Published in final edited form as:
IEEE Trans Biomed Circuits Syst. 2024 October 01; 18(5): 1100–1111. doi:10.1109/TBCAS.2024.3378973.

An Event-based Neural Compressive Telemetry with >11× Loss-less Data Reduction for High-bandwidth Intracortical Brain Computer Interfaces

Yuming He [Student Member, IEEE],
Stan van der Ven,
Hua-Peng Liaw,
Chengyao Shi [Student Member, IEEE],
Pietro Russo,
Marios Gourdouparis [Student Member, IEEE],
Mario Konijnenburg,
Stefano Traferro,
Martijn Timmermans,
Carolina Mora Lopez [Senior Member, IEEE],
Pieter Harpe [Senior Member, IEEE],
Eugenio Cantatore [Fellow, IEEE],
Elisabetta Chicca [Senior Member, IEEE],
Yao-Hong Liu [Senior Member, IEEE]

Abstract

Intracortical brain-computer interfaces offer superior spatial and temporal resolutions, but face challenges as the increasing number of recording channels introduces high amounts of data to be transferred. This requires power-hungry data serialization and telemetry, leading to potential tissue damage risks. To address this challenge, this paper introduces an event-based neural compressive telemetry (NCT) consisting of 8 channel-rotating Δ-ADCs, an event-driven serializer supporting a proposed ternary address event representation protocol, and an event-based LVDS driver. Leveraging a high sparsity of extracellular spikes and high spatial correlation of the high-density recordings, the proposed NCT achieves a compression ratio of >11.4×, while consumes only 1 μW per channel, which is 127× more efficient than state of the art. The NCT well preserves the spike waveform fidelity, and has a low normalized RMS error <23% even with a spike amplitude down to only 31 μV.

Index Terms

Intracortical neural recording; event-based; serialization; level-crossing ADC; data compression; intracortical brain-computer interfaces; serializer; SERDES; neuromorphic; wireline; address event representation

Europe PMC Funders Author Manuscripts

I Introduction

REVOLUTIONIZING the landscape of neural therapeutics and neuroscience research, brain-computer-interfaces (BCIs) represent a frontier technology in connecting the brain and external electronics. For example, BCIs help individuals with severe disabilities, such as paralysis, locked-in syndrome, etc [1], [2]. Beyond therapeutic applications, BCIs have a transformative impact on neuroscience research, facilitating the understanding of neural dynamics, neurological disorders, and even cognitive functions [3]. Among the existing BCI's neural sensing methods, intra-cortical neural recording (shown in Fig. 1) has been proved to provide the finest spatial and temporal resolutions, and is being rapidly and widely applied. As the number of recording channels doubles every 6 years, the amount of data required to be transferred also grows dramatically. For instance, the high-density CMOS-based neural probe in [4] generates a data rate of 171 Mbps across 384 channels, operating at 30 kSps/ch and a 10-bit ADC resolution.

The intracortical recording data from the cortical implant is typically transferred to a "neural hub" with a rechargeable battery, for the subsequent wireless data transmission. This neural hub is typically placed at specific location of the body with a larger volume for the heat dissipation reason, e.g., in the chest with an implant devices similar to the standard implantable pulse generator (IPG) [5], [6], as illustrated in Fig. 1. In this case, serial data transfer is required to avoid routing multiple wires in the body, and a challenging synchronization issue between signals. However, with the continuous increase of the recording channel number, the data serializer needs to handle up to 100's of Mbps of data rate, which is too power consuming (>80 mW in [7]) for a cortical implant. It requires at least one pair of high-bandwidth Low-Voltage Differential Signaling (LVDS) line driver which typically consumes more than 10's of mW [8] of static power, and a high-frequency (100's of MHz) PLL as a serializer clock, as shown in Fig. 2(a). The high-power consumption of the high-speed serializer potentially results in a high risk of tissue damage due to the increase of temperature. The literature [9], [10] suggest a temperature increase of even only 1 degree can lead to the brain tissue damage, and recommend that a power density of implant ASIC should be reduced to below 20 mW/cm^2. Hence, to avoid serializing and transferring large amounts of data, compression is highly favored.

The amplitude of extracellular neuron firing signals (spikes) can range from ~500 μV ("good" spikes) to <50 μV ("poor" spikes), and they are all important for neuroscience data analysis, e.g., spike sorting. Modern high-density neural probes capture not only good spikes but also poor spikes to have rich spatial features and improve the yield of the spike sorting. Hence, a "loss-less" compression is crucial as it can preserve the spike waveform features with low RMS error for a wide range of amplitudes.

Several loss-less data compression solutions have been presented in literature, but they need either high hardware resource, or they significantly reduce the dynamic range in the process of compression. An "Epoch" approach presented in [11] in which recorded samples need to be buffered in a memory (typically 1-10 kb/ch), leading to large hardware resources when the number of recording channels increases. The wired-OR event readout method used

in [12] achieves a low hardware complexity, but it suffers from collisions when the spike amplitude is low (<±25 LSBs).

The Δ modulation technique in [13], [14] achieves a significant data reduction by only taking the difference of the signal, while faithfully preserving the low-amplitude spike waveforms. However, the analog Δ modulator in [14] occupies a large area due to the need of large capacitors. A continuous-time level-crossing ADCs (LC-ADCs) in [13] performs the Δ modulation along the digitization, but they are not compatible with time-multiplexed analog front-ends (AFEs), which are commonly implemented in high-density and high-channel-count neural probes [4]. In addition, arbiters [13], [15] are required to serialize the asynchronous events generated from LC-ADCs before data transfer, and the power consumption and the timing uncertainty (which degrades SNR after reconstruction) are significantly increased when the number of recording channels is increased [16].

This paper proposes an “event-based” neural compressive telemetry (NCT) ASIC which leverages the sparsity and spatial correlation of neural spike signals and achieves >11× of loss-less data reduction, even for a high-density recording with high spatial redundancy. The key contributions of this work are as follows:

1) The proposed event-based NCT architecture digitizes, compresses, serializes, and transfers extracellular neural recording signals from high number of channels, with a very power- and area-efficient design;

2) A channel-rotating Δ (CR-Δ) ADC is proposed to support time-multiplexing analog input (to be detailed in section III.A);

3) Presents an event-based serializer (eSER) architecture that employs an event-driven high-speed clock, and supports a novel “ternary address event representation (AER) packet” protocol suitable for energy-efficient serial communication (to be detailed in section III.B);

4) An event-based LVDS driver (eLVDS) is presented, which can be heavily duty cycled to avoid the static power consumption of the traditional LVDS driver (to be detailed in section III.C).

The rest of this article is organized as follows: Section II introduces the proposed architecture of NCT and Section III describes the circuit implementation. The measurement results will be shown in Section IV. Finally, the conclusions will be provided in Section V.

II The Proposed Serial Telemetry

The proposed NCT architecture is illustrated in Fig. 2(b). It consists of 8 CR-Δ ADCs, each supporting a 16:1 time-multiplexed input, the event-driven serializer (eSER) with spatial-grouping event packet protocol, and an event-based LVDS driver (eLVDS). The focus of this work is serial telemetry because it typically consumes a significant portion of the power consumption of the intracortical neural implant. The AFE part (i.e., instrument amplifiers, analog filter, etc.) is not included in this work.

Europe PMC Funders Author Manuscripts

This proposed event-driven methodology employs an event-driven clock for data serialization, effectively eliminating the need for a power-hungry HF-PLL. It enables the simultaneous streaming of neural data from 128 channels with little power consumption. The proposed CR-Δ ADCs sample the signal variation between sampling and generate Δ outputs only if the change of the signal exceeds a specific threshold. It allows power-efficient data processing afterwards as the zero-Δ output does not need to be processed.

The Δ outputs of the CR-Δ ADCs are sent to the eSER and the serialization clock is enabled if non-zero Δ output is detected. After the start-up of serialization clock, the eSER "packs" the data from active-only channels (Δ≠0). This process continues until all relevant data is processed, following which the clock is disabled. The outputs from the eSER, which indicate the period of the packets and the bitstream, are sent to the eLVDS driver. With the proposed Δ-based processing, the power consumption of the serialization clock and data processing can be reduced significantly.

Leveraging the sparse nature of the spikes and the spatial correlation of the high-density recording, this proposed event-driven approach drastically reduces the data processing and transfer complexity, leading to substantially reduced energy consumption.

III Circuit Implementation

The implementation of the circuits in the proposed NCT system will be discussed in this section, including (A) a CR-Δ ADC for the multiplexed analog input; (B) a power-efficient event serializer which compresses the event outputs from the CR-Δ ADCs and generates the AER packets for data transmission; (C) an event-based LVDS driver with low energy consumption.

A Channel-rotating Δ-ADC

Most of the ADCs for neural recording presented in the literature have a sampling rate of 20-30 kSps and a resolution of 8-10 bits, resulting in large amounts of data to be either processed or transferred. Since the power and area are very constrained in implants, it is important to investigate the impact of reducing the resolution of the ADCs.

In most of the BCI applications, the recorded intracortical spikes are sorted to individual (neuron) units before further analysis, i.e., spike sorting. The impact of conventional Nyquist ADC's resolution to the spike sorting accuracy has been discussed in [17]. It reveals that the conventional ADC with a resolution even down to 7.62 bit has a negligible degradation in spike sorting accuracy, in comparison to the ADC [18] with a higher resolution (>9 bit).

To evaluate the effects of reduced resolution in Δ-ADCs on spike sorting, an in-silico analysis using a pre-recorded in-vivo dataset [19] (with original 10-bit resolution and 30-kSps sampling rate) has been conducted. A Δ-ADC model with different resolutions is employed to quantize the dataset and convert it to Δ, and the quantized data is outputted to a widely-adopted spike sorting tool, Kilo-sort [20], to evaluate the sorting performance. To determine the impact of the Δ-ADC resolutions on the sorting outcomes, the sorted spikes from the original recorded dataset are set as a reference. The percentage of sorted spikes,

from the quantized dataset, that can be matched to the reference indicates the impact of Δ-ADC quantization, as plotted in Fig. 3. In this analysis, the full scale of the Δ-ADC referred to the recording site input is set to 1.1 mV_{PP} with a sampling rate at 20 kSps. The simulated compression ratio across various Δ-ADC bit numbers is also shown in Fig. 3. Based on the investigation, the Δ-ADCs with a 7-bit resolution are chosen to ensure a negligible impact on spike sorting while having a high compression ratio of >10×.

In addition, Fig. 4 Fig. 3shows the spike waveforms of the clusters from Kilo-sort. These spikes are from both the in-vivo pre-recorded dataset and the reconstruction from a model of Δ-ADC with a 7-bit resolution. The normalized RMS error (NRMSE) between the original dataset and the reconstructed signals are also listed in Fig. 4. Across a wide range of spike amplitudes from different clusters, the NRMSEs are within 8.2%, i.e., SNR of >21.7 dB.

Discrete-time CR-Δ ADCs are designed to support a 16:1 time-multiplexed input, which is shown in Fig. 5(a). The proposed CR-Δ ADC consists of one sample-and-hold circuit, one dynamic comparator, one differential top-plate sampling capacitive DAC (C-DAC) controlled by a channel-rotating First In First Out (FIFO) and a DAC control module (DAC Ctrl.), and an event counter with a pulse generator.

The time-multiplexer from the AFE is not implemented in this chip. In order to synchronize the time-multiplexer and the ADC, they will share the same clock, i.e., *ADC CLK*. The 16 inputs are converted consecutively during a time “frame” (50 μs), and multiplexing continues in the successive frames. Fig. 5(b) explains the conceptual waveform of the proposed CR-Δ ADC output. The CR-Δ ADC quantizes the signal difference between two consecutive frames, and then it generates two outputs: polarity (*SIGN*) and the quantized delta (i.e., Δ). This will be performed for each channel per ADC clock cycle.

Fig. 6(a) shows the detailed operation of the proposed CR-Δ ADC. When *ADC CLK* is low, the time-multiplexed input is sampled by the sample-and-hold circuit. Before the conversion of each channel, i.e., on the rising edge of *ADC CLK*, the channel-rotation module sets the 7-bit DAC to the stored code from the previous sampling frame of the same channel. Then the conversion starts and the DAC output settles to the current channel input based on the comparator output. During the conversion period, asynchronous pulses are generated by the pulse generator as an internal clock signal for the comparator. Based on the comparator output, an event counter counts the number of comparisons and generates *SIGN* and Δ. After the conversion, the final 7-bit code is stored in the channel-rotation module, and the event counter is reset to zero. As illustrated in Fig. 6(b), the channel-rotation module includes a 7-bit FIFO and a DAC control module. The final code of a certain channel after the conversion will be pushed to the FIFO, and the code of next channel will be read and sent to the DAC before conversion. During the conversion, the DAC Ctrl updates the control of the DAC based on the signal from the event counter. Unlike conventional SAR-ADCs [4] that always start the conversion from the center code, the proposed CR-Δ ADC starts the conversion from the code stored in the channel-rotation module from the previous frame. This greatly reduces the number of comparisons needed, especially when the input is sparse, thus lowering the energy consumption for each conversion by 5-10×. Furthermore, for the high-density neural recording, the final code from the previous channel (as illustrated in Fig.

6(a)) is usually close to the code to be set in the next channel. Therefore, the DAC switching energy for channel rotation can be further reduced, thanks to the high spatial correlation between channels.

$Step_{Max}$ is the maximum number of events between ADC frames, which is designed to be 31 in this work. This $Step_{Max}$ determines the required bit width of the channel-rotation FIFO, and the comparator speed because these comparisons should be completed within half of the ADC clock cycle. A dynamic comparator is employed in this ADC, as shown in Fig. 6(c), because of the tradeoff between the power consumption and the comparison speed.

B Event serializer with Spatial-grouping Ternary AER

Multiple CR-Δ ADCs generate Δ outputs need to be serialized before the following (wired or wireless) data telemetry. The traditional data serialization method [4], [20], [21] scans through each ADC, which results in large amounts of data and requires a high-frequency synchronous clock. This does not benefit from the event-driven data reduction from the CR-Δ ADCs. Address-event-representation (AER) is a standard communication protocol for event-driven systems, e.g., dynamic vision sensors [22], and typically use parallel bus to transfer asynchronous events from each channel (or pixel) of an array-based sensor. A serial-AER protocol suitable for data telemetry is presented in [23], [24], which encodes multiple-bit source address into a serial form. However, serializing asynchronous events from multiple channels requires event arbitration or an acknowledgement-based protocol, thereby increasing system complexity and is not easily scaled to a high-channel-count sensing system.

In this work, a “packetized ternary AER” is proposed to bundle events from multiple channels into a serial ternary-coded packet. Furthermore, a “spatial grouping” technique is proposed to further reduce the protocol overhead of the proposed packetized AER, as conceptually illustrated in Fig. 7(a) and Fig. 7(b). With high-density neural probes, a spike from one neuron can be recorded by multiple adjacent channels. Therefore, events from nearby channels have a strong spatial correlation, and adjacent channels can be time-multiplexed to the same ADC [4] in Fig. 7(a). Given the strong spatial correlation among events from nearby channels sampled by the same ADC, instead of sending the same *ADC ID* multiple times, it will be sent out once per packet for each *ADC ID*. To indicate the start of each ADC, the *ADC HDR* is included in the packet. As shown in Fig. 7(a), this spatial grouping can reduce the protocol overhead by up to 2×. The event packet structure is shown in Fig. 7(b). It consists of distinct sequences of binary words: synchronization (*SYNC*), ADC header (*ADC HDR*), ADC address (*ADC ID*), and one or more channel ID words (*CH ID*) from the same ADC. The *CH ID*, including 1-bit *SIGN* and 4-bit channel address, can be repeated several times depending on the quantized delta (“Δ” value).

Fig. 8(a) shows the block diagram of the proposed eSER that can support the proposed ternary AER packet protocol. At the rising edge of every ADC clock, an event memory stores events from active channels, defined as those with a non-zero Δ, from each CR-Δ ADC. Each active channel requires a 10-bit memory (1-bit *SIGN*, 5-bit for Δ in binary format, and 4-bit *CH ID*) plus another 10-bit replica memory, for parallel data buffering and processing. Once all the events in a single ADC sampling frame (16 ADC clock cycles) are

stored, a finite state machine (FSM) activates a ring oscillator (RO) to provide a clock for the eSER (CLK_{SER}) at the beginning of each subsequent sampling frame if any non-zero Δ is stored in this ADC sampling frame. The FSM module scans the event information of each ADC stored in the event memory. Based on the outputs of the event memory, the FSM generates a 5-bit output word, *WRD*, every 5 CLK_{SER} cycles, each representing *SYNC, ADC HDR, ADC ID*. These words *WRD* are sent to a bitstream generator which converts them to a serial event packet. The bit-stream generator output is ternary coded, including two signals: *FLAG* and *DATA*. The *FLAG* signal indicates the active period of the packet. It is used to enable the following event-based LVDS driver (to be discussed in section III.C) and the *DATA* signal controls the output polarity of the event-based LVDS driver.

Since the oscillator is only activated when packets are generated, a proper start-up period is needed for the oscillator. Fig. 8(b) conceptually illustrates the start-up waveform of the event-driven clock generation. At the beginning of an ADC frame, the serializer clock will be enabled if a packet needs to be transferred. To avoid unstable states of the RO during the start-up, a counter is implemented to count the number of RO edges, $CNT_{STARTUP}$, during the start-up. After it reaches a preset value, the FSM in the eSER starts to generate the packet. The CLK_{SER} will be disabled once all events stored in the event memory are processed, and it will be enabled, if needed, in the next ADC frame. An overflow detector is also integrated in the eSER to identify instances where the AER packet length is longer than one ADC frame. In cases of overflow, the frequency of the CLK_{SER} can be increased to mitigate possible data loss in future frames.

A PLL is typically required in a serializer to generate a stable high-frequency clock. However, its long start-up time makes it difficult to be heavily duty-cycled to reduce power consumption. Therefore, a fully synthesized ring oscillator (RO) is employed in the presented event-driven system because of its fast start-up characteristic, and the compatibility in standard digital design and verification flows. Fig. 9 shows the gate-level circuitry of the synthesized RO. The RO frequency can be digitally adjusted by changing the number of inverters in the loop. This allows it to increase its frequency if the eSER needs to transfer more events in a packet once an overflow is detected. The frequency of the RO is designed to cover a frequency range from 84 MHz to 574 MHz.

C Event-based LVDS (eLVDS) Driver

LVDS is a widely adopted high-speed wireline digital interface because of its high bandwidth, high noise immunity, and the capability of long-distance propagation. Fig. 10(a) shows the conceptual schematic of the traditional LVDR driver. However, its high static power consumption (>10's of mW[4][21]) burdens the implants with extreme limited energy resource. Therefore, a power-efficient dynamic eLVDS driver has been proposed.

Given that the data output from the eSER is formatted as short event packets, the driver predominantly operates in idle mode. To enhance the power efficiency under this condition, an event-based LVDS driver circuit is proposed. Furthermore, a Manchester encoding scheme is adopted such that the receiver can extract the clock frequency directly from the transferred data, obviating the need of transmitting a high-frequency clock alongside the data to the receiver.

The schematic of the eLVDS driver are shown in Fig. 10(b). During the event packet transmission, the *FLAG* signal from the eSER remains high, and the differential input signals, *INP* and *INN*, are generated in response to the *DATA* signal from the eSER. In the idle period, the *INN* and *INP* signals are set to low and the outputs of the driver, *OUTP* and *OUTN*, are connected to half the supply voltage (*VDD/2*). The duty-cycling of the LVDS driver significantly reduces the power consumption of the proposed event-based telemetry system.

Fig. 11(a) illustrates the interface between the eSER and eLVDS driver on the cortical implant side, and the LVDS RX on the neural hub. The signals generated by the eSER are fed into the eLVDS driver, which produces a differential output in response to the state of the *DATA* signal when the *FLAG* signal is in high state. Fig. 11(b) illustrates the differential outputs, *OUTP* and *OUTN*, from the proposed eLVDS driver. These outputs are then sent to the neural hub with standard LVDS interface. The *FLAG* signal can be recovered from the RX side by detecting a specific pattern in the AER packet, i.e., *ADC HDR* in this work.

IV Measurement Results

The presented chip has been fabricated in 65-nm CMOS technology. The chip photo and the area breakdown of the sub-modules are shown in Fig. 12.

A CR-Δ ADC and NCT measurements

Fig. 13(a) shows the effective number of bits (ENOB) measured from a single CR-Δ ADC, utilizing a 1-kHz sinusoidal input, where it achieves an ENOB of 6.98 bits. The proposed CR-Δ ADC consumes 0.11 μW per channel and achieves a Walden's Figure of Merit (FoM_W) of 282.19 fJ/conv when the input is a 1-kHz sinusoidal signal. Note that the power consumption will be lower when the input is extracellular neural signals because of the sparsity. A 1-kHz sinusoidal signal is provided to all the 16 channel inputs of one CR-Δ ADC. The NCT output is subsequently deserialized and reconstructed for the calculation of the signal-to-noise-and-distortion ratio (SNDR) and the normalized RMS error (NRMSE). As illustrated in Fig. 13(b), the measured SNDR from the reconstructed signal is 41.6 dB SNDR, which is mere 2.2-dB deviated from that of the measured single-channel CR-Δ ADC output. The minor difference suggests that the proposed event-based compression and serialization processes introduce only a negligible error. This leads to a "system ENOB" of 6.62 bits.

Fig. 13(c) presents the measured SNDR across various signal amplitudes. To ensure the accurate sorting of low-amplitude spikes, a minimum SNDR of 12 dB, corresponding to an NRMSE of less than 25%, is recommended [25]. This demonstrates that the NCT effectively preserves the waveform features essential for spike sorting, even down to an amplitude equivalent to 3% of a 1.1-mV_{PP} full scale (i.e., ~25.5 μV).

Fig. 13(d) shows a spike-like analog signal feeding into the proposed NCT architecture, to illustrate the event-driven operation. It can be observed that the NCT generates a reduced number of event packets during 'quiet' periods, thereby showing the event-driven nature of this proposed work.

Europe PMC Funders Author Manuscripts

B Validation with in-vivo data

To assess the effectiveness of the proposed compression technique with real extracellular neural recording signals, two pre-recorded in-vivo datasets are used: a high-density Neuropixels dataset recording mouse neural activities [19] and a low-density Utah array dataset from primates [26]. To synthesize 8 time-multiplexed (16:1) input signals, the datasets are converted to 128-channel time-multiplexed analog signals by an FPGA and 8 external 12-bit DACs, before feeding to the NCT.

Fig. 14(a) shows the reconstructed waveform across 128 channels from the dataset [19]. Fig. 14(b) shows the reconstructed spikes from neighboring channels with different amplitudes from the same spike cluster, demonstrating the stability of the proposed compression method, which supports the analysis and discussion in section III.A and Fig. 4. In addition, it demonstrates that the spike amplitude down to 31 μV can be effectively reconstructed with a NRMSE of less than 23%, suggesting that the proposed NCT system can maintain the target signal fidelity for the spike sorting (i.e., 12 dB SNDR), as discussed in Section I and IV.A. Fig. 15(a) shows the measured average input-referred RMS error are 7.4 and 5.7 μV_{RMS} respectively for dataset [19][26], which are comparable to the input-referred noise typically observed in neural AFEs [4], [11], [12], [14], which range from 6-11 μV_{RMS}.

The power breakdown of this proposed NCT is shown in Fig. 15(b), highlighting a measured power consumption of just 1 μW per channel. The proposed event-driven approach reduces the power consumption of the eLVDS driver's by 30×, i.e., from 3.3 mW to 114.6 μW. The measured compression ratios including the protocol overhead is 11.4× for Neuropixels dataset and 13.7× for Utah array dataset. It is important to note that the average spike firing rate can affect the compression ratio and therefore the firing rates are specified in Fig. 16. Additionally, the proposed NCT requires little memory, only 27 bits per channel which is a 55× lower than the Epoch approach [11]. This significantly reduces the demand in memory access, which contributes to the overall small silicon area and low compression power of this proposed NCT.

C Benchmark and discussion

Table I summarizes the performance of the proposed NCT and compares it with state-of-the-art neural recording systems having either data compression or serialization. Note the presented NCT does not contain analog front-end (AFE) part, and this work intends to demonstrate the importance of power reduction in data compression and serialization, which dominate the power consumption in [4], [21]. Hence, the power and area of the AFE are not benchmarked in Table I. The proposed NCT chip achieves a high compression ratio (>11.4×) for high-density recording, while consuming the lowest power consumption and occupying the smallest digital area.

Refence [21] shows that the AFE part of the state-of-the-art neural recording systems consume a substantial portion of total power (3-30μW/ch) to ensure a large dynamic range covering wide range of spike amplitudes. However, this leads to large amounts of data to be transferred, resulting in a high power consumption in data serialization and telemetry (with either wired LVDS or wireless), which can be up to 10× higher than that of the AFE [4].

Europe PMC Funders Author Manuscripts

Given the strict constrains in power consumption and area in the targeted implant use cases, the proposed compressive telemetry drastically reduces the power consumption by more than an order, while maintaining the dynamic range required for the spike sorting analysis (6.6 dB system ENOB). For the conventional telemetry system, all data bits generated by the ADCs are transmitted out, regardless of the effective dynamic range of the ADCs, or the dynamic range reduction due to e.g., compression. In addition, the size of the streamed data from the proposed compressive telemetry system is highly signal-dependent, and the information is bear in the timing of events transmitted instead of the actual number of bits. Hence, in order to provide a good assessment of energy efficiency of the data telemetry, taking into account the effective dynamic range of compressible spike amplitude, we propose a compressive telemetry energy efficiency ($EnergyEff_{Compress\_Telemetry}$) based on system ENOB instead of using actual ADC bits as in the traditional telemetry energy efficiency,

$$EnergyEff_{Compress\_Telemetry} \equiv \frac{Power\_Cons_{Telemetry}}{f_s \cdot CH \cdot \boldsymbol{ENOB_{System}}}, \tag{2}$$

where $Power\_Cons_{Telemetry}$ is the power consumption including the ADCs, compression, serialization, and data driver (either LVDS or wireless), f_s is the sampling rate, and CH is the number of channels. Note it may not be suitable to use this metric to benchmark the energy efficiency of wireless telemetry systems, since their power consumption is highly dependent on the air transmission distance. The presented work achieves a telemetry energy efficiency of 7.5 pJ/bit, at least 127× better than the state of the art with the wired LVDS telemetry.

V Conclusion

The increasing number of recording channels in intra-cortical neural recording poses challenges in data transfer and power consumption, leading to the need for efficient compression techniques to mitigate the risk of tissue damage associated with high-power consumption.

This paper introduces an innovative low-power event-based neural compressive telemetry that includes channel-rotating CR-Δ ADCs for multiplexing analog inputs, an event-based serializer and an event-based LVDS driver. The primary objective is to efficiently digitize, serialize and transfer high-bandwidth data, while maintaining sufficient dynamic range of 6.6-bit ENOB required for further spike sporting analysis. Section III.A revealed the feasibility of reducing both the sampling rate and resolution without compromising spike sorting performance. The presented neural compressive telemetry demonstrates an ability to energy-efficiently serializing 128 recording channels with a high compression ratio of 11.4× for high-density recording. In addition, it distinguishes itself through a remarkable energy efficiency at least 127× better than state of the art, and a compact compression module design with a 55× lower memory access demand compared to traditional approaches, making it a promising solution for high-channel-count neural recording implants.

Acknowledgement

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101001448).

References

[1]. Chaudhary U, Mrachacz-Kersting N, Birbaumer N. Neuropsychological and neurophysiological aspects of brain-computer-interface (BCI) control in paralysis. J Physiol. 2021; 599 (9) 2351–2359. [PubMed: 32045022]

[2]. Bockbrader MA, Francisco G, Lee R, Olson J, Solinsky R, Boninger ML. Brain Computer Interfaces in Rehabilitation Medicine. PM&R. 2018; 10 (9S2) S233–S243. [PubMed: 30269808]

[3]. Buzsáki G, et al. Tools for Probing Local Circuits: High-Density Silicon Probes Combined with Optogenetics. Neuron. 2015; Apr; 86 (1) 92–105. DOI: 10.1016/j.neuron.2015.01.028 [PubMed: 25856489]

[4]. Lopez, CM; , et al. 22.7 A 966-electrode neural probe with 384 configurable channels in 0.13µm SOI CMOS; 2016 IEEE International Solid-State Circuits Conference (ISSCC); San Francisco, CA, USA. 2016 Jan. 392–393.

[5]. Haci, D; Mifsud, A; Liu, Y; Ghoreishizadeh, SS; Constandinou, TG. In-body wireline interfacing platform for multi-module implantable microsystems; 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS); 2019 Oct. 1–4.

[6]. Tasneem, N; Ahmed, T; Walker, RM. Design of a 180 nm CMOS transceiver for implantable wireline communication, achieving 800 Mbps at BER<1e-12 with 22.4 dB of channel loss; 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS); 2019 Aug. 1155–1158.

[7]. Maxim Integrated. MAX9271: 16-Bit GMSL Serializer with Coax or STP Cable Drive. 2013. Nov. [Online]. Available: https://www.analog.com/en/products/max9271

[8]. Tasneem, N; Ahmed, T; Walker, RM. Wireline communication over an implantable lead; 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2016 Dec. 321–325.

[9]. Borton DA, Yin M, Aceros J, Nurmikko A. An implantable wireless neural interface for recording cortical circuit dynamics in moving primates. J Neural Eng. 2013; Apr. 10 (2) 026010 doi: 10.1088/1741-2560/10/2/026010 [PubMed: 23428937]

[10]. Reichert, WM. Indwelling Neural Implants: Strategies for Contending with the In Vivo Environment. CRC Press; 2007.

[11]. Biederman W, et al. A 4.78 mm 2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels With Digital Compression and Simultaneous Dual Stimulation. IEEE J Solid-State Circuits. 2015; Apr; 50 (4) 1038–1047. DOI: 10.1109/JSSC.2014.2384736

[12]. Jang, M; , et al. A 1024-Channel 268 nW/pixel 36x36 µm ^{2}/ch Data-Compressive Neural Recording IC for High-Bandwidth Brain-Computer Interfaces; 2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits); Kyoto, Japan. 2023 Jun. 1–2.

[13]. He Y, et al. An Implantable Neuromorphic Sensing System Featuring Near-Sensor Computation and Send-on-Delta Transmission for Wireless Neural Sensing of Peripheral Nerves. IEEE J Solid-State Circuits. 2022; Oct; 57 (10) 3058–3070. DOI: 10.1109/JSSC.2022.3193846 [PubMed: 36741239]

[14]. Kim S-J, et al. A Sub-\muW/Ch Analog Front-End for \Delta -Neural Recording With Spike-Driven Data Compression. IEEE Trans Biomed Circuits Syst. 2019; Feb; 13 (1) 1–14. [PubMed: 30418918]

[15]. Stuijt J, Sifalakis M, Yousefzadeh A, Corradi F. µBrain: An Event-Driven and Fully Synthesizable Architecture for Spiking Neural Networks. Front Neurosci. 2021; May. 15 664208 doi: 10.3389/fnins.2021.664208 [PubMed: 34093116]

[16]. Carrillo, S, Harkin, J, McDaid, L, Pande, S, Morgan, F. Evolvable Systems: From Biology to Hardware. Tempesti, G, Tyrrell, AM, Miller, JF, editors. Berlin, Heidelberg: Springer; 2010. 133–144. Lecture Notes in Computer Science

[17]. Barsakcioglu DY, et al. An Analogue Front-End Model for Developing Neural Spike Sorting Systems. IEEE Trans Biomed Circuits Syst. 2014; Apr; 8 (2) 216–227. [PubMed: 24800679]

[18]. Gao H, et al. HermesE: A 96-Channel Full Data Rate Direct Neural Interface in 0.13 μm CMOS. IEEE J Solid-State Circuits. 2012; Apr; 47 (4) 1043–1055. DOI: 10.1109/JSSC.2012.2185338

[19]. Neural recording results with Neuropixels probe from primates. 2019. Mar. [Online]. Available: http://data.cortexlab.net/dualPhase3

[20]. Steinmetz NA, et al. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. doi: 10.1126/science.abf4588 [PubMed: 33859006]

[21]. Wang S, et al. A Compact Quad-Shank CMOS Neural Probe With 5,120 Addressable Recording Sites and 384 Fully Differential Parallel Channels. IEEE Trans Biomed Circuits Syst. 2019; Dec; 13 (6) 1625–1634. [PubMed: 31545741]

[22]. Moreno, JM; Madrenas, J; Kotynia, L. Synchronous Digital Implementation of the AER Communication Scheme for Emulating Large-Scale Spiking Neural Networks Models; 2009 NASA/ESA Conference on Adaptive Hardware and Systems; San Francisco, CA, USA. 2009 Jul. 189–196.

[23]. Motto Ros P, et al. A Wireless Address-Event Representation System for ATC-Based Multi-Channel Force Wireless Transmission. 2013; doi: 10.1109/IWASI.2013.6576061

[24]. Cassidy, A; Zhang, Zhaonian; Andreou, AG. Neuromorphic interconnects using Ultra Wideband radio; 2008 IEEE Biomedical Circuits and Systems Conference; Baltimore, MD, USA. 2008. 297–300.

[25]. Suner S, Fellows MR, Vargas-Irwin C, Nakata GK, Donoghue JP. Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. IEEE Trans Neural Syst Rehabil Eng. 2005; Dec; 13 (4) 524–541. [PubMed: 16425835]

[26]. Neural recording results with Utah-array from mice. [Online]. Available: https://zenodo.org/record/1488441

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

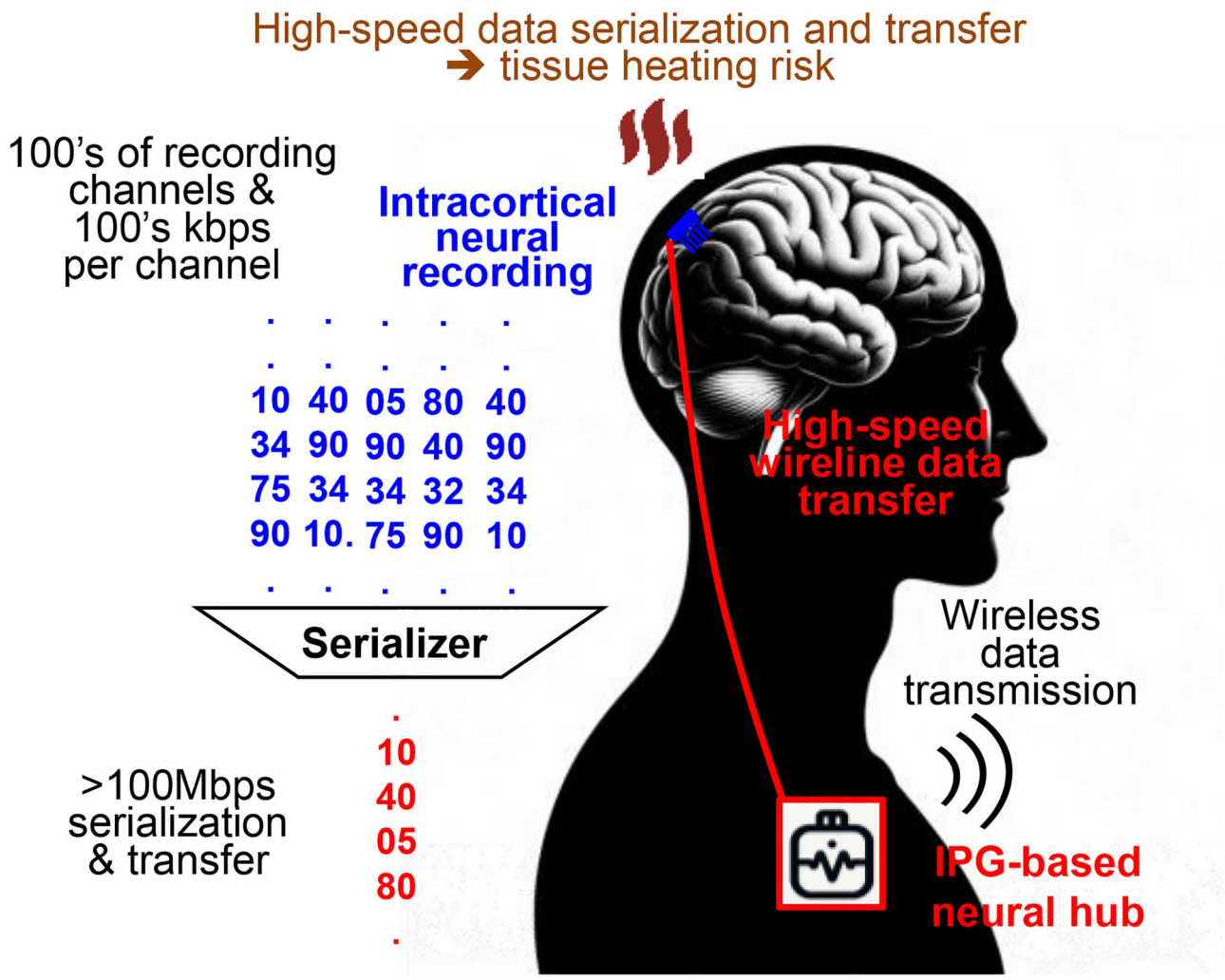

Fig. 1. Concept illustration of intracortical neural recordings with implantable data serialization.

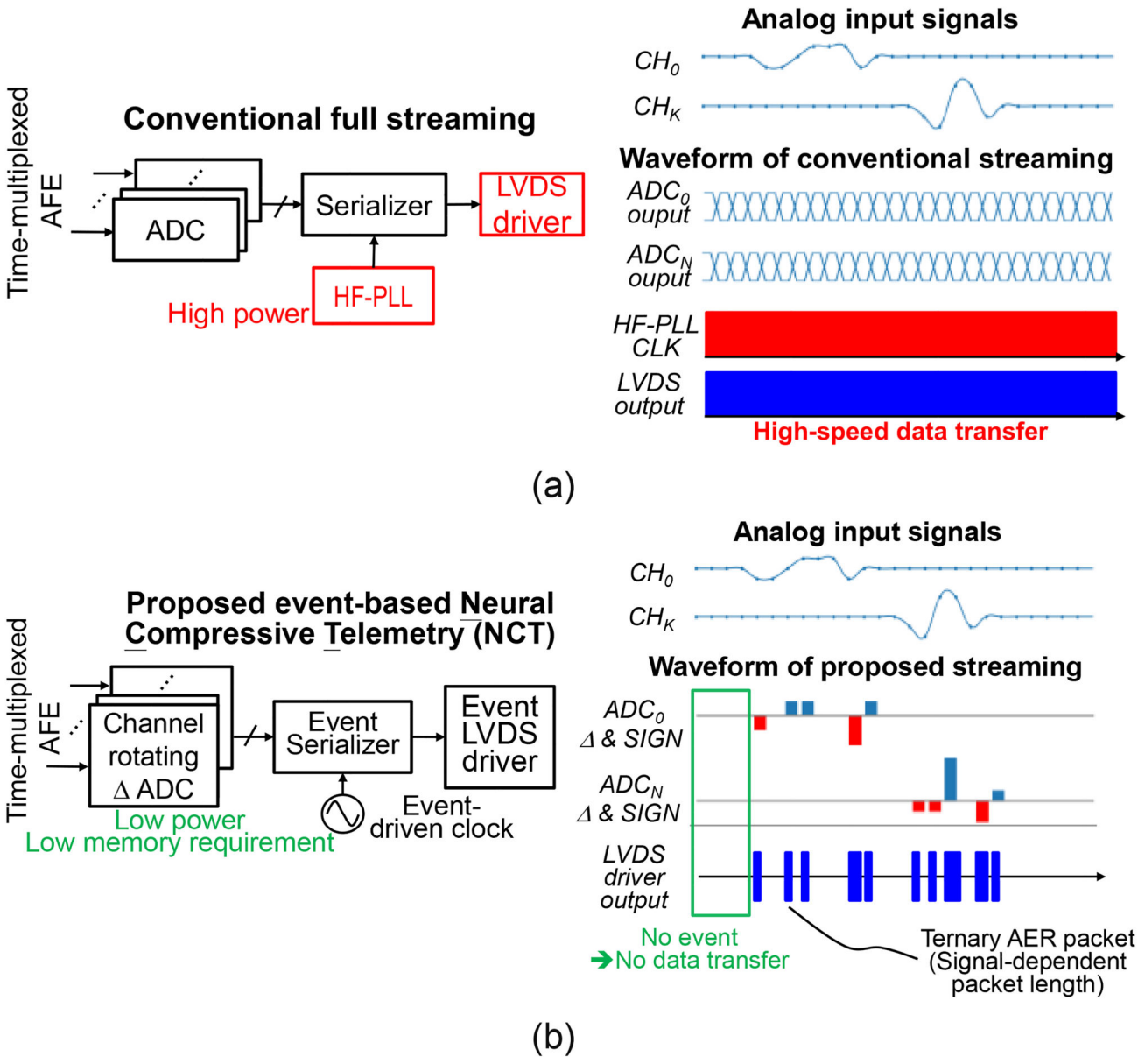

Fig. 2.
Block diagram of (a) conventional full streaming and (b) proposed NCT system.

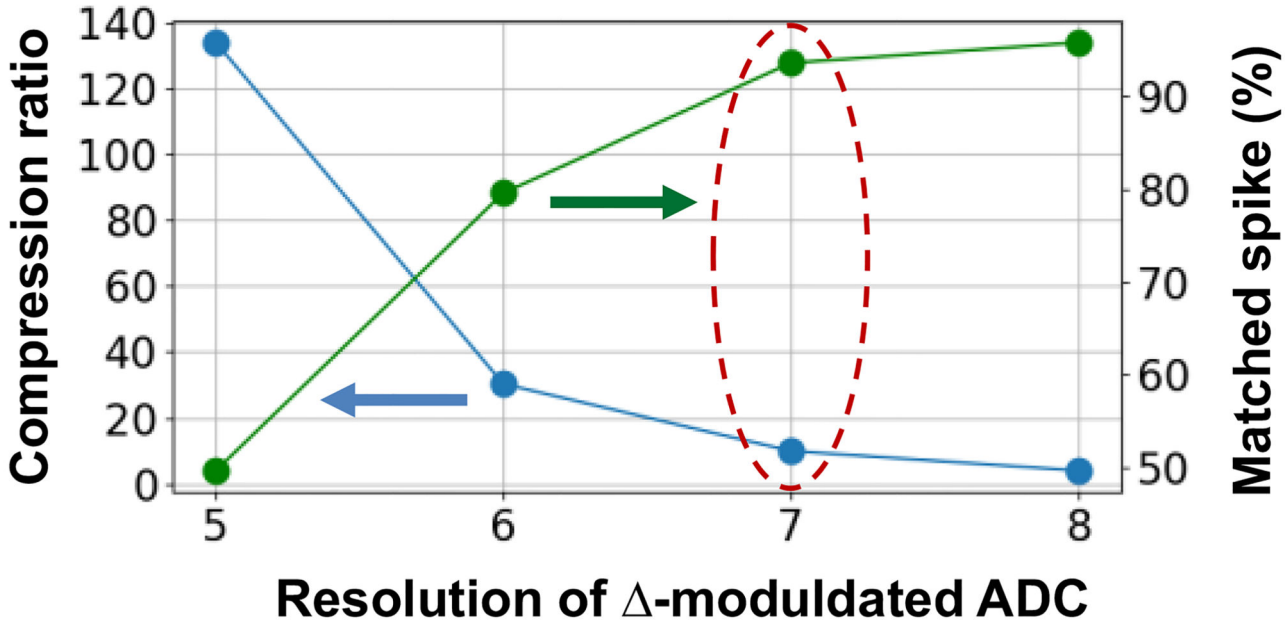

Fig. 3. Simulated compression ratio and percentage of matched spikes across a range of Δ-ADC resolutions.

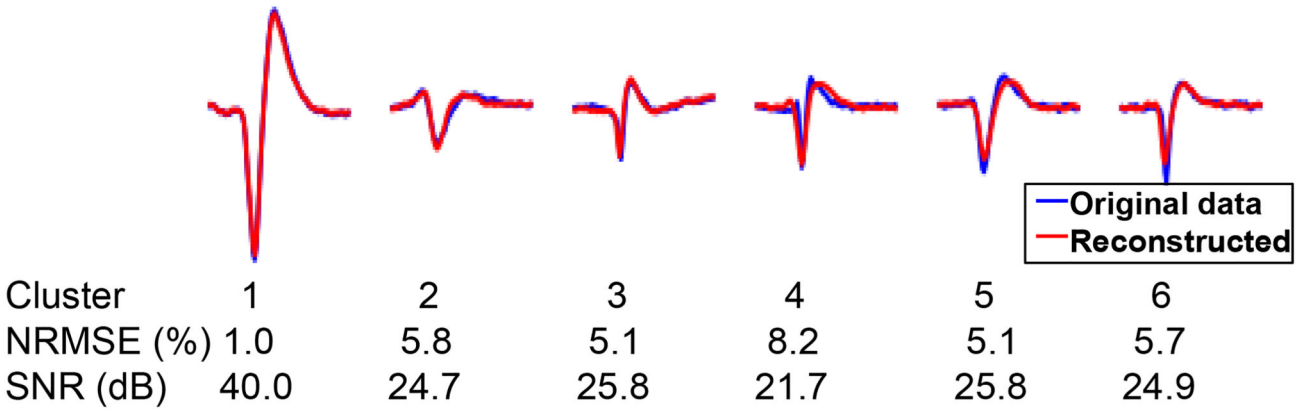

Fig. 4. Different spike clusters from raw dataset and reconstructed waveform from an equivalent model of 7-b Δ-ADCs, calculated NRSME and SNR.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

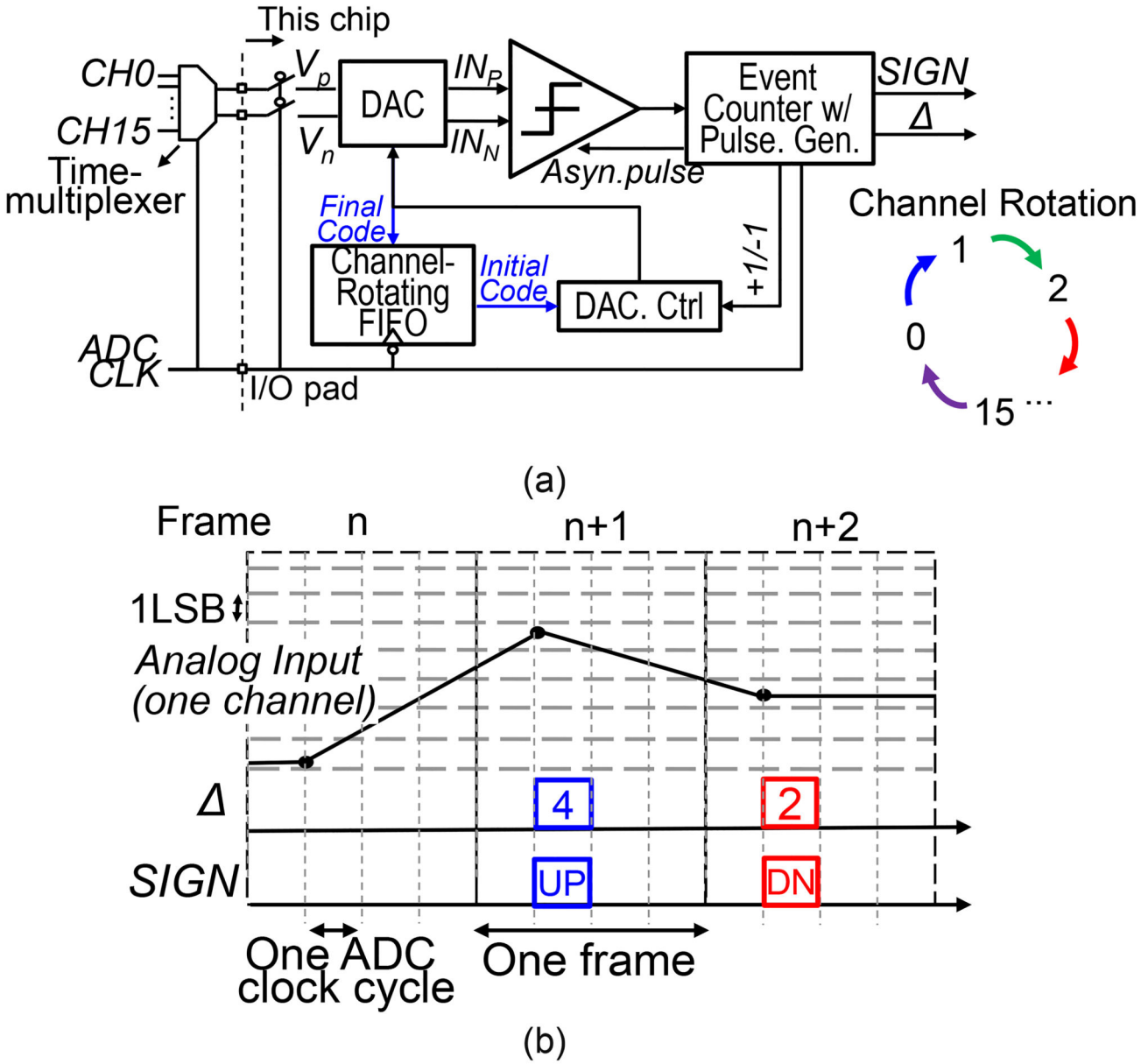

Fig. 5.

(a) Block diagram of CR-Δ ADC; (b) Conceptual waveform with an example of 4:1 time-multiplexing.

Europe PMC Funders Author Manuscripts

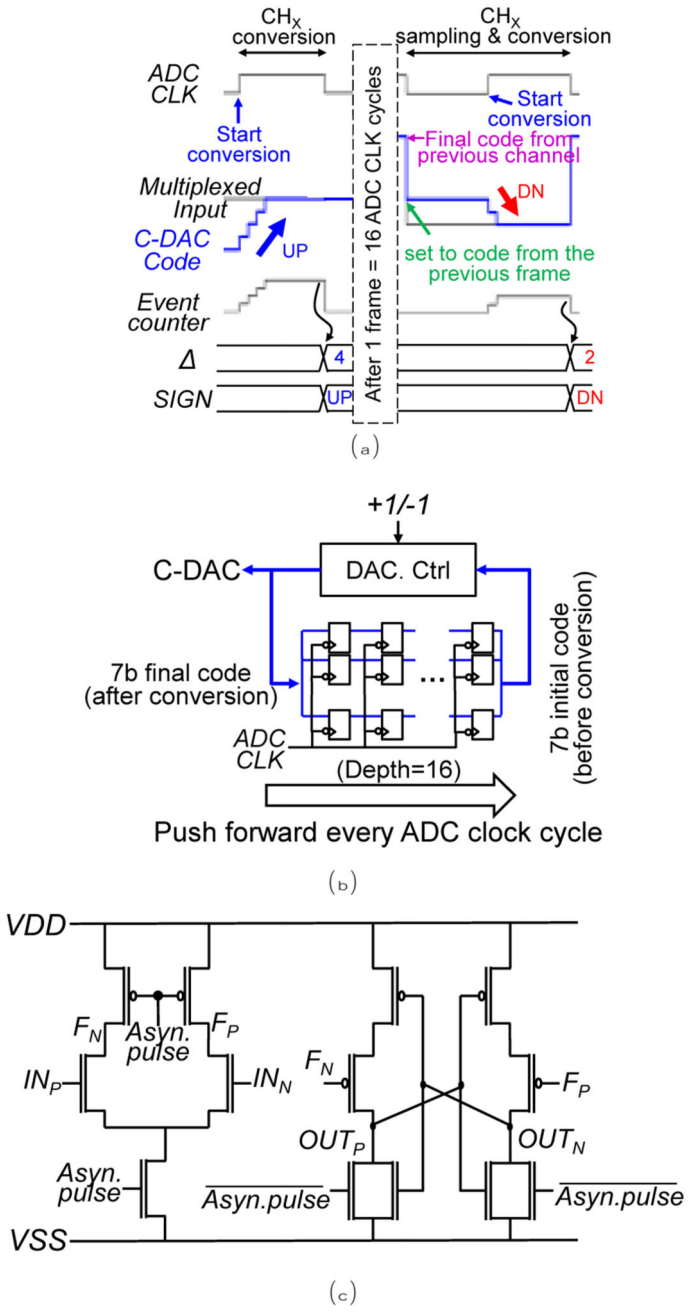

Fig. 6.

(a) Detailed operation of CR-Δ ADC; (b) Block diagram of the channel-rotation module; (c) Schematic of the comparator.

Europe PMC Funders Author Manuscripts

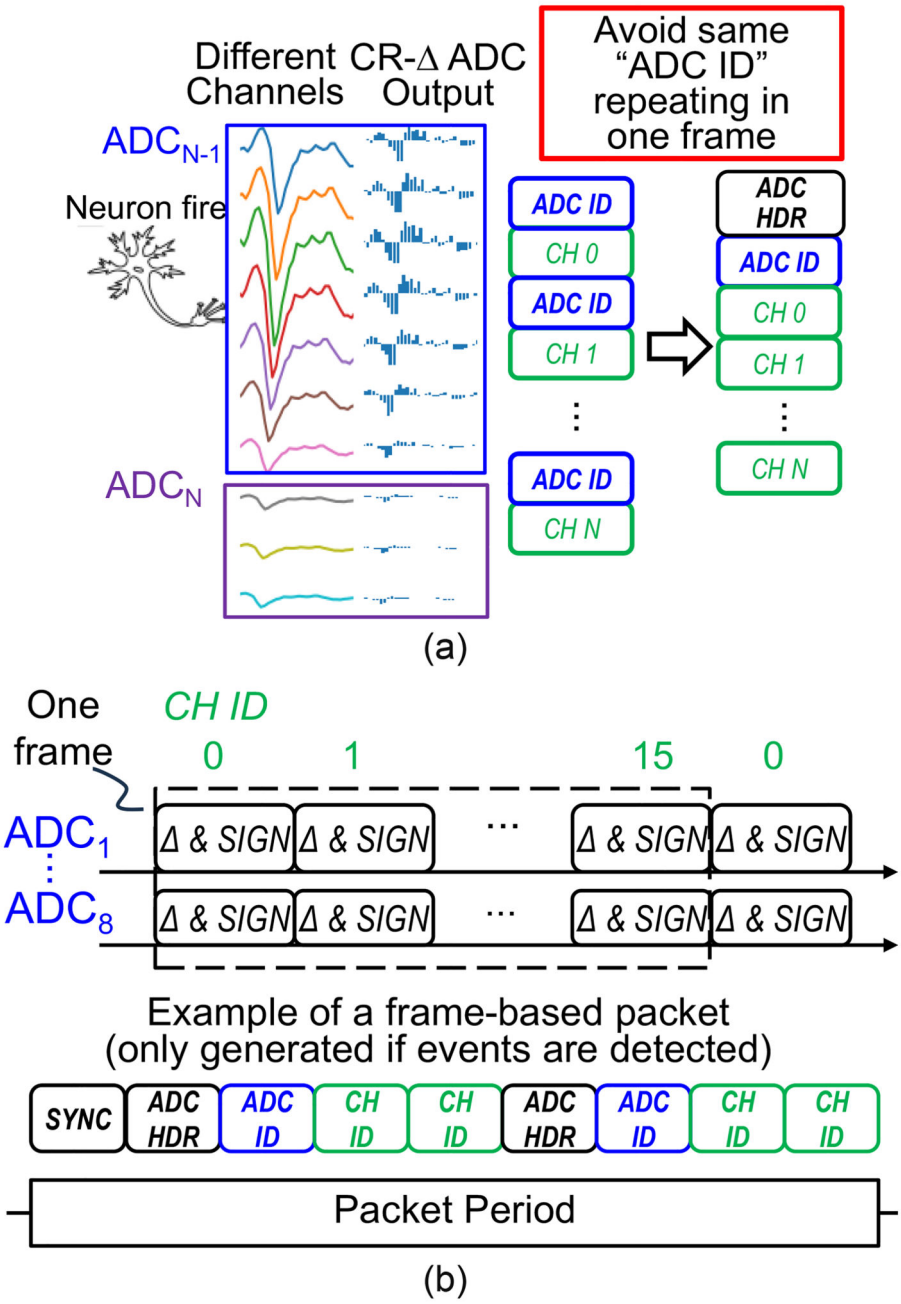

Fig. 7.

(a) Conceptual illustration of spatial grouping; (b) the proposed ternary AER packet with spatial grouping.

Europe PMC Funders Author Manuscripts

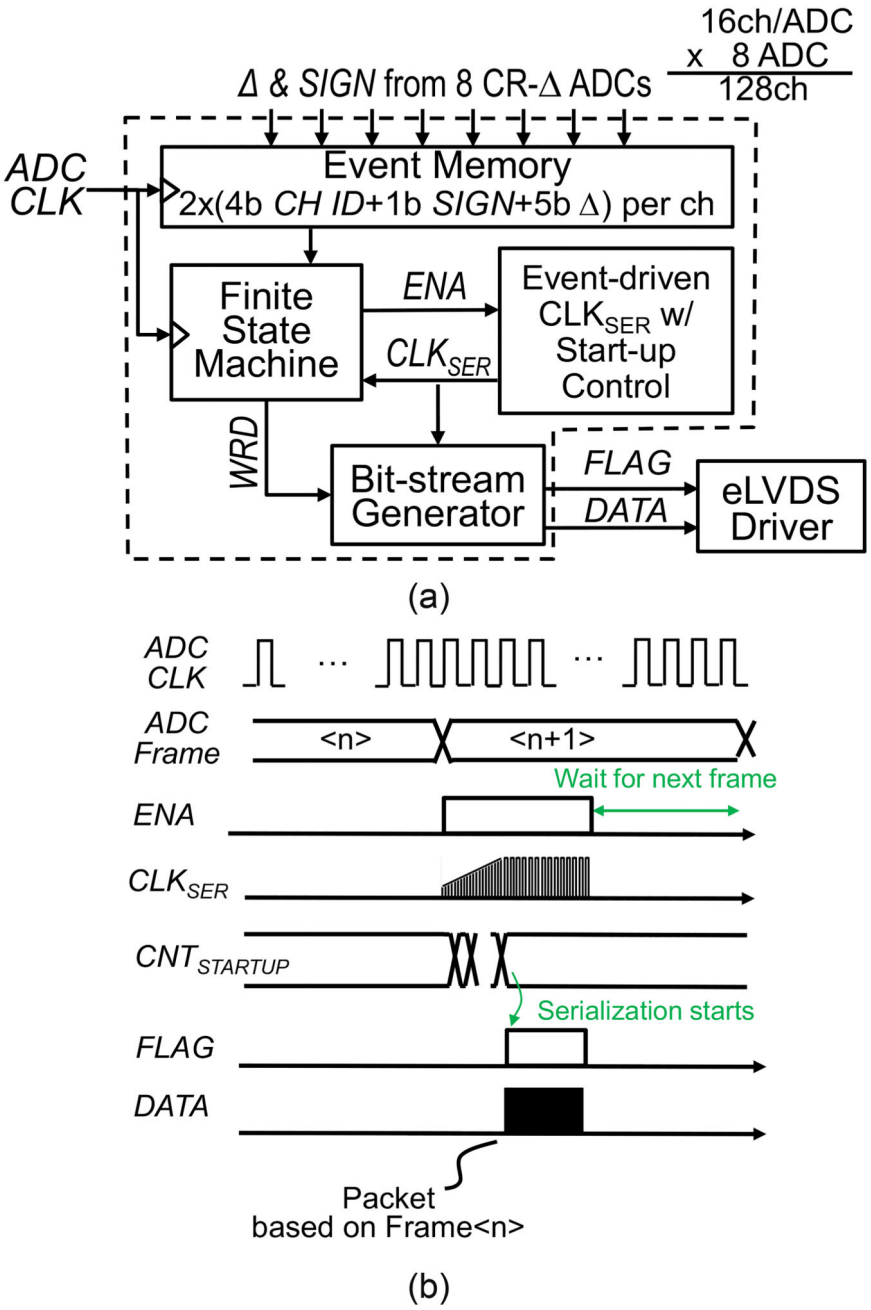

Fig. 8.

(a) Block diagram of event serializer; (b) Timing diagram of event serializer.

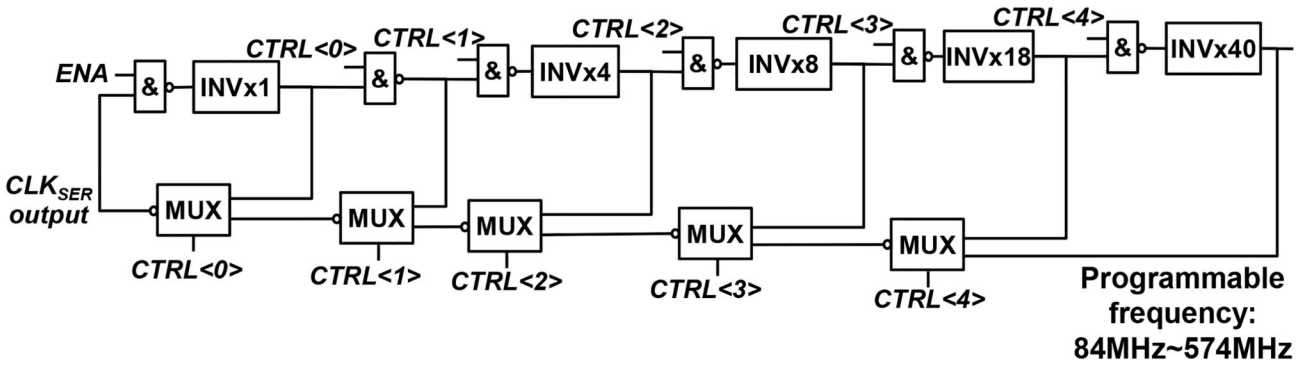

Fig. 9. Block diagram of fully-synthesized ring oscillator with frequency tuning.

Europe PMC Funders Author Manuscripts

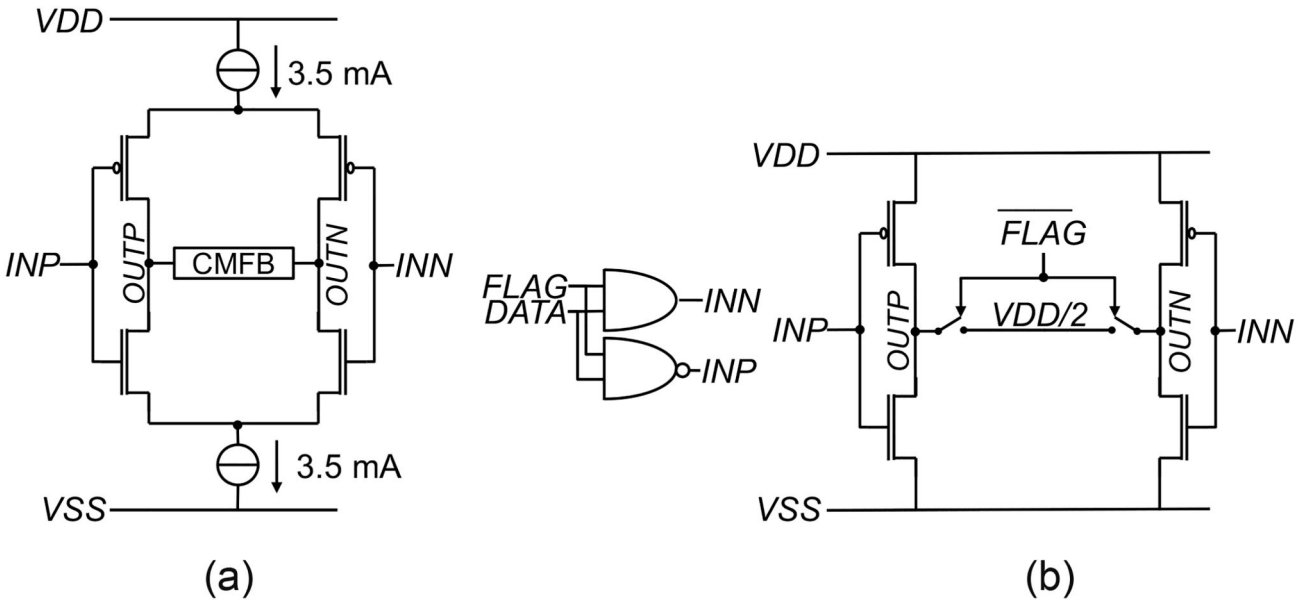

Fig. 10.
(a) Conceptual schematic of conventional LVDS driver; (b) Schematic of eLVDS driver.

Europe PMC Funders Author Manuscripts

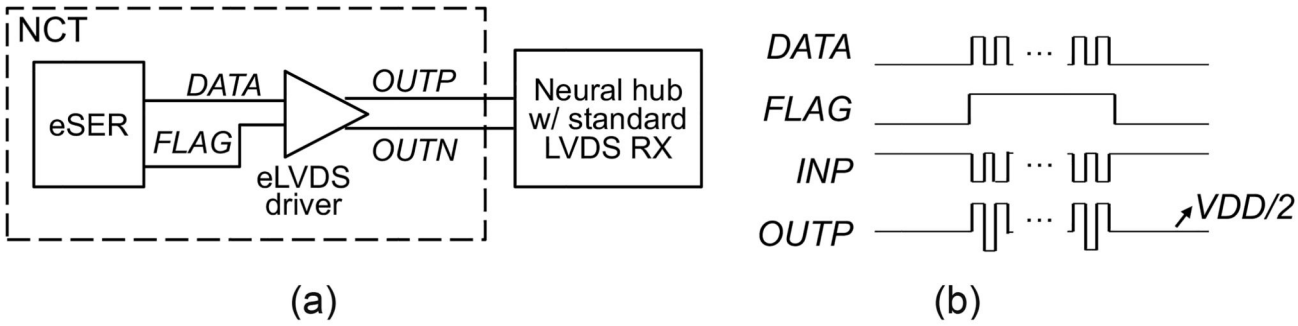

Fig. 11.

(a) Interface between eSER, eLVDS driver and neural hub; (b) Illustration waveform of eLVDS driver.

Europe PMC Funders Author Manuscripts

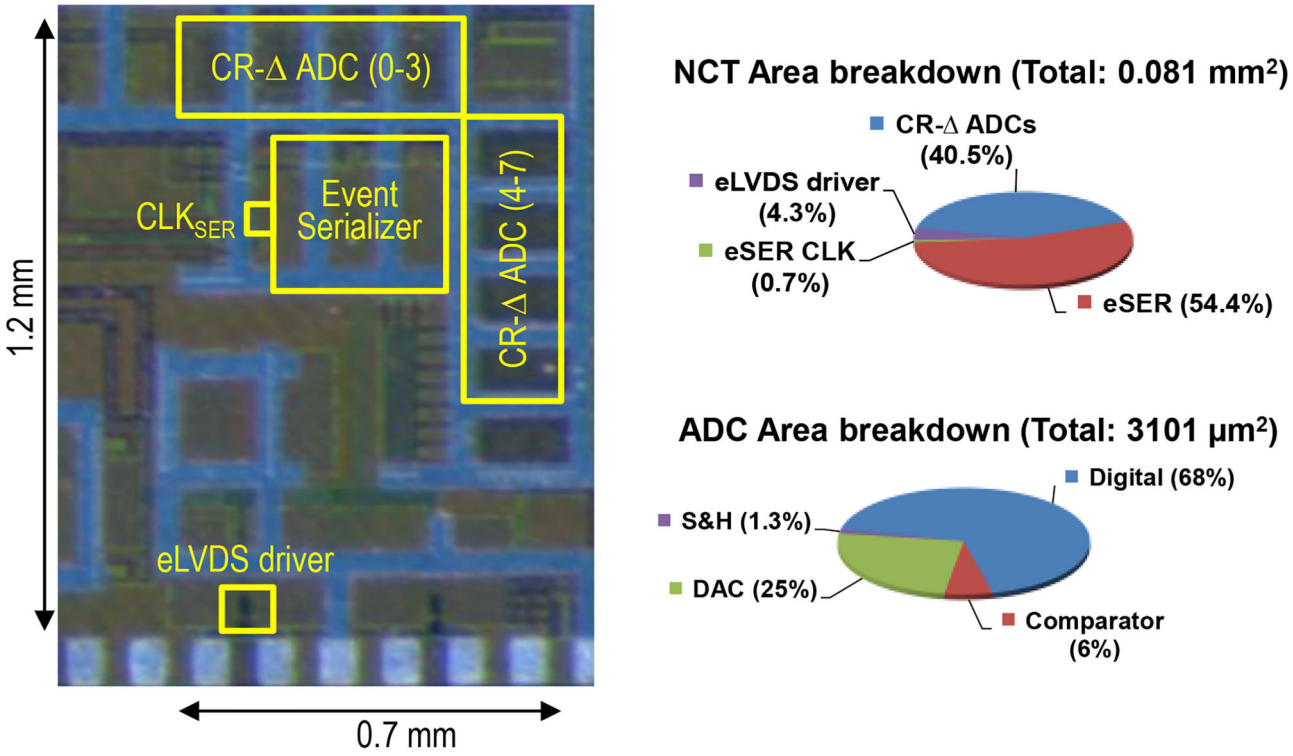

Fig. 12. Chip photo and area breakdown of NCT and single CR-Δ ADC.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

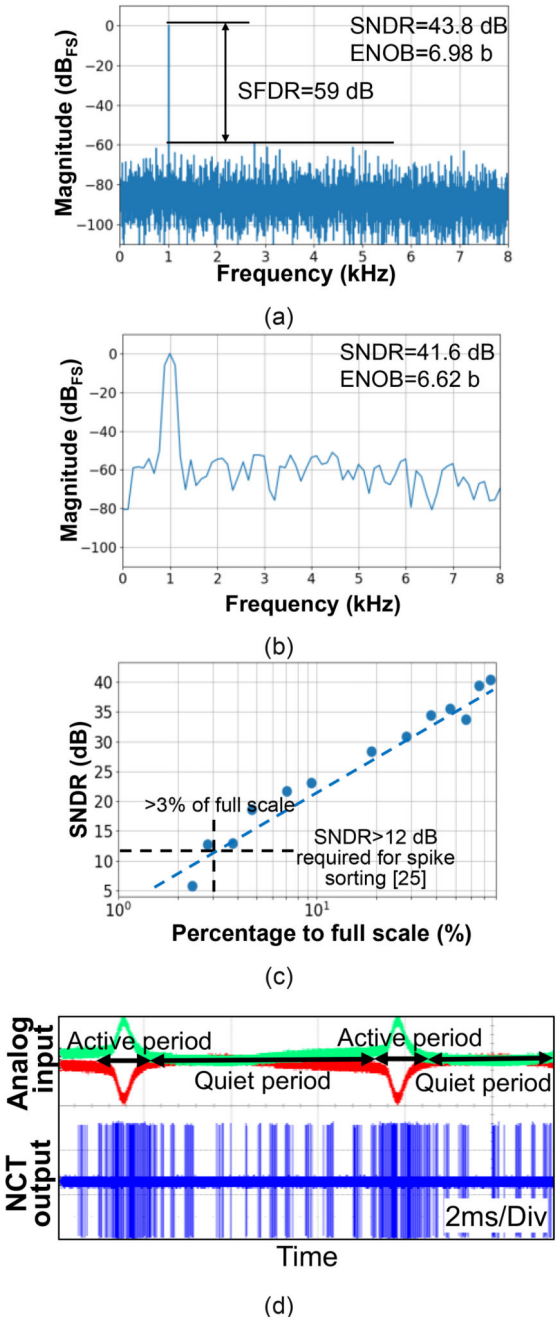

Fig. 13.

(a) Measured FFT and SNDR of the stand-alone CR-Δ ADC; (b) FFT and SNDR measured from NCT output; (c) measured SNDR across various input swing; (d) Measured NCT output with spike-like analog input.

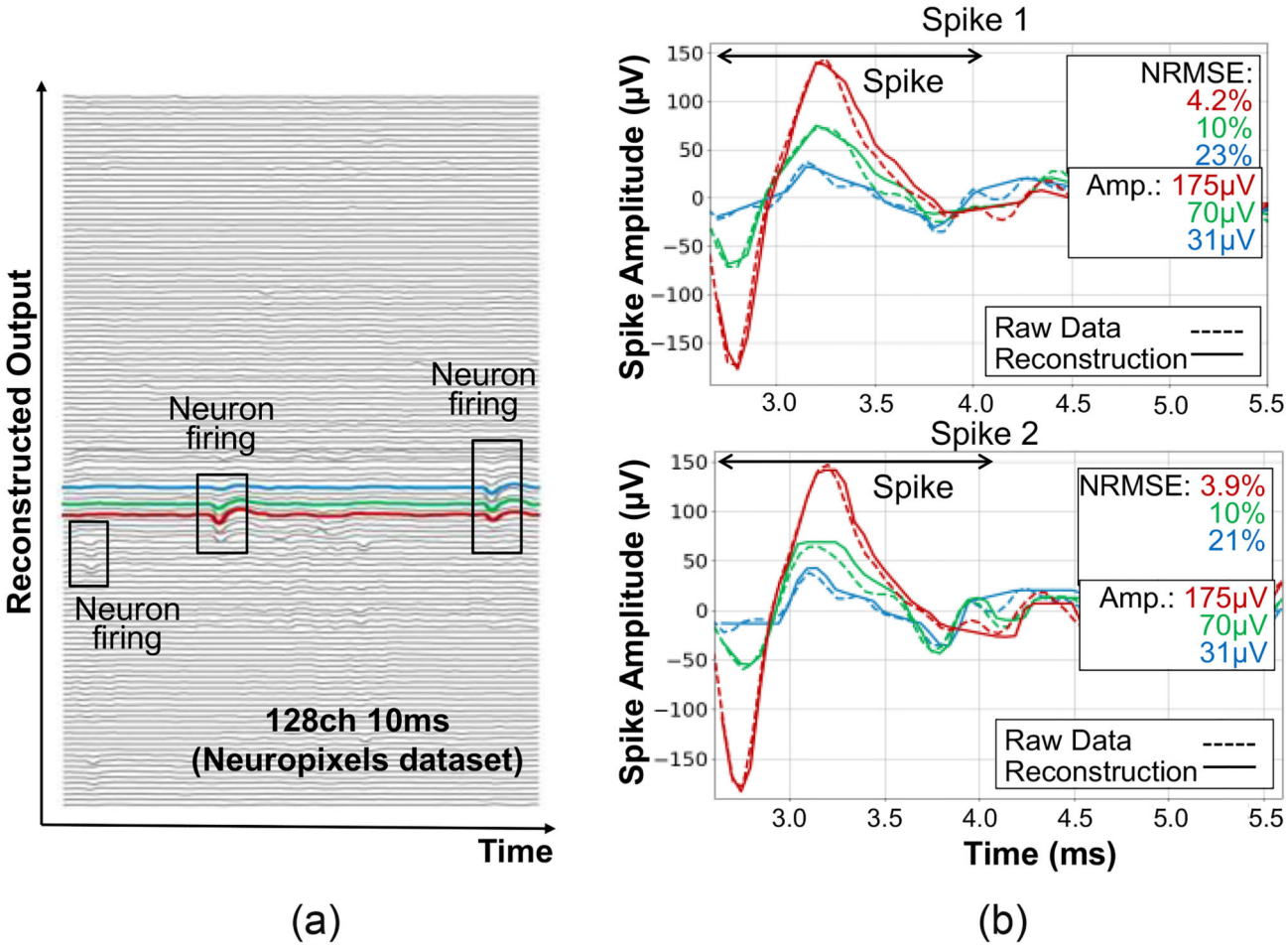

Fig. 14.

(a) Reconstructed output from dataset [19]; (b) two zoom-in version of neighboring channels with different spike amplitudes, from the same spike cluster.

Europe PMC Funders Author Manuscripts

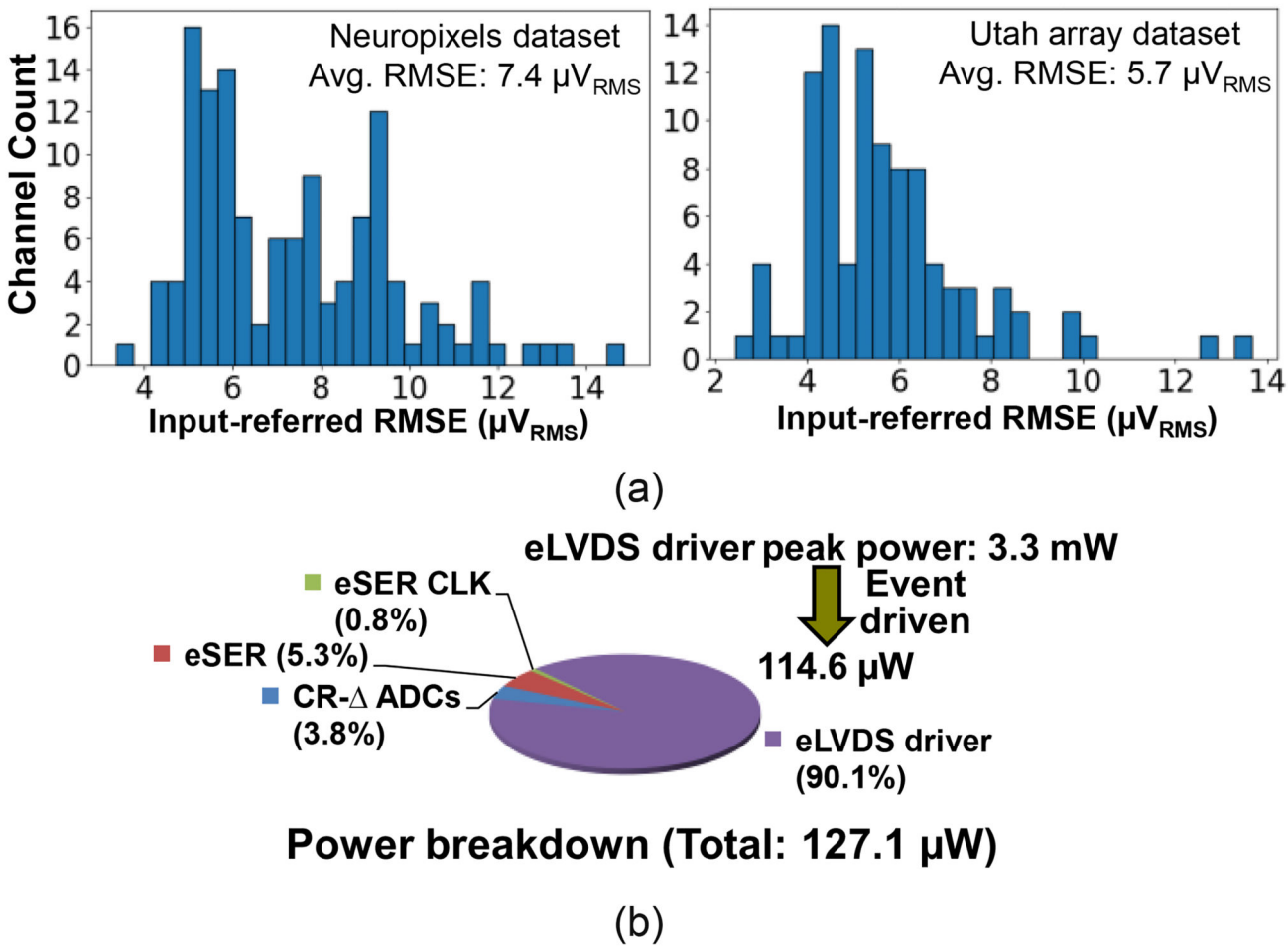

Fig. 15.

(a) Histogram of RMS error from different channels with two datasets; (b) Power breakdown of the proposed NCT.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

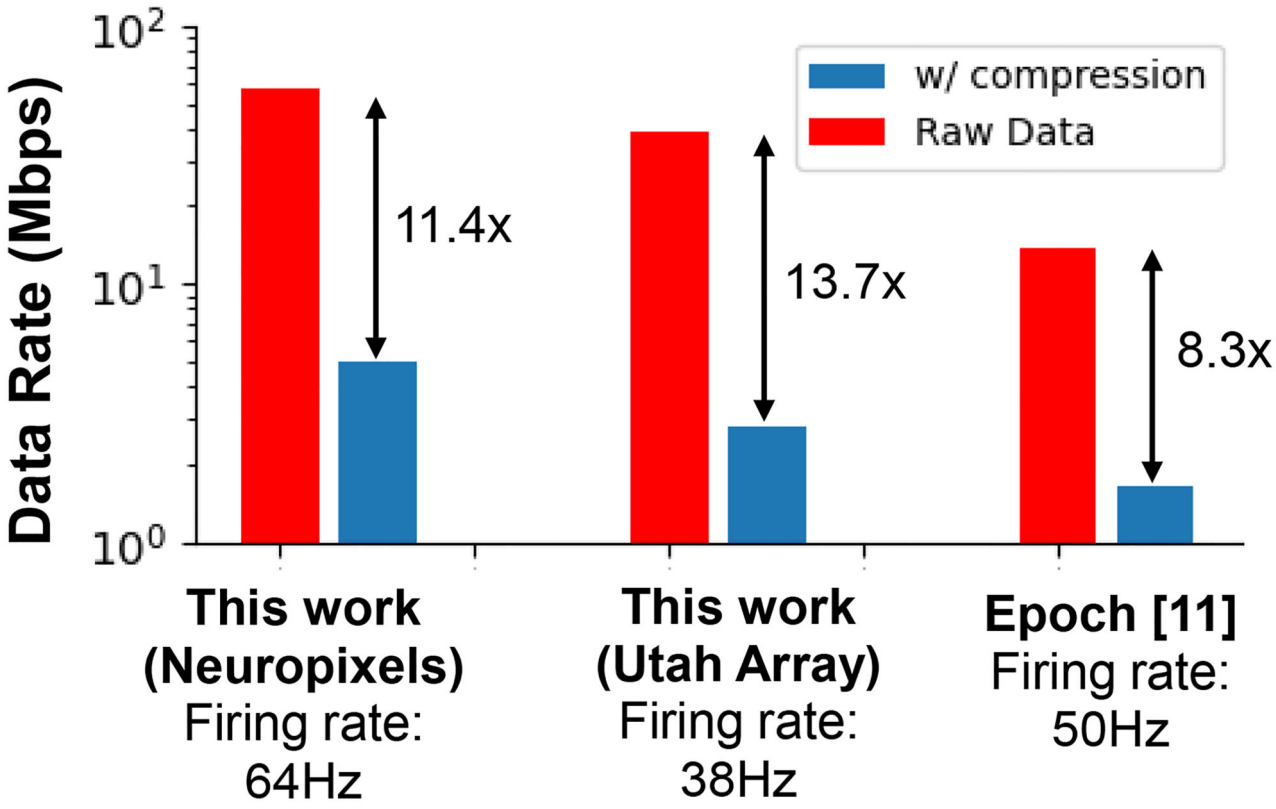

Fig. 16. Measured data compression with different datasets.

Table 1

Summarized Performance and Comparison to State-of-the-Art Neural Recording Systems with Data Compression or Serialization

		This work		Jang VLSI'23 [12]	Biederman JSSC'15 [11]	Lopez TBioCAS'17 [4]	Wang TBioCAS'19 [21]	He JSSC'22 [13]
Tech	nm	65		28	65	130	130	40
#Channel	-	128		1024	64	384	384	2
w/AFE	-	No		Yes	Yes	Yes	Yes	No
Digitization method	-	Discrete-time Δ-mod.		Ramp	SAR	SAR	SAR-assisted pineline	Continuous-time Δ -mod.
Compression method	-	Spatial group		Wired-OR	Epoch	NA	NA	NA
Power/ch								
Digitization (ADC)	μW	**0.02**		0.21	0.20	0.89	18.5	17
Compression		**0.02**		0.20	1.78	-	-	-
Serialization		**0.06**		-	-	58[A]	44[A]	7.5
LVDS driver		**0.90**		-	-	172[A]	186[A]	-
Area/ch								
Digitization (ADC)	μm^2	**256**		648	9200	9375[B]	5440	7000[B]
Compression		**70**		1903	13281	NA	NA	-
Memory/ch.	bit	**27**		-	1.5k	NA	NA	NA
MEA density of validated recording	-	**High (Neuropixels)**	Low (Utah array)	-	Low	High	High	NA
Data rate/ch	kbps	**39.1[C]**	29.3[C]	-	25.6	447[C]	420	-
Compression ratio (Avg. firing rate/ch)	-	**11.4 (64)[C]**	13.7 (38)[C]	12.5 (NA)	8 (50)	1	1	-
System ENOB	bit	6.62		~2.1[D]	-	8.1	8.6	-
Telemetry energy efficiency[E]	pJ/bit	**7.5**		NA	NA	953	963	NA

[A] MAX9271 is used as serializer with a power consumption of ~82 mW.

[B] Estimated from chip photo.

[C] Protocol overhead included.

[D] Minimal spike amplitude is limited due to a collision range of ~±25 LSB from an 8b ADC, and the system ENOB is calculated to be [20*log10(256/50)-1.76]/6.02.

[E] Based on Equation (2)