Highly Uniform All-Vacuum-Deposited Inorganic Perovskite Artificial Synapses for Reservoir Computing

Li-Wei Chen, Wei-Chun Wang, Shao-Han Ko, Chien-Yu Chen, Chih-Ting Hsu, Fu-Ching Chiao, Tse-Wei Chen, Kai-Chiang Wu,* and Hao-Wu Lin*

The development of artificial synapses is inspired by the energy-efficient recognition ability of the central nervous system of living organisms and is proposed for use as the basic building units of next-generation neuromorphic computing networks. Herein, the perovskite/MoO₃/Ag synaptic device exhibits very uniform electrical characteristics, a low operation voltage, and linear and repeatable analog switching. Important synaptic functions such as paired-pulse facilitation (PPF) and short- and long-term potentiation (STP/LTP) are all successfully and stably implemented in the device. The diffusion of Ag⁺ and I⁻ ions is believed to contribute to the synaptic behavior of the device. Furthermore, the uniform and stable volatile characteristics are utilized in a proof-of-concept reservoir-computing-based simulation program using all the experimental parameters; the results strongly suggest that the neuromorphic computing that utilizes such components can simultaneously reduce the circuitry complexity and increase its recognition accuracy with noisy inputs. The findings of this work promise to significantly facilitate the utilization of highly homogeneous vacuum-deposited perovskite/metal oxide structures in high-performance neuromorphic computing systems.

Neuromorphic computing has attracted significant attention since its successful application to machine deep learning. This has enabled the effective solution of problems that have been difficult to handle using traditional computing methods, such as pattern recognition,[1–3] voice search,[4,5] and motion detection.[6,7] However, the implementation of neuromorphic computing has so far been limited to software programming, with the synaptic weight updates involving pure mathematical calculation and summation using a traditional Von Neumann architecture. There has, nevertheless, been a gradual increase in interest in neuromorphic computing using hardware-based electronic devices that mimic the behavior of a biological synapse, with the synaptic weight updates implemented as the resistance state of the device.[11–15] The use of this type of device and calculation architecture is referred to as “in-memory computing.” The device significantly differs from a Von Neumann machine, being more similar to a biological neuro-network, with the memory and arithmetic unit indistinguishable. The electrical states of the device are usually resistance or conductance states, and the device can, thus, be manipulated in a multi-level manner, similar to the potentiation and depression of a synaptic link, and used as the basic building block for developing a hardware-based neuromorphic computing network. Because of the analog behavior of a natural synapse, the proposed device is usually referred to as an “artificial synapse.”[16] Several state-of-the-art technologies such as phase-change memory,[17–20] resistive memory,[21–24] ferroelectric memory,[25–28] and ion-diffusion memory[29–32] are under consideration for use in reproducing synaptic behavior by an electronic device. In this drive, some device features such as uniformity and programming stability[13–15] are considered to be crucial for achieving a high system performance.

Reservoir computing (RC) is a framework evolved from recurrent neural networks (RNNs) and mainly consists of two parts: a dynamic reservoir, followed by a readout layer. A reservoir with a short-term property can non-linearly transform the input data into a higher dimensional space to obtain the reservoir input. Following this, the reservoir output would enter the readout layer, where training and testing are carried out through a linear algorithm.[33] As the network training is solely carried out at the readout layer, the training cost can be significantly reduced.[34] The ability of the reservoir to pre-process the input data can aid in decreasing the network size by multiples compared with traditional NNs, thus lowering the computation power. Owing to the short-term characteristics of the reservoir, the state change is influenced by recent inputs but remains unaffected by far-past
inputs. Therefore, RC is inherently suitable for dealing with time-related input signals. In recent times, various systems are being utilized to build RC frameworks.[15] In addition, several time-series tasks, such as digit recognition and chaotic system forecasting, are resolved.[2,38]

Halide perovskites are some of the most fascinating semiconductors, exhibiting intriguing optoelectronic properties. They have been used to develop very efficient solar cells, light emitting diodes, and lasers.[39-44] The easy migration of ions in the perovskite thin film is believed to be able to mimic the biological synapse.[41] In previous works, we demonstrated the effectiveness of a vacuum process for the production of thin, large, and homogeneous high-quality perovskite films for the fabrication of solar cells.[45-50] However, due to the ease of implementation and versatility of solution systems, they are utilized in most studies on halide perovskites, including the pioneer works that revealed the synaptic characteristics of the materials[29,31,51].

In the present study, we demonstrated the high performance of artificial synapses with all-vacuum-deposited inorganic halide perovskite/metal oxide/metal structures. The devices were found to exhibit excellent synaptic characteristics, along with superior device-to-device and cycle-to-cycle uniformity and stability. These real-life device parameters were further evaluated using an RC-based handwritten digit recognition simulation program and were shown to achieve a recognition accuracy of 89%. Notably, the proposed devices attained an acceptable accuracy of 74%, regardless of exceedingly noisy images. The recognition accuracy value was close to that of the traditional non-reservoir NN, which possessed a significantly larger network size. The results also strongly suggested that the device-to-device uniformity and cycle-to-cycle variation were the two most important factors of the recognition accuracy of the device, indicating the promise of all-vacuum-fabricated perovskite artificial synapses for the development of future hardware-based neuromorphic computing systems.

All-vacuum deposition affords great flexibility for the fabrication of a variety of layered structures without having to deal with the dissolution problems associated with solution processes. Cesium-based inorganic perovskites, CsPbX₃ (where X denotes a halide), were chosen owing to their compatibility with the vacuum process.[45,50] The ability of the device resistance to change from high impedance to low impedance and vice versa, which is analogous to the potentiation and depression of a synaptic link, was evaluated. Previous studies have revealed that the strong reaction of halides with the metal eventually compromises the stability of the device.[52-55] To counter this issue, we introduced a thin MoO₃ layer between the perovskite and the Ag layer to prevent direct contact between the two, and the device was configured as: indium tin oxide (ITO)/CsPbI₂Br/MoO₃/Ag. The top-view photographs of the devices with and without the MoO₃ layer between CsPbI₂Br/Ag interfaces are shown in Figure S1a,b, Supporting Information, respectively. These devices were stored in a N₂ glove box for ≈3 months. The one without the MoO₃ layer clear exhibited a ring-like appearance around the Ag electrode, which was due to the chemical reaction between Ag and halides in perovskite under a direct contact. The corresponding current (I)–voltage (V) and current (I)–time (T) characteristics of the device are shown in Figure S2a,b, Supporting Information. The device exhibited conspicuous changes in the current under electrical loop and pulse stimulation. Nevertheless, exceedingly high current densities were observed; this is unsuitable for practical low-power applications, irrespective of the change in resistance.

By inserting another thin MoO₃ underlayer between the ITO substrate and the CsPbI₂Br layer, we found that there was scope for significant reduction in the device current. This is due to the much larger bandgap of MoO₃ and large carrier injection barrier from the perovskite to the MoO₃ layer, which efficiently decreases the overall device current.[56-59] The optimized device not only exhibited synaptic characteristics, but also showed exceedingly diminished driving voltage and operation current density, which are highly desirable for low-power computing. The structure and typical I–V sweeping characteristics of the device are presented in Figure 1a,b. The ITO anode and the Ag cathode, respectively, mimic pre-synapse and post-synapse, whereas the electrical stimulation of the device induces the mimetic synaptic behavior. The device resistance decreased under negative sweeping loops and could be restored to the high-resistance states (HRSs) via positive sweeping loops.

We investigated the lowest spiking voltage that could be used to drive the device. As is evident from Figure S3a,b, Supporting Information, the device clearly shows a gradual increase of conductance under stimulation by a very low stimulation of −0.1 V pulses and sweeping voltages. This low operation voltage is much lower than those of other two-terminal devices and three-terminal synaptic transistors and is among the lowest values observed for any type of artificial synapse (see Table S1, Supporting Information).[1,3,31,50,60-65] The energy consumption of the device is ≈2 nJ mm⁻², but the adoption of a smaller device area can be used to substantially reduce the operating current of the devices, as shown in Figure S4, Supporting Information. Hence, it is believed that a smaller device area can reduce the energy consumption to a value comparable to that of the human brain (≈10⁻¹⁴ J per synaptic event),[61,66] as listed in Figure S5, Supporting Information. It is evident from Figure S6, Supporting Information, that the device exhibits a smooth curve and analogue change under different spiking conditions. Unlike in the case of conventional memories,[67-70] the device current changes abruptly from HRS to low-resistance state (LRS); moreover, artificial synapses are capable of exhibiting continuous change between HRS and LRS. Continuous resistance changes in the present device between 1000 and 10 000 Ω are shown in Figure 1c; the high stability of the device can also be observed from the figure. The device shows a good potentiation linearity with a non-linearity factor value of −0.05 and a moderate depression linearity with a non-linearity factor of −2.8.[71]

To fully elucidate how well an artificial synapse mimics the behavior of a natural synapse, it is necessary to access several cognitive mechanisms of the artificial synapse, especially in the time domain. In a natural neuromorphic system, the synaptic plasticity, which is the ability of the synapse to adjust its synaptic weight, is an important factor of learning and memory, according to Hebbian theory.[11,71,72] Synaptic weight is the strength of a connection between neurons, which can be mimicked by different conductance states of an electronic device. When a voltage stimulates an artificial synapse, the concentration of the neurotransmitters temporarily increases, with the original state restored with time. If a second stimulating voltage spike
arrives within a short time interval (Δt) before the restoration, i.e., before the complete disappearance of the neurotransmitters induced by the first spike, the concentration of the neurotransmitters after the second spike would be higher than that after the first spike. Conversely, if Δt is sufficiently long for the synapse to recover to its original state after the first spike, the neurotransmitter concentrations induced by the two spikes would be similar. This phenomenon is referred to as paired-pulse facilitation (PPF) and is shown in Figure 2a.[73,74] As shown in Figures S7a,b, Supporting Information, for a small Δt, the current after the second pulse is higher than that after the first pulse, whereas the two are the same for a large Δt. Figure S7c, Supporting Information, shows the PPF index as a function of Δt. As can be observed, the PPF index increases with decreasing Δt, and when Δt is sufficiently large, the PPF index decreases to almost 100%. Figure 2b shows the I–V curve of the device under consecutive stimulations by a series of /C0.5/ + 0.05 V pulses. The write and read currents were both found to increase linearly with continuous pulse stimulation, indicating a highly stable PPF behavior. Another basic synaptic parameter is the spiking-rate-dependent plasticity (SRDP), which relates the plasticity of the synapse to the frequency of the input spikes.[75–77] Similar to the PPF, when a train of high-frequency spikes arrives at the neuron, the synaptic weight increases with each successive spike. The SRDP of the present device is shown in Figure 2c,d, which reveal that at sufficiently small Δt values between successive pulses, there is a significant increase in the device current; this behavior is typical of the SRDP. The device current for the application of ten pulses is plotted against Δt in Figure S7d, Supporting Information. Figure 2e shows the decay stability of the present device. The decay curves remained unchanged over 10^5 stimulations and under consecutive measurements, thus establishing the reliability of the short-term and long-term stability characteristics in further simulations.

Short-term potentiation (STP) and long-term potentiation (LTP) are two important types of plasticity of the human brains. They are considered to reflect the brain’s basic functions of memory and learning and are distinguished by their retention times, with STP lasting for tens of milliseconds and LTP lasting for minutes to hours.[78–80] We successfully implemented the STP-to-LTP transition behavior, as represented in Figure S8a, Supporting Information. Noticeably, the LTP current increases with increasing stimulation voltage, and the on/off ratio remains sufficiently high compared with the STP state. In addition, stimulation of the device by a train of −1.8 V pulses produces a sufficiently long retention time of over 10^6 s (18 days) in Figure S8b, Supporting Information. Figure S9, Supporting Information, shows that the device can still be tuned back to HRS after the application of a series of 1.5 V pulses, indicating good performance with the device remaining intact under the extreme LTP.

Figure 3a shows the cross-sectional transmission electron microscope (TEM) image of the fabricated device. The complete separation of the top metal layer from the perovskite layer by a 15 nm thick MoO3 layer can be observed. The smoothness and uniformity of the device morphology are considered to be the basis of the observed uniform electrical properties. The high quality of the vacuum-deposited CsPbI2Br layer was evident from its lattice diffraction patterns, as shown in Figure 3b. Wang et al. revealed the direct evidence on energy level alignments at the
hybrid interfaces between perovskite and MoO$_3$ layer, and the formation of a high carrier mobility (tens of cm$^2$ V$^{-1}$ s$^{-1}$) AgI layer at this interface provided a stepwise energy level alignment, which enhanced the charge conductivity of the devices. As a result, we believe that the synaptic mechanism of our device is based on the migration of Ag$^+$ and I$^-$ ions, accompanied with AgI formation/annihilation at the CsPbI$_2$Br/MoO$_3$ interface to modify the charge transfer properties. Under reverse bias, the Ag$^+$ ions experience an electrical field and migrate across the MoO$_3$ layer to encounter the easily migrating I$^-$ ions to form a thin AgI composite, which gradually improves the conductance of the device, owing to the enhanced charge transfer ability. The AgI layer is unstable under weak stimulations, and the ions drift eventually back. This corresponds to short-term behaviors, as shown in Figure 3c. To verify this hypothesis, we replaced the top electrode (Ag) with Au. The I$-$V curve of these devices is shown in Figure S10a,b, Supporting Information. In the absence of synaptic behavior, it is evident that the strongly mobilized Ag$^+$ ions play a crucial role in the synaptic phenomenon. We then substituted the MoO$_3$ underlayer with different carrier transporting layers. The I$-$V curves in Figure S2a and S10c,d, Supporting Information, exhibit synaptic behavior under distinctive current ranges, thus verifying that the mechanism is unrelated to underlayer materials. Based on the results, vacuum sublimation was used to introduce a thin 2 nm AgI layer at different positions in the device. As is evident from the results in Figure 3d, the fresh device 1) with deliberately deposited AgI at the perovskite/MoO$_3$ interface exhibits a significantly higher current compared with the device 2) without the AgI layer, 3) AgI inside the MoO$_3$, and 4) at the MoO$_3$/Ag interface. Hence, we conclude that the formation/annihilation of AgI at the CsPbI$_2$Br/MoO$_3$ interface contributes to the synaptic mechanism. The I$-$V results in Figure S10e$-$g, Supporting Information, of other device structures are also consistent with the proposed mechanism. For the STP-to-LTP transition, the AgI layer exhibits enhanced stability under stronger stimulations, as shown in Figure S11. Energy-dispersive X-ray spectroscopy (EDS) analyses of the Ag/MoO$_3$/CsPbI$_2$Br structure in the HRS (fresh device) and LRS (by applying a large stimulation to LTP) revealed trace Ag in the perovskite layer, as indicated in Figure 3e. It has been verified by various groups that halides in perovskite materials are highly reactive with Ag$^+$, hence, we believe that after the Ag migration, the AgI layer is formed at the perovskite/MoO$_3$ interface. The result of LTP was in good agreement with those of the EDS analyses and the HRS/LRS characteristics of the device.

Figure 2. a) Schematic illustration of the PPF phenomenon. The synaptic weight change after the second stimulation is correlated with the time interval ($\Delta t$) between two stimulation spikes. For a small $\Delta t$, the enhancement of the synaptic weight after the second spike is greater than that after the first spike. b) $I$–$T$ curve of the device under spike operation. Under a train of $-0.5$ V pulses with a 200 ms pulse duration, the read current (measured with a 0.05 V bias) and the write current (measured under a $-0.5$ V pulse) could be continuously linearly tuned, exhibiting a steady PPF behavior. c) Read current after ten $-0.5$ V pulses with respect to $\Delta t$ under 0.05 V and pulse durations of different $\Delta t$. The current significantly increases with continuous application of the pulses for a small $\Delta t$. d) SRDP behavior of the device with respect to the number of applied $-0.5$ V pulses and $\Delta t$, where $I_x$ is the current after $x$ pulses ($x = 1$–10). e) Stability test under continuous $-0.3$ V–100 ms pulses and decay curve measured at 0.05 V. In the first column, the decay curves after the 1st, 10th, 100th, 1000th, 10 000th, and 100 000th pulse fully overlap with each other, exhibiting a stable decay behavior for over 10$^5$ stimulations. The consecutive columns show similar decay curves for the consecutive stimulations.
STP behavior of the artificial synapses can be implemented for efficient RC. To evaluate the performance of the present device in an RC framework, we conducted a few related measurements before running a simulation program. Only volatile STP was utilized here, and the devices were operated at a relatively small stimulated voltage ($V_{stim} < 0.3$ V) to avoid them entering the LTP region.

Figure 4a shows the results of 4 bit test. We stimulated various arrangements of "1" ($V_{stim} = 0.3$ V, 100 ms pulse) and "0" (no additional pulse, device current decays naturally) to the device, and the device conductance was read with a $V_{read} = 0.05$ V, 50 ms pulse. The device responded to the different inputs with distinct final conductance values, thus verifying its ability to distinguish different 4 bit inputs. In Figure 4b, a specific 4 bit pattern is applied repetitively on the device 50 times with a 2 s time interval between successive patterns to obtain the cycle-to-cycle variation. The standard deviation ($\sigma$) of the final conductance value was appreciably low ($\sigma = 2.5\%$). In Figure 4c, a specific 4 bit pattern is applied on different devices, and the $\sigma$ of the final conductance values does not exceed 15%, thus proving a low device-to-device variation.

The variation is small compared with other studies ($\sigma = 15\%$).\[88\] The stability of the present device under the same conditions for the 4 bit tests (Figure 2e) proves the reliability of the results.

Based on the framework of RC and the measured real-device parameters, we carried out a series of experiments on handwritten digit recognition to simulate the strength of our device for real-world applications. The recognition system was trained...
and tested with the commonly used Mixed National Institute of Standards and Technology (MNIST) database, which contains a training dataset of 60,000 images and a testing dataset of 10,000 images. Before being fed to the dynamic reservoir, input data/images went through a pre-processing stage. Based on our optimization results (Figure 4a), we decided to divide each row into seven 4 bit sections, which leads to a total of 196 sections (seven sections per row multiplied by 28 rows), and each 4 bit section has 24 possible pattern combinations. This setting implies a good tradeoff between efficiency and accuracy. From the efficiency perspective, there is a tremendous reduction in the size of the recognition system (i.e., NN), and the recognition will be more efficient. From the accuracy perspective, the distinguishability of 2⁴ possible combinations is effective even in the presence of variation, and thus, high-quality recognition accuracy can be maintained.

We conducted two sets of experiments on handwritten digit recognition. In the first experiment, we focused on comparing our RC-based approach against traditional NN, in terms of recognition accuracy. The details of the RC framework setup are presented in Figure 5a and in the Experimental Section. The setup of the traditional NN is shown in Figure 5b. Furthermore, we introduced the commonly encountered salt-and-pepper noise into the testing images, to showcase the overall performance of the device in a more real-world scenario. The results are presented in Figure 5c and Table S2, Supporting Information. Upon comparison of the recognition accuracy of our RC-based approach with traditional NN, it is evident that there is some accuracy loss (∼3–5.7%) depending on the percentage of noise in the testing images. However, as mentioned earlier, based on the RC framework and our experimental settings, the RC-based simulation only requires a dynamic reservoir and a readout function. The former does not need any training, and the latter is a 196 × 10 single-layer network. On the other hand, the traditional NN is a 784 × 10 single-layer network, which is four times larger than the RC-based simulation network size. Consequently, this will lead to significantly higher costs in terms of computation and energy. Considering the overall system size, it is clear that our RC-based approach would boast a significantly reduced training time. Meanwhile, the recognition accuracy remains promising, with minimal accuracy loss in comparison with a much larger system.

In the second experiment, the aim is to demonstrate the effects of imperfect devices and noisy images on recognition accuracy. To this end, we not only considered noisy images, but also introduced device-to-device and cycle-to-cycle variations during the training and testing procedures. Figure 5d and Table S3, Supporting Information, present the results of the second experiment. The value of σ was set at 0.15, based on the measurements of our device instances. This means that device-to-device and cycle-to-cycle variations of our device instances are around the magnitude and do not exceed 0.15, when considered jointly. However, the settings of σ = 0.5, 1, and 2 represent devices with much higher variation, which significantly exceeds than that of our device.

Compared with the ideal case of σ = 0, the accuracy loss (as a result of σ = 0.15) is consistently marginal (<0.7%) across all
The results confirm that the device-to-device and cycle-to-cycle variations of our device instances are well controlled, such that the accuracy of a real-world recognition system will not be noticeably affected in the presence of variations. As a reference, when $\sigma = 0.5$, a significant drop in the accuracy can be observed. In addition, larger noise levels result in a more significant drop. For the cases of $\sigma = 1$ and 2, the resulting accuracy is rather unacceptable. Based on the experimental results and our observation, we are certain that the device uniformity plays a significant role in the success of RC-based recognition tasks. Owing to the highly uniform characteristics of our device, we achieved promising recognition accuracy, even in the case of exceedingly noisy data. Therefore, our device has a tremendous potential for real-world applications.

After elucidating the impact of variation and noise on recognition accuracy, we delved deeper into the interaction between these two properties from the perspective of software simulation. Namely, we set $\sigma$ to 0.15 and investigated two different training approaches: variation-aware training and variation-unaware training. Furthermore, we applied different percentages of noise to the testing images, ranging from 0% to 25%, to perform an in-depth testing of the two training approaches. For clarity, variation-aware training means that during the training process, the device variation is already considered. Therefore, we will
apply the variation to the reservoir outputs of all the training images, by setting the \( \sigma \) value that matches the degree of device variation. On the other hand, variation-unaware training implies that no variation is added to reservoir outputs. In other words, the \( \sigma \) value is set to zero during the training process. Figure S12 and Table S4, Supporting Information, present the experimental results. It is evident that the variation-aware training approach outperformed the other approach consistently. Moreover, as the noise percentage grows, the gap between the two lines increases. For instance, at 0% noise, the gap between the two lines is \( \approx 1\% \), and at 8% noise, it increases to \( \approx 1.5\% \), following which it continues to increase and reaches an accuracy difference of 3.5% at 25% noise. Therefore, based on the observations, variation-aware training can be concluded to be the superior approach for simulations with imperfect devices. With this approach, we can not only achieve higher recognition accuracy across the broad, but also reduce the accuracy loss caused by noisy images.

Up to this point, we already showed various promising results and valuable findings about the proposed RC-based system. However, it came to our attention that we were using a naïve assignment strategy, which assigns input patterns (from image to device’s RC conductance (from measurement)) in unaltered order, that is, input pattern \((0000)\) to device’s RC conductance of \((0000)\), input pattern \((0001)\) to device’s RC conductance of \((0001)\), etc.

We believe that with a more careful assignment, better performance for any given recognition task can be expected. Therefore, to further explore the potential of the RC-based system, we came up with a well-designed assignment strategy, as described in the sequel. First of all, we define the importance of an input pattern: if a 4 bit input pattern occurs more frequently and thus had better to be distinguished/recognized from other patterns (for better recognition of the whole image), it is more important; input patterns \((0001), (0111), (1000), (1100), (1110)\) are such important patterns. Otherwise, input patterns such as \((0010), (1001), (1010), (1101)\) are relatively less important. We can then obtain the importance of each pattern based on simple statistics and, finally, assign a larger RC conductance to a more important input pattern. By doing so, inputs patterns “which are more important and need to be clearly distinguishable” can be better distinguished/recognized. Here, we would like to emphasize again that, in the original naïve assignment strategy, RC conductance was not assigned based on some sort of fancy strategy. Instead, RC conductance was assigned in an unaltered order, no matter whether an input pattern is complicated/important or not.

Figure 5e and Table S5, Supporting Information, present the experimental results. By comparing Figure 5e (Table S5, Supporting Information) with Figure 5d (Table S3, Supporting Information), it is clear that under every single experimental configuration, the new assignment strategy outperformed the original naïve assignment strategy. Meanwhile, it is worth mentioning that the new assignment strategy is particularly beneficial to noisy images. As can be compared between Figure 5e (Table S5, Supporting Information) and Figure 5d (Table S3, Supporting Information), when \( \sigma = 0 \) and noise = 0%, there is a 0.75% accuracy improvement. However, when \( \sigma = 0 \) and noise = 25%, the accuracy improvement is up to around 7.5%, which is much more significant than the noiseless case. In summary, with the newly designed assignment strategy, we not only improved the overall recognition accuracy, but also strengthened the capability of fault tolerance in an RC-based system.

On a final note, given 1) a specific application of RC, 2) \( N \) for dividing each row of input data into multiple \( N \)-bit sections; and 3) measurement of \( 2^N \) RC conductance, there exists an optimal assignment leading to the best accuracy. Finding the optimal assignment efficiently is an algorithmic problem and not particularly addressed in this work.

However, to showcase the strength of our device and our training methodology in a more detail manner, we presented two additional sets of pairwise comparison. In both these sets, as shown in Figure 6, we presented the testing images and their corresponding distributions of output probabilities under different configurations. In the output probability distributions/histograms, the probabilities \((y\text{-axis})\) of ten output neurons represent the probabilities of recognizing the testing image as digit 0–9 \((x\text{-axis})\). Meanwhile, the highest probability determines the result of recognition.

The first set of comparison is between the simulation results of two different training methodologies: variation-aware and variation-unaware. In the training-aware methodology, device instances in the system are assumed to be imperfect and differing from one another. The deviation across all device instances follows a normal distribution with a deviation of \( \sigma = 0.15 \). In contrast, the variation-unaware methodology assumes the ideal case, where all device instances are (and behave) exactly the same. As shown in Figure 6a,b, with variation-aware training, the system can recognize the testing images accurately, even when the image is exceedingly noisy. However, with variation-unaware training, the system failed to do so. Therefore, it is clear that when it comes to RC-based simulation, variation-aware training is superior to variation-unaware training and compensates for the negative impact on variation-induced accuracy loss. A similar experiment is shown in Figure S13, Supporting Information.

The second set of comparison is between simulation results using devices with two different degrees of variation. To be precise, the two \( \sigma \) values of the devices were set to 0.15 and 0.5, and the variation-aware training methodology was used. As shown in Figure 6c,d, the recognition results are incorrect when \( \sigma = 0.5 \). On the other hand, when \( \sigma = 0.15 \), the system recognizes the same images correctly. Based on the observation mentioned earlier, we come to the conclusion that even with variation-aware training (the superior training methodology), when the device variation surpasses a certain degree, the resulting accuracy loss cannot be compensated. Therefore, inevitably, devices with large variation will lead to less satisfactory recognition accuracy. Another example is showcased in Figure S14 Supporting Information. Overall, the variation of our device is relatively small, with the \( \sigma \) value not exceeding 0.15. In conjunction with variation-aware training, we can realize an RC-based system with exceptional overall performance, which exhibits two highly desirable aspects: 1) marginal loss in accuracy and 2) significant saving in energy consumption.

In summary, we fabricated an all-vacuum-deposited perovskite/MoO$_3$/Ag artificial synaptic device. The device was characterized by an analogue switching behavior, broad dynamic
Figure 6. a,b) Output probability distributions of an image (digit “2”) with a) 0% noise and b) 25% noise. The output probabilities were obtained by two different training methodologies: variation-aware training and variation-unaware training. The image is recognized correctly as “2” with variation-aware training (tallest red bars in both histograms) and is recognized incorrectly as “13” with variation-unaware training (tallest gray bars in both histograms). c,d) Output probability distributions of an image (digit “6”) with c) 0% noise and d) 25% noise. The output probabilities were obtained by two different degrees of device variation: $\sigma = 0.15$ and $\sigma = 0.5$. The image is recognized correctly as “6” when $\sigma = 0.15$ (tallest red bars in both histograms) and is recognized incorrectly as “4” when $\sigma = 0.5$ (tallest gray bars in both histograms).

range, repeatable resistance control, and low operation voltage. All its parameters exhibited highly desirable values, comparable with those of the state-of-the-art artificial synapses. Moreover, the device exhibited crucial synaptic features, such as PPF, SRDP, and STP-to-LTP behavior. The results of TEM and EDS investigations performed on the AgI layer of the device revealed that the diffusion of Ag$^+$ and I$^-$ ions was the basis of the analogue synaptic behavior of the device. The measured parameters of the device were used for the simulation of RC-based handwritten digit recognition. By utilizing a much smaller $196 \times 10$ single-layer network, we demonstrated a surprisingly high accuracy of the recognition results, even with substantially high input noises. The results flaunted the advantage of device uniformity and low cycle-to-cycle variation of the vacuum-deposited perovskite artificial synapses. The encouraging results of the simulation and the promising device performance exhibit the strong potential of the proposed all-vacuum-deposited perovskite artificial synapse for the development of future neuromorphic computing networks.

Experimental Section

Device Fabrication: The perovskite precursors (CsBr (99.999% purity, Alfa Aesar), PbI$_2$ (99.985% purity, Alfa Aesar), PbBr$_2$ (99.998% purity, Alfa Aesar), MoO$_3$ (99.95% purity, Alfa Aesar), Ag (99.999% purity, Gredmann), AgI (99.999% purity, Alfa Aesar), Au (99.999% purity, Gredmann), NiO (>95% density, Nikko-Materials), Fullerene (C$_{60}$, 99.9% purity, Nichem), and 1,1-bis[(di-4-tolylamino)phenyl]cyclohexane (TAPC; >99.5% purity, Lumtec) were used as purchased without further purification. ITO-coated glass was used as the substrate, and the vacuum-deposited layers were fabricated in a high-vacuum chamber with a base pressure of $\approx 2 \times 10^{-5}$ Torr. The deposition rate for all materials was maintained at $1-5$ Å s$^{-1}$. Vacuum-deposited perovskite layers were synthesized by co-evaporation of metal halide precursors.$^{[45]}$ All the deposition rates were monitored by quartz crystal microbalance sensors. The area of each sample was 500 x 500 µm$^2$ and was defined by shadow masks.

Device Characterization: The electrical characteristics of the devices were measured in a nitrogen glove box using a Keithley 2636B SourceMeter and a Keysight B1500A semiconductor device analyzer with the B1330A waveform generator fast measurement unit (WGFMU). All the characterization processes were carried out in a dark environment to avoid the effects of the photocurrents generated by the perovskite film. TEM lamella samples were prepared by dual beam focused ion beam (FEI, HELIOS-660). The final TEM lamella sample thickness was $\approx 50$ nm. TEM and EDS analyses were conducted using a JEOL JEM-2800 equipment.

Network Simulation: In the pre-processing stage, each original 28 x 28 gray-scale handwritten digit image was converted into a binary-pixel (black-and-white) counterpart. Following this, we divided each row of the image into smaller sections; by doing so, compared with directly feeding the entire row undivided into the reservoir, we can effectively reduce the number of possible input patterns from $2^{28}$ to a much smaller degree. This would significantly increase the distinguishability for the reservoir. Based on our optimization results, we decided to divide each row into seven 4-bit sections.

Based on the aforementioned pre-processing steps, for each handwritten digit image, the dynamic reservoir would generate a total of 196 reservoir outputs (7 sections x 28 rows). Therefore, the second part of the RC framework (the readout) can be realized by a 196 x 10 single-layer NN with 196 input neurons from the reservoir outputs and ten output
neurons corresponding to ten classes of digits (0–9). The readout function was trained with 60,000 images from the training dataset of MNIST for 4,000 epochs and tested with 10,000 images from the testing dataset of MNIST.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Authors L.-W.C. and W.-C.W. contributed equally to this work. This work was financially supported by the Ministry of Science and Technology of Taiwan (Grant Nos. 109-2636-E-007-009; 108-2218-E-009-046; 109-2634-F-007-023), Academia Sinica (AS-SS-109-05), and the “Frontier Research Center on Fundamental and Applied Sciences of Matters” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

artificial synapses, inorganic halide perovskites, neuromorphic computing, reservoir computing, vacuum deposition

Received: August 18, 2020
Revised: September 23, 2020
Published online: November 16, 2020

[1] S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P.-Y. Chen, H. Yeon, S. Yu, J. Kim, Nat. Mater. 2018, 17, 335.
[2] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, W. D. Lu, Nat. Commun. 2017, 8, 2204.
[3] C. S. Yang, D. S. Shang, N. Liu, E. J. Fuller, S. Agrawal, A. A. Talin, Y. Q. Li, B. C. Shen, Y. Sun, Adv. Funct. Mater. 2018, 28, 1804170.
[4] S. Seo, S.-H. Jo, S. Kim, J. Shim, S. Oh, J.-H. Kim, K. Heo, J.-W. Choi, C. Choi, S. Oh, Nat. Commun. 2018, 9, 1.
[5] F. Zhou, Z. Zhou, J. Chen, T. H. Choy, J. Wang, N. Zhang, Z. Lin, S. Yu, J. Kang, H.-S. P. Wong, Nat. Nanotechnol. 2019, 14, 776.
[6] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, IEEE Signal processing magazine 2012, 29, 82.
[7] T. N. Sainath, O. Vinyals, A. Senior, H. Sak, IEEE Inte. Conf. on Acoustics, Speech and Signal Processing, IEEE, Brisbane, QLD, Australia 2015, p. 4580.
[8] R. S. Brinkworth, P. A. Shoemaker, D. C. O’Carroll, Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing, IEEE, Melbourne, VIC, Australia 2009, p. 289.
[9] P. W. McCowan, C. Benton, J. Dale, A. Johnston, Int. J. Neural. Syst. 1999, 9, 429.
[10] K. M. Song, J-S. Jeong, B. Pan, X. Zhang, J. Xia, S. Cha, T.-E. Park, K. Kim, S. Finizio, J. Raabe, Nat. Electron. 2020, 3, 148.
[11] Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H.-S. P. Wong, Nat. Electron. 2018, 1, 458.
[12] Z. H. Tan, R. Yang, K. Terabe, X. B. Yin, X. D. Zhang, X. Guo, Adv. Mater. 2016, 28, 377.

© 2020 The Authors. Published by Wiley-VCH GmbH
