

NANO REVIEW

Open Access



Recent Advances on Neuromorphic Systems Using Phase-Change Materials

Lei Wang^{1,2*}, Shu-Ren Lu^{1,2} and Jing Wen^{1,2}

Abstract

Realization of brain-like computer has always been human's ultimate dream. Today, the possibility of having this dream come true has been significantly boosted due to the advent of several emerging non-volatile memory devices. Within these innovative technologies, phase-change memory device has been commonly regarded as the most promising candidate to imitate the biological brain, owing to its excellent scalability, fast switching speed, and low energy consumption. In this context, a detailed review concerning the physical principles of the neuromorphic circuit using phase-change materials as well as a comprehensive introduction of the currently available phase-change neuromorphic prototypes becomes imperative for scientists to continuously progress the technology of artificial neural networks. In this paper, we first present the biological mechanism of human brain, followed by a brief discussion about physical properties of phase-change materials that recently receive a widespread application on non-volatile memory field. We then survey recent research on different types of neuromorphic circuits using phase-change materials in terms of their respective geometrical architecture and physical schemes to reproduce the biological events of human brain, in particular for spike-time-dependent plasticity. The relevant virtues and limitations of these devices are also evaluated. Finally, the future prospect of the neuromorphic circuit based on phase-change technologies is envisioned.

Keywords: Phase-change materials, Neuromorphic, Neuron, Synapse, STDP, Brain

Review

Background

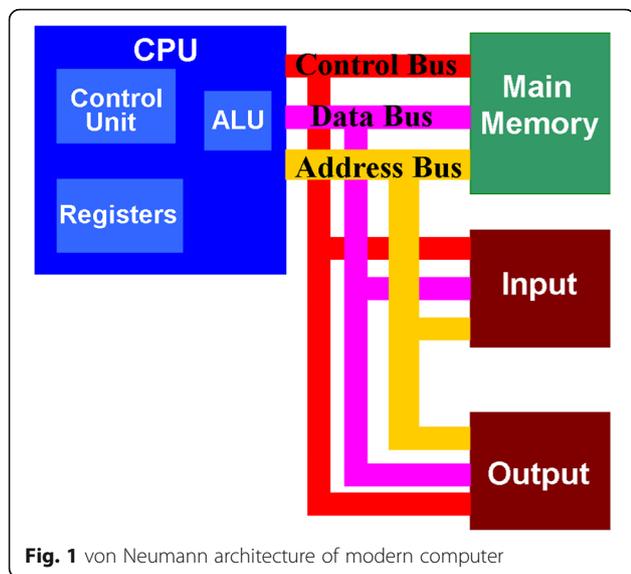
Today, digital computer, commonly considered as a milestone in the history of human life, has a pervasive influence on every citizen's daily activities involving business, education, entertainment, and sports. As a physical device while manipulated by the operational system, the prosperity of the digital computer aggressively lies on the progress of both hardware and software technologies. Recent technological developments on ultra-large-scale integration (ULSI) allow millions or even billions of electronic components to be integrated on a single semiconductor chip, significantly improving the physical performances of the modern computer. Under this circumstance, it is not too naïve to imagine that human being will be govern by computer machines one day that has been frequently described in

scientific fictions, particularly after the recent victory of 'AlphaGo' over the top human Go player [1]. However, thanks to the architectural difference between computer and human brain, it is not possible for digital computer to outperform the biological brain in the near future. It is well known that modern computer usually makes use of the so-called von Neumann architecture that consists of three main components [2], i.e. processor, main memory, and bus, as shown in Fig. 1. The processor, also known as central processing unit (CPU), comprises arithmetic logic unit (ALU), control unit, and register. As implied by their names, ALU is responsible for all the arithmetic and logical operations such as addition, subtraction, AND, and OR functions, while the control unit decodes the instructions and controls all other internal components of the CPU. The register is mainly used to store the data during execution. After all the essential computations, the processed data is sent back from the CPU to the main memory where the data is stored through the data bus, whereas address bus and control bus are employed to determine the address of the data inside

* Correspondence: LeiWang@nchu.edu.cn

¹School of Information Engineering, Nanchang HangKong University, Nanchang 330063, People's Republic of China

²Department of Automatic Control, School of Information Engineering, Nanchang Hangkong University, Nanchang 330069, Jiangxi, People's Republic of China



the memory and the type of the operations (e.g. write and read) between CPU and memory, respectively.

According to above descriptions, an apparent feature of modern computer adopting von Neumann architecture is that CPU where data is processed is separated from the main memory where data is stored by bus. As a consequence, CPU needs to retrieve data from the main memory for any necessary processing, after which data is transferred back to the main memory for storage. The fact that bandwidth between CPU and the main memory (also called data transfer rate) is much lower than the speed that a typical CPU can work severely limits the processing speed of the modern computer, which is known as von Neumann bottleneck. In order to circumvent the von Neumann bottleneck, several advanced technologies such as Cache memory, multi-threading core, and low-latency command channel have been proposed in the past to increase the processing speed of the modern computer. These approaches seem to be viable for the cases with less repeated operations that only cope with relatively small amount of the processing data, while failing to satisfy the demands from data-centric applications that usually require often-repeated transient operations on a vast amount of the digital data such as real-time image recognition and natural language processing [3]. Therefore, the current consensus is that it is inevitable to have an unprecedented revolution from von Neumann architecture to non-von Neumann architecture so as to utterly eliminate the von Neumann bottleneck.

Fortunately, an ideal computing system that adopts non-von Neumann architecture has been existing for millions of years, which is the human brain. The human brain comprises many types of cells within which the core component is called neuron [4], as shown in Fig. 2.

The highly importance of the neuron for human brain stems from its ability to process and transmit information through electrical and chemical signals. According to Fig. 2, a neuron is made up of a cell body (also called soma), dendrites, and an axon. Information usually in the form of an electrical or chemical signal is transferred from an axon of one neuron towards the conjunction of its axon and dendrites of the neighbouring neurons, also known as synapse. The synapse can evaluate the importance of the received information by integrating it with the strength of the synapse (synaptic weight), and subsequently distribute information to even more neurons through their respective axons. Differing from the digital computer, human brain performs the information processing during the transferring period, and there is only a single value by the time that information reaches the neighbouring neurons. This clearly indicates an encouraging finding that human brain allows information storage and processing to occur at the same time in the same place. Due to this attractive feature, human brain whose neural networks consists of $\sim 10^{11}$ neurons and $\sim 10^{15}$ synapses enables an operation frequency of 1–10 Hz on the power budget of 10–100 W [5], corresponding to an energy consumption of 1–10 fJ per synaptic event [5].

Thanks to the exceptional capability of the human brain, it is natural to conceive the possibility of building a super-intelligent computer that reproduces the neural networks of the human brain to completely overcome the von Neumann architecture, leading to the prosperity of artificial intelligence (AI). One possible way to achieve brain-like computer is to simulate the behaviours and connections between biological neurons inside the human brain using conventional computers or even so-called supercomputers, replying on the recent progress of the software algorithms. In spite of its advantageous flexibility and availability [6, 7], the software-based approaches fail to cope with large-scale tasks such as pattern recognition, learning, and intelligent cognition [8, 9], and also causes several orders of magnitude higher energy consumption than the human brain [10]. In this case, vast majority of research efforts has been recently devoted to exploiting a novel hardware architecture that can emulate both the biological structure and the biological function of the human brain, delivering the debut of neuromorphic engineering that can be dated back to 1980s [8]. However, the concept of neuromorphic has not received considerable attentions until the presence of the emerging non-volatile memories (NVM) such as Ferroelectric random access memory (FeRAM) [11, 12], magnetic random access memory (MRAM) [13, 14], phase-change random access memory (PCRAM) [15, 16], and resistive random access memory (ReRAM) [17, 18], as well as the physical realization of the early proposed

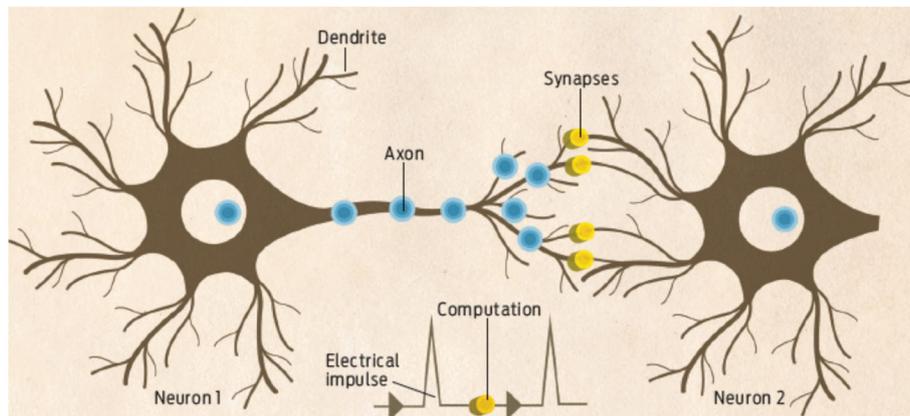


Fig. 2 Neuron structure in the human brain. Reprinted with permission from [4]

'memristor' concept [19–21]. These innovative devices are capable of storing information on an ultra-small region during ultra-short interval with ultra-low energy consumption, closely resembling the attractive features of the human brain. Perhaps most enticingly, some physical properties of these devices that depend on their previous states can be successively tailored by the external stimulus and be stored even after power-off, very analogous to the behaviours of the biological synapse. As a result, several nanoscale electronic devices that include metal oxides [22, 23], solid-electrolytes [24, 25], carbon nanotubes [26, 27], organic electronics [28, 29], spin transfer torque MRAM [30, 31], and phase-change memory (PCM) [32] based on the aforementioned technologies have been proposed to systematically mimic the biological function of the human brain, in particular the learning and memorization capabilities govern by the synapse. It should be noticed that due to the extremely complexity of the human brain, none of these available electronic devices has exhibited the promising potential to prevail or even meet the physical performances of the biological synapse to date. However, the nanoscale electronic synapse based on PCM technology seems to be more promising than its compatriots, as phase-change materials has already gained widespread applications from the conventional optical disc [33] to the recently emerging all-photonics memory [34], and PCRAM is considered as the leading candidate for the so-called universal memory for NVM applications [35, 36]. More importantly, a wealth of theoretical knowledge and practical experiences that concerned the electrical, optical, thermal, and mechanical properties of the phase-change materials has been accumulated along with the development of the PCMs that has been under intensive study for the last two decades. For above reasons, a comprehensive review about physical principles and recent developments of the nanoscale electronic synapse using phase-change

devices becomes indispensable in order to help researchers or even ordinary readers to deeply understand the physical way that the phase-change electronic synapse can imitate the biological synapse, and thus stimulate more research efforts into the establishment of the design criterion and performance requirement for the future artificial synapse device to ultimately compete with the human brain.

Here, we first introduce the fundamental biological behaviours of brain neurons and synapses that are the prerequisites for designing appropriate artificial synapses, followed by a brief overview of the phase-change materials and its applications on NVM applications. Then, we present different types of nanoscale electronic synapse using phase-change devices as well as their respective advantages and disadvantages for neuromorphic applications. The future prospect of the artificial synapse using phase-change materials is also discussed.

Biological Behaviours of the Human Brain

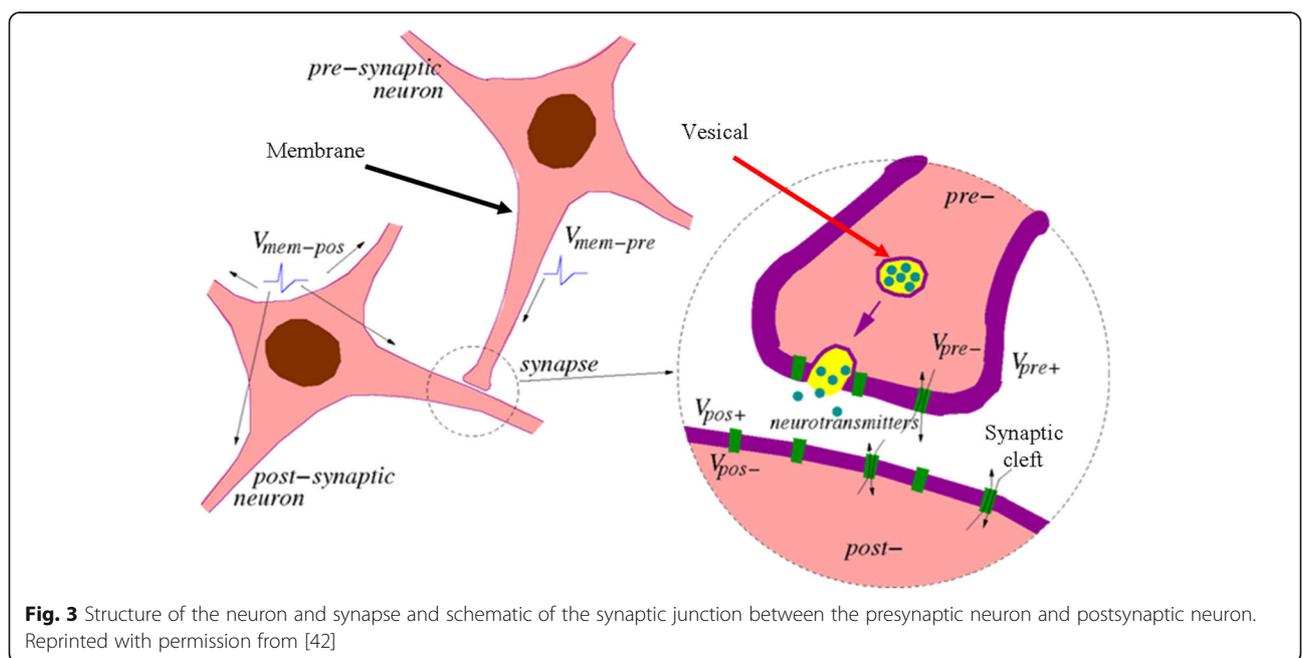
Briefly speaking, the patterns of synaptic transmissions between different neurons inside the brain can be simply classified into electrical synapse and chemical synapse. Electrical synapse transmits signal directly by the drift of ions without involving any chemical reaction. In this case, electrical synapse usually exhibits a more rapid signal transmission that allows signal to flow bidirectionally between neighbouring neurons than chemical synapse and is frequently found on a nerve system that requires quick responses. However, electrical synapse is passive, which means that the signal received in the receiver side is the same as or smaller than the initial signal from the sender side, called 'lack of gain'. Chemical synapses usually divided into excitatory synapse and inhibitory synapse allows signal to be transmitted unidirectionally between neuron sender (called presynaptic neuron) and neuron receiver (called postsynaptic neuron) through the release of neurotransmitter from presynaptic neuron

and absorption of neurotransmitter from postsynaptic neuron [37]. The certain type of chemical synapse is completely determined by the receptors on the postsynaptic membrane. It has been reported that the gamma-aminobutyric acid (GABA) neurotransmitter is an inhibitory receptor [38] that leads to an inhibitory postsynaptic potential (IPSP) [39] and moves the postsynaptic neuron away from depolarization threshold, while glutamate neurotransmitter is an excitatory receptor [40] that exhibits an excitatory postsynaptic potential (EPSP) and drives postsynaptic neuron towards the depolarization threshold that can activate an action potential [41]. Various neurons can be connected to one certain neuron via their axon terminals and release neurotransmitters that stimulate excitatory and inhibitory receptors to yield EPSPs and IPSPs; subsequently, postsynaptic neuron integrates the EPSPs with the IPSPs to determine whether to excite or ‘fire’ an action potential or not. In this case, the resulting action potential is transferred to the synapse cleft through the axon, and merges the neurotransmitter vesicles with the presynaptic membrane at the axon terminal, thereby releasing the neurotransmitters into the synapse cleft. After diffusing through the synapse cleft, these neurotransmitters bind to and activate the receptor in the postsynaptic membrane, thereby changing the synapse weight. Such a chemical scheme is schematically shown in Fig. 3 [42]. As chemical synapse is known to be in charge of learning and memory, the ‘synapse’ mentioned in this review refers to chemical synapse, unless otherwise stated. A more detailed description of the biological behaviours of

the chemical synapse is presented below from chemical point of view.

Note that neurons communicate with each other through so-called action potentials that can be interpreted as the transmitted electrical signals [43]. A liquid membrane separates neuron whose intracellular components are rich in K^+ ions from outside environment whose extracellular component has high concentration of Na^+ and Cl^- ions. This non-uniform ion distribution across the membrane polarizes the membrane with a certain resting potential between -90 and -40 mV [44], when the cell membrane does not experience any external stimulus, i.e. in its rest state. Nevertheless, when subject to an external electric stimuli above some certain threshold, the ions can exchange through voltage-gated ion channels that are activated by the change of the electrical membrane potential and ion pumps between intracellular and extracellular media to lower the chemical potential gradient [45–49]. Such a redistribution of the ions across membrane leads to the depolarization of the neuron, also known as action potential firing responsible for the signal transmission from the neuron soma to the terminal of the neuron axon [43]. However, this depolarized state would soon restore to the previously rest state after Na^+/K^+ ion pumps recover its ion distribution to the rest state, called sodium-potassium adenosine triphosphatase (Na^+/K^+ -ATPase) [50]. Therefore, the state change of the neuron or cell membrane is considered as ‘elastic’ meaning no permanent change occurring when the external stimuli exceeds the threshold value.

Although the change of neuron states exhibit elasticity, the change of synapse weight however shows plasticity, simply meaning that the change can either last for a long



time or a short time. Given its durability, a long-term activity-dependent synaptic plasticity has been widely believed to account for learning and memory [51]. The synaptic activity to induce a long-lasting increase on synapse weight is named as long-term potentiation (LTP) that is in charge of the long-term memory, whereas the synaptic activity to generate a long-lasting decrease on synapse weight is known as long-term depression. LTP is believed to be strongly related to the activities of the presynaptic and postsynaptic neurons. When the presynaptic neuron is activated, it releases various neurotransmitters within which glutamate neurotransmitter acts as a principle role in LTP [52]. The released glutamate is subsequently bound to *N*-methyl-D-aspartate receptor (NMDAR) that opens ion channels for both monovalent and divalent cations (Na^+ , K^+ , and Ca^{2+}), and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic receptor (AMPA) that opens ion channels for monovalent cations (Na^+ and K^+), which are located on the postsynaptic neuron [52]. Accordingly, the change of the electrical membrane potential causes the depolarization of the postsynaptic neuron that opens the voltage-gated ion channel of the NMDAR and consequently results in an increase of Ca^{2+} concentration that can be further enhanced through Ca^{2+} inward flux through L-type voltage-gated calcium channels (VGCCs) [53–57]. However, a weaker depolarization is unable to completely displace the Mg^{2+} ions that block NMDA ion channels and therefore prevents adequate Ca^{2+} ions from entering the postsynaptic neuron, yielding a lower intracellular Ca^{2+} concentration. This is considered as the

induction of LTD. Figure 4 shows the schematics of the described process.

The mechanisms governing synaptic plasticity shown in Fig. 4 still remain mysterious. One hypothesis attributes the induction of synaptic plasticity to the relative rates between the pre- and postsynaptic action potentials also called spikes [58]. Based on rate-dominant speculation, high-frequency stimulations induce LTP, while low frequency stimulations bring about LTD, as illustrated in Fig. 5.

An alternative to rate-dominant synaptic plasticity is time-dependent synaptic plasticity that ascribes the cause of long-term plasticity to temporal correlations between pre- and postsynaptic spikes [59]. The origin of this assumption can be dated back to 1949 when Donald Hebb proposed his famous Hebbian learning rule: ‘When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic changes take place in one or both cells such that A’s efficiency as one of the cells firing B, is increased’ [60]. Hebb suggested that the activations of pre- and postsynaptic neurons do not essentially induce the long-term plasticity, but the relative spiking time that is the time difference between presynaptic spike and postsynaptic spike plays the most important role on synaptic plasticity, usually referred to spike-timing-dependent plasticity (STDP). As illustrated in Fig. 6, the situation that presynaptic spike precedes postsynaptic spike results in LTP, while the synapse undergoes LTD when presynaptic spike lags postsynaptic spike [61].

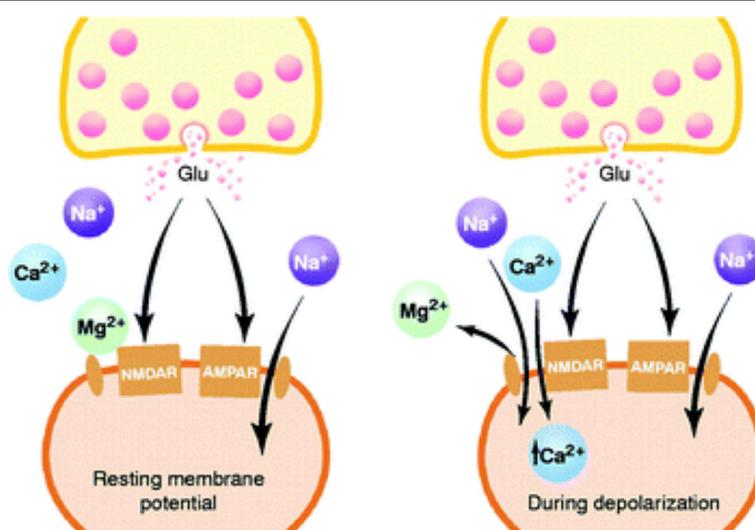


Fig. 4 Potentiation procedures for a chemical synapse. The presynaptic activation releases glutamate neurotransmitters (Glu) that bind to NMDAR and AMPAR. Only the AMPAR opens the ion channels for monovalent cations (e.g. Na^+) when the postsynaptic neuron is at rest state or polarized, while NMDAR opens the voltage-gated channels to allow Ca^{2+} to diffuse once the postsynaptic neuron is depolarized or activated. Reprinted with permission from [53]

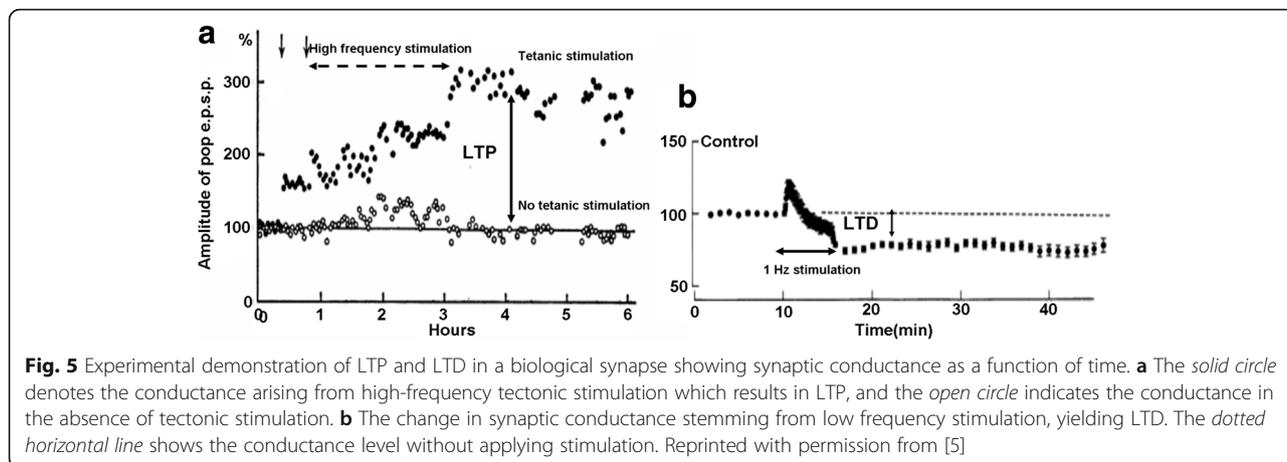


Fig. 5 Experimental demonstration of LTP and LTD in a biological synapse showing synaptic conductance as a function of time. **a** The *solid circle* denotes the conductance arising from high-frequency tectonic stimulation which results in LTP, and the *open circle* indicates the conductance in the absence of tectonic stimulation. **b** The change in synaptic conductance stemming from low frequency stimulation, yielding LTD. The *dotted horizontal line* shows the conductance level without applying stimulation. Reprinted with permission from [5]

The fundamental formula to mathematically depict the STDP learning function is shown below [42]:

$$\Delta w = \xi(\Delta T) = \begin{cases} a^+ e^{-\Delta T/\tau^+} & \text{if } \Delta T > 0 \\ -a^- e^{\Delta T/\tau^-} & \text{if } \Delta T < 0 \end{cases} \quad (1)$$

where Δw is the change on synapse weight, ξ denotes the STDP function, ΔT is the relative time difference between presynaptic spike (t_{pre}) and postsynaptic spike (t_{post}), and a_{\pm} and τ_{\pm} represent the scaling factor and the time constant of the exponential function, respectively. According to Eq. 1, four types of STDP updated functions commonly adopted by computational models of STDP synaptic learning can be readily derived [62], giving rise to Fig. 7. It should be noticed that the cases illustrated in Figs. 6 and 7a correspond to synapses with positive weights ($w > 0$). Therefore, as long as $\Delta T > 0$, the synapse weight w is always strengthened due to the

positive sign of Δw . However, this learning rule does not apply to some so-called inhibitory synapses with negative weights ($w < 0$), as an increase in weight would in turn weaken the strength of the synapse. To resolve this issue, an STDP learning rule with a similar shape to Fig. 7b is required, since in this case decreasing the weight would strengthen the synapse itself when $\Delta T > 0$. Such a relationship between Δw and ΔT , given by Fig. 7b, is usually called anti-STDP learning function. Figure 7c, d is the symmetric form of (a) and (b), respectively.

STDP can be further categorized into additive STDP and multiplicative STDP. The learning function ξ of additive STDP is irrelevant to the actual weight (w), but strongly depending on the time interval (ΔT). Additive STDP requires the weight values to be bounded to an interval because weights will stabilize at one of their boundary values [63, 64]. On the other hand, the learning rule of multiplicative STDP relies on both actual synapse weight (w) and the time interval (ΔT). In multiplicative STDP,

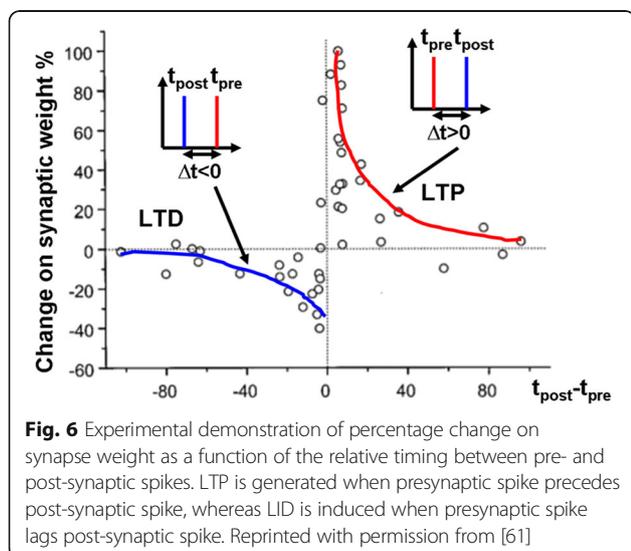


Fig. 6 Experimental demonstration of percentage change on synapse weight as a function of the relative timing between pre- and post-synaptic spikes. LTP is generated when presynaptic spike precedes post-synaptic spike, whereas LTD is induced when presynaptic spike lags post-synaptic spike. Reprinted with permission from [61]

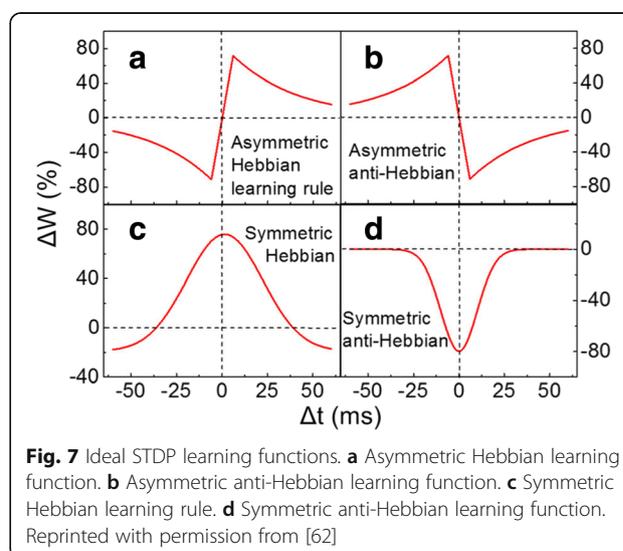


Fig. 7 Ideal STDP learning functions. **a** Asymmetric Hebbian learning function. **b** Asymmetric anti-Hebbian learning function. **c** Symmetric Hebbian learning rule. **d** Symmetric anti-Hebbian learning function. Reprinted with permission from [62]

weight can stabilize to intermediate values inside the boundary definitions, whereby it is not mandatory to enforce boundary conditions for the weight values [63–65]. It should be kept in mind that STDP has recently attained much more interest than spike rate dependent plasticity (SRDP) due to its simplicity and biological plausibility as well as acceptable computational power.

In contrast to long-term plasticity, short-term plasticity (STP) features the modification of synapse weight that can only last from tens of milliseconds to a few minutes and quickly recover to its initial state [66], as described in Fig. 9. Similar to long-term plasticity that can be classified into LTP and LTD, short-term plasticity also includes two types of activity patterns, which are short-term facilitation (STF) and short-term depression (STD). STF that transiently increases the synapse weight is primarily caused by the arriving of two or more action potentials at presynaptic neuron within a very short-time interval, thus boosting the concentration of Ca^{2+} ions. As a consequence, the subsequent presynaptic action potential can excite more neurotransmitter released, which can be further facilitated by a high-frequency train of presynaptic firings called tetanus. This leads to another form of synaptic plasticity referred to post-tetanic potentiation (Fig. 8). Repeated presynaptic activities can also cause continuous depletion of the synaptic vesicles available for release into the synaptic cleft, which results in decrease of synapse weight, i.e. STD.

Phase-Change Materials

Undoubtedly building a neuromorphic circuit that can effectively imitate the behaviours and connections between different neurons is the most prospective approach to overcome von Neumann limits and to ultimately achieve neural brain emulation. However, conventional CMOS-

based neuromorphic circuits caused much more energy consumption than the biological brain and cannot truly mimic its biological behaviours [67–70]. It is evident that a desired neuromorphic device must be able to emulate the spike-dependent synapse plasticity within the similar time interval to brain synapse while at the cost of low energy consumption. As a result, several innovative hardware architectures have been under intensive research aiming to achieve the required synapse performances. Within these technologies, PCM has been recently considered as a leading candidate in comparison with its rivals due to the analogous physical behaviours of phase-change materials to biological synapse [71, 72].

Phase-change materials adopted for neuromorphic applications commonly refers to chalcogenide alloys that mainly include the elements Ge, Sb, and Te. Attractiveness of phase-change materials arises from a phenomenon that it can be switching reversibly between a metastable amorphous phase with high electrical resistivity and low optical reflectivity and a stable crystalline phase with low electrical resistivity and high optical reflectivity when suffering from either electrical or optical stimulus [73]. Such remarkable differences on the electrical/optical properties between amorphous and crystalline phases make phase-change materials very promising for a variety of NVM applications such as optical disc [33, 74], PCRAM [35, 75–77], scanning probe phase-change memory [36, 78–80], and phase-change photonic device [37, 81–83], as shown in Fig. 9. To electrically achieve crystallization, an electric pulse is applied to phase-change materials at amorphous state to bring its temperature above glass transition temperature but below melting temperature by means of the resulting Joule heating, subsequently followed by a slow cooling. This process usually refers to ‘SET’ operation for PCRAM

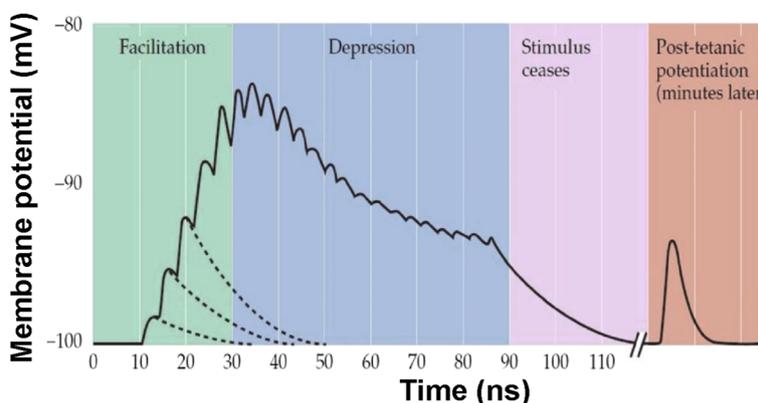
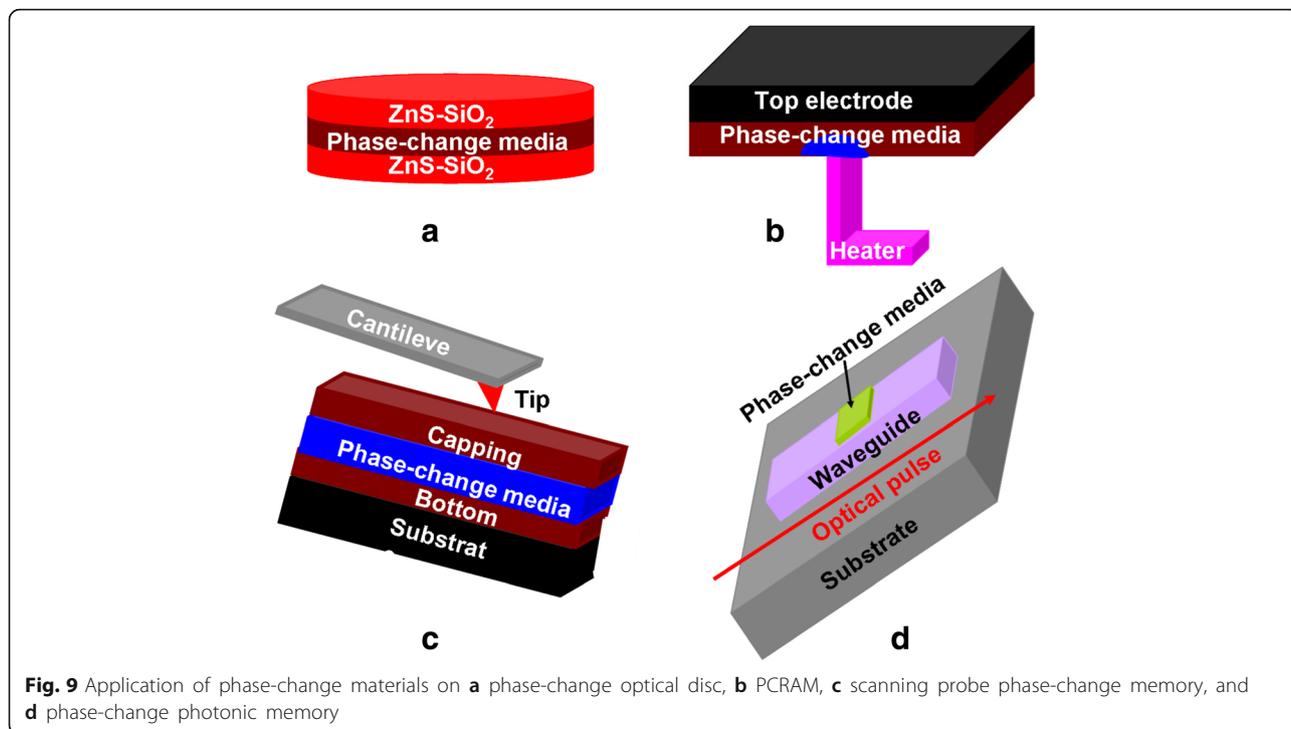
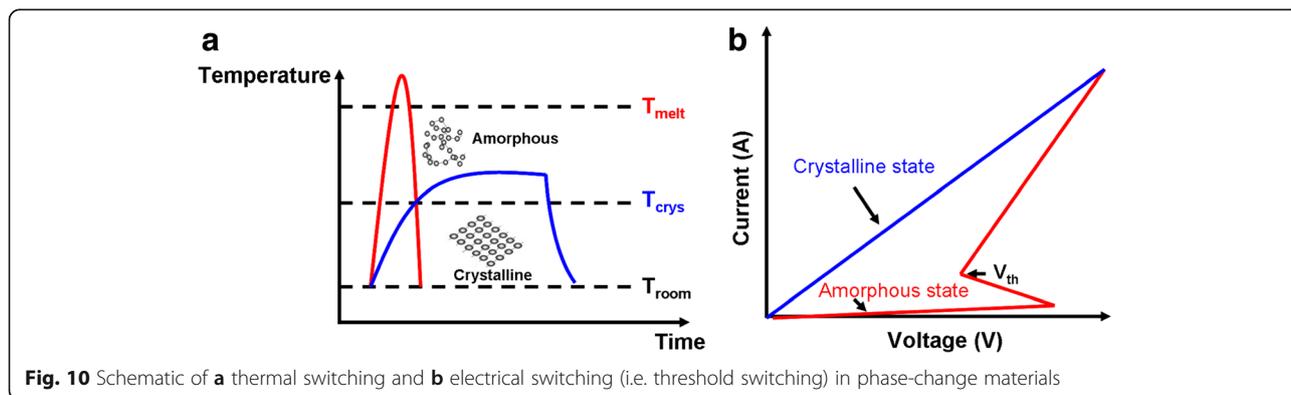


Fig. 8 Experimental demonstration of STP for neuromuscular synapse. Presynaptic motor nerve was stimulated by train of electrical pulses. The end-plate potentials (EPPs) that are the depolarizations of skeletal muscle fibres caused by neurotransmitters binding to the postsynaptic membrane in the neuromuscular junction are facilitated initially, followed by a depression. Epps return to their rest state after removing the stimulation. Reprinted with permission from [5]



operation. Amorphization of phase-change materials that is called ‘RESET’ by convention is induced by electrically heating the media above melting temperature and quickly quenching it to room temperature. In general, SET operation generally lasts a few hundreds of nanoseconds, while time taken by ‘RESET’ is in the range of a few of tens of nanoseconds or even down to picosecond regime. Such a thermal switching is schematically illustrated in Fig. 10a. In addition to thermal switching, phase-change materials also exhibit a unique electrical switching characteristic, namely ‘threshold switching’ [84–86]. Because of threshold switching effect depicted in Fig. 10b, the electrical resistance of the phase-change materials in amorphous state can be dramatically reduced once the bias voltage

exceeds the threshold value, thereby resulting in a high current beneficial for the incoming crystallization. Threshold switching is extremely crucial for phase-change materials, as it allows for the phase transition at relatively low voltage. Otherwise, phase transition can only take place using very high voltage that yields unnecessary energy consumption. These superior physical properties consolidated with considerable practical and theoretical knowledge obtained from the applications of phase-change materials for NVM field during the last two decades successfully endow phase-change materials with several advantageous features such as an excellent scalability of <5 nm [87], a fast switching speed of <1 ns [88], a long cycle endurance of >10¹² [17], and a



low energy consumption of less than a petajoule per bit [89], thus enabling the viability of using nanoscale PCMs to emulate the biological synapse that has similar characteristics to phase-change materials.

PCM-Based Synapse Models

Before the experimental study of the PCM-based artificial synapse, a comprehensively theoretical model is always preferable to assess the potential of PCM devices for large-scale spike neural network (SNN) applications. Without an effectively theoretical model, researchers may have to spend tremendous computational and experimental resources to establish the viable technological path towards the success of building PCM-based neurons and synapses. To achieve this goal, Suri et al. first proposed a behaviour model to simulate the LTP and LTD behaviours of PCMs using the Ge₂Sb₂Te₅ (GST) and GeTe medium [90–92]. In Suri’s model, the conductances of GST and GeTe medium are considered as the synapse weight which can be therefore modified through either ‘SET’ process that corresponds to LTP or ‘RESET’ process indicating LTD. In this case, proper pulse magnitude and width need to be determined carefully so as to generate the desired conductances. The change on synapse weight (conductance in this case) with respect to the applied pulse duration is described as:

$$\frac{dG}{dt} = \alpha \exp\left(-\beta \frac{G - G_{\min}}{G_{\max} - G_{\min}}\right) \tag{2}$$

where G , G_{\min} and G_{\max} denotes the device conductance, minimum device conductance, and maximum device conductance, respectively; α and β are fitting parameters.

According to above equation, the change on synapse weight during a certain time interval Δt is equal to:

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

The reported equations here give rise to a satisfactory matching with the experimentally measured data for GST device, as shown in Fig. 11.

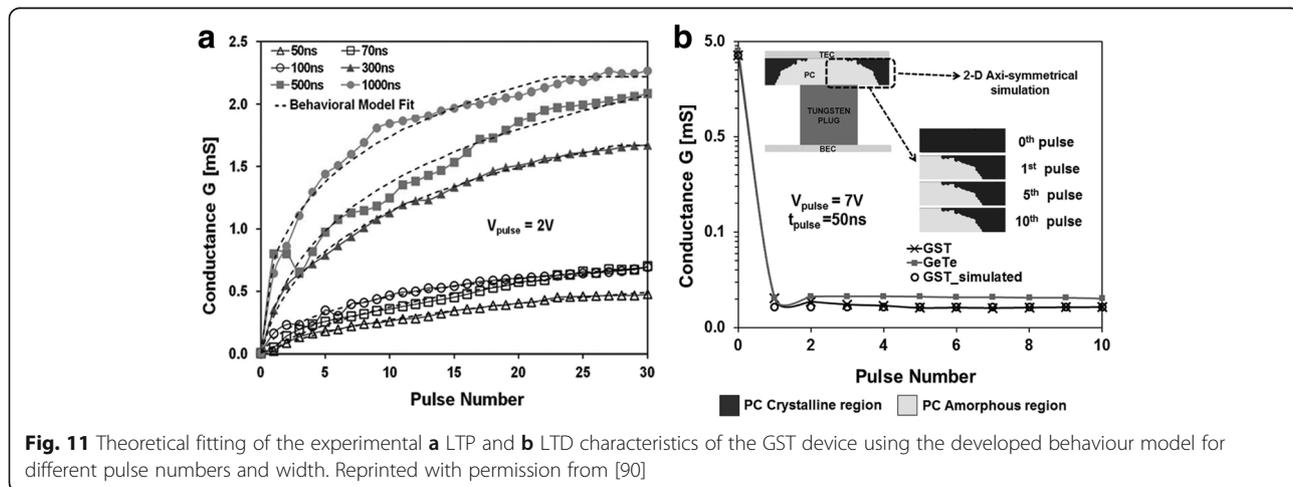
Although this model can accurately forecast the biological behaviours of brain synapses, it cannot be directly applied to a hybrid neural circuit comprising CMOS neurons and PCM-based synapse that requires a circuit-compatible model [90]. In order to resolve this issue, a more comprehensive circuit-compatible model consisting of electrical, thermal, and phase-change sections has been developed to mimic the progressive character of the LTP experiments [90]. Electrical sections is governed by the well-known Ohm’s law, and the adopted device resistance R_{dev} is the sum of the electrode resistance R_s and the resistance of the GST layer that depends on the volume fraction ratio of amorphous to crystalline phase as well as the resistance drift effect [93], given by

$$R_{\text{dev}} = R_s + R_c^{(1-C_a)} + R_{0a} \left(\frac{t}{t_0}\right)^{C_a dr} \tag{4}$$

where R_c and R_{0a} is the resistance of fully crystalline and amorphous phase respectively and C_a is the amorphous volume fraction; t_0 is the time when the initial phase-change process begins, and dr is an exponent that indicates the power-law slope. The thermal section calculates the temperature T_b inside the GST media using

$$T_b = T_0 + P_t R_{t\text{gst}} \tag{5}$$

where T_0 is the ambient temperature, P_t is the electrical power resulting from Joule heating, and $R_{t\text{gst}}$ is the thermal resistance of the device, described by



$$P_t = \frac{u^2}{R_s + R_{0c}} \quad (6)$$

$$R_{tgst} = (1 - C_a)R_{tc0} + C_a R_{ta0} \quad (7)$$

where u is the applied electric pulse and R_{tc0} and R_{ta0} are the thermal resistance of the full crystallized and amorphized phases.

The phase-change section is mainly employed to calculate amorphous volume fraction C_a that strongly relates to the electrical and thermal sections using crystallization rate equation:

$$\frac{dC_a}{dt} = -\frac{C_a^2}{\tau_c} \left(1 - \exp\left(\frac{E_a T_b - T_m}{T_b}\right)\right) \exp\left(-\frac{E_b}{T_b}\right) \text{ when } T_b > T_m \quad (8)$$

$$\frac{dC_a}{dt} = -\frac{1}{\tau_a} \frac{T_b - T_m}{T_m} \text{ when } T_m > T_b > T_g \quad (9)$$

where T_m and T_g are melting and glass transition temperature for GST; E_a and E_b are the fitting parameters for crystallization rate at high and low temperature, respectively; τ_a and τ_c are the fitting amorphization and crystallization rates, respectively.

By solving the aforementioned electrical, thermal, and phase-change sub-models simultaneously, this circuit-compatible model exhibits a less closer fitting to the observed LTP behaviour of the GST device in comparison with the previous behaviour model, as shown in Fig. 12. However, due to its capability of capturing the correct behaviour of PCMs for a relatively wide range of measurements with a small number of semi-physical parameters, it is therefore suitable for the easier circuit design that takes advantage of PCMs to emulate millions of synapses.

Another behaviour model to study the computational properties of networks of synapses based on phase-

change materials was developed by Jackson et al. [94]. In this model, the state transition is determined by a calculated probability depending on the initial conductance and pre-post spike timing. In this case, the device conduction is extracted from a log-Gaussian distribution once the probability indicates the occurrence of the state transition. The pre-before-post pairings are tackled completely separately from post-before-pre pairings, each requiring a distinct set of model parameters. The probability of this phase-change-based synapse is given for each type of pairing between the amorphous and crystalline states as a function of the pre-post spike timing Δt , tailored by the initial device resistance R_i following:

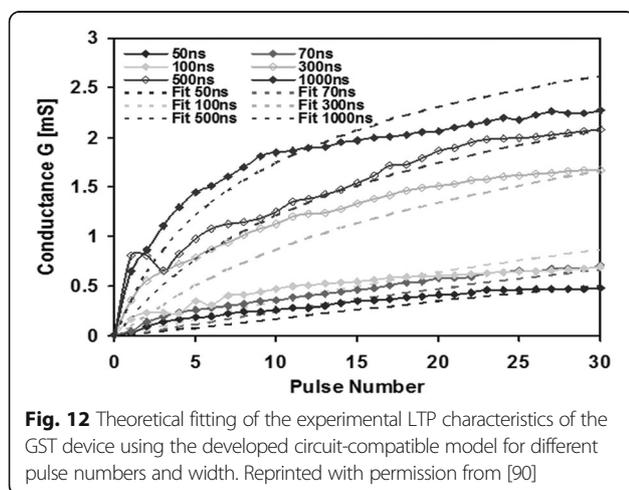
$$P(\text{transition}) = \left(1 + \exp\left(\frac{\Delta t + \alpha \log_{10} R_i + \beta}{\kappa}\right)\right)^{-1} \quad (10)$$

where α , β , and γ are timescale-dependent, threshold-dependent, and initial resistance-dependent parameters, respectively. Hence, the probability for a resistance $R_f = 10^x$ is calculated by:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu \exp(\gamma \Delta t))^2}{2\sigma^2}\right) \quad (11)$$

where μ , σ , and γ are parameters regulating the mean, standard deviation, and dependence on spike timing of the log normal distribution, respectively. It should be noticed that these parameter values rely on the physical characteristics of the device and the pulse scheme harnessed to induce phase transformations, showing a good agreement with the experimentally measured distribution of conductances and the STDP-like dynamics.

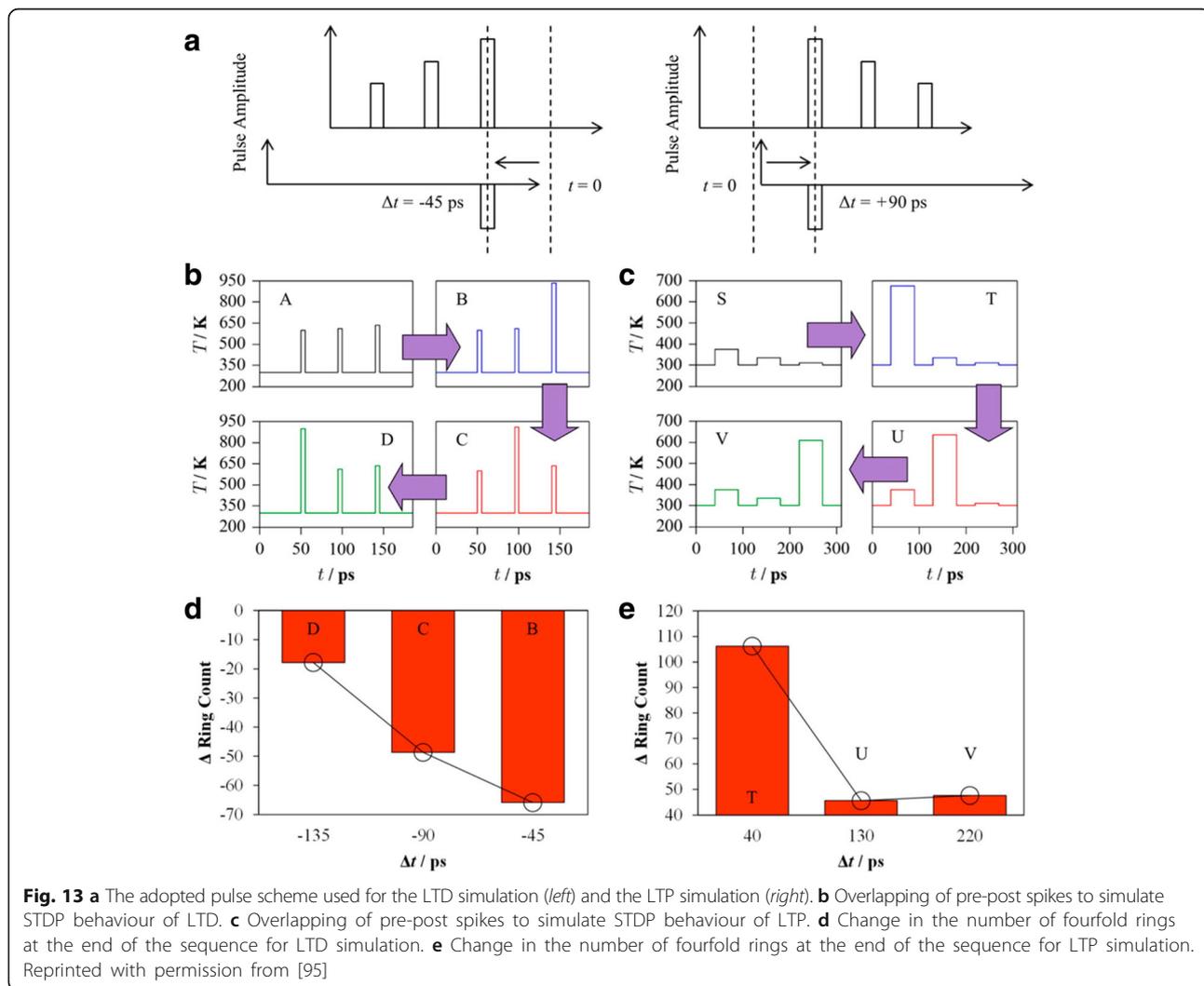
Differing from aforementioned models, the response of PCM-based synapse to the time delay between pre-post spikes has recently been re-examined by an ab initio molecular-dynamics (AIMD) model that consists of 180 atom models created in cubic supercells with periodic boundary conditions at a fixed density of 6110 kg/m³[95]. PAW pseudopotentials are deployed in this case to handle the outer s and p electrons as valence electrons [96], associated with the Perdew-Burke-Enzerhof (PBE) exchange-correlation functional and a plane-wave kinetic-energy cutoff of 175 eV [97]. During the simulation, phase-change materials, referring to GST, are initially considered to have amorphous phase obtained from a conventional melt-quench approach at a cooling rate of -15 K/ps. Crystallization in this model is achieved by heating it at 500 K for 500 ps, thus leading to the adequate number of the defined fourfold rings that represent the structural order of the GST media [98, 99]. According to this model, fourfold rings are interpreted as four atoms forming a closed path with a maximum bonding distance and with



all four three-atom bond angles, plus the angle between the planes defined by two triplets of atoms, being a maximum of 20° from the ideal angles of 90° and 180°, respectively.

In order to simulate the STDP behaviour of the designed artificial synapse, a train of temperate pulses with different width and magnitudes are introduced into this model to represent the corresponding pre- and post-spikes. For LTD simulation, the pre-spikes comprise three stepcase 5-ps temperature pulses of 600, 610, and 635 K, separated by a time interval of 40 ps at 300 K between each single pulse. The pre-spikes are set to have sufficient magnitude but with too short duration to induce the crystallization. The post-spike consists of a sole temperature pulse of 300 K. As a result, by modulating relative time window between pre-post spikes, the single post-spike can overlap with the pre-spike with different temperature magnitude. As the superposed temperature exceeds the melting point, the resulting pre-post spikes

would attenuate the degree of crystallinity of GST and thus decrease the conductivity, reflected in this model by a reduction in the number of fourfold rings. On the other hand, the pre-spikes for LTP simulation consist of a sequence of 50 ps temperature pulses of 375, 335, and 310 K with a time interval of 40 ps, while the post-spike remains the same as the LTD case. The choice for the LTP pre-spikes follows a rule that these temperature pulses have sufficiently long width but with too low amplitude to cause crystallization. Therefore, the resulting temperature possessed from the overlapping of pre-post spikes outnumbers the crystalline point and consequently boost a growth in structure order, increasing the conductivity. The described pulse scheme as well as the resulting influence on the structure order of GST is schematically shown in Fig. 13. Additionally, in order to deeply analyze and interpret the results shown in Fig. 13, the effects of various single pulse width and temperature on the characteristics of the designed PCM-

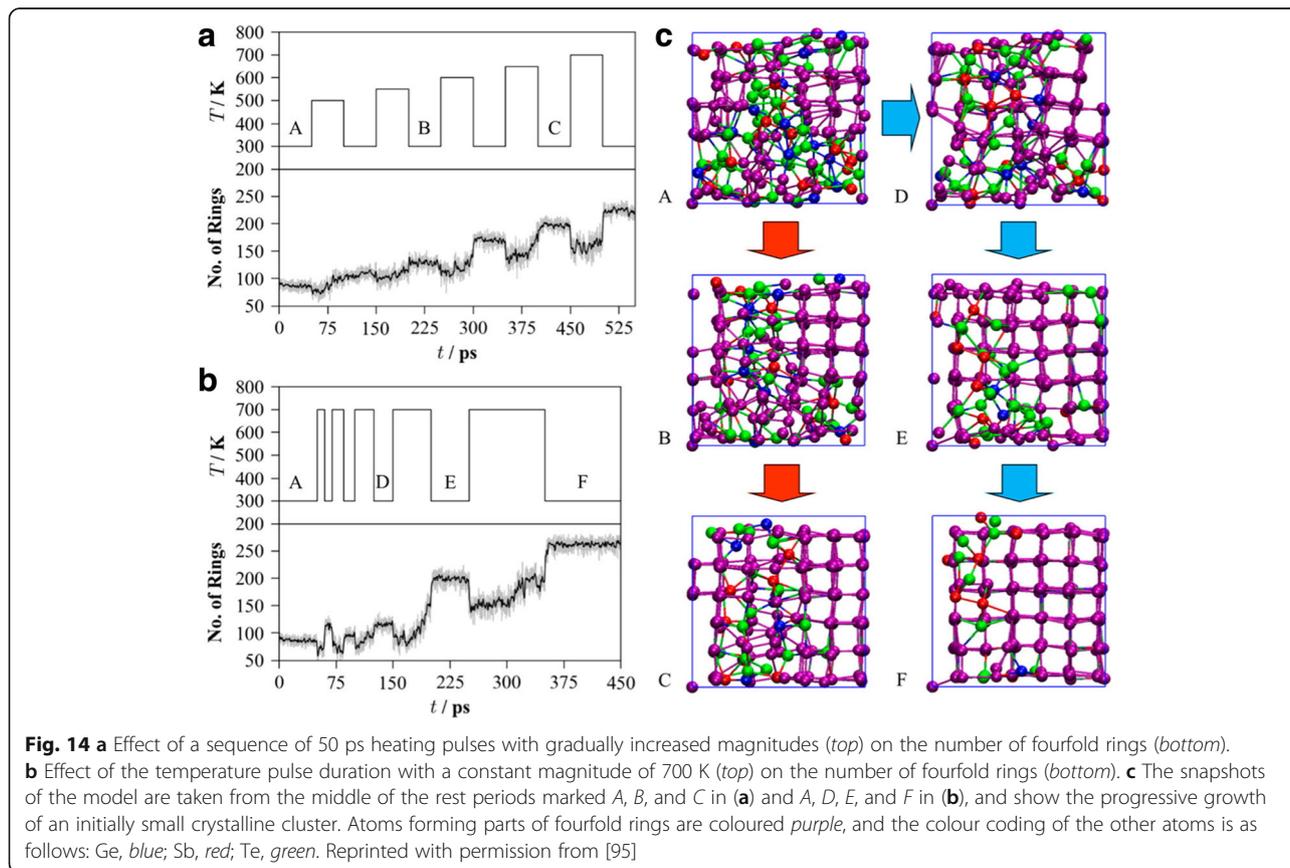


based synapse are also evaluated, giving rise to Fig. 14. It is clearly indicated in Fig. 14c that a small stable crystalline cluster is formed by a low-temperature annealing of the melt-quenched amorphous model prior to the neuromorphic simulations, followed by a subsequent growth proportional to the magnitude and duration of the applied heat pulses, as illustrated in Fig. 13.

PCM-Based Synapse Devices

The resulting theoretical feasibility secured from various *in silico* models has dramatically strengthened the confidence of worldwide researchers in physically realizing the desired PCM-based synapse. Thanks to this, the pioneering electronic synapse using phase-change materials was achieved by Kuzum et al. in 2011 to emulate biological STDP behaviour according to a very delicate control of the intermediate resistance level in the phase-change material (GST in this case) that is regarded as the synapse weight [100–102]. In their neuromorphic circuit, the pre-post spikes generated from an arbitrary waveform generator are connected to the top and bottom electrodes of an electronic synapse. Pre-spikes that consist of a train of stepwise pulses have successive larger magnitude for depression simulation corresponding to reset state, while exhibiting gradually decreasing

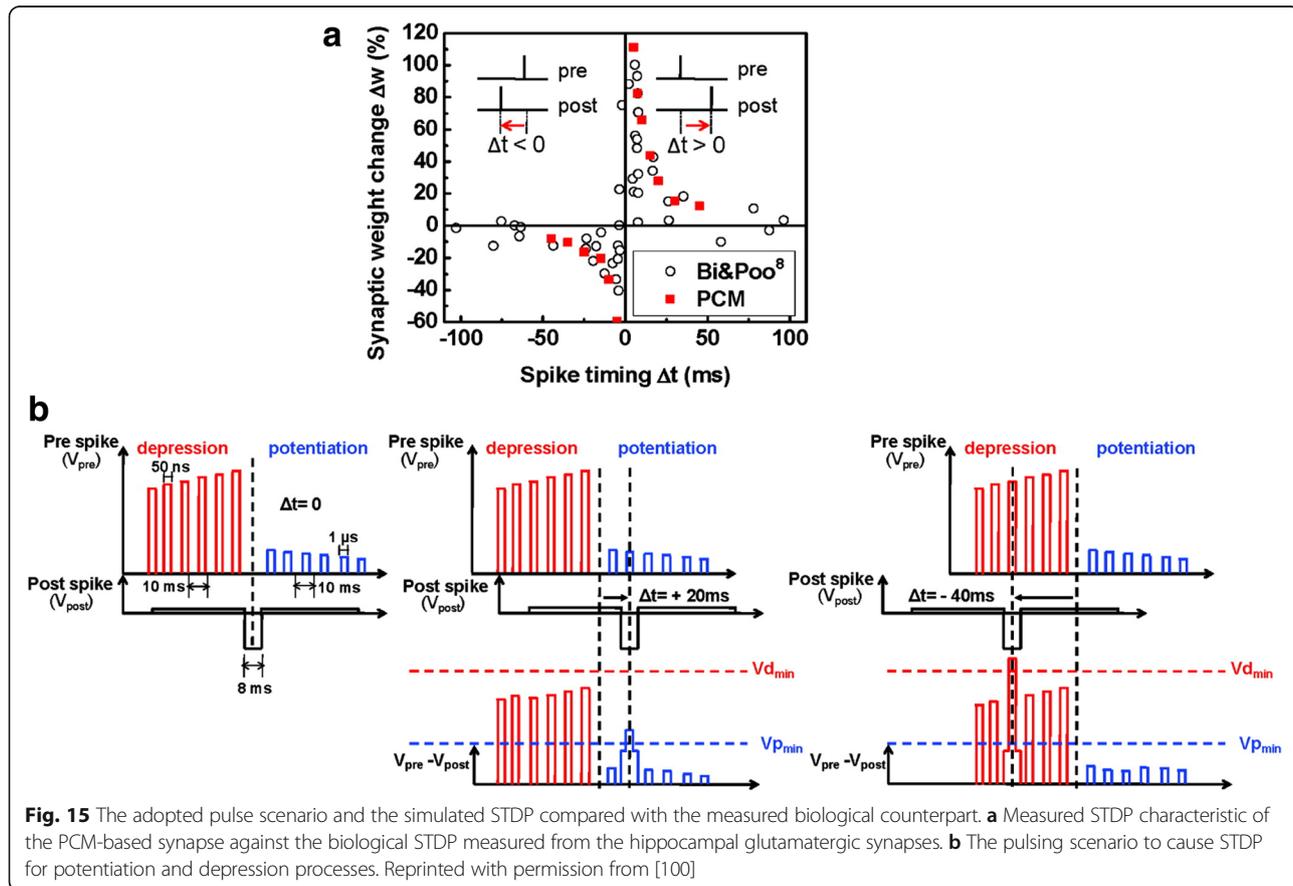
magnitude for potentiation case indicated by set state. The pulse width and rise and fall times of a single depression pulse is chosen to be 50, 10, and 10 ns, whereas they are 1 μ s, 100 ns, and 100 ns for a single potentiation pulse. The time interval between each depression or potentiation pulse is 10 ms, and the entire duration of pre-spikes last 120 ms to match the biological synapse. In contrast to pre-spikes, post-spikes that operate as gating function for the pre-spike is a single pulse with a negative low magnitude and duration of 120 ms while having 8 ms at the centre. In this case, the net electric potential applied across the electronic synapse is equal to the difference between the magnitude of pre-spike and that of post-spike. If pre-post spikes are synchronous, the post-spike superposes with neither depression nor potentiation pulses of the pre-spikes. However, for the case of pre-spikes that precede post-spikes, the net potential drop across the designed synapse would exceed the so-called potentiation threshold, i.e. the minimum voltage to induce crystallization, which can increase the device conductance, suggesting an increase on synaptic weight. On the other hand, when the pre-spike lags behind the post-spike, the post-spike is overlapped with pre-spikes having larger magnitude so that the overall voltage drop would be above the depression threshold

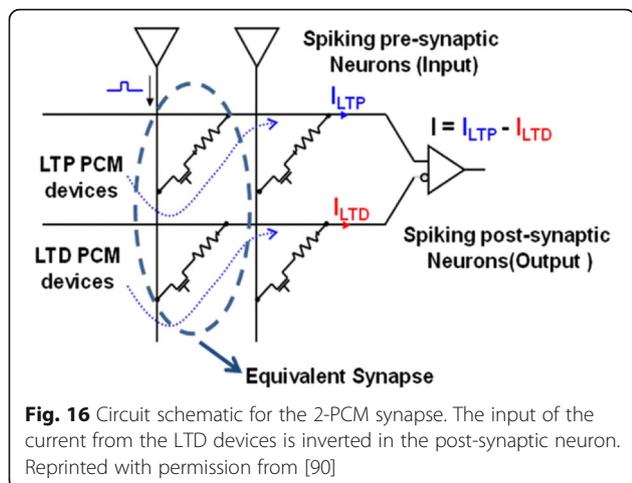


representing the minimum voltage for amorphization. This would effectively reduce the conductance, thus corresponding to a decrease on the synapse weight. More importantly, modulation of the relative time delay between pre-post spikes allows post-spike to overlap with pre-spike with different magnitudes and leads to various extents of either amorphous or crystalline states, closely mimicking the biological STDP mechanism. The presented pulsing scenario as well as the simulated STDP function is schematically shown in Fig. 15.

There is no doubt that the above approach can astonishingly imitate both LTD and LTP properties of biological synapse based on the adjustment of the intermediate resistance level. However, for implementation of large-scale neural systems, this complex resistance-control method would cause several severe issues such as capacitive line charging and high power dissipation [90]. In addition, as the emulation of the LTD process is strongly linked to amorphization, higher energy consumption is therefore required than the LTP process involving crystallization with less energy consumption. To address these issues, a novel paradigm to harness two PCM devices to simulate a single synapse, namely 2-PCM synapse, has been proposed by Suri et al. [90, 91], as illustrated in Fig. 16.

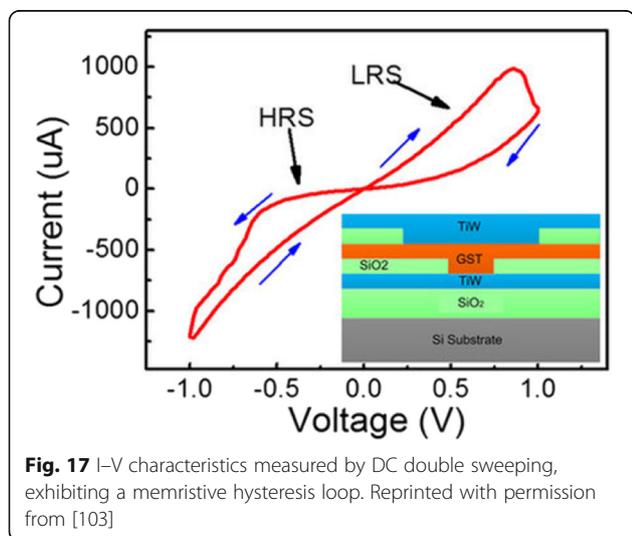
According to Fig. 16, an electronic synapse consists of one PCM device responsible for the LTP and another PCM device in charge of the LTD. The initial phases of both devices are considered to be amorphous. In contrast to the previous design, both LTP and LTD characteristics in this 2-PCM synapse are realized through the crystallization of phase-change materials. The LTP PCM device gives a positive current distribution, whereas current flowing through the LTD PCM device is negative by means of an inverter. Therefore, the potentiation of the LTD device in turn leads to a synaptic depression due to the subtracted current through it in the post-neuron. The pre-post spikes are configured in such a way that the overall impact of the two pulses only potentiate or partially crystallize the LTP device without affecting the LTD device in the case of pre-spike fired before post-spike, while the superposition of the two pulses only potentiates the LTD device rather than the LTP device when the post-spike precedes the pre-spike. Following above rule, the pre-neuron would send out a read pulse and enters 'LTP' mode for time t_{LTP} during which the LTP synapse undergoes a partial SET pulse if the postsynaptic neuron spikes; otherwise, the LTD synapse is programmed. One prominent advantage of this 2-PCM synapse obviously arises from its operations in





crystalline regime for both LTP and LTD events that consumes less energy than amorphization case. Additionally, as the information stored in the designed device is chiefly in crystalline state, it is therefore less susceptible to the resistance drift effect that often occurs in phase-change materials in amorphous states.

Li et al. proposed another PCM-based synapse whose biological events are mainly controlled by crystallization rather than amorphization. In their experiment, the biological synapse is emulated by a crystalline GST (c-GST)-based memristor whose resistance varies between 500 Ω and 10 kΩ when subject to a DC voltage sweep [103, 104], as illustrated in Fig. 17. Accordingly, the synaptic weight is denoted by the conductance of the c-GST memristor that can be modulated by the applied electric pulses. It was found that the device conductance can be increased via the negative pulses that thus represent the potentiating spikes, while the positive spikes leading to the reduction on the device conduction corresponds to the depressing



spikes. To trigger the synaptic behaviour in crystalline regime, the device was first crystallized with a resistance of 2.6 kΩ by a SET pulse. Similar to aforementioned pulse schemes, the potentiating spikes consist of a sequence of the pulse trains with successive increasing amplitude from -0.6 to -0.8 V with a 10-mV step, whereas the amplitude of the successive depressing spikes ranges from 1 to 1.8 V with a 40-mV step. The rest, rise, and trailing periods of a single spike is fixed to 30, 10, 10 ns, respectively. As a result, varying the relative time delay between pre-post spikes clearly results in various net voltage drops across the electronic synapse after the superposition of the pre- and post-spikes, thereby accurately resembling the typical or even more complex STDP forms, as shown in Fig. 18. Such an analogue phase-change synapse exhibits several advantages such as ultra-low operation voltage, ultra-fast synaptic events, and feasibility of time window tuning when compared to other emerging electronic synapses.

The practicality of using different pulsing schemes from above devices to generate STDP has recently been investigated by Jackson et al. who built an electronic synapse that includes a FET acting as an access device, connected in series with a PCM device [94]. A 200-ms electric pulse with front portion having higher amplitude and back portion having lower amplitude is injected to the gate of the synaptic FET with zero time delay after the presynaptic spike. Another 60-ns electric pulse with a time delay of 100 ms after postsynaptic spike is directly applied to the PCM device. The amplitudes of the pre-post spikes are established in such a way that the FET only programmes during the brief overlap between these two signals. According to this strategy, when the presynaptic neuron is spiked before post-neuron, the shorter pulse only coincides with the portion of the longer pulse with lower amplitude, whereby the resulting power only enables the crystallization and thus increases the device conductance. On the contrary, for the case of post-spike preceding the pre-spike, the overlapping of the shorter pulse with the portion of the longer pulse with larger amplitude would yield sufficiently high power to drive the temperature inside the phase-change materials towards the melting point, thus inducing amorphization and decreasing the conductance. Therefore, the consequent device conductance entirely depends on the coincidence of the short and long electrical signals whose characteristics can be modulated by the relative pulse time window. Such a modulation method in associated with the simulated STDP is given in Fig. 19.

Differing from the first route, Jackson et al. also devised another electronic device to produce STDP where time delay is internally determined by a simple RC circuit [94], as revealed in Fig. 20. The memory of neuron's latest firing event in this method is interpreted by the spike-timing-dependent change in synaptic

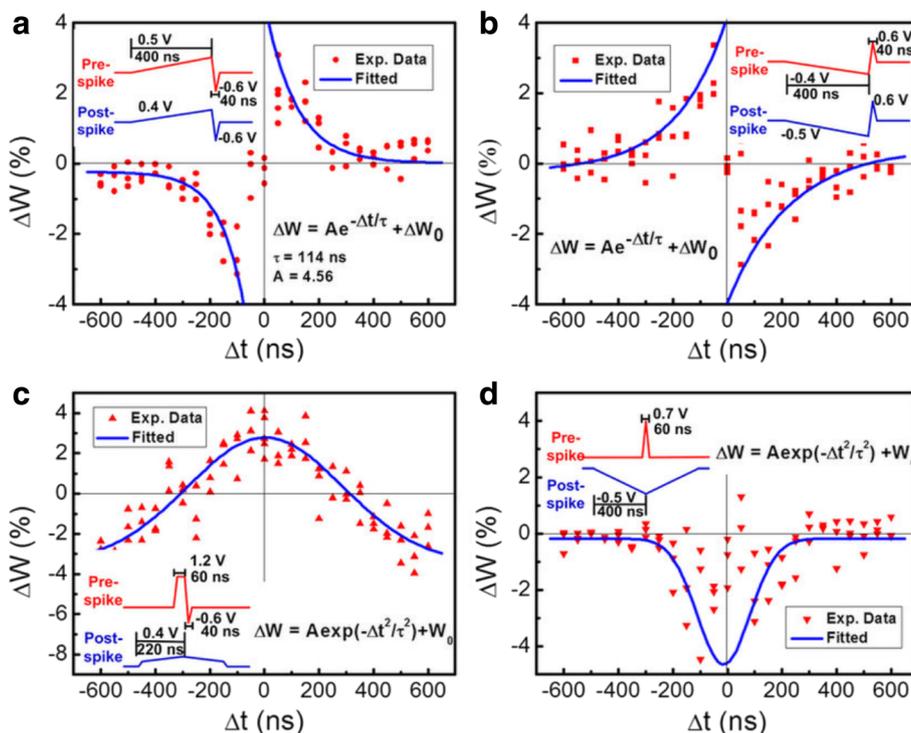


Fig. 18 Four types of the STDP characteristics obtained from the c-GST-based electronic synapse. **a** Antisymmetric Hebbian learning rule. **b** Antisymmetric anti-Hebbian learning rule. **c** Symmetric Hebbian learning rule. **d** Symmetric anti-Hebbian learning rule. Reprinted with permission from [103]

conductance determined by an asynchronous handshaking mechanism. The key insight of this mechanism is that the spiking neuron would initialize its capacitor as well as fire a short ‘alert’ pulse to its synaptic partners. The function of the alert pulse is to remind the synaptic partners to examine the instantaneous voltages across their respective capacitors and to stimulate a response pulse after a constant delay that is therefore synchronous with a gating pulse from the spiking neuron. Therefore, the device can only be programmed during the

overlapping of the gating and response pulses, which implies that only the synapses in conjunction with the spiking neuron can be programmed for large crossbar array. The width of the gating pulse offsets any unwanted temporal jitter between the arrival of the alert and response pulses; the amplitude change of the synaptic weight relies on the amplitude of the response pulse whose shape is determined by whether the response pulse is sent to an axon or a dendrite. Note that axon usually generates a square response pulse with very short

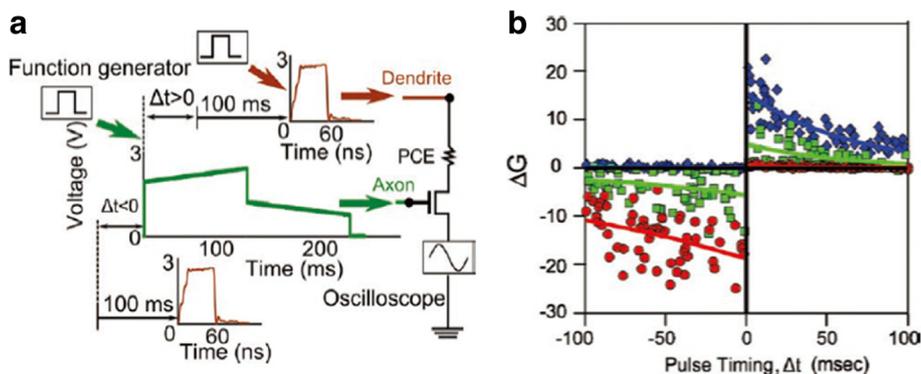
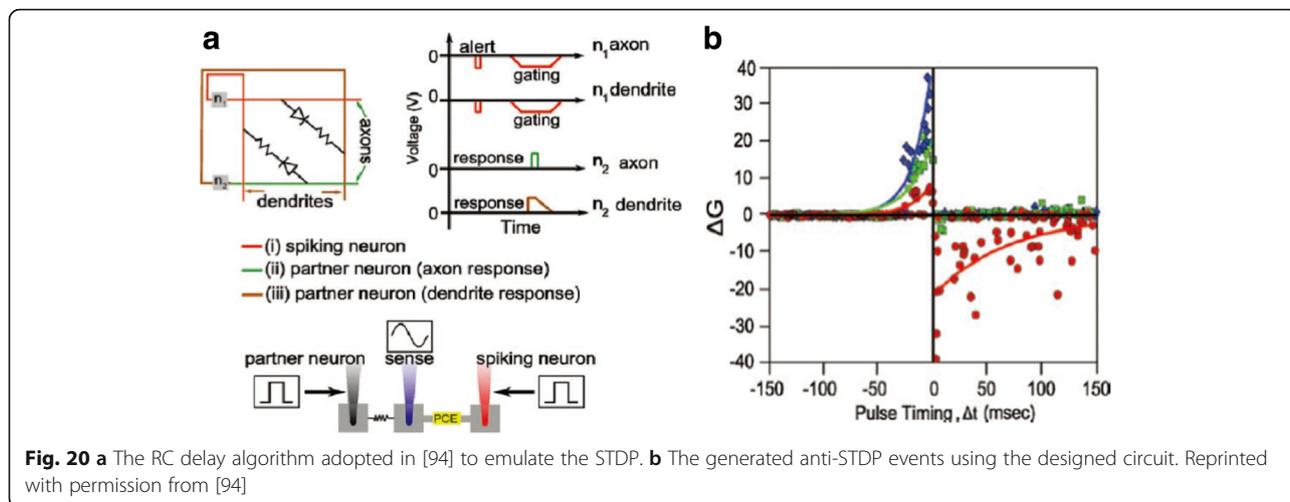


Fig. 19 a The pulse algorithm adopted in [94] to emulate the STDP. **b** The generated STDP events using the designed circuit. Reprinted with permission from [94]

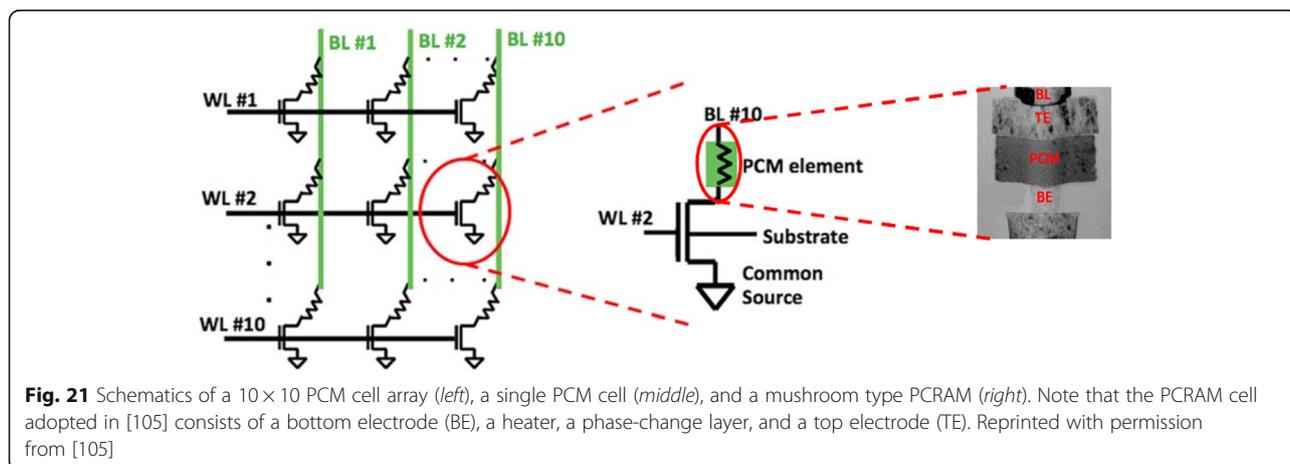


falling time, thus giving rise to a very high quenching rate and decreasing the device conductance due to the amorphization. By contrast, response pulses obtained from dendrites are triangular shaped, which would increase the device conductance because of the crystallization effect. Based on this algorithm, the exponential time dependence of RC circuits permits the neurons that have not fired recently to only send low magnitude response signals. The merit of this newly proposed method arises from the fact that the neuronal spiking duration in this case is a sum of nanosecond-scale alert, nanosecond to microsecond delay, and nanosecond to microsecond gating pulses, resulting in a total duration on the order of a few microsecond that exceeds the biological synaptic rate at the same delay level of 100 ms.

PCM-Based Neural Networks

Considering the large amount of biological synapses in human brain, it would be imperative to design a much more complex circuit than a single PCM-based synapse in order to emulate the STDP events at the network

level. To achieve this goal, the idea to make use of PCM cells in a crossbar fashion to mimic the brain neural networks has been experimentally demonstrated by Eryilmaz et al. As can be seen from Fig. 21, a 10-by-10 memory array consisting of 100 PCM cells configured in a crossbar fashion are deployed in their experiment to imitate the large networks [105]. Each cell comprises a PCM device using a conventional mushroom type PCRAM and a selection transistor. Each cell can be accessed either through a bitline (BL) that is connected to the gates of selection transistors of 10 memory cells, or a wordline connected to the top electrode of the PCM element of 10 memory cells. According to this design, biasing the corresponding BL and WL nodes allows each cell to be exclusively accessed. To simulate the synaptic events, a pulse scheme that harnesses a RESET pulse of 1.1 V, followed by nine SET pulses of 0.85 V, was implemented to induce the gradual resistance change between high resistance and low resistance states, thus leading to nine distinguishable resistance levels.



A more attractive feature of the designed phase-change synapse array stems from its ability to perform the associative learning function. Before the experiment, all the synapses are initially set to amorphous state, and the learning experiment consists of epochs during which synaptic weights are updated depending on firing neurons. The pattern is monitored after the training but with an incorrect pixel, and the incorrect pixel is expected to be recalled in the recall phase after training is performed. Therefore, a complete pattern appears during the training phase of an epoch, while an incomplete pattern with an incorrectly OFF pixel is presented during the recall phase. All patterns consist of 10 pixels, and each neuron is associated with a pixel. By means of this paradigm, test pattern was found to be stored and recalled associatively via Hebbian plasticity in a manner similar to the biological brain, as verified in Fig. 22.

Most recently, Ambrogio et al. has proposed another physics-based model to simulate the large-scaled synaptic networks that consist of numerous phase-change synapses with one transistor/one resistor (i.e. 1T1R) architecture [106, 107], as illustrated in Fig. 23. It is clearly indicated in Fig. 23 that the pre-spike defined as a train of rectangular

pulses is directly applied to the gate of the transistor, and the positive gate voltage consolidated with a constant bias maintained by the postsynaptic neuron (POST) at V_{TE} triggers a current spike in the synapse that flows towards POST. As V_{TE} is usually set to be negatively constant, the resulting current spike is also negative and thus yields a staircase-like increase of the internal potential V_{int} in the inverting integrator. Once V_{int} exceeds the threshold voltage of the comparator V_{th} , the POST excites a forward spike to the neuron in the next layer and a backward spike to V_{TE} to induce STDP through the change on the synaptic weight (PCM conductance here). V_{TE} spike in this case is made up of two rectangular pulses; the second of which is required to have higher amplitude than the first one. Therefore, the adopted shape of the VTE spike can effectively modulate the conductance of PCM device as a function of the relative time delay between pre-post spikes, clearly fitting the STDP behaviours. Furthermore, in order to prove its learning capability, simulations of pattern learning on a two-layer networks consisting of 28×28 pre- and one post-neuron with the proposed 1T1R synapse was performed, resulting in Fig. 24. The developed model allows for MNIST digit recognition probability of

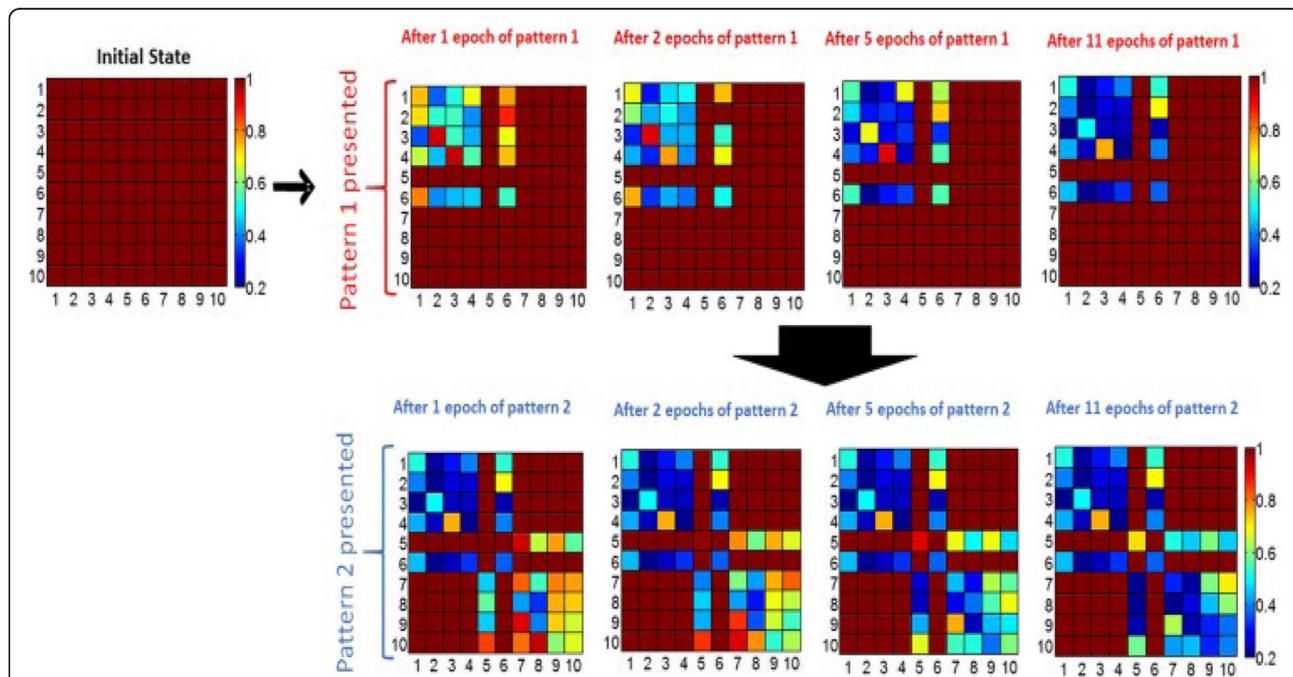
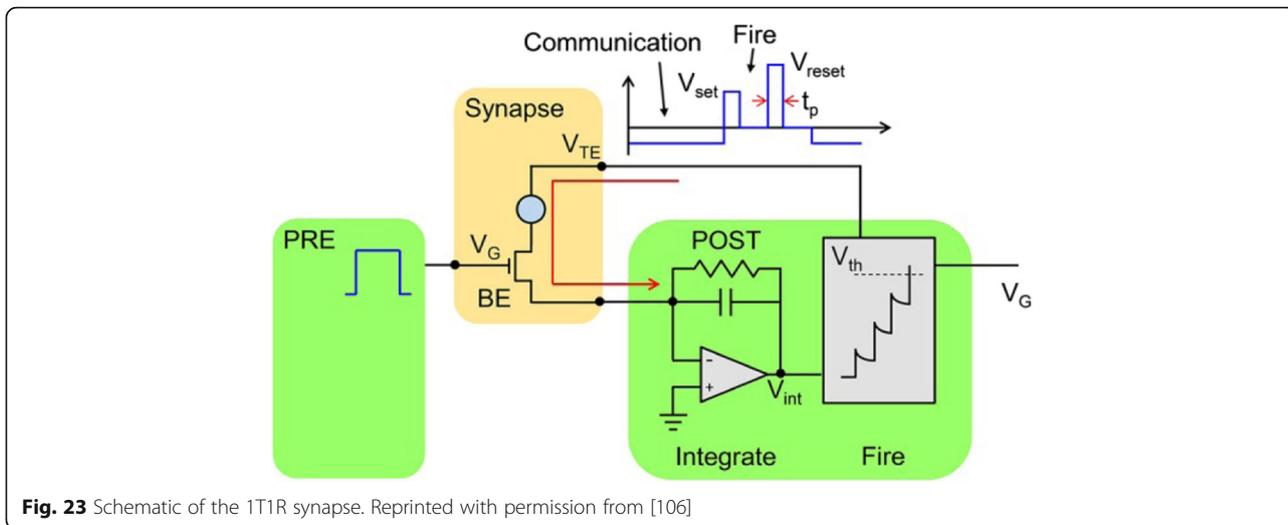


Fig. 22 Evolution of normalized resistance of synaptic devices is shown, for the 60% initial variation case. All normalized resistances are one initially since the normalized resistance map shows the current resistance of a synaptic device divided by its initial resistance. Note that the row and column numbers corresponds to BL and WL that connect the synaptic devices. For instance, the data shown in row #3 and column #6 is the normalized resistance of the memory cell that can be accessed by BL #6 and WL #3. First, pattern 1 is presented to the network. For pattern 1, ON neurons for the complete pattern during update phase are N1, N2, N3, N4, and N6, and for the recall phase, N6 is OFF and expected to be recalled (i.e. expected to fire) after training with a certain number of epochs. The gradual decrease in the normalized resistance of synaptic connections between firing neurons during the update phase can be observed. After 11 epochs, when recall phase is performed, OFF pixel #6 (neuron #6) is recalled (meaning neuron #6 fires in recall phase), and then pattern 2 is presented for training. For pattern 2, the complete pattern is represented by N5, N7, N8, N9, and N10, and N5 is missing in the recall phase. Reprinted with permission from [105]



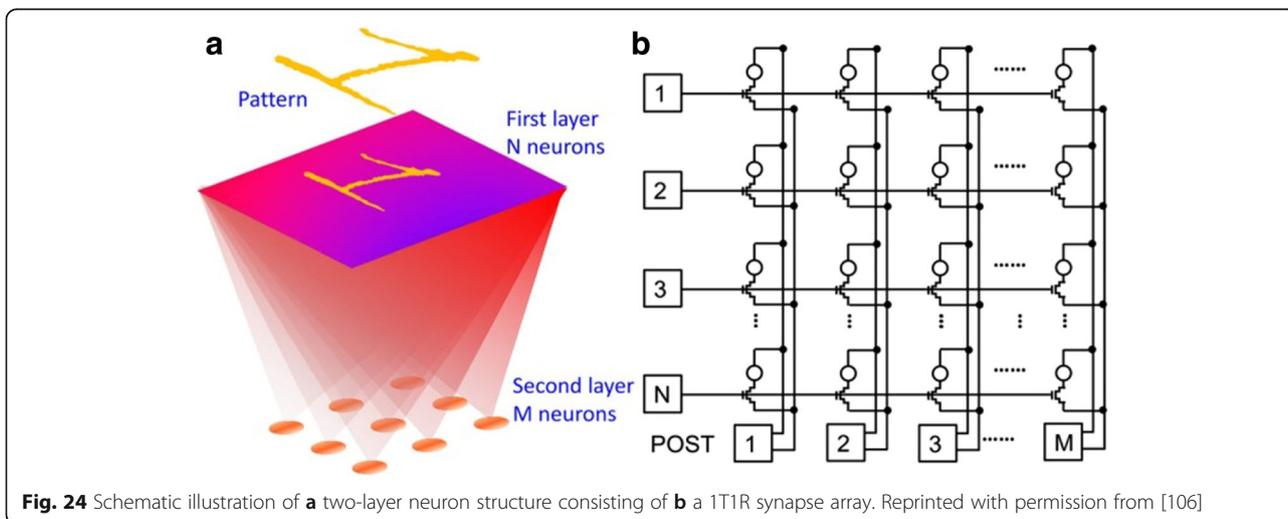
33% and a corresponding error of 6%, which can be further improved to 95.5 and 0.35% respectively by using three-layer structure with 256 neurons [106].

PCM-Based Neurons

No doubt that all results presented so far are dedicated to the emulation of biological synapse using PCM device. However, in order to ultimately achieve the brain-like neural networks, the capability of using PCM-based device to imitate biological neuron with the involvement of maintenance of the equilibrium potential, the transient dynamics, and the neurotransmission process is highly preferred. The prerequisite for PCM-based neuron realization is to effectively simplify the complex biological neuron mechanism that can be therefore readily applied to hardware [108]. Unlike the PCM-based synapse with continuously adjustable device conductance, the PCM-based neuron must be able to ‘fire’ after receiving a certain

number of pulses that can influence an internal state that does not necessarily relate to the external conductance unless the neuron fires when exceeding the threshold. Additionally, non-volatility usually required for electronic synapse is not mandatory for neuron emulation that can utilize volatility to implement leaky integrate-and-fire dynamics.

When simulating the biological neuron, particular attentions need to be paid to stochastic neuronal dynamics that was reported to account for signal encoding and transmission besides the deterministic neuronal dynamics [3]. The cause of this stochastic behaviour can be owed to several complex phenomena such as inter-neuron morphologic variabilities, chaotic motion of charge carriers due to thermal noise, ionic conductance noise, and other background noise [109]. As a result, it would be necessary for artificial neuron to reflect this stochastic firing behaviour so as to closely mimic the biological



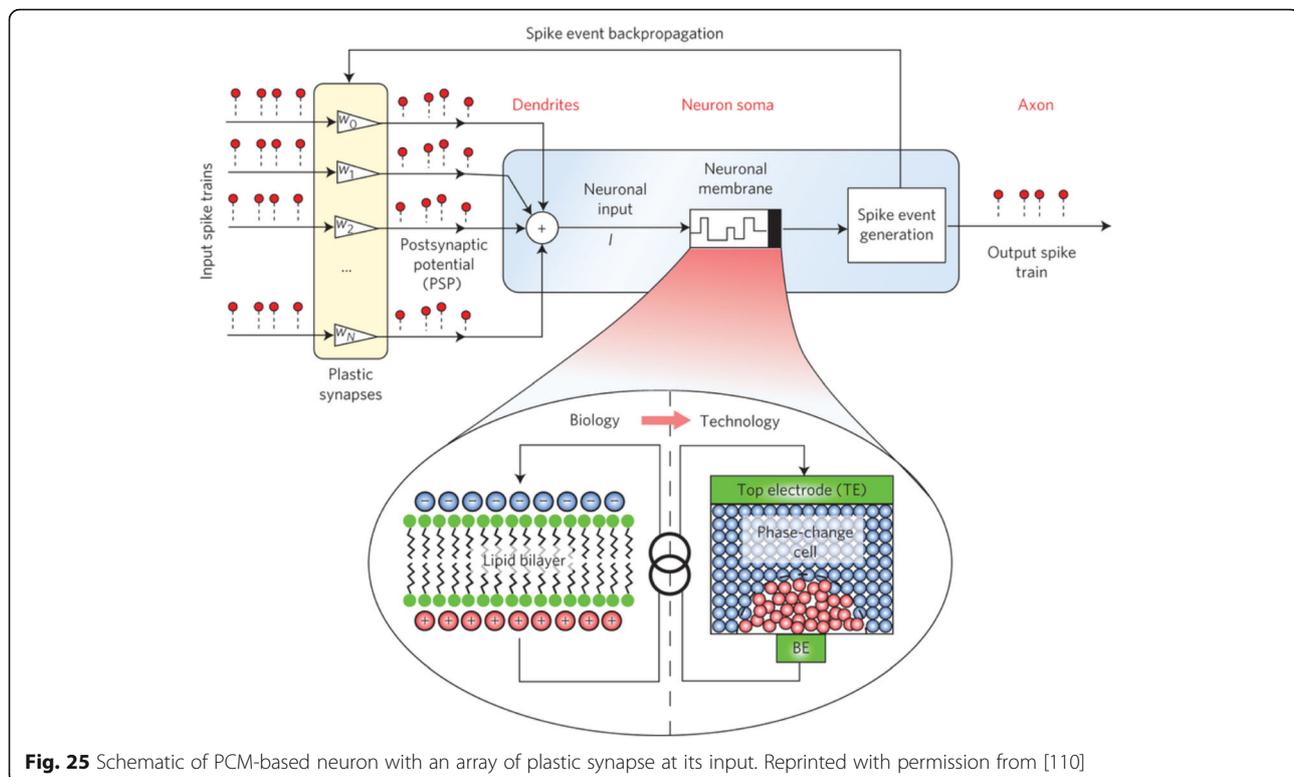


Fig. 25 Schematic of PCM-based neuron with an array of plastic synapse at its input. Reprinted with permission from [110]

brain. In spite of these stringent requirements, Tuma et al. have recently devised a PCM-based neuron shown in Fig. 25 to integrate postsynaptic inputs [110], whereby the evolution of neuronal membrane potential as encoded by phase configuration within the device was demonstrated. More importantly, the ability to present remarkable inter-neuronal and intra-neuronal randomness using the devised neuron was also verified. Intra-device stochasticity that is usually ascribed to shot-to-shot variability in both internal atomic configuration of the melt-quenched amorphous region and thickness results in multiple integrate-and-fire cycles in a single phase-change neuron to generate a distribution of the interspike intervals, leading to population-based computation. Notwithstanding the slow firing rate of the individual neurons, overall neuron population was proved to accurately indicate fast signals. Based on the designed PCM-based neuron, the detection of temporal correlations within a large number of event-based data streams was demonstrated, and a complete PCM neuromorphic circuit made up of PCM-based neurons and PCM-based synapses has also been delivered [111].

Conclusions

The ability to gradually induce a reversible switch between SET and RESET states, integrated with several superior transition properties such as fast switching speed, low energy consumption, and long retention, has

made phase-change-based devices a leading candidate to emulate the biological synaptic events. Moreover, the excellent scalability of phase-change materials down to 2-nm size [112] also forecasts its potential to reproduce the ultra-high density neurons and synapses inside human brain. In this case, majority of the current work on electronic synapse are devoted to simulate the STDP event of the biological synapse that was reported to govern the learning and memory function by gradually changing the conductance of the phase-change materials, thereby resulting in several novel pulse schemes to adjust the device conductance with respect to the relative time delay between two external stimulus applied to device that represent pre-post spikes. Most importantly, the construction of PCM elements in array level allows for the emulation of large-scale connectivity of human brain with any given neuron having as many as 10,000 inputs from other neurons, which is exemplified by an integrated hardware with 256×256 neurons and 64,000 synapses [113].

In spite of the aforementioned merits, the neurons and synapses based on PCM devices are also facing some serious issues. Although the conductance of PCM device in crystalline phase can be modulated continuously, the device conductance in amorphous phase is found to suffer from a sudden change. This can be alleviated using multiple conductances per synapse and periodic corrections [3], but still remaining questionable. Besides, the inherent weakness of amorphous phase-change

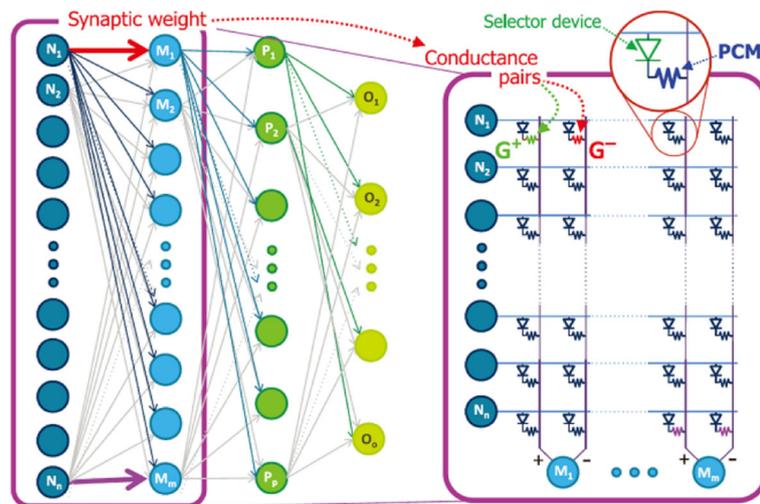


Fig. 26 Implementation of dense crossbar of PCM and selector devices for non-Von Neumann computing where neuron activates each other through dense networks of programmable synaptic weights. Reprinted with permission from [114]

materials including resistance drift or relaxation of the amorphous phase after the melt-quenching also exacerbates the application of PCM on neuromorphic circuit systems. Another issue of PCM-based neuromorphic device arises from its fairly poor performance metrics mainly due to immaturities or inefficiencies in currently developed STDP learning algorithm. Under this circumstance, such inherent imperfections of the PCM devices may pose some indiscernible problems. To boost the classification accuracy, a so-called back propagation method that has received extensive application in computer science field to train artificial neural networks has recently been introduced a three-layer perceptron network with 164,885 synapses based on 2-PCM structure that was trained on a subset of a database of handwritten digits, leading to training and test accuracies of 82–83% [114], as shown in Fig. 26. This work simply implies that classification accuracy can be achieved on the condition that either phase-change materials or the training algorithm permits PCM devices to serve more like a bidirectional NVM with a symmetric, linear conductance response of high dynamic range [115].

Despite the aforementioned challenges, the excellent physical properties of phase-change materials in conjunction with the currently mature technologies on PCM devices has provided an opportunity to envision the success of the future artificial neural networks that can perform the similar complex tasks to human brain while without sacrificing the occupied area and energy consumption. Device models suitable for neuromorphic architectures are still needed for application-specific performance evaluations of these systems.

Acknowledgements

The authors acknowledge the financial supports of the Natural Science Foundation of Jiangxi Science and Technology Department (Grant No. 20151BAB217003) and the Foundation of Jiangxi Education Department (Grant No. GJJ50719).

Authors' Contributions

LW collected and reviewed the data for phase-change synaptic models and phase-change synaptic devices and drafted the manuscript; SL collected and reviewed the data for phase-change neural network; JW collected and reviewed the data for phase-change neurons. All authors read and approved the final manuscript.

Competing Interests

The authors declare that they have no competing interests.

Publisher's Note

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

Received: 4 April 2017 Accepted: 26 April 2017

Published online: 11 May 2017

References

1. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Driessche G et al (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529:484–9
2. Von Neumann J (1993) First draft of a report on the EDVAC. *IEEE Ann Hist Comput* 15:11–21
3. Burr GW, Shelby RM, Sebastian A, Kim S, Kim S, Sidler S et al (2017) Neuromorphic computing using non-volatile memory. *Adv Phys* 2(1):89–124
4. Versace M, Chandler B (2010) The brain of a new machine. *IEEE Spectr* 47(12):30–37
5. Upadhyay NK, Joshi S, Yang JJ (2016) Synaptic electronics and neuromorphic computing. *Sci China: Info Sci* 59:061404–26
6. Ananthanarayanan R, Esser SK, Simon HD, Modha DS (2009) The cat is out of the bag cortical simulations with 109 neurons, 1013 synapses. In: *Proceedings of the Conferences on High Performance Computing Networking, Storage and Analysis*, p 1
7. Ananthanarayanan R, Modha DS (2010) Anatomy of a cortical simulator. In: *Proceedings of the ACM/IEEE Conference on High Performance Networking and Computing*, p 1
8. Mead C (1990) Neuromorphic electronic systems. *Proc IEEE* 78(10):1629–36

9. Djurfeldt M, Lundqvist M, Johansson C, Rehn M, Ekeberg O, Lansner A (2008) Brain-scale simulation of the neocortex on the IBM blue gene/L supercomputer. *IBM J Res Dev* 52:31–41
10. Izhikevich EM, Edelman GM (2008) Large-scale model of mammalian thalamocortical systems. *Proc Natl Acad Sci* 105(9):3593–98
11. Scott JF, Dearaujo CAP (1989) Ferroelectric memories. *Science* 246(4936):1400–5
12. Hiroshi I (2012) Impurity substitution effects in BiFeO₃ thin films—from a viewpoint of FeRAM applications. *Curr Appl Phys* 12:603–11
13. Prejbeanu IL, Bandiera S, Heralut JA, Sousa RC, Dieny B, Nozieres JP (2013) Thermally assisted MRAMs: ultimate scalability and logic functionalities. *J Phys D:Appl Phys* 46:074002–16
14. Chen E, Apalkov D, Smith AD, Khvalkovskiy A, Lottis D, Moon K et al (2012) Progress and prospects of spin transfer torque random access memory. *IEEE T Magn* 48:3025–30
15. Wong HS, Raoux S, Kim S, Liang J, Reifenberg JP, Rajendran B et al (2010) Phase change memory. *Proc IEEE* 98:2201–2227
16. Burr GW, Breitwisch MJ, Franceschini M, Garetto D, Gopalakrishnan K, Jackson B et al (2010) Phase change memory technology. *J Vac Sci Technol B* 28:223–263
17. Jeong DS, Thomas R, Katiya RS, Scott JF, Kohlstedt H, Petraru A et al (2012) Emerging memories: resistive switching mechanisms and current status. *Rep Prog Phys* 75:076502–31
18. Wouters DJ, Waser R, Wuttig M (2015) Phase-change and redox-based resistive switching memories. *Proc IEEE* 103(8):1274–1288
19. Chua L (1971) Memristor-The missing circuit element. *IEEE Trans Circuit Theory* 18(5):507–19
20. Strukov DB, Snider GS, Stewart DR, Williams RS (2008) The missing memristor found. *Nature* 453:80–83
21. Yang JJ, Pickett MD, Li XM, Ohlberg DAA, Stewart DR, Williams RS (2008) Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat Nanotechnol* 3:429–433
22. Jung CH, Lee H, Yoon J, Park J, Jun Seong J, Lee D et al (2009) An electrically modifiable synapse array of resistive switching memory. *Nanotechnology* 20(34):345201
23. Wu Y, Yu S, Wong HSP, Chen YS, Lee HY, Wang SM et al (2012) AlO_x-based resistive switching device with gradual resistance modulation for neuromorphic device application. In: *Proceedings of International Memory Workshop*, p 111
24. Hasegawa T, Ohno T, Terabe K, Tsuruoka T, Nakayama T, Gimzewski JK et al (2010) Learning abilities achieved by a single solid-state atomic switch. *Adv Mater* 22(16):1831–34
25. Jo SH, Chang T, Ebong I, Bhadviya BB, Mazumder P, Lu W (2010) Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett* 10(4):1297–1301
26. Gacem K, Retrouvey JM, Chabi D, Filoramo A, Zhao W, Klein JO et al (2013) Neuromorphic function learning with carbon nanotube based synapses. *Nanotechnology* 24:384013
27. Shen AM, Chen CL, Kim K, Cho B, Tudor A, Chen Y (2013) Analog neuromorphic module based on carbon nanotube synapses. *ACS Nano* 7:6117–22
28. Scott J, Bozano L (2007) Nonvolatile memory elements based on organic materials. *Adv Mater* 19(11):1452–63
29. Tour J, Van Zandt W, Husband C, Husband S, Wilson L, Franzon P et al (2002) Nanocell logic gates for molecular computing. *IEEE Trans Nanotechnol* 1(2):100–9
30. Zhang D, Zeng L, Cao K, Wang M, Peng S, Zhang Y et al (2016) All spin artificial neural networks based on compound spintronic synapse and neuron. *IEEE Trans Biomed Circuits Syst* 10(4):828–36
31. Sengupta A, Al Azim Z, Fong X, Roy K (2015) Spin-orbit torque induced spike-timing dependent plasticity. *Appl Phys Lett* 106:093704
32. Kuzum D, Yu S, Wong HSP (2013) Synaptic electronics: materials, devices and applications. *Nanotechnology* 24(38):382001–22
33. Yamada N, Ohno E, Nishiuchi K, Akahira N (1991) Rapid-phase transitions of GeTe-Sb₂Te₃ pseudobinary amorphous thin films for an optical disk memory. *Jpn J Appl Phys* 69(5):2849–2856
34. Rios C, Stegmaier M, Hosseini P, Wang D, Scherer T, Wright CD et al (2015) Integrated all-photonic non-volatile multi-level memory. *Nat Photonics* 9:725–732
35. Burr GW, Brightsky MJ, Sebastian A, Cheng H-Y, Wu J-Y, Kim S et al (2016) Recent progress in phase-change memory technology. *IEEE J Emerg Sel Topics Circuits Syst* 6(2):146–162
36. Meena JS, Sze SM, Chand U, Tseng TY (2014) Overview of emerging nonvolatile memory technologies. *Nanoscale Res Lett* 9(1):1–33
37. Pereda AE (2014) Electrical synapses and their functional interactions with chemical synapses. *Nat Rev Neurosci* 15:250–63
38. Stamatakis AM, Jennings JH, RI U, Blair GA, Weinberg RJ, Neve RL et al (2013) A unique population of ventral tegmental area neurons inhibits the lateral habenula to promote reward. *Neuron* 80(4):1039–53
39. Preffer CK, Xue MS, He M, Huang ZJ, Scanziani M (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat Neurosci* 16(8):1068–U130
40. Ebert DH, Greenberg ME (2013) Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* 493(7432):327–37
41. Tonnesen J, Katona G, Rozsa B, Nagel UV (2014) Spine neck plasticity regulates compartmentalization of synapses. *Nat Neurosci* 17(5):678
42. Zamarrano-Ramos C, Camunas-Mesa LA, Perez-Carraso JA, Masquelier T, Serrano-Gotarredona T, Linares-Barranco B (2011) On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front Neurosci* 5:1–22
43. Hodgkin AL, Huxley AF (1939) Action potentials recorded from inside a nerve fibre. *Nature* 144:710–11
44. Goldman DE (1943) Potential, impedance, and rectification in membranes. *J Gen Physiol* 27(1):37–60
45. Angew WS, Levinson SR, Brabson JS, Raftery MA (1978) Purification of the tetrodotoxin-binding component associated with the voltage-sensitive sodium channel from *Electrophorus electricus* electroplax membranes. *Proc Natl Acad Sci U S A* 75(6):2606–10
46. Tempel BL, Papazian DM, Schwarz TL, Jan YN, Jan LY (1987) Sequence of a probable potassium channel component encoded at Shaker locus of *Drosophila*. *Science* 237(4816):770–5
47. Noda M, Shimizu S, Tanabe T, Takai T, Kayano T, Ikeda T et al (1984) Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* 312:121–7
48. Bezanilla F (2005) Voltage-gated ion channels. *IEEE Trans Nanobiosci* 4(1):34–48
49. Gadsby DC (2009) Ion channels versus ion pumps: the principal difference, in principle. *Nat Rev Mol Cell Biol* 10(5):344–52
50. Shinoda T, Ogawa H, Cornelius F, Toyoshima C (2009) Crystal structure of the sodium-potassium pump at 2.4 Å resolution. *Nature* 459(7245):446–50
51. Lynch MA (2004) Long-term potentiation and memory. *Physiol Rev* 84:87–136
52. Jeong D-S, Kim I, Ziegler M, Kohlstedt H (2013) Towards artificial neurons and synapses: a materials point of view. *RSC Adv* 3:3169–83
53. Malenka RC, Nicoll RA (1999) Long-term potentiation—a decade of progress? *Science* 285(5435):1870–4
54. Cavus I, Teyler T (1996) Two forms of long-term potentiation in area CA1 activate different signal transduction cascades. *J Neurophysiol* 76(5):3038–47
55. Wankerl K, Weise D, Gentner R, Rumpf JJ, Classen J (2010) L-type voltage-gated Ca²⁺ channels: a single molecular switch for long-term potentiation/long-term depression-like plasticity and activity-dependent metaplasticity in humans. *J Neurosci* 30(18):6197–204
56. Bauer EP, Schafe GE, LeDoux JE (2002) NMDA receptors and L-type voltage-gated Calcium channels contribute to long-term potentiation and different components of fear memory formation in the lateral amygdala. *J Neurosci* 22(12):5239–49
57. Rogan MT, LeDoux JE (1995) LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* 15(1):127–36
58. Bliss TVP, Lomo T (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol* 232:331–56
59. Sjostram PJ, Gerstner W (2010) Spike-timing-dependent plasticity. *Scholarpedia* 5:1362
60. Hebb DO (1949) The first stage of perception: growth of the assembly. In: *The Organization of Behavior*. John Wiley & Sons Inc., Hoboken, pp 60–78
61. Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* 18:10464–72
62. Li Y, Zhong YP, Zhang JJ, Xu L, Wang Q, Sun HJ et al (2014) Activity-dependent synaptic plasticity of a chalcogenide electronic synapse for neuromorphic systems. *Sci Rep* 4:4906–7
63. Van Rossum MCW, Bi GQ, Turrigiano GG (2000) Stable Hebbian learning from spike timing-dependent plasticity. *J Neurosci* 20:8812–21
64. Rubin JE, Gerkin RC, G-Bi Q, Chow CC (2005) Calcium time course as a signal for spike-timing-dependent plasticity. *J Neurophysiol* 93:2600–13

65. Gutig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J Neurosci* 23:3697–3714
66. Purves D, Augustine GJ, Fitzpatrick D, Katz LC, LaMantia A-S, McNamara JO et al (2001) *Neuroscience*, 2nd edn. Sinauer Associates, Sunderland
67. Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Ak F (2014) A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345:668–73
68. Furber SB, Lester DR, Plana LA (2013) Overview of the SpiNNaker system architecture. *IEEE Trans Comput* 62:2454–67
69. Markram H (2006) The blue brain project. *Nat Rev Neurosci* 7:153–60
70. Benjamin BV, Gao P, McQuinn E, Choudhary S, Chandrasekaran AR, Bussat J-M et al (2014) Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proc IEEE* 102:699–716
71. Raoux S, Xiong F, Wuttig M, Pop E (2014) Phase change materials and phase change memory. *MRS Bull* 39:703–710
72. Wuttig M, Raoux S (2012) The science and technology of phase change materials. *Z Anorg Allg Chem* 638(15):2455–65
73. Wang L, Yang C-H, Wen J (2015) Physical principles and current status of emerging non-volatile solid state memories. *Electron Mater Lett* 11(4):505–43
74. Yamada N, Ohno E, Akahira N (1987) High speed overwriteable phase change optical disk material. *Jpn J Appl Phys* 26(26-4):61–6
75. Lankhorst M, Ketelaars B, Wolter R (2005) Low-cost and nanoscale non-volatile memory concept for future silicon chips. *Nat Mater* 4(4):347–352
76. Lacaite AL, Wouters DJ (2008) Phase-change memories. *Phys Stat Sol (a)* 205(10):2281–2297
77. Lai S, Lawrey T (2001) OUM-A 180 nm nonvolatile memory cell element technology for stand alone and embedded applications. In: Proceedings of the International Electron Devices Meeting., p 36.5.1
78. Wright CD, Wang L, Shah P et al (2011) The design of rewritable ultrahigh density scanning-probe phase-change memories. *IEEE Trans Nanotechnol* 10(4):900–912
79. Wang L, Wright CD, Mustafa A, Yang C-H, Yang G-W (2014) A physics-based three dimensional readout model for phase-change probe memory. *Curr Appl Phys* 14(9):1296–1300
80. Wang L, Yang C-H, Wen J, Gong S-D, Peng Y-X (2016) Overview of probe-based storage technologies. *Nanoscale Res Lett* 11:342
81. Stegmaier M, Rios C, Bhaskaran H et al (2016) Nonvolatile all-optical 1 × 2 switch for chipscale photonic networks. *Adv Opt Mater* 5(1):00346–5
82. Stegmaier M, Rios C, Bhaskaran H et al (2016) Thermo-optical effect in phase-change nanophotonics. *ACS Photonics* 3(5):828–835
83. Bhaskaran H, Pernice WHP (2012) Photonic non-volatile memories using phase-change materials. *Appl Phys Lett* 101:171101–4
84. Adler D, Shur MS, Silver M et al (1980) Threshold switching in chalcogenide-glass thin films. *J Appl Phys* 51(6):3289–3310
85. Ielmini D, Zhang YG (2007) Analytical model for sub-threshold conduction and threshold switching in chalcogenide-based memory devices. *J Appl Phys* 102(102):054517–13
86. Karpov VG, Kryukov YA, Savransky SD et al (2007) Nucleation switching in phase change memory. *Appl Phys Lett* 90(12):123504–3
87. Kim IS, Cho SL, Im DH, Cho EH, Kim DH, Oh GH et al (2010) High performance PRAM cell scalable to sub-20nm technology with below 4F² cell size, extendable to DRAM applications. In: Proceedings of the Symposium on VLSI Technology., p 203
88. Loke D, Shi LP, Wang WJ, Zhao R, Yang HX, Ng L-T et al (2011) Ultrafast switching in nanoscale phase-change random access memory with superlattice-like structures. *Nanotechnology* 22(25):254019–6
89. Liang JL, Jeyasingh RGD, Chen H-Y, Wong HSP (2011) A 1.4 μA reset current phase change memory cell with integrated carbon nanotube electrodes for cross-point memory application. In: Proceedings of the Symposium on VLSI Technology., p 100
90. Suri M, Bichler O, Querlioz D, Traore B, Cueto O, Perniola L et al (2012) Physical aspects of low power synapses based on phase change memory devices. *J Appl Phys* 112:054904
91. Suri M, Bichler O, Querlioz D, Cueto O, Perniola L, Sousa V et al (2011) Phase change memory as synapse for ultra-dense neuromorphic systems: application to complex visual pattern extraction. In: Proceedings of the International Electron Device Meeting., p 4.4
92. Bichler O, Suri M, Querlioz D, Vuillaume D, DeSalvo B, Gamrat C (2012) Multi-level recording in erasable phase-change media by light intensity modulation. *IEEE Trans Electron Dev* 59:2206
93. Sonoda K, Sakai A, Moniwa M, Ishikawa K, Tsuchiya O, Inoue Y (2008) A compact model of phase change memory based on rate equations of crystallization and amorphization. *IEEE Trans Electron Dev* 55(7):1672–81
94. Jackson BL, Rajendran B, Corrado GS, Breitwisch M, Burr GW, Cheek R et al (2013) Nanoscale electron synapses using phase change devices. *ACM J Emerg Technol Comput Syst* 9(2):12
95. Skelton JM, Loke D, Lee T, Elliott SR (2015) Ab initio molecular-dynamics simulation of neuromorphic computing in phase-change memory materials. *ACS Appl Mater Interfaces* 7:14223–30
96. Blochl PE (1994) Projector augmented-wave method. *Phys Rev B* 50:17953–79
97. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865–68
98. Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. *Phys Rev B* 47:558R
99. Hegedus J, Elliott SR (2008) Microscopic origin of the fast crystallization ability of Ge-Sb-Te phase-change memory materials. *Nat Mater* 7:399–405
100. Kuzum D, Jeyasingh RGD, Lee B, Wong HSP (2012) Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett* 12:2179–86
101. Kuzum D, Jeyasingh RGD, Wong HSP (2011) Energy efficient programming of nanoelectronic synaptic devices for large-scale implementation of associative and temporal sequence learning. In: Proceedings of the International Electron Device Meeting., p 30.3.1
102. Kuzum D, Jeyasingh RGD, Shimeng Y, Wong HSP (2012) Low-energy robust neuromorphic computation using synaptic devices. *IEEE Trans Electron Dev* 59(12):3489–94
103. Li Y, Zhong YP, Xu L, Zhang JJ, Xu XH, Sun HJ, Miao XS (2013) Ultrafast synaptic events in a chalcogenide memristor. *Sci Rep* 3:1619
104. Zhong YP, Li Y, Xu L, Miao XS (2015) Simple square pulses for implementing spike-timing-dependent plasticity in phase-change memory. *Phys Status Solidi (RRL)* 9(7):414–19
105. Eryilmaz SB, Kuzum D, Jeyasingh R, Kim SB, BrightSky M, Lam C et al (2014) Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front Neurosci* 8:205
106. Ambrogio S, Ciochini N, Laudato M, Milo V, Pirovano A, Fantini P et al (2016) Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Front Neurosci* 10:56
107. Ielmini D, Ambrogio S, Milo V, Balatti S, Wang ZQ (2016) Neuromorphic computing with hybrid memristive/CMOS synapses for real-time learning. In: Proceedings of the International Symposium on Circuits and Systems., p 1386
108. Indiveri G, Linares-Barranco B, Hamilton TJ, Van Schaik A, Etienne-Cummings R, Delbruck T et al (2011) Neuromorphic silicon neuron circuits. *Front Neurosci* 5:73
109. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358
110. Tuma T, Pantazi A, Le Gallo M, Sebastian A, Eleftheriou E (2016) Stochastic phase-change neurons. *Nat Nanotechnol* 11:693–699
111. Pantazi A, Wozniak S, Tuma T, Eleftheriou E (2016) All-memristive neuromorphic computing with level-tuned neurons. *Nanotechnology* 27(35):355205
112. Liang J, Jeyasingh RGD, Chen HY, Wong HSP (2012) An ultra-low reset current cross-point phase change memory with carbon nanotube electrodes. *IEEE Trans Electron Dev* 59:1155–63
113. Kim S, Ishii M, Lewis S, Perri T, BrightSky M, Kim W et al (2015) NVM neuromorphic core with 64k-cell(256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning. In: Proceedings of the International Electron Devices Meeting., p 1
114. Burr GW, Shelby RM, Di Nolfo C, Jang JW, Shenoy RS, Narayanan P et al (2014) Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. In: Proceedings of the International Electron Devices Meeting., p 29.5
115. Burr GW, Shelby RM, Di Nolfo C, Jang JW, Shenoy RS, Narayanan P et al (2015) Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses), using phase-change memory as the synaptic weight element. *IEEE Trans Electron Dev* 62(11):3498–3507