Proceedings of the 4th International Conference on Functional Materials and Civil Engineering
DOI: 10.54254/2755-2721/2026.37906

Defect Engineering, Interface Design, and Heterostructure Integration: Three Material-Level Strategies for Low-Energy Synaptic Weight Update in MoS_2-Based Two-Dimensional Memristors

Yujie Shen[1,*], Yuhao Sun[2], Jiaqi Guo[3]

[1]*The Hill School, Pottstown, USA*
[2]*Suzhou International Acdemy BFSU, Suzhou, China*
[3]*Guangzhou Foreign Language School, Guangzhou, China*
**Corresponding Author. Email: yshen27@thehill.org*

Abstract. Practical artificial synapse operation relies on continuous conductance modulation, making electrical programming conditions decisive to device energy expenditure. In real synaptic updates, working voltage, operational current and pulse duration together determine energy consumption. This review compares three material-level engineering strategies for lowering the electrical cost of synaptic weight adjustment in MoS_2-based two-dimensional memristors: (1) argon-plasma defect engineering of monolayer MoS_2, (2) graphene-interface design in a graphene/MoS_2/SiO_x/Ni stack, and (3) an all-2D van der Waals heterostructure combining MoS_2, h-BN and graphene. Under ten +20 V and ten −20 V, 1 ms potentiation/depression pulses, argon-plasma-treated monolayer MoS_2 (sulfur-vacancy density ≈3.4×10^{14} cm^{-2}) reached a minimum programming energy of approximately 30 fJ per pulse and a switching ratio of 10^4. Under 64 pulses of +4.5 V and 64 pulses of −3 V, each 50 μs wide, a graphene-barrier interface restricted the programming current to as low as 100 nA and produced a non-linearity factor of 0 in both potentiation and depression under that specific pulse condition, though linearity degraded when the pulse parameters or channel thickness changed. Under ±15 V, 40 ns pulses, an all-2D floating-gate heterostructure reached approximately 18 fJ per pulse across roughly 3,000 quasi-continuous conductance states. Together, these three strategies show that current suppression, carrier-injection barriers, and short-pulse tunnelling control are complementary, material-level routes to lowering the electrical cost of synaptic weight update, each trading differently against operating voltage, pulse duration and update linearity.

Keywords: MoS_2, graphene, artificial synapse, two-dimensional materials, defect engineering, interface engineering, synaptic update energy

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37906

1. Introduction

Neuromorphic computing aims to replicate biological neural processing through artificially constructed neuron and synapse circuits. Conventional computing systems separate memory and processing units physically. Such separation inevitably induces frequent data exchange, which constitutes one major source of system-level energy waste. Artificial synapses offer a feasible alternative by encoding synaptic weight within device conductance, allowing computation and storage to function within the same hardware framework [1, 2]. For real industrial implementation, how much energy is consumed during repeated conductance tuning largely determines the practical usability of neuromorphic devices.

Traditional artificial synapses are commonly developed based on standard CMOS circuits, resistive memory or phase-change switching materials. Despite mature technical accumulation, current resistive synaptic devices still exhibit noticeable non-ideal behaviors. These include large programming current, imbalanced and non-linear weight update trends, as well as inconsistent performance across operational cycles and individual chips [2]. It is worth clarifying that these issues represent widespread engineering challenges rather than fundamental flaws of conventional device principles. Severe energy waste also occurs when devices operate under high voltage, excessive current or prolonged pulse excitation.

For an approximately rectangular electrical pulse, the energy consumed during one synaptic update is commonly approximated as the product of the operating voltage, the average current, and the pulse duration. Because real electrical pulses often show non-ideal, transient waveforms rather than perfect rectangles, a more accurate estimate instead integrates the instantaneous voltage-current product over the full duration of the pulse. Both approaches identify the same three controllable quantities — voltage, current and pulse duration — as the main levers available for reducing update energy, and the Methodology and Discussion sections return to this basis when comparing the three material strategies examined in this review.

Two-dimensional materials provide new room for synaptic device optimization. Their ultra-thin atomic structure, inherent lattice defects, unique surface characteristics and van der Waals interfacial features strongly influence internal charge transport and ion migration behaviors. In synaptic device fabrication, such materials can act as switching channels, conductive electrodes, tunneling barriers or charge-stacking layers to support core device functions [3–5]. Even with these advantages, simply replacing traditional layers with 2D materials cannot directly deliver low-energy performance. Device geometry, interfacial quality, pulse operation protocols and energy evaluation standards all play indispensable roles in shaping final device performance.

This review analyses published experimental data to explore effective material-based strategies for lowering synaptic update energy. The discussion centers on three practical engineering routes: monolayer MoS_2 defect modification, graphene-based electrode and interface optimization, and all-2D layered heterostructure design. These three studies were selected as representative, well-documented examples of three distinct mechanisms rather than as an exhaustive survey of the field; other defect-engineering, interface-engineering and heterostructure strategies exist and are not individually covered here. The analysis explains how each strategy modifies charge transport and relates these mechanisms to the reported voltage, current, pulse duration and energy data. The three approaches are then considered together to show their different routes to lowering electrical cost.

The contribution of this review is threefold. First, it consolidates three independently reported MoS_2-based engineering strategies within one consistent analytical framework, making their underlying mechanisms directly comparable even where their absolute numbers are not. Second, it

DOI: 10.54254/2755-2721/2026.37906

identifies the specific electrical parameter each strategy manipulates to reach femtojoule-to-picojoule-scale operation — suppressed current in defect engineering, restricted carrier injection in interface engineering, and shortened pulse duration in heterostructure design — clarifying that these are complementary rather than competing routes to the same goal. Third, it highlights the trade-offs in operating voltage, pulse duration and update linearity that accompany each route, giving device designers a basis for selecting or combining strategies according to application requirements.

This paper adopts a targeted narrative review approach instead of systematic review or meta-analysis techniques. Three experimental studies were selected according to clear screening criteria. All selected works report functional neuromorphic memory or artificial synapse devices, apply two-dimensional materials for specific structural or functional purposes, and provide measurable energy-related parameters such as update energy, programming current or operational power. Basic academic concepts regarding neuromorphic computation, in-memory architecture and 2D synaptic devices are defined and standardized using established literature resources.

Multiple key dimensions were extracted from each study for analysis. The collected information covers device structural design, material functional roles, underlying physical mechanisms, typical programming electrical parameters, recorded energy performance metrics, experimental control settings and device trade-offs. Reported energy data are used when available, while current and power readings are used to explain how each material strategy influences electrical cost.

Evidence reliability is assessed based on experimental control quality and transparency of testing parameters. Energy values are calculated when voltage, current and pulse duration are available, following academic guidelines for resistive-switching research [6].

2. Background

2.1. Biological synapses and artificial synaptic weight update

A biological synapse is the junction where a signal passes from the axon terminal of one neuron, the pre-synaptic neuron, to the dendrite of another, the post-synaptic neuron [7]. When an electrical signal arriving at the pre-synaptic terminal is strong enough, it triggers the release of chemical neurotransmitters across the synaptic gap; these bind to receptors on the post-synaptic neuron and generate an electrical response there. The strength of this connection, called synaptic weight, is not fixed: repeated or patterned signaling can strengthen or weaken it, a property known as synaptic plasticity, which underlies learning and memory [7].

An artificial synapse reproduces this behavior by using a single electronic device's conductance to stand in for synaptic weight: applying a programming pulse changes the device's conductance the way a biological signal changes connection strength, and reading the device's current or conductance afterward reads out the stored weight [7]. Two plasticity regimes are commonly distinguished. Potentiation refers to an increase in conductance, corresponding to a strengthened connection; depression refers to a decrease, corresponding to a weakened one. When the change persists well after the programming pulse ends, it is termed long-term potentiation or depression (LTP/LTD); when it instead decays back toward the original value within seconds to minutes, it is termed short-term plasticity (STP) [7]. All three devices compared in this review are evaluated through LTP/LTD-type testing: a series of identical positive pulses potentiates the device, a series of identical negative pulses depresses it, and the resulting conductance trajectory is what the linearity and switching-ratio metrics discussed in Section 2.4 describe.

DOI: 10.54254/2755-2721/2026.37906

2.2. Mechanisms of resistive switching and charge storage in 2D-material devices

Artificial synapses built from two-dimensional materials generally fall into a small number of physical categories, distinguished by what physically changes inside the device when its conductance is updated [7]. In electrochemical-metallization and valence-change memories, switching happens through the formation and rupture of a conductive filament: mobile metal ions or lattice vacancies migrate under an applied field to form, or break, a narrow conducting path between electrodes [7]. This kind of filamentary, defect-mediated switching has been demonstrated even in single-layer 2D materials generally, a behavior broadly termed an atomristor [7]; in monolayer MoS_2 specifically, sulfur vacancies have been identified as one defect type that supports it [7]. Sulfur-vacancy-mediated switching is the mechanism that Rajput et al. deliberately enhance through argon-plasma treatment, discussed in Section 3.1 [8].

A related but distinct route modifies conductance not by forming a filament inside the channel, but by controlling how easily charge carriers can enter the channel in the first place. Graphene has previously been used as an electrode in oxide-based resistive memories specifically because its semimetallic, low-density-of-states surface limits carrier injection and reduces operating current compared with a conventional metal contact [7]. This interfacial route is what Krishnaprasad et al. exploit in the graphene/MoS_2/SiO_x/Ni structure discussed in Section 3.2 [9].

A third category, charge-trap or floating-gate memory, stores synaptic weight as trapped charge on an isolated conductive layer rather than as a filament or an injection barrier. Charge tunnels onto or off the floating gate through a thin insulating barrier under an applied pulse, and the trapped charge electrostatically gates a separate conducting channel, changing its resistance [7]. Two-dimensional floating-gate memories built from an MoS_2/h-BN/graphene stack, in which the graphene layer serves as the charge-storage electrode and the h-BN as the tunnelling barrier, have specifically been reported as a way to reach large on/off ratios and long retention [7]. This is the same general device category that Tang et al. develop into an all-2D heterostructure, discussed in Section 3.3 [10].

Other working principles reported for 2D-material synapses include phase-change switching, in which an applied field or heat converts the material between structural phases of different resistivity, and ferroelectric switching, in which reversal of an internal electrical dipole modulates channel conductance [7]. These are not among the three strategies examined in this review, but they illustrate that defect/filament control, interface/injection control, and charge-storage control are three of several distinct physical routes by which 2D materials can be used to build a low-energy artificial synapse.

2.3. Two-dimensional materials relevant to this review

The three strategies compared in this review draw on the same small set of two-dimensional materials, each chosen for a specific electronic property. MoS_2 is a semiconducting transition metal dichalcogenide and provides the conducting channel in all three devices; because it can be grown or thinned down to a single monolayer, its electronic properties are unusually sensitive to atomic-scale defects such as sulfur vacancies [7], a sensitivity that Rajput et al. exploit deliberately [8]. Graphene is a semimetal with no bandgap and a very low density of states near its Fermi level [7], which makes it a poor source of free carriers to inject into an adjacent semiconductor; this property is what allows it to function as a current-limiting electrode in Krishnaprasad et al.'s device [9], and as a charge-storage floating-gate layer in Tang et al.'s device [10]. Hexagonal boron nitride (h-BN) is a wide-bandgap insulator that, unlike amorphous oxide dielectrics, has an atomically flat, dangling-

DOI: 10.54254/2755-2721/2026.37906

bond-free surface [7]; this reduces the density of unwanted charge traps at the interface and makes it well suited to serving as a controlled tunnelling barrier, the role it plays in Tang et al.'s floating-gate structure [10]. More generally, this dangling-bond-free character is what distinguishes 2D materials from conventional bulk semiconductors and oxides for neuromorphic applications: bulk materials accumulate uncontrolled surface defects and scattering, whereas defects in atomically thin 2D layers can be introduced and controlled deliberately [7], which is the property that Rajput et al.'s plasma treatment relies on [8].

2.4. Key performance metrics used in this review

Several quantities recur throughout the Methodology, Results and Discussion sections and are defined here for reference. The switching ratio, or on/off ratio, is the ratio between a device's conductance (or current) in its high-conductance state and its low-conductance state; a larger ratio makes the stored weight easier to distinguish from noise. Retention describes how long a programmed conductance state persists without further stimulation, and endurance describes how many programming cycles a device can withstand before its behavior degrades. Non-linearity in a potentiation or depression curve describes how far the actual conductance-versus-pulse-number trajectory deviates from a straight line; the studies compared here quantify this deviation with a non-linearity factor (NLF) fitted from the measured curve, where a value near zero indicates a close-to-linear, more easily trainable device [9]. Update energy, the main quantity compared across the three strategies in this review, is the electrical energy delivered to the device during one programming pulse and depends on the pulse's voltage, the current it drives, and its duration. For context, Rajput et al. report, citing established neuroscience literature, that the human brain carries out each synaptic event at a cost of only about 1–100 fJ [8]; this is the biological scale that all three strategies compared here are attempting to approach or undercut.

3. Device structures and experimental methods of the three strategies

3.1. Defect-engineered monolayer MoS_2

Rajput et al. fabricated defect-engineered monolayer MoS_2 memristors and compared the same devices before and after argon-plasma treatment [8]. Polycrystalline MoS_2 monolayers were grown on 285 nm SiO_2/Si substrates by atmospheric-pressure chemical vapor deposition (APCVD) in a single-zone furnace, using sulfur and molybdenum trioxide powders as precursors; the substrate temperature was raised to 750 °C and held while the sulfur source was subsequently heated to 150 °C for 25 minutes. Source and drain electrodes (Cr/Au) were defined by photolithography and thermal evaporation with a lift-off process, and the heavily doped Si substrate served as a global back gate, giving devices with a 5 μm channel length and 25 μm channel width.

To introduce sulfur vacancies without damaging the lattice, the authors applied on-chip argon-plasma treatment using a Diener Zepto plasma cleaner. Low-power treatment (0.3 mbar, 5 W, 60 s) produced no detectable vacancies, while high-power treatment (0.3 mbar, 20 W, 5 s) etched and damaged the flakes; an intermediate, optimal-power condition (0.3 mbar, 10 W, 5 s) selectively sputtered sulfur atoms, producing a sulfur-vacancy density of approximately 3.4×10^{14} cm^{-2} (stoichiometry $MoS_{1.7}$, versus $MoS_{1.98}$ for the pristine film). This vacancy density was confirmed through X-ray photoelectron spectroscopy (XPS) of the Mo 3d and S 2p core levels, Raman spectroscopy (a red-shifted E^1_2g mode, broadened peak widths, and new defect-related peaks near

DOI: 10.54254/2755-2721/2026.37906

377 and 411 cm^{-1}), and photoluminescence (an increased B-to-A exciton intensity ratio). This process is shown in Figure 1.

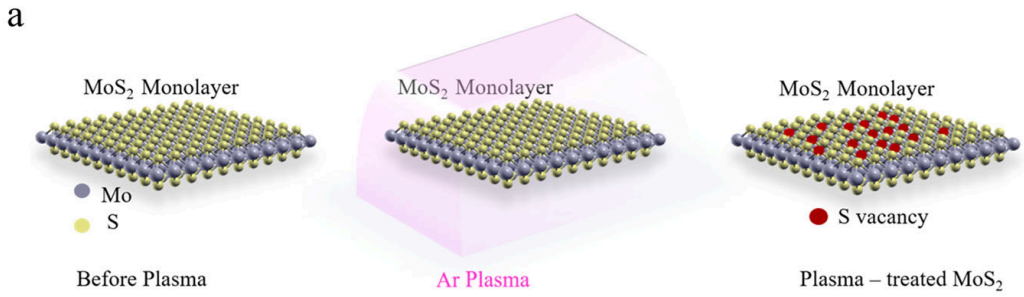

Figure 1. Schematic of the argon-plasma defect-engineering process applied to monolayer MoS_2: the pristine lattice, plasma exposure, and the resulting sulfur vacancies. Reproduced from Rajput et al. [8]

Electrical, endurance and synaptic-plasticity measurements were carried out in a home-built probe station using a Keithley 2636B dual-channel source meter under ambient conditions. Switching behavior was evaluated from current–voltage sweeps at zero gate bias (0 to 10 V for the pristine device; ±40 V after plasma treatment). Synaptic potentiation and depression were driven with ten +20 V, 1 ms pulses followed by ten −20 V, 1 ms pulses, with the resulting current read at a 0.1 V or 0.5 V bias, and pulse energy was estimated as the product of the read current, the pulse duration and the pulse voltage. A three-layer (784-256-10) artificial neural network was subsequently trained offline on the MNIST handwritten-digit dataset using the measured potentiation/depression curves as the basis for synaptic weights.

3.2. Graphene/MoS_2/SiO_x/Ni interface engineering

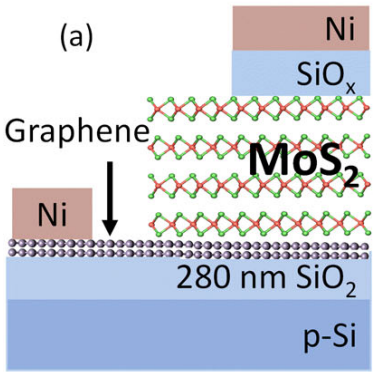

Figure 2. Schematic of the graphene/MoS_2/SiO_x/Ni cross-point memristor structure. Reproduced from Krishnaprasad et al. [9]

Krishnaprasad et al. fabricated graphene/MoS_2/SiO_x/Ni cross-point memristors in which a bilayer-graphene bottom electrode was used to restrict carrier injection into the switching layer [9]. Chemical-vapor-deposited bilayer graphene was wet-transferred onto a 285 nm SiO_2/p^+-Si substrate, patterned by photolithography, and etched into strips with oxygen plasma. Monolayer MoS_2 (approximately 8.8 nm as-grown thickness, confirmed by atomic force microscopy) was grown directly on the graphene by depositing 3–10 nm of Mo through electron-beam evaporation followed

DOI: 10.54254/2755-2721/2026.37906

by low-pressure CVD sulfurization. A SiO_x layer approximately 10 nm thick (confirmed by high-resolution transmission electron microscopy) and a Ni top contact were deposited in the same electron-beam evaporation step, giving cross-point devices with active areas ranging from 5 × 5 μm² to 50 × 50 μm². The device structure is shown in Figure 2.

Electrical characterization used a Keysight B1500A semiconductor device analyzer, with pulsed current–voltage measurements performed through WGFMU B1530A modules; devices were probed at room temperature on a 6200 Micromanipulator probe station, and a Janis cryogenic probe station was used for temperature-dependent measurements between 310 K and 370 K under vacuum (8×10^{-5} mbar). DC potentiation and depression were first examined by sweeping the current compliance from 100 nA to 10 μA and then applying increasing negative bias, giving 5 potentiation and 7 depression states with retention exceeding 10^3 s. Pulsed weight-update testing applied 64 potentiation pulses of +4.5 V and 64 depression pulses of −3 V, each 50 μs wide, while conductance was normalized to its maximum value and the non-linearity factor (NLF) was extracted using an established behavioral model. To isolate the role of each layer, the authors additionally fabricated control devices that replaced graphene with a gold electrode, and that omitted the SiO_x layer, comparing OFF-state current and RESET power across these variants. The optimized graphene/MoS_2/SiO_x/Ni synapse was finally integrated in series with a separately reported MoS_2-based leaky integrate-and-fire (LIF) neuron, driven by 60 μs, 5 V input pulse trains, to demonstrate synapse-controlled neuron spiking.

3.3. All-2D van der waals heterostructure

Tang et al. constructed a two-terminal floating-gate memory (2TFGM) entirely from 2D van der Waals materials: monolayer MoS_2 as the conducting channel, hexagonal boron nitride (h-BN) as the tunnelling barrier, and few-layer graphene (FLG) as both the floating-gate electrode and the source/drain contacts [10]. Monolayer MoS_2 was grown by chemical vapor deposition in a three-temperature-zone furnace using sulfur and MoO_3 powders as precursors on c-plane sapphire substrates, with the sulfur source, MoO_3 source and substrate held at 130 °C, 530 °C and 900 °C respectively under an Ar/O_2 carrier gas for approximately 30 minutes. h-BN and FLG flakes were mechanically exfoliated from bulk crystals and transferred layer-by-layer using a poly(propylene carbonate)-assisted home-built transfer station; the floating-gate and contact layers were then patterned by electron-beam lithography and oxygen reactive-ion etching, with devices wired out through Ti(3 nm)/Au(30 nm) electrodes. The device structure is shown in Figure 3.

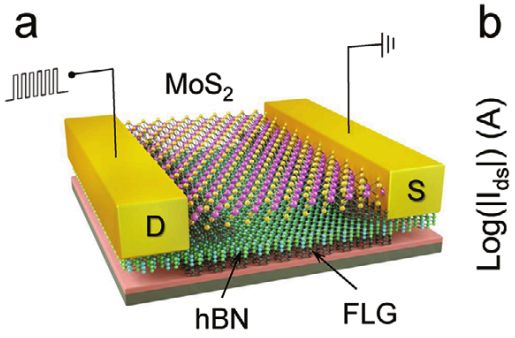

Figure 3. Schematic of the all-2D two-terminal floating-gate memory (2TFGM), showing the MoS_2 channel, h-BN tunnelling layer, and few-layer-graphene contacts/floating gate. Reproduced from Tang et al. [10]

DOI: 10.54254/2755-2721/2026.37906

Raman and photoluminescence spectra were collected with a 532 nm laser to confirm material quality, and surface morphology was verified by atomic force microscopy. Electrical measurements were performed in a closed-cycle cryogenic probe station (70–500 K) using an Agilent B1500 semiconductor parameter analyzer under vacuum. Devices with three h-BN thicknesses (7, 10 and 15 nm) were compared to examine how tunnel-barrier width affects the operating-voltage range, which fell from about ±20 V for the thickest barrier to about ±8 V for the thinnest. Retention was measured at 300, 400 and 500 K; endurance was tested over 10^5 switching cycles at ±18 V with a 100 ms pulse width; and switching speed was verified with pulses as short as 40 ns at ±23 V. For the synaptic-weight experiment, approximately 3,000 quasi-continuous conductance states were programmed using ±15 V pulses only 40 ns wide, and the energy of a single pulse was calculated as the product of the peak post-synaptic current, the pulse duration and the operating voltage. As with the other two strategies, the measured potentiation/depression data were used to simulate a three-layer (784-200-10) perceptron trained on the MNIST dataset with a standard backpropagation algorithm.

3.4. Systematic comparison of the three experimental approaches

Table 1 summarizes the device geometry, materials, key engineered layer, fabrication route, characterization tools, electrical test system and programming-pulse conditions used in each of the three studies, providing a systematic, side-by-side view of how the three research designs differ before their outcomes are compared in Section 4.

Table 1. Summary of device architecture, fabrication processes, and electrical characteristics for the three different memristor designs

Aspect	Rajput et al.(Defect Engineering)	Krishnaprasad et al.(Interface Engineering)	Tang et al.(All-2D Heterostructure)
Device geometry	Lateral back-gated FET-type memristor; 5 μm channel length × 25 μm width	Vertical cross-point memristor; 5×5 to 50×50 μm² active area	Vertical two-terminal floating-gate memory (2TFGM)
Channel material	CVD polycrystalline monolayer MoS_2	CVD monolayer MoS_2 (~8.8 nm) grown on graphene	CVD monolayer MoS_2
Key engineered layer	Ar-plasma-induced sulfur vacancies (~3.4×10^{14} cm^{-2})	Bilayer-graphene bottom electrode + ~10 nm SiO_x	h-BN tunnelling layer (7–15 nm) + few-layer-graphene floating gate
Electrodes	Cr/Au source-drain; Si back gate	Graphene bottom electrode; Ni top electrode	Few-layer-graphene contacts/floating gate; Ti/Au wiring
Key fabrication step	On-chip Ar-plasma treatment (0.3 mbar, 10 W, 5 s)	Direct CVD growth of MoS_2 on transferred graphene; e-beam evaporated SiO_x/Ni	Layer-by-layer dry transfer of exfoliated h-BN/FLG onto CVD MoS_2
Characterization tools	XPS, Raman, photoluminescence, AFM	AFM, Raman, HRTEM, EDS	Raman, photoluminescence, AFM
Electrical test system	Keithley 2636B, home-built probe station	Keysight B1500A + WGFMU B1530A; Janis cryogenic probe station	Agilent B1500, closed-cycle cryogenic probe station
Potentiation / depression pulses	10× (+20 V, 1 ms) then 10× (−20 V, 1 ms)	64× (+4.5 V, 50 μs) then 64× (−3 V, 50 μs)	±15 V, 40 ns pulses; ~3,000 states

DOI: 10.54254/2755-2721/2026.37906

Table 1. (continued)

Read voltage	0.1–0.5 V	0.3–0.5 V	~1 V (retention)

4. Results

The three studies were compared using the electrical quantities that determine the cost of changing a synaptic weight: programming voltage, current, and pulse duration. Potentiation refers to an increase in conductance caused by a positive programming pulse; depression refers to a decrease in conductance caused by a negative programming pulse.

4.1. Results of defect engineering

Using the argon-plasma-treated device described in Section 3.1, the effect of the added sulfur vacancies on synaptic weight update was examined directly [8]. To increase conductance, the device received positive 20 V pulses lasting 1 ms; to decrease conductance, it received negative 20 V pulses of the same duration.

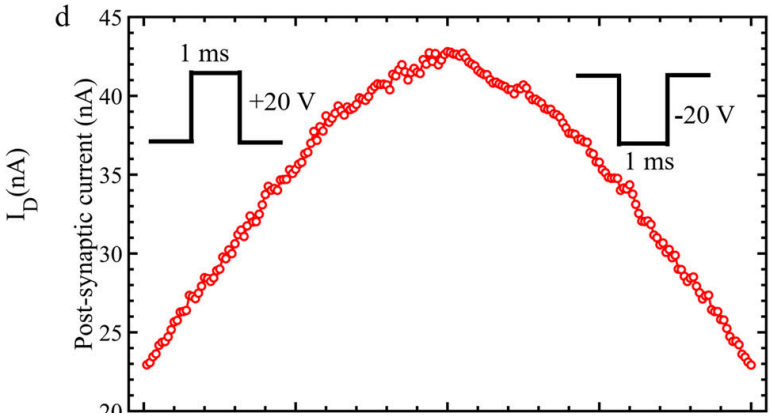

Figure 4. Potentiation and depression of post-synaptic current in the argon-plasma-treated MoS_2 device under ten +20 V and ten −20 V, 1 ms pulses. Reproduced from Rajput et al. [8]

After treatment, the high-resistance-state current measured at a 0.5 V read bias decreased from approximately 1 nA to 1 pA. The corresponding standby power decreased from approximately 0.5 nW to 0.5 pW, a reduction of about three orders of magnitude. The switching ratio increased from approximately 10^2 to 10^4, and the fitted potentiation non-linearity factor decreased from 21.79 to 2.81. For the programming sequence, the authors reported a minimum energy dissipation of approximately 30 fJ for one pulse, corresponding to an average pulse current of approximately 1.5 pA at 20 V and 1 ms. The resulting potentiation and depression are shown in Figure 4.

4.2. Results of graphene interface engineering

Using the graphene/MoS_2/SiO_x/Ni device described in Section 3.2, the effect of the graphene interface on current injection was examined through pulsed synaptic testing [9]. The device received 64 positive pulses of 4.5 V, each lasting 50 μs, to increase conductance, followed by 64 negative pulses of 3 V magnitude, also lasting 50 μs, to decrease conductance. Under this condition, both potentiation and depression had a reported non-linearity factor of zero, meaning that the conductance changed almost linearly and symmetrically with pulse number. The authors also

DOI: 10.54254/2755-2721/2026.37906

reported that weight-update linearity deteriorated when the pulse conditions or MoS_2 thickness changed. The linear and symmetric weight update is shown in Figure 5.

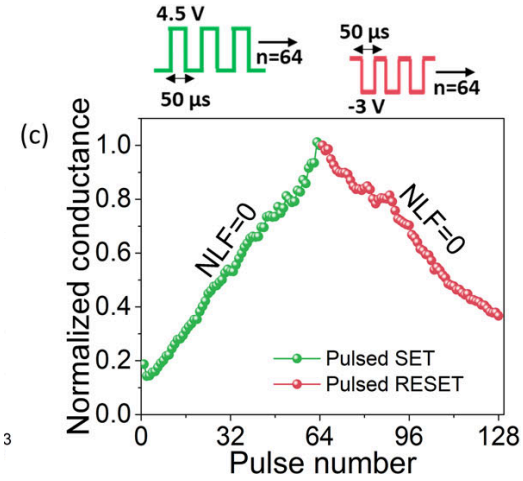

Figure 5. Linear and symmetric normalized-conductance weight update (NLF = 0) in the graphene/MoS_2/SiO_x/Ni device under 64 pulses of +4.5 V (SET) and 64 pulses of −3 V (RESET), each 50 µs. Reproduced from Krishnaprasad et al. [9]

Replacing the graphene bottom electrode with gold increased the OFF-state current by more than three orders of magnitude, supporting the explanation that graphene limits carrier injection. A graphene-barrier configuration demonstrated potentiation and depression with a lowest programming current of 100 nA. The authors also found that adding SiO_x to graphene/MoS_2/Ni reduced RESET power from approximately 18 mW to 15 µW. These results demonstrate how graphene-interface and stack engineering can restrict current and reduce RESET power.

4.3. Results of the all-2D heterostructure

Using the all-2D floating-gate device described in Section 3.3, reducing the h-BN barrier thickness from 15 nm to 7 nm reduced the approximate voltage-sweep range needed for switching from positive and negative 20 V to positive and negative 8 V [10], showing that a thinner tunnel barrier lowered the required switching voltage.

During the synaptic-weight experiment, positive 15 V pulses lasting 40 ns programmed the device toward higher conductance, while negative 15 V pulses of the same duration erased it toward lower conductance. The device produced approximately 3,000 quasi-continuous conductance states. For one positive programming event, the reported peak current was approximately 30 nA. Multiplication of the reported voltage, peak current, and pulse duration gives 18 fJ, matching the authors' reported energy for a single pulse. The fitted non-linearity values were 0.18 for potentiation and negative 0.29 for depression. The resulting conductance states are shown in Figure 6.

DOI: 10.54254/2755-2721/2026.37906

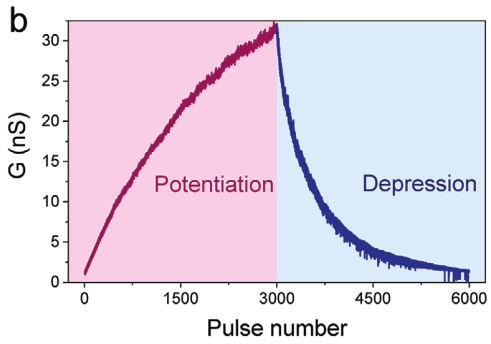

Figure 6. Potentiation and depression of conductance over 3,000 quasi-continuous states in the all-2D heterostructure device under ±15 V, 40 ns pulses. Reproduced from Tang et al. [10]

4.4. Combined results of the three strategies

Rajput and Tang both report energy for an individual programming pulse. Rajput reports a minimum of approximately 30 fJ, whereas Tang reports approximately 18 fJ for one positive pulse. Tang's reported value is about 40% below Rajput's minimum value. Both studies therefore demonstrate femtojoule-scale updates through different designs: Rajput through strong current suppression and Tang through short-pulse operation. However, Rajput's value is computed from an average pulse current while Tang's is computed from a peak pulse current, and peak current is typically higher than average current for a non-rectangular pulse; the two figures are therefore not calculated on a fully consistent basis, and the apparent 40% gap should not be read as a precise, like-for-like performance ranking.

The voltage and time results reveal how the three devices distribute electrical cost differently. Rajput used a relatively high magnitude of 20 V and a long 1 ms pulse, but its minimum average pulse current was only about 1.5 pA. Tang used a magnitude of 15 V and a reported peak current of 30 nA, but its pulse lasted only 40 ns. Krishnaprasad used lower pulse amplitudes of positive 4.5 V and negative 3 V with a 50 μs pulse. These results show that the three strategies act on different contributors to update cost rather than following one identical operating approach.

Together, the three studies demonstrate complementary methods for controlling electrical cost. Rajput primarily uses defect engineering to suppress current; Krishnaprasad uses a graphene interface to restrict carrier injection; and Tang combines controlled tunnelling with nanosecond programming in an all-2D heterostructure.

4.5. Summary of advantages and limitations

Table 2 condenses the results above into the specific mechanism, the reported metric and its test condition, and the main advantage and trade-off of each strategy.

DOI: 10.54254/2755-2721/2026.37906

Table 2. Comparison of key mechanisms, performance metrics, advantages, and limitations across the three engineering strategies

Strategy	Key Mechanism	Reported Metric (test condition)	Advantage	Trade-off / Limitation
Defect Engineering (Rajput et al.)	Current suppression via Ar-plasma-induced sulfur vacancies	~30 fJ minimum (20 V, 1 ms pulses)	Simple modification of an existing MoS_2 channel; 10^4 switching-ratio gain; NLF improved from 21.79 to 2.81	Requires relatively high voltage (20 V) and long (1 ms) pulses; low energy depends almost entirely on an extremely small current
Interface Engineering (Krishnaprasad et al.)	Carrier-injection barrier via graphene bottom electrode	100 nA lowest current (+4.5 V/−3 V, 50 μs pulses)	Linearity/symmetry (NLF = 0) under its tested pulse condition; lowest programming voltage of the three strategies	Linearity is sensitive to pulse width/amplitude and MoS_2 thickness; per-pulse energy not directly reported
All-2D Heterostructure (Tang et al.)	Combined low current and ultrashort (40 ns) pulses via h-BN tunnelling	~18 fJ per pulse (15 V, 40 ns)	Lowest reported energy on a peak-current basis; ~3,000 conductance states; fastest switching (40ns)	Still requires a relatively high voltage (15 V); more complex multi-layer stacking/transfer fabrication

Note: the energy figures in this table are not all calculated on the same basis. Rajput et al.'s ~30 fJ uses average pulse current, whereas Tang et al.'s ~18 fJ uses peak pulse current, which is typically higher than average current for a non-rectangular pulse; the two values are therefore not strictly comparable. Similarly, Krishnaprasad et al.'s NLF = 0 was measured under one specific pulse condition and is not a fixed property of the device.

5. Discussion

The results indicate that the examined MoS_2-based devices can achieve a low electrical cost when modifying synaptic weights, although the three strategies reduce this cost through different mechanisms. The energy consumed during one weight update can be approximated as $E = VIt$. Therefore, a device can reduce update energy by limiting the current, shortening the programming pulse, or reducing the required voltage.

Rajput et al. reported a minimum energy consumption of approximately 30 fJ for one conductance-update pulse after treating a monolayer MoS_2 channel with argon plasma [8]. The treatment introduced sulfur vacancies and reduced the current flowing through the device while preserving its ability to produce potentiation and depression. This result suggests that defect engineering can reduce the electrical cost of a weight update mainly by limiting current. However, the device still required pulses of up to 20 V lasting 1 ms. Its low reported energy therefore resulted primarily from its extremely small current rather than from low-voltage or high-speed operation.

Krishnaprasad et al. demonstrate a second current-reduction mechanism through graphene-interface engineering [9]. In the graphene/MoS_2/SiO_x/Ni structure, graphene restricted carrier injection into the MoS_2-containing stack, and a graphene-barrier configuration demonstrated programming current as low as 100 nA. Because energy is proportional to current when voltage and pulse duration are fixed, restricting carrier injection provides a route to lowering the electrical cost of conductance modification. The physical trade-off is that the device used pulses lasting tens of microseconds, and its weight-update linearity was sensitive to pulse conditions and MoS_2 thickness.

DOI: 10.54254/2755-2721/2026.37906

Tang et al. reported an even lower value of approximately 18 fJ for a potentiation event in an all-2D heterostructure containing MoS_2, h-BN and graphene [10]. This value can be calculated from the reported 15 V pulse, peak current of approximately 30 nA and pulse duration of 40 ns. Although the voltage was relatively high, the combination of low current and a very short pulse limited the amount of electrical energy delivered during the update. In comparison with Rajput et al., Tang et al. used a much shorter pulse but allowed a higher current. The two studies therefore demonstrate different routes to femtojoule-scale operation: Rajput et al. primarily reduced current through defect-controlled transport, whereas Tang et al. combined low current with nanosecond-scale programming. It is worth noting that Rajput et al.'s 30 fJ figure is derived from an average pulse current while Tang et al.'s 18 fJ figure is derived from a peak pulse current; because peak current typically exceeds average current for a non-rectangular pulse, this difference in calculation basis means the two values are not strictly comparable, and the apparent energy advantage of Tang et al.'s device should be read with that caveat in mind.

Overall, the three strategies aim to reduce the electrical cost of modifying synaptic weights through different methods. Defect engineering by Rajput et al. reduced the current through the MoS_2 channel. However, it required relatively high-voltage, long-duration pulses, meaning that its low energy consumption depended mainly on an extremely small current. The graphene-interface strategy examined by Krishnaprasad et al. restricted carrier injection and achieved programming currents as low as 100 nA. However, the device still required pulses lasting tens of microseconds, and its weight-update linearity deteriorated when the pulse conditions or MoS_2 thickness changed. The all-2D heterostructure developed by Tang et al. reported the lowest energy value of the three, calculated on a peak-current rather than average-current basis, by combining a low peak current with a much shorter 40 ns pulse, although its operating voltage remained relatively high. The evidence shows that particular MoS_2-based designs can combine the material's thin active region with defect, interface or heterostructure engineering to achieve very low reported update energies.

Several factors limit how far these comparisons can be generalized. First, the three studies were not conducted under identical measurement environments: Rajput et al.'s pulsed synaptic tests were performed in ambient air, Krishnaprasad et al.'s main pulsed test was likewise conducted in ambient conditions with vacuum reserved for their separate temperature-dependent mechanism study, while Tang et al.'s electrical measurements, including the synaptic-weight test, were carried out in a vacuum probe station. Because adsorbates and trapped charge at 2D-material surfaces are sensitive to ambient exposure, this difference in measurement environment could itself contribute to some of the reported differences in current and energy, independently of the underlying material-engineering strategy. Second, none of the three studies reports device-to-device or cycle-to-cycle variability for the specific pulse condition used to calculate its headline energy or current figure; the values compared in this review should therefore be read as single-condition, and in most cases single-device, results rather than as statistically characterized averages. Third, the operating voltages reported here — 20 V for Rajput et al. and 15 V for Tang et al. — are well above the sub-5 V range typical of practical CMOS peripheral circuitry, so the demonstrated low energy per pulse would need to be reconciled with the additional circuit-level cost of generating and switching such voltages before these results translate into system-level efficiency gains. Finally, Tang et al.'s heterostructure depends on layer-by-layer mechanical transfer of exfoliated h-BN and graphene flakes, a process that has not been demonstrated at wafer scale, which is a practical limitation not shared by Rajput et al.'s and Krishnaprasad et al.'s directly grown or deposited structures.

DOI: 10.54254/2755-2721/2026.37906

6. Conclusion

The examined studies form a feasible focused review because they can be organized around one shared problem: reducing the electrical cost of changing an artificial synaptic weight. Defect-engineered monolayer MoS_2 suppresses current and produces a reported minimum of approximately 30 fJ per pulse. Graphene interface engineering restricts carrier injection and produces a lowest reported programming current of 100 nA in a MoS_2-containing stack. An all-2D MoS_2/h-BN/graphene heterostructure combines controlled tunnelling with 40 ns pulses and produces a reconstructable estimate of approximately 18 fJ.

The strongest conclusion is that the three strategies are complementary rather than competitive. Defect engineering acts mainly through current suppression, graphene-interface engineering controls carrier injection, and the all-2D heterostructure combines controlled tunnelling with short-pulse operation. Together, they demonstrate how different two-dimensional material designs can address the voltage, current and pulse-duration terms that determine synaptic-update energy.

References

[1] Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE, 78*, 1629–1636. https://doi.org/10.1109/5.58356

[2] Ielmini, D., & Wong, H.-S. P. (2018). In-memory computing with resistive switching devices. *Nature Electronics, 1*, 333–343. https://doi.org/10.1038/s41928-018-0092-2

[3] Shi, Y. et al. (2018). Electronic synapses made of layered two-dimensional materials. *Nature Electronics, 1*, 458–465. https://doi.org/10.1038/s41928-018-0118-9

[4] Zhang, Z. et al. (2022). 2D materials and van der Waals heterojunctions for neuromorphic computing. *Neuromorphic Computing and Engineering, 2*, 032004. https://doi.org/10.1088/2634-4386/ac8a6a

[5] Cao, G. et al. (2021). 2D material based synaptic devices for neuromorphic computing. *Advanced Functional Materials, 31*, 2005443. https://doi.org/10.1002/adfm.202005443

[6] Lanza, M. et al. (2019). Recommended methods to study resistive switching devices. *Advanced Electronic Materials, 5*: 1800143. https://doi.org/10.1002/aelm.201800143

[7] Lee, G., Baek, J.-H., Ren, F., Pearton, S. J., Lee, G.-H., & Kim, J. (2021). Artificial neuron and synapse devices based on 2D materials. *Small, 17*: 2100640. https://doi.org/10.1002/smll.202100640

[8] Rajput, M. et al. (2024). Defect-engineered monolayer MoS_2 with enhanced memristive and synaptic functionality for neuromorphic computing. *Communications Materials, 5*: 190. https://doi.org/10.1038/s43246-024-00632-y

[9] Krishnaprasad, A. et al. (2023). Graphene/MoS_2/SiO_x memristive synapses for linear weight update. *npj 2D Materials and Applications, 7*: 22. https://doi.org/10.1038/s41699-023-00388-y

[10] Tang, J. et al. (2021). A reliable all-2D materials artificial synapse for high energy-efficient neuromorphic computing. *Advanced Functional Materials, 31*: 2011083. https://doi.org/10.1002/adfm.202011083