Physical aspects of low power synapses based on phase change memory devices

M. Suri, O. Bichler, D. Querlioz, B. Traore, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, B. de Salvo

Physical aspects of low power synapses based on phase change memory devices

Cite as: J. Appl. Phys. **112**, 054904 (2012); https://doi.org/10.1063/1.4749411
Submitted: 06 March 2012 • Accepted: 09 August 2012 • Published Online: 06 September 2012

Manan Suri, Olivier Bichler, Damien Querlioz, et al.

Physical aspects of low power synapses based on phase change memory devices

Manan Suri,[1,a)] Olivier Bichler,[2] Damien Querlioz,[3] Boubacar Traoré,[1] Olga Cueto,[1] Luca Perniola,[1] Veronique Sousa,[1] Dominique Vuillaume,[4] Christian Gamrat,[2] and Barbara DeSalvo[1]

[1]*CEA – LETI - MINATEC, 38054 Grenoble, France*
[2]*CEA- LIST, Embedded Computers Laboratory, 91191 Gif-sur-Yvette Cedex, France*
[3]*Institut d'Electronique Fondamentale, Univ. Paris-Sud, CNRS, 91405 Orsay, France*
[4]*Institute for Electronics Microelectronics and Nanotechnology (IEMN), CNRS, F-59652 Villeneuve d'Ascq, France*

(Received 6 March 2012; accepted 9 August 2012; published online 6 September 2012)

In this work, we demonstrate how phase change memory (PCM) devices can be used to emulate biologically inspired synaptic functions in particular, potentiation and depression, important for implementing neuromorphic hardware. PCM devices with different chalcogenide materials are fabricated and characterized. The asymmetry between the potentiation and depression behaviors of the PCM is stressed. Detailed multi-physical simulations are performed to study the underlying physics of the synaptic behavior of PCM. A versatile behavioral model and a multi-level circuit-compatible model are developed for system and circuit-level neuromorphic simulations. We propose a unique low-power methodology named the *2-PCM Synapse*, to use PCM devices as synapses in large scale neuromorphic systems. To show the strength of our proposed solution, we efficiently simulated fully connected feed-forward spiking neural network capable of complex visual pattern extraction from real world data. © *2012 American Institute of Physics*.
[http://dx.doi.org/10.1063/1.4749411]

I. INTRODUCTION

Research in the field of biologically inspired neuromorphic circuits, with the goal of achieving low power, highly parallel, and fault-tolerant systems, has recently gained a lot of interest.[1–8] Neuromorphic hardware contains components that can emulate functions performed by the neurons and the synapses present inside the brain. In order to build functional neuromorphic circuits comparable to the cerebral cortex, an enormous number of neurons (10^{10}) and synapses (10^{14} to 10^{15}) are desired.[9] The idea of achieving such a high synaptic density using pure CMOS synapse circuits is not practical in terms of on-chip silicon area consumption, due to the large number of transistors required (more than 10 transistors per synapse).[10] Hybrid neuromorphic architectures containing CMOS "neuronal" circuits, integrated with nanoscale "synaptic" devices[11] have thus been proposed in recent years. Some nanoscale devices considered for implementing synaptic functions include hybrid organic/nanoparticle transistors,[12] single-electron transistors,[13] carbon-nanotube based structures,[14] ionic/electronic hybrid transistors,[15] and atomic switches.[16] Neurophysiological models[17,18] suggest that a device, in order to emulate synaptic behavior, should have a conductance that can be modulated by the type of stimulus it receives. Since the storage of synaptic weights is believed to play an important role in the functionality of neural networks,[19] the devices emulating synaptic behavior should possess some kind of memory to retain their conductance states (or the so-called synaptic weights). These conditions of conductance modulation and of memory effect can be fulfilled by the class of devices known as resistive memories.[20] Several types of unipolar and bipolar resistive memory technologies such as phase change memory (PCM),[21–23] conductive-bridge (CBRAM) or programmable-metallization cell (PMC),[24,25] and oxide-resistive (OXRAM) memory[26] have been demonstrated as suitable candidates for synaptic applications.

In this work, we focus on the use of PCM devices for synapses due to advantages such as high scalability, CMOS compatibility, good endurance, and good technological maturity compared to other resistive memory technologies.[27] Section II discusses in detail the experiments demonstrating synaptic potentiation and synaptic depression on PCM devices, and some limitations emerging from the emulation of synaptic depression. Section III presents multi-physical simulations used to analyze the experiments described in Sec. II. Section IV introduces a behavioral and a circuit compatible model, for designing and simulating large scale PCM based neural networks. Section V describes a new energy-efficient methodology named the *2-PCM Synapse* to implement synaptic functionality using PCM devices and a solution to overcome the limitations described in Secs. II and III. Finally, in Sec. VI the behavioral model and the *2-PCM Synapse* approach were used to design a large scale neural network for visual pattern extraction application, presented in detail.[22]

II. ELECTRICAL CHARACTERIZATION

When used as a synapse, a PCM device is connected to two spiking CMOS circuits, acting as the pre- and post-synaptic neurons (Fig. 1). The difference between the

a)E-mail: manan.suri@cea.fr.

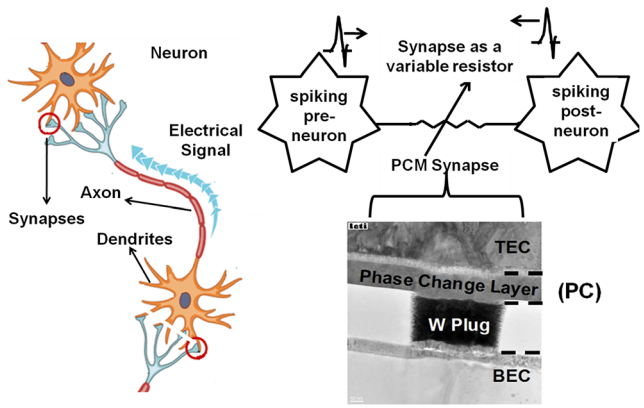

FIG. 1. Illustration of biological synapse and the equivalent PCM synapse in a neural circuit connecting a spiking pre- and post-neuron.

electrical resistivity of the amorphous and the crystalline phases of the chalcogenide can be exploited to use PCM as variable resistor or programmable memristor. The high-resistive amorphous phase is usually defined as the RESET state, while the low-resistive crystalline phase as the SET state. When a bias is applied across the two electrodes of the PCM, current flows through the metallic heater and the chalcogenide layer, causing joule-heating. Depending on the pulse-duration, fall-time edge, and the amplitude of the current flowing through the device, crystalline, amorphous, or partially crystalline and partially amorphous regions can be created inside the chalcogenide layer. If the chalcogenide layer is melted and quenched quickly, it does not get sufficient time to re-organize itself into a crystalline structure and thus amorphous regions are created. If the chalcogenide layer is heated, between the glass-transition and the melting temperature, for sufficiently long time it leads to crystallization.

In this work, Lance-type PCM test devices, with a 100 nm-thick phase change layer and a 300 nm-diameter tungsten plug, were fabricated and characterized. Two different chalcogenide materials were integrated: nucleation-dominated $Ge_2Sb_2Te_5$ (GST) and growth-dominated GeTe.[29] GST and GeTe were chosen to examine how materials with different crystallization parameters would impact the synaptic behavior. For all the measurements, a current limiting resistance of 100 Ω was connected in series with the top pad of the PCM device. The bottom pad was grounded and the signal was applied to the top pad.

Throughout this paper, we will refer to an increase in synaptic conductance as synaptic potentiation (or long term potentiation, LTP[18]), and to a decrease in synaptic conductance as synaptic depression (or long term depression, LTD[18]). Figs. 2(a) and 2(b) show the measured current-voltage (I-V) and resistance-current (R-I) curves, for GST and GeTe PCM devices. In Fig. 2(a) (I-V curve), the device was initially reset to the amorphous state. As the applied voltage is increased, electronic switching occurs at the threshold voltage, and the amorphous region becomes conductive (the so called volatile electronic-switching[30]). Fig. 2(b) (R-I curve) illustrates well the difference between a SET state (crystalline), a "potentiated" state (partially crystalline), and a strong RESET state (large amorphous region). At the beginning of the test, the device was reset to a high resistive amorphous state using a strong reset-pulse (7 V, 100 ns, rise/fall time = 10 ns). This was followed by the application of a programming pulse. After the programming pulse, a small reading voltage of 0.1 V was applied to measure the device resistance.

Fig. 2(c) shows the characteristic resistance-voltage (R-V) curves for GST based PCM devices, for different programming pulse widths. For a given pulse amplitude, the resistance decreases much faster for longer pulse widths. Thus, by using the combination of right pulse width and right pulse amplitude, PCM resistance (or conductance) can be modulated, as an analog to synaptic weights. For both Figs. 2(b) and 2(c) the rise and fall times of the set pulse are always 10 ns. The reset-to-set transition in GeTe is more abrupt

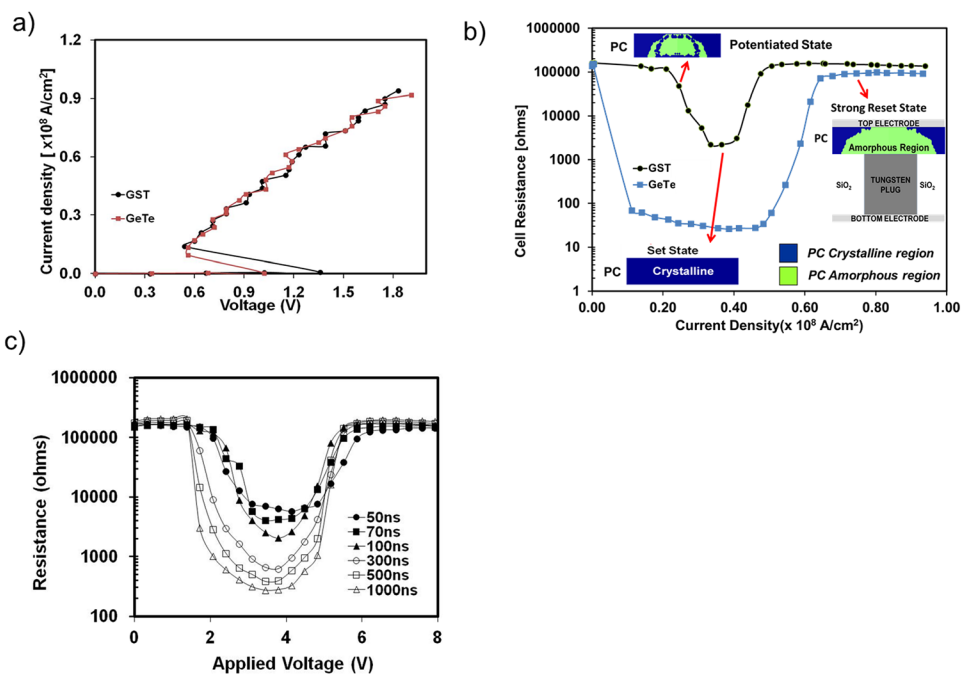

FIG. 2. (a) I–V characteristics for PCM devices with 100 nm thick GST and GeTe layer starting from initially amorphous phase. (b) R-I characteristics of GST and GeTe PCM devices, with inset showing the PCM phase of intermediate resistance states. (c) R-V curves for GST devices with six different pulse widths. Read pulse = 0.1 V, 1 ms. The legend shows pulse widths.

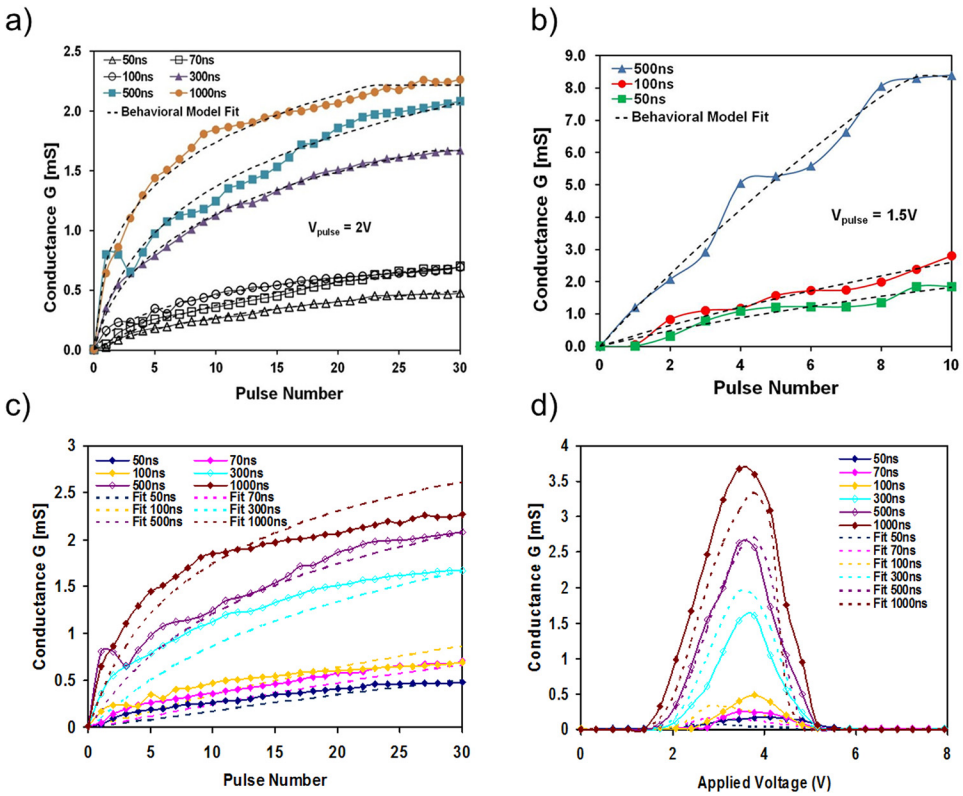

FIG. 3. (a) Experimental LTP characteristics of GST PCM devices. For each curve, first a reset pulse (7 V, 100 ns) is applied followed by 30 consecutive identical potentiating pulses (2 V). Dotted lines correspond to the behavioral model fit described in Eqs. (3a) and (3b). (b) Experimental LTP characteristics of GeTe PCM devices. (c) Circuit-compatible (Sec. IV B) based LTP simulations for GST devices. (d) Circuit-compatible (Sec. IV B) simulations of the conductance evolution as a function of the applied voltage for GST devices with six different pulse widths. The legends in Figs. 3(a)–3(d) indicate pulse widths.

compared to GST; GeTe being a growth dominated material crystallizes much faster compared to GST which is nucleation dominated. On the other hand the set-to-reset transition appears to show more gradual resistance change for both GST and GeTe as it is possible to obtain different intermediate resistance values by controlling the volume of the amorphous region created inside the phase change layer post-melting.

LTP can be emulated if the PCM is progressively programmed along the RESET-to-SET transition (or amorphous-to-crystalline). LTD can be emulated if the PCM is progressively programmed from the SET-to-RESET transition (or crystalline-to-amorphous). In order to emulate spiking neural networks (SNN), it is desired that both LTP and LTD can be achieved by application of simple and identical pulses. Generation of complex types of spikes or pulses would require additional complexity in the neuron circuit design. Two types of pulses can be defined for our purpose: "depressing" or amorphizing pulse (reset) and a "potentiating" or partially crystallizing pulse (a weak set).

Figs. 3(a) and 3(b) show LTP-like conductance variation of PCM devices with GST and GeTe, respectively. Initially, the devices were programmed to a high resistive state by using a strong depressing pulse (7 V, 100 ns). This was followed by the application of several identical potentiating pulses, which are simple rectangular voltage pulses with a rise and fall times of 10 ns (2 V for GST, 1.5 V for GeTe). The voltage amplitude of the potentiating pulses is chosen such that the resulting current flowing through the device is just sufficient to switch the device and cause minimum amount of crystallization with the application of each pulse. Nucleation-dominated behavior leads to a more gradual conductance change in GST, when compared to GeTe (GeTe being growth dominated). The saturation of the conductance-programming window in GeTe occurs in less than a third of the total number of pulses required for GST. From the viewpoint of storage capacity (in the form of synaptic weights) inside a neural network, a GST synapse seems superior to a GeTe synapse, as GST offers a higher number of intermediate conductance states.

Fig. 4 shows the emulation of LTD-like behavior on PCM devices. The devices were first crystallized by applying a strong potentiating pulse (2 V, 1 μs), followed by the application of several identical depressing pulses (7 V, 50 ns). It was not possible to obtain a gradual synaptic depression by

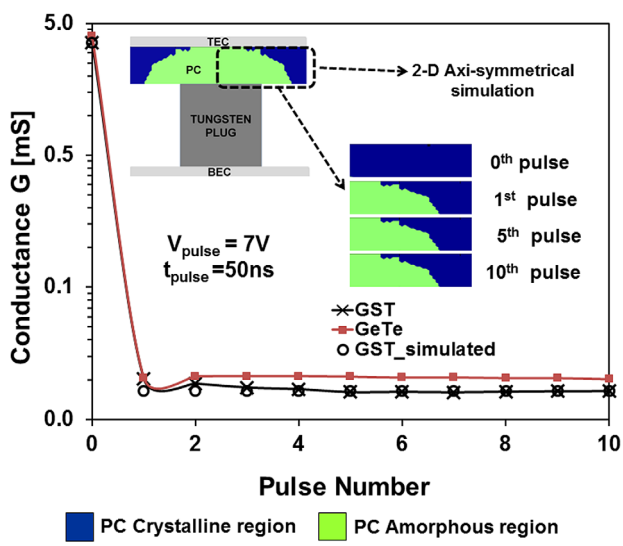

FIG. 4. Experimental LTD characteristics of GST and GeTe PCM devices. Inset shows simulated phase morphology of GST layer after the application of consecutive depressing pulses.

applying identical depressing pulses. The LTD experiment seems more like an abrupt binary process, with negligible intermediate conductance states. Multi-physical simulations (described in Sec. III) were performed to interpret the LTD experiment. From these simulations (shown embedded in Fig. 4), we observed that the volume of the molten region created after the application of each pulse remains almost the same if the pulses are identical. We also performed LTD simulations with consecutive pulses of increasing amplitude (not shown in this paper), and observed a gradual decrease in the device conductance with the application of each pulse, in agreement with other papers.[23] A strong depressing pulse would melt a larger region compared to a weak depressing pulse, creating a larger amorphous cap and a lower value of final device conductance. Thus, in order to obtain gradual synaptic depression behavior, the amplitude of the consecutive pulses should increase progressively. This is also in agreement with the set-to-reset transition seen in Fig. 2(b).

Innovative pulse sequences like the ones suggested in Ref. 23 could help in achieving LTD with multiple intermediate states. Nevertheless, implementing such complex pulse-schemes with varying amplitudes can lead to practical problems, such as capacitive line charging and high power dissipation, when implemented in large scale neural systems. Moreover, the generation of non-identical pulses would lead to an augmented complexity in the design of the CMOS neuron circuits. Additionally, emulation of LTD or synaptic depression on PCM devices is significantly more energy consuming compared to the emulation of synaptic potentiation, as the former requires amorphization (occurring at a higher temperature compared to crystallization). In Sec. V, we will propose a new solution to emulate LTD with PCM devices, thus overcoming the limitations discussed above.

III. PHYSICAL SIMULATIONS

In this section, we present electro-thermal simulations to study the underlying physics of the LTP and LTD experiments described in Sec. II. The motivation for such a study is to optimize the synaptic characteristics by engineering the crystallization properties of the phase change materials. Crystallization properties of phase change materials can be altered to some extent by doping,[33,34] modifying the stoichiometric compositions,[35–37] and interface engineering.[50] A better control over the crystallization parameters by materials engineering is also in the interest of multi-level PCM implementation[38] and enhanced data-retention characteristics.[39] All the simulations were performed for the GST based PCM devices. The electro-thermal simulator, developed in MATLAB and C++, was described in detail in Ref. 31. The simulations were performed in the 2D axi-symmetrical coordinate reference system. In all the simulations, a series load resistance, R_s, and a parasitic capacitance C_p (Fig. 5(a)) were considered. Figs. 5(b)–5(e) show the time evolution of several simulated electro-thermal parameters for a PCM device (initially amorphous), during the application of a potentiating pulse (2.1 V, 30 ns). The evolution of the voltage drop across the device, the current, the resistance, and the maximum cell temperature is shown.

Fig. 6 shows a time snapshot of the GST phases during the application of a depressing pulse (8 V, 25 ns), starting from an initially crystalline state. The selected time is the beginning of the quenching process (i.e., the falling edge of the reset pulse, Fig. 6(a)). Before this instant, a mushroom-shaped melted mass of GST (brown color) in the region right above the *W*-plug is seen. As the quenching progresses, the GST amorphizes, moving inwards from the melted-crystalline interface towards the center of the mushroom.

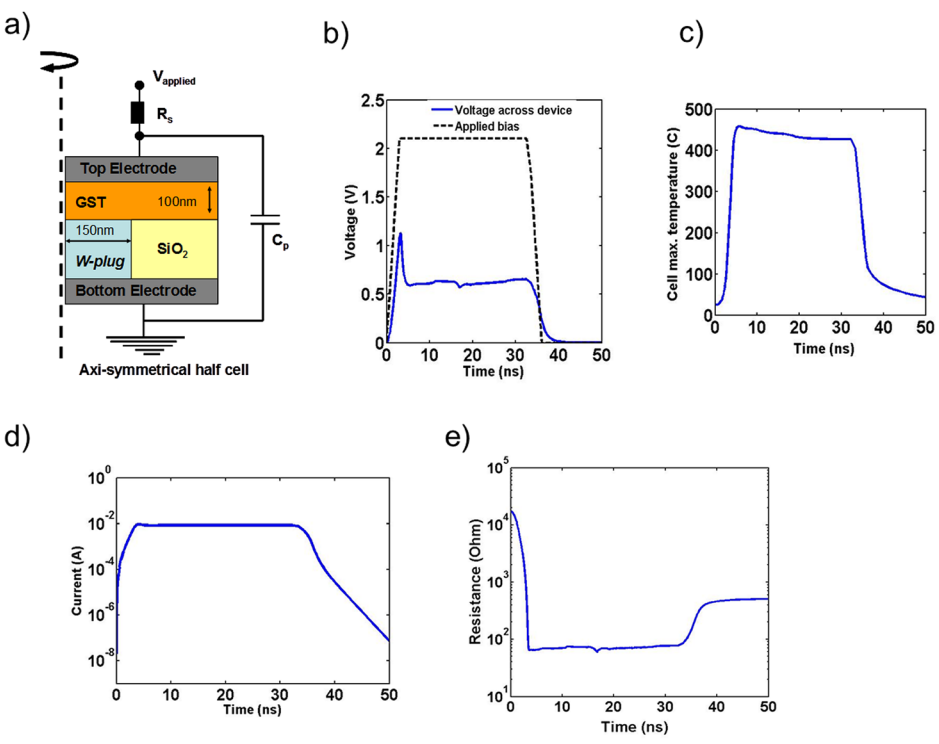

FIG. 5. (a) 2D Axi-symmetrical half cell description used for physical simulations. (b) Simulated time evolution of applied voltage pulse and drop across the device for a potentiating pulse. (c) Simulated maximum temperature in GST layer with the applied pulse. (d) Simulated current passing through the device during the applied pulse. (e) Simulated resistance of the device with the applied pulse.

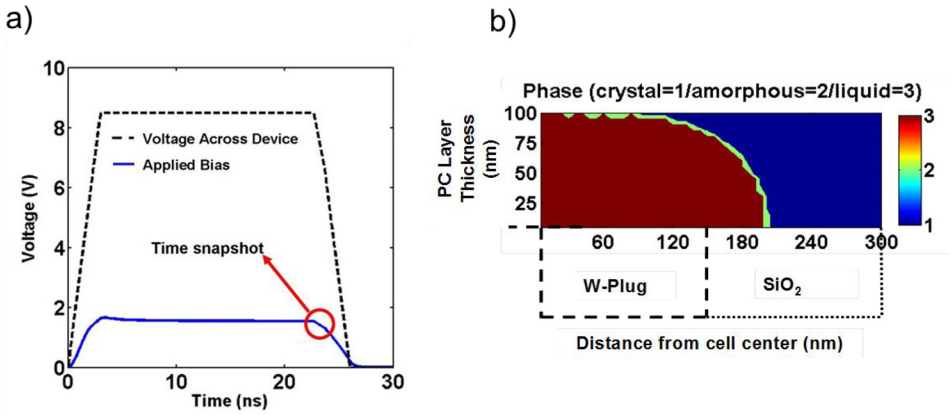

FIG. 6. (a) Simulated depressing (reset) pulse indicating the instance of time snapshot. (b) Time snapshot of the simulated phase morphology of the GST phase change layer.

Formation of thin amorphous GST (green color) can be seen at melted-crystalline interface in Fig. 6(b).

The first few points in the LTP curves (for GST) shown in (Fig. 3(a)) are crucial in determining the total number of intermediate conductance states that can be obtained for a given programming-window. The maximum change in conductance was indeed observed to occur during the application of the first 5 pulses for GST. In order to better understand the variation of conductance during the application of first few pulses in the LTP experiment, we performed the simulations shown in Fig. 7. The nucleation rate (I) and growth velocity (V) in the phase change layer were modeled using Eqs. (1) and (2), respectively, adapted from Ref. 31

$$I = N_a \gamma O_n Z \exp\left(-\frac{\Delta G^*}{kT}\right), \tag{1}$$

$$V = \gamma d \left[1 - \exp\left(\frac{-\Delta G_v}{RT} \cdot \frac{M}{\rho}\right)\right], \tag{2}$$

where N_a is the number of nucleation sites, γ an atomic vibration frequency, the ΔG are free energies, Z a Zeldovitch parameter, O_n the number of atoms at critical nucleus surface, M a molar mass, d an inter-atomic distance, ρ a volumic mass, and ΔG_v the difference in Gibbs free energy between the amorphous and the crystalline phases. The calculation of each parameter is detailed in Ref. 31.

For the simulations shown in Fig. 7, the fitting of the GST-LTP experimental data (corresponding to the 50 ns pulse width, Fig. 3(a)) was defined as the reference nucleation rate (NR = 1) and the reference growth velocity (GR = 1). LTP simulations (Fig. 7(a)) with artificially increased (GR = 10) and artificially reduced (GR = 0.1) growth velocities with respect to GST (GR = 1) were performed, keeping the nucleation rate constant. Similarly, LTP simulations with artificially increased (NR = 2) and artificially reduced (NR = 0.1) nucleation rates were also performed (Fig. 7(b)). The artificial boost or decrease in NR and GR was performed by directly multiplying Eqs. (1) and (2)

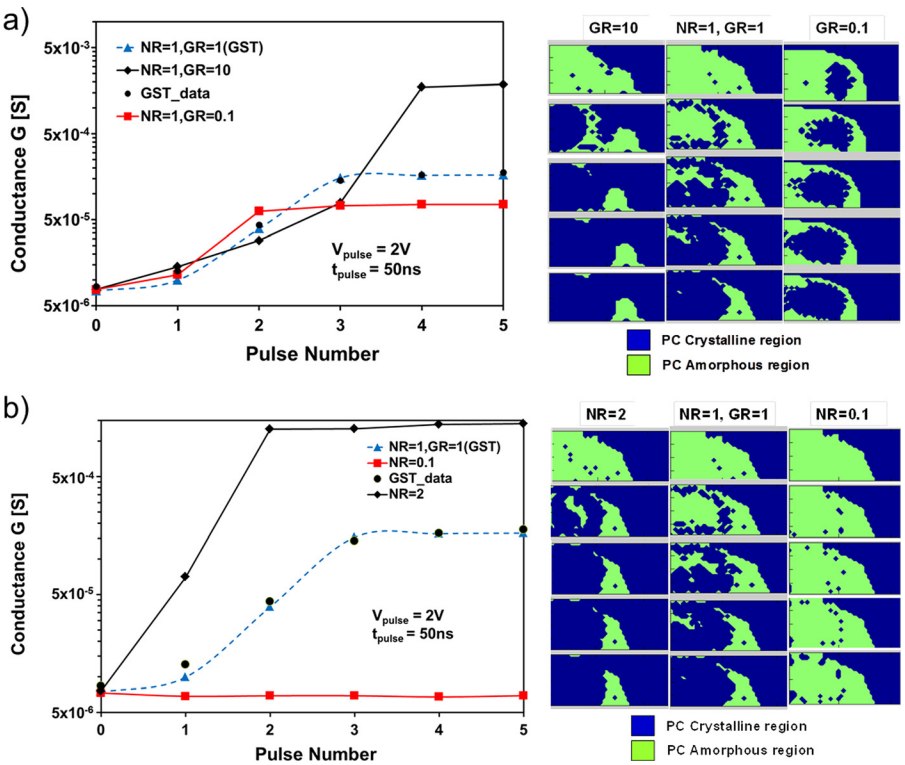

FIG. 7. (a) Simulated LTP curves while fixing the nucleation rate (NR) and varying the growth rate GR compared to GST (taken as reference: GR = 1, NR = 1). Corresponding simulations of GST layer morphology are shown (0th pulse: reset; 1st-5th: potentiating). (b) Simulated LTP curves while fixing the growth rate (GR = 1) and varying the nucleation rate (NR) compared to GST (taken as reference material: NR = 1, GR = 1). Corresponding simulation of GST layer morphology are also shown.

with a constant value. Three major observations were made. First, the maximum value of conductance was reached in fewer pulses if either the growth or the nucleation rate were enhanced. Second, the shape of the bulk amorphous region created after application of the initial reset pulse had a strong dependence upon the values of the growth and the nucleation rates. It is not straightforward to decouple the effect of the nucleation and of the growth parameters as the shape of the amorphous region or morphology changes after the application of each potentiating pulse.[32] A high growth rate (GR = 10) leads to a strong crystal growth from the amorphous-crystalline interface during the falling edge of the reset pulse, thus distorting the mushroom-like shape of the amorphous region. A low growth rate (GR = 0.1) leads to a more uniform mushroom shape of the amorphous region. Finally, after the application of the first potentiating pulse, conductance was more sensitive to changes in the nucleation rate compared to growth. Fig. 7 also shows the strong impact of nucleation rate and growth velocity on the morphological evolution of the crystalline phase inside phase change layer.

IV. MODELING

A. Behavioral model for system level simulations

In order to model the millions of PCM synapses in large scale neural networks, and thus to evaluate the potential of PCM synaptic technology, a computationally efficient model of its behavior is particularly desirable. For this purpose, we introduced the following phenomenological equation to model the LTP characteristics of the GST and GeTe devices during a LTP pulse:

$$\frac{dG}{dt} = \alpha \exp\left(-\beta \frac{G - G_{\min}}{G_{\max} - G_{\min}}\right) \tag{3a}$$

where G is the device conductance, α and β are fitting parameters. G_{min} and G_{max} are the minimum and maximum values of device conductance, respectively. This equation was originally introduced to model memristive devices.[40] To model the conductance change ΔG after the application of a short LTP pulse of duration Δt, Eq. (3a) may be integrated as

$$\Delta G = \alpha \Delta t \exp\left(-\beta \frac{G - G_{\min}}{G_{\max} - G_{\min}}\right). \tag{3b}$$

These equations gave a very satisfactory fit of our measurements for both GST and GeTe devices, as shown in Figs. 3(a) and 3(b). This is valuable because the shape of the potentiation curve should have a serious impact on the system performance as suggested by the works in computational neuroscience.[41] For GST and GeTe, we used unique set of parameters α, β, and G_{min}. G_{max}, for the different pulse widths. Table I lists the values of the parameters used for the fitting two pulse widths for GST and GeTe shown in Figs. 3(a) and 3(b), respectively.

B. Circuit-compatible model

To design hybrid neural circuits consisting of CMOS neurons and PCM synapses, a circuit-compatible model for the PCM is required. We thus developed a circuit compatible PCM model, specifically tailored to capture the progressive character of the LTP experiments shown in this paper. This simple circuit-compatible model, inspired by Ref. 42, was developed using the VHDL-AMS language and includes both LTP and LTD. The simulations were performed with the Cadence AMS simulator. Fig. 3(c) shows the simulated LTP curves for six different pulse widths for GST. Table III lists all the constants and fitting parameters used in the circuit-compatible model. The model consists of three parts: electrical, thermal, and phase-change. For the electrical part, an Ohmic relationship between current and voltage is assumed:

$$v = R_{gst}.i. \tag{4}$$

We preferred not to include the threshold switching effect in this model, as its primary purpose is to emulate LTP behavior during learning in large-scale neural networks simulations, where simulation efficiency is essential. R_{gst} is the low field resistance of the device, which consists of the sum of the GST layer resistance and the bottom and top electrodes resistance R_s. The resistance of the phase change layer is a function of the amorphous volume fraction C_a

$$R_{gst} = R_s + R_{0c}^{1-C_a}.R_{0a}{}^{C_a} \tag{5}$$

R_{0c} and R_{0a} correspond to the resistances of the fully crystallized and fully amorphized states, respectively. We used a logarithmic interpolation, which is intermediate between the series and parallel cases,[42] as this led to the best fitting for our GST devices.

In order to evaluate the impact of resistance drift[43] on the stability of the learning, we included a behavioral modeling of this phenomenon, which can be optionally enabled or disabled in the simulations. To do so, R_{0a} is replaced with R_a

$$R_a = R_{0a}.\left(\frac{t}{t_0}\right)^{C_a.dr}, \tag{6}$$

where t_0 is the time, at which the latest phase change occurred (i.e., when T_b last crossed T_g).

In the thermal part of the model, the electrical power P_t and the temperature of the phase-change material T_b are connected by the following equation, with T_0 the ambient temperature

$$T_b = T_0 + P_t.R_{tgst}. \tag{7}$$

TABLE I. Fitting parameters of the behavioral model for 300 ns GST LTP curve and 100 ns GeTe LTP curve shown in Figs. 3(a) and 3(b), respectively.

Parameters	GST (300 ns)	GeTe (100 ns)
G_{min} (μS)	8.50	8.33
G_{max} (mS)	2.3	2.9
α (S/s)	1100	3300
β	−3.8	−0.55

TABLE II. Parameters used for the GST compact model simulations shown in Fig. 3(c).

Parameter	Value	Description
Electrical model		
R_s	100 Ω	Serial resistance (top and bottom electrodes)
R_{0a}	159 kΩ	Resistance of the fully amorphized state
R_{0c}	135 Ω	Resistance of the fully crystallized state
Thermal model		
R_{ta0}	15.9×10^3 W/K	Thermal resistance of the fully crystallized state
R_{tc0}	5.07×10^3 W/K	Thermal resistance of the fully amorphized state
T_0	300 K	Ambient temperature
Phase-change model		
E_a	0.335	Fitting parameter for crystallization rate at high temperature
E_b	5.77×10^3	Fitting parameter for crystallization rate at low temperature
T_g	380 K	Lowest temperature at which crystallization can occur
T_m	698 K	Melting temperature of the phase change material
τ_a	8.17×10^{-13} s^{-1}	Amorphization rate (fitting)
τ_c	1.10×10^{-6} s^{-1}	Crystallization rate (fitting)
dr	0.03	Drift coefficient

The thermal resistance of the phase change layer R_{tgst} is described by the following equation:

$$R_{tgst} = (1 - C_a).R_{tc0} + C_a.R_{ta0}, \tag{8}$$

where R_{tco} and R_{tao} are the thermal resistances for the completely crystallized and completely amorphized states. Due to the threshold switching effect, the electrical power during phase change is essentially independent on the amorphous ratio. The electrical power is therefore calculated using the fully crystallized phase-change resistance, for which no threshold switching occurs, instead of the low field resistance

$$P_t = \frac{v^2}{R_s + R_{0c}}. \tag{9}$$

The phase-change part of the model uses behavioral equations. Amorphization occurs when T_b is higher than T_m, the melting temperature. The amorphization rate is assumed to increase linearly with the temperature, and is zero when T_b is equal to T_m, thus ensuring continuity with the crystallization rate at this temperature. This leads to the equation

$$\frac{dC_a}{dt} = \frac{1}{\tau_a}.\frac{T_b - T_m}{T_m}, \quad \text{when } T_b > T_m. \tag{10}$$

The equation modeling the crystallization rate does not attempt to model nucleation-driven and growth-driven rate separately. The expression we used is, however, reminiscent to growth rate modeling. It includes a term ${C_a}^2$, because crystallization rate typically depends on the amorphous-crystal interface surface for growth-driven process (and volume for nucleation-driven process).

$$\frac{dC_a}{dt} = -\frac{{C_a}^2}{\tau_c}.\left(1 - \exp\left(E_a.\frac{T_b - T_m}{T_b}\right)\right).\exp\left(-\frac{E_b}{T_b}\right), \quad \text{when } T_b < T_m \text{ and } T_b > T_g. \tag{11}$$

This model, with parameters listed in Table II, gives good fitting results for the LTP curves (Fig. 3(c)), although the fit is not as excellent as the behavioral model (Fig. 3(a)). Simulation of the conductance evolution as a function of the applied voltage with the same parameters is shown in Fig. 3(d). Although the fitting is less accurate for shorter pulses, all the curves in Fig. 3 were fitted with a single set of parameters. The model captures the correct behavior of the PCM for a relatively wide range of measurements with a small number of semi-physical parameters. It is therefore adapted for fast exploration and easier circuit design where PCM devices are employed to emulate millions of synapses.

V. THE "2-PCM SYNAPSE" CIRCUIT

Different feasible programming methodologies to use PCM or nanoscale resistive-memory devices as synapses in neural networks have been proposed in the literature.[48,49] Here, we propose a new methodology based on the use of two PCM devices for emulating a single synapse (Fig. 8). In this scheme, one of the PCM device implements synaptic potentiation (the LTP-device), while the other (the LTD-device) implements synaptic depression. Both devices are initialized to a high resistive amorphous state before the network undergoes learning. The contribution of the currents through the LTP device to the post-synaptic neuron is positive, while that of the LTD device is negative. Thus, when the LTD device is potentiated, the overall impact of the synapse is a synaptic depression, because the current flowing through it is subtracted in the post neuron. Since the depression is also attained by crystallization, it can be performed gradually and by using identical pulses. The detailed programming methodology for implementing a simplified form of biological spike-time-dependent-plasticity (STDP)[45] learning rule using the *2-PCM Synapse* is described in detail in Ref. 22. The pre- and the post-neuron pulses are defined in such a way that when the pre-neuron spikes before the post

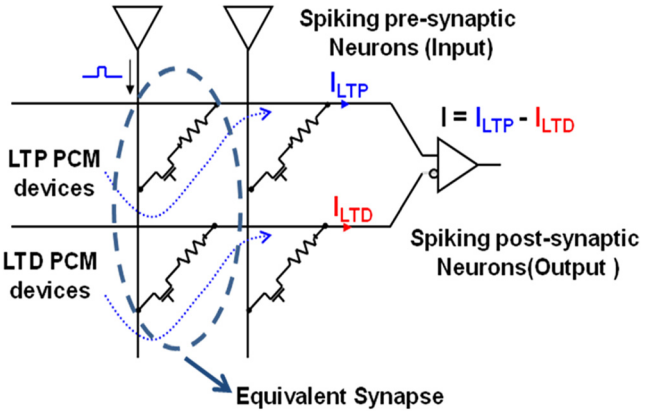

FIG. 8. Circuit schematic for the *2-PCM Synapse*. The input of the current from the LTD devices is inverted in the post-synaptic neuron.

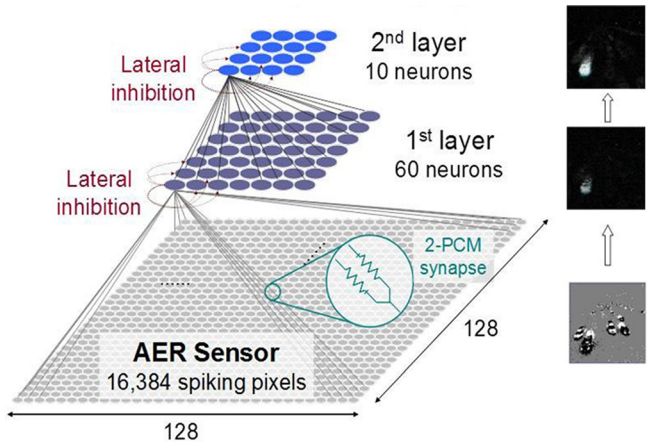

FIG. 9. Simulated two-layer fully connected feed, forward SNN with 70 fully connected neurons and about 2 million synapses.[22]

neuron, the interaction of the two pulses would only potentiate or partially crystallize the LTP device, without affecting the state of the LTD device. In the case when the post-neuron spikes before the pre-neuron, the two pulses interact and potentiate only the LTD device, without affecting the state of the LTP device. The conductance of the PCM devices increases during the learning process and eventually saturates. Thus, we developed a refresh mechanism to enable continuous learning in the system. The refresh mechanism reduces the conductance of the LTP and LTD devices while keeping the weight of the equivalent synapse unchanged. When one of the two devices reaches its maximum conductance the refresh mechanism is initiated. First, both the devices are programmed to reset. Then, a series of potentiating or partially crystallizing pulses are applied to the PCM device which initially had a higher conductance. The potentiating pulses are applied until the equivalent synaptic weight is re-programmed within a reasonably accurate range. Due to the inherent variability robustness of neural architectures, it is not necessary that the synapse be re-programmed to the exact same value of the synaptic weight before the initiation of the refresh operation. As one of the device stays at minimum conductance, this mechanism enables continued evolution of the weights. The detailed circuit implementation of the refresh methodology is described in Ref. 28. Although the *2-PCM Synapse* approach occupies more area compared to a synapse based on a single PCM device, it has several important advantages. It is particularly low power, because the majority of the synaptic events are achieved by crystallization (and not by amorphization), which is a less energy consuming phase-change phenomenon. Another inherent advantage of this approach is the decrease of the impact of resistance drift on the stored synaptic information. Since we potentiate and depress the synapses by crystallization, the majority of the synaptic information is stored or programmed in the low resistance states of the PCM devices. Crystalline or low resistance states are more stable and immune to the resistance drift phenomena compared to high resistance PCM states.[44] If the synaptic depression events were to be implemented by amorphization, all the information stored in the depressed synapses would be in the high-resistance regime and thus more susceptible to resistance-drift.

VI. NEURAL NETWORK AND LEARNING TEST CASE

To validate the approach of the *2-PCM Synapse* described in Sec. V, we simulate a two-layer fully connected feed-forward spiking neural network (Fig. 9) deeply detailed in Ref. 22. The neurons are based on a a leaky-integrate-fire (LIF) model.[46] The GST and GeTe PCM synapses were modeled using Eqs. (3a) and (3b). Overall, there are 70 neurons (60 in the first layer and 10 in the second layer) in the network. Address-event representation (AER) data (Fig.10) recorded with128×128 pixel silicon retina[47] are used as the input for the neural network. The AER data are a special format recording of cars passing on a 6-lane freeway. Each of the 128×128 pixels in Fig. 9 is connected through two synapses to every neuron in the first layer. Likewise, each neuron in the first layer is connected to every neuron in the second layer with a single synapse, leading to a total of 1 966 680 synapses and thus 3 933 360 PCM devices (2 PCM/synapse). The goal of the neural network is to detect cars passing in different lanes on a freeway in an unsupervised way. Fig. 10 and Table III show the overall learning results for the AER dataset. The average successful detection rate for the neural network was found to be greater than 90%. To demonstrate the variability robustness of our neuromorphic system, we implement 20% dispersion on standard deviation

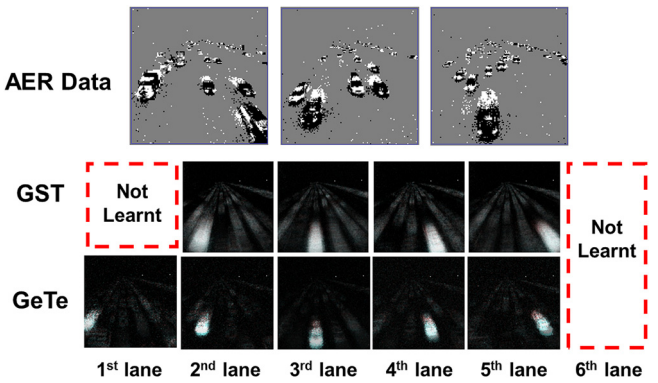

FIG. 10. Grey squares show video recorded data of cars passing on a freeway in AER format. Black squares show the sensitivity map of the neurons in the 1st layer of the neural network for both GST and GeTe. Each neuron becomes sensitive to a specific orientation of cars in a specific lane.

TABLE III. Average car detection rate for the 6 lanes for GST and GeTe PCM synapses.

Car detection %		
	PCM	
Lane	GST	GeTe
1st	N.A.	88
2nd	100	91
3rd	89	96
4th	89	97
5th	96	100
6th	N.A.	N.A.

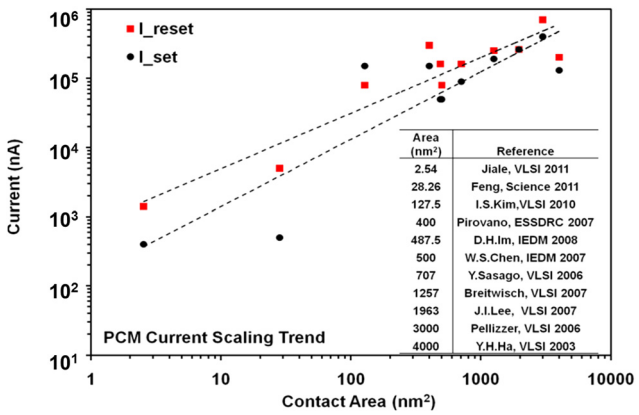

FIG. 11. Scaling trend of RESET and SET current for different PCM technologies (values extracted from the literature).

of all the parameters in Eqs. (3a) and (3b). Output neurons in the 2nd layer are able to detect cars in 4 traffic lanes out of 6 for systems based on GST-PCM synapses, and 5 out of 6 lanes for systems based on GeTe- PCM synapses, respectively. The frequency of potentiating pulses per device is about 25 times higher than the frequency of reset pulses for GST-PCM based system, and about 10 times higher for GeTe-PCM based system. Indeed, this result suggests that the efficiency of the system can be further increased by choosing the right phase-change material with the optimum number of intermediate resistance states in the LTP curve. A rough estimate of the power consumption to program the PCM synapses in the two layer neural network for the car-detection experiment is provided in Ref. 22. The total learning duration for the cars experiment is about 10 min and the total power dissipation in the learning mode for programming the PCM synapses is 112 μW. Based on results from the literature, Fig. 11 shows a strong trend of set and reset current reduction of PCM devices with scaling of the active contact area. Using the current values for the state of the art devices shown in Fig. 11, the power dissipated to program the PCM synapses in the test case can be decreased to as low as 100 nW.

VII. CONCLUSIONS

In this paper, we demonstrated that PCM devices could be used to emulate synaptic behavior such as synaptic potentiation (LTP) and synaptic depression (LTD). The reason for the abrupt or binary LTD behavior with identical depressing pulses was explained. PCM devices with two different chalcogenides materials—GST and GeTe—were characterized. The impact of nucleation rate and growth velocity on LTP experiments was studied. Growth-dominated GeTe showed much faster saturation in the synaptic potentiation experiment compared to nucleation-dominated GST. GST devices showed a higher number of intermediate resistance states for the LTP behavior compared to GeTe. Physical simulations evidenced a strong dependence of the morphological evolution of the phase change layer on the growth velocity and the nucleation rate for the LTP. Enhancing the nucleation rate or growth velocity leads to a faster conductance saturation and thus to a smaller number of intermediate resistance states in the LTP simulations. The shape of the amorphous mushroom formed after the application of the initial reset pulse had a dependence on the growth velocity. Conductance variation in the LTP simulations was found to be more sensitive to nucleation rate compared to growth velocity, after the application of the first potentiating pulse.

A versatile behavioral model and a multi-level circuit-compatible model useful for large scale neural network simulations were developed. We finally provided a new architecture *2-PCM Synapse* to overcome the limitations imposed by emulation of LTD through progressive amorphization. Our approach which mainly relies on the use of crystallization is more energy efficient, uses simple identical pulses and has a higher tolerance to loss of synaptic information through resistance drift.

Using the low power *2-PCM Synapses* and a spiking neural network, we demonstrated complex pattern extraction from real world visual data of cars passing on a freeway. The simulated neural network was able to detect moving cars with a very high average detection rate (>90%) and extremely low synaptic programming power (112 μW).

[1]J. Seo, B. Brezzo, Y. Liu, B. D. Parker, S. K. Esser, R. K. Montoye, B. Rajendran, J. A. Tierno, L. Chang, D. S. Modha, and D. J. Friedman, "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," in IEEE Custom Integrated Circuits Conference (CICC), 2011.
[2]P. Merolla, J. Arthur, F. Akopyan, N. Imam, R. Manohar, and D. S. Modha, "A digital neurosynaptic core using embedded crossbar memory with 45pJ per spike in 45nm," in IEEE Custom Integrated Circuits Conference (CICC), 2011.
[3]J. Schemmel, A. Grubl, K. Meier, and E. Mueller, "Implementing synaptic plasticity in a VLSI spiking neural network model," in The 2006 International Joint Conference on Neural Networks (IJCNN), (IEEE, 2006), pp. 1–6.
[4]G. Indiveri, E. Chicca, and R. Douglas, "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," IEEE Trans. Neural Networks **17**, 211–221 (2006).
[5]J. V. Arthur and K. Boahen, "Learning in silicon: Timing is everything," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2006), Vol. 18, pp. 281–1185.
[6]G. Snider, R. Amerson, D. Carter, H. Abdalla, M. S. Qureshi, J. Léveillé, M. Versace, H. Ames, S. Patrick, B. Chandler, A. Gorchetchnikov, and E. Mingolla, "From synapses to circuitry: Using memristive memory to explore the electronic brain," Computer **44**, 21–28 (2011).
[7]R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, L. Camunas-Mesa, R. Berner, M. Rivas-Perez, T. Delbruck, S.-C. Liu, R. Douglas, P. Hafliger, G. Jimenez-Moreno, A. C. Ballcels, T. Serrano-Gotarredona, and A. J. Acosta-Jimenez, "CAVIAR: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory–processing–learning–actuating system for high-speed visual

object recognition and tracking," IEEE Trans. Neural Networks **20**(9), 1417–1438 (2009).
[8]M. M. Khan, D. R. Lester, L. A. Plana, A. Rast, X. Jin, E. Painkras, and S. B. Furber, "SpiNNaker: Mapping neural networks onto a massively-parallel chip multiprocessor," in *The 2008 International Joint Conference on Neural Networks (IJCNN)*, (IEEE, 2008), pp. 2849–2856.
[9]D. Purves, *Neuroscience*, 3rd ed. (Sinauer Associates, Inc, Massachusetts, 2004), pp. 7–9.
[10]S. Mitra, S. Fusi, and G. Indiveri, "A VLSI spike-driven dynamic synapse which learns only when necessary," in Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), 2006.
[11]K. K. Likharev, and D. B. Strukov, "CMOL: Devices, Circuits, and Architectures," in *Introducing Molecular Electronics,* edited by G. Cuniberti (Springer, Berlin, 2005), Chap. 16, p. 447.
[12]F. Alibart, S. Pleutin, D. Guérin, C. Novembre, S. Lenfant, K. Lmimouni, C. Gamrat, and D. Vuillaume, "An organic nanoparticle transistor behaving as a biological spiking synapse," Adv. Funct. Mater. **20**, 330–337 (2010).
[13]S. Folling, O. Turel, and K. K. Likharev, "Single-electron latching switches as nanoscale synapses," in *The 2001 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2001), pp. 216–221.
[14]A. K. Friesz, A. C. Parker, C. Zhou, K. Ryu, J. M. Sanders, H.-S. P. Wong, and J. Deng, "A biomimetic carbon nanotube synapse circuit," in Biomedical Engineering Society (BMES), Annual Fall Meeting, 2007.
[15]Q. Lai, L. Zhang, Z. Li, W. F. Stickle, R. S. Williams, and Y. Chen, "Ionic/electronic hybrid materials integrated in a synaptic transistor with signal processing and learning functions," Adv. Mater. **22**, 2448–2453 (2010).
[16]T. Hasegawa, K. Terabe, T. Tsuruoka, and M. Aono, "Atomic switch: Atom/ion movement controlled devices for beyond von-neumann computers," Adv. Mater. **24**, 252–267 (2012).
[17]X. L. Zhang, "A mathematical model of a neuron with synapses based on physiology," in Nature Precedings, 2008.
[18]A. R. Sargsyan, A. A. Melkonyan, C. Papatheodoropoulos, H. H. Mkrtchian, and G. K. Kostopoulos, "A model synapse that incorporates the properties of short-term and long- term synaptic plasticity," Neural Networks **16**(8), 1161–1177 (2003).
[19]S. Song, "Hebbian learning and spike-timing-dependent plasticity," in *Computational Neuroscience*, edited by J. Feng (Chapman and Hall/CRC, 2004), Chap. 11.
[20]B. DeSalvo, *Silicon Non-Volatile Memories* (Wiley-ISTE, 2009), pp. 188–216.
[21]M. Suri, V. Sousa, L. Perniola, D. Vuillaume, and B. DeSalvo, "Phase change memory for synaptic plasticity application in neuromorphic systems," in *The International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2011), pp. 619–624.
[22]M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, and B. DeSalvo, "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," in *IEEE International Electron Devices Meeting (IEDM)*, 2011.
[23]D. Kuzum, R. G. D. Jeyasingh, B. Lee, and H.-S. P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," Nano Lett. **12**(5), 2179–2186 (2012).
[24]S. H. Jo, T. Chang, I. Ebong, B. B. Bhavidya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," Nano Lett. **10**, 1297–1301 (2010).
[25]S. Yu and H. S. P. Wong, "Modeling the switching dynamics of programmable-metallization-cell (PMC) memory and its application as synapse device for a neuromorphic computation system," in *IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2010), pp. 520–523.
[26]S. Yu, Y. Wu, R. G. D. Jeyasingh, D. Kuzum, and H.-S. P. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation," in IEEE Trans. Electron Devices **58**(8), 2729–2737 (2011).
[27]H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase change memory," in Proc. IEEE **98**(12), 2201–2227 (2010).
[28]O. Bichler, M. Suri, D. Querlioz, D. Vuillaume, B. DeSalvo, and C. Gamrat, "Visual pattern extraction using energy-efficient '2-PCM synapse' neuromorphic architecture," IEEE Trans. Electron Devices **59**(8), 2206–2214 (2012).
[29]A. Fantini, L. Perniola, M. Armand, J. F. Nodin, V. Sousa, A. Persico, J. Cluzel, C. Jahan, S. Maitrejean, S. Lhostis, A. Roule, C. Dressler, G. Reimbold, B. De Salvo, P. Mazoyer, D. Bensahel, and F. Boulanger, "Comparative assessment of GST and GeTe materials for application to embedded phase-change memory devices," in 2009 IEEE International Memory Workshop (IMW), 2009.
[30]A. Pirovano, A. L. Lacaita, A. Benvenuti, F. Pellizzer, and R. Bez, "Electronic switching in phase change memories," IEEE Trans. Electron Devices **51**(3), 452–459 (2004).
[31]A. Gliere, O. Cueto, and J. Hazart, "Coupling of Level set method with an electrothermal solver to simulate GST based PCM cells," in *International conference on Simulation of Semiconductor Process and Devices (SISPAD)* (IEEE, 2011), pp. 63–66.
[32]G. Bruns, P. Merkelbach, C. Schlockermann, M. Salinga, M. Wuttig, T. D. Happ, J. B. Philipp, and M. Kund, "Nanosecond switching in GeTe phase change memory cells," Appl. Phys. Lett. **95**, 043108 (2009).
[33]C. Sandhya, A. Bastard, L. Perniola, J. C. Bastien, A. Toffoli, E. Henaff, A. Roule, A. Persico, B. Hyot, V. Sousa, B. De Salvo, and G. Reimbold, "Analysis of the effect of boron doping on GeTe phase change memories," in IEEE International Reliability Physics Symposium (IRPS), 2012.
[34]R. M. Shelbya and S. Raoux, "Crystallization dynamics of nitrogen-doped $Ge_2Sb_2Te_5$," J. Appl. Phys. **105**, 104902 (2009).
[35]M. Boniard, D. Ielmini, A. L. Lacaita, A. Redaelli, A. Pirovano, I. Tortorelli, M. Allegra, M. Magistretti, C. Bresolin, D. Erbetta, A. Modelli, E. Varesi, F. Pellizzer, and R. Bez, "Impact of material composition on the write performance of phase-change memory devices," in IEEE International Memory Workshop (IMW), 2010.
[36]L. van Pieterson, M. H. R. Lankhorst, M. van Schijndel, A. E. T. Kuiper, and J. H. J. Roosen, "Phase-change recording materials with a growth-dominated crystallization mechanism: A materials overview," J. Appl. Phys. **97**, 083520 (2005).
[37]G. Navarro, N. Pashkov, M. Suri, V. Sousa, L. Perniola, S. Maitrejean, A. Persico, A. Roule, A. Toffoli, B. DeSalvo, P. Zuliani, and R. Annunziata, "Electrical performances of tellurium-rich GexTe1-x Phase change memories," in IEEE International Memory Workshop (IMW), 2011.
[38]S. Braga, N. Pashkov, L. Perniola, A. Fantini, A. Cabrini, G. Torelli, V. Sousa, B. De Salvo, and G. Reimbold, "Effects of alloy composition on multilevel operation in self-heating phase change memories," in IEEE International Memory Workshop (IMW), 2011.
[39]A. Fantini, V. Sousa, L. Perniola, E. Gourvest, J. C. Bastien, S. Maitrejean, S. Braga, N. Pashkov, A. Bastard, B. Hyot, A. Roule, A. Persico, H. Feldis, C. Jahan, J. F. Nodin, D. Blachier, A. Toffoli, G. Reimbold, F. Fillot, F. Pierre, R. Annunziata, D. Benshael, P. Mazoyer, C. Vallée, T. Billon, J. Hazart, B. De Salvo, F. Boulanger *et al.*, "N-doped GeTe as performance booster for embedded phase-change memories," in IEEE International Electron Devices Meeting (IEDM), 2010.
[40]D. Querlioz, P. Dollfus, O. Bichler, and C. Gamrat, "Learning with memristive devices: How should we model their behavior?," in *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)* (IEEE, 2011), pp. 150–156.
[41]J. Rubin, D. D. Lee, and H. Sompolinsky, "Equilibrium properties of temporally asymmetric Hebbian plasticity," Phys. Rev. Lett. **86**, 364–367 (2001).
[42]K. Sonoda, A. Sakai, M. Moniwa, K. Ishikawa, O. Tsuchiya, and Y. Inoue, "A compact model of phase change memory based on rate equations of crystallization and amorphization," IEEE Trans. Electron Devices **55**(7), 1672–1681 (2008).
[43]I. V. Karpov, M. Mitra, D. Kau, G. Spadini, Y. A. Kryukov, and V. G. Karpov, "Fundamental drift of parameters in chalcogenide based phase change memory," J. Appl. Phys. **102**(12), 124503 (2007).
[44]D. Ielmini, S. Lavizzari, D. Sharma, and A. L. Lacaita, "Physical interpretation, modeling and impact on phase change memory (PCM) reliability of resistance drift due to chalcogenide structural relaxation," in *IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2007), pp. 939–942.
[45]H. Markram, W. Gerstner, and P. J. Sjöström, "A history of spike-timing-dependent plasticity," Front. Synaptic Neurosci. **3**, 4 (2011).
[46]G. Indiveri *et al.*, Front. Neurosci. **5**, 73 (2011).
[47]P. Lichtsteiner *et al.*, "A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," IEEE J. Solid-State Circuits, **43**, 566–576 (2008).
[48]D. Kuzum, R. Jeyasingh, and H.-S. P. Wong, "Energy efficient programming of nanoelectronic synaptic devices for large-scale implementation of associative and temporal sequence learning," in *IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2011), pp. 30.3.1–30.3.4.
[49]G. Snider, "Spike-timing-dependent learning in memristive nanodevices," in *IEEE International Symposium on Nanoscale Architectures, NANOARCH* (IEEE, 2008), pp. 85–92.
[50]M. Suri, O. Bichler, Q. Hubert, L. Perniola, V. Sousa, C. Jahan, D. Vuillaume, C. Gamrat, and B. DeSalvo, "Interface engineering of PCM for improved synaptic performance in neuromorphic systems," in IEEE International Memory Workshop (IMW), 2012.