

Experimental Demonstration of a Second-Order Memristor and Its Ability to Biorealistically Implement Synaptic Plasticity

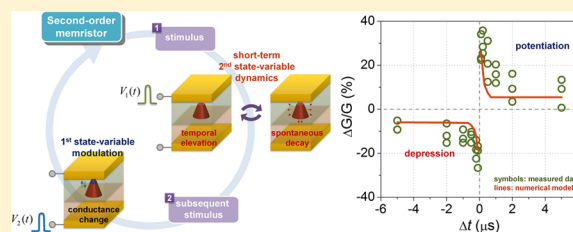
Sungho Kim, Chao Du, Patrick Sheridan, Wen Ma, ShinHyun Choi, and Wei D. Lu*

[†]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, United States

Supporting Information

ABSTRACT: Memristors have been extensively studied for data storage and low-power computation applications. In this study, we show that memristors offer more than simple resistance change. Specifically, the dynamic evolutions of internal state variables allow an oxide-based memristor to exhibit Ca^{2+} -like dynamics that natively encode timing information and regulate synaptic weights. Such a device can be modeled as a second-order memristor and allow the implementation of critical synaptic functions realistically using simple spike forms based solely on spike activity.

KEYWORDS: Memristor, resistive switching, second-order, dynamics, synapse, Ca^{2+} , synaptic plasticity



Memristors and memristive systems,^{1–5} two-terminal resistive devices with inherent memory, have been exploited intensively for memory and logic applications. In particular, memristors have been proposed as an attractive candidate to emulate synaptic functions in biologically inspired neuromorphic systems. In such an implementation, spikes, or action potentials from the presynaptic neuron, can be transmitted through the synapse (memristor) and generate a postsynaptic membrane potential with an amplitude determined by the synaptic weight, i.e., memristor conductance. Additionally, the synaptic weight can be modulated by the spikes from pre- and postsynaptic neurons and the new synaptic weight can be maintained over long-term, thus enabling learning and memory in the system.

To date, critical synaptic learning rules such as spike timing dependent plasticity (STDP) have been demonstrated in memristors in a number of studies.^{6–11} In previous studies, however, the memristors were essentially treated as programmable memory devices. To implement synaptic learning rules the shape of the spikes had to be engineered carefully so that overlapping of the spikes leads to desired effective programming pulse duration or amplitude to encode the information on the relative timing of the spikes from pre- and postsynaptic neurons.^{6–11} Overlapping of the spikes was necessary since there was no other mechanism to encode the relative timing information (in a spike-based asynchronous system) to modulate the memristor's response to the different stimulation conditions. The origin of this limitation can be traced to the fact that the modulation of the memristor conductance, determined by an internal state-variable (e.g., conduction channel size w), was solely controlled by the input (e.g., voltage V applied to the device) and w itself, i.e.

$$\frac{dw}{dt} = f(w, V, t) \quad (1)$$

As a result, to achieve the desired conductance change the pulse width and pulse height had to be carefully engineered,^{6–11} as shown in Figure 1, parts a and b. Mathematically, these devices can be categorized as first-order memristors.¹² On the contrary, in biological systems synaptic plasticity is achieved via simple, nonoverlapping spikes and controlled by the *activity* of the synapse (i.e., frequency and relative timing of the spikes) rather than the amplitude and duration of the spikes. A key factor here is that in biology, the synaptic weight (determined by factors such as receptor level or presynaptic transmitter release probability)^{13–15} is in turn modulated by other, secondary state-variable(s) such as the postsynaptic calcium ion (Ca^{2+}) concentration,^{14–16} rather than directly by the spikes. The rise and spontaneous decay of the Ca^{2+} concentration additionally provides an internal timing mechanism to encode the activity information on the spikes, since the Ca^{2+} concentration value during a spike is determined by the summation of the effect from the current spike and the residue value from the previous activity (where the relative timing between the spikes is encoded due to the natural decay of Ca^{2+} concentration). The Ca^{2+} -based model has been successfully used to explain a broad range of synaptic plasticity effects such as long-term plasticity (LTP) and STDP.^{14,15} On the other hand, previous implementations using first-order memristors require engineering of different pulse shapes for every different learning rule^{17–19} while the lack of the necessary internal dynamics implies the implementation is only phenomenological and nonbiorealistic, and important dynamic effects at both the cell level and system level may be missed. Several recent studies have explored the short-term temporal conductance changes in some memristor devices,^{20–22} although in those studies the temporal evolution is still driven by the same conductance state

Received: February 19, 2015

Published: February 24, 2015

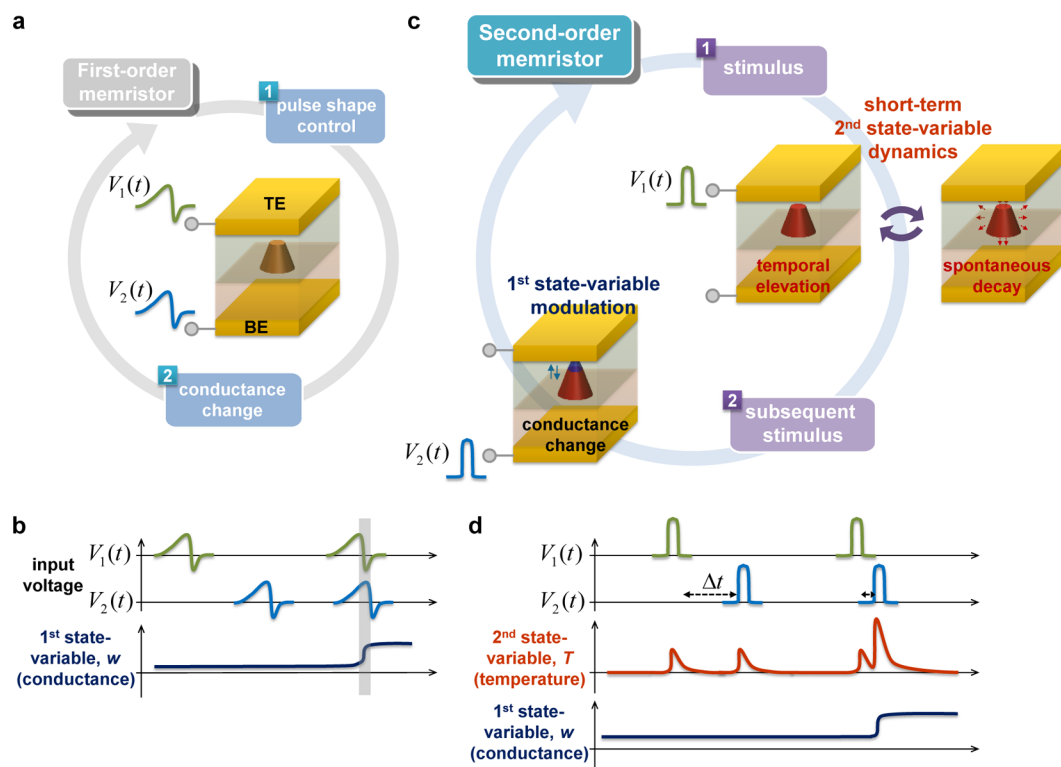


Figure 1. Conceptual operation of the second-order memristor. (a) Operation of a first-order memristor. (b) Input voltage pulse is directly used to control the conductance change and the pulse shape needs to be engineered to obtain the desired conductance evolution. (c) Operation of a second-order memristor. (d) Input pulse can affect both the 1st-order and the 2nd-order state-variables. The modulated 2nd-order state-variable exhibit short-term dynamics and naturally encodes information on the relative timing and synapse activity. The activity of the 2nd-order state-variable then determines the evolution of the 1st-order state-variable that leads to long-term conductance change. The hidden internal dynamics of the 2nd-order state-variable modulation enable biorealistic implementation of synaptic plasticity based on synaptic activity using simple, nonoverlapping spikes.

variable (i.e., size of a volatile conduction channel) and can still be described by the first-order memristor eq 1. The inability to utilize secondary, Ca^{2+} -like dynamic factors makes it impossible to implement critical bidirectional synaptic plasticity effects such as STDP in a biorealistic fashion based on first-order memristors.¹²

This problem can be addressed by realizing that a memristor should also be considered as a dynamic device governed by internal processes, as emphasized by the original memristor theoretical framework,^{1,2} rather than just a simple programmable memory device. In particular, even though in many cases a typical memristor device can be modeled by one internal state variable (i.e., the size of the conduction channel region, Figure 1, parts a and b) which is in turn directly modulated by the stimuli as described by the first-order memristor model (1); in so-called second-order memristors¹² the memristor conductance (and equivalently, the associated state variable) will be regulated by other, secondary state-variable(s) (i.e., a second-order state-variable T), thus enabling the devices to implement complex, potentially biorealistic dynamic effects.

Mathematically, a second-order memristor can be described as¹²

$$\frac{dw}{dt} = f(w, T, V, t) \quad (2)$$

where w is the first order, weight state variable, while T is another, second order state variable. In particular, if the second-order state-variable T exhibits Ca^{2+} -like internal dynamics, biorealistic emulation of synaptic functions may be achieved naturally, as shown in Figures 1c and 1d, where the dynamics of

the second-order state variable T offers an internal timing mechanism and enables *activity-dependent* modulation of the conductance state variable w .

Here we show experimentally that second-order memristor effects can be pronounced in a common oxide-based memristor, and demonstrate that by controlling the internal, short-term dynamics of the second-order state variable, different long-term synaptic functions can be achieved naturally, without the need for complex programming waveform design or the use of overlapping programming pulses. Moreover, experimental observations of these second-order memristor effects can be quantitatively explained using both detailed physics-based numerical simulation and a compact second-order memristor model including two sets of state-variables, enabling large scale system-level implementation for the realization of memristor-based neuromorphic computing systems.

The operation of the second-order memristor can be understood from carefully examining the different state-variables involved during the resistive switching (RS) process that causes the memristor conductance to change. In oxide-based memristors, the RS behavior is believed to be caused by oxygen vacancy (V_O) transport in the oxide layer, where regions with high concentrations of accumulated V_O s provide high conductance channels for electrical transport (i.e., forming conducting filaments (CFs)).²³ In this regard, the size of the conductive channels can be considered as the first-order state-variable (denoted as w), which directly determines the memristor conductance. Significantly, the evolution of the first-order state-variable w (driven by V_O drift and diffusion

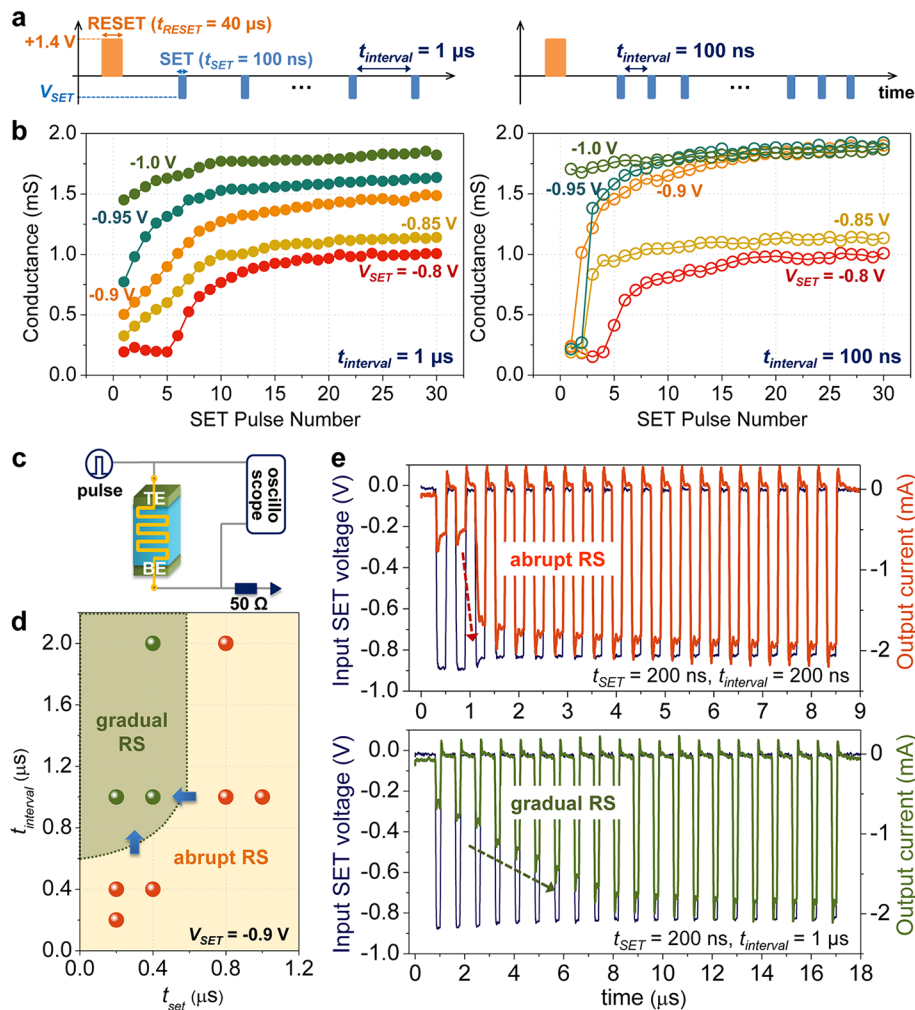


Figure 2. Effect of temporal accumulation. (a) Schematic of pulse trains used for the measurements. A pulse train consists of a single reset pulse ($V_{RESET} = 1.4$ V, pulse duration $t_{RESET} = 40$ μ s) and 100 subsequent set pulses ($V_{SET} = -0.8 \sim -1.0$ V, pulse duration $t_{SET} = 100$ ns), where the intervals between the set pulses ($t_{interval}$) are either 1 μ s or 100 ns. (b) Measured conductance changes with different V_{SET} levels as the number of applied set pulses increases for $t_{interval} = 1$ μ s and $t_{interval} = 100$ ns. The conductance was measured during the set pulse. (c) Transient response measurement setup. (d) Qualitative analysis of the RS behavior (gradual or abrupt) for various t_{SET} and $t_{interval}$ conditions. (e) An example showing abrupt RS with $t_{SET} = 200$ ns and $t_{interval} = 200$ ns (top) and an example showing gradual RS with $t_{SET} = 200$ ns and $t_{interval} = 1$ μ s (bottom).

processes) is determined not only by input signals but also by internal parameters such as the local temperature of the device,^{23–25} which can be considered as a second-order state-variable (denoted as T). For example, during the application of a set (from a high resistance state (HRS) to a low resistance state (LRS)) or reset (from LRS to HRS) pulse, the local temperature T increases due to Joule heating which in turn dramatically affects the drift and diffusion processes of V_O s leading to the growth and dissolution of the CF and the subsequent conductance change (Supporting Information, Figure S1c). More importantly, the dynamics of T is short-term and provides an internal timing mechanism; its value increases abruptly upon the application of a voltage pulse and decays spontaneously after the removal of the stimulation. As a result, the first-order state-variable w during a voltage pulse can still be affected by the activity from previous pulses (if T has not decayed to the steady-state value) and the degree of which will depend on the relative timing of the stimuli. In this sense, the second-order state-variable T plays the role analogous to the Ca^{2+} concentration^{14–16} in that it is not the weight state variable itself but instead regulates the evolution of the weight

state-variable, and allows more biorealistic implementation of activity-dependent synaptic plasticity in such second-order memristors.

The memristor devices used here are based on a common tantalum oxide-based bilayer device structure, which consists of a resistive Ta_2O_{5-x} film (as the RS layer) and a conductive TaO_x film (as the V_O reservoir) (methods in Supporting Information), section 1.^{25–27} These two layers are sandwiched by top and bottom Pd electrodes (TE and BE). Similar to other memristor devices, a pinched-hysteresis behavior can be clearly observed in the current–voltage (I – V) characteristics, as shown in Figure S1a (Supporting Information) where the device conductance can be increased with a negative voltage (applied to the TE) and decreased with a positive voltage. Although a common oxide-based memristor is studied here, we note the second-order memristor effect can be intrinsic to many other memristor devices and materials, as long as proper state variables can be identified and their dynamics can be controlled.

First, we present results in a series of timing-controlled pulse experiments to elucidate the internal dynamics that drive the resistive switching behavior, and demonstrate the second-order

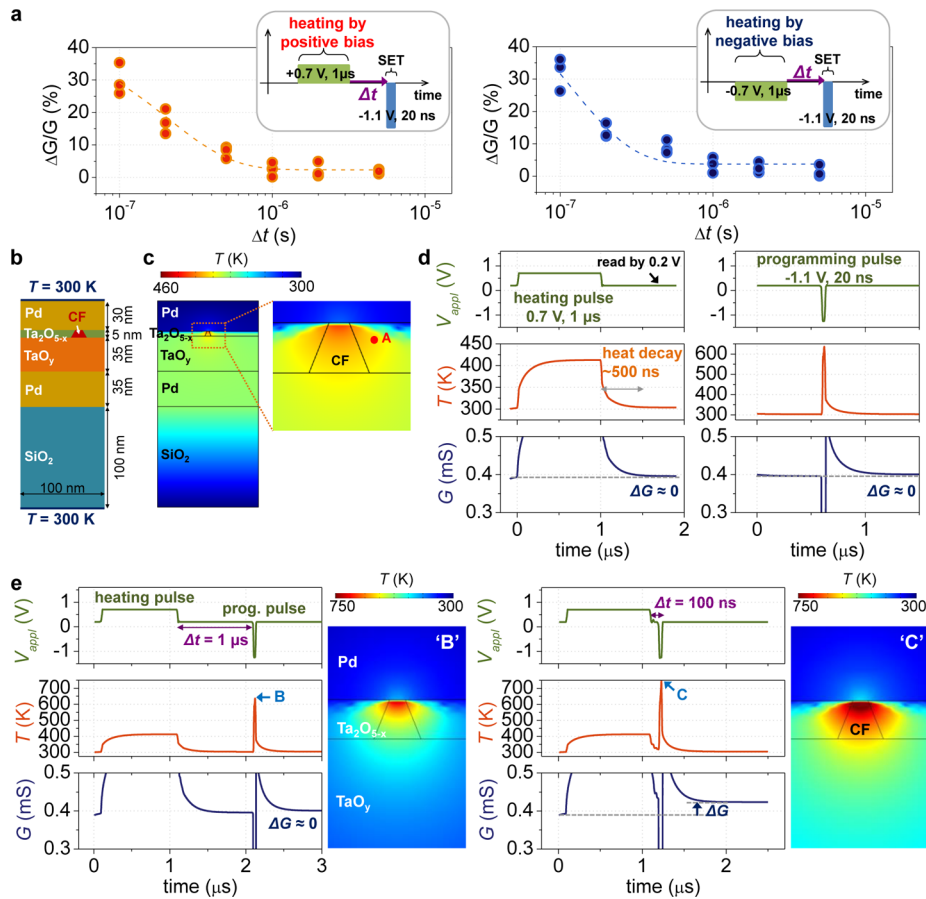


Figure 3. Revelation of different state variables in the second-order memristor. (a) Memristor conductance change as a function of the pulse interval (Δt) (the dotted line is an exponential fitting). The device is only subjected to two pulses in each test: a first heating pulse with a 0.7 V amplitude and 1- μ s duration, and a second programming pulse with a relatively higher voltage (-1.1 V) and shorter duration (20 ns). When the two pulses are close to each other (within 500 ns), a conductance change is observed. The sign of the conductance change is only affected by the polarity of the programming pulse, and independent of the polarity of the heating pulse. (b) Cross-section of the device used in the multiphysics numerical simulation. (c) Simulated temperature profile inside the device when a single heating pulse (0.7 V, 1 μ s) is applied to the device. (d) Simulated transient temperature evolution of the device (measured at point “A” as depicted in part c) during and after the application of just the heating pulse (left) and just the programming pulse (right). Natural temperature decay is clearly observed for ~ 500 ns after the applied pulse voltage is removed. Neither the heating pulse nor the programming pulse can induce a conductance change in the device alone (bottom). (e) Simulated transient temperature evolution when the pulses are applied, for $\Delta t = 1 \mu$ s and $\Delta t = 100$ ns. When the two pulses are far apart ($\Delta t = 1 \mu$ s), the temperature achieved during the programming pulse is similar to that achieved during the heating pulse alone (part d) and no significant conductance change is observed. When the two pulses are close to each other ($\Delta t = 100$ ns), the temporal heat accumulation leads to an elevated temperature during the programming pulse (marked as point C), which causes a measurable conductance change.

memristor effect in these devices. Figure 2a shows a schematic of the pulse trains used for one of such measurements. A pulse train consists of a single reset pulse ($V_{\text{RESET}} = 1.4$ V with pulse duration $t_{\text{RESET}} = 40 \mu$ s) and 100 subsequent set pulses (with pulse duration $t_{\text{SET}} = 100$ ns). The interval between the set pulses (t_{interval}) is varied between 1 μ s and 100 ns in the tests, and the test is then repeated for 5 different V_{SET} pulse heights of $-0.8 \sim -1.0$ V. When t_{interval} is 1 μ s, a gradual set transition is observed as the number of applied set pulses increases, and intermediate conductance states can be achieved with the conductance values correlating well to the different V_{SET} levels and the number of V_{SET} pulses applied (Figure 2b). In contrast, abrupt set transition is observed after the application of only a few pulses in the case of $t_{\text{interval}} = 100$ ns, and intermediate states cannot be obtained by either controlling the V_{SET} amplitude or controlling the number of V_{SET} pulses. The different behaviors for otherwise identical programming conditions can be explained by considering the device temperature T dynamics (i.e., increase due to Joule heating by the programming pulse

and spontaneous decrease due to heat dissipation). When the interval between programming pulses, t_{interval} is sufficiently long (e.g., $t_{\text{interval}} = 1 \mu$ s), Joule heating generated from each set pulse can be fully dissipated, and as a result, the programming pulses are essentially decoupled and each pulse is not sufficient to fully set the device so intermediate states can be obtained by controlling the amplitude and number of the programming pulses. On the other hand, when the interval between programming pulses, t_{interval} is short (e.g., $t_{\text{interval}} = 100$ ns), temporal heat accumulation can occur, leading to significant increase of the device temperature during the pulse which in turn dramatically speeds up the filament growth^{23,24} and leads to abrupt RS.

To systematically study the how the RS behavior is affected by the internal dynamics, the transient response of the RS was measured at different t_{SET} and t_{interval} configurations (Supporting Information, section 2), as shown in Figure 2c. In this measurement, a train of set pulses were applied with the pulse amplitude fixed at $V_{\text{SET}} = -0.9$ V while t_{SET} and t_{interval} were

changed from 200 ns to 2 μ s. Figure 2d shows that two qualitatively different RS behaviors can be observed, depending on t_{SET} and t_{interval} , and can be categorized as follows: (1) under long t_{SET} (>800 ns), abrupt RS behavior is always observed regardless of t_{interval} , because sufficient heat accumulation can occur even inside a single set pulse when the pulse is sufficiently long. The increased temperature in turn causes a positive feedback during the set process and leads to abrupt RS.²³ (2) Under short t_{SET} (e.g., 200 ns) and with short t_{interval} (e.g., 200 ns), heat generated by the set pulses can still be temporary accumulated due to the short t_{interval} . Thus, even though temperature increase caused by a single short set pulse may not be sufficient, the ability to accumulate heat from a series of set pulses still leads to sufficiently elevated temperature to drive abrupt RS, as shown in Figure 2e. (3) Finally, under short t_{SET} (e.g., 200 ns) and with long t_{interval} (e.g., 1 μ s), gradual RS is observed. This is because in this condition each set pulse does not create high enough temperature rise, while heat generated during the previous set pulses will be effectively dissipated during the long interval, so that heat accumulation is not achieved either so the device temperature is never increased to a very high level. As a result, slower V_{O} migration can be achieved leading to the more gradual RS observed in this condition (Figure 2e).

The role of the dynamics of the temperature state variable on the conductance change can be revealed more clearly from a measurement based on two simple, nonoverlapping pulses, as shown in Figure 3a. To separate the effects of heating (temperature change) and programming (filament size change), each test consists of two pulses: a first pulse of 0.7 V and 1 μ s duration (termed the heating pulse) such that the voltage is low enough and cannot induce a conductance change in the device but the pulse duration is long enough to cause measurable temperature change; and a second pulse (termed the programming pulse) with a higher voltage (-1.1 V) but a much shorter duration (20 ns) which is not sufficient to cause conductance modulation by itself either due to the short pulse duration (Supporting Information, section 3). The interval between the two pulses (Δt) was then changed systemically to determine how the memristor responds to the pulses. Significantly, when the two pulses are close to each other (e.g., within 1 μ s), a detectable conductance change is observed as shown in Figure 3a, and larger conductance changes (up to 30%) are observed when the pulse interval is shorter. Furthermore, the measured conductance changes are apparently independent of the polarity of the first pulse (the heating pulse), as shown in Figure 3a, where reversing the polarity of the heating pulse from positive to negative still leads to qualitatively and quantitatively similar conductance increase during the programming pulse. While reversing the polarity of the programming pulse leads to conductance decrease with a similar pulse interval dependence, regardless of the polarity of the heating pulse (Supporting Information, Figure S5a). This is strong evidence that the internal dynamic process resultant from the heating pulse is not caused by polarity-dependent effects (e.g., charge trapping/detrapping or ion accumulation processes which depend on the voltage polarity), and further supports the idea of local temperature T as the state-variable, since Joule heating is independent of the voltage polarity. Specifically, the modulation of the device conductance observed by the two nonoverlapping pulses can be explained by the temporal heat summation effect: the first (heating) pulse temporarily increases the device temperature T regardless of

the polarity of the pulse, which then decays spontaneously after the removal of the heating pulse; if the second pulse is applied before T has decayed to its resting value, a higher T can be obtained within the second pulse and a conductance change is more likely since the V_{O} drift and diffusion processes become significantly accelerated at the elevated T .²³ Additionally, a longer interval will lead to a smaller conductance change due to the spontaneous heat dissipation leading to a lower residue T at the arrival of the second (programming) pulse. To this regard, the dynamics of the internal device temperature T can be considered playing a role analogous to the (postsynaptic) Ca^{2+} concentration that facilitates synaptic plasticity. It has been shown experimentally that the postsynaptic Ca^{2+} concentration temporarily rises during a spike, then decays spontaneously.^{14–16,28} If another spike arrives before the Ca^{2+} concentration decays to the resting value, elevated Ca^{2+} concentration can be achieved during the second spike that leads to long-term synaptic weight changes, the extent of which depending on the value of the Ca^{2+} concentration which is in turn regulated by the relative timing between the spikes. This picture has been shown to be able to explain different rate- and timing-dependent synaptic plasticity in a unified theoretical framework.²⁹ In this sense, very similar internal dynamic processes happen inside the second-order memristor, including the temporary increase of temperature during a pulse, the spontaneous decay after the pulse, and the temporal summation effect of the temperature which leads to conductance changes that are dependent on the relative timing between the pulses.

The hypothesis of temperature as the second-order state-variable and the short-term dynamics effects including decay and temporal heat summation are further supported from detailed multiphysics numerical simulation by considering the ion drift, diffusion and Joule heating effects (Supporting Information, section 4). Figure 3b shows a 2-D cross section of the simulated cell geometry with electrothermal boundary conditions. When a single pulse (0.7 V, 1 μ s) is applied to the device, Joule heating is generated at the CF (Figure 3c) leading to a ~ 450 K temperature rise. Additionally, after the pulse is removed, the spontaneous temperature decay is observed to occur over a time scale of ~ 500 ns (Figure 3d). The relatively long decay time is due to the relatively lower thermal conductivity values assumed for the thin electrodes^{30–32} (Supporting Information, section 4), which allow for slower heat dissipation. Additionally, the SiO_2 substrate acts as a heat insulator which further slows down heat dissipation. Specifically, as shown in Figure 3e, when the two pulses corresponding to the ones used in Figure 3a are applied to the device (a first pulse of 0.7 V with 1 μ s duration for heat generation and a second pulse of -1.1 V with 20 ns duration for programming) and the interval of these two pulses (Δt) is short, a higher device temperature is obtained (e.g., $T > 750$ K is obtained when $\Delta t = 100$ ns, point “C”) during the second pulse due to the temporal heat summation effect. The elevated temperature enables fast V_{O} migration and results in detectable conductance changes during the programming pulse. On the other hand, when the pulse interval is long the internal temperature during the programming pulse is kept low (e.g., $T < 650$ K for $\Delta t = 1$ μ s, point “B”) and negligible conductance change is obtained. The simulation results confirm the key experimental findings and the second-order memristor hypothesis, that is (1) Joule heating generated by the application of voltage pulses leads to a temporary temperature increase; (2) temperature exhibits short-term dynamics and will

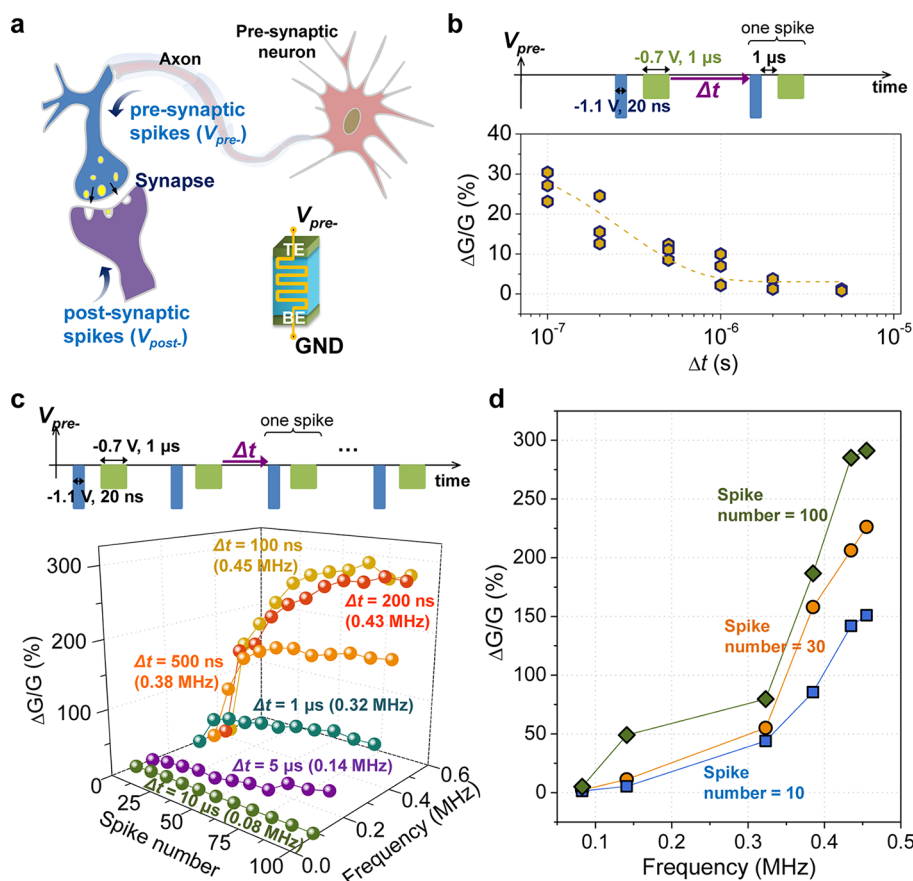


Figure 4. Frequency-dependent plasticity effects. (a) Experimental set up for frequency-dependent plasticity implementation: the spikes are applied to the presynaptic side (the TE), with the BE grounded. (b) Experiments with only two spikes (the dotted line is an exponential fitting). Each spike consists of a programming pulse and a heating pulse. The intervals between the two spikes (Δt) are changed systematically to probe memristor response. (c) Memristor conductance change as a function of the number of spikes and the spike frequency. A total of 100 spikes (pulse-pairs) are applied at each frequency, and the spike frequency is changed from 0.08 to 0.45 MHz by tuning the interval between spikes. (d) Memristor conductance change as a function of spike frequency after the application of different number of spikes.

decay spontaneously when the pulse is removed, with a decay time constant of ~ 500 ns for our devices (Figure 3d), and (3) temporal summation of the thermal effect can occur and can lead to an elevated device temperature that is higher than produced by a single pulse alone and subsequently enabling device conductance modulation, the extent of which depends on the relative timing of the pulses (Figure 3, parts a and e). These results prove that the internal temperature can be used as a second-order state-variable whose (short-term) dynamics can be used to regulate the (long term) conductance change of the device, in a way similar to how internal Ca^{2+} concentration dynamics regulate synaptic plasticity in biologically synapses. In turn, it makes it possible to emulate synaptic plasticity in a biorealistic fashion by employing internal dynamic effects with such second-order memristors.

In the following, we verify this concept by demonstrating different synaptic plasticity effects such as frequency-dependent weight changes^{33,34} and timing-dependent plasticity (STDP³⁵) in these second-order memristors using spikes consisting of simple, nonoverlapping input signals (Figure 4a) in a biorealistic fashion. We first show how the second-order memristor will respond to changes in stimulation frequency by applying a series of stimuli to the presynaptic side of the device (i.e., TE side of the device), as shown in Figure 4b. Each stimulation is a pulse pair consisting of a large but narrow first pulse (-1.1 V with 20 ns duration—the programming element)

and a small but long second pulse (0.7 V with 1 μs duration—the heating element), and can be considered together as a single spike. The first experiment involves the application of only two spikes (pulse-pairs) applied successively to the device, with the interval between the spikes Δt changed systematically. As expected, Figure 4b shows that the second spike always leads to a conductance increase, with a clear dependence on the time interval between the two spikes, i.e., a larger interval leading to a smaller conductance enhancement (potentiation). Similar effects can be observed when 100 spikes (pulse-pairs) were applied with different frequencies by controlling the interspike time interval, as shown in Figure 4c. At each stimulation frequency (determined by the inter spike interval Δt), an increase in device conductance, i.e., potentiation, is generally observed as the number of spikes is increased. However, when the stimulation frequency is low, the potentiation effect is weak and no potentiation is observed for stimulation frequency below 0.1 MHz, highlighting the effect of the spike rate on conductance modulation (synaptic plasticity). The conductance modulation with respect to the stimulation frequency is replotted in Figure 4d, for cases with different numbers of spikes, and strong frequency dependence on the conductance modulation is clearly observed. This frequency dependence can be readily explained by the second-order memristor model, since spike trains with higher frequency mean smaller intervals between spikes, which allow more pronounced heat summation

and higher device temperature experienced during the programming pulse leading to more significant conductance changes, as already explained above. Additionally, by applying stimulation to the presynaptic side (the sign of the heating element is insignificant), similar frequency-dependent effects can be obtained for long-term decrease of the memristor conductance (depression), as shown in Supporting Information, section 5. A different amplitude (1.6 V compared to 1.1 V used in the presynaptic pulses) is used in the programming element for the postsynaptic spike to account for the asymmetry in the set/reset processes.

Next, we show that the internal dynamics in the second-order memristors also naturally lead to important timing-based synaptic plasticity effects such as STDP. Contrary to previous attempts to implement STDP,^{6–11} where the relative timing between the pre- and postsynaptic neuron spikes had to be converted to differences in pulse amplitude or duration by engineering overlapping programming pulse profiles, here STDP can be implemented with simple, nonoverlapping pulses since the short-term T dynamics provide an internal timing mechanism and enables timing-dependent long-term weight change. Similar spikes used in the frequency-dependent study were also used to implement STDP, as shown in Figure 5, parts

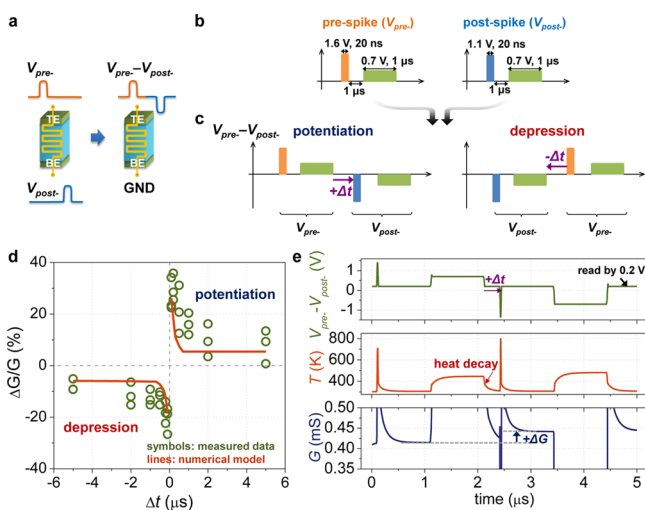


Figure 5. Implementation of spike-timing dependent plasticity in the second-order memristor. (a) Experimental set up: a pair of spikes ($V_{pre} - V_{post}$) applied to TE is equivalent to a pair of spikes V_{pre-} and V_{post-} applied to the presynaptic and postsynaptic side, respectively. (b) Presynaptic and postsynaptic spikes used in this study. (c) Equivalent pulses applied to the TE of the device, for $\Delta t > 0$ and $\Delta t < 0$. (d) Measured (symbols) and simulated (solid lines) STDP results obtained in the second-order memristor. (e) Simulation results showing the spike pair, the internal temperature evolution and the device conductance evolution during a spike pair with $\Delta t = 300$ ns.

a and b, where the presynaptic (postsynaptic) spikes consist of a programming element: 1.6 V (1.1 V) with 20 ns duration and a heating element: 0.7 V with 1 μ s duration applied to the TE (BE) respectively. Equivalently, the postsynaptic spike corresponds to a programming element of -1.1 V with 20 ns duration and a heating element of -0.7 V with 1 μ s applied to the TE, as shown in Figure 5c. Similar to earlier discussions, each programming and heating element alone cannot modulate the device conductance. However, when the presynaptic spike reaches the device earlier than the postsynaptic spike (i.e., in the case of $\Delta t > 0$), the postsynaptic spike will be affected by

the temporal heating effect from the presynaptic spike. This elevated temperature during the second spike causes the second spike (the postsynaptic spike in this case, which with a negative programming voltage will increase the device conductance) to be stronger than that from the first spike (the presynaptic spike, which with a positive programming voltage will decrease the device conductance) and thus the overall effect will be dominated by the second spike and an overall increase in device conductance (potentiation) is obtained. In the opposite case ($\Delta t < 0$), a decrease in device conductance (depression) can be obtained by similar arguments. Indeed, Figure 5d shows STDP results measured in our memristor using the non-overlapped, spike-pairing protocols, where the conductance was measured with a small (0.2 V) read voltage after the application of the spikes. Significantly, an effect analogous to STDP observed in biology is clearly observed, with the sign of conductance change determined by the sign of Δt , and the size of the conductance change determined by the value of Δt with larger relative timing between the pre- and postspikes (larger Δt) resulting in smaller conductance change and vice versa. Here again, T plays a similar role of the (postsynaptic) Ca^{2+} concentration and provides an intrinsic mechanism to encode the spike timing and activity information, which in turn causes the first-order state-variable w to change accordingly leading to corresponding memristor conductance changes. Furthermore, we want to note that even though the timing information is encoded in the short term T dynamics, the effect on the memristor conductance change is through the first-order state-variable and is thus long-term (as in biology), and the conductance change in the second-order memristors have been verified to last hours and days after stimulation (Supporting Information, section 7).

The STDP behavior observed in the second-order memristor can again be explained by detailed microscopic physical modeling of the RS process under the test conditions, as shown in Figure 5d, which shows a good quantitative agreement with the measured data. The numerical physical simulation further captures the internal dynamic evolutions of the device temperature T and the device conductance modulation at different stimulation conditions (Figure 5e), and thus not only provides plausible explanations for the observed dynamic effect but also yields accurate predictions of the device behavior under different stimulation conditions that enable systematic implementation of synaptic functions.

Finally, we show that all key findings can also be captured by a simple analytical memristor model based on two (sets of) macroscopic state-variables. The second-order memristor model highlights the rich dynamic nature of the device, and also facilitates large-scale simulation of neuromorphic systems based on these dynamic devices. Details of the model development are discussed in Supporting Information, section 8. Briefly, the first state-variable, which directly controls the device conductance, includes the CF radius (r) and the depleted gap length (g) in our oxide memristor structure. The filament size state-variables r and g are in turn regulated by the temperature state-variable (including T which represents the internal filament temperature serving as the second-order state-variable, and T_{bulk} , which represents the effective temperature in the bulk of the device outside of the filament), serving as the second-order state-variable. Figure 6a shows the calculated DC I - V characteristics during the set and reset processes using this simple analytical second-order memristor model, which show good agreement with the measured data. The physical nature of

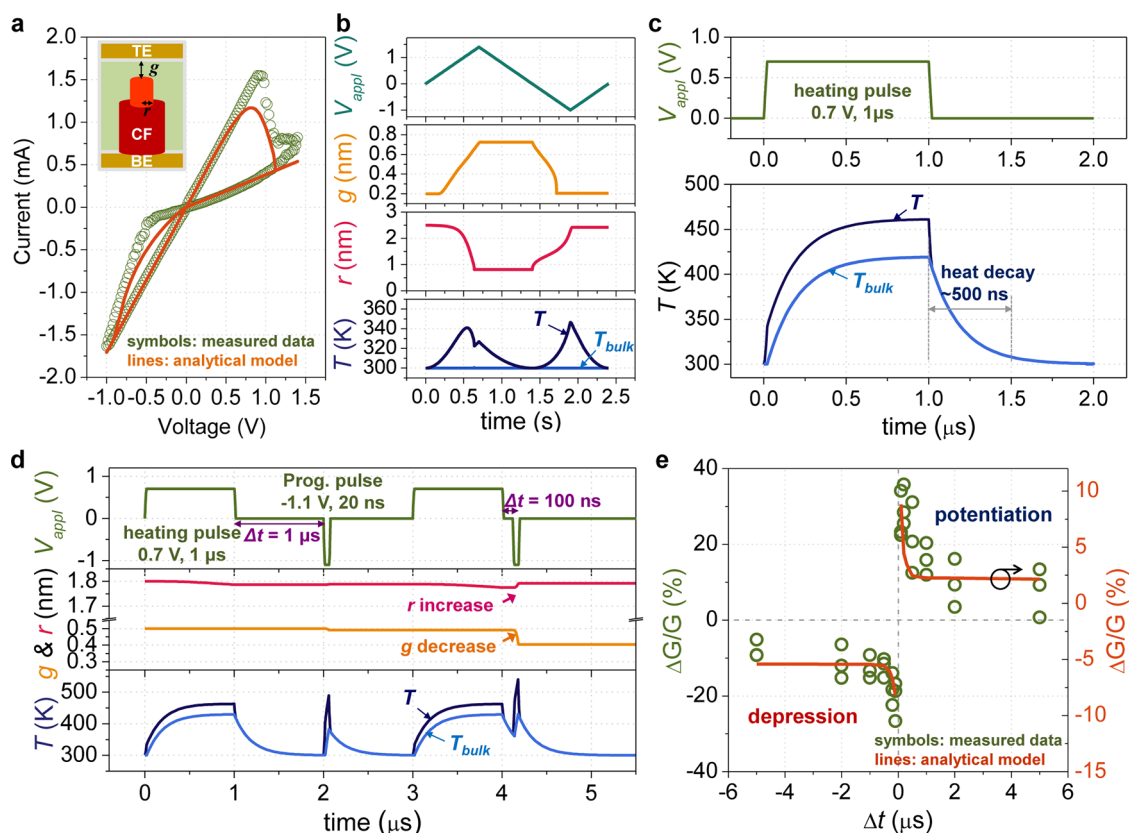


Figure 6. Analytical second-order memristor model results. (a) Measured (circles) and calculated (solid line) DC I – V characteristics of the memristor device. Inset shows the filament structure in the device. (b) The time-evolution of the parameters (g , r , T , and T_{bulk}) obtained from the analytical second-order memristor model during DC sweep. (c) Transient response of T and T_{bulk} obtained from the analytical second-order memristor model when a single pulse (0.7 V, 1 μ s) is applied to the device. (d) Evolution of T and T_{bulk} and g , r obtained using the analytical memristor model during the STDP measurement for $\Delta t = 100$ ns. (e) STDP behavior reproduced using the analytical second-order memristor model (solid lines) that shows good agreement with experimental results (symbols).

both these processes can be studied by examining the change in the parameters (i.e., g , r , T , and T_{bulk}), as shown in Figure 6b. Specifically, a depleted gap of ~ 0.8 nm is formed during reset, leading to a decrease in the device conductance. The radius of the CF also changes accordingly and the increase in r is a main factor of the increase in device conductance during set. In addition, T increases due to Joule heating, while T_{bulk} remains close to room temperature during DC sweeps because the relatively slow DC sweeps allow for sufficient heat dissipation to the environment.

The second-order memristor model also reveals the coupling of the different state variables g , r , T , and T_{bulk} during RS, and captures the nature of the short-term temperature dynamics. Figure 6c shows the calculated transient response of T and T_{bulk} when a single pulse (0.7 V, 1 μ s) is applied to the device. Natural heat decay is clearly observed for ~ 500 ns after the applied pulse voltage is removed, consistent with the simulation result from the numerical model shown in Figure 3d. This short-term dynamics of T and the temporal summation effect thus allow the rate- and timing-dependent memristor conductance change, as discussed earlier. For example, Figure 6d shows the evolution of the second-order state-variables (T and T_{bulk}) and the first-order state-variables (g , r) during the application of the spikes in the STDP experiment (using the pulses as those used in Figure 5d), and quantitative agreements can be obtained with the experimental results using this simplified second-order memristor model, as shown in Figure

6e. As a result, this analytical model, with the emphasis on internal dynamics, enables accurate prediction of the dynamic behavior of the device and allows large scale neuromorphic circuit simulation, which will be essential in realizing a neuromorphic system design.

In summary, we show second-order memristor effects can be significant in oxide-based memristors, with different state-variables governing the short-term and long-term dynamics of the device. Specifically, the short-term temperature dynamics plays a role similar to Ca^{2+} level in biological synapses and provide a native mechanism to encode timing and synaptic activity information. Following the theoretical framework based on Ca^{2+} -driven synaptic plasticity, different synaptic behaviors including frequency-dependent plasticity and timing-based plasticity can be implemented and quantitatively explained using a second-order memristor model. These findings enable the development of biorealistic neuromorphic systems based on synaptic activity, without having to engineer programming parameters for specific learning rules. A pulse pair consisting of a programming pulse and a heating pulse was used in this study to emulate a spike due to its ease of implementation since square pulses are much easier to obtain in a circuit, although one unified pulse with a fast rising edge and longer tail (e.g., similar to the shape of a biological spike) should also be able to produce the frequency- and spike-timing dependence of synaptic plasticity by taking into account the second order dynamic effects. Additionally, the demonstrated time window of

STDP ($\sim 1 \mu\text{s}$) is much narrower than the typical STDP time window in biological synapses ($\sim 10 \text{ ms}$).³⁵ The faster time window makes the device more compatible with the other fast electronic components employed in hardware implementation of neuromorphic systems, although slower time window can also be obtained through continued device structure and material engineering to control the temperature (or other types of second-order state-variable) decay rate. Finally, although a relatively high programming current ($\sim \text{mA}$) is used in this study, much lower programming current (e.g., $< 1 \text{ nA}$) has been reported in carefully designed memristor devices,³⁶ suggesting that memristor-based systems can operate at much lower power consumption.

Our studies highlight that memristors are inherently dynamic devices and should not be treated simply as programmable resistors. We expect the rich dynamic processes revealed in this study for these seemingly very simple devices will stimulate continued fundamental device development and broaden their applications. The experimental implementation of the second-order memristor system along with the application of the second-order memristor model to explain the different internal dynamic effects will also encourage the development of other second-order or higher-order memristive systems based on different types of physical mechanisms for efficient memory, computing, and other applications in the future.

■ ASSOCIATED CONTENT

● Supporting Information

Fabrication of the memristor devices, dc and transient response measurements, effects from the heating and the programming pulses alone, development of physics-based numerical model, frequency-dependent depression, implementation of the STDP behavior using only first-order effects, long-term effect of STDP, and analytical 2nd-order model development. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*(W.D.L.) E-mail: wlupee@eecs.umich.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. S.K. and C.D. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work at the University of Michigan was supported by the AFOSR through MURI Grant FA9550-12-1-0038, by the National Science Foundation (NSF) through Awards ECCS-0954621 and CCF-1217972, by DARPA through Award HR0011-13-2-0015, and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013R1A6A3A03061299). S.K. is currently with the Department of Electronic Engineering, Sejong University, Korea.

■ REFERENCES

- (1) Chua, L. O. *IEEE Trans. Circuit Theory* **1971**, *18*, 507–519.
- (2) Chua, L. O.; Kang, S. M. *Proc. IEEE* **1976**, *64*, 209–223.
- (3) Strukov, D. B.; Snider, G. S.; Stewart, D. R.; Williams, R. S. *Nature* **2008**, *453*, 80–83.
- (4) Yang, J. J.; Strukov, D. B.; Stewart, D. R. *Nat. Nanotechnol.* **2013**, *8*, 13–24.
- (5) Di Ventra, M.; Pershin, Y. V.; Chua, L. O. *Proc. IEEE* **2009**, *97*, 1717–1724.
- (6) Jo, S. H.; Chang, T.; Ebong, I.; Bhadviya, B. B.; Mazumder, P.; Lu, W. *Nano Lett.* **2010**, *10*, 1297–1301.
- (7) Yu, S.; Member, S.; Wu, Y.; Jeyasingh, R.; Kuzum, D.; Wong, H. P. *IEEE Trans. Elect. Dev.* **2011**, *58*, 2729–2737.
- (8) Alibart, F.; Pleutin, S.; Bichler, O.; Gamrat, C.; Serrano-Gotarredona, T.; Linares-Barranco, B.; Vuillaume, D. *Adv. Funct. Mater.* **2012**, *22*, 609–616.
- (9) Krzysteczko, P.; Münchenberger, J.; Schäfers, M.; Reiss, G.; Thomas, A. *Adv. Mater.* **2012**, *24*, 762–766.
- (10) Wang, Z. Q.; Xu, H. Y.; Li, X. H.; Yu, H.; Liu, Y. C.; Zhu, X. J. *Adv. Funct. Mater.* **2012**, *22*, 2759–2765.
- (11) Zamarreño-Ramos, C.; Camuñas-Mesa, L. A.; Péres-Carrasco, J. A.; Masquelier, T.; Serrano-Gotarredona, T.; Linares-Barranco, B. *Front. Neurosci.* **2011**, *5*, 26.
- (12) Pershin, Y. V.; Di Ventra, M. *Proc. IEEE* **2012**, *100*, 2071–2080.
- (13) Malenka, R. C.; Bear, M. F. *Neuron* **2004**, *44*, 5–21.
- (14) Shouval, H. Z.; Bear, M. F.; Cooper, L. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10831–10836.
- (15) Graupner, M.; Brunel, N. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3991–3996.
- (16) Yang, S.-N.; Tang, Y.-G.; Zucker, R. S. *J. Neurophysiol.* **1999**, *81*, 781–787.
- (17) Caporale, N.; Dan, Y. *Annu. Rev. Neurosci.* **2008**, *31*, 25–46.
- (18) Abbott, L. F.; Nelson, S. B. *Nat. Neurosci.* **2000**, *3*, 1178–1183.
- (19) Roberts, P. D.; Bell, C. C. *Biol. Cybern.* **2002**, *87*, 392–403.
- (20) Chang, T.; Jo, S.-H.; Lu, W. *ACS Nano* **2011**, *5*, 7669–7676.
- (21) Ohno, T.; Hasegawa, T.; Tsuruoka, T.; Terabe, K.; Gimzewski, J. K.; Aono, M. *Nat. Mater.* **2011**, *10*, 591–595.
- (22) Yang, R.; Terabe, K.; Liu, G.; Tsuruoka, T.; Hasegawa, T. *ACS Nano* **2012**, *6*, 9515–9521.
- (23) Ielmini, D. *IEEE Trans. Electron Devices* **2011**, *58*, 4309–4317.
- (24) Larentis, S.; Nardi, F.; Balatti, S.; Gilmer, D. C.; Ielmini, D. *IEEE Trans. Electron Devices* **2012**, *59*, 2468–2475.
- (25) Kim, S.; Choi, S. H.; Lu, W. *ACS Nano* **2014**, *8*, 2369–2376.
- (26) Yang, Y.; Choi, S.; Lu, W. *Nano Lett.* **2013**, *13*, 2908–2915.
- (27) Lee, M.-J.; Lee, C. B.; Lee, D.; Lee, S. R.; Chang, M.; Hur, J. H.; Kim, Y.-B.; Kim, C.-J.; Seo, D. H.; Seo, S.; Chung, U.-I.; Yoo, I.-K.; Kim, K. *Nat. Mater.* **2011**, *10*, 625–630.
- (28) Jackson, M. B.; Konnerth, A.; Augustine, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 380–384.
- (29) Rachmuth, G.; Shouval, H. Z.; Bear, M. F.; Poon, C.-S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1266–E1274.
- (30) Yairi, A.; Nakanishi, T. *J. Phys. Conf. Ser.* **2010**, *214*, 012055.
- (31) J. Schmidt, A.; Cheaito, R.; Chiesa, M. *J. Appl. Phys.* **2010**, *107*, 024908.
- (32) Lee, S. M.; Cahill, D. G. *J. Appl. Phys.* **1997**, *81*, 2590–2595.
- (33) Froemke, R. C.; Tsay, I. A.; Raad, M.; Long, J. D.; Dan, Y. *J. Neurophysiol.* **2006**, *95*, 1620–1629.
- (34) Xu, C.; Zhao, M.-X.; Poo, M.-M.; Zhang, X.-H. *Nat. Neurosci.* **2008**, *11*, 1410–1418.
- (35) Bi, G.; Poo, M. *J. Neurosci.* **1998**, *18*, 10464–10472.
- (36) Gaba, S.; Cai, F.; Zhou, J.; Lu, W. D. *Elec. Dev. Lett. IEEE* **2014**, *35*, 1239–1241.