Two-dimensional materials for artificial synapses: toward a practical application

I-Ting Wang*, Chih-Cheng Chang, Yen-Yu Chen, Yi-Shin Su and Tuo-Hung Hou*

Department of Electrical Engineering and Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: itwang@nycu.edu.tw and thhou@mail.nctu.edu.tw

Keywords: 2D material, synapse, neuromorphic computing, deep neural network, hardware implementation

Abstract

Combining the emerging two-dimensional materials (2DMs) and neuromorphic computing, 2DM-based synaptic devices (2DM synapse) are highly anticipated research topics with the promise of revolutionizing the present Si-based computing paradigm. Although the development is still in the early stage, the number of 2DM synapses reported has increased exponentially in the past few years. Nevertheless, most of them mainly focus on device-level synaptic emulations, and a practical perspective toward system-level applications is still lacking. In this review article, we discuss several important types of 2DM synapses for neuromorphic computing. Based on the cross-layer device-circuit-algorithm co-optimization strategy, non-ideal properties in 2DM synapses are considered for accelerating deep neural networks, and their impacts on system-level accuracy, power and area are discussed. Finally, a development guide of 2DM synapses is provided toward accurate online training and inference in the future.

1. Introduction

With ever-increasing data generation and the rising demand for data processing, conventional computing systems based on the von Neumann architecture can no longer satisfy the requirement of fast processing speed and high energy efficiency [1]. As a result, novel non-von Neumann architectures are now being actively investigated. Neuromorphic computing [2] is believed to be the most promising novel computing architecture to provide high energy efficiency and massive parallelism inspired by the human brain. Many synaptic devices, which act as crucial components in neuromorphic computing, have been realized using several types of nonvolatile memories (NVMs), including resistive switching memory (RRAM) [3, 4], phase-change memory (PCRAM) [5, 6], ferroelectric memory [7], etc for accelerating deep neural networks (DNNs) with in-memory computing capability [8]. While the three-dimensional bulk materials (3DMs) are commonly adopted in the aforementioned synaptic devices, 2D materials (2DMs) have drawn increasing attention in recent years, and a large number of 2DM-based synaptic devices (2DM synapses) have been reported. The use of 2DMs as the semiconducting channel with an atomically thin thickness and a larger bandgap compared to the traditional Si counterpart not only offers better immunity to short channel effects by providing superior gate control but also lowers the leakage current, making it suitable for low-power applications [9]. Moreover, the dangling-bond-free surface promises excellent charge transport. Controlling the number of layers in 2DMs allows tunable bandgaps that provide flexibility in diverse applications [10]. The newly available synthesis and advanced device fabrication methods enrich the structural design of 2DM synapses including van der Waals heterostructures [11] and 3D monolithic integration [12]. The wafer-scale integration of 2DMs enables high-density electronic circuits for more complex computing tasks [13]. These highly scalable, tunable, and stackable properties make 2DM synapses promising candidates for neuromorphic computing applications.
Although numerous 2DM synapses have been proposed and demonstrated, their development for neuromorphic computing is still in the early stage. Whether the reported synaptic behaviors of those 2DM synapses meet the requirement for practical applications remains in doubt. Many excellent review papers have systematically summarized 2DM synapses from the perspectives of materials and devices [14–25]. However, a comprehensive and quantitative study that connects to system-level requirements is still lacking. In this article, we first introduce the recent advances in 2DM synapses for neuromorphic computing applications. We categorize 2DM synapses by their device structures and operating mechanisms and discuss their synaptic characteristics. To evaluate the readiness of 2DM synapses for practical applications, the main focus of this work, we adopt a cross-layer device-circuit-algorithm co-optimization strategy, where the influence of non-ideal device properties such as limited precision, nonlinear weight update, intrinsic device variation, etc could be connected to the system-level specifications, namely performance (accuracy), power, and area, for image classification tasks [26]. Based on the co-optimization results with the special emphasis on accurate online training and inference in DNNs, not only the current status but also the future research directions of 2DM synapses will be provided in this work.

2. 2DM synapse and categorization

In neuromorphic computing systems, an artificial synapse performs multiplication of the voltage input signal given from the pre-neuron and its present synaptic state, and the product output signal is then transmitted to the post-neuron. Such multiplication can be easily realized in crossbar memory arrays using Ohm’s law [27], which at the same time promises high-density integration and massively parallel computation [28]. Therefore, implementing learning functions in synaptic devices becomes a prerequisite for neuromorphic applications [29], which ultimately aim for constructing a replica of the human brain. Synaptic plasticity is one of the key features to be obtained, and it includes potentiation and depression of synaptic weight that describes the strengthening and weakening of the connection between neurons, respectively [30]. Figure 1(a) shows the trends in the annual counts of the number of published papers focusing on 2DM synapses. The total number of reported 2DM synapses has increased exponentially within a few years, suggesting an ever-growing interest in this field. Similar to the 3DM-based synapses that have been widely realized in several types of devices [31–33], here, based on the device structure, the 2DM synapses could be categorized into the two-terminal metal–insulator–metal (MIM)-based synapses and three-terminal transistor-based synapses accordingly. The former is generally a memristive 2DM synapse showing resistive switching (RS) by applying bias. The latter is usually based on gate-modulated conductance change in the transistor structures to represent synaptic plasticity. Herein, various 2DMs are adopted in the 2DM synapses. Figure 1(b) manifests the percentage of various 2DMs in the reported 2DM synapses, where MoS₂ and graphene are most popular due to the maturity in synthesis. The ‘others’ category includes a wide variety of 2DMs, such as WS₂, MoTe₂, ReS₂, Ti₃C₂, etc. Interestingly, the proportion of this ‘others’ category of 2DMs has increased from 19.72% in 2019 [34] to 35.2% in 2021, suggesting the growing interest in exploring the rich material database of 2DMs in this emerging application.
Figure 2. Structural difference between (a) two-terminal MIM-based and (b) and (c) three-terminal transistor-based 2DM synapses. Each layer may be substituted by 2DMs. The structure in (b) is especially for the charge-storage transistor-based 2DM synapse, where the charge-storage layer (also called floating gate) is indicated. The structure in (c) is for other transistor-based 2DM synapses not relying on the additional charge-storage layer.

Table 1. Selected two-terminal MIM-based 2DM synapses based on different RS mechanisms have been proposed, where 2DMs play roles in different layers of MIM structures.

| Involved 2D material | MIM stack | Switching mechanism | References |
|----------------------|-----------|---------------------|------------|
| Graphene             | Al/AlO\(_x\)/graphene | Filament | [59] |
|                      | Au/2D perovskite/graphene | Filament | [60] |
|                      | Ag/BNO\(_x\)/graphene | Filament | [61] |
|                      | Al/AlO\(_x\)/graphene | Filament | [62] |
|                      | Ir/Gd\(_2\)O\(_y\)/graphene | Filament | [63] |
|                      | Ta/TaO\(_x\)/AlN/graphene | Filament | [64] |
|                      | Graphene/WSe\(_{2-x}\)O\(_y\)/graphene | Barrier modulation | [65] |
| h-BN                 | Metal/h-BN/metal | Filament | [66] |
|                      | Ag/h-BN/Pt | Filament | [67] |
| MoS\(_2\)            | Cu/MoS\(_2\)/Au | Filament | [68] |
|                      | Ag/MoS\(_2\)/Pt | Filament | [69] |
|                      | Ag/MoS\(_2\)/Pt | Filament | [70] |
|                      | Ag/MoO\(_3\)/MoS\(_2\)/Ag | Barrier modulation | [71] |
|                      | Au/Ti/MoS\(_2\)/Ti/Au | Barrier modulation | [72–74] |
|                      | Au/Ti/MoS\(_2\)/Ti/Au | Phase transition | [75] |
|                      | Au/Li\(_x\)MoS\(_2\)/Au | Phase transition | [76] |
| WSe\(_2\)            | Ag/WSe\(_2\)/Ag | Filament | [77] |
|                      | Graphene/WSe\(_{2-x}\)O\(_y\)/graphene | Barrier modulation | [65] |
| WS\(_2\)             | Ag/ZrO\(_x\)/WS\(_2\)/Pt | Filament | [78] |
|                      | Pd/WS\(_2\)/Pt | Barrier modulation | [79] |
| MoTe\(_2\)           | Ti/MoTe\(_2\)/Au | Filament | [80] |

Although optical synaptic responses [14, 16, 35–57] have also been presented in 2DM synapses, in the following sections, we will only focus on the electronic synapses because of their relatively mature architecture and peripheral circuit design for accelerating most DNN problems in practical applications.

### 2.1. Two-terminal MIM-based 2DM synapse

Figure 2(a) illustrates the general structure of MIM-based 2DM synapse, which is similar to several two-terminal memristors such as RRAM and PCRAM using the typical 3DMs. The compact two-terminal structure offers the ultimately scaling capability to construct synaptic crossbar arrays with the highest possible density [58]. Each layer in the MIM structure may be replaced by 2DMs. For instance, table 1 lists several representative MIM-based 2DM synapses. The semimetallic graphene layer can serve as conducting electrodes [59–65], while the insulating h-BN [66, 67] and semiconducting MoS\(_2\) [68–76], WSe\(_2\) [65, 77], WS\(_2\) [78, 79], MoTe\(_2\) [80], etc are commonly adopted as RS layers sandwiched between two electrodes. The conductance change of MIM-based 2DM synapses involve a wide range of switching mechanisms, including filament formation and rupture [59–64, 66–70, 77, 78, 80–83], barrier modulation [65, 71–74, 79], phase transition [75, 76], and ferroelectric polarization [84–86]. Similar to the conventional filamentary RRAM, the filament formation and rupture in 2DM synapses relies on mobile ions or vacancies from the metal electrodes or RS thin films that are driven by the applied electric field. On the other hand, unlike the locally formed filament, the current conduction in the barrier modulation mechanism is rather uniform, where the modulated Schottky barrier at the 2DM/electrode interface is affected by the migration of charged defects. As for phase transition in the 2DM synapse, it requires the intercalation process by metal ions such as Li\(^{+}\) to initiate the phase transition of 2DMs.
The ion-induced transition between 2H (semiconductor) and 1T′ (metal) phases in MoS2 is a good example [75, 76]. Moreover, several 2DMs such as α-In2Se3 [84], CuInP2S6 [85], and SnS [86] are reported showing spontaneous polarization. These ferroelectric 2DMs form ferroelectric tunnel junctions to represent another category in the MIM-based 2DM synapses based on their switching mechanisms.

2.2. Three-terminal transistor-based 2DM synapse

The transistor-based synapse has been long adopted since the early stage of neuromorphic implementation by using a traditional Si-based charge-storage transistor [29]. Compared to the two-terminal MIM-based synapse, the three-terminal transistor-based synapse is less compact. However, it offers another degree of freedom not only in structural design but also in read and write operating methods for synaptic property tuning due to the multiple terminals [87, 88]. Figures 2(b) and (c) illustrate the general structures of transistor-based synapses, where 2DMs may play roles in certain layers of transistor-based synapses. For instance, table 2 summarizes several reported transistor-based 2DM synapses, which are further categorized into the charge-storage transistor [89–95], electrolyte-gated transistor [96–99], ferroelectric-gated transistor [49, 100–103], and memristor [104, 105] according to different operating mechanisms. The 2DMs with superior carrier mobility are most often adopted as the channel layer. The insulating h-BN is reported with fewer charge impurities and better uniformity [106], which is suitable as the tunneling barrier [89, 91–94]. The graphene commonly acts as the floating gate [89–92] that stores different amounts of charges to represent analog synaptic weights. Moreover, it can be adopted as a thin ionic tunneling layer between the channel and electrolyte for creating a stable interface in the WO3-based electrolyte-gated 2DM synapse [98]. The electrolyte at the gate terminal provides gate-controlled ions that modulate the channel conductance through the intercalation process. As for the ferroelectric-gated 2DM synapse, the spontaneous polarization switching may arise from the conventional ferroelectric materials [49], the newly emerged HfO2-based ferroelectric materials [100], or the 2DM itself [101, 102]. On the other hand, the memristor-based 2DM synapse is a hybrid structure combining both the memristor and transistor [104], and the defects migrating along the grain boundaries of polycrystalline MoS2 result in the modulation of conductance [104, 105]. In the next section, we will focus on the synaptic plasticity realized in various 2DM synapses.

3. Synaptic plasticity in 2DM synapse

Several forms of synaptic plasticity can be found in 2DM synapses. The short-term plasticity such as paired-pulse facilitation/depression describes the decay of weight modulation immediately after removing stimuli. The long-term plasticity including spike-timing-dependent plasticity (STDP) and long-term potentiation/depression (LTP/LTD) show longer sustainability in the weight modulation. Although the emulation of short-term and long-term plasticity are both important for paving the way toward neuromorphic computing, long-term plasticity is more commonly adopted in DNNs. This is because the time-independent weight modulation is more compatible with well-developed learning rules in DNNs [107]. Besides, STDP is reported as an important learning rule by encoding the spike-based information in the time domain, which is favorable for spiking neural networks (SNNs). SNN is promising for constructing a computing system resembling the biological neural network with low power consumption and high energy efficiency [1]. However, the lack of global learning architecture supporting the STDP learning algorithm still hinders its development [32]. On the contrary, DNN has been long developed with optimized backpropagation algorithms [108], and it is now matured for most artificial intelligence applications. Therefore, in this section, we will mainly discuss the long-term synaptic plasticity, especially in LTP and LTD for weight update, where the adjustable multilevel memory states are preferable to practical DNN hardware implementation. The representative experimental results of the 2DM synapses are reviewed.

3.1. Long-term potentiation and depression

Figures 3(a)–(d) show the typical weight updates in several 2DM synapses [75, 84, 86, 92]. By applying consecutive pulses for LTP and LTD, multilevel weight modulation can be achieved bidirectionally. The tunability of LTP and LTD characteristics depends on the process and operating conditions. For example, various weight update characteristics with different conductance ranges and nonlinearity can be obtained in a MoS2 synapse by using different types of ions for intercalation [75]. Moreover, the tuning of weight update can be realized by modulating the amplitude [86, 92], width, and number [84, 92] of the input pulses. Reducing pulse amplitude and pulse width show an improvement in nonlinearity, and it is also effective when further decreasing the pulse number to achieve a perfectly linear weight update [92]. However, the consequential degradation in dynamic range and thus weight precision need to be taken into consideration.
Table 2. Several transistor-based 2DM synapses categorized into charge-storage transistor, ferroelectric-gated transistor, electrolyte-gated transistor, and memtransistor types. The 2DMs with their superior properties are involved in different layers of each device.

| Channel | Tunneling layer | Charge-storage layer (floating gate, FG) | Blocking layer | Gate electrode | Source, drain electrodes | References |
|---------|-----------------|------------------------------------------|----------------|----------------|--------------------------|------------|
| MoS₂    | h-BN            | Graphene                                | SiO₂           | Si             | Cr/Au                    | [89]       |
| MoS₂    | Al₂O₃           | Graphene                                | HfO₂           | Ti/Au          | Ti/Au                    | [90]       |
| MoS₂    | h-BN            | Graphene                                | SiO₂, h-BN     | Si, Au (double-gate) | Ni/Au                    | [91]       |
| MoS₂    | h-BN            | Graphene                                | N/A            | N/A            | Graphene                 | [92]       |
| MoS₂    | h-BN            | Au                                      | h-BN           | Graphene       | Ti/Au                    | [93]       |
| MoS₂    | h-BN            | MoS₂, Au (double-FG)                    | SiO₂           | Si             | Cr/Au                    | [94]       |
| Pentacene | TiO₂          | Ti₃C₂T₅                                 | SiO₂           | Si             | [95]                     |

*The device is based on a two-terminal floating-gate memory structure without a gate terminal.

| Channel | Gate Dielectric | Gate Electrode | Source, Drain Electrodes | References |
|---------|-----------------|----------------|----------------------------|------------|
| α-MoS₂  | Ionic liquid    | Au             | Cr/Au                      | [96]       |
| WSe₂    | Polymer electrolyte | Pd/Au         | Pd/Au                      | [97]       |
| WO₃     | Solid electrolyte/Graphene | Pt | Pt | [98]       |
| SnS₂ and reduced graphene oxide composites | Polymer electrolyte | Al | Au | [99]       |

| Channel | Gate dielectric | Gate electrode | Source, drain electrodes | References |
|---------|-----------------|----------------|----------------------------|------------|
| WS₂     | Pb(Zr₀.₅Ti₀.₅)O₃ (PZT) | SrRuO₃ | Au | [49]       |
| WS₂     | HFe₂₋ₓZrₓO₆ (HZO)/Al₂O₃ | Si | Ti/Au | [100]      |
| Fluorographene | Al₂O₃ | Al | Ni | [101]      |
| α-In₂Se₃ | Al₂O₃ | Si | Ti/Au | [102]      |
| Graphene | Ferroelectric polymer | Al | Cr/Au | [103]      |

| Channel | Gate dielectric | Gate electrode | Source, drain electrodes | References |
|---------|-----------------|----------------|----------------------------|------------|
| MoS₂    | SiO₂            | Si             | Ti/Au                      | [104]      |
| MoS₂    | HfO₂            | Ti/Au          | Ti/Au                      | [105]      |

*The graphene layer serves as an insertion layer between WO₃ channel and gate electrolyte.
3.2. Reliability and variability properties

The reliability and variability are crucial issues to be considered for any practical application, they are however less discussed in most 2DM synapses. Nevertheless, there is still no clear consensus on the reliability requirement in synaptic devices because it is strongly application-specific. Since the synaptic device exhibits both memory and computing functions in DNNs, its reliability evaluation could be similar to that of the NVMs [109], where endurance and retention are the two key metrics [110]. Figures 3(e) and (f) show the stable cycling measurement of weight updates in several 2DM synapses [76, 90]. Although non-identical rather than identical input pulses were applied, this cycling measurement still provides useful information for evaluating the endurance of synaptic devices. Besides, the cycling behavior of the WO3-based electrolyte-gated 2DM synapse is optimized by inserting a graphene layer between the electrolyte and WO3 channel, as shown in figures (g) and (h) [98]. The inserted graphene successfully suppresses the self-diffused Li\textsuperscript{+} ions into the channel and prevents unstable states. Therefore, obvious improvement in the weight update can be observed. Moreover, temporal variation in the weight update is commonly found in the reported 2DM synapses, but this cycle-to-cycle (CtC) variation is rarely the main focus because usually the higher priorities are given to the demonstration of a large number of analog states and stable cycling of LTP/LTD in the currently reported 2DM synapses. Examining the device-to-device (DtD) variation in 2DM synapses is even scarcer. Although CtC and DtD variations are seldom reported in current 2DM synapses, their impact on DNN training and inference may not be negligible and will be discussed in the next section. Figures 4(a) and (b) demonstrate outstanding cycling endurance for evaluating variation in \(\alpha\)-In\textsubscript{2}Se\textsubscript{3}-based and SnS-based ferroelectric 2DM synapses with a CtC variation of 1.91\% [102] and 6.27\% [86], respectively. Moreover, the 5.95\% DtD variation among 10 devices is further reported in [86], supporting its relatively stable and uniform characteristics. Although both devices present superior cycling endurance, the stable and symmetric weight updates were all obtained using non-identical input pulses [86, 102], which may require a write-and-verify programming scheme. Such a complex programming scheme not only increases the additional operational overhead but also compromises the advantages of high parallelism and energy efficiency in the hardware implementation of DNNs.

The retention of 2DM synapses is more frequently demonstrated compared to the variability. However, the reported retention properties are usually obtained from the DC-cycled high resistance and low resistance states that possess a much larger conductance change than those used in synaptic operations [76, 80, 111]. Analyzing the retention properties obtained by AC pulses is preferable. Figure 5(a) demonstrates the retention of 2000 s obtained using AC pulses in the MoS\textsubscript{2}-based charge-storage 2DM synapse [90]. However, only two synaptic states were characterized using stronger pulses, 35 pulses of \(-10\ V/10\ \mu s\) and \(8\ V/10\ \mu s\) for potentiation and depression, respectively, which is different from its normal programming condition in LTP (\(-10\ V/1\ \mu s\)) and LTD (\(8\ V/1\ \mu s\)). A similar case can be found in the retention behavior of the WO3-based electrolyte-gated 2DM synapse illustrated in figure 5(b) [98]. Although each of the five synaptic states maintains a retention time of 10 s, these synaptic states were programmed using incremental pulses, which is different from the
Figure 4. Reported cyclic LTP and LTD in (a) α-\(\text{In}_2\text{Se}_3\)-based \cite{102} and (b) SnS-based \cite{86} ferroelectric 2D synapses, showing superior endurance with the maximum number of allowed pulses exceeding 4000 and 10,000, and low CtC variation of 1.91\% and 6.27\%, respectively.

Figure 5. Reported retention of 2D synapses obtained by AC pulses. (a) The MoS\(_2\)-based charge-storage 2D synapse presents two synaptic states with the retention of 2000 s by applying 35 consecutive LTP and LTD pulses \cite{90}. (b) The WO\(_3\)-based electrolyte-gated 2D synapse shows the retention of 10 s in five synaptic weight states by using pulses with incremental amplitudes \cite{98}. (c) The MoS\(_2\)-based charge-storage 2D synapse reports the retention of 120 s for the 131 distinguishable states by applying different numbers of pulses \cite{92}.

favorable identical pulses used for weight update. Therefore, evaluating the retention of each synaptic weight state obtained using operating conditions similar to that used in the weight update is highly recommended. In figure 5(c), the 131 distinguishable synaptic weight states between 0.1 nS to 13 nS are successfully demonstrated in the MoS\(_2\)-based charge-storage 2D synapse, and the reported retention of each state is 120 s \cite{92}. Although some detailed information of the measurement is not disclosed, this retention measurement could be by far one of the most practical demonstrations toward real applications. However, a longer retention time of each synaptic weight state is required especially for inference-based applications. Table 3 summarizes several key device parameters obtained in representative 2D synapses \cite{72, 75, 76, 80, 84, 86, 90, 92, 96–99, 102, 111} from different categories. There is still plenty of room for exploration and improvement in 2D synapses, especially in statistic measurement and reliability analysis.

4. Discussion and prospects

While the development is still in the early stage, a quantitative analysis of the requirement of 2D synapses from the system-level perspective for practical hardware implementation is still lacking. Whether the current 2D synapses are adequate or which part of characteristics needs further improvement for real applications is still under scrutiny. Therefore, in the following sections, we will consider the impact of the key device parameters (listed in table 3) on DNN applications based on our previously proposed methodology \cite{26}. We intend to analyze the device requirement of 2D synapses by connecting with the system-level specifications, such as accuracy, power, and area. In particular, we will focus more on how to realize accurate online training and inference in DNNs. In addition, we will highlight the importance of low state variation in synaptic devices. Finally, we will discuss the system-level accuracy during DNN online training and inference quantitatively. By taking into account state variation, we aim at providing a design guideline for the future development of 2D synapses toward practical applications.
Table 3. Reported performance of long-term synaptic plasticity (LTP and LTD) in different types of 2DM synapse with key parameters summarized.

| Type                        | Involved 2DMs | Weight precision (#) | Dynamic range | Pulse amplitude/width | Endurance (# of cycle (total pulses)) | Reliability | Device dimension | References |
|-----------------------------|---------------|----------------------|---------------|-----------------------|---------------------------------------|-------------|------------------|------------|
| Long-term synaptic plasticity (LTP and LTD) |               | Gmin, Gmax | Gmax/Gmin | LTP                  | LTD                                  | Retention   |                  |            |
| Filament                    | MoTe₂         | 200             | 200 μS, 300 μS | 2.5 | 0.6 | + 1.1 V/80 μs | − 1.2 V/80 μs | 2 cycles/800 pulses | N/A        | N/A | 0.1 μm² [80] |
| Filament                    | PbSe₂         | 100             | 400 μS, 700 μS | 1.75 | 0.7 | + 0.8 V/100 μs | − 0.9 V/100 μs | 3 cycles/6000 pulses | N/A        | N/A | 4 μm²² [111] |
| Barrier modulation MoS₂     |               | 1000            | 400 nS, 700 nS² | 1.75 | 0.64 | − 15 V/5 ms | + 10 V/2 ms  | N/A | N/A | L = 2 μm [72] |
| Phase transition MoS₂       |               | 50              | 65 μS, 140 μS | 2.15 | 0.7 | + 5 V/20 ms | − 5 V/20 ms | N/A | N/A | L = 10 μm [75] |
| Phase transition MoS₂       |               | 100             | 0.1 μS, 0.3 μS | 5 | 0.5 | + 4 V/1 ms | − 4 V/1 ms | 1000 cycles/40,000 pulses | N/A | N/A | L = 2 μm [76] |
| Ferroelectric α-In₂S₃       |               | 31              | 480 nS, 10 μS | 21 | 0.7 | + 2 V/80 ns | − 2 V/80 ns | 3 cycles/10 000 pulses | CIC = 6.27% | 5 states >950 s | L = 5 μm [86] |
| Ferroelectric SnS            |               | 100             | 6 nS, 130 nS | 21 | 0 ⁰⁻ | 1 to 3.5 V/20 ms | − 1 to − 3.5 V/20 ms | 50 cycles/10 000 pulses | CIC = 5.93% | by identical pulses | L = 5 μm [86] |
| Charge storage MoS₂, graphene, h-BN | 35          | 6.2 μS, 6.8 μS⁴ | 1.1 | 0.48 | − 10 V/1 μs | + 8 V/1 μs | 19 cycles/1330 pulses | N/A | 2 states 2000 s by 35 stronger pulses | L = 5 μm | 90 |
|                           |               | 20              | 1 nS, 13.5 nS | 13.5 | 0.81 | + 13 V/100 ms | − 13 V/100 ms | N/A | N/A | L = 3 μm [92] |
| Transistor-based            | MoS₂, graphene, h-BN | 50          | 1.5 nS, 9 nS | 6 | 0.6 | + 12 V/100 ms | − 12 V/100 ms | 55 cycles/5500 pulses | N/A | 131 states 120 s by different pulse number | W = 10 μm |
| 2DM synapse                 | Electrolyte α-MoO₃ | 70 nS, 95 nS² | 1.4 | 0.265 | + 2.5 V/1 ms | − 1.8 V/1 ms | 5 cycles/500 pulses | N/A | N/A | L and W in few μm [96] |
| Electrolyte WSe₂            |               | 60              | 260 pS, 560 pS | 2.2 | 0.3 | + 1.2 V/100 ms | − 0.4 V/100 ms | 3 cycles/360 pulses | N/A | N/A | Effective gate area [97] |
| Electrolyte Graphene        |               | 50              | 800 nS, 22 μS | 28 | 0 ⁰⁻ | + 3 V/1s | − 3 V/1s | 5 cycles/500 pulses | N/A | 5 states > 10 s by incremental pulses | L = 100 μm | [98] |
| Electrolyte SnS             |               | 50              | 1.25 μS, 1.45 μS | 1.16 | 0.32 | 0 to + 2 V/50 ms | 0 to − 0.2 V/30 ms | 36 cycles/3600 pulses | N/A | N/A | L = 100 μm | [99] |
| Ferroelectric α-In₂S₃       |               | 100             | 600 nS, 2.5 μS⁴ | 4 | 0 ⁰⁻ | − 3 to − 5 V/12 μs | 0.8 to 2.8 V/12 μs | 20 cycles/4000 pulses | CIC = 1.91% | N/A | L = 1.2 μm | [102] |
|                           |               | 50              | 800 nS, 22 μS | 28 | 0 ⁰⁻ | + 3 V/1s | − 3 V/1s | 5 cycles/500 pulses | N/A | 5 states > 10 s by incremental pulses | L = 100 μm | [98] |
|                           |               | 100             | 600 nS, 2.5 μS⁴ | 4 | 0 ⁰⁻ | − 3 to − 5 V/12 μs | 0.8 to 2.8 V/12 μs | 20 cycles/4000 pulses | CIC = 1.91% | N/A | L = 1.2 μm | [102] |

⁴Conductance value is calculated based on the provided current value and read condition.
⁵ANL: asymmetric nonlinearity, which is estimated based on the equation reported in [112].
⁶Non-identical pulse train is adopted.
⁷Assume ANL = 0 for symmetric weight update.
4.1. Power prospects
The pulse amplitude and pulse width for achieving LTP and LTD weight update in synaptic devices are directly related to the required power and speed during DNN online training. Although improving training accuracy is the main concern at the current development stage, operating the synaptic device using low voltages (\(<1\) V) and short pulses (\(<\mu s\)) are highly favorable for improving energy efficiency and latency. Moreover, the operating voltage needs to be compatible with the power supply voltage of peripheral driving circuits. In advanced CMOS, the power supply voltage has been scaled to less than 1 V. The short pulse width of less than \(\mu s\) should be a reasonable target because most 3DM NVMs claim to have a fast programming speed. However, most of the reported 2DM synapses require high operating voltages (\(>5\) V) and are with slow speed (\(\sim ms\)). Therefore, there exists plenty of room for further reducing the pulse amplitude and pulse width in 2DM synapses.

4.2. Area prospects
The dimension of synaptic devices directly affects the device density and area overhead. Realizing synaptic devices in nanometer scales is still highly recommended [31]. However, most reported 2DM synapses are still in micrometer scales. Although scaling the device area is practical for improving operating voltage, current, and speed, whether similar performance can be obtained in the scaled device needs to be carefully investigated [109].

4.3. Accuracy prospects in DNN online training
4.3.1. Weight precision
Classical LTP and LTD characteristics utilize the same total number of electrical pulses for both LTP and LTD. Applying one LTP/LTD pulse results in the increase/decrease of the synaptic weight state by one. Therefore, the maximum number of synaptic weight states determines the weight precision of a synaptic device. Weight precision is one of the key parameters that define the accuracy in DNN online training. Higher weight precision enables a more precise calculation and results in higher accuracy in DNNs. Although the weight precision of synaptic devices also has an influence on DNN inference, its requirement is much relieved compared to that in online training.

4.3.2. Asymmetric nonlinearity (ANL)
The ANL of a synaptic device is an important indicator for describing the asymmetry and nonlinearity of weight update during LTP/LTD, it is therefore crucial for DNN online training. In our analysis, ANL is defined based on the following equations [112]:

\[
G_P (n_P) = G_{min} + A \times \frac{n_P}{n_P + e^k}
\]
\[
G_D (n_D) = G_{max} - A \times \frac{(N - n_D)}{(N - n_D) + e^k}
\]
\[
A = (G_{max} - G_{min}) \times \left(1 + \frac{e^k}{N}\right)
\]
\[
ANL = \left[\frac{G_P \left(\frac{N}{2}\right) - G_D \left(\frac{N}{2}\right)}{G_{max} - G_{min}}\right]
\]

where \(G_P\) and \(G_D\) functions describe the behavior of synaptic weight modulation in LTP and LTD, respectively. \(n\) is the number of applied pulses. \(N\) is the maximum number of pulses allowed for both \(P\) and \(D\), which defines the weight precision. \(k\) is the fitting parameter. ANL is normalized to between zero and one by using the dynamic range (i.e. \(G_{max} - G_{min}\)). For a perfectly symmetric and linear LTP and LTD, ANL is zero. By contrast, ANL is one for an extremely asymmetric nonlinear LTP and LTD.

ANL is particularly important to DNN online training because of the frequent weight updates required in synaptic devices. By contrast, ANL is less concerned in DNN inference where the write-and-verify programming scheme could be applied with a negligible penalty because of the infrequent weight update. The existence of ANL results in uneven modulation of each synaptic state depending on the direction of LTP/LTD and the present synaptic state, and the cumulative effect on the weight value no longer follows a simple summation and subtraction rule [112]. As a result, the training accuracy degrades when ANL is increased. Note that the synaptic device showing symmetric weight update is less influenced by the nonlinearity. Even though the nonlinearity contributes to the uneven increase/decrease of weight update depending on the present state, the symmetric LTP/LTD provides comparable modulation and mitigates the training accuracy loss [112, 113].

4.3.3. Cycle-to-cycle (CtC) variation
The CtC variation defines the temporal random variation of the conductance state change in a synaptic device when applying the same input stimuli. The conductance values during the weight update process are randomly
affected by the CtC variations. The CtC variation is the main factor affecting online training accuracy. By contrast, the DtD variation is tolerable due to the self-adaptive nature of online training [114], where the weight of each synaptic device could be adjusted appropriately through a supervised learning process even with different weight update characteristics. In our analysis, the CtC variation is normalized to the total dynamic range [26], and we use the standard deviation (σ) to describe CtC variations of synaptic devices.

4.3.4. Endurance property

For endurance analysis, the total number of input pulses applied during the cyclic LTP and LTD measurements could be roughly seen as the maximum number of input pulses allowed during weight update. The endurance property is particularly important for DNN online training with frequent weight updates, but it has less impact on DNN inference. The maximum allowed number of input pulses should be high enough to guarantee successful training without early device failure.

4.3.5. Simulation results

To analyze the performance during DNN online training, we first consider that the synaptic devices are ideal without state variation. The accuracy of VGG-9 DNN training for CIFAR-10 classification under different ANL of synaptic devices is evaluated, as shown in figure 6(a). When a perfectly linear weight update (ANL = 0) is assumed, the accuracy of 88.3%, 90.8%, and 91.84% are obtained with weight precision of 64 (6 bit), 128 (7 bit), and 256 (8 bit), respectively. The higher weight precision results in higher accuracy. However, when the ANL is increased, a clear accuracy degradation could be found. The synaptic device with a lower weight precision (N = 64) shows an even worse immunity against the increase of ANL compared to that with a higher weight precision (N = 128 and 256). This accuracy degradation induced by ANL has been a well-known issue in synaptic devices [115, 116]. Without considering state variation, pursuing linear weight updates and high weight precision in synaptic devices is feasible to increase the training performance in DNNs.

However, when CtC state variation is present in a realistic case, drastically different conclusions are obtained. Figure 6(b) shows the training accuracy when considering CtC variation in the synaptic device with linear weight update (ANL = 0). The result suggests that the synaptic devices with a higher weight precision (N = 128 and 256) sustain the advantage of high accuracy when the CtC variation is low (σ < 2%); however, when the CtC variation is increased (σ > 2%), those with higher precision show severe accuracy degradation. On the other hand, the synaptic device with a lower weight precision (N = 64) is less sensitive to CtC variation because of the wider margin between each synaptic state. The synaptic device with 64 states (accuracy = 85.02%) even outperforms that with 256 states (accuracy = 76.92%) when the CtC variation is 5%.

When considering the existence of both CtC variation and ANL in a more realistic situation, the accuracy is calculated as shown in figure 6(c). Even though the synaptic device with lower weight precision (N = 64) has almost constant accuracy with increased CtC variation, it can no longer prevail in a large ANL case (0.6) because the baseline accuracy without CtC variation is seriously compromised. Therefore, pursuing a higher weight precision is beneficial only when the CtC variation is small or when ANL is high. One should carefully consider the non-ideal properties of the synaptic device to find the best trade-offs.

Despite of the importance of ANL, the CtC variation, and their interaction, very few 2DM synapse studies reported them simultaneously, which makes the evaluation on their online training potential difficult. The two ferroelectric 2DM synapses [86, 102] are the only few with reported CtC variation, and the cycling behaviors
Table 4. Specifications of synaptic device for achieving the best accuracy in DNN online training and inference applications.

| DNN application | Best achievable accuracya | Weight precision | ANL | Dynamic rangeb | CtC variation (σ) | DtD variation (σ) | Retention |
|-----------------|--------------------------|------------------|-----|----------------|------------------|------------------|-----------|
| Online training |                          |                  |     |                |                  |                  |           |
|                 | 91.84%                   | 256              | 0   | —              | 0%               | —                | >7825     |
|                 | 85.02%                   | 64               | 0   | —              | 5%               | —                | >1535     |
|                 | 80.94%                   | 128              | 0.6 | —              | 5%               | —                | >754      |
|                 | 91.03%                   | 2                | —   | 2/100          | —                | 0%               | 0%        |
|                 | 78.62%                   | 2                | —   | 2              | —                | 20%              | 0%        |
|                 | 78.15%                   | 2                | —   | 2              | —                | 20%              | −5%       |
| Inference       | 92.18%                   | 8                | —   | 2/100          | —                | 0%               | 0%        |
|                 | 88.83%                   | 8                | —   | 100            | —                | 20%              | 0%        |
|                 | 87.54%                   | 8                | —   | 100            | —                | 20%              | −5%       |

aAccuracy after training VGG-9 DNNs for CIFAR-10 classification.
bDynamic range is assumed with a fixed high resistance $(10^7 \Omega)$ of the synaptic device.

are shown in figure 4. The accuracy of 80.54% and 87.78% based on SnS-based [86] and α-In$_2$Se$_3$-based [102] ferroelectric 2DM synapses is identified, respectively, as indicated in figure 6(b). Both devices have weight precision of 100, and their symmetric nonlinear weight update is assumed to have ANL = 0.

Table 4 summarizes the specifications of synaptic devices for accurate online training under several situations mentioned above. The total number of pulses applied to each synaptic device after completing online training on CIFAR-10 is also provided accordingly. For the ANL = 0 cases, higher/lower weight precision results in finer/rougher calculation, thus requiring more/less numbers of weight updates. However, when ANL is considered (0.6 in this case), the value of synaptic weights is often modulated close to $G_{\text{max}}$ or $G_{\text{min}}$ due to the nonlinear LTP/LTD properties. As a result, the online training requires a less number of updates due to the early convergence to a less accurate model. Note that the required endurance is a minimum criterion. Any additional retraining increases the total number of pulses required. In general, the endurance of synaptic devices should be as high as possible to perform repeated training tasks. Higher weight precision enables a more precise calculation and results in higher accuracy in DNNs.

4.4. Accuracy prospects in DNN inference

4.4.1. Dynamic range

The dynamic range of synaptic devices describes the total modulated range of synaptic weight in conductance from the lowest $G_{\text{min}}$ to the highest $G_{\text{max}}$, and it can be defined by the conductance range of $G_{\text{max}} - G_{\text{min}}$ or by the ratio of $G_{\text{max}}/G_{\text{min}}$. The latter is similar to the definition of memory window. A higher dynamic range allows a larger margin in conductance between the multilevel states of synaptic devices, which is important when DtD variations are present. Besides, the dynamic range defines the magnitude of individual synaptic current and the summing current of multiple synapses on the same bit-line (BL) [26], which directly correlates with the power consumed in DNN systems. Synaptic devices with a higher conductance value result in higher power consumption. The dynamic range and conductance value of a synaptic device also limit the synaptic array size because the summing current of BLs must not exceed the maximum allowed current of the output sensing amplifier. Although a large array could be partitioned into multiple smaller ones to lower the summing current of each BL, the additional overheads in area and power are unavoidable for maintaining the same accuracy [117]. A more detailed study on the dynamic range of synaptic devices for the area and energy co-optimization in a feasible in-memory-computing design can be found in [26].

4.4.2. Device-to-device (DtD) variation

The DtD variation defines the spatial random variation of the conductance values in different synaptic devices when applying the same input stimuli, and it needs to be considered in DNN inference. By contrast, the CtC variation can be ignored here because there is no weight update during inference. In our analysis, the DtD variation is normalized to the mean conductance value [26], and we use the standard deviation $(\sigma)$ to describe DtD variations of synaptic devices.

4.4.3. Retention property

For conventional 3DM synapses, several retention failure modes with a wide range of time constants have been discussed and their impacts have been evaluated in [110]. Here, for simplicity, we mainly consider a global and permanent conductance drifting that typically occurs after a long retention time. Because the time constant of retention degradation considered here is long, it mainly affects inference but not training accuracy because frequent weight updates are performed during training, and maintaining weight states accurately for a long
time is not critical. By contrast, the retention time for inference prefers to be as long as possible. For 3DM NVMs, a 10 year retention time is usually targeted.

4.4.4. Simulation results
In inference applications, repetitive reading without modifying the conductance of synaptic devices is required. Therefore, the stability of each weight state is crucial because any deviation of the BL summing current may lead to accuracy degradation. To analyze the performance during DNN inference, the dynamic range of 2 and 100 times are assumed with a fixed maximum resistance ($10^7 \, \Omega$) of the synaptic device. The inference accuracy is shown in figures 7(a) and (b), respectively. When DtD variation = 0%, both dynamic ranges of 2 and 100 have the same baseline accuracy of 91.03%, 92.09%, and 92.18% for weight precision of 2 (1 bit), 4 (2 bit), and 8 (3 bit), respectively, because each state could be precisely defined regardless of the magnitude of dynamic range. This result also indicates that the requirement of weight precision for inference is much relieved compared to that for training. However, in the presence of DtD variation, the dynamic range becomes relevant because the margin between each synaptic state depends on both the DtD variation and the dynamic range. Synaptic devices with lower weight precision have better immunity against variation when the dynamic range is small and the margin between states is insufficient for the high weight precision. On the other hand, synaptic devices with higher weight precision show better accuracy when the dynamic range is large because higher weight precision improves the baseline accuracy and the margin between states is less a concern. However, to increase the dynamic range of synaptic devices unlimitedly is not recommended from the perspective of power and area.

The specifications of synaptic devices for accurate inference under several situations mentioned above are summarized in table 4. Here we further consider the state variation induced by the retention degradation of 5%. The retention degradation is defined by the ratio of the amount of drifting after a certain retention time to the total dynamic range, and the sign represents the drifting direction (positive/negative: toward higher/lower conductance). Therefore, one should carefully consider the non-ideal properties of synaptic devices to achieve the best possible inference accuracy.

5. Conclusions and outlook
The 2DM synapse is the union of the abundant research in 2DM and novel computing paradigm. With the increasing interest, current research on 2DM synapses focuses more on the emulation of synaptic functions in different 2DM systems but significantly less on necessary properties enabling practical DNN applications. Based on our analysis, we made the following recommendations for future research on 2DM synapses. First, state variability is arguably the most critical factor for accurate DNNs. Statistical data such as CtC/DtD variation should be provided for any 2DM synapse of interest. Even though the process maturity of 2DM synapses is less than that of the conventional 3DM counterparts, leveraging the unique properties of 2DMs could be the key to success. For example, ferroelectric 2DMs are discovered just recently [118] and are relatively new to be implemented as synapses. They often show superior CtC and DtD variation [84, 86, 102], probably due
to the stable spontaneous polarization in the large-domain or epitaxial ferroelectric 2DMs that cannot be easily achieved using thin ferroelectric 3DMs. Retention-induced state drifting is also rarely discussed in 2DM synapses, which is critical for inference applications. Second, although many studies focus on increasing the number of analog states and improving ANL, these properties are only important for online training. The device requirements for online training are much stricter. Even the state-of-the-art in-memory computing studies using matured memory technologies focus mostly on the inference applications where a significantly smaller number of states is sufficient and there exists no ANL and endurance constraints. Third, an excessively large dynamic range is usually not necessary because it usually comes with higher conductance and penalty in power and area. Therefore, combining the first and third points, improving state variability should be the first priority for achieving not only accurate but also efficient DNN hardware. Fourth, although scaling the device to the nanometer scale is not only beneficial for achieving a high-density crossbar synaptic array but also helps to improve operating voltage, current, and speed. Nevertheless, most reported 2DM synapses are with dimensions larger than micrometer scales. Besides, the array-level demonstration is seldom seen [75, 92, 111]. The device scalability and array integration of 2DM synapses are prerequisites for large-scale DNN hardware.

Overall speaking, discovering the best synaptic device for neuromorphic computing is a holistic optimization that has never been simple. The specification of synaptic devices is strongly application-specific, and trade-offs must be made with care. There is still plenty of room for improvement and innovation in the field of 2DM synapses. But with its current strong momentum, we anticipate continuing progress and blossom in 2DM synapse research for neuromorphic computing in the future.

Acknowledgments

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant 109-2221-E-009-020-MY3, 110-2221-E-A49-088-MY3, 110-2634-F-009-017, and 110-2622-8-009-018-SB.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Tuo-Hung Hou https://orcid.org/0000-0002-9686-7076

References

[1] Merolla P A et al 2014 Science 345 668–73
[2] Mead C 1990 Proc. IEEE 78 1629–36
[3] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P and Lu W 2010 Nano Lett. 10 1297–301
[4] Yu S, Wu Y, Jeyasingh R, Kuzum D and Wong H-S P 2011 IEEE Trans. Electron Devices 58 2729–37
[5] Kuzum D, Jeyasingh R G D, Lee B and Wong H-S P 2012 Nano Lett. 12 2179–86
[6] Burr G W et al 2015 IEEE Trans. Electron Devices 62 3498–507
[7] Jerry M, Chen P-Y, Zhang J, Sharma P, Ni K, Yu S and Datta S 2017 2017 Int. Electron Devices Meeting https://doi.org/10.1109/IEDM.2017.8268338
[8] Ielmini D and Ambrogio S 2019 Nanotechnology 31 092001
[9] Yoon Y, Ganapathi K and Salahuddin S 2011 Nano Lett. 11 3768–73
[10] Fiori G, Bonaccorso F, Iannaccone G, Palacios T, Neumaier D, Seabaugh A, Banerjee S K and Colombo L 2014 Nat. Nanotechnol. 9 368–45
[11] Jariwala D, Marks T J and Hersam M C 2017 Nat. Mater. 16 170–81
[12] Wang C-H, McClellan C, Shi Y, Zheng X, Chen V, Lanza M, Pop E and Wong H-S P 2018 2018 IEEE Int. Electron Devices Meeting https://doi.org/10.1109/IEDM.2018.8614495
[13] Chen S et al 2020 Nat. Electron. 3 638–45
[14] Zhou F, Chen J, Tao X, Wang X and Chai Y 2019 Research 2019 9490413
[15] Wang C Y, Wang C, Meng F, Wang P, Wang S, Liang S J and Miao F 2019 Adv. Electron. Mater. 6 1901107
[16] Feng X, Liu X and Ang K-W 2020 Nanophotonics 9 1579–99
[17] Wang B, Sun Y, Ding H, Zhao X, Zhang L, Bai J and Liu K 2020 Adv. Funct. Mater. 30 2003732
[18] Zhao Q et al 2020 Mater. Horiz. 7 1495
[19] Huh W, Lee D and Lee C H 2020 Adv. Mater. 32 2002092
[20] Seo S, Lee J-J, Lee H-J, Lee H W, Oh S, Lee J J, Heo K and Park J-H 2020 ACS Appl. Electron. Mater. 2 371–88
[21] Ko T-J et al 2020 Science 363 101676
[22] Cao G, Meng P, Chen J, Liu H, Bian R, Zhu C, Liu F and Liu Z 2021 Adv. Funct. Mater. 31 2005443
[23] Lee G, Baek J H, Ren F, Pearton S J, Lee G H and Kim J 2021 Small 17 2100640
[24] Liu X T, Chen J R, Wang Y, Han S T and Zhou Y 2021 Adv. Funct. Mater. 31 2004733
[96] Yang C S, Shang D S, Liu N, Shi G, Shen X, Yu R C, Li Y Q and Sun Y 2017 Adv. Mater. 29 1700906
[97] Zhu J et al 2018 Adv. Mater. 30 1800195
[98] Nikam R D, Kwak M, Lee J, Rajput K G and Hwang H 2019 Adv. Electron. Mater. 6 1901100
[99] Jang E-K, Park Y and Lee J-S 2019 Nanoscale 11 15382
[100] Chen L et al 2020 Adv. Electron. Mater. 6 2000057
[101] Liu B et al 2019 Adv. Mater. Technol. 4 1900422
[102] Wang L, Wang X, Zhang Y, Li R, Ma T, Leng K, Chen Z, Abdelwahab I and Loh K P 2020 Adv. Funct. Mater. 30 2004609
[103] Chen Y, Zhou Y, Zhuge F, Tian B, Yan M, Li Y, He Y and Miao X S 2019 npj 2D Mater. Appl. 3 31
[104] Sangwan V K, Lee H-S, Bergeron H, Balla I, Beck M E, Chen K-S and Hersam M C 2018 Nature 554 500–4
[105] Wang L et al 2019 Adv. Funct. Mater. 29 1901106
[106] Lee G-H, Yu Y-J, Lee C, Dean C, Shepard K L, Kim P and Hone J 2011 Appl. Phys. Lett. 99 243114
[107] LeCun Y, Bengio Y and Hinton G 2015 Nature 521 436–44
[108] Rumelhart D E, Hinton G E and McClelland J L (the PDP Research Group) 1986 Parallel Distributed Processing: Explorations in the Microstructure of Cognition vol 1 ed D E Rumelhart and J L McClelland (Cambridge, MA: MIT Press) pp 45–76
[109] Lanza M et al 2019 Adv. Electron. Mater. 5 1800143
[110] Chen P-Y and Yu S 2018 2018 IEEE Int. Reliability Physics Symp. https://doi.org/10.1109/IRPS.2018.8353615
[111] Li Y, Loh L, Li S, Chen L, Li B, Bosman M and Ang K-W 2021 Nat. Electron. 4 348–56
[112] Chang C-C, Chen P-C, Chou T, Wang I-T, Hudec B, Chang C-C, Tsai C-M, Chang T-S and Hou T-H 2018 IEEE J. Emerg. Sel. Top. Circuits Syst. 8 116–24
[113] Sun X and Yu S 2019 IEEE J. Emerg. Sel. Top. Circuits Syst. 9 570–9
[114] Chen P-Y, Gao L and Yu S 2016 IEEE Trans. Multi-Scale Comput. Syst. 2 257–64
[115] Chen P-Y, Lin B, Wang I-T, Hou T-H, Ye J, Vrudhula S, Seo J-S, Cao Y and Yu S 2015 IEEE/ACM Int. Conf. Comput.-Aided Design https://doi.org/10.1109/ICCAD.2015.7372570
[116] Yu S 2018 Proc. IEEE 106 260–85
[117] Wu T-Y et al 2019 2019 IEEE Int. Electron Devices Meeting https://doi.org/10.1109/IEDM.2019.8993565
[118] Belianinov A et al 2015 Nano Lett. 15 3808–14