PAPER • **OPEN ACCESS**

Event-based SerDes telemetry network for distributed brain computer interfaces

To cite this article: Pietro Russo *et al* 2026 *Neuromorph. Comput. Eng.* **6** 014001

NEUROMORPHIC
Computing and Engineering

OPEN ACCESS

RECEIVED
12 September 2025

REVISED
5 December 2025

ACCEPTED FOR PUBLICATION
15 December 2025

PUBLISHED
31 December 2025

PAPER

Event-based SerDes telemetry network for distributed brain computer interfaces

Pietro Russo[1,2,*], Yuming He[1], Jac Romme[1], Stefano Traferro[1], Gert-Jan van Schaik[1], Hua-Peng Liaw[1], Zhong Ren[3], Zhenyu Gao[3], Guido Dolmans[1,2] and Yao-Hong Liu[1,2,4]

[1] IMEC, Eindhoven, The Netherlands
[2] Department of Electrical Engineering, Eindhoven University of Technology (TUe), Eindhoven, The Netherlands
[3] Erasmus Medical Center, Rotterdam, The Netherlands
[4] Department of EEMCS, Delft University of Technology, Delft, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: Pietro.Russo@imec.nl

Keywords: neural recording, implantable brain-computer interfaces, event-based, Serializer-Deserializer (SerDes), address event representation (AER), Manchester decoding, asynchronous network

Abstract

Intracortical brain computer interfaces hold the potential to revolutionize neurotherapeutics, but they must overcome technological challenges such as the high data rates generated by high-channel-count neural sensors and the stringent power and volume constraints of implantable devices. In addition, the brain-wide coverage needed for a deeper understanding of brain processes challenges the synchronization between distributed neural sensors and the central neural hub. To address these challenges, we present a deterministic-latency and power-efficient serializer–deserializer (SerDes) telemetry network that effectively mitigates the synchronization issue under strict power and volume constraints. The serializer on the sensor side employs event-based sampling and a packet-based address-event representation transmission protocol, achieving a low power consumption of only 127 μW and a low latency variation $<$10 μs. A crystal-free clock source is employed on the sensor side to minimize power consumption, with serialized data encoded using Manchester coding scheme. The deserializer on the hub handles the bit period uncertainty by counting and extracting the bit period of received data with a clock only $\sim$2.2$\times$ faster than the serializer clock. The proposed counting-based Manchester decoder achieves a wide frequency coverage up to 204 000 ppm of frequency variation. The deserializer achieves a measured Manchester decoding bit error rate (BER $<10^{-6}$), with a total estimated power consumption below 415 μW. The SerDes performance has been validated with *in vivo* pre-recorded data, demonstrating a compression ratio greater than 7, while preserving a high signal fidelity with an average RMSE $<$6 μV_{RMS}.

1. Introduction

Brain computer interfaces (BCIs) hold the potential to revolutionize medicine in treating cognitive, sensory, and motor neurological disorders. Among various BCIs methods, intra-cortical BCIs (iBCIs) have been demonstrated to offer the highest spatial and temporal resolutions [1]. Over the past decades, significant advancements in this field have led to breakthroughs such as prosthetic limb control [2, 3], speech restoration [4], etc.

Despite these advancements, the development of iBCIs still faces significant challenges, including limited neural recording channel count, large data volumes, and bulky form factors. For example, one of the high-density silicon-based neural probes, Neuropixels 2.0 [5], generates 161.3 Mbps of data across 384 recording channels. This data is typically transferred to a hub (or a headstage [6]) for further processing

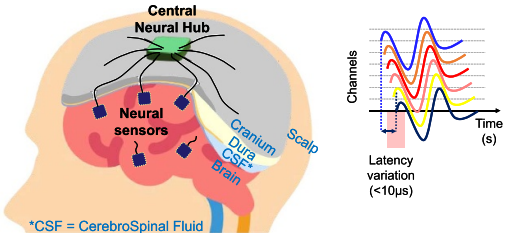

Figure 1. Distributed neural recording system, consisting of multiple neural implant sensors for data sampling and serialization, and a central neural hub for data recollection, synchronization, and transmission to external devices. Right: latency variation between spikes across different channels.

and communication with external devices. Ideally, a fully implantable iBCI would minimize invasiveness and risk of infections, reduce health risks associated with the surgery, and shorten patient's recovery time. However, processing such high data rates in a compact, implantable device presents significant challenges related to power consumption and volume constraints. In fact, implantable devices are subject to strict thermal flux limitation. In [7, 8], the recommended limit of heat flux is 40 mW cm^{-2}. This translates into the need for compression techniques to reduce data rates and optimize power efficiency while keeping the implant volume small. At the current state of the art, brain implant volume is in the order of hundreds of mm^3 [9]. With increasing number of channels, data rates, required implant volume, and power consumption will also continue to rise, further complicating the design of scalable, high-performance, and low-power BCIs.

The aforementioned approaches [2–4] focus on a single neural sensor probe. However, studies show that motor, sensory and cognitive actions involve multiple brain areas, with neural activity following specific pathways across these regions [10, 11]. Recording signals from multiple brain regions can provide a deeper understanding of brain processes, and a distributed neural recording system with multiple neural implant sensors could offer great benefits to neuroscience research. In fact, studies [12] suggest that the brain processes time-varying stimuli from multiple sources and integrates them to guide decision-making. A well-known example is the McGurk effect [13], where auditory and visual inputs interact to shape human perception [14] has shown that the two hemispheres work together to maintain short-term memory. Motor control is distributed across different regions of the motor cortex [15] and involves continuous interaction between the motor cortex and the cerebellum [16]. Therefore, since different brain regions contribute to processing distinct sensory and cognitive functions, recording from multiple areas and integrating the results, similar to how the brain naturally operates, could enhance our understanding of neural mechanisms. However, distributing the neural sensors increases the challenges related to the management of large data volumes and the synchronization between the sensors and the central hub. It also increases the system's invasiveness. Reserving a wireless link for each distributed neural sensor—to readout the data on an external machine—is impractical because of the high-power consumption and large volume of the module on or below the skull [17, 18], which needs to include also powering components (e.g., a battery), an antenna, digital signal processing, etc. A more volume- and power-efficient alternative uses a central neural hub [19] that collects the data from distributed neural sensors through wires, processes them and transmits them to the external device with a single wireless link, as illustrated in figure 1.

In [20], a distributed system of asynchronous neural implants is introduced; however, each implant supports only one channel, relies on a wireless link to an external hub—limiting scalability—and lacks compression techniques. In [21, 22] a flexible, ultra-low-power architecture for BCIs with asynchronous nodes and centralized processing is presented. However, it relies on computationally intensive methods that reduce data dimensionality by outputting aggregate features of the data. This approach compromises raw signal fidelity and temporal resolution, with latency reaching up to 50 ms. Similarly, in [23], advanced techniques based on spiking neural networks and associative learning are employed on each node to reduce data dimensionality in robotic skin for pain detection tasks. However, this approach does not preserve temporal information as well as original signal fidelity. While signal fidelity is less critical in low-dimensional decoding tasks (e.g. motor intention decoding) based on multi-unit activity, it is essential for some BCI applications with higher dimension tasks to decode with single-unit activity. In these cases, where spike sorting is necessary, maintaining high spike waveform fidelity is crucial.

Finally, the data transfer latency variation plays an important role in such a telemetry network, as illustrated in figure 1 (right). Extracellular action potential signals (i.e., spikes) have a narrow time

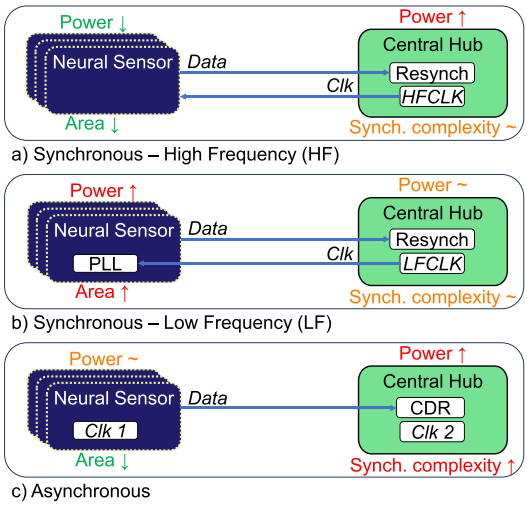

Figure 2. Star network synchronization. (a) Central hub sends a high frequency clock (*HFCLK*) to the neural sensors. (b) Central hub sends a low frequency clock (*LFCLK*) clock to the neural sensors which extract an *HFCLK* locally. (c) Each neural sensor uses its local clock.

duration, and the shortest spikes can last less than 200 ms [24]. Since the conduction velocity of a spike in the brain is ∼1 m s^{-1} [25], and given that modern high-density neural probes have electrode pitch in the order of 20 μm, the propagation delay of a spike across different channels is in the order of tens of microseconds. To preserve the precise temporal relationship between spikes from different channels, the system is designed to maintain data latency uncertainty well below 10 μs [26].

1.1. Serializer and deserializer (SerDes) for high channel count neural recording

In order to record large neural populations, modern neural sensors should support hundreds of recording channels. When handling data coming from a high number of recording channels, it is impractical to reserve a dedicated wire for each channel, due to high power dissipation, implant volume, and electromagnetic interference. For this reason, a pair SerDes is required to transfer data at high rates over a reduced number of wires. Various high-speed SerDes solutions are available in the literature and on the market, tailored for diverse application requirements [27, 28] introduced the Gigabit Multimedia Serial Link (GMSL), a SerDes solution for automotive video applications with differential signaling scheme. It enables robust serial transmission at speeds up to 12 Gbit s^{-1} over a 15 m cable length. Similar SerDes communication systems have been presented in [29–32], but they are all bulky and power consuming, which makes them not adoptable for implantable systems.

Partially-implantable high-channel count neural recording solutions have been shown in [33, 34], where active neural probes allow the simultaneously recording and reading-out of 384 channels. These solutions also use a GMSL serializer to serialize all the channels into a single pair of low-voltage differential signaling (LVDS) lines, but they consume high power (hundreds of milliwatts), which is challenging to manage in a fully implantable system.

The SerDes must be synchronized to ensure correct data transfer. Since a distributed neural recording system is typically configured as a star network, synchronization becomes even more challenging. Various approaches to address the synchronization issue in such systems are illustrated in figure 2.

The most straightforward approach synchronizes by sending a high frequency clock (*HFCLK*) signal from the hub to the neural sensors via dedicated clock lines, as shown in figure 2(a). Although this approach has relatively low complexity for the neural sensors, it requires a dedicated line for each of them, increasing the number of wires. Moreover, HFCLKs up to hundreds of MHz which are typically required in high-channel-count neural recording systems, lead to increased power consumption. Furthermore, such configuration is susceptible to skew between the *HFCLK* and *Data* signals, necessitating phase resynchronization at the receiving hub, increasing its complexity.

Another synchronization method, shown in figure 2(b) and adopted in [35], consists in transmitting a low-frequency clock (*LFCLK*) from the hub to the neural sensor, and then multiply it on the implant (with a phase-locked loop—PLL) to recreate the HFCLK needed to serialize data. However, this solution leads to higher power consumption and large area occupation on the implant sensors. Additionally, the resynchronization between the *LFCLK* and *Data* signals is still necessary in the hub.

The third alternative is the fully asynchronous approach, shown in figure 2(c), where each neural sensor encodes data using its local clock, and the hub recovers the clock from the received data (i.e., clock and data recovery—CDR). CDR circuits extract timing information from incoming data streams, ensuring synchronization without a separate clock signal. Common approaches include PLLs [36], which dynamically adjust the clock phase to align with data transitions. Alternatively, feedforward methods, such as oversampling or self-clocking coding schemes [36], recover timing by selecting the optimal transition point. The proposed solution reduces the implant's complexity, as no PLL or a precise clock generation is needed. However, this shifts the complexity to the hub to recover the clock. The hub can handle this complexity since it has less power and area constraints compared to the neural sensors, being implanted in a shallower position (e.g. above the skull or in the chest). Since the clock is not explicitly transferred, an alternative strategy to include this information within the data is needed. A straightforward approach is to encode the data with a self-clocking signal, like Manchester or 8b/10b coding [31, 37]. The clock recovery module will operate at relatively high frequencies, resulting in increased power consumption and reducing energy efficiency in an implantable neural hub.

1.2. Send-on-delta acquisition and address event representation (AER)

The synchronization challenge is further worsened by the high data rates generated by high-channel count neural sensors, which necessitate high frequency data transmission and high communication power. To mitigate these issues, low-loss compression, and efficient communication protocols, tailored to the unique characteristics of neural signals, are essential.

The firing activity of a neuron is relatively low, approximately 10's of Hz [38]. By exploiting the sparsity of intra-cortical neural signals, event-based send-on-delta acquisition can be performed [39] based on signal activity rather than at a fixed sampling frequency, reducing power consumption compared to conventional Nyquist-sampling approaches. This solution generates asynchronous events. To manage asynchronous data efficiently, event-based communication protocols have been developed. A widely adopted protocol in neuromorphic systems is the AER [40, 41]. The AER protocol encodes each event as a tuple of values (x, y, p), where (x, y) represents the spatial address of the event source (row and column) and p indicates the polarity of the event. Some approaches may include an additional value t, corresponding to the event's timestamp. This data is generated only when the source is actively producing events. In this communication approach, the time associated with each event is a crucial information for signal reconstruction.

A few advancements of AER protocols were introduced in literature: serial AER, group AER, and ternary packet-based AER.

Serial AER. Existing AER approaches typically employ an arbitration logic to serialize the data on a single bus [31, 42–44]. Compared to traditional AER, the serialized approach requires fewer wires. The arbiter with handshaking determines which channel accesses the shared bus and in what order, based on various polling methods [45, 46]. However, its complexity—in terms of power consumption and area—scales with the number of channels and the polling algorithm used to allocate the bus. Another disadvantage of these approaches is that the handshaking method required for arbitration introduces uncertainty in latency. Some approaches [43] use temporal grouping to collect multiple events within a time bin. While this reduces redundant channel address transmissions, it still requires the transmission of the full address for each channel.

Group AER. Although the AER perfectly manages asynchronous events, it introduces protocol overhead, as each event must include its source address. This overhead scales up with the number of channels, as more address bits are required. To solve this issue, an alternative solution is the 'group AER,' which combines the activity of multiple adjacent channels into a single group [47, 48]. Instead of sending the complete channel address for each event, this approach sends a group address and a fixed pattern containing the information relative to that group. Therefore, the protocol overhead is reduced. However, the aforementioned works also suffer from uncertain latency due to the arbitration, e.g., ∼0.25–250 ms in [47] depending on number of event collisions. To avoid such latency variation, additional bits for timestamping must be transferred, which further increases the overhead (e.g., extra 32 bits in [47], 34 bits in [48]).

Ternary packet-based AER. Existing AER communication approaches (including serial and grouping AER) all lack a synchronization pattern, making it challenging for the receiving side to recover data in a wireline communication system. In [49], we presented an event-based serializer that employs a new 'ternary packet-based AER' to handle these challenges. Unlike the previous approaches, this solution exploits 3 logic levels (i.e., ternary) to represent 3 different pieces of information: up, down, and no

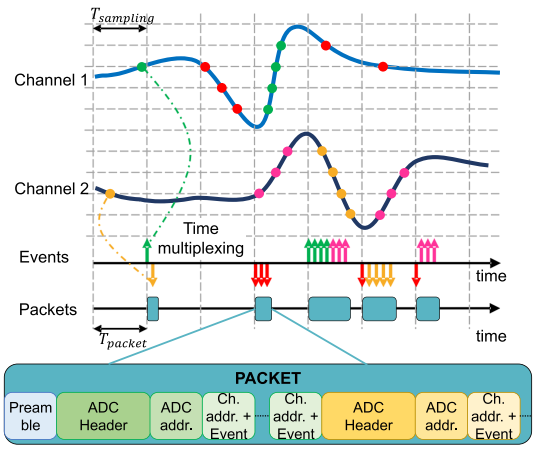

Figure 3. Ternary packet-based AER sampling and transmission. The signal is sampled when it crosses predefined thresholds. Events are collected into ternary-packets using the AER protocol. No packet will be generated when no event happens within T_{packet}.

event. Most important, asynchronous events are collected at uniform time intervals and grouped in small AER packets. Each packet includes a synchronization pattern (i.e., preamble), to overcome the challenge of synchronization between the neural sensors and the hub. Furthermore, instead of handshaking arbitration adopted in most of state-of-the-art AER readout, an agile time multiplexing is adopted to serialize the active channels. Additionally, without timing uncertainty introduced by arbitration, timestamping information is not required in the transmission packet, resulting in smaller packets that can be transmitted in real time, leading to a system with more deterministic latency.

1.3. Protocol design of the ternary packet-based AER

The ternary-packet-based AER has been adopted in [49]. The sampling is performed by 8 level-crossing ADCs (LC-ADCs) that sample the signal only on activity (i.e. when it crosses predefined voltage thresholds), thereby significantly reducing the number of bits to be transferred. The working principle of a LC-ADC and serializer [26, 39] is illustrated in figure 3. Unlike traditional approaches [50], demonstrated for a single channel and sending events as soon as they are generated, the proposed LC-ADC supports time-multiplexing across multiple channels and collects events within a short time packet [51] introduces a multiplexing strategy that shares a single ADC among four channels; however, it does not implement any compression techniques and relies on a large memory for data storage, which increases power and area requirements. In the proposed architecture, each LC-ADC is time-multiplexing among 16 channels to reduce the ASIC area (figure 3). The sampled data indicate the number of crossed thresholds and the direction of crossing (up or down) relative to the previous value, representing the quantized difference between successive signal values (i.e. delta—D).

These data points are packetized and Manchester-encoded for transmission. To maintain the time information of the event and to minimize data buffering on chip (which requires power-hungry memory access), each packet is generated in real time as soon as a new sampling point is acquired (thus sampling rate and packet rate are equivalent).

The ternary packet-based AER protocol is used to reduce the overhead: instead of the 9 bits typically required to address up to 512 channels, this approach sends 5 initial bits for the group address, followed by 4 bits per event for each channel. In fact, while traditional AER must transmit the full address for each event, the proposed approach minimizes AER protocol overhead by sharing portions of the address within groups. Our simulations with 100, 1k, and 10k channels show that the proposed AER generates fewer bits across all scenarios, achieving up to a $\sim$2.4$\times$ reduction compared to traditional AER. However, the benefit of sharing protocol parts with grouping becomes less effective when the sparsity increases (i.e. number of events is less). Simulations based on a pre-recorded *in-vivo* dataset in which events were removed randomly from the total number of events (to model a higher sparsity) show that the traditional AER can have less overhead than the proposed scheme when more than $\sim$85% of the original events are removed.

The packet length is not fixed, as it depends on activity. The complete AER packet, depicted in figure 3, consists of the following subparts (with the relative number of bits used):

- Preamble (5b): initial known sequence for packet detection and synchronization.
- ADC header (10b): golden code indicating the start of data for a new group (ADC) in the packet. The golden code is unique and will not occur within the data.
- ADC address (5b): identifies the active group (i.e., ADC).
- Channel address (4b): identifies the active channels within the active group.
- Event (1b): binary information indicating the data direction (up or down). The delta value is represented by sending the channel address and event multiple times corresponding to the number of thresholds crossed.

In conclusion, by grouping events from neighboring channels (spatial grouping) and collecting them in time bins (temporal grouping), we achieve low protocol overhead compared to traditional approaches [31, 42, 43], with acceptable impact on reconstruction quality (as detailed in section 3.3).

Based on numerical simulations, the resolution of the ADC is set to 7 bits, resulting in a quantization step of ∼10 mV. This value represents the best trade-off between reconstruction quality and event rate, as a larger quantization step would lower the event rate but impact reconstruction quality, while a smaller quantization step would generate more events triggered by noise. The proposed LC-ADCs and event-based serializer achieve a compression ratio $>11.4\times$, while preserving high reconstruction quality (RMS error $<8\ \mu V_{RMS}$) [49]. Thanks to the crystal-less event-based send-on-delta and the packet-AER approaches, the power consumption of the serializer, including a LVDS line driver, is only 127 mW.

In this work, we present a complete event-based SerDes link capable of streaming neural data from >100 channels in a power- and resource-efficient manner, making it suitable for distributed neural telemetry networks. It further extends from the previous serializer work in [49], by incorporating deserialization and addressing the critical aspects of communication and the feasibility of a multi-node star network. The key contributions of this work are as follows:

- Demonstrate the feasibility of the asynchronous star network.
- Design and optimization of an event-based deserializer.

The rest of this paper is organized as follows. Section 2 presents the proposed SerDes architecture. In section 3, SerDes system considerations and analysis supporting the implementation choices are presented. The measurement results and relative discussion are presented in section 4. Section 5 provides the conclusions. Discussion about future works is provided in section 6.

2. Proposed architecture

This section introduces the proposed distributed low-power asynchronous SerDes system, designed to handle the challenges outlined in the preceding section. The architecture shown in figure 4 illustrates the proposed concept.

Crystal-free deterministic-latency telemetry network. The system operates as an asynchronous network, where each neural sensor operates independently with its own clock. Each neural sensor employs event-based sampling and the ternary packet-based AER, so the data size is reduced. This solution eliminates the challenges of clock distribution, as no clock signal is distributed between serializers and the deserializer in the network. Instead of the crystal oscillator commonly required for proper synchronization in traditional telemetry networks, a low-power on-chip crystal-free clock, less accurate and stable than a crystal oscillator, is adopted in each implant sensor. This solution greatly relaxes power and the ASIC area required for precise clock generation and distribution. Since data is generated with an uncertain clock and the clock information is not transmitted to the hub, the self-clocking Manchester coding is employed, to allow the hub to recover the clock with low hardware complexity. In addition, no arbitration (which needs handshaking) is adopted in the network, which reduces uncertainty in data transfer latency. Finally, since no clock distribution is required, this network is also more scalable.

In this work, high-frequency data (at hundreds of MHz) is transferred using an LVDS scheme. A *Flag* signal is transmitted alongside with the data, informing the deserializer when a valid packet begins and ends, extending the event-based paradigm to the deserializer as well. Note that transmitting a *Flag* is more advantageous than transmitting a clock signal. *Flag* has very low frequency than data clock, so

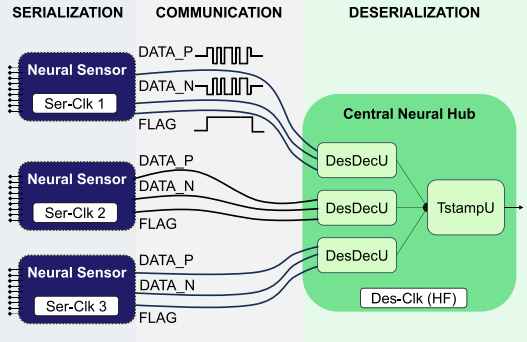

Figure 4. SerDes system proposed architecture. Each neural sensor generates differential *data* and *Flag* with a local clock. The central neural hub allocates multiple instances of the deserializer-decoding unit (*DesDecU*) to retrieve the original data. Ultimately, data are timestamped (*TstampU*) and realigned in a single output stream.

it is much more power efficient to deliver. It also does not require complex resynchronization at the deserializer side.

Event-based deserialization. An event-based deserializer is presented in this work to address the synchronization challenges introduced by the uncertain crystal-free clock source from the serializer, while minimizing the power consumption in the implant neural hub. The deserializer, consisting of a deserializer-decoding unit (*DesDecU*) and a timestamping unit (*TstampU*), retrieves signals from each neural sensor without precise clock information, handling data period uncertainty of 100 000's of parts per million (ppm) by decoding at a frequency only ∼2.2 higher than the encoding frequency. *TstampU* captures the time of arrival for each packet, enabling the reconstruction of the event-based signal on the correct temporal grid. Furthermore, the *TstampU* ensures a proper preservation of time information of events from different implant sensors. Instead of timestamping on the transmitting side as in most of the state-of-the-art AER readout, e.g. [47, 48], which requires to transfer more bits, the timestamping in this work is performed on the receiving side (i.e. deserializer). This approach can achieve more controlled latency.

The synchronization of the data with uncertain bit period at the *DesDecU* is achieved by extracting clock information from Manchester-encoded data. Manchester coding guarantees a transition in every symbol, which is essential for clock recovery. In addition, Manchester coding balances DC at the symbol level (same number of 0's and 1's), and this is critical to minimize the potential risk of electrical leakage, especially because miniature implants are more difficult to be packaged with bio-compatible materials.

Various Manchester decoders have been proposed in the literature. The 'fixed-delay' approach presented in [52], after an initial bit period estimation performed on a known sync word, detects the middle-bit transition contained in each Manchester symbol and samples the data after a precise interval of time (e.g. 1/4 of the estimated symbol period). Analog CDR-based methods can dynamically adjust the local clock frequency and phase by tracking the incoming data, but at the cost of increased power and area consumption [36, 53].

The proposed deserializer opted for the 'counter-based' approach [52] due to its capability of tolerating wide range of bit period uncertainty. This method exploits an intrinsic property of Manchester code, which ensures that a transition is always found in the middle of each symbol (mid-transition). It uses a deserializer decoding clock operating at a higher decoding frequency f_{Des} than the serializer encoding-bit frequency f_{bit}. By oversampling the incoming data with a certain oversampling ratio (OSR, i.e. the frequency ratio between f_{Des} and f_{bit}), the system can extract extra samples, allowing it to accurately decode the data even in the presence of large bit-period uncertainty. Figure 5 shows an example of deserialization of the packet-based AER. The decoder counts the extracted samples between two transitions of the incoming data (each black dashed line is a sample). If the next bit is the same as the previous bit, the time interval between two transitions is one bit period: it is a 'short transition.' Conversely, if the interval between transitions is two-bit periods, it is a 'long transition,' indicating the next bit is the complement of the previous bit. Therefore, by knowing the initial bit (e.g. last bit of the golden code—ADC header), the entire payload sequence can be decoded by counting the samples between consecutive transitions and properly setting a threshold to discriminate between short and long transition intervals.

Unlike traditional Manchester decoding approaches that use OSR $\geqslant 4$ [54, 55], the proposed approach reduces the OSR to 2.2, approaching the theoretical limit of 2, to minimize the required

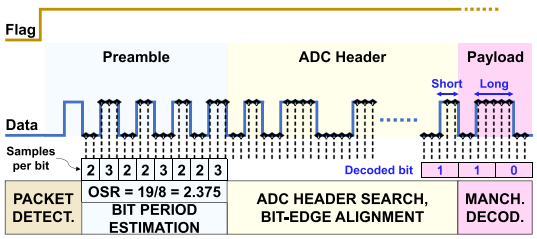

Figure 5. *DesDecU* architecture (with OSR of 2.375 as an example). It oversamples the incoming sequence extracting 2–3 samples per bit. It performs packet detection (PACKET DETECT.) based on *Flag*, bit period estimation on preamble, bit-edge alignment on ADC header and Manchester decoding (MANCH DECOD.) of the payload.

decoding frequency. Moreover, it handles bit period uncertainty of 100 000's of ppm, surpassing existing crystal-based synchronization approaches that can only support 10's ppm of uncertainty. A crystal-less wireless receiver in [56] employs a phase-tracking architecture and achieves a larger frequency tolerance of 320 ppm, but this still cannot meet the frequency tolerance required in this work.

The working principle of the *DesDecU* is shown in figure 5. The *Flag* signal is used to detect the packet and activate the decoding operations. The bit period estimation is performed based on the preamble, averaging the number of extracted samples (detailed in section 3.1). Note that since *Flag* may not be perfectly aligned with the start of preamble, i.e., the first preamble bits may be corrupted, the bit period is estimated using only four Manchester bits (i.e., two Manchester symbols) out of total ten bits of the preamble. Then, once the bit period is estimated, the ADC header sequence is searched, in order to align with its last bit and extract the payload. Finally, after knowing the last ADC header bit, the decoding of payload begins.

3. Event-based SerDes design considerations

The proposed event-based SerDes tackles the challenging trade-off between synchronization and power consumption in traditional SerDes systems. However, such a SerDes system, while offering high level of flexibility (e.g., packet rate, timestamping frequency, and OSR range) in accommodating a wide range of frequency uncertainty, also shows a trade-off among reconstruction quality, compression rate, and BERs. In this section, the considerations and analyses to support the design choices are discussed in detail.

The relationship among bit period, packet rate, OSR, and timestamping frequency can be summarized by the following three equations:

$$T_{\text{packet}} \geqslant N_{\text{bpp,max}} \times T_{\text{bit}} \rightarrow f_{\text{bit}} \geqslant N_{\text{bpp,max}} \times f_{\text{packet}} \tag{1}$$

$$\frac{f_{\text{Des}}}{f_{\text{bit}}} \geqslant \text{OSR}_{\text{min}} \tag{2}$$

$$T_{\text{stamp}} \langle T_{\text{packet}} \rightarrow f_{\text{stamp}} \rangle f_{\text{packet}} \tag{3}$$

where T_{packet} is the inverse of the packet rate (f_{packet}), $N_{\text{bpp,max}}$ is the maximum number of bits per packet (i.e. packet size), $T_{\text{bit}} = (f_{\text{bit}})^{-1}$ is the (Manchester-encoded) bit period, f_{Des} is the deserializer *DesDecU* frequency, OSR_{min} is the minimum oversampling ratio required to correctly decode the Manchester sequence, and T_{stamp} is the timestamping period.

Equation (1) ensures that the interval between two packets T_{pack} is large enough to accommodate the largest packet size $N_{\text{bpp,max}}$. Equation (2) indicates that the ratio between the decoding f_{Des} and encoding bit frequency f_{bit} must be larger than the OSR limit OSR_{min} supported by the decoder *DesDecU* (to be detailed in section 3.2), which also translates into a limitation for the encoding frequency f_{bit} (assuming the decoding speed f_{Des} will be limited by the maximum operation speed from the ASIC implementation). Equation (3) requires the timestamping frequency $f_{\text{stamp}} = (T_{\text{stamp}})^{-1}$ to be higher than the packet rate $f_{\text{pack}} = (T_{\text{pack}})^{-1}$, to avoid to under-sample the packets arrival time.

There are three primary system considerations in the proposed crystal-free telemetry network. First is the wide range of uncertainty of the packet rate and bit frequency, and this will be discussed in section 3.1. Second is the OSR range that can be supported by the prosed counter-based Manchester decoder (section 3.2). Finally, there is a trade-off among reconstruction quality, compression ratio, and packet loss, and this will be discussed in detail in section 3.3.

3.1. Uncertain packet rate and bit frequency

The system is configured as an asynchronous crystal-free star network. This gives arise to two types of challenges: bit frequency and packet rate are both unknown to the deserializer, and they can also vary over time due to many uncertainties in the environment, e.g., supply voltage, temperature, etc.

The bit period variability can be related to the process-voltage–temperature variations. Assuming a constant temperature in the body, and that the process variation effect can be mitigated via an initial calibration, voltage variations become the primary concern. The low dropout regulator, as the supply source of the clock generator of the designed serializer ASIC, has ∼5% uncertainty at a 1 V supply. This introduces frequency uncertainty. Given the measured supply sensitivity of the serializer clock of ∼500 MHz V^{-1}, a ±50 mV voltage uncertainty can induce an approximately ±25 MHz frequency uncertainty. With a central operating frequency of ∼160 MHz, this results in a frequency uncertainty of over 150 000 ppm. The total frequency uncertainty range, estimated in 100 000's of ppm, is wider compared to a typical 10's of ppm of frequency stability of crystal-based clock generation. The maximum frequency of the deserializer clock can be limited by the maximum achievable digital speed in target ASIC implementation in 65 nm CMOS process, which is approximately 500 MHz. To ensure a sufficient operation margin while consuming low power, the maximum decoding frequency is set to 400 MHz. By setting the OSR range to 2–3 (i.e. extracting from 2 to 3 samples per incoming bit), the *DesDecU* can support encoding bit frequency uncertainty from 133.33 MHz to 200 MHz, equivalent to approximately 400 000 ppm, covering the range needed to account for incoming data bit period uncertainty.

The bit frequency can be estimated from preamble, such that the *DesDecU* can operate in an OSR range that accommodates the incoming frequency. Bit period estimation is performed by averaging the number of extracted samples within a selected number of transitions (where two transitions contain a bit) in the preamble, as shown in equation (4):

$$T_{\text{bit, est.}} = \frac{\sum_{i=1}^{N_\text{b}} N_{\text{spb},i}}{N_\text{b}} \tag{4}$$

where $T_{\text{bit,est}}$ is the estimated bit period, $N_{\text{spb},i}$ is the number of samples extracted from the ith bit, and N_b is the total number of bits considered. When the signal-to-noise ratio is low, the bit period estimation may be affected by glitches in the preamble, leading to an incorrect estimation or, in the worst case, failure to recognize the preamble pattern, resulting in the packet being discarded. One example of the bit period estimation with an OSR of 2.375 is shown in figure 5. The OSR bounds are configurable, allowing different trade-offs between power consumption and frequency range to be achieved based on system requirements.

The uncertainty of the packet rate is addressed by timestamping each packet upon reception from the deserializer. This timestamp value is crucial for data reconstruction, as each data point will be associated with the corresponding timestamp: the quality of the reconstructed data is affected by timestamping resolution. Based on the analysis of signal-to-noise-and-distortion-ratio (SNDR) versus time uncertainty in [26], the timestamping frequency is set to 100 kHz.

3.2. Supported OSR range

The limits of the counter-based Manchester decoder have been investigated by simulating a sequence of random serialized data at different OSR and with an extra 10 000 ppm of random frequency fluctuation. Figure 6 shows the distribution of the extracted number of samples per bit at varying OSR. The blue bars represent the number of short transitions, and the green bars represent long transitions. The red bars indicate when it is not possible to discriminate between short and long transitions, a condition that generates wrong Manchester decoding, i.e., bit error. For fractional OSR values, the count distribution shifts to the nearest upper and lower integer values. As OSR decreases, the distributions for short and long transitions get closer. At OSR = 2, the overlapping of the distribution causes decoding errors, and the BER increases to ∼50%. Note that the state-of-the-art Manchester decoder uses a minimum OSR of 4 [54, 55]. As presented in section 4, the proposed architecture achieves a measured minimum OSR of 2.2 with a BER $<10^{-6}$, offering reliable decoding while keeping power consumption low.

3.3. Trade-off of reconstruction quality, packets loss, and compression ratio

Figure 7 presents the histogram of packet sizes obtained through numerical simulations for two *in-vivo* pre-recorded datasets, referred to as dataset 1 [57] and dataset 2 [58]. These datasets consist of 60 s *in vivo* recordings pre-collected from rodents using high-density Neuropixels sensors. The datasets are first bandpass filtered (300 Hz–6 kHz) to remove noise and long-term fluctuations affecting the measurements, and resampled at 30 kHz using the proposed LC-ADC. For dataset 2, the resulting event rate

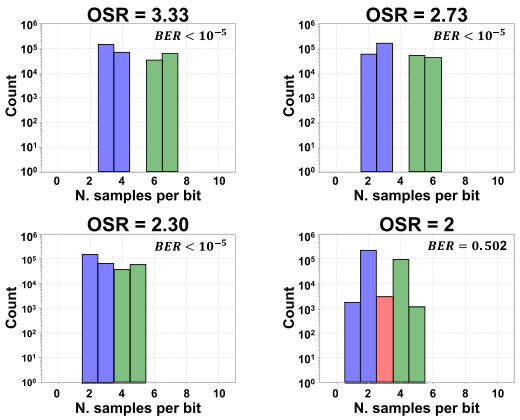

Figure 6. Number of extracted samples per bit at different OSR values. Blue: short transition intervals; green: long transition intervals; red: inconclusive interval.

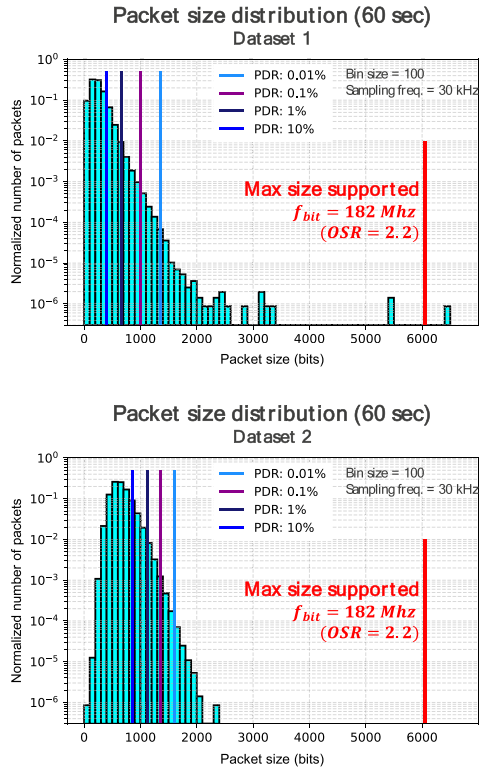

Figure 7. Packet size distribution at 30 kHz sampling frequency. The maximum supported packet size with OSR = 2.2 is 6060 bits.

is 2.81 Meps (Mega-events-per-second), with the largest packet containing 314 events, translating to 2320 bits (after including AER protocol and Manchester encoding). In contrast, dataset 1 exhibits a lower event rate of 1.31 Meps, but contains bursts of events across multiple channels, leading to a peak packet containing 1776 events. This explains the outlier packet size of 6440 bits.

Since the maximum encoding bit frequency is limited to ∼182 MHz (i.e., OSR = 2.2), packets exceeding 6060 bits will be truncated to the maximum supported size. This is ensured by a built-in saturation protection mechanism in the proposed serializer. However, according to the numerical analysis, this happens in extremely rare cases (less than 10^{-6}). Additionally, by selectively discarding the largest packets (setting a Packet Discard Rate—PDR), the system can limit the maximum packet size, potentially enabling support for a higher number of channels within the same bandwidth constraints.

As illustrated in figure 8, allowing a PDR of 1% for dataset 1 and 0.1% for dataset 2 has a negligible impact on the reconstruction quality, measured as the Root Mean Square Error (RMSE) between the

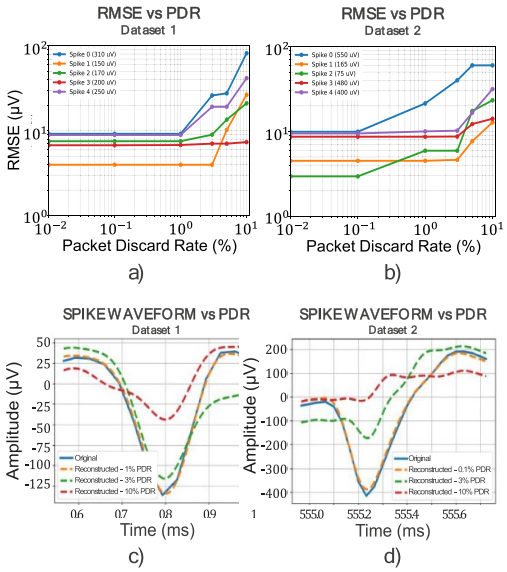

Figure 8. (a) and (b): RMSE vs Packet Discard Rate (PDR) for selected spikes of different amplitudes for datasets 1 and 2. (c) and (d): impact of PDR on selected spike waveforms from datasets 1 and 2.

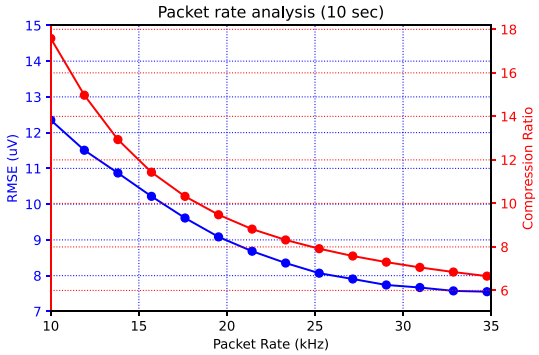

Figure 9. Simulated reconstruction quality (in terms of RMSE, blue) and compression ratio (red) at varying packet rates.

original and reconstructed data for different spikes. By allowing this packet discard, the maximum packet size is reduced to 700 bits and 1400 bits for dataset 1 and 2, respectively. Assuming a linear relationship between packet size and the number of channels, these PDR values enable the system to support 3324 and 1662 channels for dataset 1 and 2, respectively. These results demonstrate the system's scalability, supporting more than 1000 channels per implant.

Figure 9 illustrates the impact of packet/sampling rate on reconstruction quality and compression ratio, based on numerical simulations. The signal is reconstructed by accumulating the delta values over time and scaling them by the LC-ADC quantization step to restore the original mV scale. Both the reconstructed and original signals are then oversampled and interpolated at the same frequency for a point-by-point comparison. The reconstruction quality is evaluated by calculating the RMSE between the reconstructed waveform and the original *in vivo* pre-recorded data (dataset 2) given as input to the serializer. The compression ratio is defined as the ratio between the data rate of the pre-recorded data provided as the input and the data rate of the deserializer output, with the AER packet protocol overhead included.

The reconstruction quality curve (blue) shows that a higher packet rate is desirable because it has a better temporal resolution and a lower timing uncertainty of events. As explained in [26], higher event timing uncertainty degrades the SNDR. On the other hand, the compression ratio (red curve) benefits from a lower packet rate, as each generated packet includes a common protocol required for synchronization (i.e., preamble and ADC header) that contributes to the overall overhead. Since the neural recording sensor used for the input dataset (Neuropixels v1) filters the signal between 0.3–10 kHz, the packet frequency must be above 20 kHz to accurately represent the spike waveform.

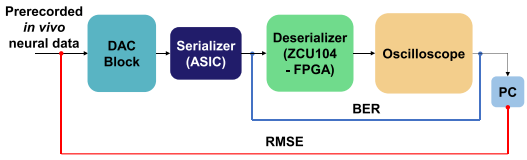

Figure 10. Experimental setup.

As shown in figure 9, the preferred packet rate is between 20 kHz and 30 kHz, to ensure a low reconstruction error while having a high compression ratio.

4. Measurement results and discussion

This section presents the hardware validation of the proposed event-based SerDes system. The serializer ASIC has been fabricated in 65 nm CMOS technology [49]. The deserializer is designed and validated with an AMD ZCU104 board, which integrates a Xilinx UltraScale+ FPGA.

The measurement setup is shown in figure 10. The serializer ASIC includes an analog input interface. To ensure fair validation of the SerDes system with neural data available in the neuroscience community, the *in-vivo* pre-recorded neural data (digital, dataset 2) is first converted into an analog signal using an off-the-shelf digital-to-analog conversion module. The decoded data from the deserializer is captured with a high-speed oscilloscope and numerically processed on a PC for reconstruction and analysis.

In the proposed measurements, the serializer is demonstrated with 128 neural recording channels. A packet rate of 30 kHz is chosen (as analyzed in section 3.3). The deserializer operates with a 400 MHz clock for the decoding operations, while a 100 kHz clock is used for timestamping.

Two key measurement indicators of the proposed SerDes system performance are BER and RMSE.

BER: assesses the quality of the Manchester decoding algorithm by comparing the originally encoded bits generated by the neural sensor at different encoding frequencies (and decoded in software) with the decoded output from the *DesDecU*.

RMSE: evaluates the end-to-end performance of the system by reconstructing the signal at the end of the data processing chain and comparing it with the original pre-recorded input to determine the amount of error introduced by the system.

To reconstruct the signal from the deserialized output, the decoded delta values are accumulated. Both the reconstructed data and the input data are oversampled and interpolated to extract the same number of timepoints from both in order to perform a point-to-point RMSE calculation.

Figure 11 displays multiple selected spike waveforms with different amplitudes. The results demonstrate that the reconstructed waveform closely matches the original waveform across various conditions, with an RMSE $<$8.50 μV_{RMS} for all the selected spikes with different amplitudes, indicating the system's ability to accurately capture the signal details. These results also highlight the benefits of the adopted deterministic-latency readout strategy, as each packet is accurately reconstructed on the correct time grid without any deformation in time, and no extra latency is incurred without transferring the timestamp information. The latency variation in the proposed SerDes telemetry is primarily determined by the short-term jitter of the packet rate clock, which is well below 10 ms. This ensure that the temporal relationship between distributed sensors can be accurately preserved.

Figure 12 shows the distribution of the RMSE between the reconstructed and input waveforms across 128 channels. All the channels present a RMSE below 10 μV_{RMS} with an average of 5.82 μV_{RMS}. This RMSE value is comparable to the noise floor of many high-channel count neural sensors [5, 34, 59], which is 7–12 μV_{RMS}.

Figure 13 shows the results of the *DesDecU* performance at different OSRs and encoding bit frequency f_{bit}. It effectively decodes data encoded at various frequencies within the range of ∼148 MHz to ∼182 MHz, corresponding to an OSR of 2.71–2.19, respectively. This capability demonstrates the decoder's ability to support a large OSR range, handling an encoding bit frequency uncertainty up to ∼204 000 ppm. This frequency tolerance is 630× wider than state of the art [56]. Above the OSR of 2.2, the system produced 0 bits error over more than 1.5 × 10^6 decoded bits. When the OSR drops below 2.2, approaching the expected limit of 2 (as analyzed in section 3.2.), bit errors due to bit period fluctuations begin to degrade the decoding performance. Additionally, the system can decode data with a

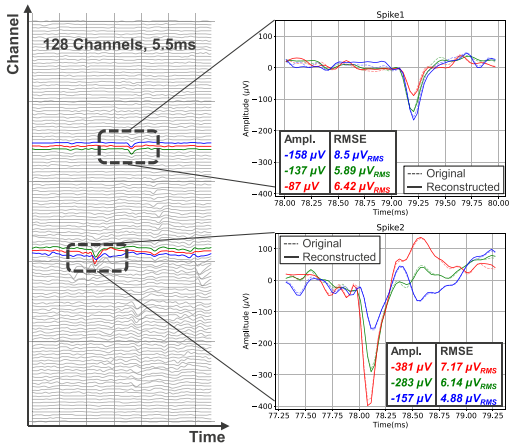

Figure 11. RMSE between reconstructed and original waveforms of spikes across neighbor channels.

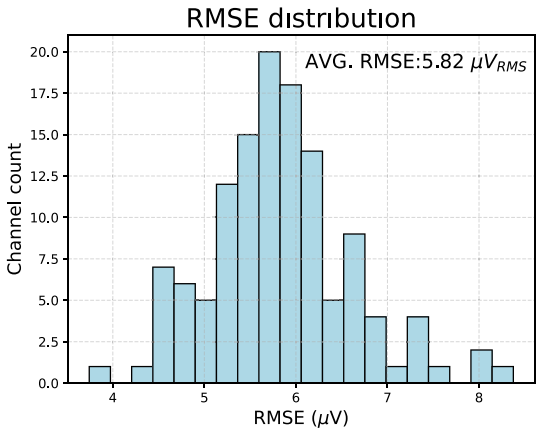

Figure 12. RMSE distribution across all the 128 channels.

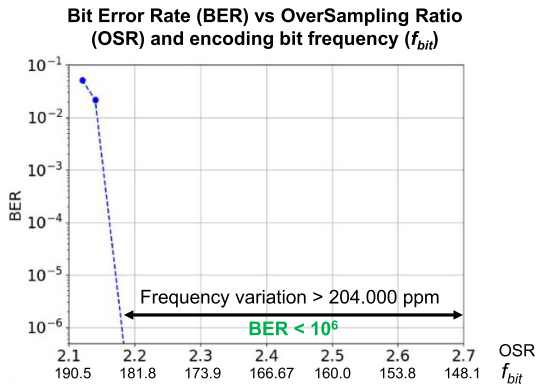

Figure 13. Decoding performance at varying OSR and bit frequency.

minimum OSR of 2.2, showing significant improvements in both power efficiency and decoding capability compared to the state-of-the-art [54, 55].

The deserializer has been designed and synthesized in ASIC 65 nm CMOS technology to extract the occupied ASIC area and the estimated power consumption in order to show the feasibility for an implantable application. The power report has been extracted by providing pre-recorded *in vivo* data to the serializer and further providing the serialized output to the deserializer. Packets are generated at 30 kHz and the serialized bits are encoded at 160 MHz, resulting in an OSR = 2.5 at the deserializer. The total deserializer power consumption is <415 μW. Considering an estimated total chip area of 2 mm^2, the power consumption generates a thermal flux of ∼20 mW cm^{-2}, well below the recommended limit for implantable devices of 40 mW cm^{-2} [7]. The dynamic power is dominant, since the low number of used cells achieves extremely low leakage power (0.15 μW). The dynamic power breakdown shows that 329 μW (79.5% of the total) is consumed by the *DesDecU* that runs in the 400 MHz clock domain, while the remaining 84.6 μW (20.5%) is associated with the 100 kHz timestamper (*TstampU)*.

To assess the risk of electronic heating, the system power consumption for neural sensor and hub is estimated. At the state-of-the-art, the power consumption of a neural sensor front-end electronics is ∼2.72 μW/channel [60], resulting in a total of 348.16 μW for 128 channels. The serializer [49] consumes a total of 127.1 μW (114.6 μW for the LVDS drivers, 12.5 μW for the LC-ADC and AER) for 128 channels. This leads to an estimated total implant power consumption of ∼475 μW (128 channels). On the hub side, the estimated power consumption amounts to 11.05 mW, where the LVDS receiver estimated power consumption is ∼0.63 mW, the deserializer unit requires 415 μW, and the hub-to-external wireless transmission module accounts for ∼10 mW, as reported in [56].

The system is designed for real-time processing, minimizing any local data storage or buffering. As detailed in [49], each neural implants requires only 27-bits register per channel to temporarily store data during serialization for packetization. Similarly, the *DesDecU* employs a single 32-bits shift register to temporarily hold serialized decoded data and a separate single 32-bits register for timestamping information, before outputting the data through an external pin for readout. This result highlights the low power and area nature of the proposed solution.

Regarding scalability, simulations from section 3.3 indicate that a single implant can support 1662–3324 channels under proper recording conditions. This result, while data- and activity-dependent, is consistent across two independent datasets, making it reasonable to conclude that over 1000 channels per site can be reliably supported. The proposed system does not inherently limit the number of recording sites, as each neural implant operates with a dedicated *DesDecU* in an independent SerDes link, maintaining stable reconstruction quality and latency. However, the practical limit on recording sites arises from factors such as area, power consumption, and surgical invasiveness. According to [61, 62], up to 20 implants have been demonstrated to be feasible for implantation in the human brain. Since each implant can support at least 1000 channels, the proposed architecture could theoretically scale to 20 000 recording channels.

In terms of power consumption, a single serializer consumes 1 μW per channel, leading to 1 mW for 1000 channels. Strategies to accommodate more channels per implant are discussed in section 6. The deserializer power, instead, remains constant at 415 μW, as the decoding bit frequency does not significantly change with the number of channels per implant. To support 1000 channels, the estimated total power consumption of each implant node and the hub is ∼3.72 mW (1 mW from serializer and 2.72 mW from neural sensor front-end) and 11.26 mW (0.415 mW from deserializer, 0.84 mW for LVDS receiver, and 10 mW from the hub-to-external wireless transmission module), respectively.

Finally, increasing the number of recording sites results in a proportional increase in total system power consumption. For 20 recording sites (20,000 channels), the estimated total power consumption of the hub increases to ∼35 mW (16,8 mW for the LVDS receivers, 8,3 mW for the deserializers, and 10 mW for hub-to-external wireless module). Given that the hub can occupy an area of a few cm^2, this power consumption remains within the thermal flux limit of 40 mW cm^{-2}.

Table 1 summarizes the performance of the proposed scheme in comparison with state-of-the-art neural recording systems with distribution or compression capabilities. With the proposed crystal-free telemetry network using event-based SerDes and ternary packet-based AER, the presented approach demonstrates the capability to scale up to 20 000 channels distributed from multiple implant nodes, each with low power consumption, while preserving high spike waveform fidelity and low latency uncertainty. It is important to note that the compression ratio could be improved to match the higher values reported in other works (for example, by increasing level crossing step sizes to generate fewer events), but this would come at the expense of reconstruction quality, which could drop dramatically.

Table 1. Summarized performance and comparison to state-of-the-art neural recording systems for BCIs with distribution or compression capabilities.

	[20]	[22]	[63]	[64]	[43]	This work
Compression	No compression	Feature extraction	Epoch	Wired-OR	Δ-modulation + serial AER	Δ-modulation + Ternary packet-based AER
Channel nr./node						
HW implemented	48	1	64	1024	—	128
Simulated capacity	770	—	—	—	Up to 10 000	1000–2000
Distribution capability & nodes	Yes 48 (meas.)	Yes 11 (sim.)	No	No	No	**Yes Up to 20 (sim.)**
Global synchronization	Required	Required	—	—	—	**Not required**
Compression ratio	1× (meas.)	100× (sim.)	8.3×[a] (meas.)	12.5× (meas.)	15-256× (sim.)	7-12× (meas.)
Compression error	—	Spikes not recoverable	N/A	N/A [b]	N/A [c]	**5-10 μV_{rms} (meas.)**
Latency (& uncertainty)	100 ms (N/A)	50 ms (N/A)	N/A	N/A	N/A	**<33 μs (10 μs)**
Memory req.	N/A	1Tb /ch	1500b /ch	N/A	N/A	**27b /ch**
Estimated power consumption						
Neural sensor	N/A	N/A	3.6 μW/ch	0.5 μW/ch	N/A	1 μW/ch
Hub	N/A	N/A	N/A	N/A	N/A	35 mW (20 nodes)

[a] Considering the Epochs streaming mode;

[b] When the spike amplitude is lower than ∼25 LSBs (75 μV_{peak} with a full scale of 0.75 mV_{pp}), compression error can be introduced by collisions (estimated from the measurements results in figure 6).

[c] A normalized RMSE of 0.0881 is reported based on simulation, but the full scale of spike amplitude is not available.

5. Conclusion

In this work, we presented a deterministic-latency and power-efficient SerDes telemetry for neural implants. The adoption of an event-based crystal-less SerDes network reduces the power consumption by eliminating the need for Nyquist-sampling and clock distribution, while preserving the temporal information associated with data, achieving a record-low latency variation well below 10 ms. Ternary packet-based AER serialization enables data compression and synchronization between neural sensors (i.e., serializers) and the central hub (deserializer), contributing to an overall communication power reduction. A wide frequency tolerance of up to 204 000 ppm is achieved by the counter-based Manchester decoder in the deserializer. This decoder has a counting frequency only 2.2× higher than the encoding bit frequency, with an estimated power consumption less than 415 μW. The system has been validated with *in-vivo* pre-recorded data, and it demonstrates a high reconstruction quality (average RMSE <6 μV_{RMS}) and a high compression ratio (> 7). The system's ability to operate without clock information from neural sensors to the central hub proves its scalability and support for a distributed architecture.

6. Future works

As part of our ongoing efforts to expand the capabilities of the proposed system, we plan to increase the number of channels. We envision the following ways to scale up the system:

- **Increase the number of channels per implant**. To support more channels in a single packet, it is possible to extend the packet period to accommodate more events, to increase the quantization step in order to reduce the event-rate, or to discard portions of packets that exceed the packet period. Extending the packet period would increase latency, negatively affecting the reconstruction quality. On the other hand, increasing the quantization step or discard packets would reduce signal quality but not affect latency (as the packet period is unchanged). As shown in figure 8, allowing up to 1% packet

loss results in a negligible impact on RMSE, suggesting that discarding the largest packets may be the most effective strategy for increasing the number of channels per implant.
- **Enhance deserializer speed**. To handle higher data rates, it is possible to increase decoding speed (moving to an ASIC implementation of the deserializer for better speed and power trade-off). This would allow for higher encoding frequencies within the same packet period, enabling linear scaling of the number of channels without increasing latency.
- **Expand the number of neural implants**. Increasing the number of neural implants (e.g., 8 probes with 128 channels each) ensures that latency per neural implant remains unchanged. The decoding hub's latency is also unaffected, as each serializer would have a dedicated *DesDecU* and temporal information would still be preserved by timestamping data upon arrival. However, this would also result in higher power consumption at the deserializer.
- **Additional encoding methods.** Further optimizations, such as pre-filtering data to reduce spatial redundancy, e.g., grouping events, or employing different DC balance schemes that has less overhead than Manchester encoding (e.g., 8b/10b), may also be explored to help manage system load and improve performance.

Future works will focus on evaluating the feasibility of these approaches and integrating them into the system architecture.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: Dataset 1: https://zenodo.org/records/14449576 [57] Dataset 2: doi: https://doi.org/10.5522/04/25232962.v2 [58].

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 and Horizon Europe research and innovation program (Grant Agreement No. 101001448 and 101138283).

Electrophysiological recordings

An adult C57Bl/6 mouse was included in the experiment [57]. The experiment was approved by the institutional animal welfare committee of Erasmus MC and the national committee CCD. An acute craniotomy was made above the mouse cerebellum. Acute Neuropixel probes (IMEC) were vertically penetrated into the cerebellar cortex via an electrode holder controlled by manipulator (Sensapex). Signals from the probe were recorded using the Open Ephys software. All animals' experiments were approved by the national Central Commissie Dierproeven and the institutional animal welfare committee of Erasmus MC.

ORCID iDs

Pietro Russo 0009-0004-2070-6809
Yuming He 0000-0002-7812-1916
Jac Romme 0000-0002-8936-2993
Stefano Traferro 0000-0002-3468-4901
Hua-Peng Liaw 0009-0003-4721-3556
Zhenyu Gao 0000-0002-4979-2366
Guido Dolmans 0000-0002-7744-5054
Yao-Hong Liu 0000-0002-3256-6741

References

[1] Brandman D M, Cash S S and Hochberg L R 2017 Review: human intracortical recording and neural decoding for brain–computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1687–96
[2] Hochberg L R *et al* 2012 Reach and grasp by people with tetraplegia using a neurally controlled robotic arm *Nature* **485** 372–5
[3] Simeral J D *et al* 2021 Home use of a percutaneous wireless intracortical brain-computer interface by individuals with tetraplegia *IEEE Trans. Biomed. Eng.* **68** 2313–25
[4] Willett F R *et al* 2023 A high-performance speech neuroprosthesis *bioRxiv* **620** 524489
[5] Steinmetz N A *et al* 2021 Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings *Science* **372** eabf4588

[6] Putzeys J *et al* 2019 Neuropixels data-acquisition system: a scalable platform for parallel recording of 10 000+ electrophysiological signals *IEEE Trans. Biomed. Circuits Syst.* **13** 1635–44
[7] Marblestone A H *et al* 2013 Physical principles for scalable neural recording *Front. Comput. Neurosci.* **7** 137
[8] Borton D A, Yin M, Aceros J and Nurmikko A 2013 An implantable wireless neural interface for recording cortical circuit dynamics in moving primates *J. Neural Eng.* **10** 026010
[9] Musk E and Neuralink 2019 An integrated brain-machine interface platform with thousands of channels *J. Med. Internet Res.* **21** e16194
[10] Alais D, Newell F and Mamassian P 2010 Multisensory processing in review: from physiology to behaviour *Seeing Perceiving* **23** 3–38
[11] Senkowski D, Schneider T R, Foxe J J and Engel A K 2008 Crossmodal binding through neural coherence: implications for multisensory processing *Trends Neurosci.* **31** 401–9
[12] Sheppard J P, Raposo D and Churchland A K 2013 Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans *J. Vis. Exp.* **13** 4
[13] MCGURK H and Macdonald J 1976 Hearing lips and seeing voices *Nature* **264** 746–8
[14] Chen G, Kang B, Lindsey J, Druckmann S and Li N 2021 Modularity and robustness of frontal cortical networks *Cell* **184** 3717–3730.e24
[15] Chen S *et al* 2024 Brain-wide neural activity underlying memory-guided movement *Cell* **187** 676–91.e16
[16] Gao Z, Davis C, Thomas A M, Economo M N, Abrego A M, Svoboda K, De Zeeuw C I and Li N 2018 A cortico-cerebellar loop for motor planning *Nature* **563** 113–6
[17] Ahmadi N *et al* 2019 Towards a distributed, chronically-implantable neural interface *2019 9th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* (IEEE) pp 719–24
[18] Moon E *et al* 2021 Bridging the 'last millimeter' gap of brain-machine interfaces via near-infrared wireless power transfer and data communications *ACS Photonics* **8** 1430–8
[19] Rozgic D *et al* 2018 A 0.338cm^3, artifact-free, 64-contact neuromodulation platform for simultaneous stimulation and sensing *IEEE Trans. Biomed. Circuits Syst.* p 1
[20] Lee J, Leung V, Lee A-H, Huang J, Asbeck P, Mercier P P, Shellhammer S, Larson L, Laiwalla F and Nurmikko A 2021 Neural recording and stimulation using wireless networks of microimplants *Nat. Electron.* **4** 604–14
[21] Sriram K, Karageorgos I, Wen X, Veselý J, Lindsay N, Wu M, Khazan L, Pothukuchi R P, Manohar R and Bhattacharjee A 2023 HALO: a hardware–software co-designed processor for brain–computer interfaces *IEEE Micro.* **43** 64–72
[22] Pothukuchi R P, Sriram K, Gerasimiuk M, Ugur M, Manohar R, Khandelwal A and Bhattacharjee A 2024 Distributed brain–computer interfacing with a networked multiaccelerator architecture *IEEE Micro.* **44** 106–15
[23] John R A, Tiwari N, Patdillah M I B, Kulkarni M R, Tiwari N, Basu J, Bose S K, Yu C J, Nirmal A and Vishwanath S K 2020 Self healable neuromorphic memtransistor elements for decentralized sensory signal processing in robotics *Nat. Commun.* **11** 4030
[24] Ritzau-Jost A, Delvendahl I, Rings A, Byczkowicz N, Harada H, Shigemoto R, Hirrlinger J, Eilers J and Hallermann S 2014 Ultrafast action potentials mediate kilohertz signaling at a central synapse *Neuron* **84** 152–63
[25] Principles of Neural Science, Fifth Edition, Book, Section vols 2014 (McGraw-Hill Education) (available at: neurology.mhmedical.com/content.aspx?aid=1101675601) (Accessed 4 February 2025)
[26] He Y *et al* 2022 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* **57** 3058–70
[27] Integrated M 2019 MAX9277/MAX9281: 3.12Gbps GMSL serializers for coax or STP output drive and LVDS Input (available at: https://datasheets.maximintegrated.com/en/ds/MAX9277-MAX9281.pdf) (Accessed 29 May 2024)
[28] Wang K Gigabit Multimedia Serial Link (gmsl) cameras as an alternative to gige vision cameras
[29] Sauerwald M 2014 FPD-Link III—doing more with less
[30] Bianchi J 2017 Next-generation FPD-Link III devices add speed and flexibility to advanced driver
[31] Berge H K O and Hafliger P 2007 High-Speed Serial AER on FPGA *2007 IEEE Int. Symp. on Circuits and Systems (ISCAS)* (IEEE) pp 857–60
[32] Safwat S, Hussein E E-D, Ghoneima M and Ismail Y 2011 A 12Gbps all digital low power SerDes transceiver for on-chip networking *2011 IEEE Int. Symp. of Circuits and Systems (ISCAS), Rio de Janeiro* (IEEE) pp 1419–22
[33] Lopez C M *et al* 2016 22.7 A 966-electrode neural probe with 384 configurable channels in 0.13μm SOI CMOS *2016 IEEE Int. Solid-State Circuits Conf. (ISSCC)* (IEEE) pp 392–3
[34] Wang S, Lopez C M, Garakoui S K, Chun H, Salinas D G, Sijbers W, Putzeys J, Martens E, Craninckx J and Van Helleputte N 2019 A compact quad-shank cmos neural probe with 5,120 addressable recording sites and 384 fully differential parallel channels *IEEE Trans. Biomed. Circuits Syst.* **13** 1625–34
[35] Ghoreishizadeh S S, Haci D, Liu Y, Donaldson N and Constandinou T G 2017 Four-wire interface ASIC for a multi-implant link *IEEE Trans. Circuits Syst. I Regul. Pap.* **64** 3056–67
[36] Ahmed S I and Kwasniewski T A 2005 Overview of oversampling clock and data recovery circuits *Canadian Conf. on Electrical and Computer Engineering* (IEEE) pp 1876–81
[37] Haci D, Mifsud A, Liu Y, Ghoreishizadeh S S and Constandinou T G 2019 In-body wireline interfacing platform for multi-module implantable microsystems *2019 IEEE Biomedical Circuits and Systems Conf. (Biocas)* (IEEE) pp 1–4
[38] Levenstein D and Okun M 2023 Logarithmically scaled, gamma distributed neuronal spiking *J. Physiol.* **601** 3055–69
[39] Weltin-Wu C and Tsividis Y 2013 An event-driven clockless level-crossing ADC with signal-dependent adaptive resolution *IEEE J. Solid State Circuits* **48** 2180–90
[40] Lichtsteiner P, Posch C and Delbruck T 2008 A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor *IEEE J. Solid State Circuits* **43** 566–76
[41] Guo M *et al* 2023 *A 3-Wafer-Stacked Hybrid 15MPixel CIS + 1 MPixel EVS with 4.6GEvent/s Readout, In-Pixel TDC and On-Chip ISP and ESP Function 2023 IEEE Int. Solid- State Circuits Conf. (ISSCC)* (IEEE) pp 90–92
[42] Cassidy A, Zhang Z and Andreou A G 2008 Neuromorphic interconnects using ultra wideband radio *2008 IEEE Biomedical Circuits and Systems Conf.* (IEEE) pp 297–300
[43] Mohan V, Tay W P and Basu A 2025 Towards neuromorphic compression based neural sensing for next-generation wireless implantable brain machine interface *Neuromorph. Comput. Eng.* **5** 014004
[44] Cartiglia M *et al* 2023 A 4096 channel event-based multielectrode array with asynchronous outputs compatible with neuromorphic processors *Nat. Commun.* **25** 7163

[45] Ros P M et al 2013 A wireless address-event representation system for ATC-based multi-channel force wireless transmission 5th IEEE Int. Workshop Advances in Sensors and Interfaces IWASI (IEEE) pp 51–56
[46] Culurciello E and Andreou A G 2003 A comparative study of access topologies for chip-level address-event communication channels IEEE Trans. Neural Netw. Learn. Syst. 14 1266–77
[47] Son B, Nam M-H, Hong C-O, Kim Y and Lee K-W 2017 4.1 A 640×480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation 2017 IEEE Int. Solid-State Circuits Conf. (ISSCC) (IEEE) pp 66–67
[48] Finateu T et al 2020 5.10 A 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86μm pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline 2020 IEEE Int. Solid- State Circuits Conf.—(ISSCC) (IEEE) pp 112–4
[49] He Y et al 2024 An event-based neural compressive telemetry with >11× loss-less data reduction for high-bandwidth intracortical brain computer interfaces IEEE Trans. Biomed. Circuits Syst. 18 1–10
[50] Liu Y, Pereira J L and Constandinou T G 2018 Event-driven processing for hardware-efficient neural spike sorting J. Neural Eng. 15 016016
[51] Luan S, Liu Y, Williams I and Constandinou T G 2016 An event-driven SoC for neural recording 2016 IEEE Biomedical Circuits and Systems Conf. (IEEE) pp 404–7
[52] Atmel 2015 Manchester coding basics (available at: https://ww1.microchip.com/downloads/en/AppNotes/Atmel-9164-Manchester-Coding-Basics_Application-Note.pdf)
[53] Perrott M H et al 2006 A 2.5-Gb/s multi-rate 0.25-μm cmos clock and data recovery circuit utilizing a hybrid analog/digital loop filter and all-digital referenceless frequency acquisition IEEE J. Solid State Circuits 41 2930–44
[54] MacEwen N C, Crockett L H, Pfann E and Stewart R W 2005 Symbol synchronisation implementation for low-power rf communication in wireless sensor networks Conf. Record of the 39th Asilomar Conf. onSignals, Systems and Computers (IEEE) pp 447–51
[55] Martens O, Liimets A and Kuusik A 2015 Synchronization algorithm for RFID-reader and its implementation 2015 Advances in Wireless and Optical Communications (IEEE) pp 82–85
[56] Song M et al 2020 30.8 A 3.5mm×3.8mm crystal-less mics transceiver featuring coverages of ±160ppm carrier frequency offset and 4.8-vswr antenna impedance for insertable smart pills 2020 IEEE Int. Solid- State Circuits Conf.—(ISSCC) (IEEE) pp 474–6
[57] Ren Z, Gao Z, Liaw H-P and Liu Y-H Ephys recording with Neuropixels v2 probe Zenodo (available at: https://zenodo.org/records/14449576) (Assessed 13 December 2024)
[58] Steinmetz N A, Carandini M and Harris K 2024 Recording with a Neuropixels probe (available at: http://doi.org/10.5522/04/25232962.v1)
[59] Raducanu B C et al 2017 Time multiplexed active neural probe with 1356 parallel recording sites Sensors 17 2388
[60] Yoon D-Y, Pinto S, Chung S, Merolla P, Koh T-W and Seo D 2021 A 1024-channel simultaneous recording neural soc with stimulation and real-time spike detection 2021 Symp. on VLSI Circuits (IEEE) pp 1–2
[61] Clinic C Stereoelectroencephalography (SEEG) (available at: https://my.clevelandclinic.org/health/diagnostics/17457-seeg-test#overview) (Accessed 02 April 2025)
[62] Children's Hospital Colorado 2025 Seeg Stereoelectroencephalography (available at: www.childrenscolorado.org/doctors-and-departments/departments/neuroscience-institute/programs/seeg-stereoelectroencephalography/)
[63] Biederman W, Yeager D J, Narevsky N, Leverett J, Neely R, Carmena J M, Alon E and Rabaey J M 2015 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 50 1038–47
[64] Jang M et al 2023 A 1024-Channel 268 nW/pixel 36x36 μm^2 /ch data-compressive neural recording ic for high-bandwidth brain-computer interfaces 2023 IEEE Symp. on VLSI Technology and Circuits (VLSI Technology and Circuits) (IEEE) pp 1–2