Editors' Suggestion

Relaxation of a Spiking Mott Artificial Neuron

Federico Tesler,[1] Coline Adda,[2,3] Julien Tranchant,[2] Benoit Corraze,[2] Etienne Janod,[2] Laurent Cario,[2] Pablo Stoliar,[3,4] and Marcelo Rozenberg[5,6,*]

[1] *Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IFIBA, CONICET, Cuidad Universitaria, Buenos Aires 1428, Argentina*

[2] *Institut des Matériaux Jean Rouxel (IMN), Université de Nantes, CNRS, 2 rue de la Houssinière, BP 32229, 44322 Nantes Cedex 3, France*

[3] *CIC nanoGUNE, Tolosa Hiribidea 76, 20018 Donostia-San Sebastian, Spain*

[4] *National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8565, Japan*

[5] *Laboratoire de Physique des Solides, CNRS, Université Paris-Sud,Université Paris-Saclay, 91405 Orsay Cedex, France*

[6] *Department of Physics and Center for Advance Nanoscience, University of California San Diego, La Jolla, California 92093, USA*

(Received 12 September 2017; revised manuscript received 8 August 2018; published 1 November 2018)

We consider the phenomenon of electric Mott transition (EMT), which is an electrically induced insulator-to-metal transition. Experimentally, it is observed that depending on the magnitude of the electric excitation, the final state may show a short-lived or a long-lived resistance change. We extend a previous model for the EMT to include the effect of local structural distortions through an elastic energy term. We find that by strong electric pulsing, the induced metastable phase may become further stabilized by the electroelastic effect. We present a systematic study of the model by numerical simulations and compare the results to experiments in Mott insulators of the AM_4Q_8 family. Our work significantly extends the scope of our recently introduced leaky-integrate-and-fire Mott neuron [P. Stoliar *et al.*, Adv. Funct. Mat. 27, 1604740 (2017)] to provide a better insight into the physical mechanism of its relaxation. This is a key feature for future implementations of neuromorphic circuits.

DOI: 10.1103/PhysRevApplied.10.054001

I. INTRODUCTION

The information age we live in is supported on a physical underlayer of electronic hardware, which originates in condensed-matter physics research. The mighty progress made in recent decades has produced faster and more power-efficient electronic devices. This has enabled a seemingly endless improvement in computer performance. However, with the smallest feature size of transistors extending down to a mere 5 nm, the present technology is reaching an unavoidable physical limit. This situation concerns the current computational paradigm, which is based on binary logic and the von Neumann architecture. While incremental progress in this direction is likely to continue, the search for disruptive technologies is under way.

On the other hand, we are currently witnessing significant progress in artificial intelligence. In fact, older algorithmic paradigms, such as neural networks combined with newer ideas such as wavelet-based filtering, have produced a remarkable improvement in the performance of computers at dealing with tasks where they were traditionally poor, such as pattern recognition. This area of research is known as Deep Neural Networks [1,2]. However, most of this progress has been reached by running the new algorithms with conventional codes on conventional (i.e., von Neumann) computers. While the performance of these algorithms with regard to many tasks is remarkable, it is unfortunately limited by the computational power of current machines. For instance, the number of synapses in a brain is 10^{15}; thus even if one allocates a single memory byte per synapse, conventional computers fall irreparably short.

Therefore, an exciting perspective is to implement the neural networks directly on hardware, by realizing novel electronic devices that may directly implement neural network functionality. These bioinspired electronic circuits have two key new components: one is a "synapse analog" and the other is a "neuron analog." The former can be implemented by a nonvolatile resistance, the value of which can be programmed (and reprogrammed) by the application of voltage pulses. This can be effectively done by the so-called resistive random-access memory (RRAM) devices, also known as memristors (for "memory-dependent resistors"). These devices have

*marcelo.rozenberg@u-psud.fr

been intensively investigated and developed over the past decade. The basic physical mechanisms for the resistive change in these oxide memristive devices are now understood in a certain amount of detail [3]. They are mainly based on achieving structural changes by inducing oxygen or metallic-ion migration [4,5]. After more than a decade of intensive research, these types of devices are beginning to reach the market.

This significant progress enables the implementation of a "synapse analog" [6,7]. Unfortunately, progress in research into and the implementation of a "neuron analog" has not been comparable. Hence, their development is becoming urgent.

In the context of resistive-change systems, we may mention a few important contributions toward devices with a "neuron"-inspired functionality: one is the "neuristor," which realizes a functionality similar to the Hodgkin-Huxley model for electric-pulse propagation in biological axons [8,9]. Another example is the self-oscillating spiking neuron, which is based on thermodynamic interactions in materials with negative differential resistance [10–12]. One more example, which concerns us more directly here, is the "artificial neuron," which implements the functionality of a leaky-integrate-and-fire (LIF) biological spiking neuron [13]. Remarkably, all these implementations are based on Mott insulators [8,14], which are strongly correlated systems that exhibit metal-insulator transitions [15].

Here, we shall focus on the latter of these types of artificial neurons. More specifically, we shall gain insight into a key feature of the LIF functionality, namely the "leaky" property, and we shall study the recovery behavior of the Mott neuron right after the firing event, which corresponds to the "refractory period" of a biological neuron. In the LIF model, a neuron is excited by a train of incoming electric pulses (spikes), which are *integrated* in time by the neuron, until it reaches a threshold level where it *fires* a spike. The "firing event" in the Mott neuron corresponds to the electric Mott transition (EMT), which is the resistance collapse of the Mott insulator. The *leaky* property models the fact that in the "down" time in between pulses, the excitation level of the neuron slowly decays, with a given relaxation-time constant. Evidently, for a neuron to reach a threshold and fire, the time delay between arriving pulses has to be shorter than the relaxation-time constant of that neuron [14]. Therefore, the control of the "leaky time" is a key issue. On the other hand, the "refractory time" of a biological neuron refers to the period of time *right after* the firing event, during which the neuron cannot fire again. As we shall see below, in our artificial Mott neuron model, both features are related to the same underlying mechanism, which leads to two characteristic relaxation times of the system.

One may naively expect that the relaxation time would be an intrinsic constant of the material; however, the situation is more complex since the artificial neurons are based on strongly correlated Mott materials that undergo insulator-metal phase transitions. In systems in equilibrium, Mott transitions can be induced and controlled by a variety of parameters, such as pressure, chemical doping, and temperature. However, in the present situation, the understanding of the relaxation is further complicated by the fact that the LIF functionality is a property of the system out of equilibrium; that is, when the EMT is induced by pulses of electric field that provoke a partial dielectric breakdown. During this breakdown, the resistance collapses [14,16] and then spontaneously recovers back to the original value some time after the application of the electric pulse is terminated. The theoretical understanding of the recovery of a Mott insulator from the collapsed resistance state is a challenging many-body out-of-equilibrium problem of actual interest [17].

EMT experiments show that the relaxation from the low to the original resistance state of a Mott insulator depends significantly on the strength of the applied electric field [18]. In fact, the systematic study of the field-induced dielectric breakdown in the Mott system $GaTa_4Se_8$ and similar compounds has revealed the existence of three different regimes: (i) at low applied electric fields there is, of course, no large resistance change; (ii) at intermediate field strength there is a *volatile* change of resistance, that is, there is a sudden collapse followed by a spontaneous recovery to the original insulating state; and (iii) at high fields there is a seemingly *nonvolatile* change of resistance, that is, the resistance collapses and does not spontaneously recover (recovery may take a very long time or be virtually infinite).

We should note that the resistive change in regime (ii) is a perfectly reproducible effect and the breakdown occurs after a delay time t_D (of the order of microseconds to milliseconds) that is strongly dependent on the applied voltage [18]. When the applied voltage pulse is terminated, the low resistance relaxes back to the original high value in a characteristic time that we shall denote by τ_R [13]. In contrast, in the nonvolatile regime (iii) the resistance change is permanent, i.e., $\tau_R \to \infty$. In this regime, the initial resistance state may still be recovered by either further pulse application or heating of the sample. Thus the resistive changes in this regime are more difficult to control and may involve structural modifications.

In previous works, we have introduced a model that captures some aspects of the basic phenomenology of the EMT [14]. However, a shortcoming of that model is that it has a single relaxation time (τ_R) for the recovery of the "broken-down" metallic regions back to the initial Mott-insulator state. Evidently, this feature cannot account for the dependence of τ_R on the applied-pulse strength in regime (ii) as is observed in experiments and that we shall report here. Thus, the main goal of the present work is to extend the previous model to capture the behavior of the relaxation time in the EMT experiments. We

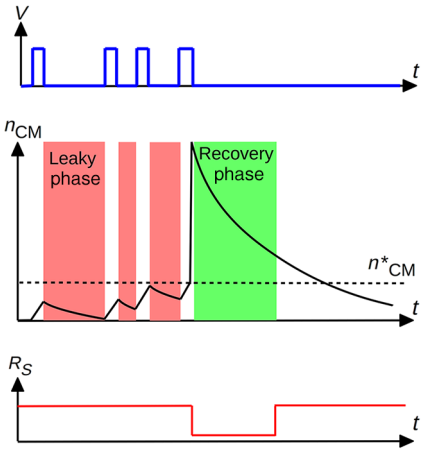

FIG. 1. The leaky and recovery phases are depicted in red and green, respectively. The first takes place in between pulses and the second after the EMT, that is, after the neuron fires.

shall show that by introducing an electroelastic effect, the relaxation time τ_R will display a nontrivial behavior, providing a qualitative explanation for the results obtained in the experiments.

Before introducing the model, we should clarify a few important points that connect the present study to a model of a neuron analog aimed at the implementation of bioinspired neural networks. As shown in Fig. 1, the *leaky time* [13] is associated with the dynamics of the neuron in between the arriving pulses *before* the firing. Experimentally, this leaky stage may be explored and characterized by the application of short electric pulses before the resistance collapse (i.e., the *firing event*) [13]. In contrast, the recovery from the low-resistance state, which begins when the pulse application is terminated (see Fig. 1), should be associated with the "refractory time" of a biological neuron. However, we should mention that there is a significant difference between the refractory time of a biological neuron and the Mott neuron. In the former, the neuron is "off" during the refractory time, that is, arriving spikes cannot induce a new action potential. In contrast, in the Mott neuron the system is relaxing from the low-resistance state, which means that arriving pulses may still generate current (fire) pulses through the device. Because of this difference, we term the recovery of the high-resistance state as the "recovery time" instead of the refractory time (see Fig. 1).

As we shall see from study of the simulation results of our model, the leaky and recovery times are associated with different microscopic physical situations. The leaky process corresponds to the relaxation of *isolated* metallic regions in an essentially insulating system, while the recovery period corresponds to the rupture or depercolation process of already formed metallic filaments. As we shall show in the experiments that we report here, our modeling work provides a useful insight into the control of these relevant neuronal model features.

The paper is organized as follows. We introduce our model in Sec. II. We describe its main features and we make a brief comparison with our previous model, presented in Ref. [14]. In Sec. III, we analyze the process of filament formation and growth (related to the leaky-integration period). First, we study the dynamics of the filament formation and the origin of the threshold voltage. Then we analyze the subsequent thickening process of the formed metallic filaments. In Sec. IV, we study the relaxation process of the low-resistance state back toward the insulating state. We see that the inclusion of electroelastic effects can induce dramatic changes in the relaxation times. In Sec. V, we analyze the relaxation process under the action of external subthreshold voltages. We compare simulations with our latest experimental results, which provide further validation for our model. Finally, in Sec. VI, we analyze the dependence of the recovery time on the strength and duration of the external applied voltage. We analyze two different voltage protocols (a single pulse and a train of pulses) and a comparison between experiments and simulations is presented for each case. The protocol of a train of pulses bring us closer to the realm of spiking neurons.

II. THE MODEL

We start from the model introduced by Stoliar *et al.* [14], which consists of a two-dimensional (2D) resistor network, where each element of the network represents a small (nanoscale) region of the physical Mott system [14,19]. The cells of the network are assumed to be large enough so that its electronic state is well defined. Imaging technique experiments with nanoscale spatial resolution have shown that across the Mott transition there is a coexistence of metallic and insulating phases with inhomogeneous distributions down to the few nanometers [20,21]. Each cell site is assumed to be in one of two electronic states: the Mott-insulator (MI) state or the correlated-metal (CM) state. These states are associated, respectively, with high and low resistance values, $R_{\rm MI}$ and $R_{\rm CM}$. The resistor-network model is shown schematically in Figs. 2(a) and 2(b). Since the experimental systems are normally in the insulator state, the model assumes that the MI state is the lowest in energy, which is defined as the reference $E_{\rm MI} = 0$. The CM state is assumed to be a *metastable* state, with a higher energy $E_{\rm CM}$, and separated from the MI state by an energy barrier E_B^E of purely electronic origin [see Fig. 3(a)]. From dynamical mean-field theory calculations of the Hubbard model for Mott systems [22,23], one may expect that the energy difference between these states may be of the order of a few tens of millielectronvolts, substantially smaller than the electronic bandwidth.

In the present numerical study, we follow the methodology of Ref. [14]. By analogy with the actual experimental setup, we adopt an electric circuit with a

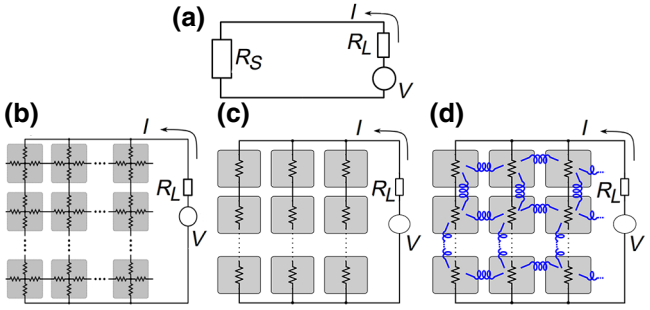

FIG. 2. (a) The circuit diagram of the system. (b) A diagram of the 2D resistive network of the model, from Ref. [14]. (c) A diagram of the 1D filament approximation: all transverse currents have been neglected. (d) A diagram of the 1D approximation with local elastic interactions (represented by the blue springs).

constant *load* resistance R_L, as shown in Fig. 2(a). Thus, the resistances of the system, R_S and R_L, form a voltage-divisor circuit. If we denote the externally applied voltage protocol by $V(t)$, then the voltage in the resistor network is given by $V_S(t) = (R_S(t)/(R_S(t) + R_L))V(t)$. The external voltage can be set as any arbitrary profile $V(t)$, which upon discretization of the time becomes $V(t_i)$ at each time step t_i. Initially, all lattice cells are assumed to be in the MI state. Then, at every time step t_i, the full resistor-network circuit is solved, obtaining all the local voltage drops ΔV_{ij} at each cell site ij, and the total resistance $R_S(t_i)$ between the top and bottom electrodes. For simplicity, the electrodes are assumed to be perfectly metallic [see Fig. 2(b)] [14].

At each time step, the state of the resistor network is updated following the procedure of Ref. [14]. The probability for a given insulator cell to undergo a local EMT—that is, from MI to CM at time step t_i—is given by $P_{\mathrm{MI}\to\mathrm{CM}} = \nu e^{-(E_B^E - e\Delta V(t_i))/k_BT}$, where the constant ν is an attempt rate, e is the charge of the electron, $\Delta V(t_i)$ is the computed local voltage drop at the given cell, k_B is the Boltzmann constant, and T is the temperature. This is

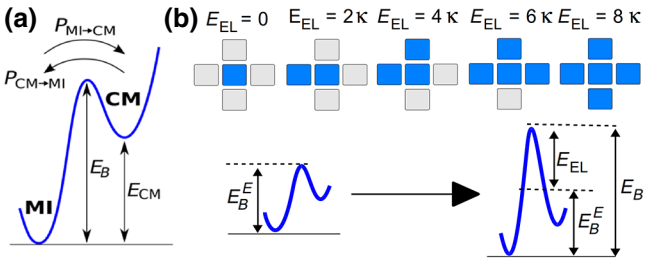

FIG. 3. (a) The energy landscape of the model. The Mott-insulator (MI) and correlated-metal (CM) states are separated by the energy barrier E_B. (b) The height of the barrier is affected by local structural distortions according to the elastic energy cost of Eq. (2). We show the different possible configurations for a first-neighbor interaction with their respective energy values. Here, κ corresponds to the elastic constant [see Eq. (2)].

an Arrhenius-like law, where the key feature is that the local electric field increases the probability that the cell will undergo a local EMT. This assumption should (hopefully) be fully supported by out-of-equilibrium many-body calculations.

A crucial ingredient of the model is the metastability of the CM state. In fact, if a given cell undergoes a MI $\to$ CM transition, then the metastable CM state may relax back to the MI one. The transition rate for the relaxation is also given by an Arrhenius-like expression $P_{\mathrm{CM}\to\mathrm{MI}} = \nu e^{-(E_B^E - E_{\mathrm{CM}})/k_BT}$; hence the model has a fixed relaxation time τ_R. We note that in this relaxation process, the term $e\Delta V$ may be omitted since $R_{\mathrm{CM}} \ll R_{\mathrm{MI}}$, so the voltage-drop effect is neglected at the CM sites.

Thus, in the numerical simulation, the 2D resistor network is solved at each time step t_i for a given protocol $V(t_i)$. The local voltage drops $\Delta V_{ij}(t_i)$ are computed for all cells ij. The state of the cells is updated according to the above probabilities. This gives a new value of the total (two-point) resistance $R_S(t_i)$ of the system and then the simulation proceeds. For simplicity, values of energies throughout the paper will be expressed in units of k_BT, voltages in units of k_BT/e, and times in units of ν^{-1} (or, equivalently, in simulation steps).

The previous numerical studies [14,19] have shown that the resistive collapse in the EMT is due to the sudden formation of filamentary structures of the CM phase, which connect the two electrodes. These structures grow, as expected, along the electric-field lines. Less evident has been the observation that while right after the percolation the filament connecting the two electrodes is quite inhomogeneous, it nevertheless continues to evolve until it rapidly reaches a rather homogeneous thickness [14]. This is illustrated in Fig. 4(a). This feature suggests that the transverse currents, i.e., those perpendicular to the direction of the applied field, may not play a crucial role, at least after some irrelevant brief transient time. Since the numerical solution of the 2D resistor network is by far the most time-consuming part of the numerical simulation, motivated by the previous observations, we numerically explore the possibility of neglecting the transverse currents altogether in our simulations, which would dramatically boost the computational speed. In Figs. 4(b)–4(d), we show a comparison of typical experimental data of resistive switching (RS) of the EMT for various applied external voltages and the respective results of the simulations for the 2D model [panel (c)] and for the case where the transverse currents are neglected [panel (d)]. We see that the latter simulations capture the behavior qualitatively, in a very similar manner compared to the 2D model [14]. This simplification leads to a dramatic speed-up of the simulations. In fact, instead of solving a 2D resistor matrix, we simply have a collection of one-dimensional (1D) series resistor chains, which is trivially solved. This enables us also to consider significantly larger systems and, more importantly,

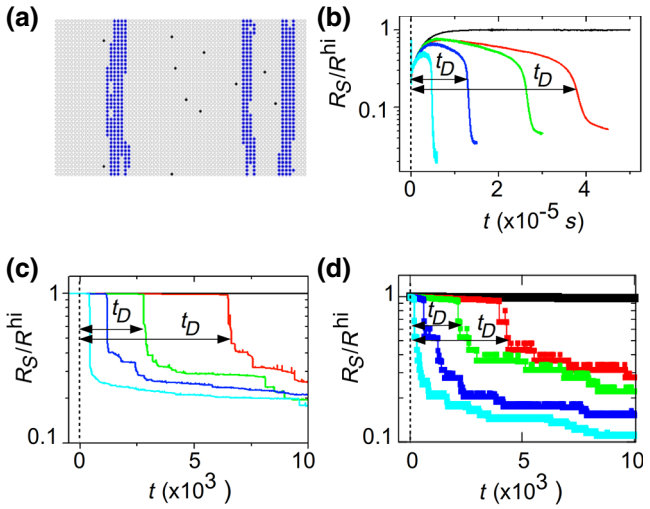

FIG. 4. (a) A snapshot of the simulated 2D resistive network after EMT, adapted from Ref. [14]. The black and blue dots indicate the sites that are in the CM state, while the gray dots correspond to the sites in the MI state. The dots forming filaments are colored for easier visualization. (b) Experimental RS curves for different applied voltages in a $GaTa_4Se_8$ device, showing the EMT or resistance collapse at a delay time t_D. The curves correspond to voltage values of 6, 40, 44, 56, and 86 V (from right to left in the plot). The delay time t_D is indicated in the plot for two different curves. We observe that the first curve at $V = 6$ V is beneath the threshold voltage and does not induce the EMT. The duration of the pulses is limited to 50, 45, 30, 15, and 6 μs, respectively. The experiments are performed at a temperature of $T = 77$K (for the experimental details, see Sec. VII). The curves show an initial transient increase due to capacitive effects. (c) Numerical simulations from the original 2D model, adapted from Ref. [19]. The simulations are performed for a 40×128 network ($N \times W$) with $E_B^E = 20$, $E_{CM} = 10$, $E_{MI} = 0$, $R_L = 0.1443$, $R_{CM} = 0.3R_L$, and $R_{MI} = 16.41R_L$. The curves correspond to V values of 400, 600, 650, 700, and 750, respectively (from right to left). (d) Numerical simulations from the 1D filament approximation: the model parameters are as in (c). The curves correspond to V values of 300, 380, 400, 450, and 500, respectively (from right to left). As mentioned in the text, energies are expressed in units of k_BT, voltages in units of k_BT/e, and times in units of ν^{-1}. The approximation provides similar qualitative behavior and enables a speed-up of the calculations by a factor of 60. This allows us to perform a systematic study of the long-time relaxation behavior.

to simulate for longer times. This turns to be crucial for the systematic study of relaxation effects, which are relatively much longer than the typical time to dielectric breakdown. We shall come back to this approximation later, when we introduce the electroelastic interactions.

This model has already provided valuable insights and very good qualitative agreement with experiments, such as the existence of a threshold field and the decrease of the delay time of the resistance collapse with the increase of the applied voltage [14]. However, as it stands, if fails to capture the existence of different relaxation regimes (ii) and (iii) as described in Sec. I. In fact, the model has a single intrinsic relaxation time, which is solely set by the energy difference between the electronic energy barrier and the metastable CM state, $\Delta E = E_B^E - E_{CM}$.

Thus, in order to qualitatively capture the observed experimental behavior where strong voltage pulsing may realize low-resistance states that are long-lived, we shall extend the model by making an additional assumption. Specifically, we shall introduce the energy cost associated with the formation of a MI-CM boundary interface. We may rationalize this as an elastic energy that originates from the structural strain created at the boundary between the two distinct electronic phases. In fact, experimentally, using scanning tunneling microscopy (STM) techniques, a large electromechanical effect has been observed, which was interpreted as a strong self-compression of the lattice structure in the metal state [21]. Therefore, we add to the model the assumption of an elastic deformation energy E_{EL}, which is a supplementary energy cost associated with the transition when two nearest-neighbor cells would end in different states. In other words, it is an energy penalty for creating spatial inhomogeneities. As we shall see, this assumption creates *spatial correlations* in the way in which multiple filaments grow. In our previous model, filaments would grow at random, in essentially uncorrelated positions, across the system. However, the addition of the elastic energy promotes the grow of new filaments neighboring previous ones, thus promoting the thickening of filamentary structures.

The simplest way to include this in the model is by adding this effect into the definition of the energy barrier.

Thus, we may write the energy of the barrier E_B of the ith site of the lattice as follows:

$$E_B(i) = E_B^E + E_{EL}(i), \tag{1}$$

with the elastic energy cost being

$$E_{EL}(i) = \kappa \sum_j (1 - q_i' q_j), \tag{2}$$

where the variable $q_i = \pm 1$ indicates the current state of the ith cell (say, MI $= 1$ and CM $= -1$), q_i' is the proposed new state for the ith cell (i.e., $q_i' = -q_i$), the index j of the sum runs over the number of nearest neighbors Q, and κ is an elastic constant.

Thus, if initially the ith cell is in the same state as all of its neighbors, then the barrier to overcome is (maximally) increased from the electronic value E_B^E by the amount of $E_{EL} = 2\kappa Q$ corresponding to the maximal mismatch. If, on the contrary, the ith cell is initially in a different state with respect to all of its neighbors, the barrier to overcome is solely the electronic one, as $E_{EL} = 0$, since all cells would end up in the same phase. For simplicity, we shall consider here that the range of the elastic term is only to the nearest

neighbors, so that in the present 2D lattice $Q = 4$. There is, of course, freedom to choose longer-range interactions. We shall come back to this point later.

As we shall see, the low-resistance states may become long-lived as the filamentary structures grow thicker. Thus, for the sake of clarity in the presentation of our results, we shall first dedicate the next section to describing the formation of CM filamentary structures and how their thickness may grow under strong applied voltages. Then, in the following section, we shall describe how the relaxation time may be significantly affected by the inclusion of the elastic deformation term and how it may qualitatively account for the various EMT regimes discussed earlier.

Here, we should come back to the issue of the approximation based on neglecting the transverse currents. This simplification allows us to avoid the numerically demanding task of solving the 2D resistor-network array. While this was not a problem for the noninteracting cells, one may wonder if this assumption remains reasonable in the presence of intercell elastic interactions. In fact, elastic interactions penalize the growth of inhomogeneities, which emerge in the initial stages of the filamentation. Thus, we may expect that this approximation is less reliable at the onset of filament formation. For instance, we may expect our estimates of the voltage threshold value to be somewhat overestimated. However, the main qualitative features of the process should remain unchanged. Indeed, the elastic cost of inhomogeneities (2) is bounded by the maximum mismatch that we quoted above, $2\kappa Q$, while the electric energy of the field can be arbitrarily increased by augmenting the applied voltage. Hence, the filaments will eventually form even in the presence of the elastic interactions. Our approximation becomes better at higher applied fields, where the relative importance of elastic terms diminishes. Finally, the 1D approximation does not concern the study of the relaxation, since that process occurs at zero applied voltage and current.

III. FILAMENT FORMATION AND GROWTH

The EMT proceeds through two distinct stages: the first is the creation of low-resistance filamentary structures that interconnect the electrodes; the second is the successive reabsorption of the filaments as the system recovers the high-resistance state. The first of these stages occurs under the application of a strong external voltage, while the second takes place when the applied voltage is switched off. The present section is devoted to describing the first process.

A. Dynamics of filament formation

The formation of the filaments under the application of an external voltage in the present model has already been qualitatively described in Ref. [14]. The main features are the existence of a threshold field for the formation of filaments, the existence of a delay time t_D before the EMT takes place, and that the number of filaments (or the fraction of phase change in the system) grows with the magnitude of the applied field via a process of subsequent formation of multiple filaments. In this section, we will provide an analytic insight into these key features, which are crucial for future applications. As we will see, the inclusion of the elastic term does not change the main basic mechanisms.

We shall now consider in detail the formation of the filament from a dynamical-system point of view. This will allow us to derive approximate analytic expressions for the delay time t_D that signals the EMT and the collapse of the resistance of the insulating state. The following will be a mostly mathematical detour. Readers not interested in the details of the derivation may skip to Eqs. (5) and (8).

The formation of filaments is a sudden event; thus, *previous* to the EMT breakdown, the system still has the vast majority of cells in the initial MI state. Hence the system before the rupture is quite homogeneously MI. This implies that the role of the elastic term before the rupture is essentially to renormalize the value of the electronic barrier height E_B^E to $E_B \approx E_B^E + 2\kappa Q$ for the MI sites, which upon transition to the CM state find themselves surrounded by neighboring MI sites. In contrast, for any site that has already become CM, under relaxation to the MI state it does not see any elastic barrier, since all its neighbors are MI. This then allows us to understand that the elastic term just produces slight changes in the value of the threshold voltage $V_{\rm th}$ and the time t_D [14].

One may gain an analytic insight into the process of the initial filament formation and the threshold behavior by considering the dynamical behavior given by the deterministic partial differential equation that follows from the probability rates of our model in a continuum limit. The production rate of CM cells is given by the difference between a gain and a loss (leaky) term, with their respective probabilities:

$$\frac{dn_{\rm CM}}{dt} = [n_{\rm MI}P_{\rm MI\to CM}] - [n_{\rm CM}P_{CM\to MI}] \approx n_{\rm MI}e^{-[E_B^E+E_{\rm EL}^{\rm MI}-\Delta V(n_{\rm CM})]} - n_{\rm CM}e^{-(E_B^E+E_{\rm EL}^{\rm CM}-E_{\rm CM})}, \tag{3}$$

where $n_{\rm CM}$ and $n_{\rm MI}$ are the time-dependent numbers of cells in the CM and MI states, respectively. N is the number of cells between the top and bottom electrodes, i.e., the linear longitudinal dimension of the system. The elastic energy terms in the model depend on the specific local configuration, so for the present analysis we have approximated them by the parameters $E_{\rm EL}^{\rm CM}$ and $E_{\rm EL}^{\rm MI}$, which would correspond to the average elastic energies of a CM and a MI cell, respectively. The voltage $\Delta V(n_{\rm CM})$ is the voltage drop on the MI cells, $\Delta V \approx V/(N - n_{\rm CM})$. Since $R_{\rm MI} \gg R_{\rm CM}$, the total applied voltage V essentially

falls on the high-resistivity cells MI, with $n_{\mathrm{MI}} = N - n_{\mathrm{CM}}$. Hence, the gain transition rate increases exponentially with increasing n_{CM}.

The rate equation has a stationary solution when $dn_{\mathrm{CM}}/dt = 0$. It is given by the balance between the gain and the loss term. We shall consider the existence of such a solution as a function of the external voltage V. A geometric solution of the problem is shown in Fig. 5, where we plot the gain and loss terms as a function of the variable n_{CM} at increasing values of V and all other parameters are left fixed. The way to interpret this plot is as follows: at any given value of n_{CM}, either the gain (red) or the loss (blue) rate is greater. If the gain is greater, then n_{CM} will evolve to the right, that is, to higher values. If the loss is greater, it evolves to the left. The equilibrium of n_{CM} is where the gain and loss transition rates are equal. We observe in the two top panels of the figure that there are two points where the gain and loss terms cross. However, it is not difficult to see that only the lowest one (black dot) corresponds to a stationary *stable* equilibrium state, while the higher one (empty dot) is an *unstable* equilibrium point. Varying the parameter V, there is a critical value at which the two crossing points "collide," at n^*_{CM}. The interpretation of this critical V is the finite threshold value V_{th}. For higher values of $V > V_{\mathrm{th}}$, as shown in the last panel, there is no stationary state for n_{CM}, which means that the gain rate is never compensated by the loss rate and there is a runaway evolution of n_{CM} (indicated by the arrows). This instability signals the formation of the first filament. This process is a not uncommon situation found in dynamical analysis and is termed a saddle-node bifurcation [24]. For further details, an analysis of the numerical integration of the rate equation is shown in the Supplemental Material (see Ref. [25]). We shall now derive two approximate expressions for the delay time t_D. One is simpler than the other and is valid at shorter delay times and higher applied V, while the other is for longer times and for V closer to V_{th}.

If the parameters of the model are such that $n^*_{\mathrm{CM}} \ll N$, then one may approximate the rate Eq. (3) as follows:

$$\frac{dn_{\mathrm{CM}}}{dt} \approx Ne^{-(E_B^E + E_{\mathrm{EL}}^{\mathrm{MI}} - V/N)} - n_{\mathrm{CM}} e^{-(E_B^E + E_{\mathrm{EL}}^{\mathrm{CM}} - E_{\mathrm{CM}})}. \quad (4)$$

If $V > V_{\mathrm{th}}$, then an avalanche event will occur, as described above. Before the avalanche can take place, there is a period during which the evolution of the system (i.e., of $n_{\mathrm{CM}}(t)$) is dominated by the slow dynamics in the proximity of the former stationary point, where $n_{\mathrm{CM}}(t) \approx n^*_{\mathrm{CM}}$. This is the region where the gain- and loss-rate curves are approximately parallel (the red and blue curves in Fig. 5). Once the evolution of n_{CM} overcomes this regime, it rapidly increases to N, i.e., the filament forms with an avalanche. Thus, we may neglect the duration of the sudden event for the estimation of the delay time for the filament formation and obtain an approximate expression

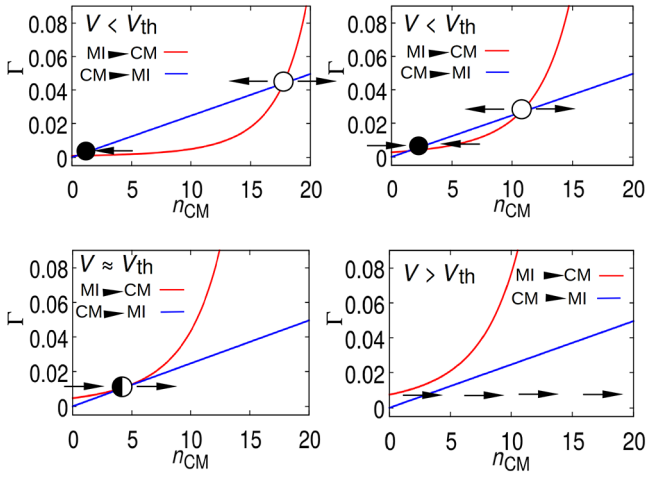

FIG. 5. The dynamical analysis of the model: EMT rates (Γ) according to Eq. (3) as a function of n_{CM}. The red curves correspond to the gain rate [the first term of Eq. (3)] and the blue curves correspond to the loss or "leaky" rate [the second term of Eq. (3)]. We show the results for four different applied voltages (increasing from left to right and from the top to the bottom panels). For $V < V_{\mathrm{th}}$ (top panels), the system exhibits two equilibrium points (where the two curves cross each other). At $V = V_{\mathrm{th}}$ (bottom left), the two equilibrium points "collide" at n^*_{CM}, while for $V > V_{\mathrm{th}}$ (bottom right), no crossing points are present and no equilibrium can be achieved. The black dots indicate the stable ("attractor") solution, the empty dots the unstable ("repeller") solution, and the half-filled dots the collapse of both solutions. In the last panel, the arrows indicate the divergent dynamic. This evolution is known as a "saddle-node bifurcation."

by integration of Eq. (4) from 0 to n^*_{CM}:

$$t_D(V) = \frac{1}{B} \ln \left[\frac{A(V)}{A(V) - n^*_{\mathrm{CM}} B} \right], \quad (5)$$

where $A(V)$ is the first term on the right-hand side of Eq. (4), B is the exponential factor in the second term of the same equation, and n^*_{CM} is assumed to be a constant.

We now consider the region $V \gtrsim V_{\mathrm{th}}$, where the previous approximation is not sufficient. To describe the region close to the threshold, we need to more accurately describe Eq. (3) by expanding around the bifurcation point $(n^*_{\mathrm{CM}}, V_{\mathrm{th}})$. Thus, defining $dn_{\mathrm{CM}}/dt \equiv f\,(n_{\mathrm{CM}}, V)$, and using the fact that at the bifurcation point $f\,(n^*_{\mathrm{CM}}, V_{\mathrm{th}}) = 0$ and $(\partial f\,/\partial n_{\mathrm{CM}})(n^*_{\mathrm{CM}}, V_{\mathrm{th}}) = 0$, we get, to the lowest order:

$$\begin{aligned} \frac{dn_{\mathrm{CM}}}{dt} \approx\ & \bar{A}(n_{\mathrm{CM}} - n^*_{\mathrm{CM}})^2 \\ & + \bar{B}(V - V_{\mathrm{th}})(n_{\mathrm{CM}} - n^*_{\mathrm{CM}}) \\ & + \bar{C}(V - V_{\mathrm{th}}), \end{aligned} \quad (6)$$

where

$$\bar{A} = \frac{\partial^2 f}{2\partial n_{\rm CM}^2} = \frac{V_{\rm th}^2 e^{-[E_B^E + E_{\rm EL}^{\rm MI} - V_{\rm th}/(N-n_{\rm CM}^*)]}}{2(N-n_{\rm CM}^*)^3},$$
$$\bar{B} = \frac{\partial^2 f}{\partial V \partial n_{\rm CM}} = \frac{V_{\rm th} e^{-[E_B^E + E_{\rm EL}^{\rm MI} - V_{\rm th}/(N-n_{\rm CM}^*)]}}{(N-n_{\rm CM}^*)^2}, \quad (7)$$
$$\bar{C} = \frac{\partial f}{\partial V} = e^{-[E_B^E + E_{\rm EL}^{\rm MI} - V_{\rm th}/(N-n_{\rm CM}^*)]},$$

in which the derivatives are evaluated at $(n_{\rm CM}^*, V_{\rm th})$. Integrating, we obtain for t_D,

$$t_D = \left. \frac{2a \tan\left(\frac{[2\bar{A}(n_{\rm CM} - n_{\rm CM}^*) + \bar{B}(V - V_{\rm th})]}{\sqrt{4\bar{A}\bar{C}(V-V_{\rm th}) - \bar{B}^2(V-V_{\rm th})^2}} \right)}{\sqrt{4\bar{A}\bar{C}(V-V_{\rm th}) - \bar{B}^2(V-V_{\rm th})^2}} \right|_{n_{\rm CM}=0}^{n_{\rm CM}^*}. \quad (8)$$

Equations (5) and (8), which we have just derived, provide approximate expressions for the dependence of the time delay t_D on the model parameters and the applied voltage.

The validity of the approximations can be tested on our numerical simulations. This is shown in Fig. 6, where we observe that the two expressions produce rather good fits of the numerical data of t_D that span several orders of magnitude. We plot the data as a function of the voltage applied to the sample (V_S) and of the initial voltage drop in the cells $\Delta V_0 = V_S/N$. This last magnitude, which is proportional to the electric field, is the one that enters into the model that consists of Eqs. (3) and (4) and is the one that directly affects the effective barrier height and is ultimately responsible for inducing the local EMT. To obtain the voltage in physical units, V (or V_S) has to be multiplied by k_BT/e; and for the field V_S/N, it has to be multiplied by $k_BT/e\Delta X$, where ΔX is the size of a cell (see Table I).

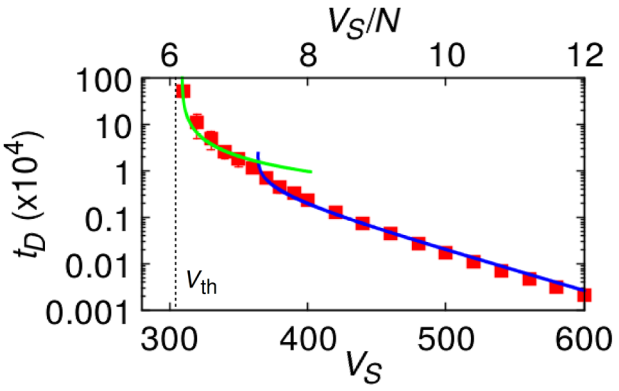

FIG. 6. The commutation delay time t_D as a function of the external voltage applied on the sample (V_S, bottom axis) and the associated initial voltage drop in the cells ($\Delta V_0 = V_S/N$, top axis) obtained from our model simulations. We show the fits according to Eq. (5) (solid blue line) and Eq. (8) (solid green line), the latter for voltages close to $V_{\rm th}$ and the former for the rest of the curve. The fitting parameters are shown in Table I. The simulations are performed on a 50×140 network with $E_B^E = 11$, $E_{\rm CM} = 5$, $E_{\rm MI} = 0$, $R_L = 500$, $R_{\rm CM} = 200$ $(0.4R_L)$, $R_{\rm MI} = 200 \times 10^3$ $(400R_L)$, and $\kappa = 1$. The same model-parameter values will be used for all simulations in this work, except where specifically indicated.

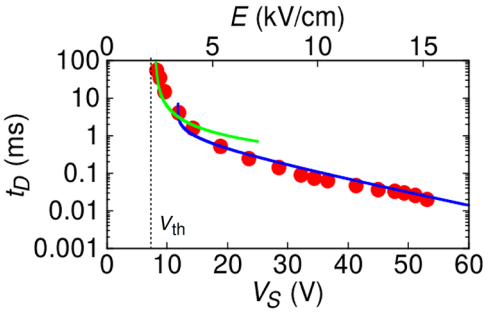

FIG. 7. The commutation delay time t_D as a function of the external voltage applied on the sample (V_S, bottom axis) and the associated electric field (E, top axis) obtained from experiments on a $GaMo_4S_8$ sample. The local electric field is calculated as V_S/l, where $l = 35$ μm is the thickness of the sample (distance between electrodes). As in the simulated case (Fig. 6), we present the fits according to Eqs. (5) (solid blue line) and (8) (solid green line), the latter for voltages close to $V_{\rm th}$ and the former for the rest of the curve. The fitting parameters are summarized in Table I.

We may further use these validated expressions to fit actual experimental results performed in $GaMo_4S_8$ at 74 K. These are shown in Fig. 7 and the fitting parameters are summarized in Table I. By analogy with the simulated case, the results in Fig. 7 are given in terms of the applied voltage V_S and the electric field $E = V_S/l$, where $l = 35$ μm is the distance between electrodes.

We note that the parameters obtained for the experiments are in the order of 100 meV (equivalent to a temperature of $T \approx 1100$ K), indicating that a well-defined EMT should still be observed at room temperature. This is of crucial importance for potential technological applications. The exploration of the temperature dependence of the EMT is beyond the scope of the present study.

B. Filament growth

We now turn to describing one of the features introduced by the elastic term in our model, namely the thickening of the filament with time. The width of the initial filament is just one cell. Thus, after it forms, the resistance of the system R_S drops from a high value of $R_S^{\rm hi} \approx R_{\rm MI}N/W$ to a lower value of $R_S^{\rm lo} \approx [R_S^{\rm hi}//R_{\rm CM}N]$, where W is the width of the resistor network and N is the thickness (i.e., the

TABLE I. The parameters used for the fits on the delay time from Figs. 6 (simulations) and 7 (experiments). The free parameters used in each case are indicated in bold. For reference, the fitting parameters of the simulated data are also given in physical units, written within parentheses. Physical units of energy are obtained by multiplying by k_BT and for voltages by k_BT/e. We assume $T = 74$ K as in the experimental case. For the experimental fit, we use $V_{\rm th} = 8\ V$ (see Fig. 7). The parameter N extracted from the fit (the number of cells) corresponds to a physical cell size of $\Delta x = l/N = 35\ \mu$m$/2000 = 17.5$ nm (where l is the distance between electrodes), which is of the order of magnitude of the domains observed in STM (see Fig. 8).

	Simulations		Experiments	
	Low-voltage region ($V \approx V_{\rm th}$)	High-voltage region	Low-voltage region ($V \approx V_{\rm th}$)	High-voltage region
E_B^E	11 (70.15 meV)	11 (70.15 meV)	**110 meV**	**100 meV**
$E_{\rm CM}$	5.0 (31.85 meV)	5.0 (31.85 meV)	**15 meV**	**15 meV**
$E_{\rm MI}$	0	0	0	0
$E_{\rm EL}^{\rm MI}$	**7.8 (49.7 meV)**	**7.5 (47.8 meV)**	**10 meV**	**10 meV**
$E_{\rm EL}^{\rm CM}$	**2.0 (12.7 meV)**	**2.0 (12.7 meV)**	0	0
$\frac{n^*_{\rm CM}}{N}$	**0.07**	**0.04**	**0.08**	**0.05**
N	50	50	**2000**	**2000**
T	74 K	74 K	74 K	74 K
$V_{\rm th}$	**309.0 (1.97 V)**	...	8.0 V	...
ν^{-1}	1	1	**1 μs**	**1 μs**

distance between the electrodes) as before. The decrease in R_S provokes a reduction of the voltage applied on the network according to $V_S = (R_S/(R_S + R_L))V$, where R_L is the load resistance. Hence, the new value of V_S may be larger or smaller than $V_{\rm th}$. In the former case, the system will undergo a second filament formation, again according to Eq. (5) but for a new (lower) value of V_S. After the formation of the second filament, V_S will drop further and the process of filament formation will continue until $V_S \approx V_{\rm th}$. This feature was already present in the previous model without an elastic term. In that case, the position at which the filaments form was essentially random and did not have any spatial correlation.

In contrast, in the present model the elastic-energy term favors the growth of adjacent filaments or, in other words, it favors the thickening of the filaments. This is easy to see, since an isolated filament requires a phase change of the cells into the CM state, while being surrounded by the other MI phase. That implies a maximal local elastic cost (see above). However, if a filament is already in place, the phase change of a cell adjacent to it is less penalized and hence has a higher probability of occurring. This feature is seen in the panels of Fig. 8(a), where we show snapshots of the system before the formation of the first filament ($t < t_D$), just after its formation ($t \approx t_D$), and at a later stage ($t \approx 5t_D$). We see that both processes have occurred and more filaments are present but also that the filaments have grown thicker. Evidently, the higher the value of the elastic constant κ and/or the greater the range of the elastic interaction, i.e., Q, the more favorable this would be for the thickening process with respect to new individual filament formation. Results are shown for two different values of κ. In addition, the increase in κ may enhance the formation of CM clusters before the generation of filaments, which, in turn, may affect the value of t_D. As seen in Fig. 6, this last effect plays little role for $\kappa = 1$.

This feature may allow us to interpret the experimental results reported by STM on a cleaved surface of a $GaTa_4Se_8$ compound [18] after the EMT. The image data in Fig. 8(b) show the variations in the conductance map due, presumably, to the cross sections of the conductive filamentary structures that were created by the EMT (in red). The distribution of areas of these conductive cross sections, which we associate with the filament thickness in our model, are quantified in the histogram in Fig. 8(c). We can compare these data to our model simulation results. In the three panels of Fig. 8(d), we show the evolution of the histogram of the filament thickness distribution during the application of a continued pulse voltage. We observe that, consistent with our previous discussion, the histogram distribution slowly drifts to higher thickness and is in qualitative agreement with the experimental data. We note here that when typical thicknesses are greater than about $0.1W$ (W being the width of the sample), the nucleation of separated filaments has been shown to be a relevant effect (i.e., the creation of large filaments via the percolation of smaller ones). In the STM image, we observe cross sections with areas up to about $0.01A$ (where $A \approx 500 \times 500$ nm^2 is the total area), suggesting that this effect is not relevant for the experiments analyzed in this work [31]. Thus, we will restrain our study to filament thicknesses below 0.1 W.

The systematic thickness growth of filaments in our model is further characterized in Fig. 9. We show a color map [panel (a)] of the largest thickness of the formed filament as a function of the duration of the applied pulse voltage ($t_{\rm ON}$) and its magnitude (V). The two panels (b) and (c), respectively, show cuts of the color map at fixed

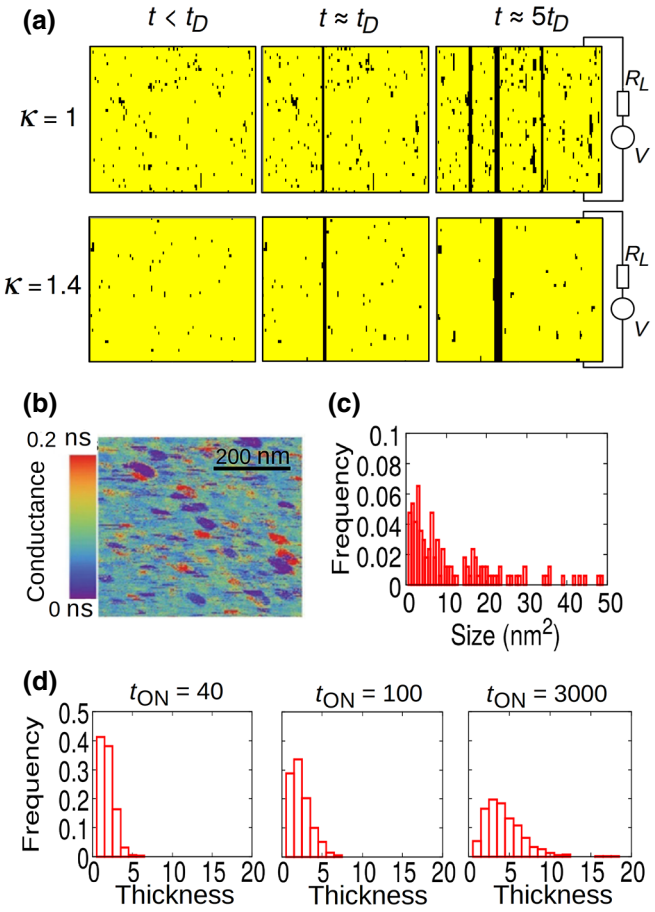

FIG. 8. (a) Snapshots of the simulated resistive map during the application of an external voltage ($V = 600$, $\Delta V_0 = V/N = 12$), which is maintained during the full duration of the simulation. The three panels illustrate the qualitative evolution before and at about and well after the formation of the first filament (i.e., the EMT). The black dots indicate the CM sites and the yellow ones the MI sites. We show the results for two different values of κ. The left panels show the system at a time before the formation of the first filament ($t < t_D$), the central panels correspond to times shortly after the first filament formation ($t \approx t_D$), and the right panels show the system at times much longer than t_D. Higher values of κ favor the thickening process. (b) An experimental conductive STM image of a cleaved $GaTa_4Se_8$ surface after the EMT (reproduced from Ref. [18]). The cleave provides a transverse section of the three-dimensional (3D) bulk sample. The spatial inhomogeneities indicated by the red regions may correspond to sections of metallic filamentary structures. The green and blue regions are insulating. (c) The distribution of the cross-section areas of the inhomogeneous metallic structures observed in the STM image of panel (b). (d) The distribution of the filament thickness from numerical simulations of our model. We show the distribution for three different pulse durations (t_{ON}). The applied V is as in panel (a). The data correspond to an average over 100 realizations.

V and fixed t_{ON}. The results are obtained as the mean over 100 realizations. The parameters used for the simulations are the same as in Fig. 6.

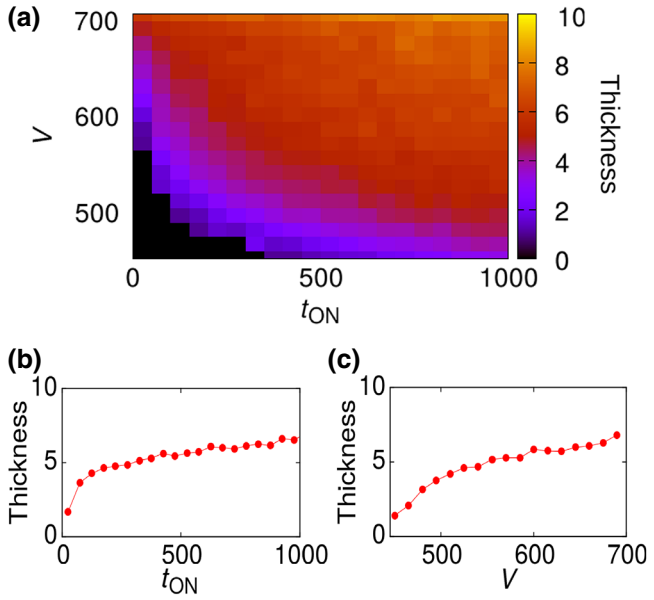

FIG. 9. Filament growth. (a) A color map of the filament thickness as a function of the applied voltage and the pulse length (t_{ON}). The results are obtained as the average over 100 realizations. The thickness is measured considering the thickest filament in each realization. The parameters used for the simulations are the same as in Figs. 6(b) and 6(c). Vertical and horizontal cross sections of panel (a): the maximum filament thickness as a function of (b) the pulse length and (c) the applied voltage. The latter is taken after a fixed $t_{ON} = 500$ and the former for fixed $V = 600$ ($\Delta V_0 = V/N = 12$).

IV. FILAMENT REABSORPTION: RETENTIVITY IN THE RECOVERY PHASE

We now turn to a central part of this study, which is the relaxation of the low-resistance state back into the Mott-insulator state after the applied voltage pulse is terminated. Physically, it corresponds to the reabsorption of the metallic filament back to the insulating phase. The proper description of this process is crucial for the understanding of the recovery phase in neuromorphic applications of the EMT for neuron-analog electronic devices [13]. We shall denote the time that the system requires to spontaneously relax back to the original insulator by τ_R. We also refer to this as the retention or recovery time.

The previous modeling of the EMT, which did not include the elastic term, predicted that the conductive low-resistance state always returns back to the high-resistance state with an essential constant relaxation-time rate. The cells in the CM metastable state relax back to the MI with a single characteristic time and they do that randomly. Therefore, the filaments disintegrate randomly as well. We shall now study how the addition of the electroelastic effect in the model modifies this feature.

In a low-resistance state, there are filamentary structures that are composed of a relatively high number of cells that

have switched into the metastable CM state. These structures are induced by an external applied voltage. After the voltage application is terminated, the CM cells start to relax back to the high-resistivity MI state. The relaxation time for the *total* number of cells to return to the MI state is long. However, this is not the physically relevant time for the experiments. In fact, the experimental observation of the relaxation of the system is through the behavior of the recovery of the resistance $R(t)$. Therefore, the important quantity to monitor in our model is when the filamentary structure of the CM cells breaks up or loses its percolation.

In our previous model, the cells forming the filament relaxed in a random fashion. In contrast, in the present case where we have added the electroelastic term in our model, we shall see that the filaments relax differently. Rather inversely to the case of filament growth by thickening, the dissolution of filaments occurs by thinning. In fact, as we shall see, the cells on the border of the filament have a higher probability of relaxing to the MI state compared to those in the interior of the filament.

In order to perform a systematic study of this process, we shall start from "artificially" well-defined initial states consisting of homogeneous filaments of cells in the CM state that we let relax back to the insulator state. We shall consider initial filaments of different thicknesses and then record the time it takes for them to break up. Since in our model this process is subject to statistical fluctuations, we average our results over a large number of realizations (typically 100). As mentioned before, we shall denote this time by τ_R, as it indicates the time for which a filament keeps its high conductive state until it relaxes to the insulator state. This quantity may also be denoted as the *retentivity*.

In Fig. 10(a), we show the results of our simulation studies for the dependence of τ_R on the initial filament thickness d, between $d = 1$ and $d = 10$. We observe that there are three different regimes: an initial exponential increase, which then becomes linear and finally saturates for higher values of d. Overall, τ_R increases by 2 orders of magnitude for a filament thickness of just a few units. The fast initial increase, of a full order of magnitude when comparing $d = 1$ to 2, follows from the presence of the elastic term. This renders much less probable the relaxation of cells in any one of the two adjacent filaments with respect to a single one. In the former case, any cell of a filament is surrounded by three other cells of the same type, while in the latter only by two (out of four). This effect enters into an exponential in the probability (see Sec. II). As $d > 2$, one may distinguish between central filament regions and border filament regions. Cells belonging to central filaments have all their 4 neighbors in the same state. Hence the eventual elastic mismatch maximally penalizes the relaxation. In contrast, cells in the external boundary of the filament only have three neighbors in the same state. Hence the boundary cells will relax

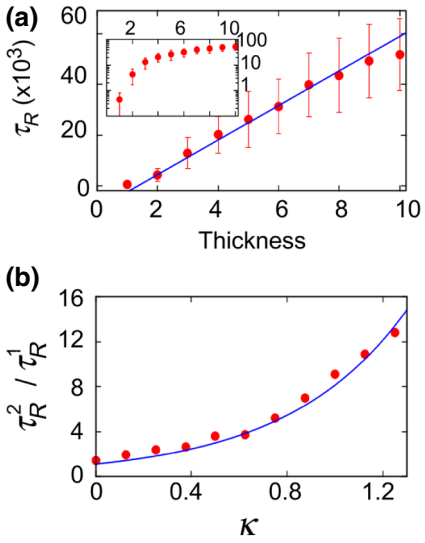

FIG. 10. (a) The retentivity time as a function of the filament thickness. The results are the average over 100 realizations. A fit according to Eq. (12) (solid blue line) is shown (for details of the fit, see the Supplemental Material [25]). Inset: a semilog-scale version of the main panel. A difference of about 1 order of magnitude can be seen between the first two points. (b) The retentivity ratio between filaments of thickness $d = 2$ and $d = 1$ (red circles) together with a fit according to Eq. (13): $f(\kappa) = A\exp(2\kappa)$, with A the only fitting parameter. The fit corresponds to $A = 1.1$. In these simulations, the initial state is a filament of a given thickness. No voltage pulse is applied. The system is left to evolve and one records the time it takes to break up.

faster. An internal cell would only acquire a good probability of relaxing by becoming part of the boundary, i.e., after the outermost neighboring cells relax. This feature leads to a continuous "thinning" effect of the filaments, and it progresses at an approximately constant rate, which results in the observed linear dependence with d. At higher values of d, a saturation effect becomes evident, as the thinning time starts to compete with the probability for relaxation of internal filament cells (i.e., with those inside a homogeneous CM phase).

We may easily estimate the relaxation time of a unitary filament from the loss term of the dynamical evolution of the cells of Eq. (3):

$$\frac{\partial n_{\mathrm{CM}}}{\partial t} = -n_{\mathrm{CM}} \exp[-(E_B^E + E_{\mathrm{EL}}^F - E_{\mathrm{CM}})], \qquad (9)$$

where E_{EL}^F is the elastic part of the energy barrier of the ith cell in the filament, which is assumed constant for all the cells in the same unitary filament. The rupture condition for a unitary filament is given by the relaxation of a single cell (since we neglect transverse currents), i.e., $\Delta n_{\mathrm{CM}} = 1$. A conductive individual filament initially has $n_{\mathrm{CM}} = N$. Therefore,

$$\Delta n_{\mathrm{CM}} = \Delta t N \exp[-(E_B^E + E_{\mathrm{EL}}^F - E_{\mathrm{CM}})] = 1 \qquad (10)$$

and we simply obtain

$$\tau_R^F = \frac{\exp[(E_B^E + E_{\rm EL}^F - E_{\rm CM})]}{N}, \tag{11}$$

where τ_R^F denotes the time for rupture of a unitary filament. From our previous qualitative discussion, we obtain that $E_{\rm EL}^F = 4\kappa$ for cells in an isolated filament of $d = 1$. Then $E_{\rm EL}^F = 6\kappa$ for cells in the outer boundary of filaments of $d \geq 2$ and $E_{\rm EL}^F = 8\kappa$ for cells in the interior of a thick filament ($d \geq 3$). The exponential dependence is explicit in Eq. (11).

Then, recalling from our discussion that the filaments relax by progressive thinning, we may simply estimate the total time for the full rupture of a filament of thickness d as follows:

$$\tau_R \propto d\tau_R^F, \tag{12}$$

where τ_R^F refers here to unitary border filaments. From the above equations, we may focus on the simple case with $d = 1$ and 2, which produces the largest relative retentivity gain. The ratio for the respective τ_R^d gives the following:

$$\frac{\tau_R^2}{\tau_R^1} = \exp[2\kappa(Q_2 - Q_1)] = \exp[2\kappa(3-2)], \tag{13}$$

which corresponds to the blue-solid-line fit of Fig. 10(b). Here, Q_d corresponds to the number of neighbors in the CM state for each thickness d. This expression highlights the exponential dependence on the elastic constant parameter κ.

We have just seen that according to the model, the duration of the metallic state may be (exponentially) long but always remains metastable. However, experimentally, the metallic states in region (iii), obtained by stronger pulsing, seem to be actually stable. One way to further increase the stability of the metallic (filamentary) state in the model is to increase the range of the elastic term. One may consider, for instance, next-nearest neighbors or any longer finite range. Indeed, elastic strain involves several lattice units in real materials [26]. At the level of our model, we find that this effect is so severe that, even upon increasing the range of the elastic term to the next-nearest neighbors, our simulations become prohibitively slow. The mathematical analysis of the behavior and dependence of the model parameters can be carried on along similar lines as we did above (see the Supplemental Material [25]). Thus our results suggest that the origin of the long (and possibly infinite) retention times observed experimentally in regime (iii) may be due to long-range strain induced by the filament growth.

V. RESISTANCE RELAXATION WITH AN APPLIED VOLTAGE BIAS

We shall now discuss an interesting effect due to the introduction of the elastic energy barrier term in our model

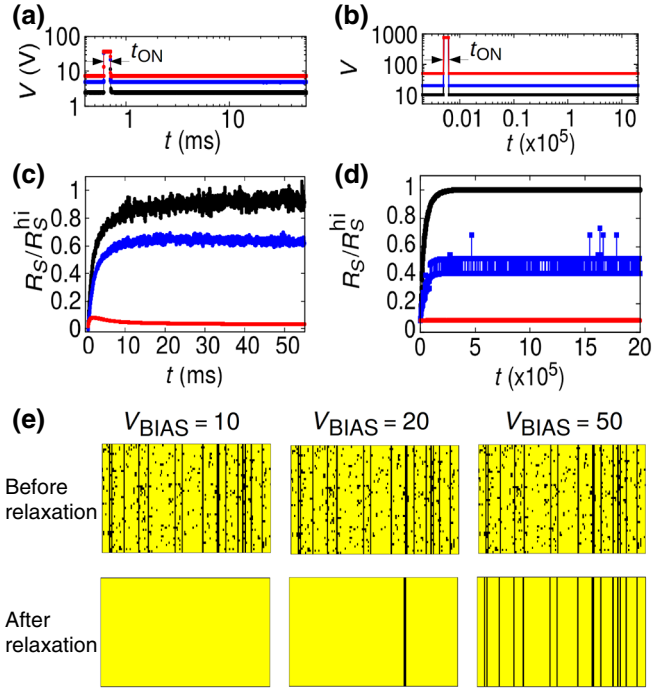

FIG. 11. The relaxation of the sample resistance after EMT under different bias voltages. (a),(b) The voltage protocol for (a) experiments on a $GaMo_4S_8$ sample and (b) simulations. An EMT is induced by a voltage pulse of magnitude $V_P > V_{\rm th}$ and duration $t_{\rm ON}$, after which the system is left to evolve under the action of a voltage $V_{\rm BIAS} < V_{\rm th}$. (c) The experimental relaxation for bias voltages of $V_{\rm BIAS} = 2.4$, 4.8, and 7.1 V (top to bottom). (d) The simulation results for $V_{\rm BIAS} = 10$, 20, and 50 (top to bottom). In panels (c) and (d), the time is measured from the beginning of the relaxation. (e) Snapshots of the resistive network before and after the relaxation process shown in panel (b).The snapshots correspond to the initial and final states of each curve in panel (b). We see that the state before relaxation is essentially the same for the three curves. The final state, however, depends strongly on the applied $V_{\rm BIAS}$. For the lowest bias, all filaments are reabsorbed; for the intermediate bias, only thicker filaments survive; and for the highest bias, all filaments survive. The presence of $V_{\rm BIAS}$ stabilizes (prevents relaxation of) the metallic filaments. Thicker (more stable) filaments require a lower external bias.The simulations correspond to a 50×140 network with $E_B^E = 15$, $E_{\rm CM} = 5$, $R_L = 500$, $\kappa = 0.5$, $R_{\rm MI} = 20000$, $R_{\rm CM} = 200$, $t_P = 500$, and $t_{\rm ON} = 120$. For a better visualization, the ratio $R_{\rm MI}/R_{\rm CM}$ is reduced with respect to previous sections. The EMT is generated with an external voltage pulse of $V_P = 750$ in the simulations and $V_P = 37.5$ V in the experiments.

related to the electric stabilization of filaments. In other words, how the elastic interaction may allow to electrically prevent the filament break-up and reabsorption. In previous sections, we have seen that a minimum threshold voltage $V_{\rm th}$ is necessary for filament formation. In this section, we consider a different issue, namely, what is the minimally low voltage that needs to be applied to sustain an already fully formed filament? The interest of this question is that

it may be experimentally tested as further validation of the present model.

In Fig. 11, we show our simulation results for a resistive-switching protocol where a small bias voltage ($V_{\rm BIAS} < V_{\rm th}$) remains applied during the relaxation process. Along with our numerical results, we also show in the figure actual experimental data obtained in a $GaMo_4S_8$ sample under a qualitatively similar voltage protocol. Specifically, the experiment is performed as follows: (i) at $t = 0$, a bias voltage $V_{\rm BIAS}$ is turned on; (ii) at $t = t_P$, a high-voltage pulse V_P is applied for a time $t_{\rm ON}$; and (iii) at $t_{\rm OFF} = t_P + t_{\rm ON}$, the applied voltage returns to $V_{\rm BIAS}$ and the evolution of the resistance of the system is observed. The bias voltage remains applied during the entire process. The times t_P and $t_{\rm ON}$ and the total applied voltage during the pulse $V_{\rm BIAS} + V_P = \text{constant} > V_{\rm th}$ are kept constant so that the switching voltage is the same in all realizations. In the figure, we show the results of this protocol for three different values of $V_{\rm BIAS}$. We observe that in both the experiments and the simulations, the asymptotic relaxation value of the resistance depends on the applied $V_{\rm BIAS}$. In fact, at the highest applied bias (red curve), we see that the resistance does not recover to the initial state but remains at a lower value. To get a better understanding of this behavior, we show in Fig. 11(c) snapshots of the simulated resistive network for each $V_{\rm BIAS}$ at two different times: a short one, right after the pulse is terminated at $t = t_{\rm OFF}$, and a longer one when the asymptotic resistance state is reached [32]. We observe from the short-time snapshots (upper panels) that the $R_S^{\rm hi} \to R_S^{\rm lo}$ resistive change is rather independent of the applied (subthreshold) bias level, as very similar multifilament structures are realized in all three cases. On the other hand, for the asymptotic state, the system shows qualitatively different states. In the lowest-bias case, the whole filamentary structure is reabsorbed; in the intermediate-bias case, only a single filament persists, which is the thickest; and in the higher-bias case, all filaments formed during the pulse remain at asymptotic times. We also note that all isolated metallic regions (clusters) do relax to the insulator state in all three cases. The fluctuations observed for the intermediate voltage in panel (b) (blue curve) are due to the rapid rupture and reconnection of small sections of a filament (i.e., a transitory decay to the insulating state). This effect may actually happen in real materials; however, in our simulations it is likely overemphasized due to the finite size.

VI. THE COMPLETE PULSE-VOLTAGE PROTOCOL

In this final section, we shall consider the full-time behavior of the system, that is, the whole process of formation and subsequent reabsorption of filaments. We shall adopt two different applied voltage protocols: a single pulse and a train of pulses. The former case is simplest to analyze as it is closely related to the discussions in the previous sections. In contrast, the interest of applying a train of pulses [13] is that it brings us closer to the situation of actual neurons, which are excited by electric spikes.

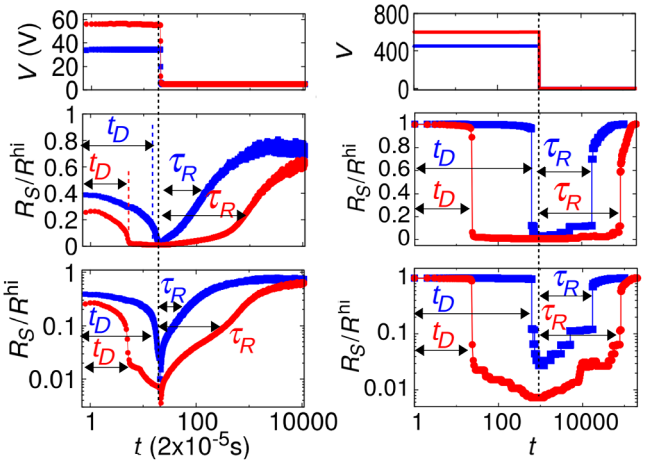

FIG. 12. The evolution of the resistance during a complete switching process. We show (a) the experimental results on a $GaMo_4S_8$ sample and (b) the results of our simulations. The voltage protocols are shown in the top panels. In each case, an external voltage is applied for a time $t_{\rm ON}$, after which the system is left to relax freely. The red and blue curves represent the results for two different applied voltages. The evolution of the resistance is shown in semilog and log-log scale in the middle and bottom panels, respectively. The experimental curves correspond to $t_{\rm ON} = 100$ μs, $V = 34$ V (blue curve), and $V = 56$ V (red curve). The simulations correspond to $t_{\rm ON} = 1000$, $V = 450$ (blue) and $V = 600$ (red). The characteristic times t_D and τ_R are indicated for each case. We note that in the experimental data, the initial value of $R_S/R^{\rm hi}$ is smaller than unity due to a capacitive transient introduced by the measurement circuit.

In Fig. 12, we show the full-time evolution of the resistance during the application of a single pulse for both simulations and experiments on a $GaMo_4S_8$ sample (the same sample as in Fig. 7). A single voltage is applied for a fixed duration $t_{\rm ON}$ and then V is set to zero. We show data for two different applied voltages. We observe the significant dependence of the retention time with the applied voltage. We see that a relatively small change in the strength of the applied pulse produces a great change of the retention time. For example, in the experiments a change in the voltage from 34 to 56 V produces a change in the retention time of a full order of magnitude (from approximately 2×10^{-3} to approximately 2×10^{-2}). Similarly in the simulations, a change in the voltage from 450 to 600 V produces a fivefold increase of the retention time (from approximately 2×10^4 to approximately 10^5).

These results permit us to rationalize the existence of regimes (ii) and (iii), which we described in Sec. I. In Fig. 13, we show our simulation results for the retention time τ_R as a function of the parameters of a single voltage

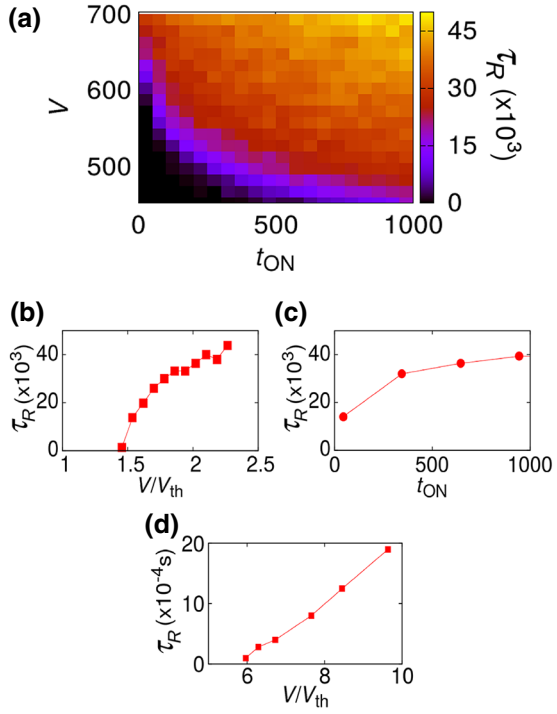

FIG. 13. (a) A color map of the retentivity as a function of the pulse length (t_{ON}) and the applied voltage. (b) A vertical cross section of panel (a): the evolution of τ_R as a function of the applied voltage for fixed $t_{ON} = 500$. The voltage is measured in units of the threshold voltage with $V_{th} = 309$. (c) A horizontal cross section of panel (a): the evolution of τ_R as function of the pulse length for fixed $V = 600$. The results are obtained as the average over 100 realizations. (d) The experimental results: the evolution of τ_R on a $GaMo_4S_8$ sample as a function of the applied voltage for fixed $t_{ON} = 100$ μs.

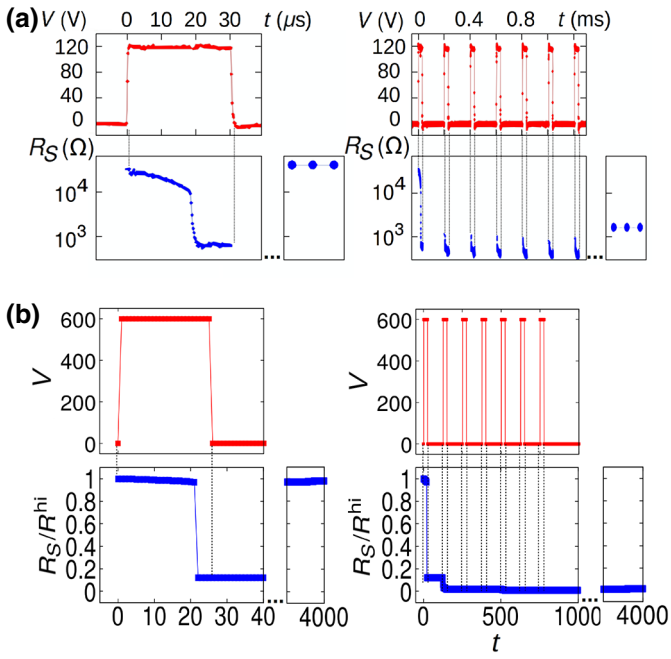

FIG. 14. The manipulation of retentivity in RS via a pulsed protocol. (a) The experimental results in a GaV_4S_8 crystal for a single pulse of $t_{ON} = 30$ μs with $V = 120$ V (left) and for a train of seven pulses of $t_{ON} = 30$ μs with $V = 120$ V separated by 200 μs (right). (b) The results from the simulations for a similar pulsed protocol. A single pulse of $t_{ON} = 25$ with $V = 600$ (left) and a train of seven pulses of $t_{ON} = 25$ with $V = 600$ separated by 100 simulation steps (right) are shown.

pulse: the intensity V and the duration time t_{ON}. In the color plot of panel (a), we observe that the retention of the filaments may be increased by incrementing either V or t_{ON}, the key point being that any one of these two parameters contributes to thickening of the filaments. The dependence of the threshold on the duration of the applied pulse can also be clearly observed. This is basically the effect of the delay time that has been discussed in previous sections: At a given applied voltage, there is a delay time for the transition, which implies that the pulse duration must be above the delay time to actually produce the filament formation. The two panels (b) and (c) show two cuts of the color map for fixed t_{ON} and V, respectively. All these results correspond to the mean over 100 realizations. Finally, in panel (d), we show the evolution of τ_R on a $GaMo_4S_8$ sample as a function of the applied voltage for fixed $t_{ON} = 100$ μs.

We may finally illustrate the behavior of the model upon the application of a train of voltage pulses. We show in Fig. 14 the comparison of our model simulations with actual experimental data of the resistive change in a crystal of GaV_4S_8 upon application of voltage pulses. Two types of pulse protocol are applied: either a single pulse or a train of identical pulses with a relatively long separation, which mimics a train of spikes that may be arriving at the dendrites of a neuron [13]. In actual situations, the arrival of the pulses is mostly random, although here we focus on regular trains for the sake of simplicity. In the top panels of the figure, we show the experimental data and in the bottom panels the simulations. Each pulse of the train is identical to the single one. The data shown in red correspond to the applied voltage and the data in blue to the respective resistance value $R(t)$. The small subpanels to the right of the main $R(t)$ panels indicate the asymptotic value of the resistance after the voltage pulse protocol is terminated. In the experimental data, we observe that the single pulse is sufficient to produce a resistive change, which takes place after about 20 μs. However, this low-resistance state is volatile, as the corresponding asymptotic value has returned to the original high-resistance value. In contrast, as the data in the right panel show, the application of seven identical pulses is enough to dramatically increase the retention time. Thus, seven pulses are found to be sufficient to drive the resistance change into the nonvolatile regime (on the time scale of the experiment).

In the bottom panels, we see that, qualitatively, the simulations capture these effects well. On the basis of the

discussions in the previous sections, we can understand this behavior. As shown in the left panels of the figure, a single pulse is enough to induce an EMT with filament formation. However, that single pulse only produces a thin filament with a short retention time, which rapidly gets reabsorbed. Thus the system returns back to the initial high-resistance state. In contrast, the train of multiple pulses produces gradual thickening of the filament, with a consequently much longer retention time, and persists beyond $t = 4000$.

This behavior may be considered an initial step toward the modeling of the recovery period of an artificial neuron under a spiking stimulus. It may be interesting to mention that this variation in the recovery time may find potential useful applications in neural networks for tasks such as neural coding [27,28].

We may note that the interpulse time duration (200 μs in the experiment) provides a rough indication of the reabsorption time of the filaments. The systematic investigation of the retention time with the parameters of the applied voltage pulsing protocol is an important issue but it certainly involves some technical challenges and is outside the scope of the present study.

VII. EXPERIMENTAL DETAILS

The experimental setup used for the application of pulses is displayed in Fig. 2(a). It consists of a pulse generator (Agilent 8114A) placed in series with a load resistor and home-synthesized GaV_4S_8, $GaMo_4S_8$, or $GaTa_4Se_8$ crystals [29]. The crystals (typical dimensions $300 \times 150 \times 50$ μm^3) used for the resistive-switching measurements are contacted on a freshly cleaved face in metal-insulator-metal (MIM) geometry using two 17 μm gold wires and carbon paste (Electrodag PR-406). A subsequent annealing in vacuum at 150°C for 30 min is necessary to achieve good-contact resistances. The low-bias resistance of the AM_4Q_8 is measured using a Keithley 6430 source-measurement unit. We check that the contact resistances are much smaller than the sample resistance. Voltage pulses are applied using an Agilent 8114A. During the pulse, the voltage and current across the sample are measured with a Tektronix DPO3034 oscilloscope, associated with IeS-ISSD210 differential probes. Pulse experiments conducted on $GaTa_4Se_8$ and displayed in Fig. 4(b) are performed at $T = 77$ K with a load resistance of $R_L = 1$ kΩ. The variation of the resistive-switching time delay versus the sample voltage displayed in Fig. 7(a) and the resistance relaxations displayed in Figs. 11(a), 12(a), and 13(d) are measured on the same $GaMo_4S_8$ crystal at $T = 74$ K using a load resistance of $R_L = 5$ kΩ. Finally, the multipulse experiments displayed in Fig. 14(a) are conducted at $T = 153$ K on a GaV_4S_8 crystal mounted in series with a load resistance of $R_L = 5$ kΩ. These last data are also presented in Ref. [30].

VIII. CONCLUSIONS

In the present study, we extend a model for the EMT to incorporate electroelastic effects. This feature leads to spatial correlations in the growth patterns of filamentary structures in the resistive transition. In particular, it leads to the thickening of the filaments during longer applied pulses or multiple pulses. The increased thickness of the filaments is directly related to a dramatic increase in the retention time of the low-resistance state.

We should mention that our model does not predict any sharp or discontinuous transition between the experimentally observed volatile regime (ii) and the nonvolatile regime (iii) discussed in Sec. I. Nevertheless, it does indicate that the retention times may be exponentially increased by means of the applied pulsing protocol. Specifically, the strength, duration, and interpulse time may be tailored to optimize the growth and thickening of the filaments. The results of the previous section may be considered as an initial step in that direction. However, it should be realized that there are a large number of variables to take into account and that systematic investigation remains an important experimental challenge.

We also briefly refer to the qualitative similarities between our present model study and the behavior of spiking leaky-integrate-and-fire neurons. We are aware of the relative simplicity of our system with respect to the biological equivalents but, nevertheless, this may be taken as a contribution toward future implementation of neuromorphic electronic systems. In this regard, it is very exciting to envision that strongly correlated Mott materials may realize unexpected bioinspired functionality.

ACKNOWLEDGMENTS

P.S. acknowledges the support of the Spanish Ministry of Economy through the Ramón y Cajal subprogram (Grant No. RYC-2012-01031). This work was supported by the Région Pays de la Loire through the funding of the "pari scientifique Neuro-Mott." The work was also supported by the ANR project Elastica No. 16-CE30-0018. B.C., E.J., and L.C. thank all the participants of the French-Japanese International Associated Laboratory "IM-LED Impacting Materials with Light and Electric Fields and Watching Real Time Dynamics" for stimulating discussions dealing with the relaxation processes. Work at UCSD was supported by Q-MEEN-C, an Energy Frontier Research Center funded by DOE, Office of Science, BES under award DE-SC0019273.

[1] S. Mallat, Group invariant scattering, Commun. Pure. Appl. Math. **65**, 1331 (2012).

[2] J. Schmidhuber, Deep learning in neural networks: An overview, Neural. Netw. **61**, 85 (2015).

[3] H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, Metal-Oxide RRAM, Proceedings of the IEEE **100**, 1951 (2012).

[4] S. Tang, F. Tesler, F. G. Marlasca, P. Levy, V. Dobrosavljevicć, and M. Rozenberg, Shock waves and commutation speed of memristors, Phys. Rev. X **6**, 011028 (2016).

[5] F. Tesler, S. Tang, V. Dobrosavljević, and M. Rozenberg, Spintronics X (International Society for Optics and Photonics, San Diego, CA, 2017), Vol. 10357, p. 103572L.

[6] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, Nanoscale memristor device as synapse in neuromorphic systems, Nano Letters **10**, 1297 (2010).

[7] Y. V. Pershin and M. D. Ventra, Experimental demonstration of associative memory with memristive neural networks, Neural. Netw. **23**, 881 (2010).

[8] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, A scalable neuristor built with Mott memristors, Nature Mat **12**, 114 (2013).

[9] Y. Babacan, F. Kacar, and K. Gürkan, A spiking and bursting neuron circuit based on memristor, Neurocomputing **203**, 86 (2016).

[10] D. Lepage and M. Chaker, Thermodynamics of self-oscillations in VO_2 for spiking solid-state neurons, AIP Adv. **7**, 055203 (2017).

[11] L. Gao, P.-Y. Chen, and S. Yu, NbO_X based oscillation neuron for neuromorphic computing, Appl. Phys. Lett. **111**, 103503 (2017).

[12] J. Lin, A. Annadi, S. Sonde, C. Chen, L. Stan, K. V. L. V. Achari, S. Ramanathan, and S. Guha, *2016 IEEE International Electron Devices Meeting (IEDM)* (IEEE, San Francisco, 2016), p. 34.5.1.

[13] P. Stoliar, J. Tranchant, B. Corraze, E. Janod, M.-P. Besland, F. Tesler, M. Rozenberg, and L. Cario, A Leaky-Integrate-and-Fire Neuron Analog Realized with a Mott Insulator, Adv. Funct. Mater. **27**, 1604740 (2017).

[14] P. Stoliar, L. Cario, E. Janod, B. Corraze, C. Guillot-Deudon, S. Salmon-Bourmand, V. Guiot, J. Tranchant, and M. Rozenberg, Universal Electric-Field-Driven Resistive Transition in Narrow-Gap Mott Insulators, Advanced Materials **25**, 3222 (2013).

[15] M. Imada, A. Fujimori, and Y. Tokura, Metal-insulator transitions, Rev. Mod. Phys. **70**, 1039 (1998).

[16] V. Guiot, L. Cario, E. Janod, B. Corraze, V. T. Phuoc, M. Rozenberg, P. Stoliar, T. Cren, and D. Roditchev, Avalanche breakdown in $GaTa_4Se_{8-x}Te_x$ narrow-gap Mott insulators, Nat. Commun. **4**, 1722 (2013).

[17] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, and P. Werner, Nonequilibrium dynamical mean-field theory and its applications, Rev. Mod. Phys. **86**, 779 (2014).

[18] E. Janod, J. Tranchant, B. Corraze, M. Querré, P. Stoliar, M. Rozenberg, T. Cren, D. Roditchev, V. T. Phuoc, M.-P. Besland, and L. Cario, Resistive switching in Mott insulators and correlated systems, Adv. Funct. Mater. **25**, 6287 (2015).

[19] P. Stoliar, M. Rozenberg, E. Janod, B. Corraze, J. Tranchant, and L. Cario, Nonthermal and purely electronic resistive switching in a Mott memory, Physical Review B **90**, 045146 (2014).

[20] M. M. Qazilbash, M. Brehm, B.-G. Chae, P.-C. Ho, G. O. Andreev, B.-J. Kim, S. J. Yun, A. V. Balatsky, M. B. Maple, F. Keilmann, H.-T. Kim, and D. N. Basov, Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging, Science **318**, 1750 (2007).

[21] V. Dubost, T. Cren, C. Vaju, L. Cario, B. Corraze, E. Janod, F. Debontridder, and D. Roditchev, Resistive Switching at the Nanoscale in the Mott Insulator Compound $GaTa_4Se_8$, Nano. Lett. **13**, 3648 (2013).

[22] A. Camjayi, C. Acha, R. Weht, M. G. Rodrǵuez, B. Corraze, E. Janod, L. Cario, and M. J. Rozenberg, First-Order Insulator-to-Metal Mott Transition in the Paramagnetic 3D System $GaTa_4Se_8$, Phys. Rev. Lett. **113**, 086404 (2014).

[23] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Rev. Mod. Phys. **68**, 13 (1996).

[24] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (CRC Press, Boca Raton, 2014).

[25] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.10.054001 for details on the dynamical analysis of the behavior of the model, some minor experimental details and results of simulations with thick filaments and longer range elastic interactions.

[26] A. J. Wilkinson, G. Meaden, and D. J. Dingley, High-resolution elastic strain measurement from electron backscatter diffraction patterns: New levels of sensitivity, Ultramicroscopy **106**, 307 (2006).

[27] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, USA, 1999).

[28] W. Gerstner, A. K. Kreiter, H. Markram, and A. V. M. Herz, Neural codes: Firing rates and beyond, Proceedings of the National Academy of Sciences **94**, 12740 (1997).

[29] C. Vaju, L. Cario, B. Corraze, E. Janod, V. Dubost, T. Cren, D. Roditchev, D. Braithwaite, and O. Chauvet, Electric-Pulse-Driven Electronic Phase Separation, Insulator-Metal Transition, and Possible Superconductivity in a Mott Insulator, Advanced Materials **20**, 2760 (2008).

[30] J. Tranchant, E. Janod, B. Corraze, P. Stoliar, M. Rozenberg, M.-P. Besland, and L. Cario, Control of resistive switching in AM_4Q_8** narrow gap Mott insulators: A first step towards neuromorphic applications, Phys. Stat. Solidi (a) **212**, 239 (2015).

[31] For a better comparison with the simulations, in the histogram of Fig. 8(c) we consider cross sections up to a few tens of nm^2. The full distribution is shown in the Supplemental Material [25].

[32] As a technical side note, we observe that we keep the same "seed" for the pseudorandom number generator in all three simulations for the sake of a more meaningful comparison.