Conductive preferential paths of hot carriers in amorphous phase-change materials

Andrea Cappelli,[1] Enrico Piccinini,[2,a)] Feng Xiong,[3] Ashkan Behnam,[3] Rossella Brunetti,[1] Massimo Rudan,[2] Eric Pop,[3,b)] and Carlo Jacoboni[1]
[1]*Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/A, I-41125 Modena, Italy*
[2]*"E. De Castro" Advanced Research Center on Electronic Systems (ARCES), Università di Bologna, Via Toffano 2/2, I-40125 Bologna, Italy*
[3]*Department of Electrical and Computer Engineering, Micro and Nanotechnology Lab, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

(Received 28 June 2013; accepted 2 August 2013; published online 20 August 2013)

We study charge transport properties of amorphous phase-change materials (PCM) using a set of balance equations applied to a three-dimensional random network of sites. In the context of trap-limited conduction, model results are checked against experimental data on PCM devices near the limits of scaling (~10 nm), explaining the main features of the current-voltage characteristics. The stochastic nature of the network also allows us to investigate the statistical variability of the sub-threshold PCM operation. Simulations of batches of similar samples show a standard deviation for the threshold condition of the order of few percent for the threshold voltage and of ten percent for the threshold current. The analysis of the network at the microscopic level near threshold reveals the formation of high-current paths, connecting the two contacts of the device through network nodes hosting the hottest carriers. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4819097]

Understanding the transport properties of disordered materials is of great interest from both the fundamental and the applied points of view. At present, the electronic industry is manufacturing an increasing number of devices whose operation relies on the properties of amorphous materials. Within the field of data storage, materials like chalcogenides,[1,2] silicon dioxide,[3] metal oxides,[4] graphene,[5] and amorphous carbon[6] can controllably be switched between a low- and a high-resistance state. In many cases, the low conductivity state stems from a disordered and defective structure. These properties allow for reversible information encoding and its readout.[7–9]

Such amorphous materials often show an onset of so-called threshold-switching phenomena when proper conditions are reached. The microscopic knowledge of the physical processes involved is of the utmost importance for any technological exploitation. In the particular case of the Ge-Sb-Te (GST) chalcogenide compounds, electronic switching leads to a large increase in conductivity triggered by the application of a voltage above a threshold value V_{th}.[9]

Although thermal interpretations for Ovonic switching have not been completely abandoned,[10] the most often referenced models ascribe this threshold-switching effect to electronic phenomena, such as hot-electron trap-limited transport[11–13] or field-induced crystallization.[14] Due to the complexity of the phenomenon, these theories usually introduce a quasi one-dimensional (1D) approximation or consider a three-dimensional (3D) continuous medium, averaging out the intrinsic disorder of the amorphous material. With the miniaturization of nanoelectronics, very small numbers of carriers and traps are involved in the device operation. Rough estimates indicate that only about one hundred defect states are present in phase-change memories at the 32 nm technological node. Thus, a more realistic 3D description of the positions of the defects paves the way to better and more precise modeling.

In this work we describe amorphous GST by means of the nodes of a random network. The node sites represent the actual positions of the traps of the real system, or the average coordinates of clusters of neighboring traps. Details on the positions of the traps should be provided by specific experimental investigations or by first-principle calculations that are not available within the required precision. Therefore, positions are randomly generated within a 3D domain according to uniform distributions along each coordinate. A minimum distance r_{min}, representing the typical dimension of the clusters, separates any pair of nodes. The transport picture complies with the hydrodynamic-like (HD) model for trap-limited conduction[12,13] and assumes carriers sitting at the node sites and moving among them according to an emission-travel-recapture process. For a prescribed input current flowing across the device, we find the corresponding voltage drop at its ends, calculating a characteristic current-voltage (I–V) curve. The effect of different node configurations on the conduction behavior and on the switching point is also assessed. The calculated I–V curves are compared to experimental data from carbon-nanotube-contacted GST-based (CNT-GST) devices.[15,16] Finally, by considering the carrier flow among the network nodes, we observe the appearance of preferential hot-carrier conduction paths, occurring in proximity of the electrical switching.

We consider a device made of an amorphous GST layer sandwiched between two conductive surfaces and model it

a)Author to whom correspondence should be addressed. Electronic mail: enrico.piccinini@unimore.it
b)Present address: Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

as a 3D network of N randomly placed nodes. At two opposite edges, two special planar equipotential nodes (0 and $N+1$) play the role of the contacts. We associate to the i-th node its electrostatic potential φ_i and the carrier population n_i. The average energy ε_i is attributed to all members of the population sitting at node i. The zero of ε_i corresponds to the equilibrium distribution.

Following the concept of trap-limited conduction and considering for simplicity only the case of electrons, carrier motion comprises three stages: emission of a carrier from the i-th node to a traveling state with energy just above the conduction-band mobility edge E_C, propagation toward the j-th node, and finally recapture by the j-th node. To represent the overall process we assume the transition rate

$$S_{ij} = \frac{f(r_{ij})}{\tau_0} \exp\left(-\frac{E_C - \varepsilon_i}{k_B T}\right) \exp\left[\frac{-q(\varphi_i - \varphi_j)\ell}{r_{ij} k_B T}\right], \quad (1)$$

where $\tau_0 \sim 10-100$ fs is the characteristic time of the transition process (dominated by the detrapping process, most likely due to the interaction with phonons),[17] k_B is the Boltzmann constant, T the lattice temperature, q the elementary charge, and r_{ij} the distance between nodes i and j. The term $f(r_{ij})$ in Eq. (1) is a function of the internode distance r_{ij}, describing propagation and recapture. For simplicity we take $f(r_{ij}) = 0$ for r_{ij} larger than a cutoff distance r_{cut}, $f(r_{ij}) = 1$ otherwise. Furthermore, the two exponentials account for the thermally activated nature of the excitation process.[18,19] The first one accounts for the energy barrier in the absence of field; the second one is an effective correction given by the average electric field $F = -(\varphi_j - \varphi_i)/r_{ij}$ between nodes i and j at a characteristic distance $\ell \lesssim r_{cut}$. When $r_{ij} < 2\ell$, the correction is evaluated at $\ell = r_{ij}/2$, similarly to Ref. 20.

The connections among the nodes define a network of transition rates; given any pair of nodes (i, j), the current between them is $I_{ij} = -q(n_i S_{ij} - n_j S_{ji})$. In order to find the steady-state value of the node variables $(\varphi_i, \varepsilon_i, n_i)$, we impose charge- and energy-flux balances at each node. As for the latter, three contributions are taken into account: the energy lost at the node site due to carriers moving away, the energy gained from the carriers arriving at the node corrected by the effect of the field experienced during the transition, and the energy transferred to the lattice via phonon scattering with a relaxation time τ_R. Neglecting higher order-terms, for a given current I, the two balance conditions[21] yield

$$\frac{I}{(-q)}\delta_{0,i} + \sum_j n_i S_{ij} = \sum_j n_j S_{ji} + \frac{I}{(-q)}\delta_{N+1,i}, \quad (2)$$

$$\sum_j n_j S_{ji}[\varepsilon_j - q(\varphi_j - \varphi_i)] + \frac{I}{(-q)}\varepsilon_i \delta_{0,i} = \varepsilon_i \left[\sum_j n_i S_{ij} + \frac{I}{(-q)}\delta_{N+1,i}\right] + n_i \frac{\varepsilon_i - \varepsilon_{i,eq}}{\tau_R}, \quad (3)$$

with $i = 0, 1, \ldots, N+1$. Since the model is current driven, the two equations for nodes 0 and $N+1$ carry two Kronecker's δ extra terms that prescribe the fluxes from and to the external circuit. The two sets of non-linear equations above are solved along with the Poisson equation to find φ_i, ε_i, and n_i in all the device nodes, employing a Newton-Raphson scheme coupled to a finite-element procedure. Given the network configuration, the algorithm leads to an I–V curve, with $V = \varphi_{N+1} - \varphi_0$.

The configuration-dependent nature of the model allows one to study the impact of intrinsic disorder on the threshold conditions. Figure 1 shows a set of one hundred I–V curves, differing only in the particular random node positions. The statistical analysis of the different threshold voltages and currents shows variations of up to ~15% for V_{th} and ~40% for I_{th}, with respect to the corresponding average values $\langle V_{th}\rangle$ and $\langle I_{th}\rangle$. Despite quantitative differences, the three known conduction regimes identified by the analytical models (Refs. 11 and 12)—Ohmic, exponential, and super-exponential—are recovered in all curves, together with the onset of the voltage snap-back. As in Ref. 12, the existence of two stable states for a given voltage is associated with the realization of two microscopic configurations, one with a near-equilibrium carrier mean energy, the other with a non-equilibrium hot-carrier distribution.

To validate our model against experimental data, we consider a theoretical I–V curve obtained for a given set of parameters averaging the potential drops of 30 simulations with random configurations. The standard deviation of the potential drops provides an estimate of the expected variability. In Fig. 2, we show the calculated average curves and standard deviations best fitting three experimental memory device I–V curves. The experimental data are taken from recent GST devices with CNT electrodes, which represent the smallest GST volumes probed to date. The CNT electrodes are ~2 nm in diameter, enabling probing of ~10 nm scale GST bits, which switch with sub-1 μA currents, nearly 100× lower than previous state-of-the-art.[16] Valuable models for similar devices would therefore enhance the

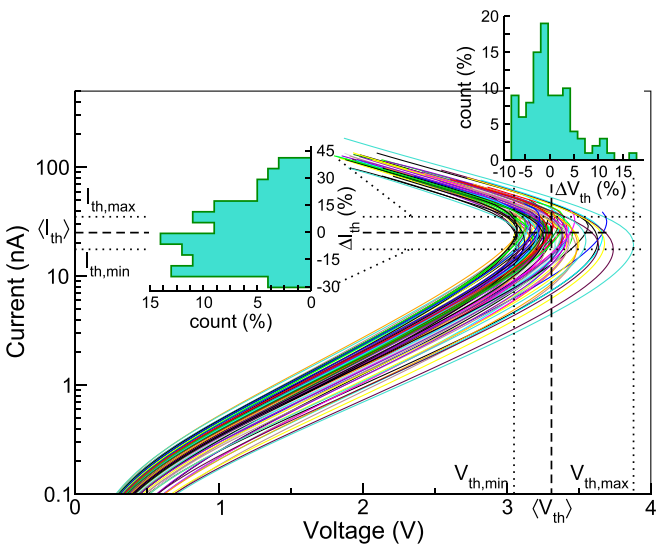

FIG. 1. Statistical fluctuations in the I–V curves for different spatial configurations of the network. The dashed and the dotted lines show the average and the minimum and maximum values for V_{th} and I_{th}, respectively. The two insets report the frequency analysis for the occurrence of the deviations from the average values $\langle V_{th}\rangle$ and $\langle I_{th}\rangle$. The analysis has been performed on a test device of $10 \times 10 \times 40\,\text{nm}^3$ setting the following parameters: $\tau_0 = 10^2$ fs, $\tau_R = 10^3$ fs, $r_{cut} = 6$ nm, $\ell = 2$ nm, node concentration $1.2 \times 10^{19}\,\text{cm}^{-3}$.

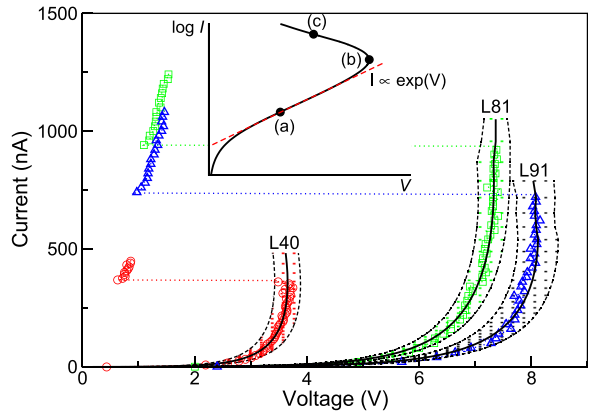

FIG. 2. Experimental *I*–*V* characteristics of three different CNT-GST devices (see Ref. 16), whose lengths are 40 nm (red circles), 81 nm (green squares), and 91 nm (blue triangles). The estimation of the device size is provided in the supplementary material.[21] The black solid lines and the dashed side lines show, up to the threshold region, the simulated average characteristics and the standard deviations obtained running a set of 30 simulations in each case. The inset shows a typical *I*–*V* characteristic in a semilog scale; the dashed lines are intended as an eye-guide for the exponential regime. Points (a), (b), and (c) used in the microscopic analysis of Fig. 3 are also highlighted.

technological development of next-generation non-volatile memories. In the subthreshold region the agreement is good. The experimental points lie within the dispersion range, confirming that the fitted parameters are reasonable. The onset of the threshold switching in the theoretical curve, i.e., the point where $dV/dI \approx 0$, corresponds quite accurately to the experimental threshold point. The abrupt voltage drop of the experimental curves above threshold eventually involves the crystallization of the system,[9] which is not addressed by the model in its present form.

The parameters used for the fit are listed in Table I and compare fairly well to those obtained with the HD model for GST memory devices.[12,21] Parameters τ_0, r_{cut}, and ℓ are similar in the three cases, so that they can be taken as best fit values of material properties. A larger variability is instead found for parameter τ_R. On the other hand, we notice that the variability may also depend on the degree of amorphization and on geometric effects of the sample that have not been considered here. The values obtained for the characteristic time τ_0 and the relaxation time τ_R confirm the link of these parameters with phononic interactions.[17] The connection of ℓ and r_{cut} with the mean propagation distance is detailed in the supplementary material.[21]

The possibility of tracking conductive preferential paths is one of the features of the present model. Figure 3 shows three frames of a ball-and-stick animation[21] from a single network configuration, representing the mean energies ε_i on the nodes and the normalized currents I_{ij}/I, for the three points (a), (b), and (c) indicated in the inset of Fig. 2. From the comparative analysis of the images, one easily sees how the different transport behaviors originate from different microscopic conditions. Figure 3(a) refers to point (a) in Fig. 2 and is obtained for a current in the exponential regime, sufficiently far from the threshold point. In this case, the current is almost evenly distributed among several paths, each of them carrying a similar fraction of the total current, roughly inversely proportional to the number of links crossing a section. Slight non-uniformities in the flux distribution are generated by geometrical factors due to the particular positions of the nodes. The carrier energies are all close to the equilibrium value. As the switching point is approached, the picture rapidly changes [Fig. 3(b)]: the non-uniformities become more evident, and the current flowing through the geometrically favored nodes is enough to induce heating of their carrier populations. This, in turn, enhances their transition rates and allows an even larger current to flow on those particular paths. The outcome of the process is the positive feedback needed to trigger the switching event: more current implies more heating and faster transition rates, and so on. Such a loop determines a steady-state condition where one or a few highly conductive continuous routes connect the contacts

TABLE I. Model parameters for highly scaled GST memory devices with CNT electrodes, as from the three best fits of Fig. 2. A cross-section $10 \times 10\,\text{nm}^2$ and a node concentration $1.2 \times 10^{19}\,\text{cm}^{-3}$ have been adopted in all cases.

Device length (nm)	τ_0 (fs)	τ_R (fs)	r_{cut} (nm)	ℓ/r_{cut}
$L = 40$	49	56.5	6	0.48
$L = 81$	49.5	25	6.5	0.5
$L = 91$	58.3	33.3	6.5	0.5

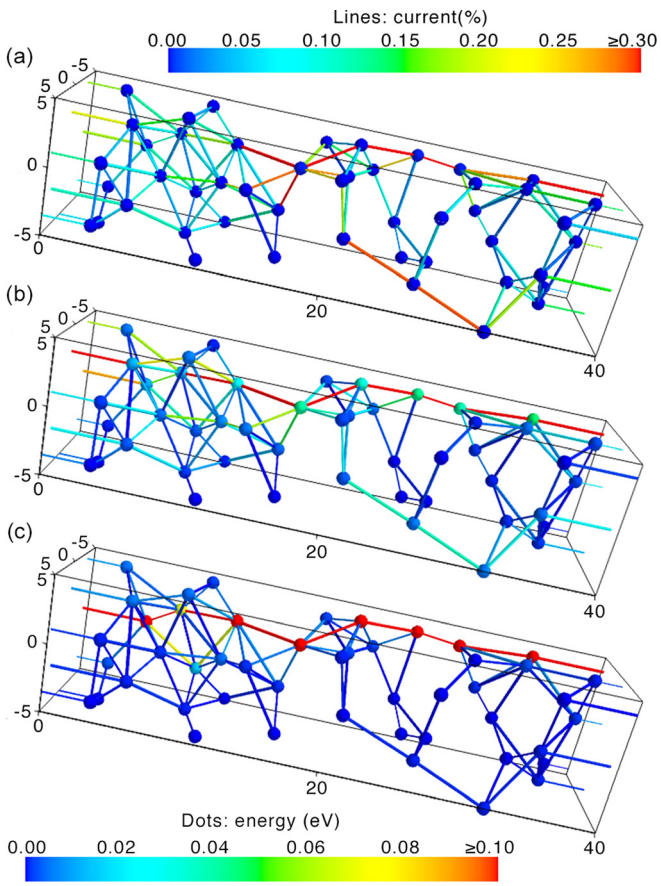

FIG. 3. Snapshots of the microscopic network configurations (lines – top scale: current; dots – bottom scale: node energy) obtained for the three points indicated in Fig. 2 for a $10 \times 10 \times 40\,\text{nm}^3$ device. The dynamics of formation of a preferential path is evidenced by: (i) the increasing number of the low-resistance connections; (ii) the progressive reduction of the current flowing in the rest of the device; and (iii) the growing energy of the nodes along the preferential path. Images have been created with the visualization software Mayavi (see Ref. 22).

TABLE II. Effect of the interaction length r_{cut} on the threshold point and its variability. Mean values and standard deviations are reported for both V_{th} and I_{th} for groups of 100 simulated devices.

r_{cut} (nm)	$\langle V_{th}\rangle \pm \sigma_{V_{th}}$ (V)	$\langle I_{th}\rangle \pm \sigma_{I_{th}}$ (nA)
6	3.31 ± 0.16 ($\pm 4.8\%$)	25.17 ± 4.35 ($\pm 17.3\%$)
7	2.94 ± 0.07 ($\pm 2.4\%$)	33.71 ± 4.48 ($\pm 13.3\%$)
8	2.74 ± 0.05 ($\pm 1.8\%$)	40.90 ± 5.71 ($\pm 14.0\%$)
10	2.48 ± 0.04 ($\pm 1.6\%$)	52.50 ± 6.86 ($\pm 13.1\%$)
12	2.29 ± 0.03 ($\pm 1.3\%$)	65.21 ± 8.19 ($\pm 12.5\%$)

[Fig. 3(c)], carrying most of the prescribed current. The lower resistance of these paths leads to a decrease in the voltage dropped across the device.

We remark in passing that the power dissipated along these preferential paths could be enough to trigger crystallization near the threshold condition. In order to further confirm this hypothesis, the present model must be enhanced with the incorporation of the Fourier heat equation and a local lattice temperature. Such an extension would also help in describing the abrupt voltage drop of the experimental *I–V* characteristics of Fig. 2.

The concept of filamentation was introduced by Adler and coworkers[23] for large samples. However, the sizes of filaments predicted in this classical case do not appear compatible with nanoscale devices.[16] Although the effect on the *I–V* characteristic is similar, the preferential paths identified by our model have a different origin than the filaments of Adler's model: in the present case, they are associated with non-equilibrium energy transport, and not with a change in the quality or charge of the traps. The paths just indicate preferred sequences of connected nodes without any reference to the filament size.

This microscopic picture complies with the aforementioned statistical variability. We can, in fact, interpret the network as a circuit of non-linear resistances that can conceptually be reduced to a single resistor, whose *I–V* characteristic depends on the node positions. Thus, prescribing the same current gives origin to different potential drops for different node placements. The sensitivity of the threshold condition to the particular location of the nodes is significantly reduced for longer cutoff distances, becoming of the order of $\sim$1% for V_{th} and $\sim$13% for I_{th} in the limit of large r_{cut} (see Table II). Experimentally, subsequent amorphizations of the same device after burn-in[16] or, equivalently, the amorphization of a number of geometrically identical devices, may originate different configurations of defects inside the amorphous matrix, producing a dispersion in V_{th} and I_{th}.

In conclusion, this letter presents a 3D network model accounting for the effect of disorder in amorphous phase-change materials. A set of balance equations coupled to the Poisson equation is employed to predict the *I–V* relationships; good agreement with recent experimental data is found for phase-change memory devices probing just hundreds of cubic nanometers of GST.[16]

The microscopic analysis of inter-node transitions correlates the threshold-switching event with the concentration of current into preferential hot-electron paths, shunting the electrodes. The stochastic network generation allows one to study statistical fluctuations in the threshold, also pointing out that this variability depends on the interplay between the network configuration and the typical interaction length of the nodes. The model can provide a valid, fast simulation and design tool for more complex geometries, as the node distribution criteria can be adapted to represent arbitrary device architectures.

Finally, we remark that the high-energy conductive paths may produce crystalline nucleation sites. This suggests a thermal extension of the model, possibly establishing a link between the points of view of the crystalline filament and the purely electronic switching theories.

The Italian authors acknowledge valuable discussions with Dr. F. Buscemi and Professor L. Zanni and financial support from the Intel Corporation under Contract No. 3477131/2011. The Illinois authors acknowledge support from the U.S. Office of Naval Research (Grant No. N00014-10-1-0853) and the U.S. National Science Foundation (Grant No. ECCS 1002026).

[1]S. R. Ovshinsky, Phys. Rev. Lett. **21**, 1450 (1968).
[2]M. Wuttig and N. Yamada, Nature Mater. **6**, 824 (2007).
[3]J. Yao, Z. Sun, L. Zhong, D. Natelson, and J. M. Tour, Nano Lett. **10**, 4105 (2010).
[4]H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. Chen, and M.-J. Tsai, Proc. IEEE **100**, 1951 (2012).
[5]A. Shindome, Y. Doioka, N. Beppu, S. Oda, and K. Uchida, Jpn. J. Appl. Phys. **52**, 04CN05 (2013).
[6]A. Sebastian, A. Pauza, C. Rossel, R. M. Shelby, A. F. Rodríguez, H. Pozidis, and E. Eleftheriou, New J. Phys. **13**, 013020 (2011).
[7]R. G. Neale and J. Aseltine, IEEE Trans. Electron Devices **20**, 195 (1973).
[8]D. Kau, S. Tang, I. Karpov, R. Dodge, B. Klehn, J. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K. wei Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro, and G. Spadini, IEDM - Tech. Dig. **2009**, 1–4.
[9]H. -S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, Proc. IEEE **98**, 2201 (2010).
[10]K. Tsendin, Phys. Status Solidi B **246**, 1831 (2009).
[11]D. Ielmini, Phys. Rev. B **78**, 035308 (2008).
[12]E. Piccinini, A. Cappelli, F. Buscemi, R. Brunetti, D. Ielmini, M. Rudan, and C. Jacoboni, J. Appl. Phys. **112**, 083722 (2012).
[13]C. Jacoboni, E. Piccinini, F. Buscemi, and A. Cappelli, Solid-State Electron. **84**, 90 (2013).
[14]M. Nardone, V. G. Karpov, D. C. S. Jackson, and I. V. Karpov, Appl. Phys. Lett. **94**, 103509 (2009).
[15]F. Xiong, A. D. Liao, D. Estrada, and E. Pop, Science **332**, 568 (2011).
[16]F. Xiong, M.-H. Bae, Y. Dai, A. D. Liao, A. Behnam, E. A. Carrion, S. Hong, D. Ielmini, and E. Pop, Nano Lett. **13**, 464 (2013).
[17]N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1979).
[18]D. Ielmini and Y. Zhang, Appl. Phys. Lett. **90**, 192102 (2007).
[19]D. Ielmini and Y. Zhang, J. Appl. Phys. **102**, 054517 (2007).
[20]G. B. Beneventi, L. Guarino, M. Ferro, and P. Fantini, J. Appl. Phys. **113**, 044506 (2013).
[21]See supplementary material at http://dx.doi.org/10.1063/1.4819097 for the complete derivation and discussion of the equations and the solution scheme, the estimation of the amorphous bit size, the comparison of the results from the present model with those obtained with the HD model, and three animations describing the microscopic evolution of the network as a function of the prescribed current.
[22]P. Ramachandran and G. Varoquaux, Comput. Sci. Eng. **13**, 40 (2011).
[23]D. Adler, M. S. Shur, M. Silver, and S. R. Ovshinsky, J. Appl. Phys. **51**, 3289 (1980).

Supplementary materials for Conductive preferential paths of hot carriers in amorphous phase-change materials

Andrea Cappelli,[1] Enrico Piccinini,[2, a)] Feng Xiong,[3] Ashkan Behnam,[3] Rossella Brunetti,[1] Massimo Rudan,[2] Eric Pop,[3] and Carlo Jacoboni[1]

[1)] *Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/A, I-41125 Modena, Italy*

[2)] *"E. De Castro" Advanced Research Center on Electronic Systems (ARCES), Università di Bologna, Via Toffano 2/2, I-40125 Bologna, Italy*

[3)] *Department of Electrical and Computer Engineering, Micro and Nanotechnology Lab University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

[a)] Author to whom correspondence should be addressed. Electronic mail: enrico.piccinini@unimore.it

I. DERIVATION OF THE CHARGE-FLUX AND OF THE ENERGY-FLUX BALANCE EQUATIONS

Let us consider a network of $N+2$ sites as in Figure S1. The two nodes labeled 0 and $N+1$ are special nodes playing the role of contacts for the device.

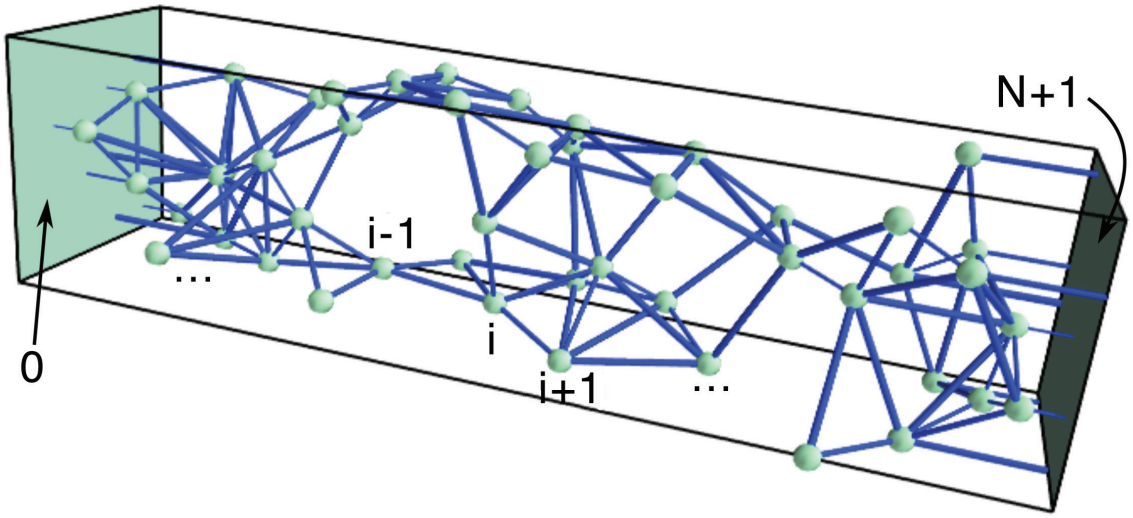

FIG. S1. Schematics of the simulated domain. The shaded surfaces represent the source and the drain contact, respectively. Distances between contacts and internal nodes are calculated along the perpendicular line.

We want to write balance equations for the charge- and energy- fluxes among these nodes. The symbols have the same meaning as in the main text. Let $\mathrm{d}t$ be an infinitesimal time during which a fraction of the population of each node transfers to other nodes. We define the following quantities with respect to the generic i-th node:

1. population flowing out of node: $\mathrm{d}t\left[n_i\sum_j S_{ij}+\frac{I}{(-q)}\delta_{N+1,i}\right]$;
2. population flowing into node: $\mathrm{d}t\left[\sum_j n_j S_{ji}+\frac{I}{(-q)}\delta_{0,i}\right]$;
3. energy flowing out of node: $\mathrm{d}t\,\varepsilon_i\left[n_i\sum_j S_{ij}+\frac{I}{(-q)}\delta_{N+1,i}\right]$;
4. energy flowing into node: $\mathrm{d}t\Big\{\sum_j n_i S_{ji}\Big[\varepsilon_j-q(\varphi_j-\varphi_i)\Big]+\frac{I}{(-q)}\varepsilon_i\delta_{0,i}\Big\}$;
5. energy to relax: $\varepsilon_i^R=\mathrm{d}t\sum_j n_j S_{ji}\Big[(\varepsilon_j-\varepsilon_{i,eq})-q(\varphi_j-\varphi_i)\Big]+\Big[n_i\left(1-\mathrm{d}t\sum_j S_{ij}\right)-\mathrm{d}t\frac{I}{(-q)}\delta_{N+1,i}\Big](\varepsilon_i-\varepsilon_{i,eq})$;

6. relaxed energy: $\varepsilon_i^R f_R = \varepsilon_i^R\Big[1-\exp(-\mathrm{d}t/\tau_R)\Big]$.

where the summations do not take the case $i=j$ into account. By means of these definitions we can write the population balance for the i-th node:

$$n_i(t+\mathrm{d}t) = n_i(t) - \mathrm{d}t\left[n_i\sum_j S_{ij} + \frac{I}{(-q)}\delta_{N+1,i}\right] + \mathrm{d}t\left[\sum_j n_j S_{ji} + \frac{I}{(-q)}\delta_{0,i}\right], \tag{S1}$$

and the energy balance:

$$\begin{aligned} n_i(t+\mathrm{d}t)\varepsilon_i(t+\mathrm{d}t) = n_i(t)\varepsilon_i(t) - \mathrm{d}t\, n_i\varepsilon_i\sum_j S_{ij} + \mathrm{d}t\sum_j n_j S_{ji}\Big[\varepsilon_j - q(\varphi_j-\varphi_i)\Big] + \\ + \mathrm{d}t\frac{I}{(-q)}\varepsilon_i\Big[\delta_{0,i}-\delta_{N+1,i}\Big] - \Bigg\{\mathrm{d}t\sum_j n_j S_{ji}\left[(\varepsilon_j-\varepsilon_{eq}) + q_e(\varphi_i-\varphi_j)\right] + \\ + \left[n_i\Big(1-\mathrm{d}t\sum_j S_{ij}\Big) - \mathrm{d}t\frac{I}{(-q)}\delta_{N+1,i}\right](\varepsilon_i-\varepsilon_{i,eq})\Bigg\} \times \Big[1-\exp(-dt/\tau_R)\Big]. \end{aligned} \tag{S2}$$

The l.h.s. and the last factor at the r.h.s. of Eq. S2 can be expanded to the first order to yield:

$$\begin{aligned} n_i(t+\mathrm{d}t)\varepsilon_i(t+\mathrm{d}t) &= \left[n_i(t)+\frac{\mathrm{d}n_i}{\mathrm{d}t}\mathrm{d}t\right]\left[\varepsilon_i(t)+\frac{\mathrm{d}\varepsilon_i}{\mathrm{d}t}\mathrm{d}t\right] \\ &= n_i(t)\varepsilon_i(t) + n_i(t)\frac{\mathrm{d}\varepsilon_i}{\mathrm{d}t}\mathrm{d}t + \frac{\mathrm{d}n_i}{\mathrm{d}t}\varepsilon_i(t)\mathrm{d}t + \frac{\mathrm{d}n_i}{\mathrm{d}t}\frac{\mathrm{d}\varepsilon_i}{\mathrm{d}t}\mathrm{d}t^2, \end{aligned} \tag{S3}$$

and

$$f_R = 1-\exp(-\mathrm{d}t/\tau_R) \approx 1-1+\frac{\mathrm{d}t}{\tau_R} = \frac{\mathrm{d}t}{\tau_R}. \tag{S4}$$

Neglecting the infinitesimal terms, and remembering that $\mathrm{d}\left(n_i\varepsilon_i\right)/\mathrm{d}t = n_i\mathrm{d}\varepsilon_i/\mathrm{d}t + \varepsilon_i\mathrm{d}n_i/\mathrm{d}t$, Eqs. S1 and S2 can be rewritten as:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = -n_i\sum_j S_{ij} - \frac{I}{(-q)}\delta_{N+1,i} + \sum_j n_j S_{ji} + \frac{I}{(-q)}\delta_{0,i} = 0, \tag{S5}$$

$$\begin{aligned} \frac{\mathrm{d}(n_i\varepsilon_i)}{\mathrm{d}t} = &- n_i\varepsilon_i\sum_j S_{ij} + \sum_j n_j S_{ji}\Big[\varepsilon_j - q(\varphi_j-\varphi_i)\Big] + \frac{I}{(-q)}\varepsilon_i\Big[\delta_{0,i}-\delta_{N+1,i}\Big] + \\ &- n_i\frac{\varepsilon_i-\varepsilon_i^*}{\tau_R} = 0, \end{aligned} \tag{S6}$$

where we have imposed $\mathrm{d}n_i/\mathrm{d}t = 0$ and $\mathrm{d}(n_i\varepsilon_i)/\mathrm{d}t = 0$ since we are interested in the steady-state solution. With little, straightforward manipulation Eqs. S5 and S6 become Eqs. 2 and

3 of the main text. The latter expression can then be recast into a more compact and handy formula by multiplying Eq. S5 times ε_i and substituting it into Eq. S6:

$$\varepsilon_i = \varepsilon_{i,eq} + \tau_R \sum_j \frac{n_j}{n_i} S_{ji} \Big[\varepsilon_j - \varepsilon_i - q(\varphi_j - \varphi_i) \Big] \tag{S7}$$

II. SOLUTION SCHEME

Formulas S5 and S7 represent two sets of equations whose unknowns are the nodal populations n_i's, electrostatic potentials φ_i's, and energies ε_i's. Since the number of nodes is $N+2$, we need $3(N+2)$ independent equations to solve the problem.

Equation S5 represents the well-known Kirchoff's Current Law applied to each node. Since any network, despite its complexity, can always be reduced to a single resistor, only a set of $N+1$ useful equations can be derived, the remaining one not adding any further information.

We already pointed out that nodes 0 and $N+1$ play the role of the two contacts. They are assumed to be planar, in the sense that distances between one of them and any of the internal nodes is calculated as the minimum distance between the internal node and the appropriate face of the device. Additional boundary conditions apply to the contacts. In particular, we can assume that their population is always the equilibrium population ($n_0 = n_{N+1} = n_{eq}$). Moreover, we can also suppose that carriers enter the device from the source contact being *cold*, i.e., $\varepsilon_0 = \varepsilon_{eq}$. On the other side of the device the drain contact is supposed to receive *hot* carriers, and its nodal energy depends on the efficiency of the relaxation mechanism at the contact interface. For a maximally efficient energy transfer we would suppose $\varepsilon_{N+1} = \varepsilon_{eq}$; otherwise, this value is unknown and must be determined by solving the problem. For the sake of truth, since the fraction of carriers entering back into the device from the drain contact is negligible at any conduction regime, very similar results are found in both cases. In order to discuss the problem under the most general conditions we consider in the followings the energy of the drain contact as unknown.

The considerations above reduce the number of unknowns to $3N+3$, a number that can be further reduced to $3N+2$ as we recall the arbitrariness of the electrostatic potential, which allows to set, for instance, $\varphi_0 = 0$. From Eqs. S5 and S7 only $2(N+1)$ equations are

available. The missing N relationships stem from the Poisson equation:

$$\nabla^2\varphi(x,y,z) = -\frac{\rho(x,y,z)}{\epsilon}, \tag{S8}$$

which links the electrostatic potential to the charge, thus the population, of each node.

A finite-element implementation of the latter over a grid of M points ($M \gg N$) is adopted, and a buffer region surrounds laterally the device to avoid the influence of artificial boundary conditions. If the electrostatic potentials of some nodes are known from an independent source of information, the finite-element solution of Eq. S8 yields the charge on these nodes and the electrostatic potential elsewhere, providing the required N supplementary relationships.

Algorithm 1 Finds the voltage drop, for a given input current I

Require: Device and material parameters, network topology, prescribed input current I

while $\left\| n_i^{(s+1)} - n_i^{(s)} \right\| > \Delta_n$ **do**
  $s \leftarrow s + 1$
  while $\left\| f(\varphi_i, \varepsilon_i, n_i^{(s)}) \right\| > \Delta_{tol}$ **do**
    NRP for Eq. S5 at fixed $(\varepsilon_i, n_i^{(s)}) \rightarrow \varphi_i$
    NRP for Eq. S7 at fixed $(\varphi_i, n_i^{(s)}) \rightarrow \varepsilon_i$
  end while
  PS using $(\varepsilon_i, \varphi_i) \rightarrow n_i^{(s+1)}$
end while
return $V = \varphi_{N+1}^{(s+1)} - \varphi_0^{(s+1)}$

Equations S5 and S7 are strongly non-linear due to the exponential form of S_{ij} and S_{ji} (see Eq. 1 in the main text), and also depend on the populations, that come from the Poisson equation, which needs the electrostatic potentials of the nodes as input values. Iterative methods of a different kind must be applied for finding potentials and energies on one side, and for populations on the other side. The Newton-Raphson procedure (NRP) for sets of non-linear equations has been implemented in the first case, assuming a constant population. It proved to be convenient to address the two sets S5 and S7 separately, keeping in the first case the energies of the nodes fixed to obtain an updated solution for φ_i, and keeping the new potentials fixed in the second case in order to update ε_i. These steps are cycled until a global convergence is achieved for both potentials and energies (overall NRP,

or oNRP). Once the oNRP phase is finished, the population of each node is updated by evaluating the charge Q_i of the nodes through a Poisson step (PS), and by applying the relationship $n_i = Q_i/(-q) + n_{eq}$. Pairs of oNRP's and PS's are repeated until all of the unknowns are determined within their desired precision (see Algorithm 1). The voltage drop $V = \varphi_{N+1} - \varphi_0$ is the final outcome of the method for any prescribed input current I.

III. ESTIMATE OF THE DIMENSIONS OF GST-CNT DEVICES

The model has been validated by checking calculated current-voltage characteristics against experimental data of carbon-nanotube-contacted self-aligned $Ge - Sb - Te$ memory devices (CNT-GST), fabricated as in Ref. 1. In order to achieve reliable comparisons, it is important to feed the model the appropriate size of the chalcogenide volume.

The manufacturing process of CNT-GST devices implies an inherent random distribution of their lengths[1,2]. A first-order estimate of these values stems from the observation that the onset of a negative differential resistance is obtained when a threshold field F_{th} is reached. Literature values for such a field report typically $F_{th} \approx 0.5$—1 MV/cm (50—100 V/μm) for CNT-GST devices[1]. For the present case, we have assumed the intermediate field $F_{th} \sim 0.9$ MV/cm (90 V/μm) for all cases considered. The simple calculation V_{th}/F_{th} leads to the three estimated lengths 40 nm, 81 nm and 91 nm for the three devices included in the analysis.

An estimate of the cross-sectional area is more complicated, as the active chalcogenide spot is expected to extend on a section larger than the CNT area[1]. Typical values for CNT-GST devices lie in the range 90—180 nm^2. In the calculations we have assumed a constant 10×10 nm^2 cross section for all devices, also considering the trade-off with the computational burden. This value may sometime underestimate the real effective cross-section, as in the case of devices featuring the highest threshold currents. As we have also commented in the main text, such underestimation may be compensated by a larger variability of the relaxation time τ_R. However, different choices of the threshold field F_{th} or of the cross-sectional area would affect results only quantitatively without influencing the physics underneath.

IV. COMPARISON WITH THE MICROSCOPIC HYDRODYNAMC-LIKE MODEL

The network model and the microscopic HydroDynamic (HD) model[3] share the same physical background for charge transport. It is therefore possible to qualitatively compare the parameters used in the two models to fit the I–V characteristics. It is important to recall that the HD model is a continuum-medium model and the momenta of transport equations are obtained integrating a distribution function, whereas the network model makes use of a spatially-discrete approach and considers directly the average energy of the carriers. As a consequence, the common background of the models makes the comparison of the parameters possible only from a qualitative standpoint.

The experimental I–V characteristic of device L40 (see main text) has been fitted also with the HD model (Fig. S2); parameters are listed in Table TI along with those obtained with the same model for traditional mushroom cells[3]. For comparison purposes, we have also reported the parameters of the statistical best-fit of device L40 with the network model.

The HD model considers a density of states of traps Γ populated according to the distribution function. By integrating it over the energy and the device volume we obtain a carrier concentration of the order of 1—5×10^{19} cm^{-3}, when $\Gamma \sim 10^{20}$ cm^{-3} eV^{-1}. In the network model, a node can in principle host an arbitrary number of carriers. If we assume the simplifying hypothesis of one carrier per node under the initial equilibrium condition, we obtain the carrier concentration $\sim 10^{19}$ cm^{-3}, which well compares to that obtained with the HD model for device L40.

TABLE TI. Parameters of the HD mode for device L40 and for a mushroom cell with cross-sectional area 1000 nm^2 and length 27 nm, and for the network model for device L40 (from the main text).

HD model			Network model	
	L40	Ref. 3		L40
Γ (cm^{-3} eV^{-1})	0.4×10^{20}	1.18×10^{20}	n (cm^{-3})	1.2×10^{19}
τ_0 (fs)	50	12	τ_0 (fs)	49.5
τ_R (fs)	17	78	τ_R (fs)	56.5
Δz (nm)	3.2	7	r_{cut} (nm)	6
			ℓ (nm)	2.9

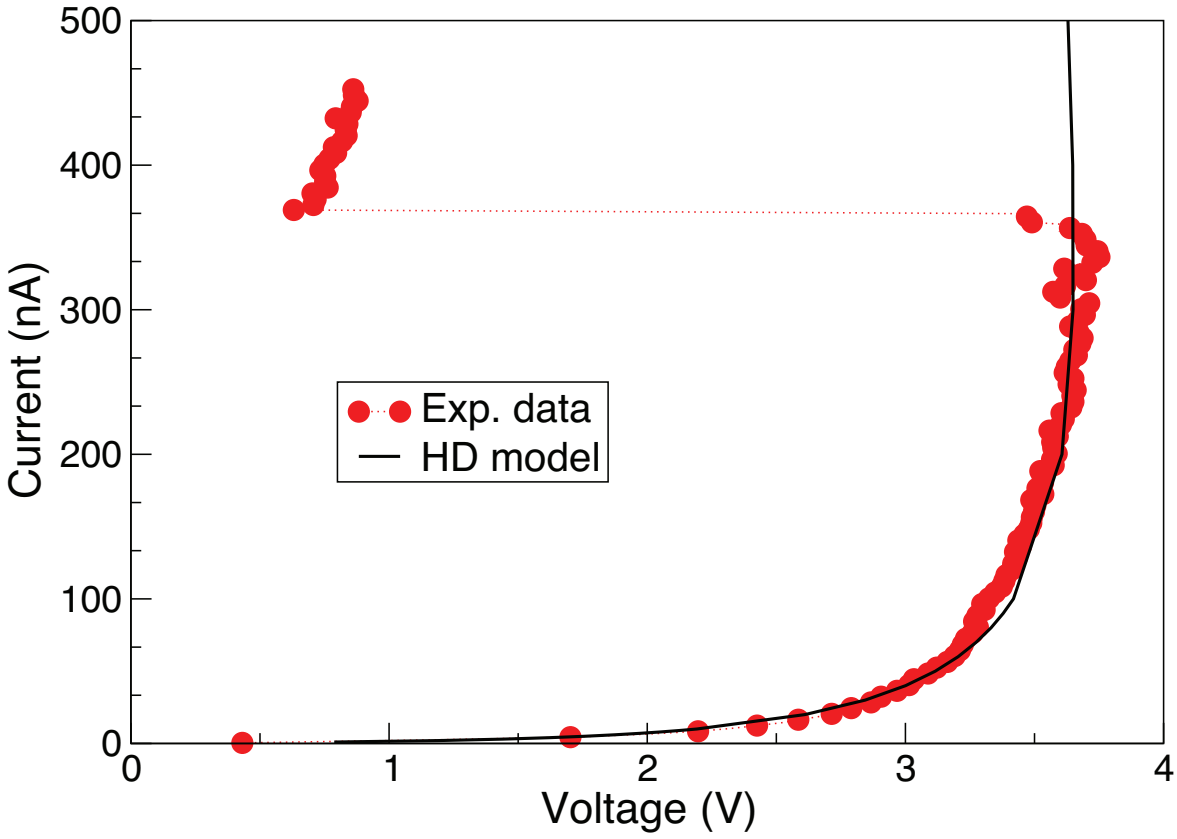

FIG. S2. I–V characteristic (line) obtained with the HD model and the parameters listed in Table TSTI and experimental data for device L40 (dots).

Next, we find similar values also for the characteristic time of the transition process τ_0 and the energy-relaxation time τ_R, that are estimated by both models in the range 10-100 fs. This result confirms once again the common nature of the physical processes of the transport framework, with numerical differences depending only on the specific details of their implementation.

A slightly more complicated reasoning stands behind parameter Δz in the HD model, which makes the comparison with distances r_{cut} and ℓ in the network model less direct. Following Ref. 4, Δz is the average transition distance and is basically determined fitting the slope of the exponential portion of the I–V curve when one assumes that the top of the energy barrier governing the transition is at the midpoint of the transition line. In the network model, these two ideas have been split and the two parameters r_{cut} and ℓ must be determined independently. Going into further details, r_{cut} represents the maximum

transition distance. Given the constraint of a minimum distance r_{min} between two nodes, which is a condition to avoid node overlap, and assuming a uniform transition probability within a spherical shell from r_{min} to r_{cut}, the average transition distance can be roughly estimated as $3/4 \times (r_{cut}^4 - r_{min}^4)/(r_{cut}^3 - r_{min}^3)$ which is up to 7% larger than $3/4 r_{cut}$ for $r_{min} \lesssim r_{cut}/2$. For the test case of device L40, according to the reasoning above, we can estimate a transition distance of 4.7 nm for $r_{min} \sim r_{cut}/2$, a value 50% larger than what predicted by the HD model.

As for the parameter ℓ (effective correction to the height of the energy barrier), we recall that in the network model this correction depends on the distance of the nodes, being ℓ its maximum value. In other words, for transitions where $r_{ij} < 2\ell$ we assume that the maximum of the energy barrier is located at the midpoint of the transition line, i.e., $r_{ij}/2$, as in the HD model; on the contrary, for longer transitions we assume a fixed position at distance ℓ. Though the procedure is tricky, we can still indirectly estimate ℓ from to the slope of the exponential portion of the I–V characteristic by weighting short and long transitions. The ratio ℓ/r_{cut} close to 0.5 indicates that the hypothesis on the position of the maximum of the energy barrier in the HD model is confirmed also in the network model.

V. MOVIES

Three movies are provided along with this work. Each of them shows the microscopic network for a particular different random node configuration, with color codes representing the energies ε_i on the nodes and the intensities I_{ij} of the net currents among the nodes, normalized to the total current I. Next to the movie, an I–V characteristic curve is reported, with the moving dot showing, for each configuration frame, its corresponding values of voltage V and current I. The 3D visualization software Mayavi[5] was employed to produce these movies and the screenshots in the main text.

Movie #1 is the movie from which the screenshots of the main text are extracted. For low currents, no unique continuous path is observed to directly connect the electrodes.This movie clearly highlights the effect of "channeling" when higher currents are prescribed: even for a current level of a few nanoamperes, the path in the upper part of the device begins to get enriched. For slightly higher I, about $\sim$ 1—50 nA, it becomes evident that such current increase in the main path happens at the expense of a flow decrease through other

secondary links, particularly the two consecutive ones in the bottom part of the sample.In the last frames, the "channeling" is almost total, and a clear path connects the electrodes while little or no flow is observed in the other lines.

Movie #2 is obtained for a value of $r_{cut} = 8$ nm, higher than the 6 nm used for the first movie. This has a clear effect: the current can be spread among more available connections, thus a more uniform low-current picture is obtained; initially no paths are strongly favored over the others simply on a geometrical basis. Anyway, a region of low node density in the middle of the device ($y = 15$—22 nm) makes it necessary to have slightly above-average values of transition rates, in order to sustain the prescribed current, since not many parallel viable paths are available in that portion of the sample. For higher currents ($I \geq 10$ nA) the current distribution starts to change rapidly. Electrons in the middle region heat up as a result of the higher "current densities" associated to it, thus allowing even more current to flow, and triggering a positive feedback. Indeed, this area is the point where the formation of a single preferential path begins for this device. In the subsequent frames, the path spreads from the center toward both contacts, linking high-energy nodes, and finally connects the device end-to-end for $I \sim 7$ nA.

Movie #3 shows the case of a 91-nm-long device, with $r_{cut} = 6.5$ nm. When r_{cut} is small compared to the device length, segmentation of the device is present, i.e., regions densely populated with nodes intertwined with others with a very limited number of nodes. For the case at hand, three highly-populated segments can be identified, two close to the two contacts and one in the middle of the device. As the current is increased from low values to the threshold condition, one path in the poorly populated regions progressively drains almost the entire current. This situation makes transport within highly-populated segments independent from each other, as if long devices were consecutive sequences of short ones. Moreover, it is also worth noticing that the preferred route leading to threshold switching can be further modified by the transport conditions. In particular, we point out that before the threshold condition there are two competitive paths converging to a node at $y \sim 45.5$ nm that gathers the current from the surrounding paths. After switching, the interplay between them is such that only one path continues to increase its current, draining it from one link that contributed to create the preferential path up to that moment. This is likely due to the strongly non-linear nature of the balance equations.

REFERENCES

[1]F. Xiong, M.-H. Bae, Y. Dai, A. D. Liao, A. Behnam, E. A. Carrion, S. Hong, D. Ielmini, and E. Pop, Nano Lett. **13**, 464 (2013).

[2]F. Xiong, A. D. Liao, D. Estrada, and E. Pop, Science **332**, 568 (2011).

[3]E. Piccinini, A. Cappelli, F. Buscemi, R. Brunetti, D. Ielmini, M. Rudan, and C. Jacoboni, J. Appl. Phys. **112**, 083722 (2012).

[4]D. Ielmini and Y. Zhang, J. Appl. Phys. **102**, 054517 (2007).

[5]P. Ramachandran and G. Varoquaux, Comp. Sci. Eng. **13**, 40 (2011).