

An electronic neuromorphic system for real-time detection of High Frequency Oscillations (HFOs) in intracranial EEG

Mohammadali Sharifhazileh

University of Zurich and ETH Zurich

Karla Burelo

University of Zurich and ETH Zurich <https://orcid.org/0000-0003-4807-5223>

Johannes Sarnthein

UniversitätsSpital Zürich <https://orcid.org/0000-0001-9141-381X>

Giacomo Indiveri (✉ giacomo@ini.uzh.ch)

University of Zurich and ETH Zurich <https://orcid.org/0000-0002-7109-1689>

Article

Keywords: intracranial EEG(iEEG), High Frequency Oscillations (HFOs), epilepsy surgery, neural recording circuits

Posted Date: October 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-83699/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on May 25th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-23342-2>.

An electronic neuromorphic system for real-time detection of High Frequency Oscillations (HFOs) in intracranial EEG

Mohammadali Sharifhazileh^{1,2*}, Karla Burelo^{1,2*}, Johannes Sarnthein², and Giacomo Indiveri¹

¹Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, 8057, Switzerland

²Klinik für Neurochirurgie, UniversitätsSpital und Universität Zürich, 8091 Zurich, Switzerland

*These authors contributed equally.

ABSTRACT

The analysis of biomedical signals for clinical studies and therapeutic applications can benefit from compact and portable devices that can process these signals locally, in real-time, without the need for off-line processing. An example is the recording of intracranial EEG (iEEG) during epilepsy surgery with the detection of High Frequency Oscillations (HFOs, 80-500 Hz), which are a biomarker for the epileptogenic zone.

Conventional approaches of HFO detection involve the offline analysis of prerecorded data, often on bulky computers. However, clinical applications during surgery or in long-term intracranial recordings demand a self-sufficient embedded device that is battery-powered to avoid interfering with other electronic equipment in the operation room.

Mixed-signal and analog-digital neuromorphic circuits offer the possibility of building compact, embedded and low-power neural network processing systems that can analyze data on-line and produce results with short latency in real-time. These characteristics are well suited for clinical applications that involve the processing of biomedical signals at (or very close to) the sensor level.

In this work, we present a neuromorphic system that combines for the first time a neural recording headstage with a signal-to-spoke conversion circuit and a multi-core spiking neural network (SNN) architecture on the same die for recording, processing, and detecting clinically relevant HFOs in iEEG from epilepsy patients.

The device was fabricated using a standard $0.18 \mu\text{m}$ CMOS technology node and has a total area of 99 mm^2 . We demonstrate its application to HFO detection in the iEEG recorded from 9 patients with temporal lobe epilepsy who subsequently underwent epilepsy surgery. The total average power consumption of the chip during the detection task was $614.3 \mu\text{W}$. We show how the neuromorphic system can reliably detect HFOs: the system predicts postsurgical seizure outcome with state-of-the-art accuracy, specificity and sensitivity (78%, 100%, and 33% respectively).

This is the first feasibility study towards identifying relevant features in intracranial human data in real-time, on-chip, using event-based processors and spiking neural networks. By providing “neuromorphic intelligence” to neural recording circuits the approach proposed will pave the way for the development of systems that can detect HFO areas directly in the operation room and improve the seizure outcome of epilepsy surgery.

1 Introduction

The amount and type of sensory data that can be recorded is continuously increasing due to the recent progress in microelectronic technology¹. This data deluge calls for the development of low-power embedded edge computing technologies that can process the signals being measured locally, without requiring bulky computers or the need for internet connection and cloud servers. In particular, biomedical signal processing and clinical applications will greatly benefit from a direct and local processing of physiological measurements using compact and portable devices. Spiking neural networks (SNNs) have been proposed as a powerful computing paradigm for processing spatio-temporal signals and detecting complex patterns within them^{2,3}, so they are well suited for biomedical applications.

Neuromorphic engineering⁴ has produced mixed signal neuromorphic circuits that can directly emulate the physics of real neurons to implement faithful models of neural dynamics and SNNs^{5,6}. By integrating many silicon neurons onto microelectronic chips, it is possible to build compact and low-power SNN processing systems to efficiently process time-varying signals in real time. Examples of neuromorphic systems developed with this goal in mind have already been applied to processing Electrocardiogram (ECG) or Electromyography (EMG) signals^{7–10}.

However, these systems were sub-optimal, as they required external bio-signal recording, frontend devices, and data conversion interfaces. Bio-signal recording headstages typically comprise analog circuits to amplify and filter the signals being measured, and can be highly diverse in specifications depending on the application¹¹. For example, neural recording headstages for experimental neuroscience target high-density recordings^{12–15} and minimize the circuit area requirements, while devices used for clinical studies and therapeutic applications require a small number of recording channels and the highest possible signal-to-noise ratio (SNR)^{16–19}.

In this work, we present a neuromorphic system that combines for the first time a neural recording headstage with a signal-to-spoke conversion circuit and a multi-core SNN architecture on the same die for recording, processing, and detecting clinically relevant biomarkers in intracranial EEG recordings (iEEG) from epilepsy patients.

Epilepsy is the most common severe neurological disease. In about one-third of patients, seizures cannot be controlled by medication. Selected patients with focal epilepsy can achieve seizure freedom if the epileptogenic zone (EZ), which is the brain area generating the seizures, is correctly identified and surgically removed in its entirety. Presurgical and intraoperative measurement of iEEG signals is often needed to identify the EZ precisely^{20–23}. High Frequency Oscillations (HFOs) have been proposed as a new biomarker in iEEG to delineate the EZ^{17,18,24–28}. While HFOs have been historically divided into “ripples” (80–250 Hz) and “fast ripples” (FR, 250–500 Hz), detection of their co-occurrence was shown to enable the optimal prediction of postsurgical seizure freedom²⁴. In that study, HFOs were detected automatically by a software algorithm that matched the morphology of the HFO to a predefined template (Morphology Detector)^{24,29}. An example of such an HFO is shown in Fig. 1a. While software algorithms are used for detecting HFOs offline, compact embedded neuromorphic systems that can record iEEGs and detect HFOs online, in real time, would be able to provide valuable information during surgery, and simplify the collection of statistics in long-term epilepsy monitoring^{30,31}. Here we show how our neuromorphic system paves the way toward the development of a compact embedded battery-operated system that can be used for the real-time online detection of HFOs.

In this paper we first describe the design principles of the HFO detection architecture and its neuromorphic circuit implementation. We discuss the characteristics of the circuit blocks proposed and present experimental results measured from the fabricated device. We then show how the neuromorphic system performs in HFO detection compared to the Morphology Detector²⁴ on iEEG recorded from the medial temporal lobe³².

2 Results

Figure 1 shows how prerecorded iEEG³² was processed by the frontend headstage and the SNN multi-core neuromorphic architecture. Signals were band-passed filtered into Ripple and Fast Ripple bands (Fig. 1a,b,f). The resulting waveforms were converted into spikes using asynchronous delta modulator (ADM) circuits^{33,34} (Fig. 1c,f) and fed into the SNN architecture (Fig. 1d,g). Neuronal spiking signalled the detection of an HFO (Fig. 1e bottom). All stages were first simulated in software to find optimal parameters and then validated with the hardware components. The HFO detection was validated by comparing the HFO rate across recording intervals (Fig. 1i) and against postsurgical seizure outcome²⁴.

¹A short video describing the rationale underlying the study is available on the Annual BCI award 2020 webpage at: https://youtu.be/Pw83Mrza_rg

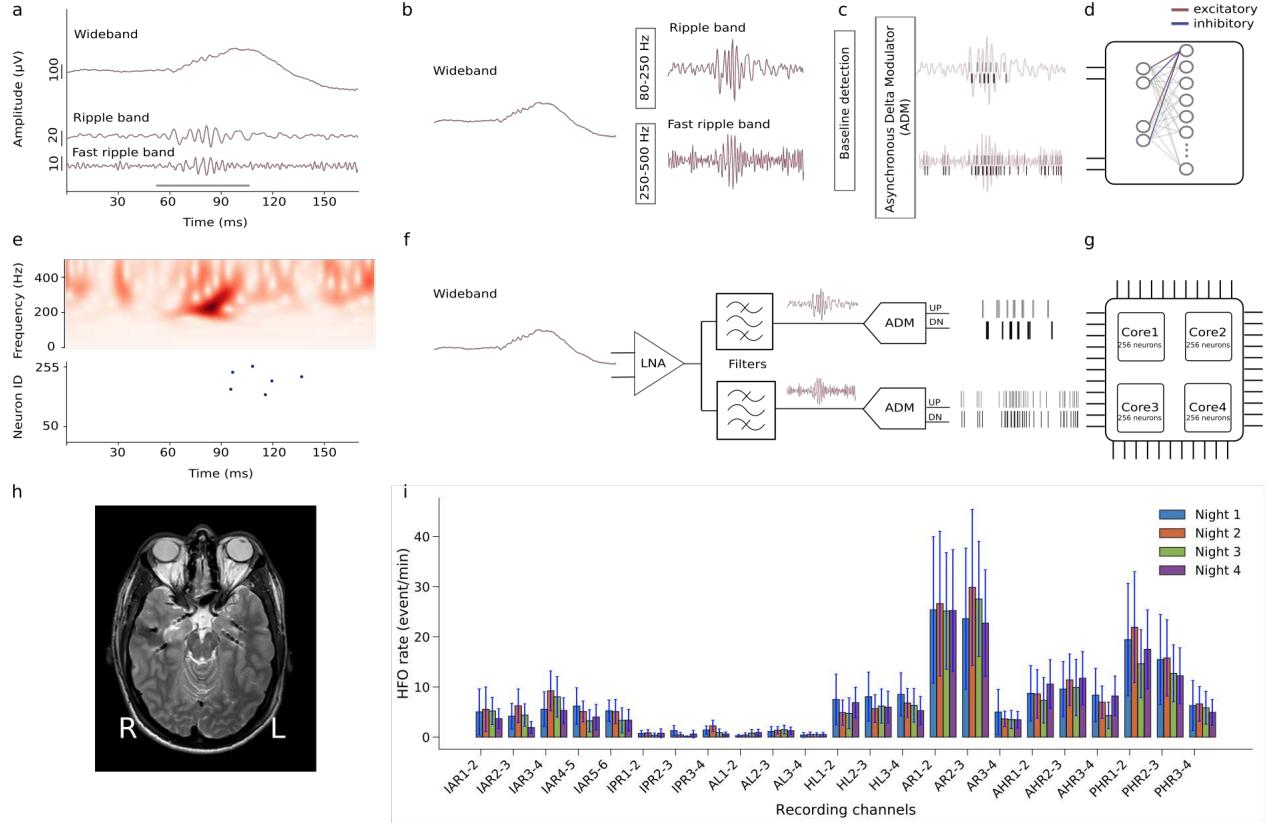


Figure 1. Automatic HFO detection using a bio-inspired SNN. (a) The pre-recorded iEEG signal in wideband, Ripple band (80-250Hz) and Fast Ripple band (250-500Hz). HFOs stand out of the baseline in the signal. The period marked by the gray bar represents a clinically relevant HFO^{24,32}. (b,c,d) Software simulated spiking neural network (SNN). For preprocessing the wideband EEG is filtered in Ripple band and Fast Ripple band. A baseline detection stage finds the optimum threshold that is applied in an Asynchronous Delta Modulator (ADM) which converts the signal to spikes. Signal traces are encoded by UP spikes (gray bars) and DOWN spikes (black bars), which are then fed as input into the SNN. The SNN is implemented as a two-layer spiking network of integrate and fire neurons with dynamic synapses. Each neuron in the second layer receives four inputs: two excitatory spike trains from UP channels and two inhibitory ones from DOWN channels. The parameters of the network were chosen to exhibit the relevant temporal dynamics and tune the neurons to produce output spikes in response to input spike train patterns that encode clinically relevant HFOs. (e, top) Time-frequency spectrum of the Fast Ripple iEEG of panel a. (e, bottom) Firing of SNN neurons indicate the occurrence of the HFO. (f) Block diagram of the neuromorphic system input headstage. The headstage comprises a low noise amplifier (LNA), two configurable bandpass filters and two ADM circuits that convert the analog waveforms into spike trains. (g) The spikes produced by the ADMs are sent to a multi-core array of silicon neurons that are configured to implement the desired SNN. (h) MRI with 7 implanted depth electrodes that sample the mesial temporal structures of a patient with drug resistant temporal lobe epilepsy (Patient 1). (i) Rates of HFOs detected by the neuromorphic SNN for recordings made across four nights for Patient 1. HFO rate and variability across intervals within a night is indicated by standard error bars. Recording channels AR1-2 and AR2-3 in right amygdala showed the highest HFO rates which were stable over nights. Thus, the neuromorphic system would predict that a therapeutic resection, which should include the right amygdala, would achieve seizure freedom for >1 year. Indeed, a resection including the right amygdala achieved seizure freedom for >1 year.

2.1 The neuromorphic system

An overview of the hardware neuromorphic system components is depicted in Fig. 2. The chip (Fig 2c) was fabricated using a standard 180 nm Complementary Metal-Oxide-Semiconductor (CMOS) process. It comprises 8 input channels (headstages) responsible for the neural recording operation, band-pass filtering and conversion to spikes³⁵, and a multi-core neuromorphic processor with 4 neurosynaptic cores of 256 neurons each, which is a novel type of Dynamic Neuromorphic Asynchronous Processor (DYNAP) based on the DYNAP-SE device³⁶. The total chip area is 99 mm². The 8 headstages occupy 1.42 mm² with a single headstage occupying an area of 0.15 mm² (see Fig. 2a). The area of the four SNN cores is 77.2 mm² with a single SNN core occupying 15 mm². For the HFO detection task, the total average power consumption of the chip was 614.3 μ W. The total static power consumption of

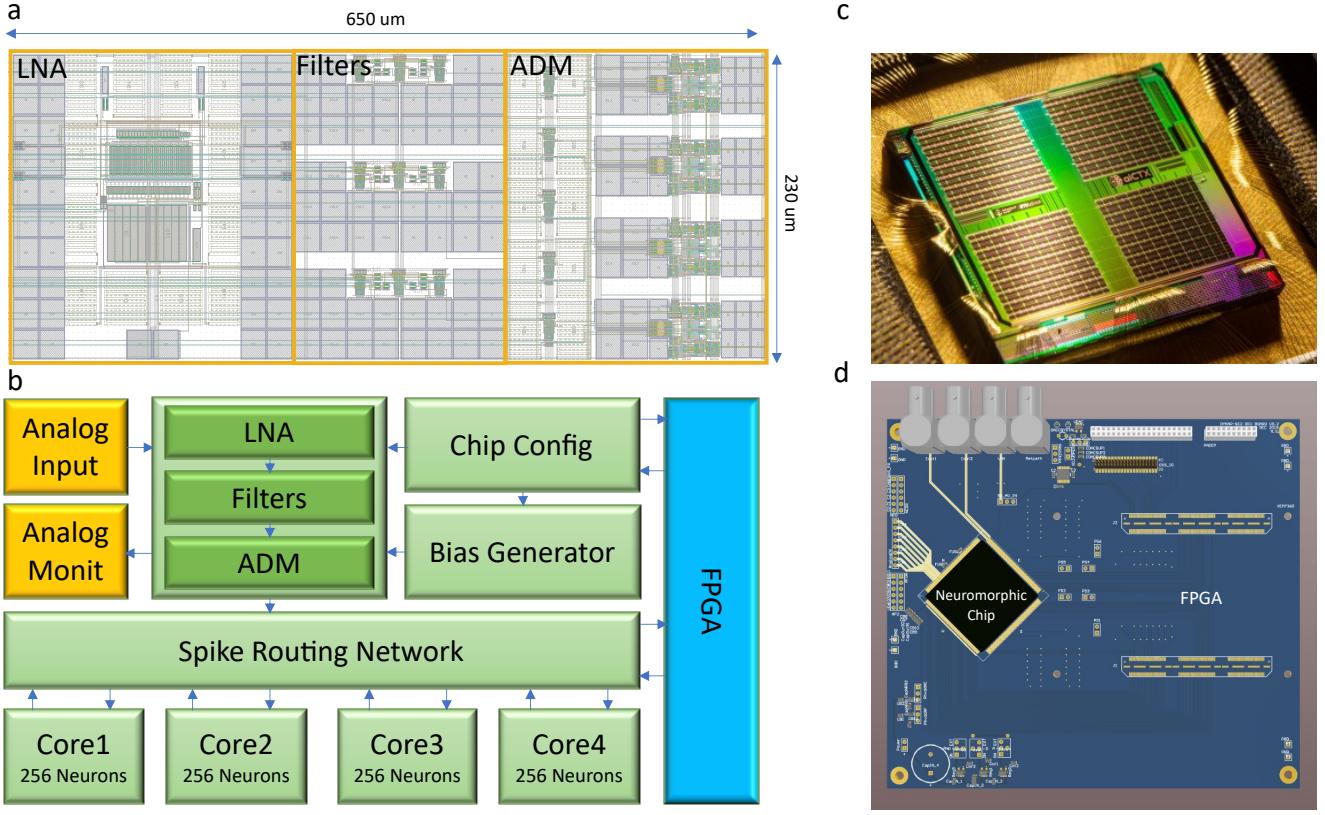


Figure 2. Neuromorphic electronic system overview (a) Physical layout of a single channel of the analog headstage array, including the LNA, three low-pass/band-pass filters, and four ADM signal to spike encoders. (b) Reduced block diagram of the neuromorphic platform. Analog signals from electrodes are fed into the input headstage that converts them into spike trains and sends them to the SNN implemented on the multi-neuron cores, via a spike routing network. The spike routing network routes the spikes within on-chip SNN and to an external FPGA device used for data logging and prototyping. The FPGA is also used for setting circuit parameters (c) Chip photograph showing the 11 mm x 9 mm silicon die. (d) Prototyping Printed Circuit Board (PCB) used to host the chip and the infrastructure to implement the test setup. The setup is composed of a prototyping FPGA board mounted on the same PCB that hosts the chip, and of probe points to evaluate the characteristics of both input headstage and SNN multi-core network.

a single headstage was $7.3 \mu\text{W}$. The conversion of filtered waveforms to spikes by the ADMs consumed on average $109.17 \mu\text{W}$. The power required by the SNN synaptic circuits to process the spike rates produced by the ADMs was $497.82 \mu\text{W}$, while the power required by the neurons in the second layer of the SNN to produce the output spike rates was 0.2nW . The block diagram of the hardware system functional modules is shown in Fig. 2b. The Field Programmable Gate Array (FPGA) block the right of the figure represents a prototyping device that is used only for characterizing the system performance. Figure 2c shows the chip photograph, and Fig. 2d represents a rendering of the prototyping Printed Circuit Board (PCB) used to host the chip.

Figure 3 shows the details of the main circuits used in a single channel of the input headstage. In particular, the schematic diagram of the Low-Noise Amplifier (LNA) is shown in Fig. 3a. It consists of an Operational Transconductance Amplifier (OTA) with variable input Metal Insulator Metal (MIM) capacitors, C_{in} that can be set to $2/8/14/20 \text{ pF}$ and a Resistor-Capacitor (RC) feedback in which the resistive elements are implemented using MOS-bipolar structures³⁷. The MOS-bipolar pseudoresistors $MPR1$ and $MPR2$, and the capacitors $C_f=200 \text{ fF}$ of Fig. 3a were chosen to implement a high-pass filter with a low cutoff frequency of 0.9 Hz . Similarly, the input capacitors C_{in} , C_f , and the transistors of the OTA were sized to produce maximum amplifier gain of approximately 40.2 dB but can be adjusted to smaller values by changing the capacitance of C_{in} .

Figure 3b shows the schematic of the OTA, which is a modified version of a standard folded-cascode topology³⁸. The currents of the transistors in the folded branch (M5-M10) are scaled to approximately 1/6th of the currents in the original branch M1-M4. The noise generated by M5-M10 is negligible compare to that of M1-M4 due to the

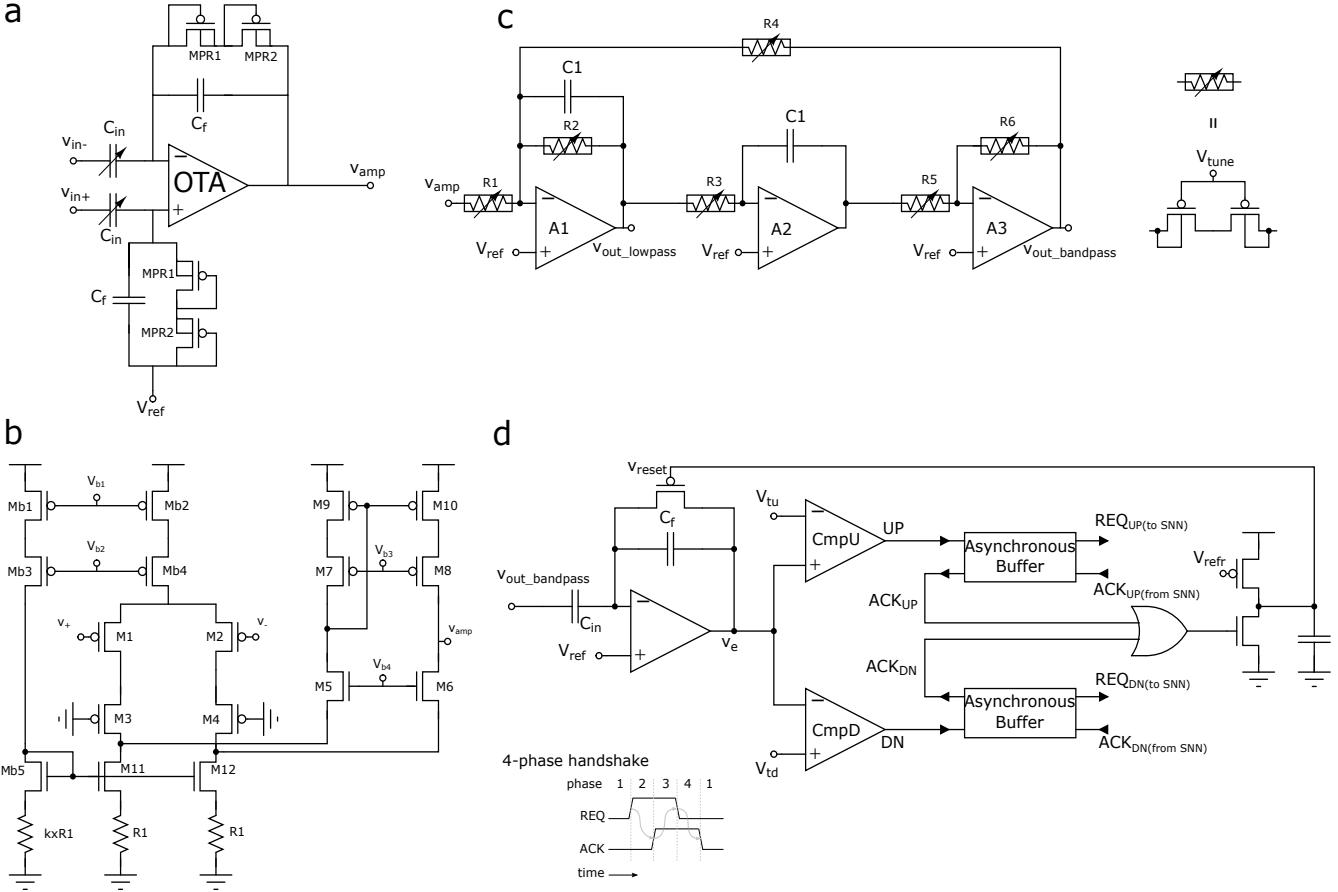


Figure 3. Schematic diagrams of the input headstage circuits. (a) Variable-gain LNA using variable input capacitor array and pseudo-resistors. The gain of this stage is calculated by C_{in}/C_f ; the use of the pseudo-resistors allows to reach small low cut-off frequencies. (b) Folded cascode OTA using resistive degeneration to reduce the noise influence of nMOS devices. Note that the current flowing through the bias branch is k times smaller than the tail branch of the amplifier. (c) band-pass (Tow-Thomas) filters for performing second-order filtering in both ripple and fast-ripple bands as well as low-frequency component of the iEEG. Tunable pseudo resistors are used to adjust the filter gain, center-frequency and band-width. The same basic structure can be used to provide both low-pass and band-pass outputs, thus is desirable in terms of design flexibility. (d) Asynchronous Delta Modulator (ADM) circuit to convert the analog filter outputs into spike trains. The ADM input amplifier has a gain of C_{in}/C_f in normal operation when V_{reset} is low and the feedback PMOS switch is off. As the amplified signal crosses one of the two thresholds, V_{tu} or V_{td} , an UP or DN spike is produced by asserting the corresponding REQ signal. A 4-phase handshaking mechanism produces the corresponding ACK signal in response to the spike. Upon receiving the ACK signal, the ADM resets the amplifier input and goes back to normal operation after a refractory period determined by the value of V_{refr} . The asynchronous buffers act as 4-phase handshaking interfaces that propagate the UP/DN signals to the on-chip Address-Event Representation (AER) spike routing network of Fig. 2.

low current in these transistors. As a result, the total current and the total input-referred noise of the OTA was minimized.

To ensure accurate bias-current scaling, the currents of Mb2 and Mb4 in Fig. 3b were set using the bias circuit formed by Mb1, Mb3, and Mb5. The voltages V_{b1} and V_{b2} in the biasing circuit can be set by a programmable 6526-level integrated bias-generator, integrated on chip³⁹. The current sources formed by Mb1 and Mb2 were cascaded to increase their output impedance and to ensure accurate current scaling. These devices operate in strong inversion to reduce the effect of threshold voltage variations. The source-degenerated current mirrors formed by M11, M12, Mb5 and resistors R1 and $k \times R1$ assure that the currents in M5 and M6 are a small fraction of the currents in M3 and M4. The R1 gain coefficient k was chosen at design time to be $k = 8.5$. Thanks to the use of this source-degenerated current source scheme, the $1/f$ noise in the OTA is limited mainly to the effect of the input differential pair. Therefore, the transistors of the input-differential pair were chosen to be pMOS devices and to have a large area.

The active filters implemented in our system are depicted in Fig. 3c. They comprise three operational amplifiers, configured to form a Tow-Thomas resonating architecture⁴⁰. This architecture consists of a damped inverting integrator that is stabilized by R_2 and cascaded with another undamped integrator, and an inverting amplifier for adjusting the loop-gain by a factor set by the ratio R_6/R_5 . The center frequency f_0 of the bandpass filter can be calculated as $f_0 = 1/2\pi\sqrt{R_3R_4C_1^2}$. By choosing $R_3 = R_4 = R$, we can then simplify it to $f_0 = 1/2\pi RC_1$. Similarly, the gain of the filter is $|T_{BP}| = R_4/R_1$ and its bandwidth $BW = 2\pi f_0 \sqrt{R_3R_4}/R_2$, but with our choice of resistors we can show that $|T_{BP}| = R/R_1$ and $BW = 1/(R_2C_1)$. Therefore, this analysis shows that R is responsible for setting f_0 , R_1 for adjusting the gain, and R_2 for tuning the bandwidth. Moreover, due to the resistive range of the tunable double-PMOS pseudo resistors used in this design⁴¹, f_0 was set in the sub-hundred Hertz region by choosing $C_1 = 10\text{ pF}$.

Figure 3d shows the schematic diagram of the ADM circuit³⁴. There are four of such circuits per headstage channel. One for converting the wideband signal V_{amp} into spike trains; one for converting the output of the low-pass filter $V_{out_lowpass}$; and two for converting the output of the two band-pass filters $V_{out_bandpass}$. The amplifier at the input of the ADM circuit in Fig. 3d implements an adaptive feedback amplification stage with a gain set by C_{in}/C_f that in our design is equal to 8 when V_{reset} is high, and approximately zero during periods in which V_{reset} is low. In these periods, defined as “reset assertion” the output of the amplifier V_e is clamped to V_{ref} , while in periods when V_{reset} is high, the output voltage V_e represents the amplified version of the input. The V_e signal is then sent as input to a pair of comparators that produce either “UP” or “DN” digital pulses depending if V_e is greater than V_{tu} or lower than V_{td} . These parameters set the ADM circuit sensitivity to the amplitude of the Delta-change. The smallest values that these voltages can take is limited by the input offset of the ADM comparators (see $CmpU, CmpD$ in Fig. 3d), which is approximately $500\text{ }\mu\text{V}$.

Functionally, the ADM represents a Delta-modulator that quantizes the difference between the current amplitude of V_e and the amplitude of V_e at the previous reset assertion. The precise timing of the UP/DN spikes produced in this way are deemed to contain all the information about the original input signal, given that the parameters of the ADM are known⁴². The UP and DN spikes are used as the request signals of the asynchronous AER communication protocol^{43–45} used by the spike routing network for transmitting the spikes to the silicon neurons of the neuromorphic cores (Fig. 2). We call this event-based computation. These signals are pipelined through asynchronous buffers that locally generates $ACK_{UP/DN}$ to reset the ADM with every occurrence of an UP or DN event. The output of the asynchronous buffer $REQ_{UP/DN(toSNN)}$ conveys these events to the next asynchronous stages. The bias voltage V_{refr} controls the refractory period that keeps the amplifier reset and limits the maximum event rate of the circuit to reduce power consumption. The bias voltages V_{tu} and V_{td} control the sensitivity of the ADM and the number of spikes produced per second, with smaller values producing spike trains with higher frequencies. Small V_{tu} and V_{td} settings lead to higher power consumption and allow the faithful reconstruction of the input signal with all its frequency components. The ADM hyper parameters V_{refr}, V_{tu} , and V_{td} can therefore be optimized to achieve high reconstruction accuracy of the input signal and suppress background noise (e.g., due to high-frequency signal components), depending on the nature of the signal being processed (see Methods). All of the 32 ADM output channels are then connected to a common AER encoder which includes an AER arbiter⁴⁶ used to manage the asynchronous traffic of events and convey them to the on-chip spike routing network.

2.2 Analog headstage circuit measurements

Figure 4 shows experimental results measured from the different circuits present in the input headstage. Fig. 4a shows the transient response of the LNA to a prerecorded iEEG signal used as input. The signal was provided to the headstage directly via an arbitrary waveform generator programmed with unity gain and loaded with a sequence of the prerecorded iEEG data with amplitude in the mV range. We also tested the LNA with an input sine wave of 100 Hz with a 1 mV peak-to-peak swing, revealing <1% of total harmonic distortion at maximum gain, and an output swing ranging between 0.7 V and 1.4 V. To characterize the input referred noise of the LNA, we shorted input terminals of the LNA, V_{in+}, V_{in-} , captured LNA output, V_{amp} , on a dynamic signal analyzer and divided it by the gain of the LNA, set to 100, and plotted output power spectral density (see Fig. 4b). The LNA generates a $\leq 100\text{ nV}/\sqrt{\text{Hz}}$ noise throughout the spectrum. As the $1/f$ noise dissipates with the increase in frequency, the LNA noise only scales down to the thermal component. Thus, the noise for the Ripple band is $< 10\text{ nV}/\sqrt{\text{Hz}}$ and $< 5\text{ nV}/\sqrt{\text{Hz}}$ for the Fast Ripple band. Figure 4b shows how the noise generated by the LNA is well below the pre-recorded iEEG power, throughout the full frequency spectrum. The LNA features a programmable gain that can be set to 20 dB, 32 dB, 36 dB, or 40 dB. It has a > 40 dB common mode rejection ratio; it consumes 3 μW of power per channel; and it has a total bandwidth, defined as $Gm/(A_M \cdot C_L)$, approximately equal to 11.1 KHz, when the capacitive load is set to $C_L=20\text{ fF}$, the OTA transconductance to $Gm=20\text{ nS}$ and the amplifier gain to 40 dB (see Fig. 4c).

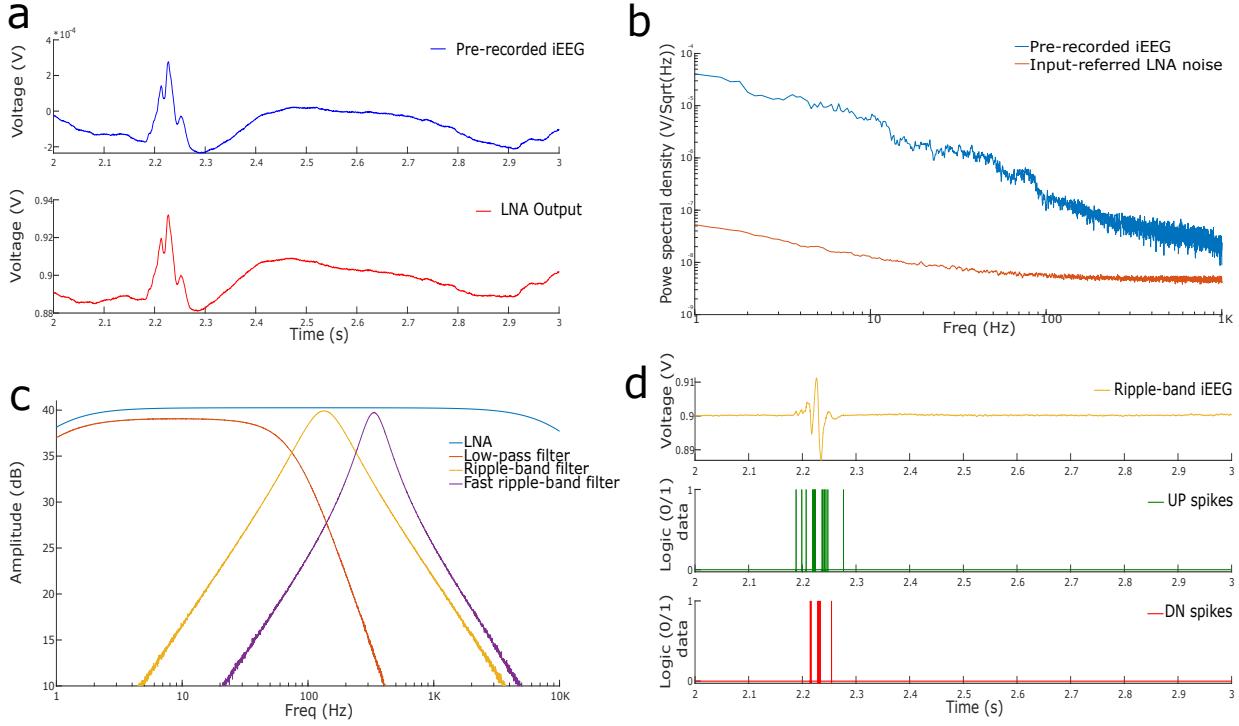


Figure 4. Measurements from the analog headstage. (a) An iEEG sample³² and the LNA amplified output. (b) Noise floor of the headstage LNA and iEEG power spectral density. In the HFO range (80–250 Hz) the noise level of the LNA is below the iEEG noise floor by an order of magnitude. (c) Frequency response of the implemented filters in the headstage. The band-pass filters are tuned to highlight HFO frequency bands. (d) ADM output spikes in response to the ripple-filtered signal. The top plot shows the analog filter output, the middle plot shows the UP spikes generated by the ADM and the bottom plot shows the DN spikes.

The on-chip bias generator can be used to set the filter frequency bands. For HFO detection, we biased the tunable pseudoresistors of the filters to achieve a cutoff frequency of 80 Hz for the low-pass filter, a range between 80 Hz and 250 Hz for the first bandpass filter, appropriate for Ripple detection, and between 250 Hz and 500 Hz for the second bandpass filter, to detect Fast Ripples (see Fig. 4c). As we set the tail current of each single-stage OpAmp of Fig. 3c to 150 nA, each filter consumed 0.9 μ W of power.

Figure 4d plots the spikes produced by the ADM circuit in response to Ripple-band data obtained from the pre-recorded iEEG measurements. We set the ADM refractory period to 300 μ s, making it the longest delay, compared to those introduced by the comparator and handshaking circuits, that are typically <100 μ s. The spike-rate of the ADM can range from few hundreds of Hz to hundreds of kHz depending on the values of V_{tu} , V_{td} , and V_{refr} . Each ADM consumed 1.5 nJ of energy per spike, and had a static power dissipation of 96 nW.

2.3 System performance

The system-level performance is assessed by measuring the ability of the proposed device to correctly measure the iEEG signals, to properly encode them with spike trains, and to detect clinically relevant HFOs²⁴ via the SNN architecture.

The SNN architecture is a two-layer feed-forward network of integrate and fire neurons with dynamic synapses, i.e., synapses that produce post-synaptic currents that decay over time with first-order dynamics. The first layer of the network comprises four input neurons: the first neuron conveys the UP spikes that encode the iEEG signal filtered in Ripple band; the second neuron conveys the DN spikes derived from the same signal; the third neuron conveys the UP spikes derived from the Fast-ripple band signal, while the forth neuron conveys the DN spikes. The second layer of the network contains 256 neurons which receive spikes from all neurons of the input layer (see Fig. 5b). The current produced by the dynamic synapses of the second layer neurons decay exponentially over time at a rate set by a synapse-specific time-constant parameter. The amplitude of these currents in response to the input spikes depends on a separate weight parameter, and their polarity (positive or negative) depends on the synapse type (excitatory or inhibitory). In the network proposed, UP spikes are always sent to excitatory synapses and DN

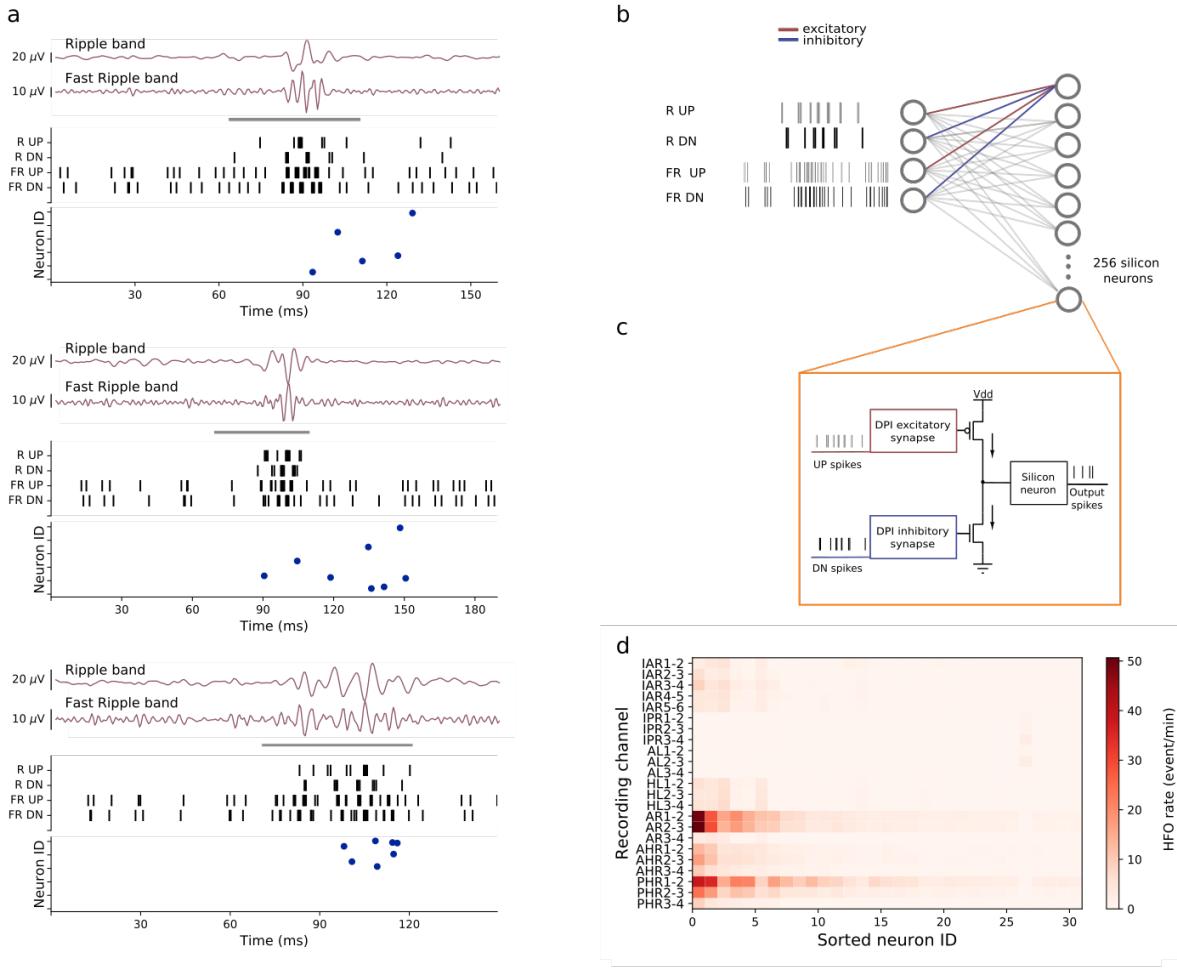


Figure 5. (a) Examples of HFOs that the hardware SNN detected in the iEEG of Patient 1. The periods marked by the gray bar represent clinically relevant HFO^{24,32}. The signals in Ripple and Fast-ripple band were transformed to UP and DN spikes. These spike trains were sent to the neurons in the hardware through the R_{UP} , R_{DN} , FR_{UP} and FR_{DN} channels. The bottom panel of each example shows the raster plot of the silicon neurons. Each neuron responds to different HFOs depending on the characteristics of the pattern. (b) The SNN architecture consist of a two-layer network of 256 integrate and fire neurons with dynamic synapses. Each neuron in the second layer receives excitatory connections from the R_{UP} and FR_{UP} channels, and inhibitory connections from the R_{DN} and FR_{DN} channels. The synaptic parameters time constants and weights are distributed randomly within a predetermined optimal range. (c) Hardware building blocks used for the implementations of the SNN: the DPI synapse is a “Differential-Pair Integrator” circuit⁴⁷, and the silicon neuron is an Adaptive Exponential Integrate and Fire (AdExp IF) circuit⁴⁸. (d) HFO rates computed for Patient 1. The neurons are sorted according to their average firing rate. Only a small number of neurons fire across all the recordings, even for channels with high HFO rates (e.g. AR1-2).

spikes to inhibitory ones. All neurons in the second layer have the same connectivity pattern as depicted in Fig. 5b with homogeneous weight values. An important aspect of the SNN network lies in the way it was configured to recognize the desired input spatio-temporal patterns: rather than following the classical Artificial Neural Network (ANN) approach of training the network by modifying the synaptic weights with a learning algorithm and using identical settings for all other parameters of synapses and neurons, we fixed the weights to constant values and chose appropriate sets of parameter distributions for the synaptic time constants and neuron leak variables. Because of the different time-constants for synapses and neurons, the neurons of the second layer produce different outputs, even though they all receive the same input spike trains.

The set of parameter distributions that maximized the network’s HFO detection abilities was found heuristically

by analyzing the temporal characteristics of the input spike trains and choosing the relevant range of excitatory and inhibitory synapse time constants that produced spikes in the second layer only for the input signals that contained HFOs as marked by the Morphology Detector²⁴ (Figure 1a, Figure 5a). This procedure was first done using software simulations with random number generators and then validated in the neuromorphic analog circuits, exploiting their device mismatch properties. The software simulations were carried out using a behavioral-level simulation toolbox⁴⁹ based on the neuromorphic circuit equations, that accounts for the properties of the mixed-signal circuits in the hardware SNN. The hardware validation of the network was done using a single core of the DYNAP-SE neuromorphic processor³⁶, which is a previous generation chip functionally equivalent to the one proposed, implemented using the same CMOS 180 nm technology node. The 256 neurons in this core received spikes produced by the ADM circuits of the analog headstage, as described in Fig. 5c. The ADM UP spikes were sent to the excitatory synapses, implemented using a DPI circuit^{6,49} that produce positive currents, and DN spikes were sent to complementary versions of the circuit that produce inhibitory synaptic currents. Both excitatory and inhibitory currents were summed into the input nodes of their afferent leaky integrate-and-fire silicon neuron circuit⁶, which produced output spikes only if both the frequency and the timing of the input spikes was appropriate (see Fig. 5c). The bias values of the excitatory and inhibitory DPI circuits and of the neuron leak circuits were set in a way to match the mean values of the software simulation parameters. All neuron and synapse circuits in the same core of the chip share global bias parameters, so nominally all excitatory synapses would have the same time-constant, all inhibitory ones would share a common inhibitory time-constant value and all neurons would share the same leak parameter value. However, as the mixed-signal analog-digital circuits that implement them are affected by device mismatch, they exhibit naturally a diversity of behaviors that implements the desired variability of responses. Therefore, in the hardware implementation of the SNN, the distribution of parameters that produce the desired different behaviors in the second layer neurons emerges naturally, by harnessing the device mismatch effects of the analog circuits used and without having to use dedicated random number generators⁵⁰.

Figure 1e (bottom panel) shows an example of the activity of the hardware SNN in response to an HFO that was labeled as clinically relevant by the Morphology Detector²⁴. The iEEG traces in the Ripple band and Fast ripple band (Fig. 1a) and the time frequency spectrum (top panel of Fig. 1e) show the HFO shortly before the SNN neurons spike in response to it (bottom panel of Fig. 1e). The delay between the beginning of the HFO and the spiking response of the silicon neurons is due to the integration time of both excitatory and inhibitory synapse circuits, which need to accumulate enough evidence for producing enough positive current to trigger the neuron to spike.

To improve classification accuracy and robustness, we adopted an *ensemble* technique⁵¹ by considering the assembly response of the 256 neurons in the network: the system is said to detect an HFO if at least one neuron in the second layer spikes in a 15 ms interval. We counted the number of HFOs detected per electrode channel and computed the corresponding HFO rate (Section 5). Examples of HFOs recorded from Patient 1 and detected by the hardware SNN are shown in Fig. 5a; several neurons respond within a few milliseconds after initiation of the HFO. Different HFOs produce different UP and DN spike trains, which in turn lead to different sets of second layer neurons spiking. Figure 5d shows the HFO rates calculated for each electrode from the recordings of this patient. Observe that not all the neurons in the second layer respond to the HFOs. Even for electrode channels with high HFO rates, a very small number of neurons fire at high rates.

The robustness of the HFO rate measured with our system can be observed in Fig 1i, where the relative differences of HFO rates across channels in Patient 1 persisted over multiple nights. To quantify this result we performed a test-retest reliability analysis by computing the scalar product of the HFO rates across all recording intervals (0.95 in Patient 1), where the scalar product is 1 for highly overlapping spatial distributions, indicating that the HFO distribution persists over intervals.

2.4 Predicting seizure outcome

In Patient 1, the electrodes were implanted in right frontal cortex (IAR, IPR), the left medial temporal lobe (AL, HL) and the right medial temporal lobe (AR, AHR, PHR). The recording channels AR1-2 and AR2-3 in the right amygdala produced the highest HFO rates persistently. We included all channels with persistently high HFO rate in the 95% percentile to define the “HFO area”. If the HFO area is included in the resection area, resection area (RA), we would retrospectively ‘predict’ for the patient to achieve seizure freedom. Indeed, right selective amygdalohippocampectomy in this patient achieved seizure freedom for >1 year.

We validated the system performance across the whole patient group by performing the test-retest reliability analysis of all the data. The test-retest reliability score ranges from 0.59 to 0.97 with a median value of 0.91. We compared the HFO area detected by our system with the RA. For each individual, we then retrospectively determined whether resection of the HFO area would have correctly ‘predicted’ the postsurgical seizure outcome

Table 1. Patient characteristics and postsurgical seizure outcome. We ‘predict’ seizure outcome for each patient based on resection of the HFO area that was delineated by the Morphology Detector²⁴ and the hardware SNN of our system. HS hippocampal sclerosis; ILAE seizure outcome classification of the International League Against Epilepsy.

Patient	Histology/ Pathology	Intervals of 5 min	Test-retest intervals	Outcome (ILAE)	Follow-up (months)	Morphology Detector prediction	Hardware SNN prediction
1	HS	28	0.95	1	12	TN	TN
2	Glioma	13	0.97	1	29	TN	TN
3	HS	39	0.83	1	13	TN	TN
4	HS	34	0.96	1	41	TN	TN
5	HS	35	0.91	1	14	TN	TN
6	HS	35	0.59	1	11	TN	TN
7	HS	1	–	3	42	FN	FN
8	HS	16	0.74	3	15	FN	TP
9	HS	12	0.90	5	46	FN	FN

Table 2. Comparison of postsurgical outcome prediction between the Morphology Detector and our system. TP True Positive; TN True Negative; FP False Positive; FN False Negative; N = TP + TN + FP + FN = number of patients. The Morphology Detector did not classify a TP so that sensitivity and PPV can not be calculated.

	Morphology Detector prediction [%]	Hardware SNN prediction [%]
Specificity = $TN / (TN + FP)$	100	100
Sensitivity = $TP / (TP + FN)$	–	33
Negative Predictive Value = $TN / (TN + FN)$	67	75
Positive Predictive Value = $TP / (TP + FP)$	–	100
Accuracy = $(TP + TN) / N [\%]$	67	78

(Table 1). Seizure freedom (ILAE 1) was achieved in 6 of the 9 patients. To estimate the quality of our ‘prediction’, we classified each patient as follows: we defined as “True Negative” (TN) a patient where the HFO area was fully located inside the RA and who became seizure free; “True Positive” (TP) a patient where the HFO area was not fully located within the RA and the patient suffered from recurrent seizures; “False Negative” (FN) a patient where the HFO area was fully located within the RA but who suffered from recurrent seizures; “False Positive” (FP) a patient where the HFO area was not fully located inside the RA but who nevertheless achieved seizure freedom.

The HFO area defined by our system was fully included in the RA in patients 1 to 6. These patients achieved seizure freedom and were therefore classified as TN. In Patients 7 and 9, the HFO area was also included in the RA but these were classified as FN since these patients did not achieve seizure freedom. In Patient 8, the HFO area was not included in the RA and the patient did not achieve seizure freedom (TP).

We finally compared the predictive power of our detector to that of the Morphology Detector²⁴ for the individual patients (Table 1) and over the group of patients (Table 2). The overall prediction accuracy of our system across the 9 patients is comparable to that obtained by the Morphology Detector. The 100% specificity achieved by both detectors indicates that HFO analysis provides results consistent with the current surgical planning.

3 Discussion

The device presented here demonstrates the potential of neuromorphic computing for extreme-edge use cases that cannot rely on computation performed on remote computers (e.g., in the “cloud”), and that require compact and very low power embedded systems (e.g., for battery-operated hand-held operations). By integrating on the same chip both the signal acquisition headstage and the neuromorphic multi-core processor, we developed an integrated system that can demonstrate the advantages of neuromorphic computing in clinically relevant applications.

Although numerous neural recording headstages have been already developed and optimized for their specific application domain (e.g., for very large scale arrays^{12–15}, or for intracranial recordings^{38,52–57}), to our knowledge,

this is the first instance of a headstage design that has the capability of adapting to numerous use cases requiring different gain factors and band selections, on the same input channel.

The approach followed to process the signals in the SNN and determine the right architecture for detecting HFOs is radically different from the deep-learning one: rather than using arrays of neurons with homogeneous parameters and no temporal dynamics, and relying on the changes in the synaptic weights, we emulated the dynamic properties of real neurons with mixed-signal analog/digital circuits and exploited their variability to find the right set of randomly distributed parameters for tuning the network to detect HFOs.

The simulations of the SNN not only allowed us to define the optimal architecture for HFO detection, but also gave us solutions for setting the hyperparameters of the analog headstage, such as the refractory period V_{ref} and the threshold (V_{tu} and V_{td}) for the signal-to-spike conversion of the ADM. While mismatch effect is generally a concern in modeling hardware designs based on software simulations, we show here that the mismatch among the silicon neurons resulted in a key feature for the implementation of our SNN architecture. This advantage allowed us to generate the normal distribution of parameters without manually defining the distribution of neuronal time-constants found in simulations or requiring extra memory to allocate these values.

By averaging over both time and the number of neurons recruited by the *ensemble* technique, the SNN network was able to achieve robust results: the accuracy obtained by the SNN are compatible with those obtained by state-of-the-art software algorithms implemented using complex algorithms on powerful computers²⁴. Overall, the high specificity (100%) achieved with our system not only generalizes the value of the detected HFO by the SNN across different types of patients, but still holds true at the level of the individual patients, which is a prerequisite to guide epilepsy surgery that aims for seizure freedom.

4 Conclusions

This is the first feasibility study towards identifying relevant features in intracranial human data in real-time, on-chip, using event-based processors and spiking neural networks. By providing “neuromorphic intelligence” to neural recording circuits the approach proposed will lead to the development of systems that can detect HFO areas directly in the operation room and improve the seizure outcome of epilepsy surgery.

5 Methods

5.1 Design and setup of the hardware device

The CMOS circuit simulations were carried-out using the Cadence® Virtuoso ADE XL design tools. All circuits including the headstage, the bias generator, and the silicon neurons were designed, simulated and analyzed in analog domain. The asynchronous buffers, spike routing network and chip configuration blocks were simulated and implemented in the asynchronous digital domain. The layout of the chip was designed using the Cadence® Layout XL tool. The design rule check, layout versus schematic and post-layout extraction were performed using the Calibre tool. We packaged our device using a ceramic 240-pin quadratic flat package. The package was then mounted on an in-house designed six-layer printed circuit board. The programming and debugging of the System-on-Chip (SoC) was performed using low-level software and firmware developed in collaboration with SynSense Switzerland, and implemented using the XEM7360 FPGA (Opal Kelley, USA). The pre-recorded iEEG was fed to the chip using a Picoscope 2205A MSO (Picotech, UK). All frequency-domain measurements were performed using a Hewlett-Packard 35670A dynamic signal analyzer.

5.2 Simulation and validation of the neural architecture

For the software simulation of the network we used the Spiking Neural Network simulator Brian2⁵⁸ and a custom made toolbox⁴⁹ that makes use of equations which describe the behavior of the neuromorphic circuits. To find the optimal parameters of the SNN, we were guided by the clinically relevant HFO marked by the Morphology Detector²⁴: Around the HFOs marked in the iEEG²⁴ we created snippets of data ± 25 ms. These snippets were used to select the parameters for the ADM and the SNN network (see Methods). The SNN architecture was validated using the previous generation of the neuromorphic processor DYNAP-SE³⁶, for which a working prototyping framework is available. The high-level software-hardware interface used to send signals to the SNN, configure its parameters, and measure its output was designed in collaboration with SynSense AG, Switzerland.

5.3 Patient data

We included long-term iEEG recordings from 9 patients of a data set that is publicly available³². Patients had drug-resistant focal epilepsy as detailed in Table 1. Presurgical diagnostic workup at Schweizerische Epilepsie-Klinik

included recording of iEEG from the medial temporal lobe. The iEEG data set has been analysed for HFO with state-of-the-art HFO detectors^{23,24}.

The surgical planning technique²⁰ was independent of HFO. Patients underwent resective epilepsy surgery at UniversitätsSpital Zürich. After surgery, the patients were followed-up for >1 year. Postsurgical outcome was classified according to the International League Against Epilepsy (ILAE):

Class 1 Completely seizure free; no auras.

Class 2 Only auras; no other seizures.

Class 3 One to three seizure days per year; ± auras.

Class 4 Four seizure days per year to 50% reduction of baseline seizure days; ± auras.

Class 5 Less than 50% reduction of baseline seizure days to 100% increase of baseline seizure days; ± auras.

Class 6 More than 100% increase of baseline seizure days.

In a previous publication²⁴, we detected HFOs with the Morphology detector^{24,29} and compared the HFO area to the resected area to predict seizure outcome. The dataset with the HFO markings³² is publicly available and was downloaded for the analyses presented here. The data consist of up to six intervals of approximately five minutes that were recorded in the same night during interictal slow-wave sleep, which promotes low muscle activity and high HFO rates. The intervals were at least three hours apart from epileptic seizures to eliminate the influence of seizure activity on the analysis. The amount of nights and intervals varied across patients (see Table 1). Because of the higher signal-to-noise ratio, in this study we focused on the recordings from the medial temporal lobe. Only the 3 most mesial bipolar channels were included in the analysis.

5.4 HFO detection

HFO detection was performed independently for each channel in each 5-min interval of iEEG. The signal pre-processing steps consisted of bandpass filtering, baseline detection and transforming the continuous signal into spikes using the ADM block. The ADM principle of operation is as follows: whenever the amplitude variation of the input waveform exceeds an upper threshold V_{tu} a positive spike on the UP channel is generated; if the change in the amplitude is lower than a threshold V_{td} , a negative spike in the DN channel is produced.

As the amplitude of the recordings changed dramatically with electrode, patient data and recording session, we introduced a baseline detection mechanisms that was used to adapt the values of the V_{tu} and V_{td} thresholds in order to produce the optimal number of spikes required for detecting HFO signals while suppressing the background noise and outliers in the recordings (see Fig. 6). This baseline was calculated for each iEEG channel in software: during the first second of recording, the maximum signal amplitude was computed over non-overlapping windows of 50 ms. These values were then sorted and the baseline value was set to the average of the lowest quartile. This procedure excluded outliers on one hand, and suppressed the noise floor on the other hand. This procedure was optimal for converting the recorded signals into spikes.

Spikes entered the SNN architecture as depicted in Fig. 5bc. The SNN parameters used to maximize HFO detection were selected by analyzing the Inter-Spike-Intervals (ISIs) of the spike trains produced by the ADM and comparing their characteristics in response to inputs that included an HFO event versus inputs that had no HFO events. This analysis was then used to tune the time constants of the SNN output layer neurons and synapses. Specifically, the approach used was to rely on an *ensemble* of neurons in the output layer and to tune them with parameters sampled from a uniform distribution. The average time constant for the neurons was chosen to be 15 ms, with a coefficient of variation set by the analog circuit devise mismatch characteristics, to approximately 20%. Similarly, the excitatory synapse time constants were set in the range [3 6] ms and the inhibitory synapse time constants in the range [0.1 1] ms.

After sending the spikes produced by the ADMs to the SNN configured in this way, we evaluated snippets of 15 milliseconds output data produced by the SNN and signaled the detection of an HFO every time spikes were present in consecutive snippets of data. Outlier neurons in the hardware SNN that spiked continuously were considered uninformative and were switched off for the whole study. The activity of the rest of the neurons faithfully signaled the detection of HFOs (see Table tab:metrics). For the HFO count, spikes with inter-spike-intervals <15 ms were aggregated to mark a single HFO.

5.5 Post-surgical outcome prediction

To retrospectively ‘predict’ the postsurgical outcome of each patient in this data set, we first detected the HFOs in each 5-min interval by measuring the activity of the silicon neurons in the hardware SNN. We calculated the rate of HFO per recording channel by dividing the number of HFOs in the specific channel by the duration of the interval.

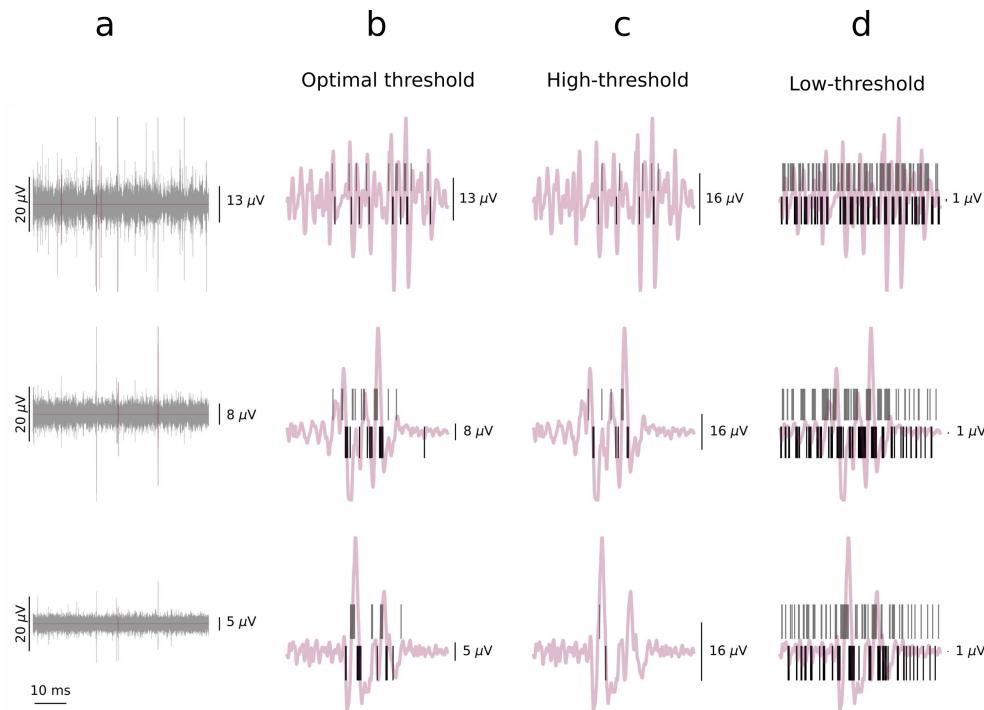


Figure 6. Optimal spike generation threshold for the asynchronous delta modulator (ADM). (a) iEEG from 3 channels in one electrode. The different noise floors are captured by the different baseline levels of [5, 8, 13] μV that we computed. (b) With the optimal channel-wise threshold, the number of UP and DN spikes is optimal for the example HFOs taken from each channel. (c) With a higher threshold (16 μV), the number of UP and DN spikes is too low for the example HFOs. (d) With a lower threshold (1 μV), the number of UP and DN spikes is too high for the example HFOs.

The distribution of HFO rates over the list of channel defines the HFO vector. In this way we calculated an HFO vector for each interval in each night. We quantified the test-retest reliability of the distribution of HFO rates over intervals by computing the scalar product of all pairs of HFO vectors across intervals (Table 1). We then delineated the “HFO area” by comparing the average HFO rate over all recordings and choosing the area under the electrodes with HFO rates exceeding the 95 percentile of the rate distribution. Finally, to assess the accuracy of the patient outcome prediction, we compared the HFO area identified by our procedure with the area that was resected in surgery, and compared it with the postsurgical seizure outcome (Table 1).

Acknowledgements

This project has received funding from Swiss National Science Foundation (SNSF 320030_176222) and from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program grant agreement No 724295.

Author contributions statement

JS and GI conceived the experiments, MS and KB conducted the experiments and analysed the results. All authors wrote and reviewed the manuscript.

References

1. Editorial-team. Big data needs a hardware revolution. *Nature* **554**, 145–146 (2018).
2. Maass, W. & Sontag, E. Neural systems as nonlinear filters. *Neural Comput.* **12**, 1743–72 (2000).
3. Kasabov, N., Dhoble, K., Nuntalid, N. & Indiveri, G. Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition. *Neural Networks* **41**, 188–201, DOI: [10.1016/j.neunet.2012.11.014](https://doi.org/10.1016/j.neunet.2012.11.014) (2013).

4. Indiveri, G. & Horiuchi, T. Frontiers in neuromorphic engineering. *Front. Neurosci.* **5**, 1–2, DOI: [10.3389/fnins.2011.00118](https://doi.org/10.3389/fnins.2011.00118) (2011).
5. Mead, C. Neuromorphic electronic systems. *Proc. IEEE* **78**, 1629–36 (1990).
6. Chicca, E., Stefanini, F., Bartolozzi, C. & Indiveri, G. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proc. IEEE* **102**, 1367–1388 (2014).
7. Bauer, F., Muir, D. & Indiveri, G. Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor. *Biomed. Circuits Syst. IEEE Transactions on* **13**, 1575–1582, DOI: [10.1109/TBCAS.2019.2953001](https://doi.org/10.1109/TBCAS.2019.2953001) (2019).
8. Corradi, F. *et al.* ECG-based heartbeat classification in neuromorphic hardware. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE, 2019).
9. Donati, E., Payvand, M., Risi, N., Krause, R. & Indiveri, G. Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network. *Biomed. Circuits Syst. IEEE Transactions on* **13**, 795–803, DOI: [10.1109/TBCAS.2019.2925454](https://doi.org/10.1109/TBCAS.2019.2925454) (2019).
10. Rahimi Azghadi, M. *et al.* Hardware implementation of deep network accelerators towards healthcare and biomedical applications. *arXiv e-prints arXiv–2007* (2020).
11. Harrison, R. The design of integrated circuits to observe brain activity. *Proc. IEEE* **96**, 1203–1216 (2008).
12. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
13. Frey, U. *et al.* Switch-matrix-based high-density microelectrode array in CMOS technology. *IEEE J. Solid-State Circuits* **45**, 467–482 (2010).
14. Ballini, M. *et al.* A 1024-channel CMOS microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro. *IEEE J. Solid-State Circuits* **49**, 2705–2719 (2014).
15. Khazaei, Y., Shahkooh, A. A. & Sodagar, A. M. Spatial redundancy reduction in multi-channel implantable neural recording microsystems. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 898–901 (IEEE, 2020).
16. Boran, E. *et al.* High-density ECoG improves the detection of high frequency oscillations that predict seizure outcome. *Clin. Neurophysiol.* **130**, 1882–1888 (2019).
17. Zijlmans, M. *et al.* How to record high-frequency oscillations in epilepsy: a practical guideline. *Epilepsia* **58**, 1305–1315 (2017).
18. Fedele, T. *et al.* Prediction of seizure outcome improved by fast ripples detected in low-noise intraoperative corticogram. *Clin. Neurophysiol.* **128**, 1220–1226 (2017).
19. Mohammadi, R. *et al.* A compact ECoG system with bidirectional capacitive data telemetry. In *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, 600–603 (IEEE, 2014).
20. Zijlmans, M., Zweiphenning, W. & van Klink, N. Changing concepts in presurgical assessment for epilepsy surgery. *Nat. Rev. Neurol.* **15**, 594–606 (2019).
21. Lesko, R. *et al.* The clinical utility of intraoperative electrocorticography in pediatric epilepsy surgical strategy and planning. *J. Neurosurgery: Pediatr.* **1**, 1–10 (2020).
22. Jobst, B. C. *et al.* Intracranial eeg in the 21st century. *Epilepsy Curr.* **20**, 180–188, DOI: [10.1177/1535759720934852](https://doi.org/10.1177/1535759720934852) (2020). PMID: 32677484, <https://doi.org/10.1177/1535759720934852>.
23. Farahmand, S., Sobayo, T. & Mogul, D. J. Using automated detection and classification of interictal hfos to improve the identification of epileptogenic zones in preparation for epilepsy surgery. *bioRxiv* 680280 (2019).
24. Fedele, T. *et al.* Resection of high frequency oscillations predicts seizure outcome in the individual patient. *Sci. Reports* **7**, 13836, DOI: [10.1038/s41598-017-13064-1](https://doi.org/10.1038/s41598-017-13064-1) (2017).
25. Frauscher, B. *et al.* High-frequency oscillations: the state of clinical research. *Epilepsia* **58**, 1316–1329 (2017).
26. Demuru, M. *et al.* The value of intra-operative electrographic biomarkers for tailoring during epilepsy surgery: from group-level to patient-level analysis. *Sci. Reports* **10**, 1–18 (2020).
27. van't Klooster, M. A. *et al.* Tailoring epilepsy surgery with fast ripples in the intraoperative electrocorticogram. *Annals neurology* **81**, 664–676 (2017).

28. Fedele, T. *et al.* Automatic detection of high frequency oscillations during epilepsy surgery predicts seizure outcome. *Clin. Neurophysiol.* **127**, 3066–3074, DOI: [10.1016/j.clinph.2016.06.009](https://doi.org/10.1016/j.clinph.2016.06.009) (2016).
29. Burnos, S., Frauscher, B., Zelmann, R., Haegelen, C. & Sarnthein, J. The morphology of high frequency oscillations (HFO) does not improve delineating the epileptogenic zone. *Clin. Neurophysiol.* **127**, 2140–2148 (2016).
30. Beniczky, S., Karoly, P., Nurse, E., Ryvlin, P. & Cook, M. Machine learning and wearable devices of the future. *Epilepsia* (2020).
31. Stirling, R. E., Cook, M. J., Grayden, D. B. & Karoly, P. J. Seizure forecasting and cyclic control of seizures. *Epilepsia* (2020).
32. Fedele, T. *et al.* High frequency oscillations detected in the intracranial EEG of epilepsy patients during interictal sleep, patients' electrode location and outcome of epilepsy surgery. *Collab. Res. Comput. Neurosci.* DOI: <http://dx.doi.org/10.6080/K06Q1VD5> (<http://dx.doi.org/10.6080/K06Q1VD5>, 2017).
33. Yang, M., Liu, S.-C. & Delbruck, T. A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding. *IEEE J. Solid-State Circuits* **50**, 2149–2160 (2015).
34. Corradi, F. & Indiveri, G. A neuromorphic event-based neural recording system for smart brain-machine-interfaces. *Biomed. Circuits Syst. IEEE Transactions on* **9**, 699–709, DOI: [10.1109/TBCAS.2015.2479256](https://doi.org/10.1109/TBCAS.2015.2479256) (2015).
35. Sharifshazileh, M., Burelo, K., Fedele, T., Sarnthein, J. & Indiveri, G. A neuromorphic device for detecting high-frequency oscillations in human iEEG. In *IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 69–72 (IEEE, 2019).
36. Moradi, S., Qiao, N., Stefanini, F. & Indiveri, G. A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs). *Biomed. Circuits Syst. IEEE Transactions on* **12**, 106–122, DOI: [10.1109/TBCAS.2017.2759700](https://doi.org/10.1109/TBCAS.2017.2759700) (2018).
37. Harrison, R. & Charles, C. A low-power low-noise CMOS amplifier for neural recording applications. *IEEE J. Solid-State Circuits* **38**, 958–965, DOI: [10.1109/JSSC.2003.811979](https://doi.org/10.1109/JSSC.2003.811979) (2003).
38. Wattanapanitch, W., Fee, M. & Sarpeshkar, R. An energy-efficient micropower neural recording amplifier. *IEEE Transactions on Biomed. Circuits Syst.* **1**, 136–147 (2007).
39. Delbruck, T. & Van Schaik, A. Bias current generators with wide dynamic range. *Analog. Integr. Circuits Signal Process.* **43**, 247–268 (2005).
40. Fleischer, P. & Tow, J. Design formulas for biquad active filters using three operational amplifiers. *Proc. IEEE* **61**, 662–663 (1973).
41. Shiue, M.-T., Yao, K.-W. & Gong, C.-S. A. A bandwidth-tunable bioamplifier with voltage-controlled symmetric pseudo-resistors. *Microelectron. J.* **46**, 472–481 (2015).
42. Lazar, A. A., Pnevmatikakis, E. A. & Zhou, Y. The power of connectivity: identity preserving transformations on visual streams in the spike domain. *Neural networks* **44**, 22–35 (2013).
43. Mahowald, M. *VLSI analogs of neuronal visual processing: a synthesis of form and function*. Ph.D. thesis, California Institute of Technology Pasadena (1992).
44. Deiss, S., Douglas, R. & Whatley, A. A pulse-coded communications infrastructure for neuromorphic systems. In Maass, W. & Bishop, C. (eds.) *Pulsed Neural Networks*, chap. 6, 157–78 (MIT Press, 1998).
45. Boahen, K. Communicating neuronal ensembles between neuromorphic chips. In Lande, T. (ed.) *Neuromorphic Systems Engineering*, 229–259 (Kluwer Academic, Norwell, MA, 1998).
46. Boahen, K. Point-to-point connectivity between neuromorphic chips using address-events. *IEEE Transactions on Circuits Syst. II* **47**, 416–34 (2000).
47. Bartolozzi, C. & Indiveri, G. Synaptic dynamics in analog VLSI. *Neural Comput.* **19**, 2581–2603, DOI: [10.1162/neco.2007.19.10.2581](https://doi.org/10.1162/neco.2007.19.10.2581) (2007).
48. Qiao, N. *et al.* A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses. *Front. neuroscience* **9**, 141 (2015).
49. Milde, M. *et al.* teili: A toolbox for building and testing neural algorithms and computational primitives using spiking neurons (2018). Unreleased software, Institute of Neuroinformatics, University of Zurich and ETH Zurich.

50. Payvand, M., Nair, M. V., Müller, L. K. & Indiveri, G. A neuromorphic systems approach to in-memory computing with non-ideal memristive devices: From mitigation to exploitation. *Faraday Discuss.* **213**, 487–510 (2019).
51. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **6**, 21–45 (2006).
52. Harrison, R. *et al.* A low-power integrated circuit for a wireless 100-electrode neural recording system. *IEEE J. Solid-State Circuits* **42**, 123–133, DOI: [10.1109/JSSC.2006.886567](https://doi.org/10.1109/JSSC.2006.886567) (2007).
53. Angotzi, G. N. *et al.* SiNAPS: An implantable active pixel sensor CMOS-probe for simultaneous large-scale neural recordings. *Biosens. Bioelectron.* **126**, 355–364 (2019).
54. Kim, S. *et al.* A sub- μ w/ch analog front-end for δ -neural recording with spike-driven data compression. *IEEE Transactions on Biomed. Circuits Syst.* **13**, 1–14 (2019).
55. Wang, S. *et al.* A compact quad-shank CMOS neural probe with 5,120 addressable recording sites and 384 fully differential parallel channels. *IEEE Transactions on Biomed. Circuits Syst.* **13**, 1625–1634 (2019).
56. Valtierra, J. L., Delgado-Restituto, M., Fiorelli, R. & Rodríguez-Vázquez, A. A sub- μ w reconfigurable front-end for invasive neural recording that exploits the spectral characteristics of the wideband neural signal. *IEEE Transactions on Circuits Syst. I: Regul. Pap.* **67**, 1426–1437 (2020).
57. Türe, K., Dehollain, C. & Maloberti, F. Implantable monitoring system for epilepsy. In *Wireless Power Transfer and Data Communication for Intracranial Neural Recording Applications*, 11–23 (Springer, 2020).
58. Goodman, D. & Brette, R. Brian: a simulator for spiking neural networks in Python. *Front. Neuroinformatics* **2**, DOI: [10.3389/neuro.01.026.2009](https://doi.org/10.3389/neuro.01.026.2009) (2008).

Figures

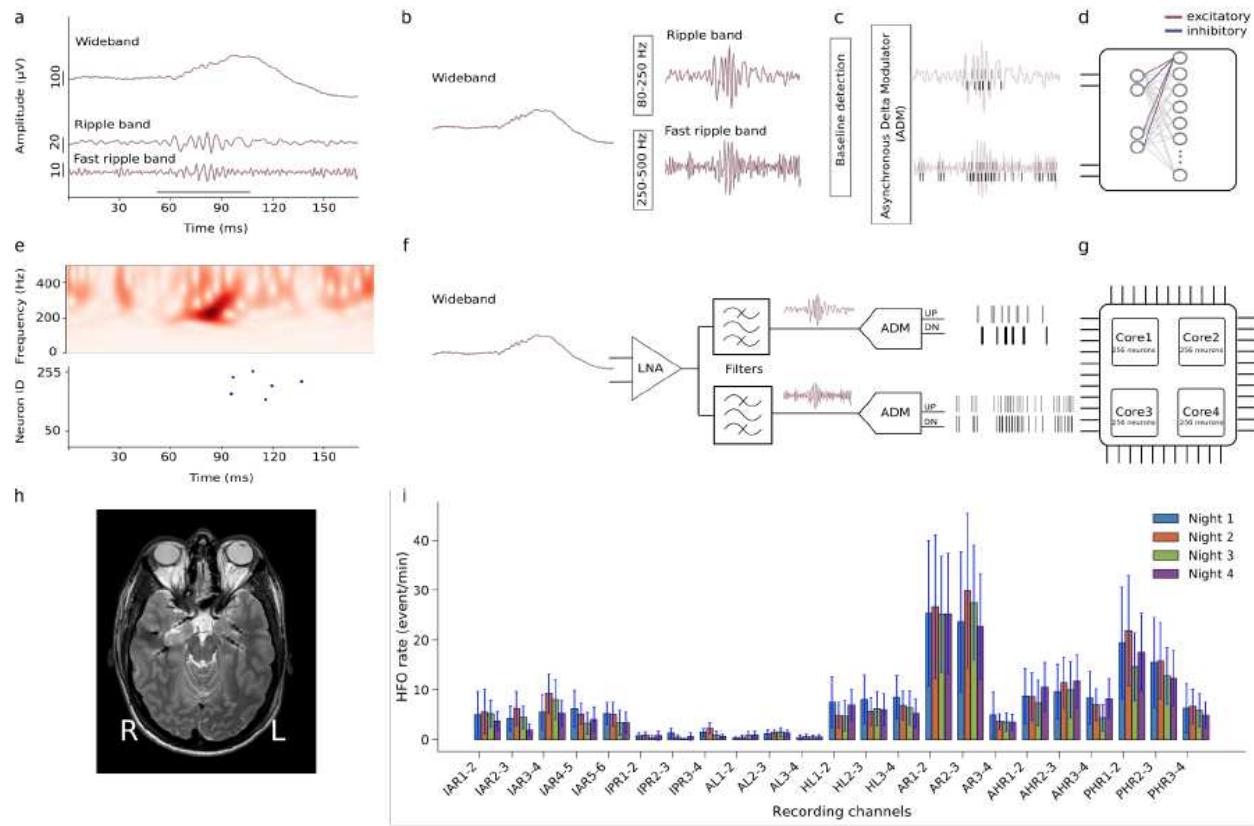


Figure 1

Automatic HFO detection using a bio-inspired SNN. (a) The pre-recorded iEEG signal in wideband, Ripple band (80-250Hz) and Fast Ripple band (250-500Hz). HFOs stand out of the baseline in the signal. The period marked by the gray bar represents a clinically relevant HFO24, 32. (b,c,d) Software simulated spiking neural network (SNN). For preprocessing the wideband EEG is filtered in Ripple band and Fast Ripple band. A baseline detection stage finds the optimum threshold that is applied in an Asynchronous Delta Modulator (ADM) which converts the signal to spikes. Signal traces are encoded by UP spikes (gray bars) and DOWN spikes (black bars), which are then fed as input into the SNN. The SNN is implemented as a two-layer spiking network of integrate and fire neurons with dynamic synapses. Each neuron in the second layer receives four inputs: two excitatory spike trains from UP channels and two inhibitory ones from DOWN channels. The parameters of the network were chosen to exhibit the relevant temporal dynamics and tune the neurons to produce output spikes in response to input spike train patterns that encode clinically relevant HFOs. (e, top) Time-frequency spectrum of the Fast Ripple iEEG of panel a. (e, bottom) Firing of SNN neurons indicate the occurrence of the HFO. (f) Block diagram of the neuromorphic system input headstage. The headstage comprises a low noise amplifier (LNA), two configurable bandpass filters and two ADM circuits that convert the analog waveforms into spike trains. (g) The spikes

produced by the ADMs are sent to a multi-core array of silicon neurons that are configured to implement the desired SNN. (h) MRI with 7 implanted depth electrodes that sample the mesial temporal structures of a patient with drug resistant temporal lobe epilepsy (Patient 1). (i) Rates of HFOs detected by the neuromorphic SNN for recordings made across four nights for Patient 1. HFO rate and variability across intervals within a night is indicated by standard error bars. Recording channels AR1-2 and AR2-3 in right amygdala showed the highest HFO rates which were stable over nights. Thus, the neuromorphic system would predict that a therapeutic resection, which should include the right amygdala, would achieve seizure freedom. Indeed, a resection including the right amygdala achieved seizure freedom for >1 year.

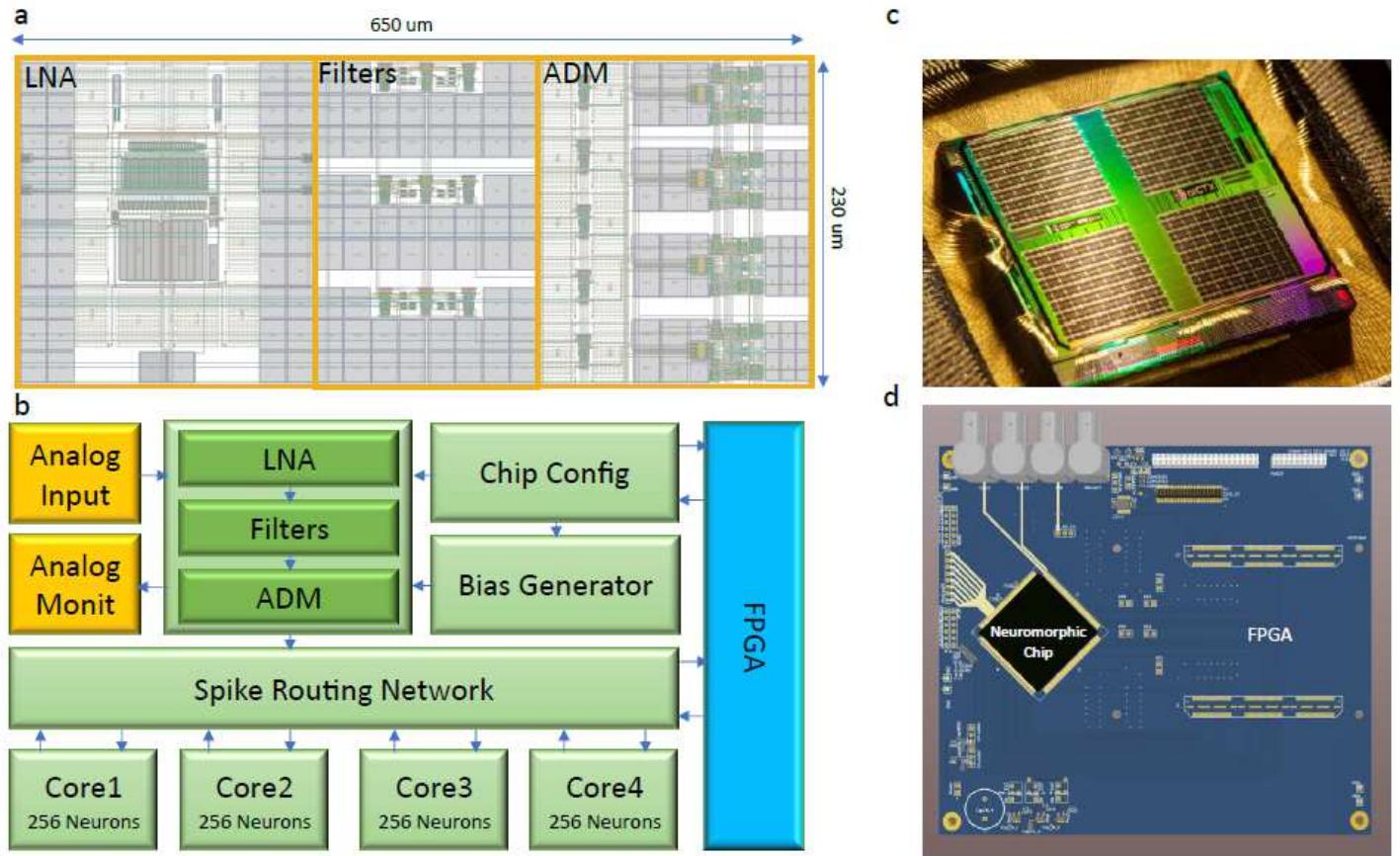


Figure 2

Neuromorphic electronic system overview (a) Physical layout of a single channel of the analog headstage array, including the LNA, three low-pass/band-pass filters, and four ADM signal to spike encoders. (b) Reduced block diagram of the neuromorphic platform. Analog signals from electrodes are fed into the input headstage that converts them into spike trains and sends them to the SNN implemented on the multi-neuron cores, via a spike routing network. The spike routing network routes the spikes within on-chip SNN and to an external FPGA device used for data logging and prototyping. The FPGA is also used for setting circuit parameters (c) Chip photograph showing the 11mm x 9mm silicon die. (d) Prototyping Printed Circuit Board (PCB) used to host the chip and the infrastructure to implement the test setup. The setup is composed of a prototyping FPGA board mounted on the same PCB that hosts

the chip, and of probe points to evaluate the characteristics of both input headstage and SNN multi-core network.

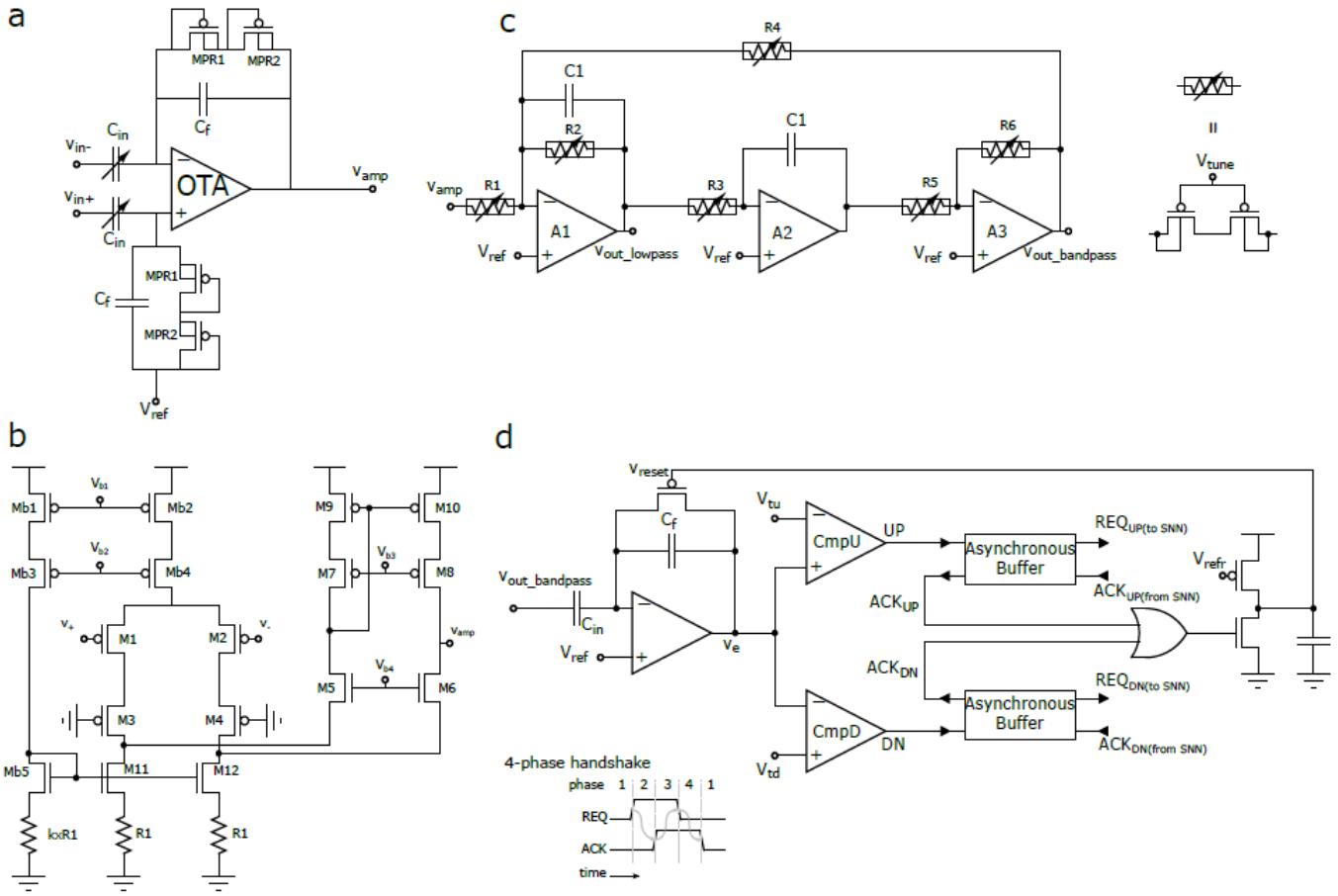


Figure 3

Schematic diagrams of the input headstage circuits. (a) Variable-gain LNA using variable input capacitor array and pseudo-resistors. The gain of this stage is calculated by $\times 8 = \times 5$; the use of the pseudo-resistors allows to reach small low cut-off frequencies. (b) Folded cascode OTA using resistive degeneration to reduce the noise influence of nMOS devices. Note that the current flowing through the bias branch is : times smaller than the tail branch of the amplifier. (c) band-pass (Tow-Thomas) filters for performing second-order filtering in both ripple and fast-ripple bands as well as low-frequency component of the iEEG. Tunable pseudo resistors are used to adjust the filter gain, center-frequency and band-width. The same basic structure can be used to provide both low-pass and band-pass outputs, thus is desirable in terms of design flexibility. (d) Asynchronous Delta Modulator (ADM) circuit to convert the analog filter outputs into spike trains. The ADM input amplifier has a gain of $\times 8 = \times 5$ in normal operation when $+A4$ $B4C$ is low and the feedback PMOS switch is off. As the amplified signal crosses one of the two thresholds, $+CD$ or $+C3$, an UP or DN spike is produced by asserting the corresponding REQ signal. A 4-phase handshaking mechanism produces the corresponding ACK signal in response to the spike. Upon receiving the ACK signal, the ADM resets the amplifier input and goes back to normal operation after a refractory period determined by the value of $+A4$ $5 A$. The asynchronous buffers act as 4-phase

handshaking interfaces that propagate the UP/DN signals to the on-chip Address-Event Representation (AER) spike routing network of Fig. 2.

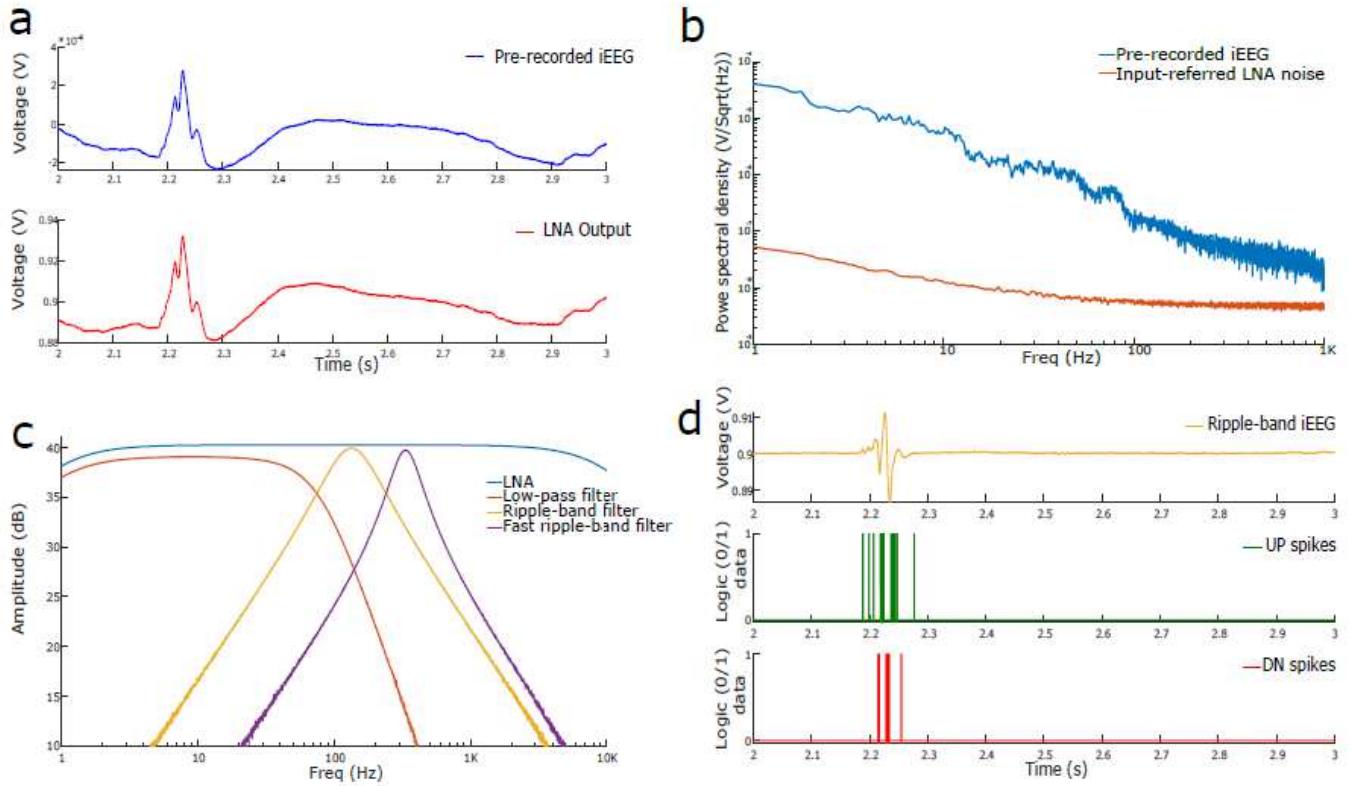


Figure 4

Measurements from the analog headstage. (a) An iEEG sample32 and the LNA amplified output. (b) Noise floor of the headstage LNA and iEEG power spectral density. In the HFO range (80-250 Hz) the noise level of the LNA is below the iEEG noise floor by an order of magnitude. (c) Frequency response of the implemented filters in the headstage. The band-pass filters are tuned to highlight HFO frequency bands. (d) ADM output spikes in response to the ripple-filtered signal. The top plot shows the analog filter output, the middle plot shows the UP spikes generated by the ADM and the bottom plot shows the DN spikes.

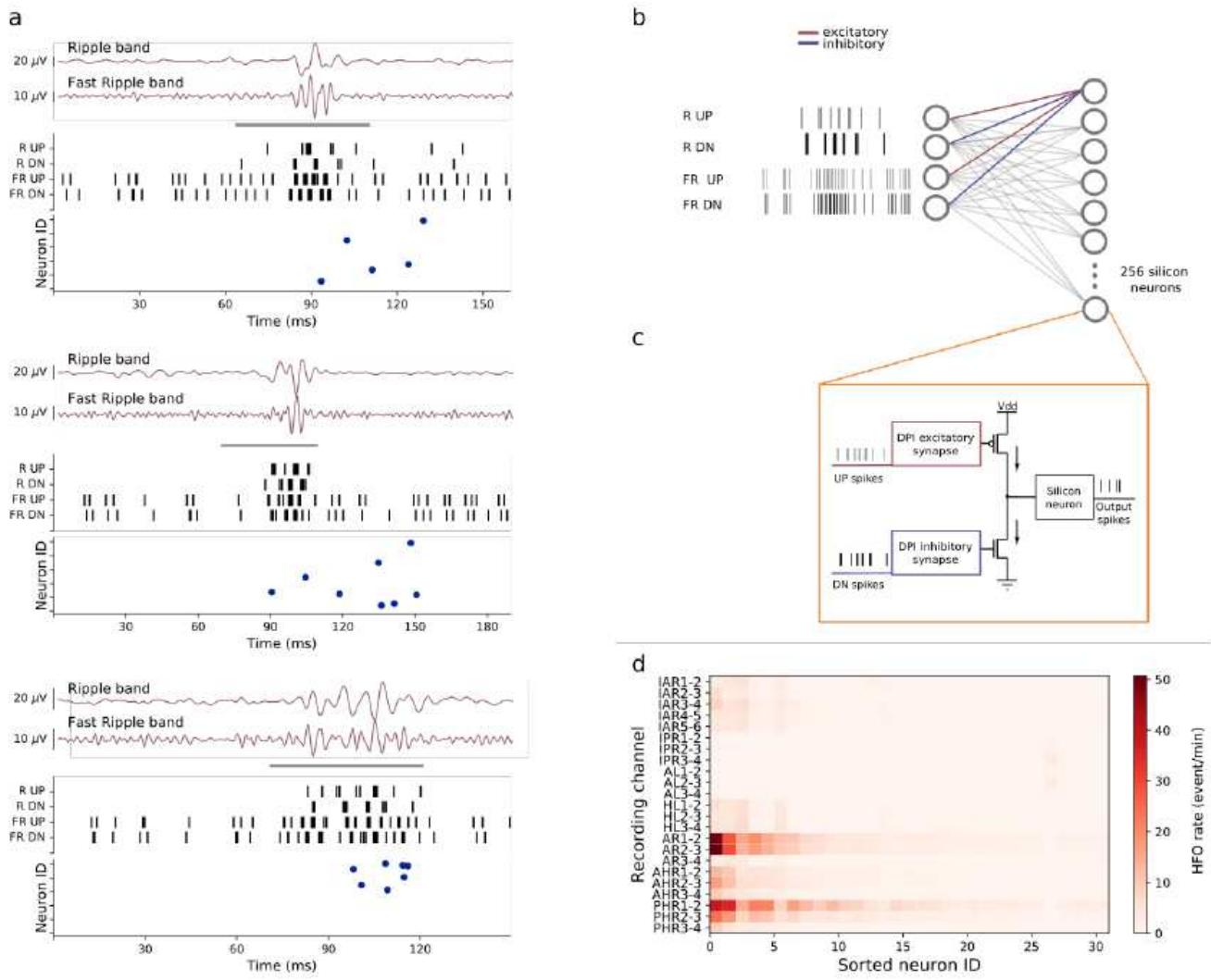


Figure 5

(a) Examples of HFOs that the hardware SNN detected in the iEEG of Patient 1. The periods marked by the gray bar represent clinically relevant HFO24, 32. The signals in Ripple and Fast-ripple band were transformed to UP and DN spikes. These spike trains were sent to the neurons in the hardware through the '%', '#', '%%' and '#%' channels. The bottom panel of each example shows the raster plot of the silicon neurons. Each neuron responds to different HFOs depending on the characteristics of the pattern. (b) The SNN architecture consist of a two-layer network of 256 integrate and fire neurons with dynamic synapses. Each neuron in the second layer receives excitatory connections from the '%' and '%%' channels, and inhibitory connections from the '#' and '#%' channels. The synaptic parameters time constants and weights are distributed randomly within a predetermined optimal range. (c) Hardware building blocks used for the implementations of the SNN: the DPI synapse is a “Differential-Pair Integrator” circuit⁴⁷, and the silicon neuron is an Adaptive Exponential Integrate and Fire (AdExp IF) circuit⁴⁸. (d) HFO rates computed for Patient 1. The neurons are sorted according to their average firing rate. Only a small number of neurons fire across all the recordings, even for channels with high HFO rates (e.g. AR1-2).

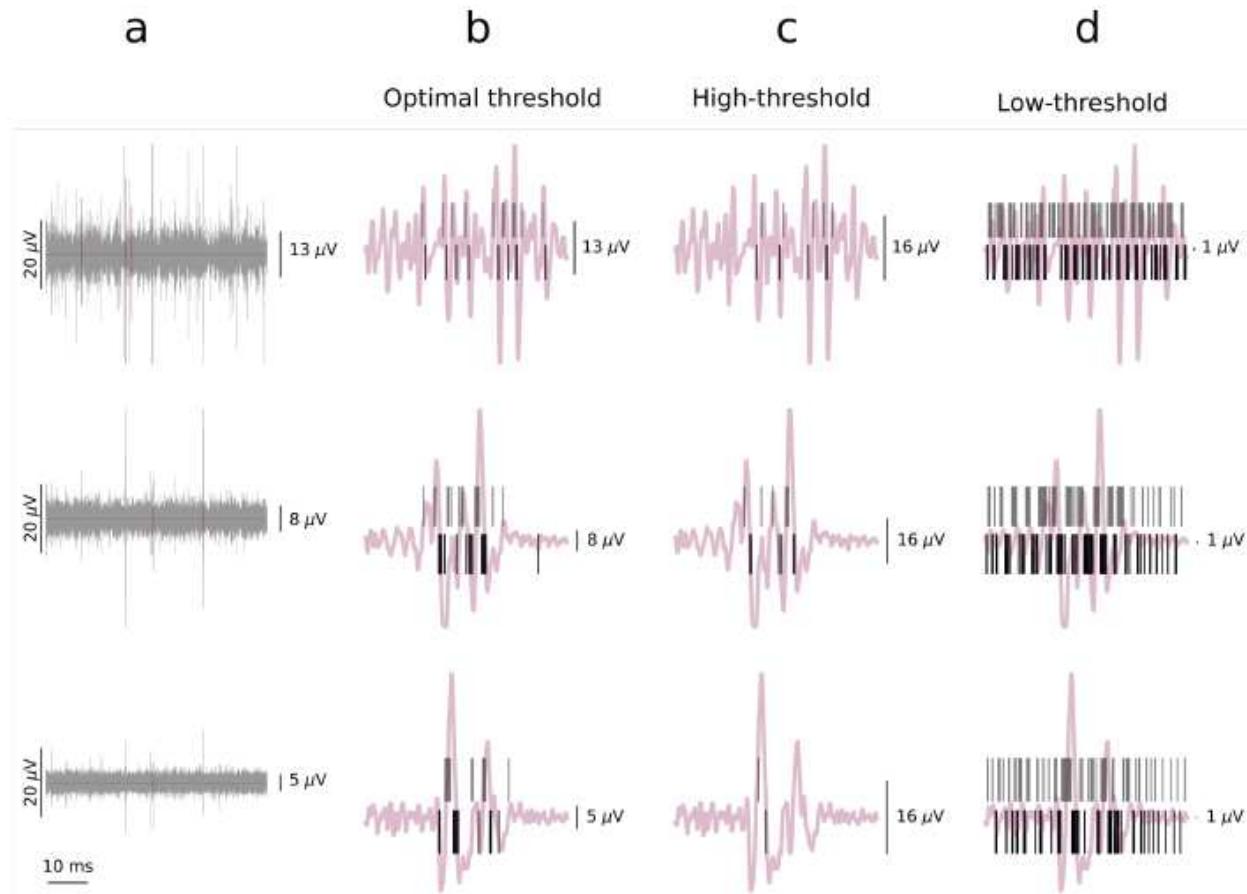


Figure 6

Optimal spike generation threshold for the asynchronous delta modulator (ADM). (a) iEEG from 3 channels in one electrode. The different noise floors are captured by the different baseline levels of [5, 8, 13] μV that we computed. (b) With the optimal channel-wise threshold, the number of UP and DN spikes is optimal for the example HFOs taken from each channel. (c) With a higher threshold (16 μV), the number of UP and DN spikes is too low for the example HFOs. (d) With a lower threshold (1 μV), the number of UP and DN spikes is too high for the example HFOs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [BCIINI.mp4](#)