OPEN

A caloritronics-based Mott neuristor

Javier del Valle[1,2*], Pavel Salev[1], Yoav Kalcheim[1] & Ivan K. Schuller[1]

Machine learning imitates the basic features of biological neural networks at a software level. A strong effort is currently being made to mimic neurons and synapses with hardware components, an approach known as neuromorphic computing. While recent advances in resistive switching have provided a path to emulate synapses at the 10 nm scale, a scalable neuron analogue is yet to be found. Here, we show how heat transfer can be utilized to mimic neuron functionalities in Mott nanodevices. We use the Joule heating created by current spikes to trigger the insulator-to-metal transition in a biased VO_2 nanogap. We show that thermal dynamics allow the implementation of the basic neuron functionalities: activity, leaky integrate-and-fire, volatility and rate coding. This approach could enable neuromorphic hardware to take full advantage of the rapid advances in memristive synapses, allowing for much denser and complex neural networks.

Machine learning has experienced an unprecedented growth in recent years, often referred to as an "artificial intelligence revolution"[1,2]. Its fundamental approach is inspired by biological systems: using neural networks to classify large amounts of data into sorting categories. Classic examples are speech and image recognition[1,2]. Neural networks are composed of two basic elements: neurons and synapses. Current machine learning schemes implement these elements at a software level: neurons and synapses are simulated on standard computers based on a von Neumann architecture[1,2]. This approach is inefficient in terms of computation speed and energy consumption, motivating a search for hardware-based systems that imitate the brain. This idea was initially proposed more than fifty year ago[3-5], and attained widespread popularity with the works of Carver Mead[6]. Since then, CMOS-based circuitry has been successfully used to realize neuromorphic systems, allowing to build tuneable and efficient neural networks[7-9]. Unfortunately, CMOS-based components rely on combinations of multiple transistors and capacitors that make them complex and large[9]. This limits circuit scalability and, hence, poses a limitation to achieve dense neural networks which could eventually rival the brain.

A solution to this problem might be found in "neuromorphic materials", whose intrinsic properties mimic those of neurons and synapses[10,11]. Resistive switching (RS), a phenomenon in which an applied electric field modifies the resistance of a material[12-14], offers a unique opportunity to achieve this goal. RS can be volatile[15-21] or non-volatile[22,23], which can be used to emulate neuron or synapse behaviours, respectively. Multiple groups have used RS to achieve synaptic functionalities[24,25], and memristor crossbar arrays have already been used to perform pattern recognition[26-28]. These synapse realizations, however, still rely on traditional electronics to play the role of neurons (neuristors). This approach does not take full advantage of the scalability and simplicity offered by memristive synapses, and motivates the search for a simpler and more scalable neuristor.

A neuristor must feature the most basic functionalities of real neurons[29]: i) leaky integrate-and-fire, ii) activity (outputting a current), iii) volatility (resetting after a firing), and iv) rate coding of the external stimuli. Performing leaky integrate-and-fire is one of the key functionalities of a neuristor. It must sum all the input stimuli coming from previous neurons and fire an output spike when the excitation is above a certain threshold[29]. In the case of biological neurons, the cell membrane acts as a capacitor that integrates incoming ionic currents. The firing mechanism, based on voltage gated sodium and potassium channels, is activated once the membrane potential exceeds a certain threshold. CMOS neuristors use a similar approach (Fig. 1a): a capacitor plays the role of cell membrane by integrating current from incoming pulses[9], while the CMOS circuitry produces the firing events. Pickett et al.[30], Ignatov et al.[31] and Yi et al.[32] also use capacitive integration, but in their case volatile RS in a Mott insulator is utilized in the firing stage. While this parallelism with biological systems is appealing[33], the use of capacitors to store the internal state of the neuron limits the circuit scalability. In order to avoid malfunctioning, their capacitance must be much larger than the parasitic capacitance of the electrode lines. Integration capacitors

a Charge integration **b** Thermal integration

$$C_{Mem} \cdot \frac{\partial V}{\partial t} = I_{Input} - \frac{V}{R_L} \qquad C_T \cdot \frac{\partial T'}{\partial t} = Q_{Input} - \frac{T'}{R_T}$$
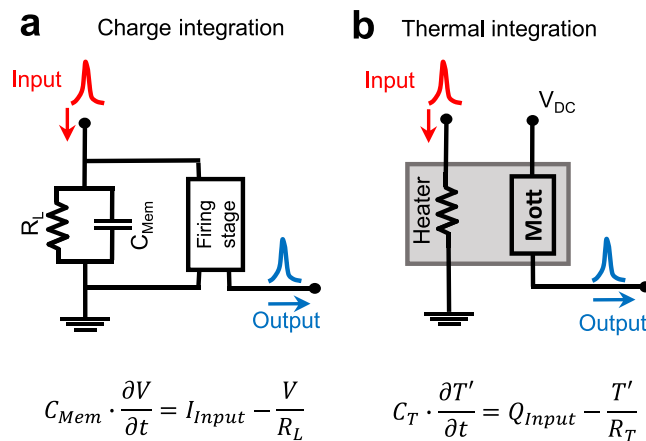
Figure 1. Charge vs thermal signal integration. (**a**) Circuit representation of the leaky integrate-and-fire neuron model. Its dynamics are described by the equation shown below the circuit. A capacitor C_{Mem} represents the neuron membrane capacitance, and accumulates charge from input current pulses I_{Input} (red). The leaking resistor R_L represents the leaky term of the equation. If the voltage V across C_{Mem} rises above a threshold, the firing stage will produce an output spike (blue). (**b**) Schematic representation of a thermal transfer-based neuristor. Incoming current pulses (red) dissipate power (Q_{Input}) in a heating resistance. A Mott nanodevice is kept under a DC voltage V_{DC}, and will create a current spike (blue) if a threshold temperature is exceeded. The equation below describes the thermal dynamics of the system, where C_T is the thermal capacitance, R_T the thermal resistance and $T' = T - T_{Eq}$; being T the local temperature and T_{Eq} the equilibrium (substrate) temperature.

cover a large area in current CMOS neurons, leading to typical sizes in the order of 10–100 μm[9]. Capacitor downscaling is one of the most challenging issues in other technologies such as DRAM, where the industry has dedicated intense effort towards developing complex 3D capacitive structures to circumvent this problem[34].

Wang et al.[35], Tuma et al.[36] and Stoliar et al.[37] successfully implement integrate and fire dynamics without the use of capacitors by utilizing diffusive memristors, phase-change materials and Mott insulators, respectively. However, these systems are not active as they do not generate a current. Moreover, they are not volatile i.e. do not reset automatically, a characteristic needed for spiking dynamics. This limits their autonomy and their practical implementation as standalone neurons. Yajima et al.[38] introduced a dedicated circuit to reset a Mott integration device. Although their neuristor is capable of performing all basic neuron tasks, the use of multiple operational amplifiers makes this implementation rather complex. A fully autonomous and scalable neuristor is yet to be found.

Instead of electrical currents, we propose using heat flow to perform computing tasks, an approach known as caloritronics. Temperature substitutes charge as the integrating variable, as depicted in Fig. 1b. Current spikes coming from previous neurons induce Joule heating while passing through a resistive element (heater), increasing local temperature with every spike. The heater is thermally coupled to a firing element that is very sensitive to temperature changes and fires once a threshold temperature is exceeded. In this work, we use VO$_2$, a well-known correlated oxide with a sharp insulator-to-metal transition (IMT) around 340 K[39] (Supplementary Figure S1a), as the firing element. We realize leaky integrate-and-fire using the thermal dynamics, which are governed by similar equations to those describing the charge dynamics of a leaky capacitor (see Fig. 1a,b). Adopting local temperature as the internal state allows building simple-design neuristors that can be downsized to the nanoscale, as thermal dynamics equations preserve the same form independent of the system size.

We fabricated and tested a proof-of-concept neuristor that performs all basic neuronal functionalities. It consists of a VO$_2$ thin film on top of which two layers of electrodes are patterned (the detailed fabrication process can be found in methods section and Supplementary Figure S2). The first layer consists of two Ti/Au electrodes (running vertically in Fig. 2a) separated by a 50 nm gap. These electrodes are used to apply voltage to the VO$_2$, and provide the source of the neuristor's active output. If the voltage is high enough, a transition into the metallic phase can be electrically triggered[15,16,40] (Supplementary Figure S1b). Figure 2b shows the current as a function of time when a voltage pulse is applied to the gap. It illustrates the threshold nature of the voltage triggered IMT: the device becomes metallic once a threshold voltage (V_{Th}) is exceeded[40]. The second electrode layer is a Ti/Au nanowire (running horizontally in Fig. 2a) which acts as a heater. It is separated from the bottom electrodes by a 70-nm-thick Al$_2$O$_3$ layer which provides electrical insulation (resistance larger than 20 MΩ), but ensures thermal coupling between the nanowire heater and VO$_2$ gap.

Figure 2c shows V_{Th} of the VO$_2$ gap as a function of temperature. Two cases are shown: zero (black) and 12.5 mA (red) current flowing through the nanowire heater (horizontal electrode). In the second case, Joule heating locally increases the temperature of the gap reducing its V_{Th}. To work as a neuristor, the gap is kept under a DC bias just below its threshold voltage ($V_{DC} < V_{Th}$), as represented in Fig. 2c by a green dot. When a high enough current pulse I_{Input} is passed through the heater, it lowers V_{Th} below V_{DC}, and the gap turns metallic, generating an output current through the bottom electrodes (I_{Output}). This situation is presented in Fig. 2d, where a 30 ns current pulse is applied to the heater, triggering the IMT in the gap. We must emphasize that the output and
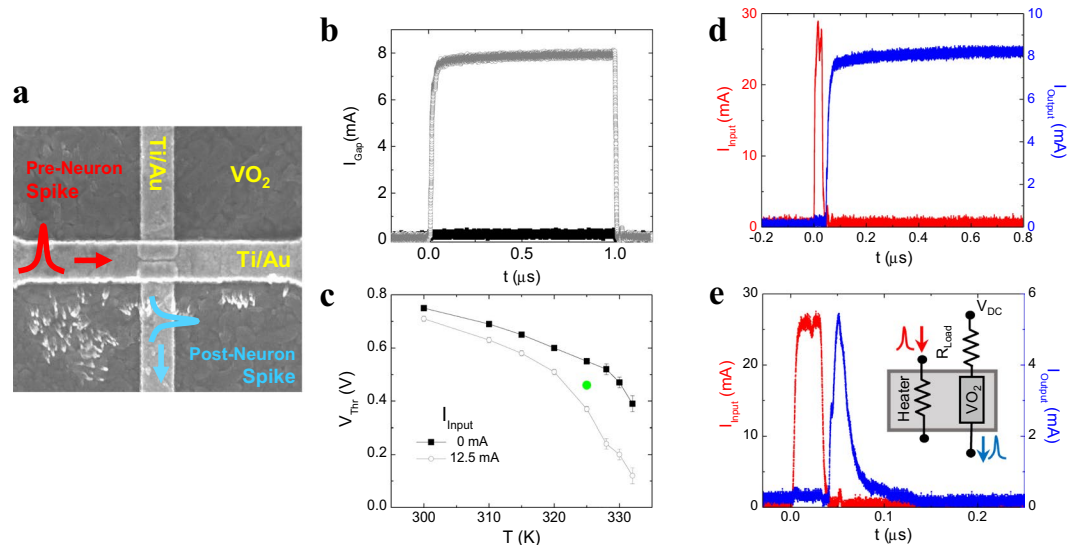
Figure 2. Experimental realization of the proposed neuristor. (**a**) SEM image of the device. Two layers of electrodes, electrically isolated by an Al_2O_3 spacer, are visible. Bottom layer consist of two Ti/Au electrodes, running vertically in the image. A small gap (~50 nm) is left in between. The upper electrode is kept at V_{DC} bias. The top electrode layer is a Ti/Au nanowire, running horizontally in the image. Input current pulses (red) are sent through that heating electrode, while the neuristor output (blue) is collected though the lower bottom electrode. (**b**) Current through a VO_2 gap vs time, when a 1 μs voltage pulse is applied. $T = 315$ K. Two different pulse amplitudes are shown, 0.87 V (black) and 0.92 V (grey). (**c**) Threshold voltage vs temperature in a gap, when two different DC currents are applied to the heater: 0 mA (black) and 12.5 mA (grey). The green dot is an example of the V_{DC} and temperature conditions for the device to behave as a neuristor. (**d**) Current vs time. Left axis (red curve) shows the input current through the heater. Right axis (blue curve) shows the output through the gap. Conditions were $V_{DC} = 0.85\ V_{Th}$ and $T = 325$ K. **e** Current vs time, with a load resistor in series with the gap. Left axis (red) shows the input current pulse. Right axis (blue) shows the output through the gap. $R_{Load} = 10$ kΩ, $V_{DC} = 0.88\ V_{Th}$ and $T = 325$ K. Inset: schematic circuit showing R_{Load} connected in series with the gap.

input electrodes are electrically isolated, and an output current is generated when the neuristor fires, making it an active element (as it must be powered to operate, and outputs power during operation). Our device releases energy when stimulated, a crucial property to avoid a re-amplification stage after each neural layer. This goes one step beyond previous RS-based neuristors[30–32], which become conducting after performing integrate-and-fire but do not create an output on their own. Furthermore, electrically decoupled input and output of our neuristor would allow building a multilayer neural network without using buffer circuits to prevent current backflow from post-synaptic to pre-synaptic neuristors.

Volatility is a necessary feature to implement spiking dynamics. As presented so far, our device would remain conductive once triggered. To reset the VO_2 gap after the firing event, we added a resistor (R_{Load}) in series with the gap (inset Fig. 2e). The role of this resistor is to lower the voltage across the gap once the VO_2 becomes metallic[41–43], which in turn reduces heat dissipation and decreases local temperature after the firing event. As a result, the VO_2 returns to its insulating state after firing (Fig. 2e) giving the desired effect: spike in – spike out.

Leaky integrate-and-fire (LIF) dynamics are governed by the characteristic thermal times of our device, given by its specific heat and thermal resistance to the substrate. Figure 3a shows the warm up times of the neuristor as a function of I_{Input}, that is, how long it takes the device to warm and fire once the current flows through the heater. Typical times are on the order of 10–100 ns. Using pulse widths and rates around that timescale allow us to implement LIF dynamics. Figure 3b shows the response of the neuristor (I_{Output}) when a train of current pulses is sent to the heater (I_{Input}). The first pulse does not raise the temperature enough to fire the device, but the cumulative effect of several pulses adds up to trigger the IMT after an integration period. The number of pulses necessary to produce a firing event depends on the pulse amplitude. Figure 3c shows the probability of the neuristor firing after a certain number of pulses are applied to the input. Several current amplitudes are shown, offering a clear visualization of the LIF dynamics. For high current pulses, the device is always triggered with just one pulse. For lower currents, the probability of it firing with just one pulse decreases, and more pulses are necessary to induce an output spike. The overall mean integration time is shifted to longer values as the current amplitude is decreased. When the input current is too low, heat leakage into the environment overcomes the dissipated power, and the device does not fire for any number of pulses.

Another basic feature of biological neurons is rate coding: the frequency at which a neuron spikes depends on the amplitude of its stimulus. Strong stimuli produce high frequency spiking, while weak stimuli yield slower patterns[29,44]. Our neuristor reproduces that feature, as shown in Fig. 4a. A constant current is passed through the heater, resulting in a repetitive spiking output. The frequency of the output increases with the input current (Fig. 4b). After the neuristor fires, both the temperature and the voltage across the gap drop, leaving the system in a refractory period until they increase back to their initial values.
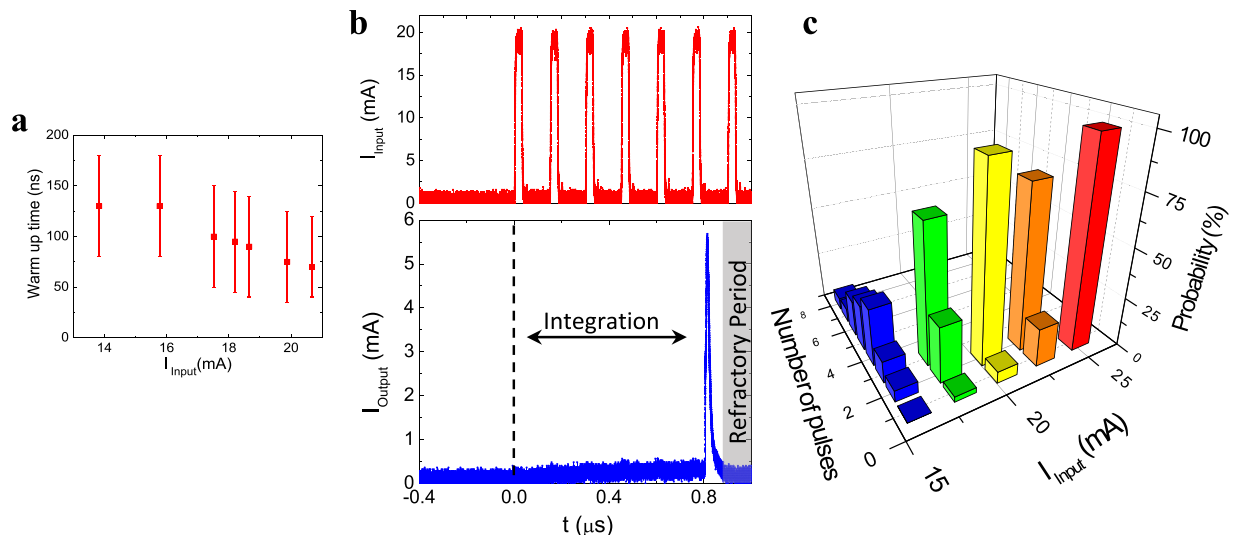
Figure 3. Leaky integrate and fire dynamics. (**a**) Warm up time as a function of the input current. $V_{DC} = 0.85$ V_{Th} and $T = 325$ K. (**b**) Current vs time when the neuristor is stimulated with a train of pulses. Upper panel shows the input current, consisting on 30 ns pulses with a 150 ns period. Bottom panel shows the output current through the gap. There is an integration time between the moment the pulses are applied and the moment the IMT is triggered. $R_{Load} = 10$ kΩ, $V_{DC} = 0.88$ V_{Th} and $T = 325$ K. No firing occurs after the last pulse because the neuron is in its refractory period (indicated with a grey shaded area). (**c**) Probability that the device will fire after a certain number of pulses. Pulses are 150 ns apart from each other. Several pulse amplitudes are shown. $V_{DC} = 0.91$ V_{Th} and $T = 325$ K.

The duration of the refractory period depends on the interplay between neuristor's thermal and electrical properties. The electrical charging time is determined by the RC constant, being R and C the resistance and capacitance of the system. While we do not use external capacitive elements, some intrinsic capacitance will always be present due to the experimental set up. The warm up time is given by $R_{th}C_{th}$, where R_{th} is the thermal resistance and C_{th} the thermal capacitance. Whether the refractory dynamics are dominated by thermal or electrical effects depends on the $RC/R_{th}C_{th}$ ratio. Considering the geometry and materials of our device, we estimate $R_{th}C_{th}$ to be around 10^{-8} s (See methods). Since R is in the 10^4 Ω range, electric parasitic capacitance is expected to be dominant for C above 10^{-12} F.

To gain a better understanding of the interplay between electrical and thermal properties, we performed lumped-element simulations of the neuristor operation. Figure 5a,b show the electrical equivalent circuit and a schematic of the thermal model used in our simulations (see methods section for more details). In the equivalent circuit, we explicitly include the parasitic capacitance associated with the device. The thermal model takes into account the heating provided by both the nanowire heater and the V_{dc} bias, and the heat loss into the environment. Figure 5c shows I_{Output} vs t when a dc I_{Input} is applied at $t = 0$, for a system with a relatively large electric capacitance $C = 10^{-10}$ F. Repetitive spiking, similar to the experimental result is observed. This suggests that in our particular devices, recovery dynamics during the refractory period are determined mainly by the parasitic capacitance. However, such capacitance is not necessary to produce spiking dynamics. Figure 5d shows I_{Output} vs t when a dc I_{Input} is applied at $t = 0$, for a system with no parasitic capacitance: spiking behaviour is still observed, although with clear differences in the shape and time separation between the individual spikes. The mechanism behind the spiking dynamics can be understood by considering the stability points of the system[45]. Figure 5e shows time derivative of the temperature, $\partial T/\partial t$, vs T for different I_{Input}. Sharp discontinuities in $\partial T/\partial t$ are present due to the IMT, resulting in a hysteresis curve. When there is no current through the heater, the system stabilizes at a certain temperature and does not oscillate. Adding an input shifts the curve in a way in which the hysteresis jumps discontinuously between $\partial T/\partial t > 0$ and $\partial T/\partial t < 0$, so $\partial T/\partial t$ is never equal to zero. This traps the system in a persistent oscillation state, purely due to thermal dynamics. In this way, the spiking behaviour of the VO$_2$ gap can be externally controlled with a heat current. Both electrically, $RC > R_{th}C_{th}$, and thermally dominated, $RC < R_{th}C_{th}$, systems produce the rate coding property, as shown in Fig. 5f, where the spiking rate is plotted as a function of I_{Input} for several values of C.

Many of the relevant parameters of the proposed neuristor depend on the particular device design, as well as on the intrinsic properties of the chosen materials. This gives plenty of room to explore and improve its functionalities. Different substrates, insulating spacers or geometric designs will strongly change the device properties. For instance, a smaller device and less thermally conductive substrate and contact pads would reduce heat leakage to the environment, allowing for the use of lower currents. For instance, using Ti instead of Au as metallic contact would drastically reduce heat loss into the pads. Similarly, using TiO$_2$ instead of Al$_2$O$_3$ as a substrate can decrease thermal conductance by a factor of 20. A simple calculation (see methods), shows that by changing the materials, thermal conductance and hence Joule heating can be reduced by more than two orders of magnitude.
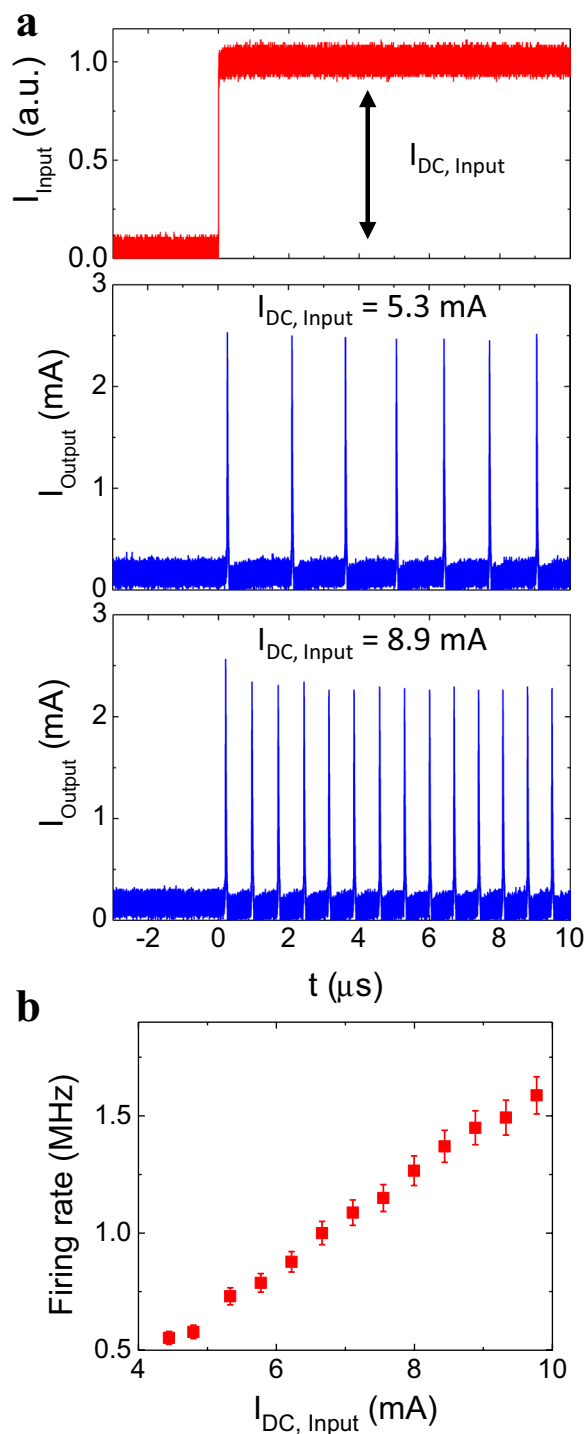
Figure 4. Rate coding. (**a**) Current vs time when a DC current is applied as input. Top panel shows the input current $I_{DC,Input}$ through the heater (red). Middle and lower panels show the output spiking pattern (blue) induced in the gap. The responses to two $I_{DC,Input}$ values are shown: 5.3 mA (middle panel) and 8.9 mA (lower panel). $R_{Load} = 10$ kΩ, $V_{DC} = 0.88$ V_{Th} and T = 325 K. (**b**) Firing rate of the output spiking pattern as a function of $I_{DC,Input}$.

Although the currents used in our proof-of-concept device are large, the short duration of each spike (~30 ns) yields an energy consumption of 3.10^{-9} J/spike. By reducing parasitic capacitance and optimizing materials and geometry, this number could be brought down to around 10^{-11} J/spike, which is comparable to the performance of biological neurons. Other design strategies, such as extremely localized Joule heating[46], could also be used to further decrease energy consumption. Regarding size, our device occupies less than 1 μm², reducing neuron area
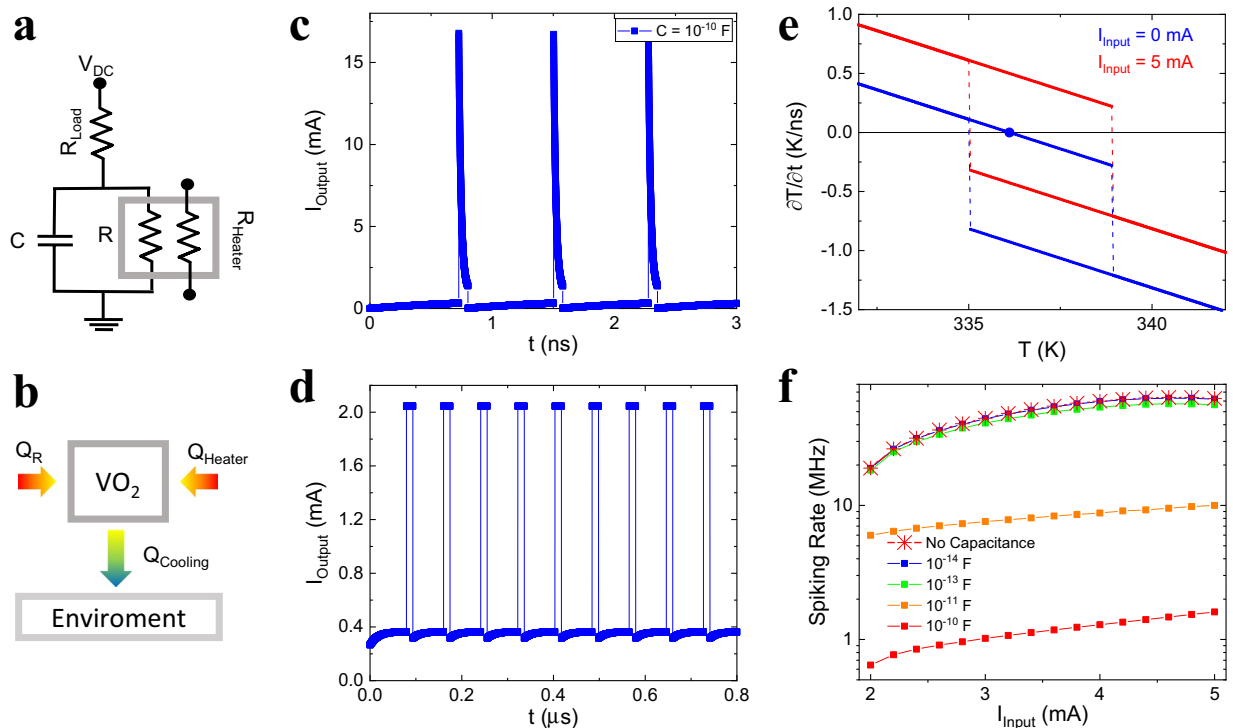
Figure 5. Simulation of the neuristor dynamics. (**a,b**) Schematic representation of the simulated electrical (**a**) and thermal (**b**) circuits. V_{DC} is the bias voltage. C is the parasitic capacitance, R is the resistance of the VO_2 gap, R_{Load} is the resistance of the load resistor and R_{Heater} is the heater resistance. Q_R and Q_{Heater} are the powers dissipated in the VO_2 gap and heater, respectively. $Q_{Cooling}$ is the power lost to the substrate and contact pads though thermal conduction. For a complete description of the model, see the methods section. (**c,d**) Output current vs time for $C = 10^{-10}$ F (**c**) and $C = 0$ (**d**). The neuristor shows spiking behaviour both with and without the series capacitor. (**e**) Thermal stability analysis derived from the model. $\partial T/\partial t$ vs T is showed for two different cases, $I_{Input} = 0$ mA and $I_{Input} = 5$ mA. An external input from the nanoheater is capable to drive the system into a permanent spiking cycle, with no stability points. (**f**) Spiking rate vs I_{Input} for five different capacitance values. Rate coding is observed in all cases.

by more than an order of magnitude compared to biological neurons, and almost four orders compared the most compact silicon neuron circuits[9].

Another attractive feature of our approach is the potential of signal amplification without needing further elements: since the input and output are electrically isolated, it is possible for the output current to be larger than the input current. This is of fundamental importance; real neural networks propagate signals across several neuron layers, and the signal must be amplified after each layer. Previous neuristor implementations are passive, and therefore the output is always smaller than the input. This makes the use of CMOS based electronics mandatory, partially defeating the purpose of building a purely oxide electronics. We must note that extensive device optimization must be done before this is experimentally possible in a caloritronics-based device. Nevertheless, our lumped-element simulations show that this is a feasible scenario if heat conductance between the heater and the gap is improved. In fact, the results presented earlier in Fig. 5d also demonstrate self-amplification: the device generates a ~2 mA output current out of a 1.3 mA input.

The use of a heat transfer-based device may also come with some drawbacks that must be considered. One of them is crosstalk between neurons which could limit device density. Let's consider two neuristors placed next to each other. With our current device dimensions, the distance between the two VO_2 gaps could be as low as 1 μm, comparable to the typical pitch of memristor crossbar arrays[26]. A simple estimation shows that firing one of the devices can locally rise temperature up to 5 K in the other one, enough to make it fire too. By optimizing device dimensions and materials this problem could be largely avoided, and such temperature increment could be limited to a few mK (see methods). Another potential drawback is that, due to the proximity to a phase transition, temperature must be precisely controlled when working with IMT-based neuristors. Although this could be hard to implement in very large circuits, it could also be a positive feature. The human brain operates close to criticality and it is only functional in a very narrow temperature range[47]. It has been argued that this critical behaviour is what allows to perform the complex cognitive task of an intelligent system[48]. In this sense, working at the edge of a phase transition might be an ideal platform to explore new and more complex phenomena in neuromorphic computing.

Caloritronics and resistive switching can be combined to create scalable and autonomous neuristors. We demonstrated four basic neural functionalities: activity, volatility, leaky integrate-and-fire dynamics and rate coding using simple devices that can be downscaled well below the μm scale. Combined with the fast advances in memristor technology, this could pave the way to develop dense neuromorphic hardware, allowing for deeper and

more complex neural networks. Our approach could be generalized to other physical phenomena. Other systems at the edge of a phase transition are very sensitive to external stimuli and might show a similar behaviour. On a broader scope, we show that, although often regarded as an undesirable consequence, power dissipation might actually enable new ways of computing, by taking advantage of the rich phenomenology of correlated systems.

Methods

Sample preparation. A 70 nm VO$_2$ film was grown by reactive sputtering on top of an R-cut Al$_2$O$_3$ substrate. A 4 mtorr Argon/Oxygen mix (8% O$_2$) was used during deposition. The substrate temperature was kept at 520 °C, and cooled down after sputtering at a rate of 12 °C/min. X-ray diffraction shows textured orientation along $\langle 100 \rangle$ for VO$_2$. Transport measurements show a four orders of magnitude IMT, confirming the high quality of the film. The device was fabricated in two lithographic steps (layers). In the first layer, e-beam lithography end e-beam evaporation was used to pattern two Ti (20 nm)/Au (30 nm) electrodes. A small gap (~50 nm) was left between both electrodes, so large electric fields could be generated by applying a few volts. The second layer consists on an Al$_2$O$_3$ (70 nm)/Ti (20 nm)/Au (30 nm) nanowire, patterned on top of the gap and running perpendicular to the first layer electrodes. E-beam lithography and e-beam evaporation was used for this purpose. Several of such devices are patterned in a single sapphire substrate. Optical lithography and reactive ion etching was used to remove the VO$_2$ outside of the gap area and isolate the different devices from each other. More information on the device fabrication process can be found in Supplementary Figure S2.

Fast transport measurements. Measurements were carried out in a TTPX Lakeshore cryogenic probe station. The station is equipped with high-speed (20 GHz) probes, with ground/line/ground geometry and 50 Ω characteristic impedance. In order to avoid reflections (the insulating resistance of the device is in the 10^4 Ω range) a 50 Ω termination to ground was installed before the sample. A 240 MHz Tektronix function generator was used to create the voltage pulses and a 50 Ω terminated Tektronix broadband oscilloscope (20 GHZ) was used to monitor the current. The electrical circuit set up ensured a rise time around 5 ns.

Simulations of the device dynamics. A simple, lumped-element model was used to investigate the electro-thermal dynamics of the device (Fig. 5a,b). The electrical part of the model considers a VO$_2$ gap as a resistor in parallel with a capacitor C, playing the role of parasitic capacitance of the circuit. The charge accumulated in such capacitor is Q. A load resistor R_L is placed in series, and V_{DC} constant bias voltage is applied across the gap and the load. We label the total current as I, while I_R and I_C are the currents flowing through the VO$_2$ and the capacitor. The heating resistor R_{Heater} is electrically isolated from the rest of the circuit, with an input current I_{Input} passing through it.

The value R depends on the VO$_2$ state: metallic or insulating. $R = R_{met} = 200$ Ω in the metallic state, while $R_{ins} = \alpha \cdot e^{\beta/T}$ in the insulating state. The values of $\alpha = 0.0178$ and $\beta = 4500$ are such that $R_{ins} = 10$ kΩ at 339 K, and $R_{ins} = 58.4$ kΩ at 300 K. Whether VO$_2$ is metallic or insulating depends on the device current temperature as well as its thermal history. A hysteresis is set between 335 K and 339 K to mimic the first-order nature of the IMT.

At each simulation step i, $R[i]$ is evaluated depending on the thermal history, and the current through the gap and the capacitor are calculated:

$$I_{gap}[i] = Q[i]/(R[i] \cdot C)$$

$$I_C[i] = (V_{DC} - I_{gap}[i] \cdot R[i])/R_L - I_{gap}[i]$$

With this, we can calculate the evolution of Q using the Euler method.

$$Q[i + 1] = Q[i] + I_C[i] \cdot \delta t$$

where δt is the simulation step.

The temperature T is governed by the thermal part of the model. The model considers the VO$_2$, the Al$_2$O$_3$ barrier and the heating element as a single system with homogeneous temperature. Although simple, it accurately mimics the experimental results. The are two heat sources: Joule heating in the gap ($Q_{gap} = R \cdot I_{gap}^2$) and in the heater ($Q_{Heater} = R_{Heater} \cdot I_{Input}^2$). Heat is evacuated from the device into the environment, consisting on the metallic pads and the sapphire substrate. Such heat loss will depend on the temperature difference between the device and the base temperature of the environment T_{Base}.

The temperature evolution is calculated using the Euler method:

$$T[i + 1] = T[i] + \frac{1}{C_{th}} \cdot (R \cdot I_{gap}^2 + R_{Heater} \cdot I_{Input}^2 - S_{th} \cdot (T[i] - T_{Base})) \cdot \delta t$$

where C_{th} and S_{th} are the thermal capacitance and conductance of the system respectively.

We must note that for simplicity we consider the powers dissipated in the gap and the heater to contribute equally to the temperature change in the VO$_2$. We treat the whole neuristor as a single thermal element with the same temperature.

By considering the individual thermal capacitances of the VO$_2$, Ti/Au pads and Al$_2$O$_3$ barrier enclosed in the 400 nm × 400 nm area of the device, we estimated $C_{th} \approx 1.3 \cdot 10^{-13} J/K$. For this calculation we used the following density d and specific heat c values[49,50]: $d_{VO_2} = 4.34 \cdot 10^3 \frac{kg}{m^3}$, $c_{VO_2} = 690 \frac{J}{K \cdot kg}$, $d_{Au} = 19.3 \cdot 10^3 \frac{kg}{m^3}$, $c_{Au} = 129 \frac{J}{K \cdot kg}$, $d_{Ti} = 4.54 \cdot 10^3 \frac{kg}{m^3}$, $c_{Ti} = 523 \frac{J}{K \cdot kg}$, $d_{Al_2O_3} = 3.97 \cdot 10^3 \frac{kg}{m^3}$ and $c_{Al_2O_3} = 854 \frac{J}{K \cdot kg}$

To estimate the thermal conductance, we took into account two contributions: the thermal conductance of the metallic pads, and the vertical thermal conductance through the substrate. The most thermally resistive part of the pads is the 1.5 μm long stretch closer to the device center. Considering only this part of the pads, we estimated $S_{pads} \approx 1.3 \cdot 10^{-5}$ W/K. Vertical heat transport will go through the VO$_2$ into the sapphire substrate and we estimate it to be $S_{vertical} \approx 8.5 \cdot 10^{-6} W/K$. This gives a total conductance value $S_{th} \approx 2.2 \cdot 10^{-5} W/K$.

For this calculation we used the following thermal conductivity values[49,50]: $\sigma_{VO_2} = 6\frac{W}{m \cdot K}$, $\sigma_{Au} = 310\frac{W}{m \cdot K}$, $\sigma_{Ti} = 21.9\frac{W}{m \cdot K}$ and $\sigma_{Al_2O_3} = 30\frac{W}{m \cdot K}$.

The effect of downsizing and material choice can be explored by calculating the thermal conductance of a similar device in which Au is substituted by Ti as metallic contact, a TiO$_2$ substrate is used, the pad size has been reduced one order of magnitude (down to 40 nm), while the thickness has been decreased to half. Using[50] $\sigma_{TiO_2} = 9\frac{W}{m \cdot K}$, we get:

$$S_{vertical} \approx 9.4 \cdot 10^{-8} W/K, \; S_{pads} \approx 3.7 \cdot 10^{-7} W/K, \; S_{Total} \approx 4.6 \cdot 10^{-7} W/K$$

This is a two orders of magnitude reduction in heat leakage towards the environment, which allows reducing the Joule heating.

Such changes would also decrease the specific heat of the system to $C_{th} \approx 1.3 \cdot 10^{-13} J/K$. This would increase the ratio S_{th}/C_{th}, and hence, the system dynamics by a factor of 4.5.

Parameters used in the simulation. Despite the simplicity of the model, we observe spiking patterns very similar to the experiments. Although the parameters were adjusted to observe oscillatory behavior, they were kept as close as possible to the device characteristics:

$T_{Base} = 325$ K, $V_{DC} = 4.4$ (Fig. 5d)–5.5 V (Fig. 5c), R_{ins} (339 K) = 10 kΩ, $R_{met} = 200$ Ω, $R_L = 2.0$ (Fig. 5d) – 5.0 kΩ (Fig. 5c), $I_{Input} = 0$–5 mA (variable), $C = 0$–10^{-10} F (variable), $R_{Heater} = 20$ Ω, $C_{th} = 10.10^{-13}$ J/K, $S_{th} = 10.10^{-5}$ W/K, $T[0] = 325$ K, $Q[0] = 0$ C and $\partial t = 10^{-13}$ s.

Estimation of temperature rise outside of the neuristor. An estimation of the temperature rise can be obtained by considering the heat flow into a substrate coming from a point source at the surface. In this case, the point source is the proposed neuristor. In an isotropic case, the temperature at a distance *r* will be given by:

$$T = T_{Base} + \; Q/2\pi r\sigma_{substrate}$$

where $Q = S_{Th} \cdot (T_{Neuristor} - T_{Base})$ is the total heat flow coming from the device.

In our experimental device, we estimated $S_{Th, vertical} \approx 8.5 \cdot 10^{-6}$ W/K. According to our simulations, for a parasitic capacitance $C = 10^{-10}$ F, $T_{Neuristor}$ rises 100 K during a spike. Such spike would increase the temperature 5 K at in a point 1 μm away from the device. With the optimization proposed in the simulations part of this methods section, $S_{Th, vertical} \approx 9.4 \cdot 10^{-8}$ W/K. For a system with no parasitic capacitance, the simulations show that $T_{Neuristor}$ rises just 10 K during a spike. Such spike would increase temperature just 15 mK in a point 1 μm away.

We must point out that this calculation is approximate and does not take into account factors such as cooling from the contact pads, which will depend on the particular device design. We must also note that this estimation is an upper limit to the temperature increment, since it considers a steady state in which the device is constantly at the maximum temperature that it reaches during firing. We expect the actual temperature change to be lower.

Data availability
The data supporting the plots and claims of this manuscript are available from the corresponding authors upon reasonable request.

References
1. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Editorial. More than machines. *Nat. Mach. Intell.* **1**, 1–1 (2019).
3. Crane, H. Neuristor-A Novel Device and System Concept. *Proc. IRE* **50**, 2048–2060 (1962).
4. Mattson, R. H. A neuristor realization. *Proc. IEEE* **52**, 618–619 (1964).
5. Nishizawa, J. I. & Hayasaka, A. Two-line neuristor with active element in series and in parallel. *Int. J. Electron.* **26**, 437–469 (1969).
6. Mead, C. Analog VLSI and Neural Systems. Addison-Wesley Longman Publishing Co., Inc. (Addison-Wesley, 1989).
7. Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
8. Furber, S. Large-scale neuromorphic computing systems. *J. Neural Eng.* **13**, (2016).
9. Indiveri, G. *et al.* Neuromorphic Silicon Neuron Circuits. *Front. Neurosci.* **5**, 1–23 (2011).
10. Zhou, Y. & Ramanathan, S. Mott Memory and Neuromorphic Devices. *Proc. IEEE* **103**, 1289–1310 (2015).
11. Romera, M. *et al.* Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).
12. del Valle, J., Ramírez, J. G., Rozenberg, M. J. & Schuller, I. K. Challenges in materials and devices for resistive-switching-based neuromorphic computing. *J. Appl. Phys.* **124**, 211101 (2018).
13. Waser, R. & Aono, M. Nanoionics-based resistive switching memories. *Nat. Mater.* **6**, 833–840 (2007).
14. Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nat. Nanotechnol.* **8**, 13–24 (2013).
15. Zimmers, A. *et al.* Role of Thermal Heating on the Voltage Induced Insulator-Metal Transition in VO$_2$. *Phys. Rev. Lett.* **110**, 056601 (2013).

16. Brockman, J. S. *et al*. Subnanosecond incubation times for electric-field-induced metallization of a correlated electron oxide. *Nat. Nanotechnol.* **9**, 453–458 (2014).
17. Wang, Z. *et al*. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* **16**, 101–108 (2017).
18. Wang, Z., Kumar, S., Nishi, Y. & Wong, H.-S. P. Transient dynamics of NbO_X threshold switches explained by Poole-Frenkel based thermal feedback mechanism. *Appl. Phys. Lett.* **112**, 193503 (2018).
19. Jiang, R. *et al*. Total-Ionizing-Dose Response of Nb_2O_5-Based MIM Diodes for Neuromorphic Computing Applications. *IEEE Trans. Nucl. Sci.* **65**, 78–83 (2018).
20. Pergament A. *et al*. Vanadium Dioxide: Metal-Insulator Transition, Electrical Switching and Oscillations. A Review of State of the Art and Recent Progress. EMN Meeting on Computation and Theory, Energy Materials and Nanotechnology, Istambul (2015).
21. Pergament, A., Stefanovich, G., Malinenko, V. & Velichko, A. Electrical Switching in Thin Film Structures Based on Transition Metal Oxides. *Adv. Cond. Matt. Phys.* **2015**, 1–26 (2015).
22. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453**, 80–83 (2008).
23. Beck, A., Bednorz, J. G., Gerber, C., Rossel, C. & Widmer, D. Reproducible switching effect in thin oxide films for memory applications. *Appl. Phys. Lett.* **77**, 139 (2000).
24. Jo, S. H. *et al*. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
25. Ohno, T. *et al*. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591–595 (2011).
26. Prezioso, M. *et al*. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
27. Pi, S. *et al*. Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension. *Nat. Nanotechnol.* **14**, 35–40 (2018).
28. Boybat, I. *et al*. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **9**, 2514 (2018).
29. Koch, C. *Biophysics of Computation*. (Oxford University Press, 1999).
30. Pickett, M. D., Medeiros-Ribeiro, G. & Williams, R. S. A scalable neuristor built with Mott memristors. *Nat. Mater.* **12**, 114–117 (2013).
31. Ignatov, M., Ziegler, M., Hansen, M., Petraru, A. & Kohlstedt, H. A memristive spiking neuron with firing rate coding. *Front. Neurosci.* **9**, 376 (2015).
32. Yi, W. *et al*. Biological plausibility and stochasticity in scalable VO_2 active memristor neurons. *Nat. Commun.* **9**, 4661 (2018).
33. Feali, M. S. & Ahmadi, A. Realistic Hodgkin–Huxley Axons Using Stochastic Behavior of Memristors. *Neural Process. Lett.* **45**, 1–14 (2017).
34. Kim, S. K. *et al*. Capacitors with an Equivalent Oxide Thickness of <0.5 nm for Nanoscale Electronic Semiconductor Memory. *Adv. Funct. Mater.* **20**, 2989–3003 (2010).
35. Wang, Z. *et al*. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **1**, 137–145 (2018).
36. Tuma, T., Pantazi, A., Le Gallo, M., Sebastian, A. & Eleftheriou, E. Stochastic phase-change neurons. *Nat. Nanotechnol.* **11**, 693–699 (2016).
37. Stoliar, P. *et al*. A Leaky-Integrate-and-Fire Neuron Analog Realized with a Mott Insulator. *Adv. Funct. Mater.* **27**, 1604740 (2017).
38. T. Yajima, T. Nishimura & A. Toriumi, Analog spike processing with high scalability and low energy consumption using thermal degree of freedom in phase transition materials, 2018 IEEE Symposium on VLSI Technology (2018).
39. Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
40. del Valle, J. *et al*. Subthreshold firing in Mott nanodevices. *Nature* **569**, 388–392 (2019).
41. Lee, Y. W. *et al*. Metal-insulator transition-induced electrical oscillation in vanadium dioxide thin film. *Appl. Phys. Lett.* **92**, 162903 (2008).
42. Lepage, D. & Chaker, M. Thermodynamics of self-oscillations in VO_2 for spiking solid-state neurons. *AIP Adv.* **7**, 055203 (2017).
43. Driscoll, T. *et al*. Current oscillations in vanadium dioxide: Evidence for electrically triggered percolation avalanches. *Phys. Rev. B* **86**, 094203 (2012).
44. Benda, J. & Herz, A. V. M. A Universal Model for Spike-Frequency Adaptation. *Neural Comput.* **15**, 2523–2564 (2003).
45. Kumar, S., Strachan, J. P. & Williams, R. S. Chaotic dynamics in nanoscale NbO_2 Mott memristors for analogue computing. *Nature* **548**, 318–321 (2017).
46. Bohaichuk, S. M. *et al*. Fast Spiking of a Mott VO_2-carbon nanotube composite device. *Nano Lett.* **19**, 6751–6755 (2019).
47. Trastoy, J. & Schuller, I. K. Criticality in the brain: evidence and implications for neuromorphic computing. *ACS Chem. Neurosci.* **9**, 1254–1258 (2018).
48. Chialvo, D. R. Emergent complex neural dynamics. *Nat. Phys.* **6**, 744–750 (2010).
49. Stefanovich, G., Pergament, A. & Stefanovich, D. Electrical switching and Mott transition in VO_2. *J. Phys. Condens. Matter.* **12**, 8837 (2000).
50. CRC Handbook of Chemistry and Physics, 84th Edition (Ed. Lide, D. R. *et al*.) Sections 4 and 12. (CRC Press Boca Raton 2003).

Acknowledgements

Author contributions

J.d.V and I.K.S. conceived the idea. J.d.V, and Y.K. designed and fabricated the devices. J.d.V. and P.S. performed the transport measurements and analyzed the data. J.d.V. performed the lumped-element simulations. J.d.V. and I.K.S. wrote the manuscript. All authors participated in the discussion of the results and corrected multiple iterations of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-61176-y.

Correspondence and requests for materials should be addressed to J.d.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.