crystals

MDPI

Article

A Comparison Study on Multilayered Barrier Oxide Structure in Charge Trap Flash for Synaptic Operation

Minkyung Kim [1,2], Eunpyo Park [1,3], In Soo Kim [4], Jongkil Park [1], Jaewook Kim [1], YeonJoo Jeong [1], Suyoun Lee [1], Inho Kim [1], Jong-Keuk Park [1], Tae-Yeon Seong [2] and Joon Young Kwak [1,*]

1 Center for Neuromorphic Engineering, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea; mkk020497@kist.re.kr (M.K.); toyson1@kist.re.kr (E.P.); jongkil@kist.re.kr (J.P.); jaewookk@kist.re.kr (J.K.); jeongyeonjoo@kist.re.kr (Y.J.); slee_eels@kist.re.kr (S.L.); inhok@kist.re.kr (I.K.); jokepark@kist.re.kr (J.-K.P.)
2 Department of Materials Science and Engineering, Korea University, Seoul 02841, Korea; tyseong@korea.ac.kr
3 Department of Electrical Engineering, Korea University, Seoul 02841, Korea
4 Nanophotonics Research Center, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea; isk@kist.re.kr
* Correspondence: jykwak@kist.re.kr

Abstract: A synaptic device that contains weight information between two neurons is one of the essential components in a neuromorphic system, which needs highly linear and symmetric characteristics of weight update. In this study, a charge trap flash (CTF) memory device with a multilayered high-κ barrier oxide structure on the MoS_2 channel is proposed. The fabricated device was oxide-engineered on the barrier oxide layers to achieve improved synaptic functions. A comparison study between two fabricated devices with different barrier oxide materials (Al_2O_3 and SiO_2) suggests that a high-κ barrier oxide structure improves the synaptic operations by demonstrating the increased on/off ratio and symmetry of synaptic weight updates due to a better coupling ratio. Lastly, the fabricated device has demonstrated reliable potentiation and depression behaviors and spike-timing-dependent plasticity (STDP) for use in a spiking neural network (SNN) neuromorphic system.

Keywords: synaptic device; neuromorphic; charge trap flash; multilayered oxide film; MoS_2

Citation: Kim, M.; Park, E.; Kim, I.S.; Park, J.; Kim, J.; Jeong, Y.; Lee, S.; Kim, I.; Park, J.-K.; Seong, T.-Y.; et al. A Comparison Study on Multilayered Barrier Oxide Structure in Charge Trap Flash for Synaptic Operation. *Crystals* **2021**, *11*, 70. https://doi.org/10.3390/cryst11010070

Received: 29 December 2020
Accepted: 15 January 2021
Published: 16 January 2021

1. Introduction

The artificial intelligence (AI) processor has been one of the most important technologies in the 4th industrial revolution [1,2]. AI hardware consisting of conventional, complementary metal-oxide semiconductor (CMOS) circuits based on the von Neumann architecture have been widely implemented, while the system consumes a quite amount of power due to the von Neumann bottleneck [3,4]. To overcome the von Neumann bottleneck, a neuromorphic computing system has been one of the promising candidates to emulate a human brain that consumes approximately 20 Watts of power.

This paper provides a study on an artificial synaptic device, which is one of the essential components of a neural network for building a hardware-based spiking neural network (SNN) system. A biological neural network consists of numerous neurons and synapses. A neuron acts as a spike generator to transfer a spike signal to the next neuron, and a synapse connects two neurons to control the connection strength, which is modified by the biological process called synaptic plasticity in the biological nervous system [5,6]. Since the number of synapses in a neural system is a few orders of magnitude bigger than the number of neurons in general, designing a proper synaptic device is critical in terms of size and power consumption [7,8]. In addition, previous studies have shown that achieving a high accuracy system is determined by synaptic devices with proper symmetricity, nonlinearity(ν), and high conductance on/off ratio in large-scale artificial neural networks (ANN) [9,10].

In the current industry, a charge trap flash (CTF) memory has been fabricated with a silicon nitride dielectric film as a charge trapping layer to trap electrons and SiO_2 as a

tunneling barrier oxide layer [11]. For the use in neuromorphic hardware, we fabricated a CTF device and engineered the barrier oxide layer by proposing a multilayered high-κ structure on the MoS_2 channel to mimic some of the key synaptic-like behaviors. Conductance of the device represents a synaptic weight, controlled by program/erase operation. The device demonstrates the improved symmetric curves and conductance on/off ratios in potentiation and depression behaviors. We also present a spike-timing-dependent plasticity (STDP) to demonstrate a popular learning rule used in SNN-based neuromorphic computing systems.

2. Fabrication Methods

A MoS_2-based CTF memory device was fabricated on a SiO_2 (300 nm)/P^+-Si substrate, as shown in the schematic in Figure 1a. MoS_2 flakes were mechanically exfoliated onto a substrate by using the "scotch tape" method to make a FET channel [12]. For the tape exfoliation method, a piece of cellulose tape was applied on the top surface of the bulk MoS_2, and several layers of flakes from the crystals were peeled off. This leads to the clean surface that had not been previously exposed to the ambient condition [13]. As shown in Figure 1b,c, the AFM image shows the thickness of the exfoliated MoS_2 flake. The flakes were dipped in acetone for 2 h to completely remove the tape residue on the substrate. Then, the sample was rinsed with isopropyl alcohol (IPA) and deionized water (DI water) for 10 min, respectively [14]. The two peaks of MoS_2 (E^1_{2g} and A_{1g}) in Raman spectroscopy were clearly observed using 532 nm wavelength in the device region in Figure 1d. Drain and source electrodes (Ti 5 nm/Au 70 nm) were patterned using photolithography, followed by an e-beam evaporator deposition process. Since two dimensional layered materials have dangling bond-free surface [15], 0.5 nm of Al were deposited first as a seed layer, creating a native thin Al_2O_3 layer for the atomic layer deposition (ALD) process [16]. For the tunneling oxide layer, 4 nm of SiO_2 were deposited by plasma-enhanced chemical vapor deposition (PECVD) at 250 °C. A multilayer oxide/nitride stack was formed to increase the trap sites rather than a single layer by alternately depositing a 2 nm Si_3N_4 film as a charge trapping layer and a 2 nm Al_2O_3 film as an intermediate barrier oxide layer by PECVD at 250 °C and ALD at 150 °C, respectively. For the top-gate dielectric blocking oxide, 6 nm of Al_2O_3 was deposited using ALD. Finally, the top-gate electrode was deposited with Ti 5 nm/Au 70 nm by e-beam evaporator. The samples were annealed at 300 °C for 2 h by flowing H_2 5%/N_2 95% mixed gas. The cross-sectional transmission electron microscopy (TEM) image of the device in Figure 1e identified the MoS_2 channel layer and the multilayered charge trapping structure.

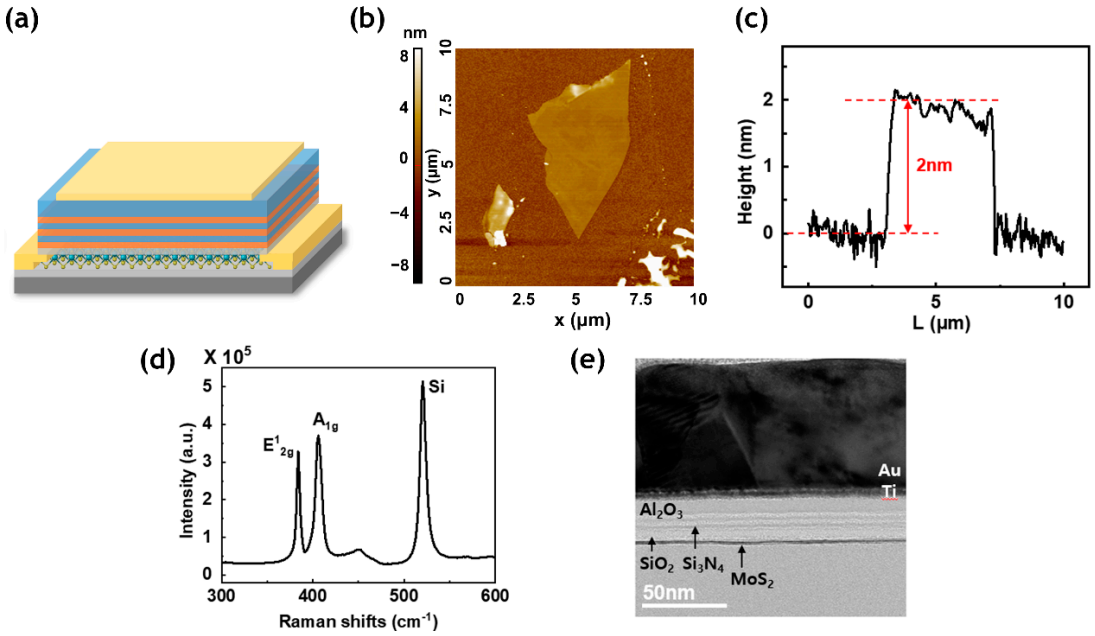

Figure 1. (**a**) Schematic of the CTF memory device with multilayered charge trapping structure. (**b**,**c**) AFM scan image and height profile of the MoS_2 flake. (**d**) Raman spectra of MoS_2 peaks; E^1_{2g} at 384 cm^{-1} and A_{1g} at 406 cm^{-1}. (**e**) Cross-sectional TEM image of the device.

3. Results and Discussion

Electrical properties were measured by a vacuum chamber probe station (MS Tech (Hwaseong, Korea), M5VC) and a semiconductor characterization system (Keithley, 4200A-SCS) at room temperature and ambient condition. For the measurement, two source measure units (SMUs) at drain/top-gate terminals were used to sweep and measure voltages with the source terminal grounded. The output and transfer curves of the CTF memory device are shown in Figure 2. The linear and symmetric drain-source current (I_{ds}) of the output curve ($I_{ds} - V_{ds}$) implies an ohmic contact between MoS_2 and Ti/Au as shown in Figure 2a. Figure 2b shows the transfer curve ($I_{ds} - V_{gs}$) measured by sweeping the top-gate at the range of –10 V to 10 V at the various drain-source voltages (V_{ds}) of 0.1 V, 1 V, and 2 V. This device exhibits n-type operation due to the unique characteristics of MoS_2 [17] and has the field-effect mobility of 2.1 $cm^2V^{-1}s^{-1}$. The mobility was extracted from the transfer curve in Figure 2b using Equation (1);

$$\mu = \left(dI_{ds}/dV_{gs}\right) \times \left(L/WV_{ds}C_i\right) \quad (1)$$

where L is length (5 μm), W is channel width (5 μm) for the device, and C_i is the capacitance for gate insulator per unit area.

Voltage pulses (V_{gs}) representing a neuronal spike train were consecutively applied to the top-gate to mimic the potentiation and depression behaviors of a biological synapse. The operating mechanism of the CTF device is related to the tunneling hot electrons and the charge trapping states. When a negative voltage pulse is applied to the top-gate electrode, a portion of the trapped electrons in the three nitride layers are released to the channel. As a result, the device is moved to a low threshold voltage state (erase state), and the channel current (conductance) increases. On the contrary, when a positive voltage pulse is applied to the top-gate electrode, the three nitride layers are gradually charged by the electrons tunneled from the channel, resulting in a high threshold voltage state (program state). This state effectively decreases the channel current (conductance). In this study, two different devices were fabricated with different intermediate barrier oxide stacks to compare the synaptic device properties. Schematics of the two different multilayered charge trapping thin film stacks are illustrated in Figure 3a.

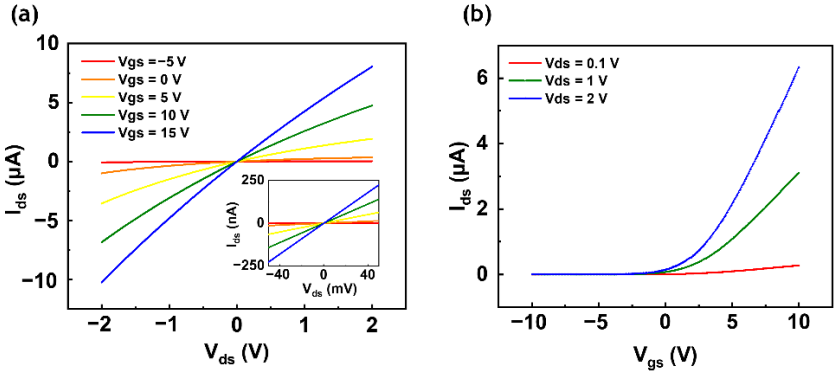

Figure 2. Output and transfer curves of the memory device. (**a**) Output curve (I_{ds}–V_{ds}) with gate voltage from −5 V to 15 V with 5 V step. Inset shows an ohmic contact of the device. (**b**) Transfer curve (I_{ds}–V_{gs}) measured at drain voltages of 0.1 V, 1 V, and 2 V with the top-gate voltage sweep from −10 V to +10 V.

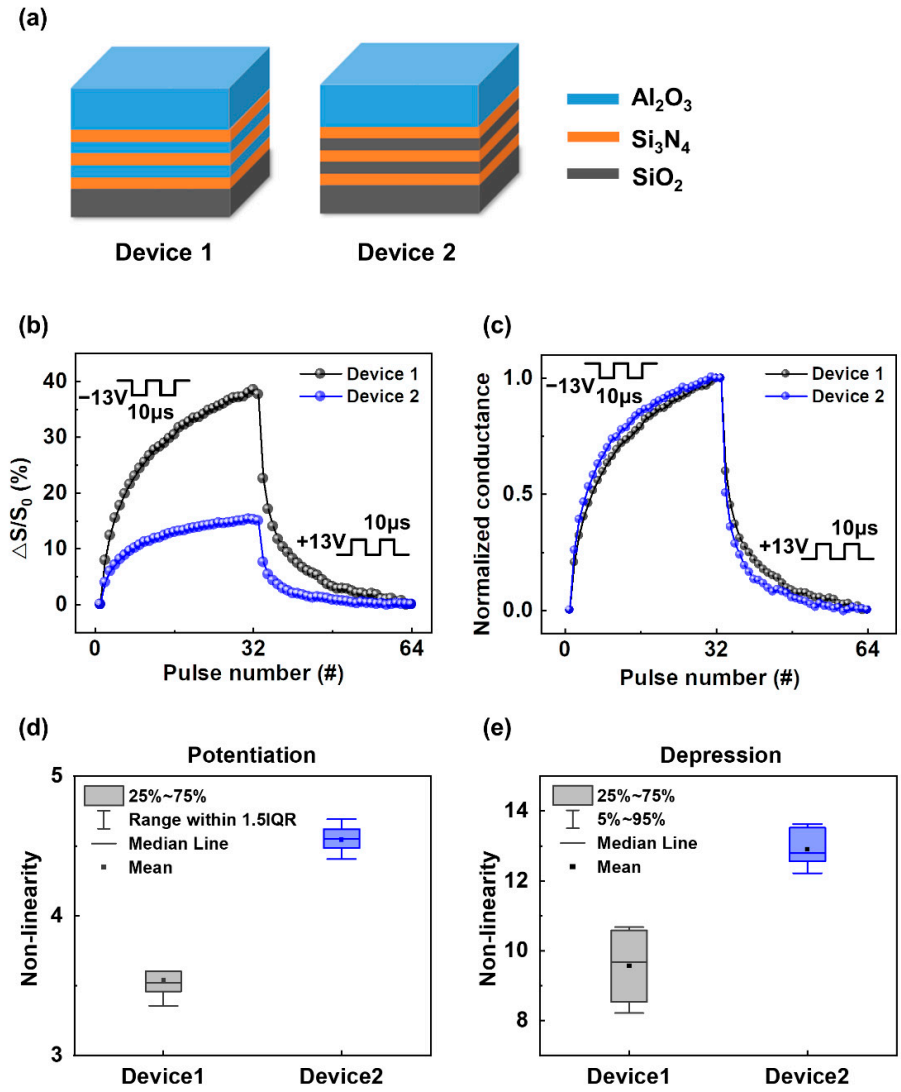

Figure 3. Schematics of multilayered charge trapping stacks and the synaptic properties of the devices. (**a**) Charge trapping layers (Si_3N_4) with different intermediate barrier oxide layers of Al_2O_3 for Device 1 and SiO_2 for Device 2. (**b**,**c**) Potentiation and depression curves of Device 1 and 2. The higher coupling ratio (Device 1) shows a higher conductance on/off ratio and a more linear synaptic update. (**d**,**e**) Nonlinearity values for fourteen cycles of potentiation and depression operations.

Both structures have the same 4 nm of SiO_2 for the tunneling oxide, 6 nm of Al_2O_3 for the blocking oxide, and 2 nm of Si_3N_4 for each charge trapping layer, while they have different intermediate barrier oxide layers (2 nm of Al_2O_3 and 2 nm of SiO_2 for Device 1 and Device 2, respectively) for this comparison. Each intermediate barrier oxide layer and charge trapping nitride layer was deposited alternately three times for both devices. Figure 3b shows the potentiation and depression curves of the synaptic devices with thirty-two synaptic weight states. For the measurement, thirty-two consecutive voltage pulses of −13 V with the pulse width of 10 µs for potentiation and +13 V with the pulse width of 10 µs for depression were applied sequentially to the top-gate to update the weight, and the device channel current (or conductance) was read after each pulse was applied. Device 1 with Al_2O_3 as the intermediate barrier oxide layers shows a higher conductance on/off ratio than that of Device 2 with SiO_2 as the intermediate barrier oxide layers, attributed to the coupling ratio differences between the two devices. A device coupling ratio determines a portion of the applied voltage across the tunneling oxide at the given top-gate voltage. Therefore, the higher coupling ratio allows for the higher tunneling voltage applied between the channel and the charge trapping layer. A coupling ratio equation is given by [18,19];

$$\alpha_{cg} = C_{Block}/(C_{Block} + C_{Tunnel}) \tag{2}$$

where α_{cg} is coupling ratio, C_{Block} and C_{Tunnel} are capacitances of the blocking oxide and the tunneling oxide, respectively.

The estimated values of the coupling ratio of Device 1 and Device 2 are 0.34 and 0.28, respectively. From the measurements, Device 1 with a high coupling ratio shows a higher conductance on/off ratio and more linear potentiation and depression curves than Device 2 as shown in Figure 3b,c. A high coupling ratio enhances the tunneling efficiency and the memory window for threshold voltage shift, and hence, the device's synaptic behavior is improved. Figure 3d,e show the nonlinearity values for fourteen cycles of potentiation and depression operations, defined by Equations (3) and (4) [20].

Potentiation:

$$G = G_1\left(1 - e^{-\nu P}\right) + G_{min} \tag{3}$$

Depression:

$$G = G_{max} - G_1(1 - e^{-\nu(1-P)}) \tag{4}$$

where

$$G_1 = (G_{max} - G_{min})/\left(1 - e^{-\nu}\right),$$

G_{max} and G_{min} are the maximum and minimum conductance of the device, respectively, ν is a parameter of nonlinearity, and P is the normalized pulse number.

The average nonlinearity values at thirty-two conductance steps of Device 1 and Device 2 in the potentiation operations were estimated as 3.49 and 4.48, respectively; and the average nonlinearity values of the devices in the depression operations were estimated as 8.27 and 12.76, respectively. Figure 4 shows the robustness of the synaptic device through an iterative process of 896 voltage pulses in potentiation and depression operations.

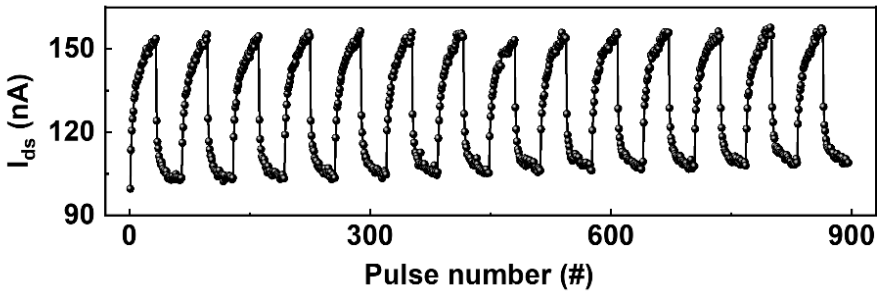

Figure 4. Iterative measurement of 896 pulses in potentiation and depression operations.

Synaptic weight updates based on the STDP learning rule are demonstrated in Figure 5 to mimic a biological process in a SNN-based neuromorphic system [21]. The STDP behavior modulates the synaptic weight information between the pre- and post-synaptic neurons according to a relative spike timing difference. To show the STDP behavior, a set of differently shaped pulses ($V_{applied}$), containing the timing difference information ($\Delta t = t_{post} - t_{pre}$) between a pre-synaptic pulse (V_{pre}) and a post-synaptic pulse (V_{post}), were applied to the top-gate (Figure 5a,b).

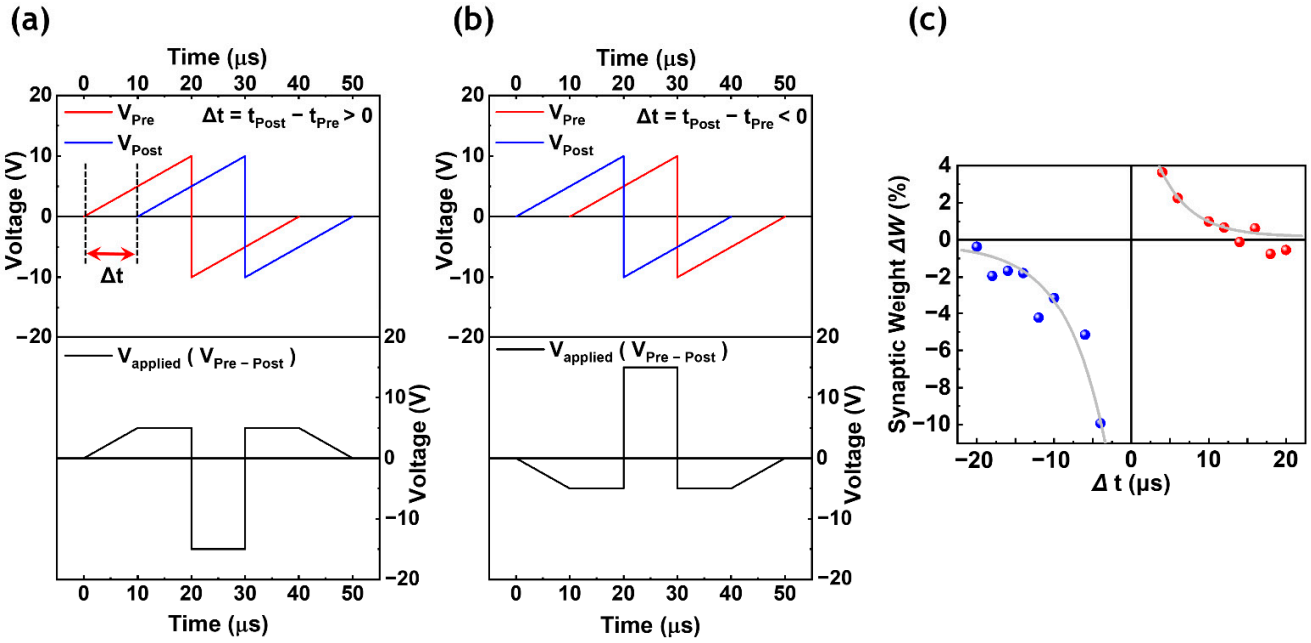

Figure 5. Spike-timing-dependent plasticity (STDP) with synaptic weight change according to time dependence. (**a**,**b**) Applied pulses ($V_{applied} = V_{pre} - V_{post}$) calculated from the pre- and post-synaptic neuron spikes at different times (Δt). (**c**) STDP measurement with fitting curves.

The pre- and post-synaptic neuron spike shapes (V_{pre} and V_{post}) were adapted from a previous study [22]. When V_{pre} spikes before V_{post} (positive Δt case), the synaptic weight is enhanced, whereas, when V_{post} spikes before V_{pre} (negative Δt case), the synaptic weight is weakened. Figure 5c shows the amount of synaptic weight (conductance) change at each Δt. The exponential fitting lines at both polarities indicate a well-behaved STDP learning rule; the shorter the Δt, the greater the amount of weight change.

4. Conclusions

In summary, we fabricated a flash memory device with a multilayered charge trapping structure to mimic synaptic operations. A high on/off ratio and symmetric conductance updates have been demonstrated using high-κ intermediate barrier oxide film stacks. The device with Al_2O_3 as intermediate barrier oxide layers exhibited a higher on/off ratio and better linear conductance change than the device with SiO_2 as intermediate barrier oxide layers. This result is attributed to the enhanced coupling ratio from the high-κ material used in the intermediate barrier oxide layers, which increases the memory threshold voltage shift window of the device. The iterative experiment and STDP data from the fabricated synaptic device suggest that a CTF device with the multilayered high-κ barrier oxide structure can be one of the viable synaptic device candidates for an SNN-based neuromorphic system.

Author Contributions: M.K. conceived and designed the research and drafted the manuscript under the supervision of J.Y.K. E.P., and I.S.K. prepared MoS_2 and ALD oxides. J.P., J.K., Y.J., S.L., I.K., J.-K.P., and T.-Y.S. participated in the design of the study and data analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Institute of Science and Technology (KIST) (Grant No. 2E30610, 2E30761). The authors acknowledge the National Research Foundation of Korea (NRF-2019M3F3A1A02072175).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nilsson, N.J. *Principles of Artificial Intelligence*; Morgan Kaufmann: Burlington, MA, USA, 2014.
2. Mitchell, R.; Michalski, J.; Carbonell, T. *An Artificial Intelligence Approach*; Springer: Berlin/Heidelberg, Germany, 2013.
3. Calhoun, B.H.; Wang, A.; Chandrakasan, A. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE J. Solid State Circuits* **2005**, *40*, 1778–1786. [CrossRef]
4. Hasler, J.; Marr, H.B. Finding a roadmap to achieve large neuromorphic hardware systems. *Front. Neurosci.* **2013**, *7*, 118. [CrossRef] [PubMed]
5. Goelet, P.; Castellucci, V.F.; Schacher, S.; Kandel, E.R. The long and the short of long–term memory—A molecular framework. *Nature* **1986**, *322*, 419–422. [CrossRef] [PubMed]
6. Striedter, G.F. *Neurobiology: A Functional Approach*; Oxford University Press: Oxford, UK, 2016.
7. Bi, G.-q.; Poo, M.-m. Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature* **1999**, *401*, 792–796. [CrossRef] [PubMed]
8. Nithianantharajah, J.; Hannan, A.J. The neurobiology of brain and cognitive reserve: Mental and physical activity as modulators of brain disorders. *Prog. Neurobiol.* **2009**, *89*, 369–382. [CrossRef] [PubMed]
9. Chen, P.-Y.; Peng, X.; Yu, S. NeuroSim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures. In Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2–6 December 2017; pp. 6.1.1–6.1.4.
10. Chen, P.-Y.; Peng, X.; Yu, S. NeuroSim: A circuit-level macro model for benchmarking neuro-inspired architectures in online learning. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **2018**, *37*, 3067–3080. [CrossRef]
11. Jung, M.-H.; Kim, K.-S.; Park, G.-H.; Cho, W.-J. Dependence of charge trapping and tunneling on the silicon-nitride (Si_3N_4) thickness for tunnel barrier engineered nonvolatile memory applications. *Appl. Phys. Lett.* **2009**, *94*, 053508. [CrossRef]
12. Lee, G.-H.; Cui, X.; Kim, Y.D.; Arefe, G.; Zhang, X.; Lee, C.-H.; Ye, F.; Watanabe, K.; Taniguchi, T.; Kim, P. Highly stable, dual-gated MoS_2 transistors encapsulated by hexagonal boron nitride with gate-controllable contact, resistance, and threshold voltage. *ACS Nano* **2015**, *9*, 7019–7026. [CrossRef] [PubMed]
13. Frindt, R. Single crystals of MoS_2 several molecular layers thick. *J. Appl. Phys.* **1966**, *37*, 1928–1929. [CrossRef]
14. Zheng, X.; Calò, A.; Albisetti, E.; Liu, X.; Alharbi, A.S.M.; Arefe, G.; Liu, X.; Spieser, M.; Yoo, W.J.; Taniguchi, T. Patterning metal contacts on monolayer MoS_2 with vanishing Schottky barriers using thermal nanolithography. *Nat. Electron.* **2019**, *2*, 17–25. [CrossRef]
15. Qian, Q.; Lei, J.; Wei, J.; Zhang, Z.; Tang, G.; Zhong, K.; Zheng, Z.; Chen, K.J. 2D materials as semiconducting gate for field-effect transistors with inherent over-voltage protection and boosted ON-current. *NPJ 2D Mater. Appl.* **2019**, *3*, 1–9. [CrossRef]
16. Bertolazzi, S.; Krasnozhon, D.; Kis, A. Nonvolatile memory cells based on MoS_2/graphene heterostructures. *ACS Nano* **2013**, *7*, 3246–3252. [CrossRef] [PubMed]
17. Bao, W.; Cai, X.; Kim, D.; Sridhara, K.; Fuhrer, M.S. High mobility ambipolar MoS_2 field-effect transistors: Substrate and dielectric effects. *Appl. Phys. Lett.* **2013**, *102*, 042104. [CrossRef]
18. Sah, R.E. Silicon Nitride, Silicon Dioxide, and Emerging Dielectrics 11. In Proceedings of the 219th ECS Meeting, Montreal, Canada, 1–6 May 2011.
19. Puzzilli, G.; Govoreanu, B.; Irrera, F.; Rosmeulen, M.; Van Houdt, J. Characterization of charge trapping in SiO_2/Al_2O_3 dielectric stacks by pulsed C–V technique. *Microelectron. Reliab.* **2007**, *47*, 508–512. [CrossRef]
20. Agarwal, S.; Plimpton, S.J.; Hughart, D.R.; Hsia, A.H.; Richter, I.; Cox, J.A.; James, C.D.; Marinella, M.J. Resistive memory device requirements for a neural algorithm accelerator. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 929–938.
21. Shouval, H.Z.; Wang, S.S.-H.; Wittenberg, G.M. Spike timing dependent plasticity: A consequence of more fundamental learning rules. *Front. Comput. Neurosci.* **2010**, *4*, 19. [CrossRef]
22. Xu, R.; Jang, H.; Lee, M.-H.; Amanov, D.; Cho, Y.; Kim, H.; Park, S.; Shin, H.-j.; Ham, D. Vertical MoS_2 double-layer memristor with electrochemical metallization as an atomic-scale synapse with switching thresholds approaching 100 mV. *Nano Lett.* **2019**, *19*, 2411–2417. [CrossRef] [PubMed]