# A 1024-Channel 268 nW/pixel 36x36 μm$^2$/ch Data-Compressive Neural Recording IC for High-Bandwidth Brain-Computer Interfaces

Jang, Moon Hyung; Yu, Wei-Han ; Lee, Changuk; Hays, Maddy; Wang, Pingyu; Vitale, Nick; Tandon, Pulkit; Chae, Youngcheol; Muratore, Dante G.; More Authors

**DOI**
10.23919/VLSITechnologyandCir57934.2023.10185288

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Proceedings of the 2023 IEEE Symposium on VLSI Technology and Circuits, VLSI Technology and Circuits 2023

# A 1024-Channel 268 nW/pixel 36x36 µm²/ch Data-Compressive Neural Recording IC for High-Bandwidth Brain-Computer Interfaces

MoonHyung Jang[1], Wei-Han Yu[2], Changuk Lee[3], Maddy Hays[1], Pingyu Wang[1], Nick Vitale[1], Pulkit Tandon[1], Pumiao Yan[1], Pui-In Mak[2], Youngcheol Chae[3], E.J. Chichilnisky[1], Boris Murmann[1], Dante G. Muratore[4]
[1]Stanford University, USA; [2]Macau University, Macau; [3]Yonsei University, South Korea; [4]TU Delft, The Netherlands

**Abstract**

This paper presents a neural recording IC featuring lossy compression during digitization, thus preventing data deluge and enabling a compact active digital pixel design. The wired-OR-based compression discards unwanted baseline samples while allowing the reconstruction of spike samples. The IC features a 32x32 MEA with 36µm pixel pitch and consumes 268nW per pixel from a single 1V supply. It achieves 9.8µ$V_{RMS}$ input-referred noise and 0.3-5kHz bandwidth, resulting in NEF/PEF of 3.7/14.1.

**Keywords:** neural, recording, brain, interface, compression

## Introduction

Next-generation brain-computer interfaces will benefit from dense, high channel count (>1k) microelectrode arrays (MEAs) at the pitch of neurons (~30µm) to effectively capture spatiotemporal patterns of neural activity. However, as the number of channels increases, the data rate of recording becomes intractable (10k channels digitized at 8b and 20kSps generate 1.6Gbps). While a spike detector (SD) can be used to compress the raw data and transmit a snippet around the spike [1, 2], this approach incurs significant overhead in threshold management (typically per channel) as well as memory to compensate for SD latency. Also, as the MEA density increases, routing congestion worsens, limiting today's interfaces to sub-array digitization [3]. Active digital pixels (ADP) have been used to simplify routing and improve chip area efficiency, but come at the cost of larger pixels [4]. This paper presents a recording IC based on a wired-OR compression scheme [5, 6] that addresses both the data deluge and routing congestion problems.

## Chip Architecture and Circuit Design

Fig. 1 shows the chip block diagram. Each ADP contains an amplifier, a sampler ($f_s$=20kHz), and a continuous-time comparator that drives the local row & column using open-drain outputs (wired-OR). The pixel functions as a ramp ADC, i.e. the acquired samples are compared to a global ramp, leading to an 8b pulse position modulation. The wired-OR array outputs are interpreted by a collision decoder at each of the 256 ramp steps. A collision occurs when two or more pixels in a row (or column) sample the same input at the same time. In this case, the samples cannot be recovered and are discarded. On the other hand, if a pixel samples a unique voltage, the output pulse can be traced back to the pixel location, and the sample is stored. For neural signals, this compression discards a large number of unwanted samples near the signal's baseline while retaining the more important spike samples [5, 6]. Fig. 2 shows an example with primate retina recordings *ex vivo*.

Fig. 3 shows the pixel and ramp generator circuits. The pixel occupies 36x36µm², with one-third allocated for ESD protection. To minimize area, the circuits use only 2.8pF of MOM capacitance on top of ESD and active devices. An AC-coupled boxcar sampler minimizes the noise penalty from noise folding. It uses an inverter-based $G_m$ with a duty-cycled resistor for DC biasing and setting the high-pass corner to $f_{HP}$=300Hz. The input is integrated on $C_{INT}$ ($296T_{ck}$) and then sampled on $C_{LPF}$ ($8T_{ck}$) to implement a switched-capacitor low pass filter (SC-LPF). The SC-LFP pole and the null from the boxcar result in an overall $f_{LP}$=5kHz. $6T_{ck}$ are required to reset $C_{INT}$ between samples. Hence, the main clock frequency is $f_{ck}=310f_s$=6.2MHz. During integration, the previous sample is compared to the global ramp. The 8-bit conversion phase lasts $(256+40)T_{ck}$ to compensate for comparator latency. Comparator auto-zero and offset calibration are used to minimize the offset between channels to the level required by the wired-OR compression. The ramp is generated by integrating a fixed current on a capacitor at each clock cycle. The ramp range ($V_{TOP}$-$V_{BOT}$) and slope can be set to change the ADC resolution between 6-10b. Another option available on chip is to divide the array into subarrays with up to 8 wires per row & column. This allows us to vary the collision rate and, thus, the degree of compression.

## Measurement Results

The IC was fabricated in 28nm CMOS. Fig. 4 (top) shows the measured frequency response and output spectrum of a single channel. The -3dB BW is 0.3-5kHz, and SNR/SFDR=34.1/63dB with a 500$\mu V_{pp}$ input. The measured input referred noise is 9.8$\mu V_{RMS}$ (1Hz-10kHz). Fig. 4 (bottom) shows the noise and offset distributions. Pt electrodes (d=17µm) were deposited on each pixel post-fabrication, and pre-recorded neural signals were injected through a Pt wire in saline (Fig. 5, top). The measured spike waveform (Fig. 5, bottom) shows that even with small high-impedance electrodes, the IC can accurately record neural spikes. Fig. 6 (top) shows sinewave measurements to visualize the wired-OR operation. In the first plot, only a single channel carries a sinewave. All samples outside the baseline are captured, while missing samples near the baseline are reconstructed using an interpolation filter. The second case shows two active channels. All critical samples for reconstructing the two signals are still captured, since there is no coincidence of seeing the same digital value at the same time (i.e., no collisions). Fig. 6 (bottom) shows measurements of a neural signal applied to a test channel, where only spike samples are captured at a compression rate of 12.5x (proportional to the spike rate). The die photo is shown in Fig. 7. The pixel area is 0.0013mm², and the total area is 3.27mm². Each pixel consumes 268nW, and the total power consumption is 508.7$\mu$W (Fig. 8) with a 1V supply. Table I shows a performance comparison. This IC achieves the smallest area per channel and highest power efficiency (NEF/PEF) while providing rail-to-rail electrode DC offset (EDO) tolerance and dealing with data deluge and routing congestion problems through wired-OR compression.

**Acknowledgements:** Chip fabrication was provided by the TSMC University Shuttle Program. Funding support: Wu Tsai Neurosciences Institute, Stanford Nanofabrication Facility, NIH Grants EY021271, EY032900.

## References

[1] S. Park, *IEEE JSSC*, 2018. [2] D.-Y. Yoon, *VLSI*, 2021. [3] C.M. Lopez, *IEEE ISSCC*, 2018. [4] D. Wendler, *IEEE JSSC,* 2022. [5] D.G. Muratore, *IEEE TBCAS*, 2019, [6] P. Yan, *IEEE BioCAS*, 2022. [7] X. Yang, *VLSI*, 2022. [8] S. Wang, *VLSI*, 2021. [9] S. Wang, *IEEE TBCAS*, 2019.

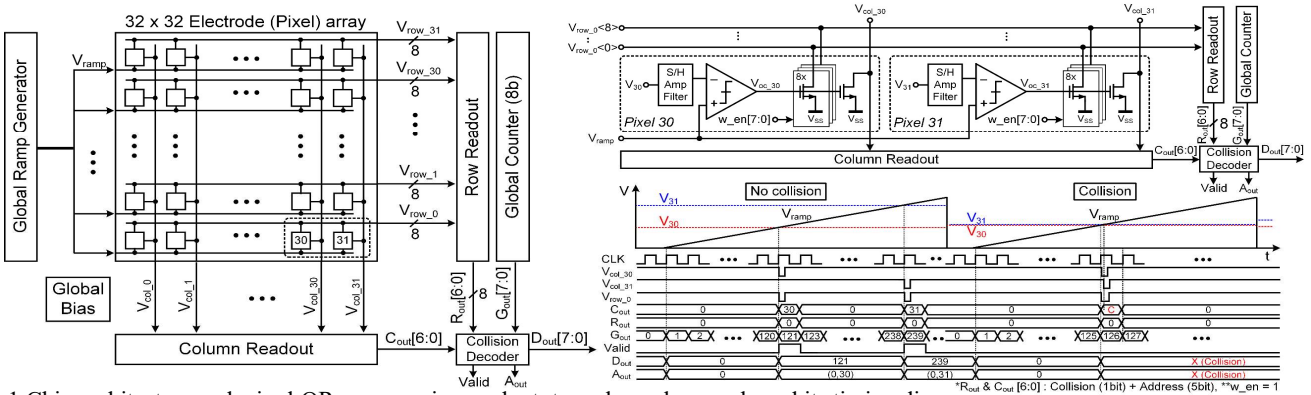


Fig.1 Chip architecture and wired-OR compressive readout: two-channel example and its timing diagram.

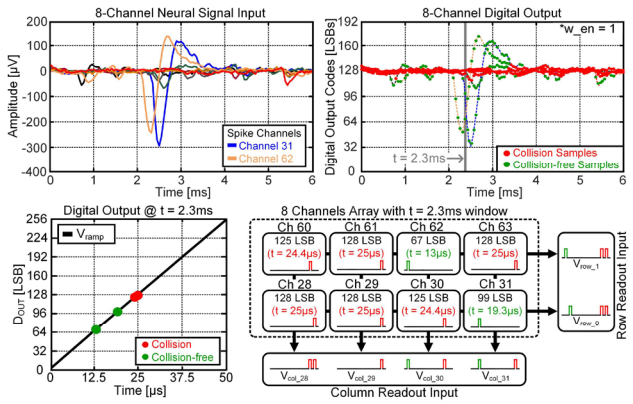


Fig.2 Wired-OR data compression example with primate retina recording (*ex vivo*).

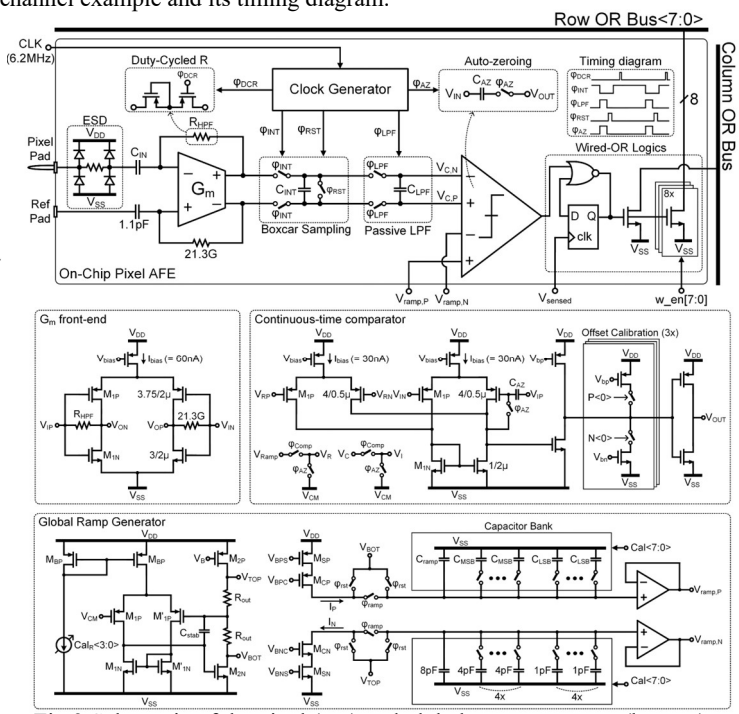


Fig.3 Schematic of the pixel (top) and global ramp generator (bottom).

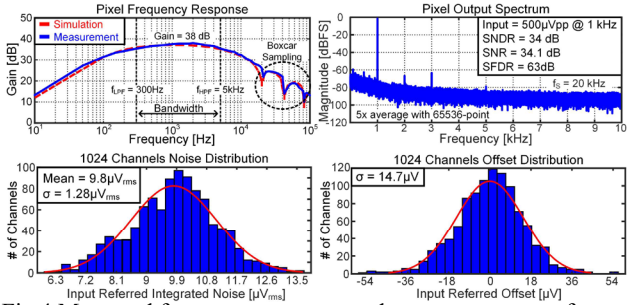


Fig.4 Measured frequency response and output spectrum of one pixel (top). Measured noise and offset distribution across the array (bottom).

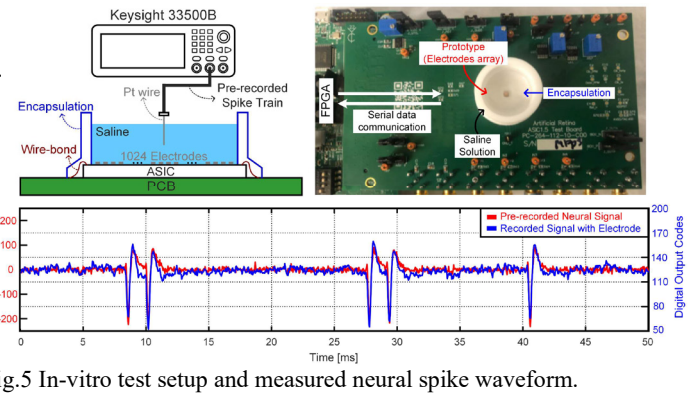


Fig.5 In-vitro test setup and measured neural spike waveform.

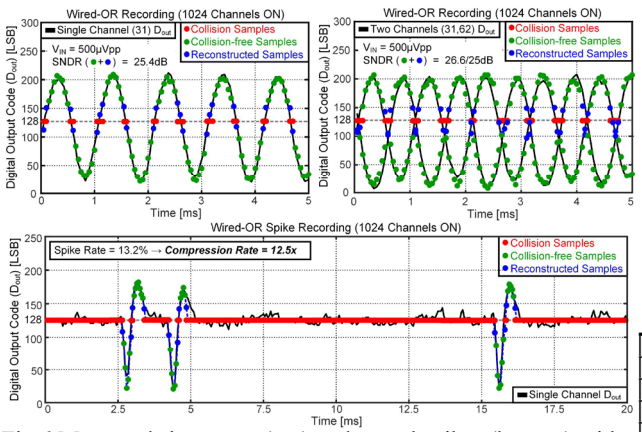


Fig.6 Measured sinewaves (top) and neural spikes (bottom) with wired-OR compression.

Table I. Comparison with state-of-the-art neural recording ICs

| | This work | VLSI'22 [7] | ISSCC'21 [4] | VLSI'21 [2] | VLSI'21 [8] | TBCAS'19 [9] | JSSC'18 [1] | ISSCC'18 [3] |
|---|---|---|---|---|---|---|---|---|
| Technology / Supply | 28nm / 1 V | 22nm / 0.8V | 180nm / 1.8V | 65nm / 1.2V | 55nm / 1.2V | 130nm / 1.2V | 180nm /0.5/1/1.8V | 130nm / 1.2V |
| Channel topology | AC-coupled Boxcar + SS ADC | AC-coupled 1st order ΔΔΣ | DC-coupled 2 Step IΔΣ | AC-coupled IA + SAR ADC | DC-coupled 2nd order ΔΔΣ | AC-coupled IA + SAR ADC | AC-coupled IA + 1st order ΔΔΣ | AC-coupled IA + SAR ADC |
| # Of Channels | 1024 | 128 | 8-24 | 1024 | 16 | 128 | 1024 | 1024 |
| Bandwidth [Hz] | 300-5k | 0.5-10k | 0.5-10k | 5-10k | 0.5-10k | 0.5-10k | 0.4-9.2k | 0.5-10k |
| ADC Resolution | 8 | - | 11 | 10 | - | 14 | 11/8 | 10 |
| Power/Ch [μW] | 0.268 | 6.02 | 8.59 | 2.72 | - | 48.7 | - | 47.42 |
| Chip Total Power/Ch [μW] | 0.496 | 8.34 | 14.94 | 24.08 | 61.2 | 95.1 | 15.35 | 92.77 |
| Input Referred Noise [$\mu V_{rms}$] | 9.8 (1Hz – 10kHz) | 7.71 (0.3 - 10kHz) 11.9 (0.5 - 1kHz) | 4.37 (0.3 - 10kHz) 2.72 (0.5 - 1kHz) | 8.89 (0.3 - 10kHz) 6.8 (5 - 1kHz) | 5.53 (0.3 - 10kHz) 2.88 (0.5 - 1kHz) | 7.43 (0.3 - 10kHz) 7.78 (0.5 - 1kHz) | 5.18 (0.4 -10.3kHz) | 12 (0.5 - 10kHz) 7.5 (300 - 1kHz) |
| [1]NEF / PEF (AP band) | 3.7 / 14.1 | 9.6 / 73.7 | 4.85 / 42.4 | 15.3 / 282.8 | 15.2 / 278.2 | 25.5 / 650.3 | - / 59.4 | 29 / 1015 |
| Input range | 0.75-2.25mVpp | 43mVpp | 14mVpp | 0.75-4.87mVpp | 148mVpp | 12.5mVpp | - | - |
| THD [%] | 0.097 @-3dBFS | 0.15 @21.5mVpp | 0.078 @10mVpp | 0.57 @-0.79dBFS | 0.05 @-20mVpp | 0.17 @10mVpp | 0.062 @3.2mVpp | 0.89 @20mVpp |
| Area/Ch [mm²] | 0.0013 | 0.0045 | 0.0046 | 0.0062 | 0.0077 | 0.035 | 0.098 | 0.125 |
| EDO Tolerance [mV] | Rail-to-Rail | Rail-to-Rail | ±60 | Rail-to-Rail | ±70 | Rail-to-Rail | Rail-to-Rail | Rail-to-Rail |

[1]NEF / PEF are calculated using Chip Total Power/Ch

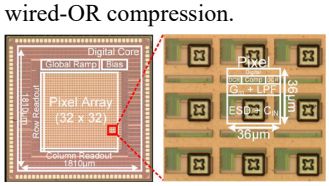


Fig.7 Die photo.

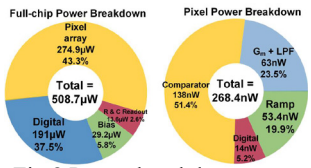


Fig.8 Power breakdown.