduces the concept of spatial zooming. In its default mode, mode 1 (Fig. 15.5.1, left), the IC detects spikes on all 64 electrodes using only one high-gain cascoded inverter (INV) per input, configured as a CT comparator dissipating 4nW for a practical spike rate of 25/sec. The input capacitor is precharged to the comparator threshold voltage $TH_n=V_{ref}-V_n$ and is inserted into the signal path, so that the difference $V_{in}-TH_n$ is presented to the INV for comparison, where n=1,…,64, and V_{ref} is the reference electrode voltage. Voltage V_n is adaptively set by a low-duty-cycle DAC (0.001%, 50ns ON time) based on previous spikes morphology. When a spike is detected, the IC dynamically switches to mode 2 – the spike recording mode (Fig. 15.5.1, center). In this mode, the 64 inverters are regrouped into four subsets of 16 inverters each, to form four CT flash ADCs, each with N=16 uniformly distributed thresholds TH_n. These four ADCs are time-multiplexed among all electrodes where spikes were detected (up to all 64). For a realistic spatial spike sparsity of 1/32, power dissipation in mode 2 is also 4nW/electrode. In mode 3 (Fig. 15.5.1, right), the IC uses two subsets of 32 inverters each, to implement two CT subranging ADCs. This increases the input range from 15mV to 200mV as needed to tolerate infrequent but large artifacts. Each ADC is implemented as two 16-inverter CT flash ADC stages, coarse and fine, for a total of N = 16×16 = 256 quantization levels. In the first stage, the INV with the input voltage closest to its TH_n is dynamically configured to act as a residue amplifier, while the other 15 INVs act as comparators that are saturated and do not pass direct-path current. Thus, each narrow-input-range INV also acts as a duty-cycled residue amplifier, with a shifted transfer characteristic, that amplifies only a portion of the input signal, resulting in a rail-to-rail input range without the corresponding power penalty. Assuming that 120Hz stimulation takes place less than 1% of the time, mode-3 power dissipation is below 3.5nW/electrode. In modes 2 and 3, the use of NU CT sampling and quantization boosts the ENOB of the ADCs from the nominal resolution values of $\log_2 16$=4b and $\log_2 256$=8b, to approximately 8b and 14b, respectively (by factors of α≈4b and β≈6b, respectively) as discussed next.

Figure 15.5.2 illustrates how NU CT sampling and quantization are implemented in both modes 2 and 3. As shown in Fig. 15.5.2 (top), in contrast to conventional neural ADCs, we perform NU sampling - only when the input signal waveform crosses a quantization level, reducing power. NU sampling prevents the folding of the quantization noise into the baseband [19-20], boosting the ADC ENOB far beyond the conventional $\log_2 N$ resolution – by ~3.5b and ~5b in modes 2 and 3, respectively. It also eliminates inherently imprecise and area-inefficient analog anti-aliasing filters and allows for compact precise digital filters [21]. As depicted in Fig. 15.5.2 (middle), the ADC also periodically evaluates the input's probability density function (PDF) and dynamically adjusts its quantization levels to enhance precision for relevant levels, reducing the quantization noise [22]. This adds ~0.5b and ~1b of ADC resolution in modes 2 and 3, respectively. Figure 15.5.2 (bottom) shows the improved ENOB/SNDR and power scalability of the design.

Figure 15.5.3 depicts the neural ADC schematic and its experimental results. The mode of each INV is dynamically selected based on the ADC output. The circuits highlighted in green and red are OFF in modes 1 and 2 but are ON in mode 3. In modes 1 and 2, each INV functions as a CT comparator. As opposed to energy inefficient constant-current biasing of OTA-based CT comparators, the INV employs dynamic self-biasing. During phase P_1, capacitor C_1 is pre-charged to the comparator threshold voltage $TH_n=V_{ref}-V_n$. Concurrently, the INV undergoes auto-zeroing, as shown in Fig. 15.5.3 (top, right). The upper and lower segments of the INV form current sources which set all biasing voltages. Additionally, C_2 and C_3 sample the low-frequency (LF) noise, which is nullified during P_2. In P_2, when V_{in} is near V_{cm}, pre-charged nodes V_1 and V_4 bias the input transistors into the subthreshold region to reduce the direct-path current. If V_{in} deviates from V_{cm} by >60μV, M_1 or M_2 cut off the direct-path current, further saving energy. In mode 3, the green block initiates a negative feedback loop to amplify the residual voltage for the fine ADC stage, and the red block establishes a positive feedback loop to enhance Z_{in}. In mode 1, a threshold (TH) adaptation circuit modulates V_n for optimum detection of spikes. In modes 2 and 3, a DC servo loop [10] suppresses LF noise. It also boosts the input dynamic range (DR) in mode 3.

Figure 15.5.4 illustrates the digital ADC mode selection. First, a spike is detected on input A. It is then recorded in mode 2, while periodic fast switching to mode 1 monitors for spikes on other electrodes. Other spikes are then detected on inputs B-C, which are also then switched to mode-2 readout. The neural activity on input D is a large artifact, beyond 15mV, so that input is switched to the 200mV-range mode 3 readout to avoid the saturation dead-time. All other electrodes continue to be monitored for spikes by periodic fast switching to mode 1. Input D returns to mode 2 and then to mode 1, when it drops below 15mV and then below the spike's TH, respectively.

Figure 15.5.5 presents the system block diagram and key results for the application of responsive neuromodulation for memory enhancement/restoration. Non-uniformly sampled ADC outputs are mapped to an address representation [18] and time-synchronized by an arbiter to prevent data collision. Arbitrated data are sent to a wearable control hub by a 915MHz TX with both pulse-width- [23] and pulse-position-modulated XOR-based [24] PA with high efficiency of 8.6pJ/b [12,25]. The maximum data rate is 52Mb/s at 440μW, scalable to 0.2Mb/s at ~1.7μW (BER=9×10^{-5}). Upon receipt, the hub decodes the address and recovers data. A digital anti-aliasing filter reconstructs these non-uniformly sampled data into the uniform format, followed by in-the-loop local or remote digital inference. For local inference, spike sorting is first performed to get the neural spike counts during memory encoding and recognition events. These spike counts are then processed by eXtreme Gradient Boosted ensemble of decision trees (XGBoost) – a leading algorithm for BMI [26] – to classify if the declared patient response is correct. XGBoost predicts memory recall with a 92.9% mean accuracy on a human single-neuron activity dataset [27]. A stimulation burst can then be triggered during an encoding event to enhance the patient's memory, which has shown to improve retention and recall [28-29]. The chip is wirelessly powered by a 2.8cm 4-element steerable-beam antenna array built into smart glasses temple, transmitting up to 1W at 915MHz. The 15×15mm^2 implantable RX coil within a human head phantom receives up to 400μW at a depth of 5cm. The phased array tracks the implant's position using backscattered data on the received power [30].

Figure 15.5.6 depicts extracellular activity recorded by the neural ADC from the mouse brain in the three modes. Figure 15.5.7 includes the comparison table, chip micrograph and mini-PCB carrier prototype. This principally digital design is well suited for scaling to advanced technology nodes for further integration and power savings.

2025 IEEE International Solid-State Circuits Conference (ISSCC) | 979-8-3315-4101-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISSCC49661.2025.10904733

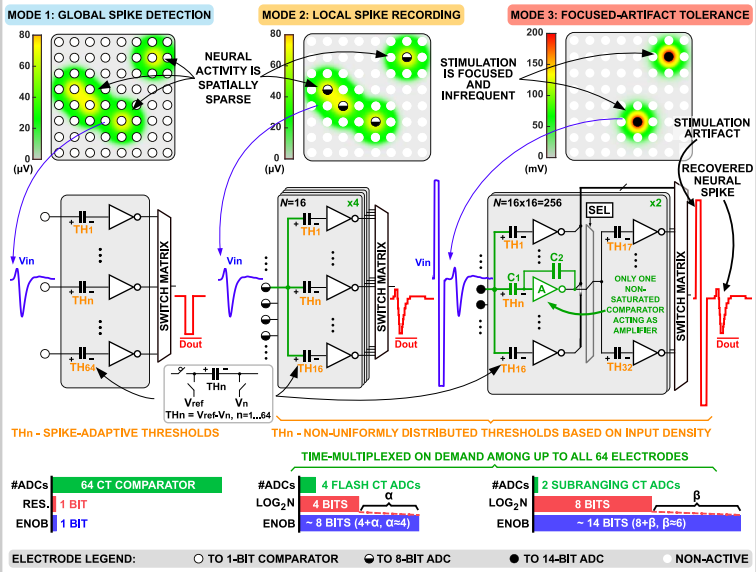


Figure 15.5.1: Spatially zooming neural ADC architecture and its implementation on a continuous-time (CT) reconfigurable-inverter fabric.

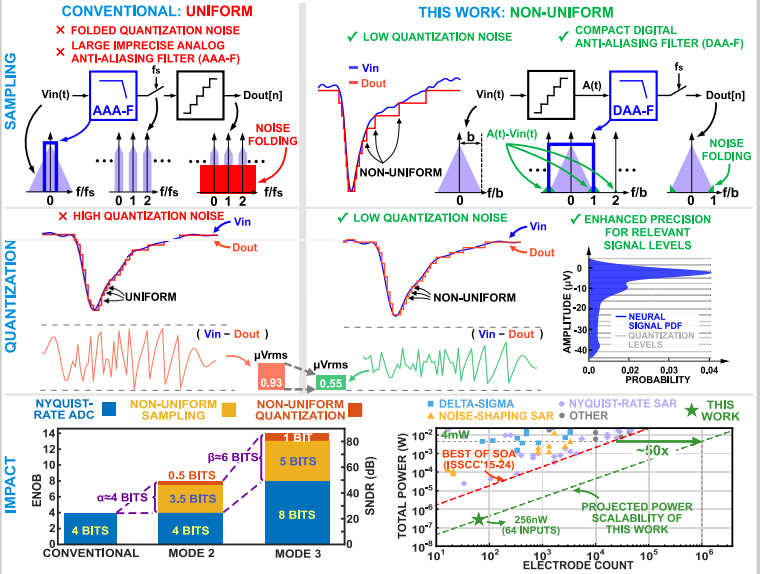


Figure 15.5.2: Non-uniform continuous-time sampling and quantization in the spatially zooming neural ADC.

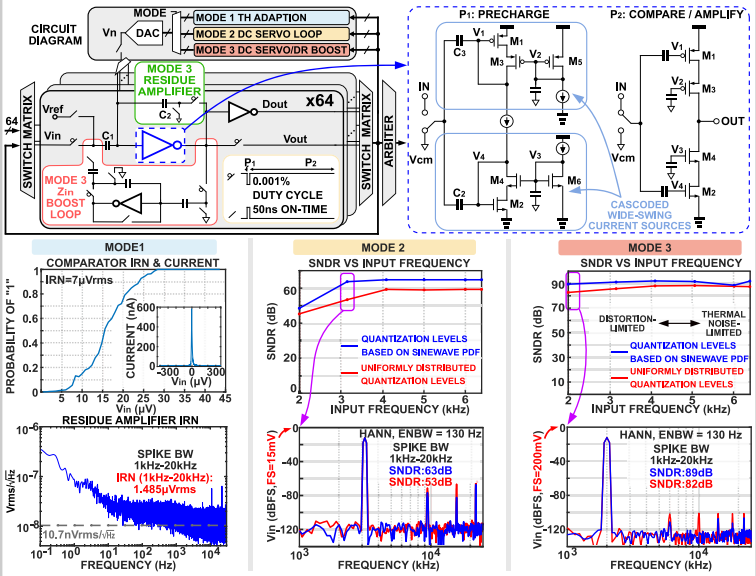


Figure 15.5.3: Spatially zooming neural ADC schematic and experimentally measured results.

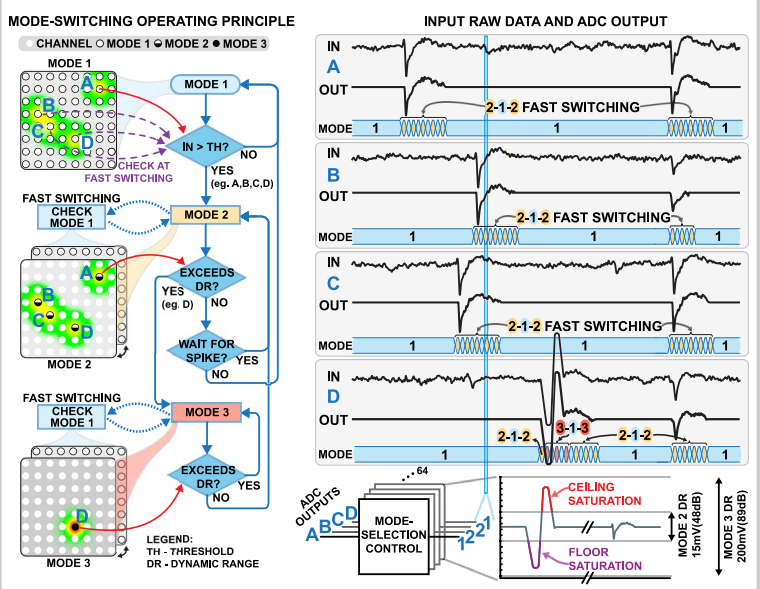


Figure 15.5.4: Dynamic mode selection control in the spatially zooming ADC.

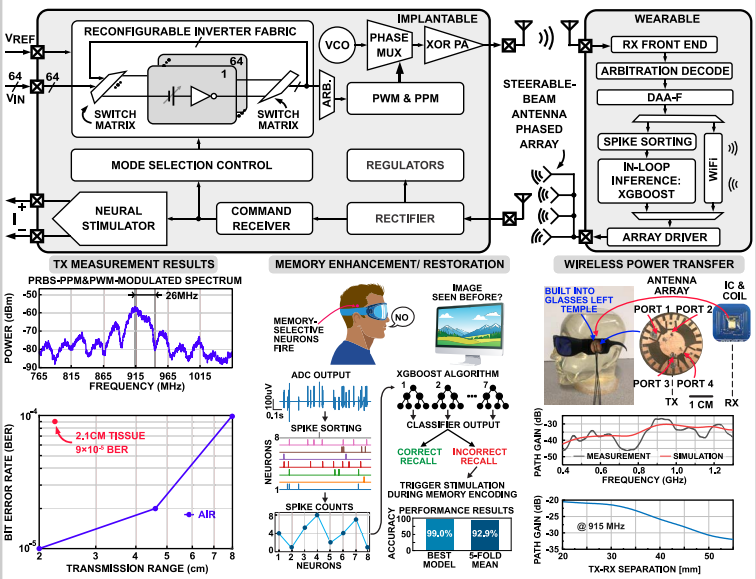


Figure 15.5.5: System block diagram for the application of wireless closed-loop neuromodulation and experimental results.

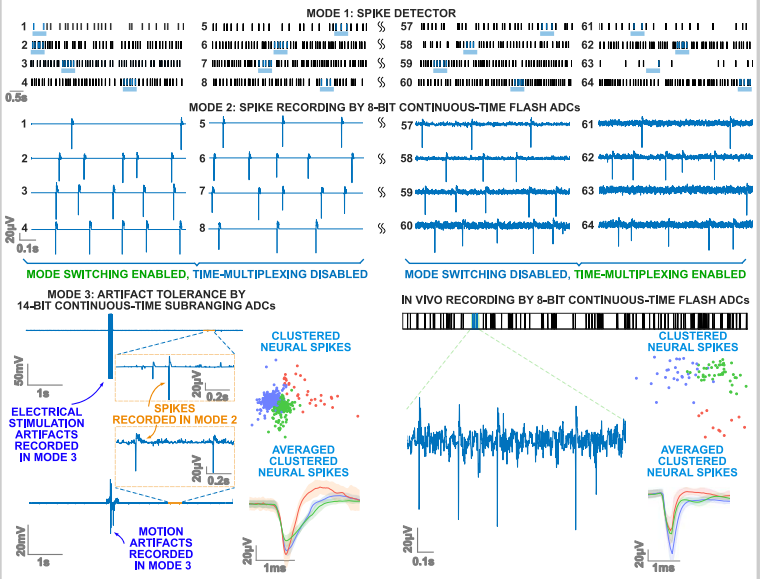
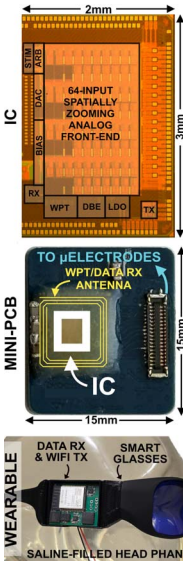


Figure 15.5.6: Experimentally measured results from the mouse brain in vitro (4AP-induced spikes) and in vivo (naturally occurring spikes).



		THIS WORK			[18]	[8]	[1]	[3]	[13]	[12]
	EVENT-BASED	YES					NO			
IC	TECHNOLOGY (nm)	65			180		130		65	
IC	SUPPLY VOLTAGE (V)	0.7 - 3.3			1.8		1.5, ±2.5	-	1.2	0.8
IC	AREA (mm²)	6			25	11.8	144	-	8.2	2
	PROCESSING METHOD	CONTINUOUS-TIME (ASYNCHRONOUS)			DISCRETE-TIME					
	SIMULTANEOUS CH. #	64			4096	12	256	1024	192	1
	TEMPORAL RESOL. (ns)	10^a			<1000[b]	-	29000[b]	33333[b]	50000[b]	48077[b]
	FUNCTIONALITY	MODE 1: SPIKE DETECTOR	MODE 2: SPIKE RECORDING	MODE 3: ARTIFACT TOLERANCE	SPIKEs		SPIKEs		LFPs & SPIKEs	LFPs
ADC	ADC	ADAPTIVE COMPAR.	CT FLASH	CT SUB-RANGING	LEVEL CROSSING	SAR	SAR			
ADC	ENOB (BITS)	1	7.68 (=8)	14.49 (=14)	6[e]	12	10			9.7
ADC	ARTIFACT TOL. (mV)	N/A	15	200	-		-			
ADC	IRN (µVrms) /BW (Hz)	7 / 1k - 20k	7.3 / 1k - 20k	1.48 / 1k - 20k	19.04 / 500-3k	3.41 / 10-10k	4.3 / 10-4k	7.5 / 300-10k	8.8 / 300-10k	6.77 / 0.9-6.9k
ADC	PWR / ELECTRODE (µW)	0.004[c]	0.004[c,h] / 0.3[d]	0.003[g] / 7.1[d]	0.8	3.6	-	-	3.44	88
ADC	BW / ELECTRODE (kHz)	20			10	14	4	4	10	6.9
ADC	NEF	N/A	1.1[d]	1.1[d]	-	1.86	-	-	-	6.9
ADC	SNDR (dB)		48[d]	89[d]	37.2	-				61[e]
ADC	FOMs (dB)		156[d]	183[d]	-					-
ADC	FOMw (fJ/CONV.STEP)		38[d]	8[d]		78.6				
ADC	HP CORNER (Hz)	1 - 2000[f]			5	5.7	5	0.5/300		0.9
ADC	INPUT IMPEDANCE (GΩ)	>1 @ 2kHz			-	-	0.1	>1		
WPT	POWER LINK TYPE	RF			N/A		IND.	N/A		RF
WPT	DISTANCE(CM)	5					1.52			3
WPT	FREQUENCY(MHz)	915					13.56			408
WPT	LINK GAIN (dB)	-29					-14			-28
TX	DATA LINK TYPE	915 MHz PWM & PPM					UWB	N/A		B-OOK
TX	ENERGY/Bit (pJ)	8.4					50			140
TX	MAX. DATA RATE (Mbps)	52					108			0.25
TX	MAX. POWER (µW)	440					38800			35
TX	RANGE (CM) / BER	8 / 9x10[-5]					-			3 / 8x10[-6]
	DBE PWR / ELECTRODE (µW)	0.006[i]					N/A			
	MAX. STIM CURRENT (mA)	1					1	-	N/A	

BOLD text denotes best performance
N/A : Not applicable
-: Not reported
FOMw (fJ/Conv.Step) = POWER / (2^{ENOB}×2BW)
FOMs (dB) = SNDR + $10\log_{10}$ (BW / POWER)
CT : Continuous-time
LFP : Local field potential
a : The smallest detectable time interval
b : 1 / sampling frequency
c : For an average spike rate of 25 spikes/second (BW = 20kHz)
d : For 2kHz sinusoidal wave input (BW = 20kHz)
e : Estimated
f : Digitally programmable DC servo loop for HPF
g : For 120Hz stimulation 1% of the time
h : For 1/32 spatial spike sparsity
i : Switching power

Figure 15.5.7: Comparison table, chip micrograph, and wireless power/data TX/RX embodiment prototypes.

References:

[1] N. Zeng, et al., "A Wireless, Mechanically Flexible, 25µm-Thick, 65,536-Channel Subdural Surface Recording and Stimulating Microelectrode Array with Integrated Antennas," *IEEE Symp. VLSI Technology and Circuits*, June 2023.

[2] D. Tsai, et al., "A very large-scale microelectrode array for cellular-resolution electrophysiology," *Nat. Commun.*, vol. 8, no. 1, p. 1802, Nov. 2017.

[3] C. M. Lopez, et al., "A 16384-electrode 1024-channel multimodal CMOS MEA for high-throughput intracellular action potential measurements and impedance spectroscopy in drug-screening applications," *ISSCC*, pp. 464-465, Feb. 2018.

[4] D. Kleinfeld, et al., "Can One Concurrently Record Electrical Spikes from Every Neuron in a Mammalian Brain?," *Neuron*, vol. 103, no. 6, pp. 1005-1015, Sept. 2019.

[5] X. Huang, et al., "A 256-Channel Actively-Multiplexed µECoG Implant with Column-Parallel Incremental ΔΣADCs Employing Bulk-DACs in 22-nm FDSOI Technology," in *ISSCC*, pp. 200-201, Feb. 2022.

[6] C. Wang, et al., "Extremely Bendable, High-Performance Integrated Circuits Using Semiconducting Carbon Nanotube Networks for Digital, Analog, and Radio-Frequency Applications," *Nano Lett.*, vol. 12, no. 3, pp. 1527-1533, Mar. 2012.

[7] X. Liu, et al., "Flexible high-density microelectrode arrays for closed-loop brain-machine interfaces: a review," *Front. Neurosci.*, vol. 18, p. 1348434, Apr. 2024.

[8] H.-S. Lee, et al., "A Multi-Channel Neural Recording System With Neural Spike Scan and Adaptive Electrode Selection for High-Density Neural Interface," *IEEE TCASI*, vol. 70, no. 7, pp. 2844-2857, July 2023.

[9] M. Reza Pazhouhandeh, et al., "Track-and-Zoom Neural Analog-to-Digital Converter With Blind Stimulation Artifact Rejection," *IEEE JSSC*, vol. 55, no. 7, pp. 1984-1997, July 2020.

[10] R. Muller, et al., "A 0.013mm², 5µW, DC-Coupled Neural Signal Acquisition IC With 0.5 V Supply," *IEEE JSSC*, vol. 47, no. 1, pp. 232-243, Jan. 2012.

[11] U. Shin, et al., "A 16-Channel 60µW Neural Synchrony Processor for Multi-Mode Phase-Locked Neurostimulation," *IEEE CICC*, Apr. 2022.

[12] Y. Zhang, et al., "An 8-Shaped Antenna-Based Battery-Free Neural-Recording System Featuring 3 cm Reading Range and 140 pJ/bit Energy Efficiency," *IEEE JSSC*, vol. 58, no. 11, pp. 3194-3206, Nov. 2023.

[13] M.A. Shaeri, et al., "MiBMI: A 192/512-Channel 2.46mm² Miniaturized Brain-Machine Interface Chipset Enabling 31-Class Brain-to-Text Conversion Through Distinctive Neural Codes," *ISSCC*, pp. 546-547, Feb. 2024.

[14] V. Valente, "Evolution of Biotelemetry in Medical Devices: From Radio Pills to mm-Scale Implants," *IEEE TBioCAS*, vol. 16, no. 4, pp. 580-599, Aug. 2022.

[15] R. Eskandari and M. Sawan, "Challenges and Perspectives on Impulse Radio-Ultra-Wideband Transceivers for Neural Recording Applications," *IEEE TBioCAS*, vol. 18, no. 2, pp. 369-382, Apr. 2024.

[16] K.V. Saboo, et al., "Unsupervised machine-learning classification of electrophysiologically active electrodes during human cognitive task performance," *Sci. Rep.*, vol. 9, no. 1, p. 17390, Nov. 2019.

[17] J. Choi, et al., "Optimal Adaptive Electrode Selection to Maximize Simultaneously Recorded Neuron Yield," *Advances in Neural Information Processing Systems*, Oct. 2020.

[18] M. Cartiglia, et al., "A 4096 channel event-based multielectrode array with asynchronous outputs compatible with neuromorphic processors," *Nat. Commun.*, vol. 15, article no. 7163, 2024.

[19] J. Van Assche and G. Gielen, "A 10.4-ENOB 0.92-5.38 µW Event-Driven Level-Crossing ADC with Adaptive Clocking for Time-Sparse Edge Applications," *ESSCIRC*, pp. 261-264, Sep. 2022.

[20] B. Schell and Y. Tsividis, "A Clockless ADC/DSP/DAC System with Activity-Dependent Power Dissipation and No Aliasing," ISSCC, pp. 550-551, Feb. 2008.

[21] T.-F. Wu, et al., "A Nonuniform Sampling ADC Architecture With Reconfigurable Digital Anti-Aliasing Filter," *IEEE TCASI*, vol. 63, no. 10, pp. 1639-1651, Oct. 2016.

[22] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129-137, Mar. 1982.

[23] K. Cho and R. Gharpurey, "A Digitally Intensive Transmitter/PA Using RF-PWM With Carrier Switching in 130 nm CMOS," *IEEE JSSC*, vol. 51, no. 5, pp. 1188-1199, May 2016.

[24] H.M. Nguyen, et al., "An Edge-Combining Frequency-Multiplying Class-D Power Amplifier," *IEEE TCASII*, vol. 70, no. 2, pp. 471-475, Feb. 2023.

[25] C. Ding, et al., "A 49.8mm² Fully Integrated, 1.5m Transmission-Range, High-Data-Rate IR-UWB Transmitter for Brain Implants," *CICC*, May. 2024.

[26] M. Shoaran, et al., "Energy-Efficient Classification for Resource-Constrained Biomedical Applications," *IEEE JETCAS*, vol. 8, no. 4, pp. 693-707, Dec. 2018.

[27] N. Chandravadia, et al., "A NWB-based dataset and processing pipeline of human single-neuron activity during a declarative memory task," *Sci. Data*, vol. 7, no. 1, p. 78, Mar. 2020.

[28] Y. Ezzyat, et al., "Closed-loop stimulation of temporal cortex rescues functional networks and improves memory," *Nat. Commun.*, vol. 9, no. 1, p. 365, Feb. 2018.

[29] B.M. Roeder, et al., "Developing a hippocampal neural prosthetic to facilitate human memory encoding and recall of stimulus features and categories," *Front. Comput. Neurosci.*, vol. 18, p. 1263311, Feb. 2024.

[30] S. Sharma, et al., "Location-aware ingestible microdevices for wireless monitoring of gastrointestinal dynamics," *Nat. Electron.*, vol. 6, no. 3, pp. 242-256, Feb. 2023.

979-8-3315-4101-9/25/$31.00 ©2025 IEEE