ACS NANO

www.acsnano.org

ARTICLE

Pattern Recognition Using Carbon Nanotube Synaptic Transistors with an Adjustable Weight Update Protocol

Sungho Kim,[†,∥] Bongsik Choi,[‡,∥] Meehyun Lim,[§,∥] Jinsu Yoon,[‡] Juhee Lee,[‡] Hee-Dong Kim,[†] and Sung-Jin Choi*[,‡]

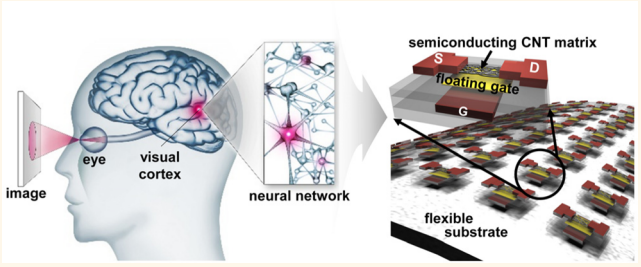
[†]Department of Electrical Engineering, Sejong University, Seoul 05006, Korea

[‡]School of Electrical Engineering, Kookmin University, Seoul 02707, Korea

[§]Mechatronics R&D Center, Samsung Electronics, Gyonggi-do 18448, Korea

Supporting Information

ABSTRACT: **Recent electronic applications require an efficient computing system that can perform data processing with limited energy consumption. Inspired by the massive parallelism of the human brain, a neuromorphic system (hardware neural network) may provide an efficient computing unit to perform such tasks as classification and recognition. However, the implementation of synaptic devices (*i.e.*, the essential building blocks for emulating the functions of biological synapses) remains challenging due to their uncontrollable weight update protocol and corresponding uncertain effects on the operation of the system, which can lead to a bottleneck in the continuous design and optimization. Here, we demonstrate a synaptic transistor based on highly purified, preseparated 99% semiconducting carbon nanotubes, which can provide adjustable weight update linearity and variation margin. The pattern recognition efficacy is validated using a device-to-system level simulation framework. The enlarged margin rather than the linear weight update can enhance the fault tolerance of the recognition system, which improves the recognition accuracy.**



KEYWORDS: *analog switching, carbon nanotube, neuromorphic system, pattern recognition, synaptic transistor, weight update*

Neuromorphic engineering is the design and construction of systems that replicate the capabilities of biological systems and their advantages, such as robustness and power efficiency, by mimicking both the functional and structural characteristics of the biological systems.[1] Although relatively little is known about the principle of information processing in the brain, it is certain that biological neural systems for sensory perception (such as the visual system shown in Figure 1a) are more energy efficient than the common von Neumann architecture of computers.[2] Therefore, inspired by the structure of biological brains,[3] neuromorphic systems have been adopted to develop innovative artificial neural network computing architecture that is adaptive, massively parallel, and fault-tolerant.[4]

A unique feature of neuromorphic systems is that in a massive parallel network, the processing and storing of information can be performed simultaneously by modulating the connection strength of synapses, which is referred to as the synaptic weight.[5] Spikes from the presynaptic neurons can be transmitted through the synapses and generate a membrane potential and thus presynaptic spikes based on the relative strengths of the synapses (*i.e.*, the synaptic weights). These synaptic weights can be modulated by either potentiating or depressing spikes from pre- and postsynaptic neurons following appropriate learning rules, such as spike-timing-dependent plasticity (STDP).[6] Therefore, a key element in the neuromorphic system is the implementation of an ideal synaptic device that can emulate the functionality of biological synapses. Unfortunately, the current lack of a highly scalable and low-power synaptic device is still a major obstacle to realize neuromorphic systems.[7]

To date, several efforts have been made to implement an appropriate synaptic device by exploiting complementary metal-oxide-semiconductor (CMOS) technology and emerging nanoelectronic devices. Although the capability of CMOS circuits is sufficient to capture synaptic functionalities,[8−10] the chip area and power consumption required would be prohibitively large for large-scale integration. To overcome the challenges of CMOS-based approaches,[11] attention has

Received: November 24, 2016
Accepted: February 21, 2017
Published: February 21, 2017

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

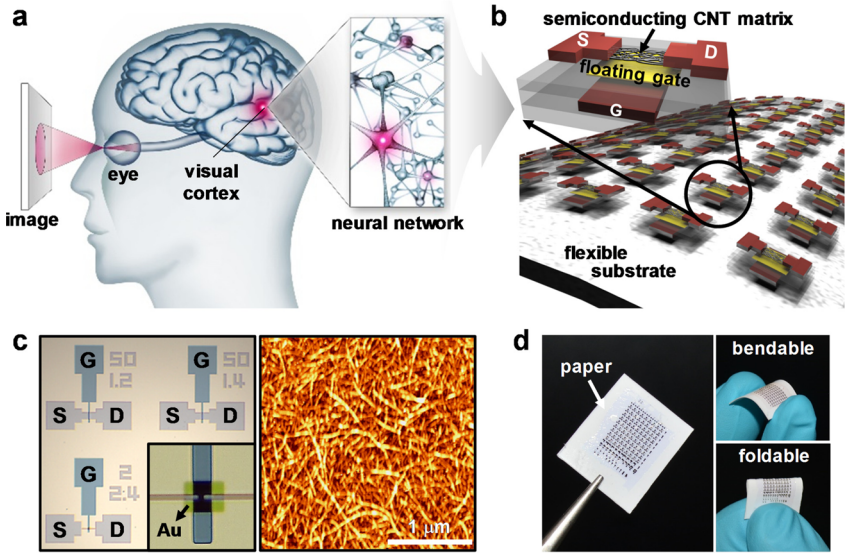


Figure 1. (a) Processing of visual input received through the eye's retina is performed in the so-called visual cortex. The genetic background of an individual's neuronal development and the activity-driven learning process connects neurons in this part of the brain with each other by means of synapses and, in so doing, forms an energy efficient neural network. (b) Schematic of the demonstrated flexible synaptic transistors based on a random matrix of semiconducting CNTs. (c) Microscopy (left) and AFM (right) images of the fabricated CNT synaptic transistors and a random matrix of CNTs, respectively. (d) Image of the fabricated flexible CNT synaptic transistors being transferred onto the paper substrate (photo paper, averaged RMS of 22.5 nm).

recently turned to the attractive characteristics of emerging devices. One of the most promising candidates is two-terminal resistive switching devices (known as memristors). Memristors are nonvolatile analog memory.[12] With memristors, analog conductance states can be maintained over the long-term, and only a minuscule amount of energy is required to modulate distinct states. Synaptic plasticity (*e.g.*, STDP, spike-rate-dependent plasticity or paired pulse facilitation) has been demonstrated using memristors in a number of studies,[13−18] indicating the possibility of establishing next-generation computation paradigms.

However, the sustainability of memristors is still in doubt, particularly with regard to the uncontrollable weight update protocol that is common to all memristor technologies. The specific physical mechanism of the conductance change in most prospective metal-oxide-based memristors, which uses an atomic-scale modulation of oxygen vacancies based on an electro/thermodynamic feedback process,[19] is responsible for the unwanted abrupt conductance change (*i.e.*, the nonlinear response of the resistance to consecutive potentiation or depression spikes) and the limited conductance variation margin.[20] Moreover, the effect of nonlinearity and the limited variation margin to the neuromorphic system operations, such as pattern recognition tasks, has not been analyzed quantitatively,[21−23] which leads to a bottleneck in the continuous design and optimization of synaptic devices.

This study breaks through these issues by demonstrating a synaptic device based on a carbon nanotube (CNT) transistor in which the randomly networked CNTs produced from highly purified, preseparated 99% semiconducting CNT solutions were used as channels in the transistor. We intentionally embedded a thin Au layer as a floating gate (*i.e.*, charge storage layer) inside a gate dielectric to control both the linearity and variation margin of the weight update. In addition, we demonstrate a device- to system-level simulation framework based on a simplified STDP scheme that has the potential for unsupervised online learning and consequent pattern recognition ability in a synaptic transistor array. Therefore, it is expected that our results can provide guidelines for the quantitative design and optimization of synaptic devices and, in particular, the required linearity and variation margin of the weight update to improve the recognition accuracy of the neuromorphic system. In addition, we demonstrate the CNT synaptic transistor on a paper substrate that has great potential to meet the demand for a popular, flexible, foldable, low-cost, mass-producible, disposable, retrievable, and easily processable neuromorphic system.

RESULTS AND DISCUSSION

Transistors based on CNTs have been investigated as synaptic devices in previous studies.[24−27] However, those studies only concentrated on the characterization of a single synaptic device and therefore could not effectively provide guidelines for system-level design, such as how to implement the functional neuromorphic system (*e.g.*, the implementation of a pattern recognition system with the proper learning algorithm). Similarly, in our previous work,[28] a synaptic transistor based on a random matrix of single-walled CNTs was demonstrated to reproduce synaptic functions. Preseparated semiconducting CNTs enabled high uniformity and sustainability of the synaptic transistors with a reliable analog modulation of channel conductance. However, the updates of the synaptic weight (*i.e.*, change in the channel conductance in the devices) were shown to be highly nonlinear for both potentiation and depression; consequently, the weight change was prohibitively large during the first few potentiation/depression pulses and became saturated as the number of applied pulses increased. Moreover, the variation margin of the weight was only a factor of 2, which limits further improvement of the pattern recognition accuracy.[28]

To alleviate these issues, a randomly networked highly purified, preseparated 99% semiconducting CNTs were utilized as a channel in the CNT synaptic transistor, and a thin Au layer

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

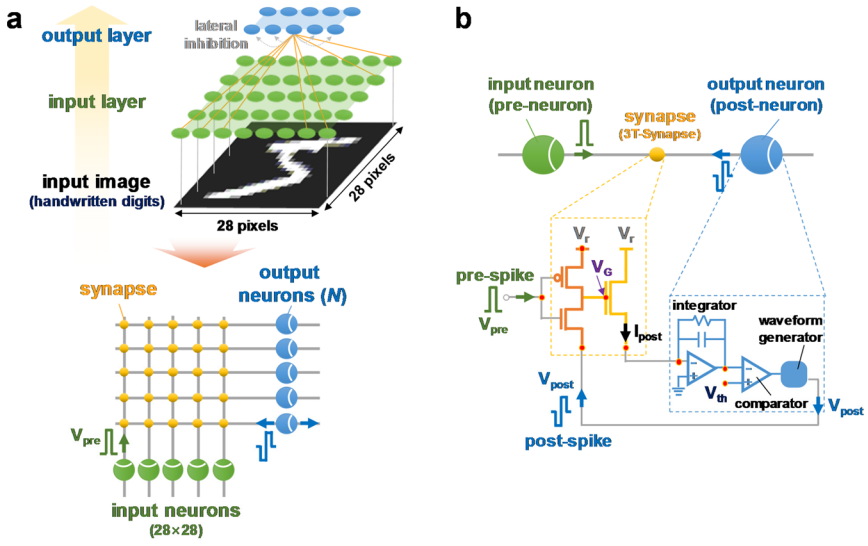


Figure 2. (a) Hierarchical neural network for pattern recognition of 28 × 28 grayscale images consisting of the input and output layers. The input neuron is fully connected to the input image pixel in a one-to-one manner. The synaptic devices are located at the junctions between the input and output neurons. (b) Neuron circuit with each synapse. The neuron circuit is composed of a leaky integrator, comparator, and waveform generator. The synapse (*i.e.*, 3T-Synapse) consists of three CNT transistors: an initial p-channel CNT transistor, PEI-doped n-type CNT transistor, and CNT transistor with a thin Au layer as a floating gate. The former two transistors were connected to form the inverter (orange), and the latter CNT transistor was used to implement the synaptic transistor (yellow). Using the inverter circuit, a postsynaptic spike (V_{post}) was selectively applied to the synaptic transistor, *i.e.*, the gate electrode of the synaptic transistor (V_G), only when presynaptic spike (V_{pre}) is the input.

was embedded as a floating gate in this work, as shown in Figure 1b,c. In this structure, the channel conductance can be modulated according to the charges stored at the thin Au floating gate, enabling the variation margin to be enlarged due to the additional charge storage node. In particular, the amount of charge at the floating gate can be adjusted more accurately by designing the amplitude and duration time of the gate voltage pulse, which enables the weight update linearity and variation margin to be controlled as desired. Additionally, a CNT is highly suitable for flexible electronic materials because of its high bendability and chemical stability, which originates from the excellent material properties.[29,30] Flexible devices are becoming increasingly significant in a wide variety of novel applications that enable integration with soft, curvilinear, and even time-variant substrates. We believe that a computing system with high flexibility could be an essential part of flexible electronics for data processing. Therefore, the CNT synaptic transistors proposed in this work were fabricated on a paper substrate that offers many advantages, including flexibility, inexpensiveness, lightweight, disposability, and recyclability (Figure 1d), as a proof-of-concept for the implementation of a flexible paper neuromorphic system (the detailed fabrication process and device performance are described in the Methods section and Supporting Information Note 1, respectively).

Figure 2a shows a neural network architecture for our conceived pattern recognition system. With the crossbar layout, the system consists of an input layer followed by an output layer. The pixels in the image housing the definite patterns to be recognized constitute the input layer. Each input neuron is connected with one pixel of the image; a total of 28 × 28 input neurons emit presynaptic spikes (V_{pre}), wherein the timing of the presynaptic spikes represents the analog information on the pixel intensities (the detail timing information on the presynaptic spike will be discussed later). Subsequently, presynaptic spikes from the input neurons can trigger multiple CNT synaptic transistors simultaneously, and postsynaptic currents (I_{post}) determined by the channel conductance of each CNT synaptic transistor are collected and accumulated at an output neuron. If the accumulated postsynaptic current level is greater than a given threshold value (V_{th}), one output neuron fires a postsynaptic spike (V_{post}); then, the synaptic weight can be modulated to any analog state according to the correlation of the pre- and postsynaptic spikes. Additionally, when an output neuron fires a spike, it sends inhibitory signals to the other output neurons of the output layer that prevent the other output neurons from firing during the refractory time. Thus, lateral inhibition promotes competitive learning and effectively prevents all of the output neurons from learning similar patterns, which can afford a winner-takes-all mechanism.[31]

In terms of circuit-level design, a synaptic function is emulated by the configuration of three CNT transistors (3T-Synapse), as shown in Figure 2b, which is the combination of the inverter (shown in orange) and the synaptic transistor (shown in yellow). The role of an inverter is a selector; the spike timing correlation between the pre- and postsynaptic spikes is converted into various pulse amplitudes through the selector. This pulse is applied to the gate electrode of the synaptic transistor (V_G) and enables the modulation of the channel conductance. In addition, as noted above, one output neuron is connected with 28 × 28 synaptic transistors in parallel; each synaptic transistor generates the postsynaptic current (I_{post}) based on the channel conductance. The total sum of the postsynaptic currents is accumulated by a leaky integrator, and finally, the output neuron generates a postsynaptic spike through a waveform generator depending on the comparison between the integrated postsynaptic current level and V_{th}. In the proposed neural network, there is no a sneak path issue in a crossbar array due to the use of three-terminal synaptic devices. This issue typically limits the practicability of crossbar architecture[32] and requires complex

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

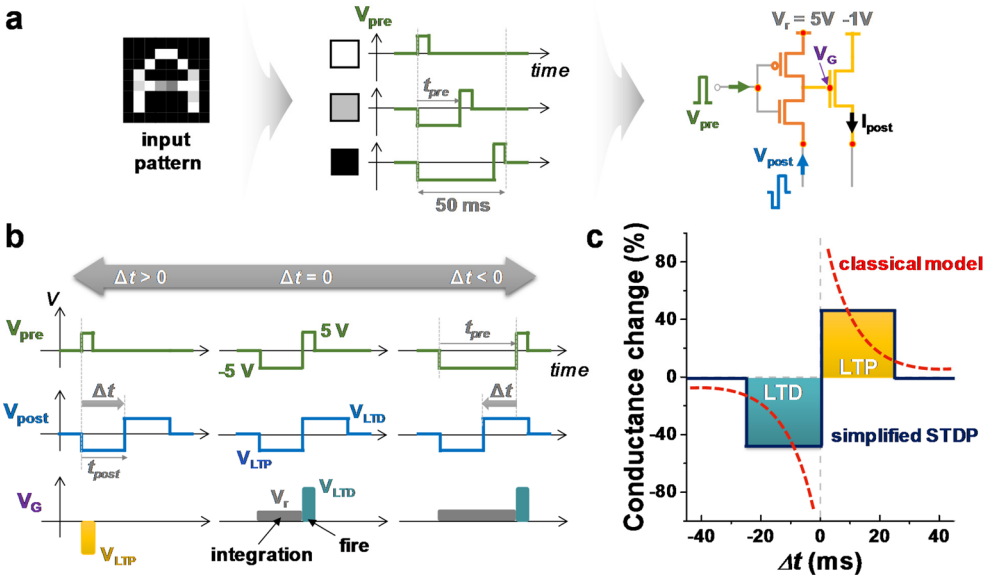


Figure 3. (a) Pulse timing for presynaptic spikes. Each pixel's intensity is converted into the timing of V_{pre}. (b) Pulses for a simplified STDP (voltage pulses as functions of time). The net applied voltage to the synaptic transistor (V_G) is determined by the timing correlation between V_{pre} and V_{post}. The polarity of V_G is determined by Δt (= $t_{post} - t_{pre}$); V_G is negative when $\Delta t > 0$ for potentiation and positive when $\Delta t < 0$ for depression. (c) Developed simplified STDP rule compared with the standard biological STDP rule.

solutions, such as complementary resistive switches[33] or the use of nonlinear devices[34] in the case of common synaptic devices based on two-terminal resistive switches.

Subsequently, we implement the simplified STDP scheme[28,35] by designing an appropriate timing correlation between the pre- and postsynaptic spikes to facilitate the learning/recognition operation of the proposed neural network system. By assuming that an image sensor senses the external pattern, the sensed information is converted into presynaptic spikes with timings. We define a presynaptic spike timing (t_{pre}) from 0 to 50 ms for 256-shade greyscale pixel intensities, as shown in Figure 3a. An early voltage spike timing of 0 ms correlates to a white pixel, and a late voltage spike timing of 50 ms correlates to a black pixel. The presynaptic spike (V_{pre}) is a succession of a negative bias and positive bias; when a negative bias part of V_{pre} is applied to the 3T-Synapse, the p-channel transistor, which is a pull-up network in the inverter, passes the certain read bias (V_r) to the gate electrode of the synaptic transistor ($V_G = V_r$), which generates the postsynaptic current (I_{post}). Conversely, when a positive bias part of V_{pre} is applied, the n-channel transistor, which is a pull-down in the inverter, passes V_{post} ($V_G = V_{post}$), which leads to the channel conductance modulation at the synaptic transistor. Similarly, the postsynaptic spike is composed of consecutive negative and positive biases, as shown in Figure 3b. If the presynaptic spike arrives at the 3T-Syanpse earlier than the postsynaptic spike (*i.e.*, $t_{post} - t_{pre} = \Delta t > 0$), then the polarity of V_G is negative (= V_{LTP}), which increases the channel conductance (which is referred to as a long-term potentiation (LTP)). Conversely, when the presynaptic spike is later than the postsynaptic spike ($\Delta t < 0$), the channel conductance of the synaptic transistor is decreased (long-term depression (LTD)) by positive V_G (= V_{LTD}) (the channel conductance change depending on the V_G polarity is discussed in Supporting Information Note 2). Therefore, the timing correlation between pre- and postsynaptic spikes is converted into various pulses (V_r, V_{LTP}, or V_{LTD}) applied to the synaptic transistor; here, V_{LTP} and V_{LTD} modulate the synaptic weight for the learning of patterns, whereas V_r only generates the postsynaptic current to determine the occurrence of firing at the output neuron for the recognizing of patterns. This simple learning rule, which is easily implemented with three CNT transistors (3T-Synapse), is the ground for the weight update protocol. Compared with the purely bioinspired and more complex scheme introduced in previous studies,[13−18] no delay matching is necessary between the pre- and postsynaptic spikes; thus, this simplified STDP learning rule (Figure 3c) should make driving circuitry considerably easier to design. Moreover, a peripheral driving circuitry that is equal to those of 3T-Synpase can also be implemented by CNT transistor technology, enabling a considerably easier fabrication process for high-density integration.

In the following, we present the synaptic learning rule from 3T-Synapse. First, complementary n- and p-channel CNT transistors were developed to implement the aforementioned CMOS inverter circuit. In general, the CNT transistor shows initial p-type behavior under ambient conditions; thus, the n-type CNT transistor could be achieved by doping electron-donating groups.[36] Here, n-type doping was accomplished by coating the prepared polyethylenimine (PEI) solution dissolved in methanol (50 vol %) on a desired CNT network channel (see the Methods section). With n-type doping, the channel current at positive V_{GS} values began to show a clear increase, and the complete type conversion was achieved with a similar on/off ratio as that of the initial p-channel CNT transistor (Figure 4a). Such a conversion of CNT transistors by PEI doping results from the shift of the Fermi level toward the conduction band due to electron donation (see Supporting Information Note 3 for a detailed explanation). The additional Au layer as a floating gate was not embedded into the inverter, *i.e.*, both the n- and p-channel CNT transistors, for stable inverter operation without significant hysteresis. In addition, highly purified, preseparated, 99% semiconducting CNTs were used for the formation of the CNT network channels, which gave rise to uniformly distributed electrical performances, such as the threshold voltage, on/off current ratio, and mobility

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

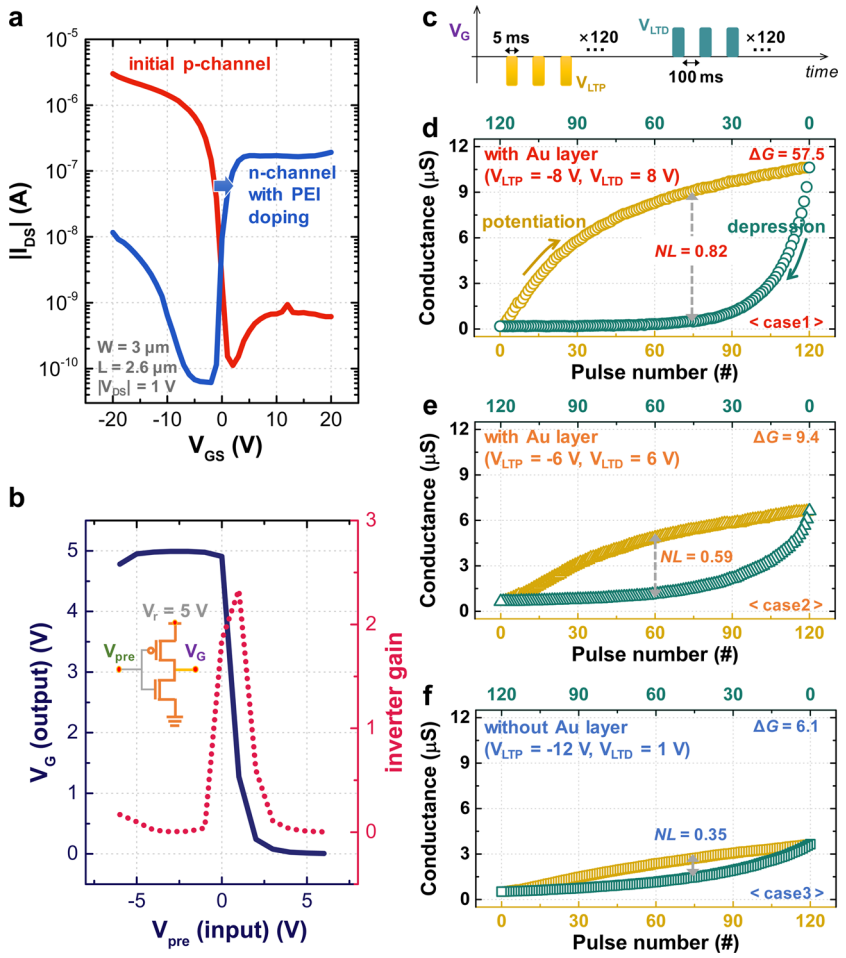


Figure 4. (a) Transfer characteristics for n- and p-channel CNT transistors for the inverter. (b) Voltage transfer curves of an inverter consisting of CMOS CNT transistors showing a voltage gain of approximately 2.3 at an operating voltage (V_{DD}) of 5 V. (c) Schematics of the applied pulse trains used to measure the analog channel conductance modulation. Each pulse train consists of 120 potentiation or depression pulses applied to the gate (V_{LTP} and V_{LTD} for 5 ms) followed by small, nonperturbative read voltage pulses (5 V for 100 ms) within the intervals. Measured analog conductance-switching behaviors in three different cases: (d) Case 1: the amplitudes of V_{LTP} and V_{LTP} are greater than other cases; thus, *NL* is the highest and ΔG is the largest. (e) Case 2: the amplitudes of V_{LTP} and V_{LTP} are smaller than in case 1; thus, *NL* and ΔG are lower. (f) Case 3: if the CNT transistor without the Au floating gate is used for the synaptic transistor, *NL* and ΔG are considerably smaller than in the other cases due to the limited charge storage space.

(Supporting Information Note 1). As a result, the operation of the complementary inverter was achieved by combining n- and p-channel CNT transistors, as shown in Figure 4b. With V_r = 5 V, clear output on and off states were observed in which these output characteristics were realized with a small leakage current at the off state of n- and p-channel CNT transistors.

As discussed above, the inverter selects the signal V_G applied to the gate electrode of the CNT synaptic transistor according to the timing correlation of the pre- and postsynaptic spikes; accordingly, V_G is either V_r, V_{LTP}, or V_{LTD}. When $V_G = V_r$, V_r (= 5 V) is not sufficient to modulate the channel conductance of the CNT synaptic transistor, it can generate a channel current (*i.e.*, postsynaptic current (I_{post})) depending on the channel conductance. Therefore, when $V_G = V_r$, only I_{post} is generated and accumulates due to the leaky integrator. Conversely, when $V_G = V_{LTP}$ or V_{LTD}, the channel conductance of the CNT synaptic transistor is modulated, and this modulation is caused by the tunneling process of carriers into the specially embedded Au floating gate (the Au layer is embedded only into the synaptic transistor in three CNT transistors in 3T-Synapse). We identified the analog channel conductance-switching behavior in the CNT synaptic transistor. Figure 4c shows the schematics of the pulse trains used for the measurement. Each pulse train consists of 120 pulses (negative polarity pulse for potentiation and positive polarity pulse for depression), followed by nonperturbative read voltage pulses at 1 V within the intervals. Although the synaptic update responses of different devices were reasonably uniform, the weight updates were highly nonlinear for both potentiation and depression, as shown in Figure 4d. The change in the channel conductance was more dramatic during the first few potentiation/depression pulses and became saturated as the number of pulses increased. Every training pulse resulted in a different response in the weight update depending on the current weight state, and the cumulative effect on the weight update does not follow a simple linear relation, which is attributed to the nonlinearity (*NL*) of the weight update. Here, *NL* and ΔG were defined quantitatively as[37]

$$NL = \frac{\max|G_p(n) - G_d(n)|}{G_p(120) - G_p(1)} \quad \text{for } n = 1-120$$

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

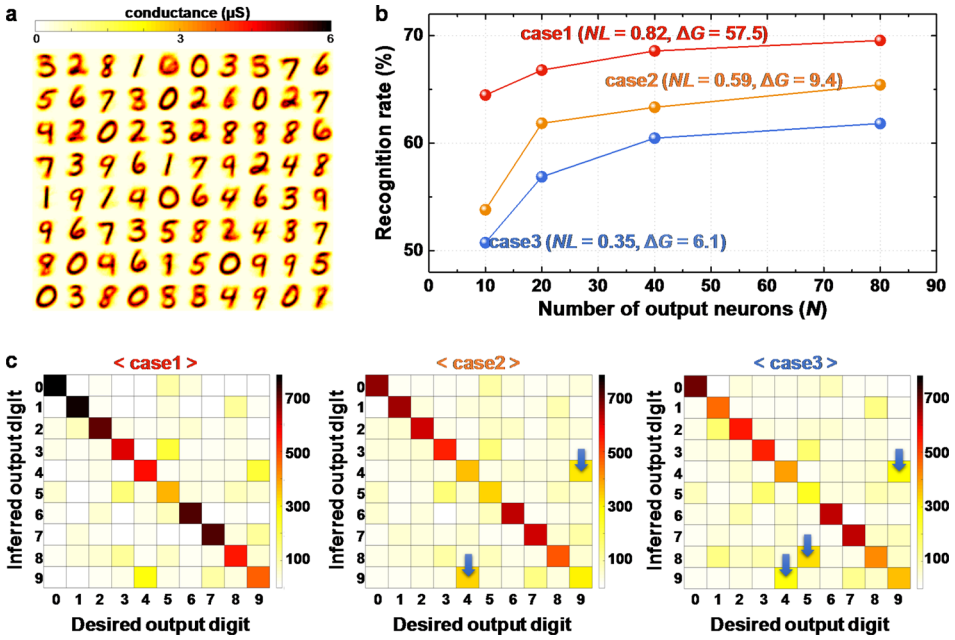


Figure 5. (a) Rearranged weights (from 784 to 28 × 28) of the connections from the input to output neurons for a network with 80 output neurons. (b) Performance as a function of the number of output neurons. Each dot shows the recognition rate for a certain network size as an average over 10 digits of the entire MNIST test set. (c) Average confusion matrix of the testing results over 10 digits of the 10,000 MNIST test set. High values along the identity indicate correct identification, whereas high values anywhere else indicate confusion between two digits, for example, the digits 4 and 9.

$$\Delta G = (G_p(120) - G_p(1))/G_p(1)$$

where $G_p(n)$ and $G_d(n)$ are the conductance values after the nth potentiation pulse and nth depression pulse, respectively. NL should be zero for a completely linear update. Interestingly, in the case of the proposed CNT synaptic transistor with a thin Au layer (floating gate), the channel conductance is modulated depending on the number of trapped charges at the Au floating gate, which can be adjusted by designing the amplitude and time of the gate voltage pulse; thus, the weight update nonlinearity (NL) and total variation margin (ΔG) are controllable. In our measurement, when V_{LTP} and V_{LTD} were −8 V and +8 V, respectively, with a fixed pulse width (5 ms), NL and ΔG were quite high (NL = 0.82, ΔG = 47.0), as shown in Figure 4d (case 1). These high NL and ΔG values are due to the abrupt carrier injection into the Au floating gate. Furthermore, when the amplitudes of V_{LTP} and V_{LTD} are reduced (Figure 4e, case 2), the conductance change becomes more gradual due to the alleviation of the carrier injection process, which enables smaller NL and ΔG values. Moreover, the CNT synaptic transistor without the Au floating gate provides the smallest NL and ΔG values (Figure 4f, case 3). Interestingly, compared with more complex schemes introduced in previous studies for two-terminal resistive switches (*i.e.*, nonidentical training pulses with a state-dependent pulse width and the amplitude have been required to control the weight update behavior, NL and ΔG),[21] with the thin Au layer in our CNT synaptic transistors, NL and ΔG can be easily controlled by simply adjusting the amplitude of the pulse, which will decrease the complexity of the peripheral circuit designs.

However, although NL and ΔG can be adjusted as desired, it is unclear what NL and ΔG values are appropriate for the reliable operation of a neuromorphic system. In other words, an investigation on the impact of high/low NL and ΔG is necessary, as is particularly evident when investigating the effect on the pattern recognition accuracy. In recent pattern recognition studies based on implemented synaptic device arrays,[21−23,38] high NL has been shown to degrade the recognition accuracy. Nevertheless, no study has performed a quantitative analysis on how a correlated NL and ΔG affect the accuracy. Therefore, in the following, we represent a device- to system-level simulation by using the implemented simplified STDP scheme and demonstrate how different NL and ΔG values affect the learning and recognition process. The detailed simulation procedure, parameters, and model used in this study are described in Supporting Information Note 4.[28] Briefly, to demonstrate pattern recognition in a CNT synaptic transistor array, we use the widely studied case of handwritten number recognition using the MNIST database, which consists of handwritten numbers that are 28 × 28 pixels.[39] We input the full MNIST training database, which consists of 60,000 digits, into the system to guide the learning process. Each input neuron is connected with one pixel of the image; thus, a total of 28 × 28 input neurons emit presynaptic spikes such that their timings are proportional to the pixel intensity. Input presynaptic spikes generate postsynaptic currents based on the synaptic weight of each synaptic transistor and are integrated by the output neurons. Then, the one output neuron whose integrated postsynaptic current is the highest fires postsynaptic spikes; correlated pre- and postsynaptic spikes result in channel-conductance potentiation or depression (*i.e.*, the learning phase). After completing the learning process, the network is tested on the MNIST test database, which consists of 10,000 digits that were not available during training (*i.e.*, the recognition phase).

Figure 5a shows the final simulated conductance states of the CNT synaptic transistors connecting the input neurons to each of the output neurons. The synaptic weights that were randomly initialized eventually learned to encode the input

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

patterns. Figure 5b shows the recognition rate for the test data set as a function of the number of output neurons (N) and that the classification accuracy (*i.e.*, recognition rate) can be improved by increasing the number of output neurons (N); with 80 output neurons, the recognition rate reaches 60−70%. Additionally, Figure 5c shows the detail of the misclassification in three different cases; it is the average confusion matrix over 10 digits of the MNIST test set, *i.e.*, every single classification of the test inputs belongs to one of the 10 × 10 tiles, and its position is determined by the actual digit and inferred digit. Given a recognition rate of approximately 70% in case 1, the majority of the inputs are on the identity that corresponds to correct classification. Conversely, more misclassifications occurred in case 3 given the lower recognition rate, where the most common confusions were that 4 was identified as 9, 5 was identified as 8, and 9 was identified as 4. Interestingly, case 1 (NL is the highest and ΔG is the largest) always exhibits a better recognition rate than the other cases regardless of the number of output neurons, as shown in Figure 5b. This result is distinctive from the results of previous studies,[21−23,38] in which a smaller NL could improve the accuracy. In contrast, in our simulation results, a larger ΔG rather than a smaller NL is the determining factor to improve the recognition accuracy.

To understand the nature of the recognition accuracy affected by NL and ΔG, we re-investigate the influence of NL and ΔG on the recognition rate according to the number in the learning phase. Figure 6 shows the simulation results with

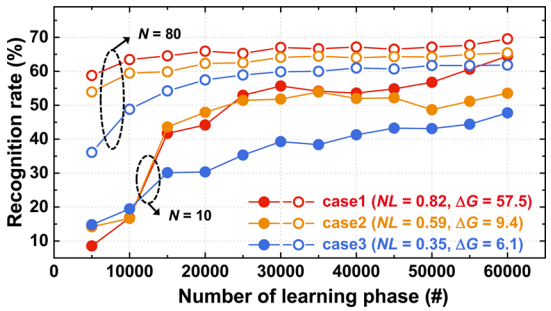


Figure 6. Recognition rate as a function of the number of learning phases.

different cases and shows again that the recognition rate is significantly affected by ΔG regardless of the number in the learning phase and the number of output neurons; as noted above, a larger ΔG instead of a smaller NL produces a better recognition rate. Although a rapid and unstable evolution of the recognition rate was obtained when N = 10 (depicted by the filled circles) compared with N = 80, the tendency of the recognition rate is predominately determined by ΔG instead of NL. This opposite result compared to previous studies is caused by the fact that the conductance variation margin of common synaptic devices based on two-terminal resistive switches is below 10 (*i.e.*, ΔG < 10).[21−23,38] With this small ΔG value, NL was the sole controlling factor; thus, previous studies have concluded that a smaller NL produces a better recognition rate. In contrast, in our CNT synaptic transistors, ΔG is 57.5 in case 1, and it can be further enhanced by increasing the amplitude of V_{LTP} or V_{LTD}. This larger ΔG provides more analog conductance states to store information on the input pattern to be more clearly distinguished, which leads to a better distinction between the previously learned pattern and other test patterns. Although a smaller NL can improve the recognition rate under a fixed ΔG, a larger ΔG is the more dominant factor in the pattern recognition procedure. Therefore, when improving the pattern recognition accuracy in a neuromorphic system, the larger variation margin of the weight states is more important than the linearity of weight update; this conclusion has not been drawn in previous studies.

CONCLUSION

In summary, we have experimentally demonstrated a synaptic device based on three CNT transistors (3T-Synapse) constructed from highly purified 99% semiconducting CNT solutions with reliable, analog, conductance-modulated behavior. Specifically, the embedded Au floating gate of the CNT synaptic transistor enables the synaptic-weight plasticity to be encoded by adjusting the amount of carrier injection, which is simpler and more accurate than previous approaches based on two-terminal resistive switches. We fabricated the CNT synaptic transistors on a paper substrate, demonstrating the feasibility of a flexible, inexpensive, lightweight, disposable, and recyclable neuromorphic system. In addition, the developed complementary n- and p-channel CNT transistors provide the CMOS inverter circuit, which emulates the simplified STDP mechanism from the timing correlation between the pre- and postsynaptic spikes. Additionally, a simplified STDP scheme was used to simulate the pattern recognition task at a system level, where the 3T-Synapses associated with peripheral neuron circuits could perform unsupervised learning. The larger margin of conductance modulation in the CNT synaptic transistor enables a better recognition accuracy; this simulation result is an important step toward effective analog hardware implementation for more complex neuromorphic systems.

Although the proposed 3T-Synapse, which is based on highly purified, preseparated 99% semiconducting CNTs, requires greater energy consumption to change the analog conductance states than previous two-terminal resistive switches or existing silicon-based floating gate memory, the existing two-terminal resistive switches and silicon-based technology are now facing reliability issues and the physical limit of device scaling, respectively. On the other hand, the CNT has been regarded as the next-generation material with excellent material properties for high-performance, low-power electronics; hence, the CNT-based synaptic device has been reported by other groups,[24−27] including our previous work.[28] In this work, we employed the highly purified, preseparated 99% semiconducting enriched CNTs-based synaptic transistor with excellent electrical performance, in particular, high on/off current ratio resulting in the large analog conductance changes. In addition, a peripheral driving circuitry (neuronal circuit) as well as synaptic devices can also be monolithically implemented using the equivalent CNT transistor technology, which enables a considerably easier fabrication process with low cost. Importantly, the integration to three-dimensionally stacked high-density array might be much easier than silicon-based conventional floating gate memory;[40] it will be an important merit toward the effective analogue hardware implementation of more complex neuromorphic networks.

METHODS

Fabrication of CNT Transistors and Transfer to the Paper Substrate. To transfer the device onto the paper substrate, CNT transistors were initially fabricated on highly p-doped rigid silicon substrates with a thermally grown 50 nm-thick SiO_2 layer. First, the copper (Cu) (300 nm) and SiO_x (300 nm) layers were sequentially

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

deposited on the substrate using evaporation and a plasma-enhanced chemical vapor deposition process, respectively. The layers served as a sacrificial layer to detach the CNT transistors from the rigid donor substrate using a water-assisted transfer technique.[36] We used the local back-gate structure for efficient local modulation of the channels in the CNT transistors. To form the local back-gate, the palladium (Pd) layer was first deposited and subsequently patterned using evaporation and a lift-off process, respectively. Next, a 50 nm-thick SiO_x layer, 10 nm-thick Au layer, and 20 nm-thick SiO_x layer were deposited sequentially. The thin Au layer served as a floating gate for charge storage. Then, the top surface of the SiO_x layer was functionalized with a 0.1 g/mL poly-L-lysine solution to form an amineterminated layer, which acted as an effective adhesion layer for the deposition of the CNTs. Subsequently, the CNT network channel was formed by immersing the chip into a 0.01 mg/mL 99% semiconducting CNT solution (NanoIntegris, Inc.) for several hours, followed by a thorough rinse with isopropanol and DI water. Subsequently, the source/drain electrodes consisting of Ti and Pd layers (each 2 and 40 nm, respectively) were deposited and patterned using conventional thermal evaporation and a lift-off process, respectively. Finally, additional photolithography and oxygen plasma steps were conducted to remove unwanted electrical paths, which isolated the devices from one another.

We used the water-assisted transfer printing technique to transfer the fabricated CNT transistors on the rigid substrate to the paper substrate. First, a thermal release tape (TRT) was attached to the fabricated CNT transistors as a temporary holder. Next, the entire structure was soaked in deionized water at room temperature, and an edge of the TRT was peeled off to initiate water penetration. Within several minutes, the entire structure was detached from the donor substrate. A Cu etchant ($FeCl_3$) was then used to eliminate the Cu layer on the backside of the detached substrate. At the end of the process, the TRT holding the structure was pasted onto the photo paper without any surface treatment of the paper surface.

Electron Doping for the n-Channel CNT Transistor. We used a branched PEI polymer (average M_W: 800, Sigma-Aldrich) to produce n-channel transistors. First, a typical branched PEI polymer with the chemical formula $H(NHCH_2CH_2)_nNH_2$ was dissolved in methanol (50 vol %) and then spin-coated on the top surface of the devices using low-rpm spin-coating. Next, the baking process was performed at 65 °C to evaporate the methanol. Finally, the remaining PEI on the CNTs was removed by rinsing again with methanol. The PEI was irreversibly adsorbed on the CNT network *via* thorough rinsing with a solvent (methanol). A relatively high concentration of PEI (50 vol %) was selected due to the high n-channel current. All experiments were conducted under ambient conditions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.6b07894.

Additional discussions about (1) the electrical properties of CNT synaptic transistors and their variability, (2) the mechanism of the channel conductance modulation in the CNT synaptic transistors, (3) doping in the CNT transistors, and (4) details concerning the pattern recognition simulation procedures (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: sjchoiee@kookmin.ac.kr.

ORCID

Sung-Jin Choi: 0000-0003-1301-2847

Author Contributions

‖These authors equally contributed to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by Nano·Material Technology Development Program (NRF-2016M3A7B4910430) and Basic Science Research Program (NRF-2016R1D1A1B03930162, 2016R1A2B4011366, and 2016R1A5A1012966) through the National Research Foundation of Korea funded by the Ministry of Science, ICT, and Future Planning. This work was partially supported by the Future Semiconductor Device Technology Development Program (grant 10067739) funded by MOTIE (Ministry of Trade, Industry, and Energy) and KSRC (Korea Semiconductor Research Consortium).

REFERENCES

(1) Mead, C. Neuromorphic Electronic Systems. *Proc. IEEE* **1990**, *78*, 1629−1636.

(2) Javed, F.; He, Q.; Davidson, L. E.; Thornton, J. C.; Albu, J.; Boxt, L.; Krasnow, N.; Elia, M.; Kang, P.; Heshka, S.; Gallagher, D. Brain and High Metabolic Rate Organ Mass: Contributions to Resting Energy Expenditure beyond Fat-Free Mass. *Am. J. Clin. Nutr.* **2010**, *91*, 907−912.

(3) Seung, H. S. Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission. *Neuron* **2003**, *40*, 1063−1073.

(4) Indiveri, G.; Horiuchi, T. K. Frontiers in Neuromorphic Engineering. *Front. Neurosci.* **2011**, *5*, 118.

(5) Abbott, L. F.; Regehr, W. G. Synaptic Computation. *Nature* **2004**, *431*, 796−803.

(6) Bi, G. Q.; Poo, M. M. Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *J. Neurosci.* **1998**, *18*, 10464−10472.

(7) Kuzum, D.; Yu, S.; Philip Wong, H.-S. Synaptic Electronics: Materials, Devices and Applications. *Nanotechnology* **2013**, *24*, 382001.

(8) Rasche, C.; Hahnloser, R. H. R. Silicon Synaptic Depression. *Biol. Cybern.* **2001**, *84*, 57−62.

(9) Diorio, C.; Hsu, D.; Figueroa, M. Adaptive CMOS: From Biological Inspiration to Systems-on-a-Chip. *Proc. IEEE* **2002**, *90*, 345−357.

(10) Bartolozzi, C.; Indiveri, G. Synaptic Dynamics in Analog VLSI. *Neural Comput.* **2007**, *19*, 2581−2603.

(11) Rajendran, B.; Liu, Y.; Seo, J. S.; Gopalakrishnan, K.; Chang, L.; Friedman, D. J.; Ritter, M. B. Specifications of Nanoscale Devices and Circuits for Neuromorphic Computational Systems. *IEEE Trans. Electron Devices* **2013**, *60*, 246−253.

(12) Chua, L. Resistance Switching Memories Are Memristors. *Appl. Phys. A: Mater. Sci. Process.* **2011**, *102*, 765−783.

(13) Zamarreño-Ramos, C.; Camuñas-Mesa, L. A.; Perez-Carrasco, J. A.; Masquelier, T.; Serrano-Gotarredona, T.; Linares-Barranco, B. On Spike-Timing-Dependent-Plasticity, Memristive Devices, and Building a Self-Learning Visual Cortex. *Front. Neurosci.* **2011**, *5*, 26.

(14) Serrano-Gotarredona, T.; Masquelier, T.; Prodromakis, T.; Indiveri, G.; Linares-Barranco, B. STDP and sTDP Variations with Memristors for Spiking Neuromorphic Learning Systems. *Front. Neurosci.* **2013**, *7*, 2.

(15) Jo, S. H.; Chang, T.; Ebong, I.; Bhadviya, B. B.; Mazumder, P.; Lu, W. Nanoscale Memristor Device as Synapse in Neuromorphic Systems. *Nano Lett.* **2010**, *10*, 1297−1301.

(16) Yu, S.; Wu, Y.; Jeyasingh, R.; Kuzum, D.; Wong, H. S. P. An Electronic Synapse Device Based on Metal Oxide Resistive Switching Memory for Neuromorphic Computation. *IEEE Trans. Electron Devices* **2011**, *58*, 2729−2737.

(17) Kim, S.; Du, C.; Sheridan, P.; Ma, W.; Choi, S.; Lu, W. D. Experimental Demonstration of a Second-Order Memristor and Its

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX

Ability to Biorealistically Implement Synaptic Plasticity. *Nano Lett.* **2015**, *15*, 2203−2211.
(18) Du, C.; Ma, W.; Chang, T.; Sheridan, P.; Lu, W. D. Biorealistic Implementation of Synaptic Functions with Oxide Memristors through Internal Ionic Dynamics. *Adv. Funct. Mater.* **2015**, *25*, 4290−4299.
(19) Ielmini, D. Modeling the Universal Set/reset Characteristics of Bipolar RRAM by Field- and Temperature-Driven Filament Growth. *IEEE Trans. Electron Devices* **2011**, *58*, 4309−4317.
(20) Kim, S.; Choi, S.; Lu, W. Comprehensive Physical Model of Dynamic Resistive Switching in an Oxide Memristor. *ACS Nano* **2014**, *8*, 2369−2376.
(21) Park, S.; Sheri, A.; Kim, J.; Noh, J.; Jang, J.; Jeon, M.; Lee, B.; Lee, B. R.; Lee, B. H.; Hwang, H. Neuromorphic Speech Systems Using Advanced ReRAM-Based Synapse. Proceedings from the *International Electron Devices Meeting, IEDM*, Washington, DC, December 9−11, 2013; IEEE: New York, 2013; pp 25.6.1−25.6.4.
(22) Burr, G. W.; Shelby, R. M.; di Nolfo, C.; Jang, J. W.; Shenoy, R. S.; Narayanan, P.; Virwani, K.; Giacometti, E. U.; Kurdi, B.; Hwang, H. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165,000 Synapses), Using Phase-Change Memory as the Synaptic Weight Element. Proceedings from the *2014 IEEE International Electron Devices Meeting*, San Francisco, CA, December 15−17, 2014; IEEE: New York, 2014; pp 29.5.1−29.5.4.
(23) Yu, S.; Chen, P. Y.; Cao, Y.; Xia, L.; Wang, Y.; Wu, H. Scaling-up Resistive Synaptic Arrays for Neuro-Inspired Architecture: Challenges and Prospect. Proceedings from the *International Electron Devices Meeting, IEDM*, San Francisco, CA, December 3−7, 2016; IEEE: New York, 2016; pp 17.3.1−17.3.4.
(24) Shen, A. M.; Chen, C.-L.; Kim, K.; Cho, B.; Tudor, A.; Chen, Y. Analog Neuromorphic Module Based on Carbon Nanotube Synapses. *ACS Nano* **2013**, *7*, 6117−6122.
(25) Kim, K.; Chen, C. L.; Truong, Q.; Shen, A. M.; Chen, Y. A Carbon Nanotube Synapse with Dynamic Logic and Learning. *Adv. Mater.* **2013**, *25*, 1693−1698.
(26) Gacem, K.; Retrouvey, J.-M.; Chabi, D.; Filoramo, A.; Zhao, W.; Klein, J.-O.; Derycke, V. Neuromorphic Function Learning with Carbon Nanotube Based Synapses. *Nanotechnology* **2013**, *24*, 384013.
(27) Chen, C.-L.; Kim, K.; Truong, Q.; Shen, A.; Li, Z.; Chen, Y. A Spiking Neuron Circuit Based on a Carbon Nanotube Transistor. *Nanotechnology* **2012**, *23*, 275202.
(28) Kim, S.; Yoon, J.; Kim, H. D.; Choi, S. J. Carbon Nanotube Synaptic Transistor Network for Pattern Recognition. *ACS Appl. Mater. Interfaces* **2015**, *7*, 25479−25486.
(29) Artukovic, E.; Kaempgen, M.; Hecht, D. S.; Roth, S.; Grüner, G. Transparent and Flexible Carbon Nanotube Transistors. *Nano Lett.* **2005**, *5*, 757−760.
(30) Sun, D.; Timmermans, M. Y.; Tian, Y.; Nasibulin, A. G.; Kauppinen, E. I.; Kishimoto, S.; Mizutani, T.; Ohno, Y. Flexible High-Performance Carbon Nanotube Integrated Circuits. *Nat. Nanotechnol.* **2011**, *6*, 156−161.
(31) Maass, W. On the Computational Power of Winner-Take-All. *Neural Comput.* **2000**, *12*, 2519−2535.
(32) Liang, J.; Wong, H. S. P. Cross-Point Memory Array without Cell Selectors-Device Characteristics and Data Storage Pattern Dependencies. *IEEE Trans. Electron Devices* **2010**, *57*, 2531−2538.
(33) Linn, E.; Rosezin, R.; Kügeler, C.; Waser, R. Complementary Resistive Switches for Passive Nanocrossbar Memories. *Nat. Mater.* **2010**, *9*, 403−406.
(34) Joshua Yang, J.; Zhang, M. X.; Pickett, M. D.; Miao, F.; Paul Strachan, J.; Li, W.-D.; Yi, W.; Ohlberg, D. A. A.; Joon Choi, B.; Wu, W.; Nickel, J. H.; Modeiros-Ribeiro, G.; Williams, R. S. Engineering Nonlinearity into Memristors for Passive Crossbar Applications. *Appl. Phys. Lett.* **2012**, *100*, 113501.
(35) Querlioz, D.; Bichler, O.; Dollfus, P.; Gamrat, C. Immunity to Device Variations in a Spiking Neural Network with Memristive Nanodevices. *IEEE Trans. Nanotechnol.* **2013**, *12*, 288−295.
(36) Derycke, V.; Martel, R.; Appenzeller, J.; Avouris, P. Carbon Nanotube Inter- and Intramolecular Logic Gates. *Nano Lett.* **2001**, *1*, 453−456.
(37) Wang, I.-T.; Chang, C.-C.; Chiu, L.-W.; Chou, T.; Hou, T.-H. 3D Ta/TaO_x/TiO_2/Ti Synaptic Array and Linearity Tuning of Weight Update for Hardware Neural Network Applications. *Nanotechnology* **2016**, *27*, 365204.
(38) Chen, P. Y.; Lin, B.; Wang, I. T.; Hou, T. H.; Ye, J.; Vrudhula, S.; Seo, J. S.; Cao, Y.; Yu, S. Mitigating Effects of Non-Ideal Synaptic Device Characteristics for on-Chip Learning. Proceedings from the *2015 IEEE/ACM International Conference on Computer-Aided Design, ICCAD 2015*, Austin, TX, November 2−6, 2015; IEEE: New York, 2016; pp 194−199.
(39) Shen, A. M.; Chen, C. L.; Kim, K.; Cho, B.; Tudor, A.; Chen, Y. Analog Neuromorphic Module Based on Carbon Nanotube Synapses. *ACS Nano* **2013**, *7*, 6117−6122.
(40) Wei, H.; Shulaker, M.; Wong, H.-S. P.; Mitra, S. Monolithic Three-Dimensional Integration of Carbon Nanotube FET Complementary Logic Circuits. Proceedings from the *2013 IEEE International Electron Devices Meeting*, Washington, DC, December 9−11, 2013; IEEE: New York, 2013; p 19.7.1−19.7.4.

DOI: 10.1021/acsnano.6b07894
ACS Nano XXXX, XXX, XXX−XXX