Sub 100 nW Volatile Nano-Metal-Oxide Memristor as Synaptic-Like Encoder of Neuronal Spikes

Isha Gupta , *Member, IEEE*, Alexantrou Serb , *Member, IEEE*, Ali Khiat, Ralf Zeitler, Stefano Vassanelli, and Themistoklis Prodromakis , *Senior Member, IEEE*

Abstract—**Advanced neural interfaces mediate a bioelectronic link between the nervous system and microelectronic devices, bearing great potential as innovative therapy for various diseases. Spikes from a large number of neurons are recorded leading to creation of big data that require online processing under most stringent conditions, such as minimal power dissipation and on-chip space occupancy. Here, we present a new concept where the inherent volatile properties of a nano-scale memristive device are used to detect and compress information on neural spikes as recorded by a multielectrode array. Simultaneously, and similarly to a biological synapse, information on spike amplitude and frequency is transduced in metastable resistive state transitions of the device, which is inherently capable of self-resetting and of continuous encoding of spiking activity. Furthermore, operating the memristor in a very high resistive state range reduces its average in-operando power dissipation to less than 100 nW, demonstrating the potential to build highly scalable, yet energy-efficient on-node processors for advanced neural interfaces.**

Index Terms—**Integrating sensor, memristors, metastable resistive state, neural recordings, RRAM, volatility, volatility module.**

I. INTRODUCTION

REVERSE engineering the human brain and decoding the underlying information processes of biological systems requires integrated efforts from researchers with different scientific backgrounds [1]. Towards enabling this vision, advances in neural recording techniques [2]–[5] target the reliable acquisition of electrophysiological data from multiple neurons in-vitro and in-vivo. This has impacted our understanding of information processing by brain microcircuits [6] and brought new prospects for novel therapies based on adaptive neural stim-

ulation [7]. To date, state-of-art implementations can simultaneously record in-vivo [8] from up to thousand sites and from up to 30 k [9] sites in-vitro using Complementary Metal Oxide Semiconductor (CMOS) based High Density Microelectrode Arrays (HDMEAs). Such advances in micro-sensors technology have been paralleled by considerable progress in neural processing microsystems [10], [11] which are capable of detecting neural spiking activity on-node [12], [13]. The relevant spike-detected information is then transmitted off-line wirelessly and techniques such as the Template Matching System (TMS) or Principle Component Analysis (PCA) [14] are used off-line for spike-sorting [15]. These methods, by mapping the recorded neural activity to the source active neurons, offer insights in neural coding principles [16] and support novel neuroprosthetic applications [17], [18]. Thus, further advances in the fast developing field of implantable neural interfaces [19] are hampered by key bottlenecks in the processing of neuronal spikes including: a) computational power required to process the ever increasing volume of neural signals (Gb/s range presently) on-node and in real-time [20]–[23], b) bandwidth [24] and, c) scalability.

Recently, we proposed a new spike-detection approach [25] based on metal-oxide resistive switching memory devices, also known as memristors [26]–[28]. Fundamentally, memristive devices undergo non-volatile resistive state transitions as a function of the integral of the input voltage, thus behaving as thresholded input integrators [29]. Taking advantage of this property, we demonstrated that "TiO_2-based" memristive devices can be employed for spike-detection [25], as extracellular neural spikes recorded from retinal ganglion cells [30]–[32] were encoded in gradual, non-volatile resistive state transitions, whereas the sub-threshold events (i.e. noise) were naturally filtered-off [25]. This property makes these devices suitable as noise-suppressing integrating sensors and are thus termed as Memristive Integrating Sensors (MIS). Non-volatility, however, was strongly limiting detection performance, as "after saturation of the resistive state" of the devices, it 'failed' to register any significant neural activity [33]. Consequently, performance was optimised by manual operation through frequent resets to the initial devices resistive state, which however impacts negatively on the overall power consumption [34].

In this work, we advance on our previous findings by exploiting an often overlooked crucial property of memristive devices that is "volatility" [35]–[39]. This approach recalls the way of operation of biological synapses that translate spiking frequency in gradual changes of postsynaptic conductance
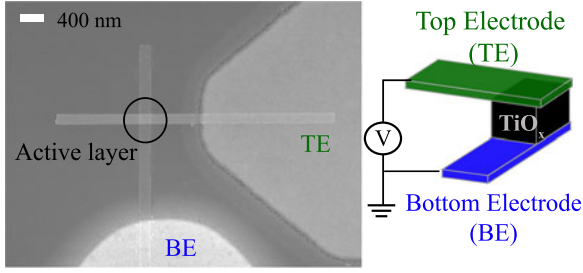
Fig. 1. Device architecture of nanoscale TiOx memristive devices. (a) Schematic (right) and Scanning Electron Microscopic (SEM) image (left) of the employed 200 nm × 200 nm memristive device.

subject to a continuous self-resetting process [40]. When used in the volatile operating region, memristive devices exhibit metastable memory state transitions following which they inherently relax to their initial resistive state range. We demonstrate that volatility enables for naturally encoding spiking events into transient resistive changes. The self-resetting mechanism of our devices, ensures that these operate far from their resistive saturation region, which overall enhances the attained spike-detection accuracy. Moreover, we particularly exploit the fact that volatile phenomena are more pronounced at higher resistive states [37] reducing the overall power dissipation to less than 100 nW, setting a new state-of-art in spike detectors using nano-scale devices.

II. CONCEPT, METHODS AND DATA PROCESSING

A. Device Fabrication

In this manuscript, two sets of devices i.e. nano- and micro-structure devices were experimentally tested. The fabrication procedure for the two sets is described as follows: All nano-devices exploited in this work (Fig. 1), Ti/Pt/TiO_{2-x}/TiN (5/10/10/40 nm) were fabricated as follows: 6-inch wafer was thermally oxidised to grow 200 nm SiO_2, which serves as an insulating layer. Then, direct write e-beam lithography method was adopted, to pattern the bottom electrodes (BEs) nanowires. Double layer resists were used to facilitate lift-off process of the BEs, which are constituted of 5 nm adhesive Ti layer and 10 nm Pt film. BEs were deposited using e-beam evaporation. Bottom access-electrodes (large features) were then defined via conventional photolithography patterning, e-beam evaporation of Ti/Au (5 nm/25 nm) and lift-off process. Access-electrodes connect the pads to the nanowires. To pattern the active layer, optical lithography, reactive sputtering and lift-off process were also used. 10 nm near-stoichiometric TiO_{2-x} active layer was sputtered from a Ti metal target with $x = 0.06$. Next, 40 nm thick TiN top electrode (TEs) nanowires and 25 nm thick Au top access-electrodes were obtained in a similar manner to BEs and to the bottom access-electrodes, respectively, with the TiN films.

Micrometre sized devices, Ti/Pt/TiO_{2-x}/Pt (5/10/25/10 nm) were fabricated on top of Si/SiO_2 wafers, where the oxide layer was 200 nm thick. In each layer, three main patterning steps were processed, optical lithography, film deposition and lift-off process. For the first layer, 5 nm Ti and 10 nm Pt films were

deposited via electron-beam evaporation to serve as BEs. In the second, magnetron reactive sputtering system was used to deposit the 25 nm near stoichiometric TiO_{2-x} active film. In the final step, 10 nm Pt TEs were deposited using electron-beam evaporation system, as well.

B. Hardware Infrastructure/Instrumentation

The electrical characterisation in pulsing mode was implemented using in-house fabricated electronic hardware infrastructure based on mBED LPC1768 micro-controller board [41]. This instrument is capable of addressing single, up to 32 devices, or crossbar arrays up to 1 kb in size (32 × 32 devices). The instrument is used to directly test the devices on-wafer interfaced via a multi-channel probe card. The hardware platform is supported by custom-made Graphical User Interface (GUI) that permits device-by-device fully automated testing.

C. Electrical Characterisation in Non-Volatile Regime

In the non-volatile regime, the devices are electrically characterised in two stages. Initially, the devices undergo an electroforming step. This involves application of ramp of voltages on a pristine sample until a sudden, non-volatile change in the resistive state of the device is observed. This typically occurs in the range of +6 to 8 V [25] and the device is then considered to be in low resistive state that is "ON" state. The resistive state of the device decreases from 10's of MΩ down to 10 kΩ. Thereafter, a train of input programming pulses in alternating polarities is applied at a fixed duration of 100 μs which leads to reversible resistive switching [42]. The resistive state of the Device Under Test (DUT) is read after each programming pulse at 0.5 V. The device is switched to low resistive state (ON/SET state) and high resistive state (OFF/RESET state) with positive and negative polarity respectively after the applied stimulus exceeds the DUTs inherent threshold voltage (V_{th}).

Importantly, when operated in non-volatile region, devices gradually switch within a 2 to 15 kΩ range. The response of the DUT for one specific input stimulus can be fitted to a second order exponential function indicating saturation of its resistive state due to continued operation [34]. The devices are asymmetric and demonstrate slightly different threshold for distinct polarities. The inherent threshold voltage of the employed devices varies in the range of $\approx\pm0.6$–2.5 V [25].

D. Device Electrical Characterisation in Volatile Regime

In the volatile-regime, the devices tend to undergo metastable resistive state transitions following which they relax to an equilibrium resistive state in a finite time window [43]. The devices are electrically characterised in two stages. The electroforming stage is similar to the one described earlier (Section II-C). For prototype devices with μm scale active areas we used negative polarity pulses (-6 to -8 V), while positive polarity pulses ($+4$ to $+6$ V) were employed in the case of devices with nm scale (200 nm × 200 nm) active core areas. Importantly, in the volatile region devices operate in a rather high resistive state range of $\approx$300 kΩ–3 MΩ. In the second stage, the devices are
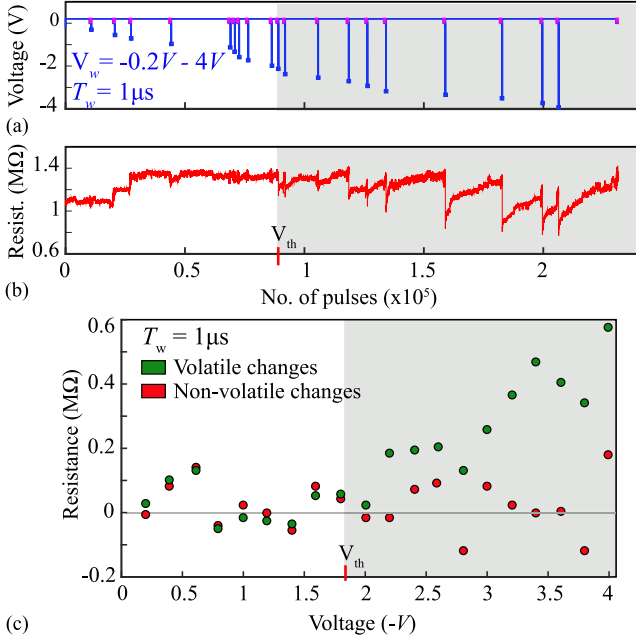
Fig. 2. Electrical characterisation of memristive device exhibiting volatility. (a) An automatic volatility characterisation module is utilised and the resistive state evolution of the DUT in response to the applied stimuli is illustrated in (b) [43]. After a first electroforming stage, the algorithm applies a series of progressively increasing input pulses (blue bars) of amplitude (V_w) and width (T_w) following which its resistive state is monitored. Pink rectangles indicate 60 s of retention time. (c) Determination of the operating voltage range of the DUT in volatile region. For every step of input stimulus applied, volatile (green circles) and residual non-volatile (red circles) resistive state changes were measured. The grey band indicates the voltage region for ensuring operation of the DUT in volatile conditions, with ≈ -1.8 V being the inherent threshold voltage of the DUT (V_{th}).
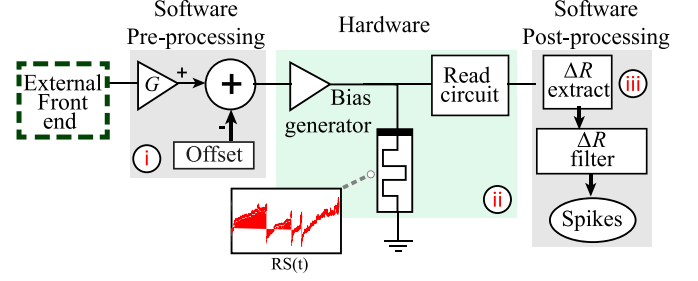

Fig. 3. Schematic for the implementation of our memristor-based spike-detection platform [25]. (a) G and V_{off} stage, (b) Hardware and (c) post-processing stage as explained in Section II-F.

characterised using an algorithm developed specifically to evaluate the retention characteristics of our prototypes (Fig. 2 and [43]).

The module applies a series of progressively more invasive pulses and then estimates the resistive state of the DUT using the standard two mean t-test method over a fixed interval of time. In practice, the t-test captures the resistive state progressive decay of the DUT and the module terminates when an equilibrium state is achieved. Subsequently, the t-test is followed by a retention condition test (Fig. 2). In this test, the equilibrium state of the device is checked for a user-defined period of time. At the end, the output of the algorithm determines the time elapsing to achieve the equilibrium condition and the voltage ranges under which the devices can be safely operated in the volatile region, as estimated by comparing non-volatile with volatile changes (Fig. 2(c)). More clearly, output of the volatility module is estimate of resistive changes between the steady state and measurements taken immediately before and after the applied voltage stimulus. Notably, the algorithm makes no assumption on what the equilibrium resistive state should be.

E. Front-End Neural Recording System

In this work, neural activity was recorded from slices of dissected mid-peripheral rabbit retinal ganglion cells using an extended-CMOS technology [32]. The CMOS based multi-transistor array (MEA) consists of 128×128 sensor sites, which

records the data at a sampling rate of 12.2 kHz and outputs a current time series containing approximately 63 k samples. The sensor sites of the CMOS-MEA are insulated by an inert TiO_2/ZrO_2 layer and a thin metal layer beneath the oxide layer is connected to the gate of the field-effect transistor. The voltage changes due to the interfaced neural tissue/cells above the recording sites are used for modulating the source-drain current in the MOSFET. Trans-Impedance Amplifiers (TIA) fabricated on-chip convert the signal into voltage and amplify the signal from a 0.1 mV–1 mV range up to 10–500 mV in range. This amplified signal is then used as an input for our platform. The CMOS MEA was kept external to the presented platform and in this work the CMOS MEA is termed as the "front-end" system.

F. Neural Recording Biasing Strategy

The input to our experimental platform is pre-processed using a software-based gain (G) and offset stage (V_{off}). This suitably amplified neural trace is then used to bias individual memristive devices using the customised hardware and the signal is fed to target devices in batches of 1000 data-points (Fig. 3).

In each processing batch, the resistive state of the device is assessed five times that is at the beginning of each batch, then after every 300 samples and finally at the end of each batch. Four consecutive measurements are thus obtained at 300th, 600th, 900th and 1000th data-point, and one measurement is made at the end of each batch and before the beginning of next batch without any neural data point in between.

Our method transforms a batch of 1000 data-points in five bins and thus results into an overall data compression rate of 200. Resistive state changes can be extracted from the consecutive measurements (bins) whilst the measurement uncertainty (N) can be estimated from the measurements made at the end of each batch and the beginning of the next batch. As a result, for a single neural recording ($\approx$63k points) we obtain 316 resistive state measurements corresponding to 252 consecutive resistive state changes and 64 noise-level measurements. These resistive state changes help in estimating the threshold voltage of the target device and consequently differentiating the significant resistive state changes from the insignificant ones [25].

G. Neural Signal Processing

Estimation of the spikes detected by our platform involves post-processing of the resistive state measurements. The nor-
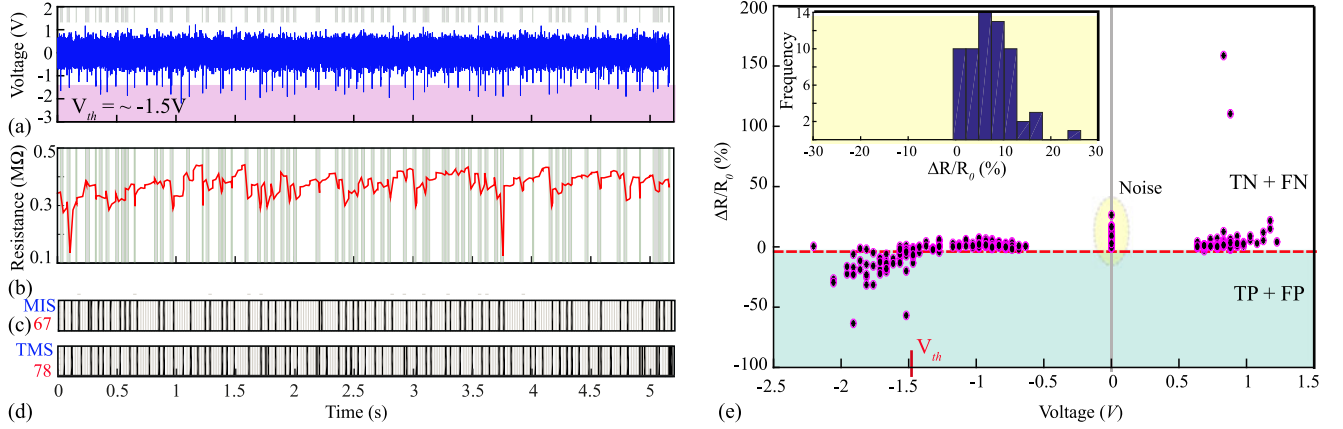
Fig. 4.   Spike-detection via volatile metal-oxide memristive device and benchmarking of the results against the TMS. (a) Neural recording used for biasing the DUT. The pink band indicates the inherent threshold voltage of the DUT (V_{th}) which is $\approx$ equal to -1.5 V. (b) Resistive state evolution of the DUT with time in response to the neural recording in (a). Time intervals where our spike-detector identifies events are indicated in grey. (c) total number of spikes detected by our system and (d) via the TMS. (e) Normalised changes in the resistive state of the device ($\Delta R/R_0$) in each bin are plotted as a function of highest voltage magnitude in each bin. The resistive state measurements in the yellow eclipse represents the noise measurements. The inset represents the histogram for the noise measurements. The red dashed line on the horizontal axis represents the boundaries of the noise band estimated using the 4σ method whilst the green band indicates the significant resistive state changes as detected by our system.

malised resistive state changes in each bin are plotted as a function of the highest voltage magnitude in each bin (see Fig. 4(e)). In the volatile region of operation, all the noise measurements made in the positive polarity are discarded. Only noise measurements in negative polarity are used for estimating the noise band boundaries. Absolute values of the noise measurements in the negative polarity are generated and then standard deviation (σ) is calculated. 4σ is used and the noise band boundary is set as shown by the horizontal red dashed line in Fig. 4(e). All the resistive state changes outside this boundary are considered to be significant whilst the resistive state changes falling within this band are disregarded.

We argue that these resistive state modulations occur due to weak amplitude neural signals and cannot be differentiated from the noise measurements. On the other hand, if noise measurements in the positive direction are included, estimation of mean and standard deviation values will be significantly affected leading to higher probability of inclusion of noise. The noise band and an ideal threshold voltage divides the resistive state measurements in four parts and the data is quantified as follows Fig. 4(e) [34]: True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN). These results are benchmarked against state-of-the-art TMS and these parameters are then redefined. TP: indicates an agreement between the two systems for spike detected in a given bin. TN: means the two systems agree that there is no activity in a given bin, FP: our system detects an event whilst TMS system doesnt and FN: TMS system detects a spike whereas our system does not. Using these values, rate of TP (TPR) and FP (FPR) is determined using following equations:

$$TPR = TP/(TP + FN) \qquad (1)$$

$$FPR = FP/(FP + TN) \qquad (2)$$

III. RESULTS

A. Operation of Nanoscale Memristive Devices as Volatile Cells

A preliminary characterisation of nano scale TiO_2 based metal-oxide memrisitve devices, as shown in Fig.1(a) is needed to identify appropriate operation conditions in the volatile region [43]. A custom hardware infrastructure was used for the electrical characterisation (Section II-B). The devices are electroformed and then the devices are operated in non-volatile or volatile manner depending upon the polarity of the voltage stress applied during the electroforming procedure and the strength of stimuli (Sections II-C and II-D). Importantly, the devices feature inherent threshold levels, below which there is no change in the resistive state and above which resistive state transitions are observed.

When operated in the volatile region, devices undergo metastable resistive transitions within a high resistive state range and are capable of inherently re-attaining their approximate initial resistive state. To enable this study, a volatility characterisation algorithm was developed (Section II-D). As shown in Fig. 2(a), (b), the operating resistive state region of the DUT was approximately 700 kΩ–1.4 MΩ, using negative as the dominant stimulus polarity with 1 μs pulses.

The volatility module results in discrimination between non-volatile and volatile resistive changes for a given voltage stimulus as exemplified in Fig. 2(c), where a -1.8 V threshold voltage marks the transition of the DUT to a prevalent volatility state. The resistive state changes in the sub-threshold region are considered as insignificant and are mainly attributed to the background fluctuations caused due to the measurement noise. The range of identified inherent threshold voltages for the employed TiO_2-based memristive devices varied from ≈ -0.6 V to -2.5 V. In conclusion, through the volatility module we determined the range in which the devices could be safely operated

in the volatile region and roughly estimated the relaxation times to equilibrium. The estimated volatility parameters were subsequently used when pre-processing the neural recordings in the spike-encoder platform (Section II-F).

The platform's schematic is illustrated in Fig. 3 and an overall picture of the spike-encoding system including the front-end system is presented in [25]. The neural recordings were obtained from the front-end system (Section II-E) and processed as described in section (Sections II-F and II-G).

In the first stage (Fig. 3(a)) the signals are pre-processed using a suitable G and V_{off} value. The neural signals are amplified such that the spikes but not the noise are above the volatility threshold of the DUT determined using volatility module. Pre-processed neural recordings are then passed through the memristive devices in batches (Fig. 3(a)) and the resistive state of the DUT is read periodically in real-time. This is done using the standard signal processing schematic described in Section II-F, thus requiring lower bandwidth as opposed to transmitting the full voltage-time series. This leads to segmentation of the neural data into smaller bins. Pairs of consecutive measurements in each batch are used to estimate the resistive state changes, whilst the pair of resistive state measurements taken at the end of each batch and the beginning of the following is used to estimate noise or reference data values (Section II-G). Finally, resistive state changes compressed in this way are processed offline and compared to noise (Fig. 3(c)), with significant changes being recognized as spikes. Notably, this concept allows a tradeoff between sampling rates and timing resolution making it suitable for applications where only a measure of overall activity within the given time bins is required [44].

B. Memristive Devices as Volatile Spike-Detectors

Employing TiOx memristive devices as volatile spike-detectors requires pre-processing any neural recording input to match the volatile operating region of devices as determined using the volatility characterisation algorithm. To illustrate the concept, we chose a neural recording that contains a dense spiking pattern as illustrated in Fig. 4(a). The operational parameters i.e. G and V_{off} used for pre-processing the input neural recording were for this case fixed at 3.2 and 0, respectively. On biasing the target device with the neural recording the intrinsic reset capability of the DUT can be clearly noted in Fig. 4(b). For instance, the initial resistive state of the DUT is approximately 350 kΩ following which the device demonstrates metastable resistive state transitions towards a low resistive state in response to supra-threshold spiking events. In the case where no subsequent supra-threshold events occur, the state of the device relaxes back towards the initial device state, demonstrating an inherent reset.

The number and approximate timing of spikes are estimated after post-processing of the resistive state measurements obtained using the standard schematic described in the methods Section II-F. As shown in Fig. 4(e), resistive state change (ΔR/R_0), R_0 being the initial resistive state value in each bin is plotted as a function of highest voltage magnitude in each bin. The resistive state change magnitude distribution of noise

measurements is used for identifying significant resistive state modulation that corresponds to spiking events. The inset of Fig. 4(e) represents the histogram for the noise measurements indicating an excessive inclination towards the positive polarity. Since the dominant stimulus polarity is negative and as a result neuronal activity-induced resistive changes are in the negative direction, the noise measurements in the positive direction are completely discarded and only the measurements in the negative direction are used to estimate meaningful noise band boundaries. This filters out the intrinsic reset transitions, which occur exclusively in the positive direction. Thus, noise band boundaries are estimated using only negative polarity noise measurements with a 4σ method (assuming Gaussian distribution), as indicated by the horizontal dashed line in Fig. 4(e). Everything outside this band in the negative region is considered as significant modulation corresponding to a spiking event whilst all events registered within this band are disregarded as these do not correspond to state modulations due to spiking events. Following this methodology, the total number of spikes detected by our system, as shown in Fig. 4(c), is equal to 67.

The performance of our memristor spike-detector is benchmarked against the established state-of-the-art TMS [34]. As depicted in Fig. 4(d), the total number of spikes determined by the TMS is equal to 78. In Fig. 4(e), approximately -1.5 V represents the inherent threshold voltage of the DUT. The negative quadrant (green) represents the spikes detected by our platform i.e. sum total of TP and FP whilst the positive quadrant indicates the FN and TN. These quantification parameters are used to evaluate the sensitivity of our system through the TPR and FPR of detection (Section II-G). Assuming TMS to be a perfect spike detector the two values are estimated to be 74.35% and 5.14% respectively. Both values represent a significant improvement in performance compare to when the devices are operated in the non-volatile region. For example, for the same neural recording in [34], TPR and FPR after introducing optimised manual frequent resets is equal to 60% and 30% respectively.

The concept of volatile spike-detection via memristive devices was initially validated in Fig. 4 using a memristive device of dimensions 60 μm $\times$ 60 μm. The robustness of the devices for the same dimensions was also evaluated. A neural recording with significantly different spiking pattern in comparison to Fig. 4(a) was used to bias a memristive device. TPR and FPR in this case were equal to 65% and 0%, respectively. Besides, similar results from fifteen different devices are tabulated in Table I, where the highest TPR and FPR obtained is equal to 88.4% and 13% respectively.

C. Nanoscale Memristive Devices as Volatile Spike-Detectors

Memristive device technologies offer huge advantages in terms of scalability and can be accommodated in Back-End-Of-Line (BEOL) of CMOS technologies, thus greatly benefiting future implantable neuroprosthetic platforms. We further support the presented approach with downscaled memristive devices of 200 nm $\times$ 200 nm dimensions. The devices were prototyped using the fabrication procedure described in the methods section (see Section II-A).

TABLE I
ROBUSTNESS OF TiO_x MEMRISTIVE DEVICES

Device Dimensions	Vth-(V)	Gain	Offset	VMS	TMS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)
μm^2											
60 × 60	−1.4	3.2	0	67	78	58	9	166	20	74.35	5.14
60 × 60	−1.4	2.6	0	39	20	13	26	207	7	65	11.15
60 × 60	−1.17	2.4	−0.6	22	20	11	11	222	9	55	4.7
60 × 60	−1.7	2.4	−0.4	15	20	13	2	231	7	65	0.85
60 × 60	−1.1	2	0	15	20	13	2	231	7	65	0.85
60 × 60	−0.7	2.2	0	74	78	47	27	148	31	60.25	15.4
60 × 60	−0.7	2.2	−0.2	52	78	42	10	165	36	53.8	5.71
60 × 60	−1.34	2.2	0	12	20	12	0	233	8	60	0
60 × 60	−1	4.4	0	56	78	25	31	144	53	32	17.7
60 × 60	−1	4.4	0	57	78	40	17	158	38	51.3	9.7
60 × 60	−1	4.8	0	76	78	53	23	152	25	70	14
60 × 60	−1.2	4.8	−0.2	102	78	69	33	142	9	88.46	18.8
60 × 60	−1.2	2.9	0	62	78	59	3	172	19	75.6	1.71
60 × 60	−1.2	2.6	0	63	78	60	3	172	18	77	1.71
60 × 60	−2.41	4.4	0	13	20	13	0	233	7	65	0
60 × 60	−0.7	2.6	−0.4	92	78	69	23	152	9	88.4	13
nm^2											
200 × 200	−1.1	2.8	−0.4	54	78	31	24	151	47	40	13.7
200 × 200	−1.25	2.6	−0.6	46	78	35	11	164	43	44.8	6.28
200 × 200	−1.3	2.6	−0.6	30	78	21	9	166	57	27	5
200 × 200	−1.63	2.8	−0.4	10	78	10	0	175	68	12.8	0
200 × 200	−1.25	2.6	−0.6	42	78	32	10	165	46	41	5.74
200 × 200	−1.25	2.6	−0.6	78	78	54	24	151	24	70	13.7
200 × 200	−1.3	Max = 0	Min = −2.2	44	78	22	22	153	56	28.2	12.5
200 × 200	−1.13	2.6	−0.6	47	78	27	20	155	51	34.6	11.4
200 × 200	−1.2	3	−0.6	47	78	27	20	155	51	34.6	11.4
200 × 200	−1.3	3	−0.6	42	78	27	15	160	51	34.6	8.5
200 × 200	−1.3	2.6	−0.6	74	78	43	31	144	35	55.12	17.71
200 × 200	−1.4	2.6	−0.6	72	78	37	35	140	41	47.43	20
200 × 200	−1.45	2.6	−0.6	43	78	27	16	159	51	34.61	9.14

For this experiment, devices with different dimensions i.e. 60 μm × 60 μm and 200 nm × 200 nm and different neural recordings with significantly different spiking pattern were used. For the pre-processing of the neural recording, the operational parameters that is G and V_{off} were varied. The quantification parameters are indicated in the table with the estimated rate of true positives and false positives. VMS: Volatile Spike Detection platform.

In this case, we employed the same neural recording used previously, as shown in Fig. 4(a) and similarly benchmarked the obtained results against the TMS. For this case the neural recording was amplified using G and V_{off} values of 2.6 and −0.6 respectively, as illustrated in Fig. 5(a). The transient response of the target device resistance is illustrated in Fig. 5(b) along with the corresponding neural recording time series. Compared to the previous case, this device was operated at a higher resistive state region of 1 MΩ to 1.5 MΩ. The spikes detected from both this and the TMS system was calculated to be 78, as shown in Fig. 5(c) and (d) respectively. The response of the target device can be more closely scrutinized in Fig. 5(e) and (f), which illustrates the neural recording and resistive state response of the DUT during the 4–5 sec window in Fig. 5(a) and (b) respectively. Marker "x" is used to indicate the resistive state measurements, following the standard schematic as described in the methods section. Between each pair of measurements the system is blind to the behaviour of the devices, however the bin size is a user-defined design parameter. Sampling rate (detection accuracy) and power consumption can be traded against each other.

The memristive devices undergo a resistive drop in response to the supra-threshold events. For instance, in the first bin shown in Fig. 3(f), the state of the device drops from approximately 1.3 to 1.15 MΩ. The asterisk * in Fig. 5(g), (h) further confirms concurrence between our system and the TMS that is true for the majority of instances. More specifically, the two systems agree for 13 over 17 instances detected, as shown in Fig. 5(g). The symbol "ϕ" indicates an instance of mismatch between the two systems. Interestingly, at this instance our system detects a neural event that closely resembles a spike, while the TMS fails. On further careful examination of the neural recording it was observed that the TMS also fails to detect apparent spiking events at approximately 1.1 s, 1.6 s, 1.9 s, 2.2 s, 2.5 s, 2.7 s, 3.3 s, 3.6 s, 3.9 s, some of which are detected by our system. On the other side, in Fig. 5(g) our system fails to detect spikes occurring at 4.55 s. From these observations it can be safely concluded that, although we assume the TMS to be a perfect spike-detector for our experiments, in practical operation this is not the case. Benchmarking of the detected spikes revealed a TPR and FPR of 70% and 13.7% respectively. Spike detection results carried out by thirteen different nano-devices with pre-processed neural recordings are reported in Table I. Notably, the table also indicates the performance of five different memristive devices on the same input neural data.

A Receiver Operating Characteristic sensitivity curve, that is defined as the rate of TP versus FP, is illustrated in Fig. 6(a) for devices of different dimensions. The details for the quantification parameters are presented in the Table II. It can be seen that as the dimensions of the devices are reduced from 60 μm × 60 μm down to 200 nm × 200 nm the detection accuracy of
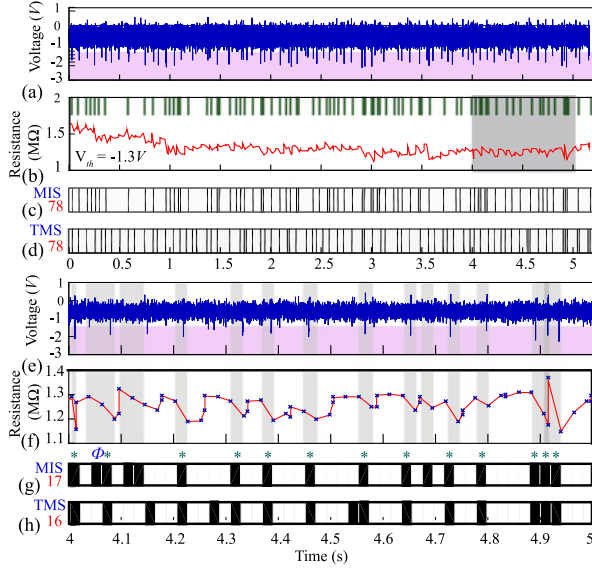
Fig. 5. Scalability of the TiOx devices to the nanoscale dimensions and operation of nano-devices as volatile spike detectors. (a) The employed neural recording pre-processed using G and V_{off} value of 2.6 and −0.6 respectively. The pink band indicates the inherent threshold voltage of the DUT (V_{th}). (b) Resistive state changes in the target DUT in response to the neural data in (a). The active core area of the nano-devices used for this experiment is 200 nm × 200 nm. The threshold voltage of the DUT is −1.3 V. (c), (d) Total number of spikes detected by our system and the TMS system is equal to 78 for both the systems. The black bins indicate the spike positions. (e) and (f) are close-ups for the neural recording and the resistive state evolution shaded grey in (a) and (b) respectively. 'X' cross mark in red indicates the positions where the resistive state measurements are taken. Time intervals where an event is detected are represented in grey. (g) spikes detected by our system and (h) the TMS system. The asterisk "*" indicates the positions where the two systems agree and "ϕ" symbol discusses one specific case of discrepancy between the two systems.

the system is reduced. One way for optimising the performance of the downscaled devices is to tune the gain and offset parameters to an optimum level. We thus proceeded by performing additional tests with three different gain settings 2.2, 2.4 and 2.6 and a fixed offset setting at −0.6 V, as illustrated in Fig. 6(b). For every round of gain, the experiment was repeated for five times indicating the performance of the same memristive device over multiple times of operation and the corresponding details of the quantification parameters are illustrated in Table II. The asterisk symbol, '*' represents the average of the quantification parameters for every round of gain. Thus, an improvement in the gain resulted into an increase in rate of TP from 23.1% to 46.7%, and these numbers can be plausibly further improved in the future by estimating the optimum bin size for practical implementations of our volatile spike-detector. We finally note that signal artefacts related with the employed recording methodology will overall contribute to an increase in detecting false positive events. Our approach can be tailored via gain adjustment to be immune to power-line noise and EM interference, nonetheless, large-amplitude or stimulation artefacts would require additional pre-filtering that is commonly used in other spike-detection implementations.

IV. DISCUSSION AND CONCLUSION

A major prerequisite in the implementation of fully implantable neural interfaces is to develop electronic platforms
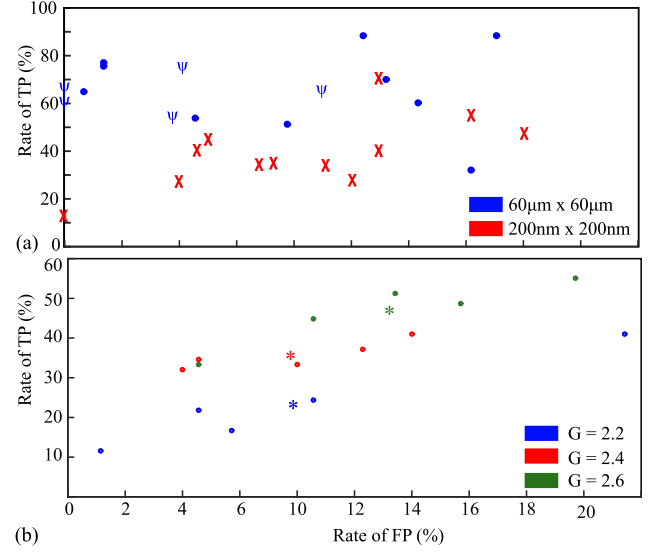

Fig. 6. Comparison of Receiver Operating Characteristics (ROC) that is rate of true positives versus rate of false positives for memristors of micro and nano dimensions. (a) The blue and red colour indicates the ROC curves for 60 μm × 60 μm and 200 nm × 200 nm devices respectively. Symbol psi "ψ" and circle in blue colour indicates different neural recordings with different spiking pattern used for the DUT. (b) Optimisation of spike detection capability for one of the 200 nm × 200 nm device. Blue, red and green colour indicates the three different gain parameters i.e. 2.2, 2.4 and 2.6 respectively chosen for the experiment where the offset was kept constant at −0.6. For every gain the experiment was repeated five times. Asterisk symbol "*" indicates the average of the quantification parameters for each round of gain.

TABLE II
OPTIMISATION OF 200 NM × 200 NM TiO_x MEMRISTIVE DEVICES

Gain	VMS	TMS	TP	FP	TN	FN	Rate of TP (%)	Rate of FP (%)
2.2	66	78	32	34	141	46	41.02	19.42
2.2	25	78	17	8	167	61	21.79	4.57
2.2	23	78	13	10	165	65	16.67	5.71
2.2	34	78	19	15	160	59	24.35	8.57
2.2	11	78	9	2	173	69	11.53	1.14
							23.07	**7.88**
2.4	32	78	25	7	168	53	32.05	4
2.4	47	78	29	18	157	49	37.18	10.28
2.4	40	78	26	14	161	52	33.34	8
2.4	53	78	32	21	154	46	41.02	12
2.4	35	78	27	8	167	51	34.61	4.57
							35.64	**7.77**
2.6	62	78	38	24	151	40	48.71	13.71
2.6	50	78	35	15	160	43	44.87	8.57
2.6	60	78	40	20	155	38	51.28	11.42
2.6	34	78	26	8	167	52	33.34	4.57
2.6	74	78	43	31	144	35	55.128	17.71
							46.66	**11.19**

Gain and offset parameters for one of the 200 nm × 200 nm device was optimised. Three different values of gain i.e. 2.2, 2.4 and 2.6 were used with constant offset values i.e. −0.6. For each round of gain the experiment was repeated five times and the quantification parameters are indicated in this table.

capable of executing low-power, on-site signal processing of the acquired neural activity. In this work, we experimentally demonstrated that metal-oxide memristive nanodevices can be operated in their volatile region as scalable spike-detection elements. The optimised performance is primarily due to the intrinsic resetting capability of such elements that eliminates the need for manually resetting the device to avoid its saturation. The major advantage of the presented work lies in the

prospects for dramatically reducing the power consumption per channel of implantable neural spike-detection circuits. Consider the example illustrated in [34]: for a neural recording consisting of approximately 63 k data-points the target device is manually reset eleven times using a pulse of positive polarity of 100 μs pulse width. Realistically assuming +3 V as the operating voltage and 10 kΩ as the resistive state of the device, the amount of power dissipated including the reset operation by this system will be approximately equal to 3 mW per channel. The reset operation consumes about 250 nJ of energy and 11 such resets for one neural recording would consume 2.7 μJ of energy which will be conserved in the volatile region.

Additional power savings also stem from the fact that the volatile regime exploited in our case occurs at relatively high resistive state regions (e.g., consider the electrical characterisation results of nano-devices in Fig. 2(a), (b). Assuming the operating resistive state and pulse width of the DUT to be 1 MΩ and 1 μs respectively, the amount of energy dissipated per channel can be estimated using the standard batch processing schematic. The read operation with standard read out voltage of +0.2 V will be equal to 0.2 pJ (0.04 pJ multiplied by 5 reads per batch) and the write operation at 3 V will cost 9 nJ (9pJ multiplied by 1000 samples per batch). The average power dissipated per channel at 12.2 kHz as the sampling frequency can thus be estimated to be approximately 100 nW. Voltage-time tradeoff [45] can further assist in reducing the power dissipation by one order of magnitude, for instance, operating the same with 100 ns pulse widths will further reduce the power dissipated to 10 nW per channel per device. Our measured results are already significantly lower than current state-of-the-art spike detectors projected at approximately 700 nW [23]. Naturally, a full system application would include other power overheads required by the memristor read-out and biasing circuitry, which are not considered in the present work as such circuitry will be cited in the periphery and would be shared (multiplexed) by each memristive device.

One can also argue that the usability of the memristive devices could be limited due to the recurrently observed device-to-device variation. Nonetheless, our approach exploits normalised changes in each bin instead of absolute changes, therefore minimising any performance compromises due to device variability. Besides, the introduced volatility characterisation module leads to an automatic en-masse characterisation of memristive devices and determines the safe region for operation of devices in the volatile region. Clearly, if higher and more invasive voltages were applied, the devices may switch to the non-volatile region and change their baseline operating region. The focal point for building on the present work is to push the scaling limits for the employed devices towards deep-submicron arrays [46], along with optimised spike-detection capability. We note that the use of alternative or engineered materials as the devices active cores could potentially allow achieving smaller range of variation in threshold voltages, tuning the volatile characteristics of devices and in turn the self-reset achieved by the spike-detector. On the other hand, these parameters will also crucially depend upon the specific application under study, accounting for the sensitivity required for a particular application.

We further note that as the resistive state modulation encountered in our prototype devices is related to the amplitude and polarity of the input neural signal, the richness of the signal is in principle preserved and can possibly open a new avenue for scalable and power efficient on-node spike-sorting [47], [48]; a prerequisite for fulfilling the electroceuticals [49] and more broadly the bioelectronics vision. In summary, taking the cue from how biological synapses compress spiking information in post-synaptic conductance changes, we have demonstrated a novel concept for neural-spike detection and encoding using intrinsic volatile behaviour of nanoscale metal-oxide memristive devices. Our results prove that single nanoscale volatile devices are capable of identifying significant spiking activity in the input neural waveform in a highly power efficient manner, thus paving the way towards advanced neuroprostheses or applications such as bioelectronics medicines where the power dissipation remains as the major challenge.

REFERENCES

[1] G. Buzsáki, "Large-scale recording of neuronal ensembles," *Nature Neurosci.*, vol. 7, no. 5, pp. 446–451, May 2004.

[2] T. J. Blanche *et al.*, "Polytrodes: High-density silicon electrode arrays for large-scale multiunit recording," *J. Neural Eng.*, vol. 93, no. 5, pp. 2987–3000, 2005.

[3] F. Franke *et al.*, "High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments: an emerging technology to study neural plasticity," *Front Neural Circuits*, vol. 6, pp. 1–7, 2012.

[4] R. Lambacher *et al.*, "Identifying firing mammalian neurons in networks with high-resolution multi-transistor array (MTA)," *Appl. Phys. A*, vol. 102, no. 1, pp. 1–11, Sep. 2010.

[5] S. Frey *et al.*, "Microelectronic system for high-resolution mapping of extracellular electric fields Appl. to brain slices." *Biosensors Bioelectronics*, vol. 24, no. 7, pp. 2191–2198, Mar. 2009.

[6] Z. Berényi *et al.*, "Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals." *J. Neurophysiology*, vol. 111, no. 5, pp. 1132–11349, Mar. 2014.

[7] X. Navarro *et al.*, "A critical review of interfaces with the peripheral nervous system for the control of neuroprotheses and hybrid bionic systems," *J. Peripheral Nervous Syst.*, vol. 10, pp. 229–258, 2005.

[8] C. M. Lopez *et al.*, "A 966-electrode neural probe with 384 configurable channels in 0.13 um SOI CMOS," in *Proc. 2016 IEEE Int. Solid-State Circuits Conf.*, 2016, pp. 392–393.

[9] B. Eversmann, A. Lambacher, T. Gerling, A. Kunze, P. Fromherz, and R. Thewes, "A neural tissue interfacing chip for in-vitro applications with 32 k recording/stimulation channels on an active area of 2.6 mm^2," in *Proc. ESSCIRC*, Sep. 2011, pp. 211–214.

[10] B. Gosselin, "Recent advances in neural recording microsystems," *Sensors*, vol. 11, no. 5, pp. 4572–4597, 2011.

[11] A. Hierlemann *et al.*, "Growing cells atop microelectronic Chips: Interfacing electrogenic cells in vitro with CMOS-based microelectrode arrays," *Proc. IEEE*, vol. 99, no. 2, pp. 252–284, Feb. 2011.

[12] H. G. Rey *et al.*, "Past, present and future of spike sorting techniques," *Brain Res. Bull.*, vol. 119, pp. 106–117, Apr. 2015.

[13] M. S. Lewicki, "A review of methods for spike sorting: The detection and classification of neural action potentials classification of neural action potentials," *Netw.: Comput. Neural Syst.*, vol. 9, no. 4, pp. R53–R78, 1998.

[14] E. M. Schmidt, "Computer seperation of multi-unit neuroelectric data: A review," *J. Neurosci. Methods*, vol. 12, pp. 95–111, 1984.

[15] R. Q. Quiroga, "Spike sorting," *Current Biol. : CB*, vol. 22, no. 2, pp. R45–6, Jan. 2012.

[16] M. A. L. Nicolelis *et al.*, "Principles of neural ensemble physiology underlying the operation of brain-machine interfaces," *Nature Rev. Neurosci.*, vol. 10, no. 7, pp. 530–540, Jul. 2009.

[17] N. G. Hatsopoulos *et al.*, "The science of neural interface systems," *Annu. Rev. Neurosci.*, vol. 32, pp. 249–266, 2009.

[18] L. F. Nicolas-Alonso *et al.*, "Brain computer interfaces, A review," *Sensors (Basel, Switzerland)*, vol. 12, no. 2, pp. 1211–1279, Jan. 2012.

[19] B. K. D. Wise *et al.*, "Microelectrodes, microelectronics, and implantable neural microsystems," *Proc. IEEE*, vol. 96, no. 7, pp. 1184–1202, Jul. 2008.

[20] G. Baranauskas, "What limits the performance of current invasive brain machine interfaces?" *Frontiers Syst. Neurosci.*, vol. 8, Jan. 2014, Art. no. 68.

[21] K. P. Stevenson and I. HKording, "How advances in neural recording affect data analysis." *Nature Neurosci.*, vol. 14, no. 2, pp. 139–142, Feb. 2011.

[22] C. Pedreira *et al.*, "How many neurons can we see with current spike sorting algorithms?" *J. Neurosci. Methods*, vol. 211, pp. 58–65, 2012.

[23] S. E. Paraskevopoulou *et al.*, "A sub-1uW neural spike-peak detection and spike-count rate encoding circuit," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Nov. 2011, pp. 29–32.

[24] A. Eftekhar *et al.*, "Towards a next generation neural interface: Optimizing power, bandwidth and data quality," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2010, pp. 122–125.

[25] I. Gupta *et al.*, "Real-time encoding and compression of neuronal spikes by metal-oxide memristors," *Nature Commun.*, vol. 7, pp. 1–16, 2016.

[26] L. Chua *et al.*, "Memristive devices and systems," *Proc. IEEE*, vol. 64, no. 2, pp. 209–223, Feb. 1976.

[27] T. Prodromakis *et al.*, "Two centuries of memristors," *Nature Mater.*, vol. 11, no. 6, pp. 478–481, Jun. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22614504

[28] J. J. Yang *et al.*, "Memristive devices for computing." *Nature Nanotechnol.*, vol. 8, no. 1, pp. 13–24, Jan. 2013.

[29] D. Strukov *et al.*, "Resistive switching phenomena in thin films: Materials, devices, and applications," *MRS Bull.*, vol. 37, no. 2, pp. 108–114, Feb. 2012.

[30] R. Zeitler *et al.*, "Extracellular voltage noise probes the interface between retina and silicon chip," *Appl. Phys. Lett.*, vol. 99, no. 26, 2011, Art. no. 263702.

[31] G. Zeck *et al.*, "Axonal transmission in the retina introduces a small dispersion of relative timing in the ganglion cell population response." *PloS One*, vol. 6, no. 6, Jan. 2011, Art. no. e20810.

[32] C. Eversmann *et al.*, "A 128 × 128 CMOS biosensor array for extracellular recording of neural activity," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2306–2317, Dec. 2003.

[33] I. Gupta *et al.*, "Practical operation considerations for memristive integrating sensors," *Proc. Int. Symp. Circuits Syst.*, 2016, pp. 2322–2325.

[34] G. Isha *et al.*, "Improving detection accuracy of memristor-based biosignal sensing platform," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 203–211, Feb. 2016.

[35] D. B. Strukov *et al.*, "Exponential ionic drift: Fast switching and low volatility of thin-film memristors," *Appl. Phys. A: Mater. Scie. .ing*, vol. 94, no. 3, pp. 515–519, 2009.

[36] R. Berdan *et al.*, "Emulating short-term synaptic dynamics with memristive devices," *Sci. Rep.*, vol. 6, 2016, Art. no. 18639.

[37] I. Valov *et al.*, "Nanobatteries in redox-based resistive switches require extension of memristor theory," *Nature Commun.*, vol. 4, 2013, Art. no. 1771.

[38] S. Cortese *et al.*, "On the origin of resistive switching volatility in Ni/TiO2/Ni stacks," *J. Appl. Phys.*, vol. 120, no. 6, 2016, Art. no. 065104.

[39] A. Wedig *et al.*, "Nanoscale cation motion in TaOx, HfOx and TiOx memristive systems," *Nature Nanotechnol.*, vol. 11, no. 1, pp. 67–74, 2015.

[40] H. Markram *et al.*, "Redistribution of synaptic efficacy between neocortical pyramidal neurons," *Nature*, vol. 382, pp. 807–810, 1996.

[41] A. Serb *et al.*, "Live demo: A versatile, low-cost platform for testing large ReRAM cross-bar arrays," *Proc. IEEE Int. Symp. Circuits Syst.*, p. 441, 2014.

[42] A. Serb *et al.*, "An RRAM biasing parameter optimizer," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3685–3691, Nov. 2015.

[43] I. Gupta *et al.*, "Volatility characterisation for RRAM devices," *IEEE Electron Device Lett.*, vol. 38, no. 1, pp. 28–31, Jan. 2017.

[44] R. A. Normann, "Technology insight: future neuroprosthetic therapies for disorders of the nervous system," *Nature Clinical Practice Neurol.*, vol. 3, no. 8, pp. 444–452, 2007.

[45] J. Xing *et al.*, "An FPGA-based instrument for en-masse RRAM characterisation with NS pulsing resolution," *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 63, no. 6, pp. 818–826, Jun. 2016.

[46] A. Khiat *et al.*, "High density crossbar arrays with sub-15 nm single cells via liftoff process only," *Sci. Rep.*, vol. 6, 2016, Art. no. 32614.

[47] I. Gupta *et al.*, "Towards a memristor-based spike-sorting platform," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2016, pp. 408–411.

[48] G. Isha, S. Alexantrou, K. Ali, T. Maria, and P. Themistoklis, "Spike sorting using non-volatile metal-oxide memristors," 2017, arXiv:1707.08772.

[49] K. Famm, "A jump-start for electroceuticals," *Nature*, vol. 496, no. 7444, pp. 159–161, 2013.

Isha Gupta is currently working toward the Ph.D. degree with the Department of Electronics and Computer Science, University of Southampton, Southampton, U.K. Her research interests include bio-signal processing with RRAM devices, device electrical characterisation, and process development.

Alexantrou Serb is a Research Fellow with the Department of Electronics and Computer Science, University of Southampton, Southampton, U.K. His research interests include instrumentation, algorithms and applications for RRAM testing, and neuro-inspired engineering.

Ali Khiat is a Senior Experimental Officer with Southampton Nanofabrication Centre, University of Southampton, Southampton, U.K. His current main research interests include micro/nanofabrication, optimisation, metrology, and characterization of memristors and memristive devices.

Stefano Vassanelli received the Medical degree from the University of Padova, Padova, Italy, in 1992, where he also received the Ph.D. degree in molecular and cellular biology and pathology in 1999. He is currently a Professor of physiology and teaching both for the Engineering and Medical faculties. He has been working on the development of neurotechnologies for recording, stimulation, and processing of signals generated by neuronal networks.

Ralf Zeitler received the Diploma degree in physics from the Technical University of Munich, Munich, Germany, in 2004. In 2009, he received the Doctor's degree in physics from Max Planck Institute for Biochemistry, Planegg, Germany, for the research of the noise processes in the cell-semiconductor interface. From 2012 to 2015, he was a Research Group Leader with Max Planck Institute for Intelligent Systems working on a CMOS-based sensor platform for living cells, which led to spin-off Venneos GmbH, Stuttgart, Germany, in 2014, where he currently is the Technical Managing Director.

Themis Prodromakis (SM'08) is a Professor of nanotechnology and EPSRC and Royal Society Industry Fellow affiliated with the Southampton Nanofabrication Centre, University of Southampton, Southampton, U.K. He previously held a Corrigan Fellowship in Nanoscale Technology and Science within the Centre for Bio-Inspired Technology, Imperial College and a Lindemann Trust Visiting Fellowship in EECS UC Berkeley. He is a Fellow of the IET and a Fellow of the Institute of Physics. His background is in electron devices and nanofabrication techniques, with his research being focused on bio-inspired devices for advanced computing architectures and biomedical applications.