0.5T0.5R—An Ultracompact RRAM Cell Uniquely Enabled by van der Waals Heterostructures

Dujiao Zhang, *Student Member, IEEE*, Chao-Hui Yeh, *Member, IEEE*, Wei Cao, *Member, IEEE*, and Kaustav Banerjee, *Fellow, IEEE*

***Abstract*— Conventional designs of the extensively studied resistive-random access-memory (RRAM) cell involve one transistor and one RRAM—"1T1R," i.e., two separate devices thereby constraining its integration density. In this work, we overcome this longstanding limitation by experimentally demonstrating a novel memory architecture that combines the 1T and 1R into a single hybrid device by uniquely leveraging both lateral and vertical van der Waals (vdW) heterostructures. This ultracompact device, which can be considered as a "0.5T0.5R" memory cell, reduces the device count by half—the first of its kind in RRAM technology history, and simultaneously allows higher lateral as well as vertical (3-D) integration density w.r.t. the conventional 1T1R architecture. The unique "smart" device that can retain information after power is turned off is structurally designed by utilizing a shared graphene edge-contact and resistively switchable hexagonal boron nitride (*h*-BN) insulator. Aided by design optimization, record performance (<10 ns switching-speed), energy- (~0.07 pJ/bit), and area-efficiency (smallest footprint among all reported vdW-material-based RRAM memory units), as well as great retention (10^6 s) and endurance (>1000), benchmarked against current vdW-material-based RRAM devices, have been achieved by this 0.5T0.5R memory cell. Moreover, the RRAM's fine tunability with ultrashort pulsewidth, pulse amplitude, and gate voltage, enables synaptic plasticity and makes it an integrated three-terminal RRAM with considerable potential for neuromorphic and in-memory computing applications.**

***Index Terms*— 1T1R, 2-D-FET, 3-D-integration, graphene, *h*-BN, in-memory computing, memristor, neural circuit, neuromorphic computing, plasticity, RRAM, smart transistor, switching-energy, switching-speed, synapse, transition metal di-chalcogenide, van der Waals (vdW) materials, WS_2.**

Manuscript received November 21, 2020; revised January 11, 2021 and January 25, 2021; accepted January 27, 2021. Date of current version March 24, 2021. This work was supported in part by the ARO under Grant W911NF1810366, in part by the AFOSR (DURIP) under Grant FA9550-18-1-0448, in part by the JST CREST under Grant SB180064, and in part by the UC MRPI Research Program under Grant MRP-17-454999. This article is an extended version of a paper presented at IEDM 2020. The review of this article was arranged by Editor D. Triyoso. *(Dujiao Zhang and Chao-Hui Yeh contributed equally to this work.) (Corresponding author: Kaustav Banerjee.)*

Dujiao Zhang is with the Nanoelectronics Research Laboratory, Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA, and also with the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dujiao@ucsb.edu).

Chao-Hui Yeh, Wei Cao, and Kaustav Banerjee are with the Nanoelectronics Research Laboratory, Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: chyeh@ucsb.edu; weicao@ece.ucsb.edu; kaustav@ece.ucsb.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2021.3057598.

Digital Object Identifier 10.1109/TED.2021.3057598

I. Introduction

TRADITIONAL von Neumann architecture with physically separated processing and memory units is undergoing a significant paradigm-shift since transistor-based serial logic computing technology can no longer provide historical energy-efficiency and noticeable performance benefits for the recent flourishing growth in highly data-driven artificial intelligence applications. Nonetheless, an alternative, parallel computation-in-memory architecture [1], has been proposed to perform with lower computational power and resolve the bottlenecks of abundant data movement [2] for boosting data processing speed. To achieve such a potent memory-based matrix network for next-generation computing [Fig. 1(a)], robust and high-performance memory cell is essential [3], [4]. RRAM is considered as one of the most promising next-generation memory devices due to its nonvolatility, high switching-speed, low switching-energy, and small footprint [5]–[8], especially as the necessity of data-centric applications is exponentially growing in this twenty-first century. These merits make it a strong contender not only for conventional digital memory such as cache memory (SRAM) and dynamic RAM (DRAM), etc., but also for analog memristors in the emerging neuromorphic computing domain [9]–[12]. However, the formation of unstable oxygen-ions in the metal oxide switching materials during long-term SET/RESET cycling operation, in the form of undesired redox reactions with electrodes, overgrowth of filaments, or unwanted diffusion (or loss) of filament atoms, leads to severe endurance failure [13]. In contrast, due to the relatively larger formation energy of boron vacancy [14]–[16] and dangling-bond-free surface [17], the oxygen-less hexagonal-boron nitride (*h*-BN) as switching layer provides superior chemical stability, alleviating any oxidation reaction to metallic filaments and preventing redundant creation of undesirable vacancies. Therefore, those aforementioned troublesome issues in metal oxide switching materials induced by undesirable oxygen ions can be avoided. On the other hand, monolayer graphene with an atomically thin body, high conductivity, and strong sp^2 bonding serving as source line (SL) electrode can mitigate the oxidation at the interface between the electrode and the switching layer. In order to fully exploit the advantages of RRAM in large-scale memory array [Fig. 1(b)], one of the practical solutions is to connect a transistor in series with each RRAM, i.e., in the form of 1T1R, in order to suppress the sneak current from the inactive memory units [18] [Fig. 1(c) and (d)], thereby improving the reliability and sta-

tic energy efficiency. However, with doubled device count, the area efficiency of RRAM cell technology is inevitably degraded. In this work, we revisit the structures of the traditional 1T and 1R, and demonstrate an innovative way to merge the two devices into one by uniquely exploiting vdW heterostructures to share the dielectric layer and contacts between the transistor and RRAM components, owing to the extraordinary properties of van der Waals (vdW) solids [19], [20]. We found that the graphene edge contact scheme and resistively switchable *h*-BN insulator not only structurally support this design, but also electrically benefit the device performance [21]. In particular, the built-in *h*-BN RRAM cell embedded in 2-D-vdW-based transistor exhibits the following tailored characteristics such as low programming current, multiple resistance-state modulation, and large dynamic working range, which are highly capable of providing the foundation for memristive-based computing technology. In fact, our hybrid device represents a "smart" transistor that can retain information after the power is turned off. This article begins by presenting the fabrication process of the 0.5T0.5R cell and heterostructure material characterization in detail (Section II), and then manifesting device features including conductive filament (CF) formation and electrical properties of the solitary transistor and the RRAM component, as well as that of the combined memory cell (0.5T0.5R) (Section III). Finally, conclusions are drawn in Section IV augmented by a benchmarking analysis that establishes the superior characteristics of the 0.5T0.5R cell.

II. Device Fabrication and Material Characterization

The device fabrication starts from a canvas of chemical vapor deposition (CVD) grown monolayer graphene [Fig. 2(a)] on a sapphire substrate. Using inductively coupled plasma (ICP), a narrow gap [Fig. 2(b)] is formed in the graphene canvas, and later bridged, by area-selective CVD under 850 °C growth temperature [22], with monolayer (1L) WS_2 that forms a graphene-semiconductor-graphene (GSG) heterojunction [Fig. 2(c)]. The gate length of the GSG-heterojunction field-effect transistor (GSG-HFET) in our fabricated 0.5T0.5R is 1 μm. To prepare the graphene edge electrode for the *h*-BN RRAM component, part of the graphene as well as the substrate underneath on one side of the WS_2 strip is etched down by ~15 nm, thereby forming a step [Fig. 2(d)]. ICP tool was employed to form the ultrasteep Gr/Al_2O_3 step using BCl_3 as reacting gas for anisotropic etching. Subsequently, 10/50 nm Pt/Au is deposited with electron beam deposition (eBeamD) on the graphene on the other side of the WS_2 strip [Fig. 2(e)], serving as the electrode at the transistor component side. Next, a piece of a few-layer *h*-BN is transferred on top [Fig. 2(f)], functioning as both the gate dielectric [23] and active switching layer for the transistor and RRAM components, respectively. The *h*-BN film was synthesized on electro-polished Cu foil by the low-pressure CVD method [24], with ammonia borane (NH_3-BH_3, Sigma-Aldrich, 97%, roughly 30 mg) as the precursor. Note that before growth, the Cu foils were annealed at 1000 °C for 60 min at the pressure of 500 mTorr with the carrier gas of

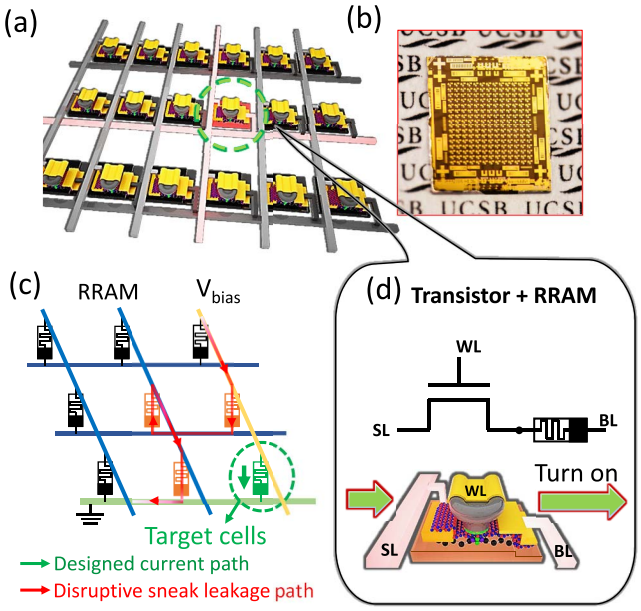

Fig. 1. (a) Schematic and (b) optical image of a large-scale 0.5T0.5R memory array demonstrated at UCSB. (c) In a large-scale memory array, leakage current of each memory cell (e.g., RRAM) in the inactive memory units can lead to a large sneak current that may disturb the normal operation of the active cell. (d) Although adding a series transistor to each RRAM can effectively suppress the leakage current, but at the penalty of additional area cost. By uniquely exploiting the unique properties of vdW materials, an ultra-compact 0.5T0.5R cell can be constructed to alleviate this penalty.

30-sccm H_2 to create a smooth surface [25], then the precursor was heated up to 90 °C and kept for 20 min to deposit a 5 nm thick *h*-BN film. To suppress potential gate leakage (bandgap of *h*-BN is not large enough, ~5.5 eV), an additional Al_2O_3 gate dielectric layer is formed by oxidizing 3 nm Al thin film deposited on top of the *h*-BN/WS_2 region [Fig. 2(g)] with standard eBeamD and lift-off process. Next, using the double-layer eBeam resist liftoff process, a T-shape Al top gate is made, which also serves as the mask for the self-aligned final metallization step (5/15 nm Ti/Au), to form the electrode at the RRAM component side [Fig. 2(h)]. Before electrical characterization, we annealed fabricated devices at 200 °C for 2 hr with H_2 carrier gas to improve their electrical contact and to remove the residue H_2O molecules introduced during *h*-BN transfer. This thermal annealing treatment can relieve the stress introduced during fabrication and remove contaminants and humidity to enhance the inter-layer adhesion [26]. Note that although the intrinsic interlayer attractive force has been reported to be strong (>10 N cm^{-2}) [27], we believe process innovations are desired to further improve the adhesion, thereby ensuring the robustness of large-scale integration of layered materials for mass production. Fig. 2(i) and (j) show the schematic illustration and optical image of a finalized 0.5T0.5R memory cell, respectively, together with three test pads, in which the active part of the device within the yellow rectangle is graphene domain. The footprint of our fabricated 0.5T0.5R memory cell is dominated by the transistor size, which is also the case in conventional 1T1R architecture with RRAM on top of the transistor. However, our 0.5T0.5R exhibits: 1) enhanced electric field arising from

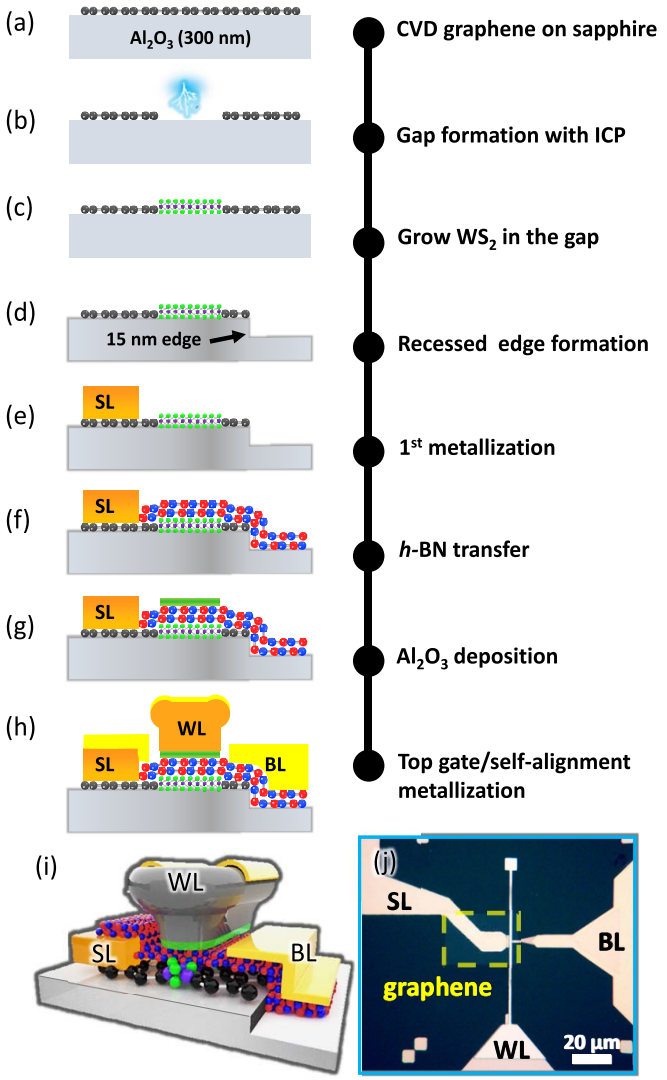

Fig. 2. Fabrication process flow of the proposed 0.5T0.5R memory cell. (a) Prepare CVD graphene on sapphire substrate. (b) Form a gap in the graphene with ICP. (c) Grow WS_2 in the gap, forming a GSG heterojunction channel. (d) Etch graphene and substrate on one side of the WS_2 to form a step. (e) Deposit Pt/Au on the graphene on the other side as the SL with electron beam deposition. (f) Cover with a few-layer CVD *h*-BN film on top serving as both the gate dielectric and resistive switching layer. (g) Deposit 3 nm Al on WS_2 and oxidize to Al_2O_3 as the second layer gate dielectric. (h) Make a T-shape Al gate with double-layer eBeam resist based liftoff process, and deposit 5/15 nm Ti/Au as the bitline (BL) using the T-gate as a self-aligned mask. (i) Schematic and (j) optical image of the finalized device.

the recessed edge, which can lower switching voltage as well as switching energy; 2) confined CFs at recessed edge region, resulting in lower off current; and 3) the potential for higher vertical 3-D integration density.

During the device fabrication, the quality of the synthesized materials was carefully examined. The WS_2 channel in the GSG heterojunction is around 1 μm wide, as shown in the optical image in Fig. 3(a), and the SEM image in Fig. 3(b). Raman mapping at 353.1 cm^{-1} (E^1_{2g} mode of WS_2) shows clear feature signals for high-quality WS_2 [Fig. 3(c)]. The *as-grown* wafer-scale CVD *h*-BN film shows a continuous and uniform surface on SiO_2/Si substrate, as shown in Fig. 3(d). Raman spectrum of the *h*-BN film shows a characteristic peak

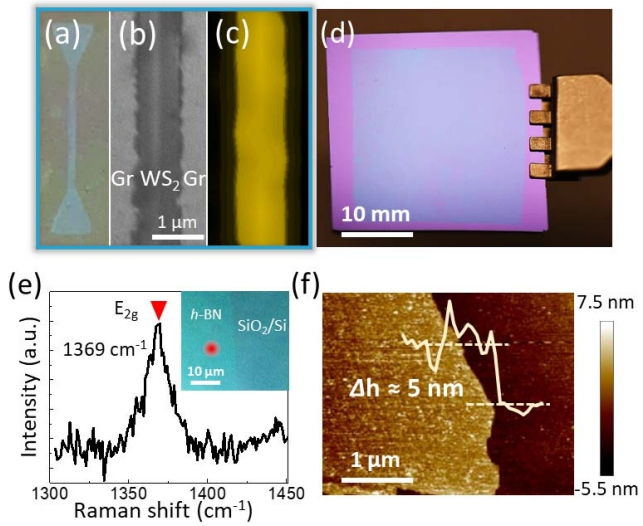

Fig. 3. (a) Optical image, (b) SEM image, and (c) Raman mapping of a GSG heterostructure fabricated through area-selective CVD technique. (d) Photograph of the wafer scale CVD *h*-BN film on SiO_2/Si substrate. (e) Raman spectrum of *h*-BN film shows a characteristic peak at 1369 cm^{-1} corresponding to the E_{2g} vibration mode of the hexagonal B-N bonds. (f) AFM image of the *h*-BN film on SiO_2/Si substrate shows the thickness of *h*-BN film to be ~5 nm.

at 1369 cm^{-1} corresponding to the E_{2g} vibration mode of the hexagonal B-N bonds [Fig. 3(e)]. The thickness of the *h*-BN film is measured to be ~5 nm by an atomic-force microscope (AFM), as shown in Fig. 3(f).

III. Device Characterization

Before diving into the 0.5T0.5R memory cell, we first characterize the two components separately to examine their individual performance.

A. Transistor Component

Fig. 4(a) and (b) shows the transfer curve I_d–V_g, and output curve I_d–V_d, respectively, exhibiting high ON-current of 127 μA/μm at V_d = 3 V, V_g–V_t = 2.5 V, large ON/OFF current ratio of $>10^7$, negligible drain induced barrier lowering (DIBL), ohmic behavior in low V_d region, and good current saturation, which are attributed to the excellent electrostatic integrity and the feature of dynamically tunable Schottky barrier uniquely enabled by the graphene edge contact [22]. This high-quality transistor component guarantees that the sneak current can be effectively suppressed, and thereby the performance, energy efficiency, and robustness of the 0.5T0.5R memory cell will not be limited. The lower-right inset in Fig. 4(a) shows the transmission electron microscope (TEM) image of the gate-stack of the transistor component, verifying the thicknesses of the WS_2 channel, *h*-BN, and Al_2O_3 dual-layer gate dielectrics to be ~1, 5, and 5 nm, respectively.

B. RRAM Component

Also, discrete RRAMs are fabricated, for the purpose of design and optimization. It is well known that the electric field can be enhanced near the sharp edge of a conductor, which is expected to benefit the RRAM performance. The cross-sectional image of the recessed edge structure of a

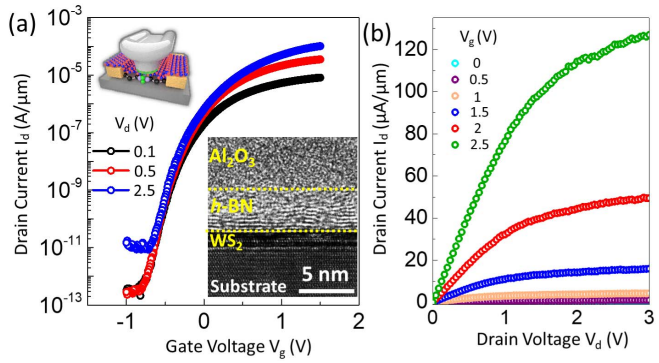

Fig. 4. (a) Transfer characteristics I_d–V_g and (b) output characteristics I_d–V_d of a solitary GSG-heterojunction-FET (GSG-HFET). The upper inset in (a) shows the structure of the separately fabricated GSG-HFET, and lower inset in (a) shows TEM image of the gate-stack of the transistor component, verifying the thicknesses of WS_2 channel, *h*-BN and Al_2O_3 dual-layer gate dielectrics to be ~1, 5, and 5 nm, respectively.

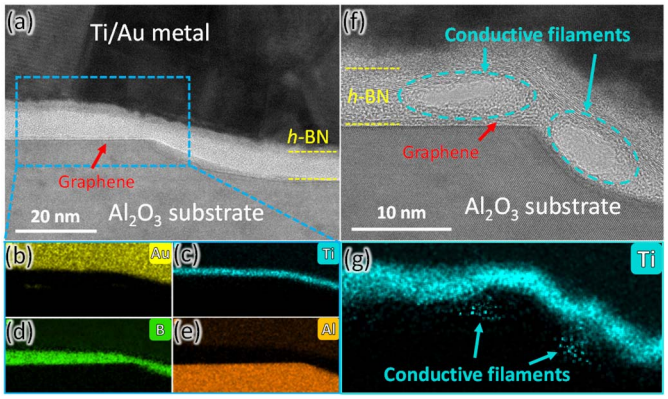

Fig. 5. (a) Cross-sectional TEM image of RRAM component. The energy dispersive X-ray (EDX) analyses from blue rectangular region confirm the (b) Au, (c) Ti, (d) B, and (e) Al elements in corresponding layers. (f) Two CFs are clearly visible around the graphene edge. (g) EDX element mappings of Ti uncovers that these filaments are attributed to Ti.

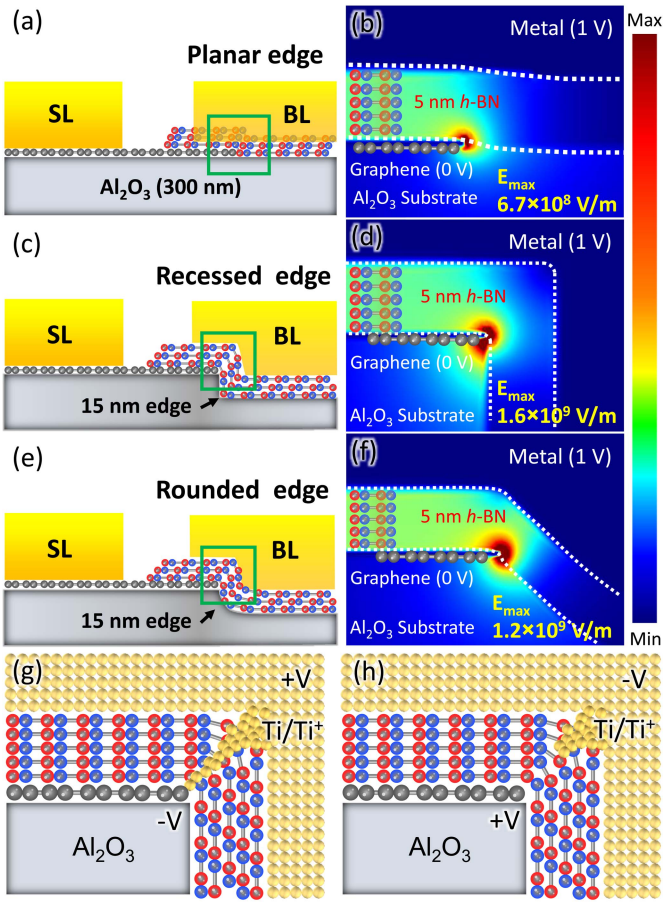

Fig. 6. (a) Schematic of an RRAM with "planar edge" structure and (b) simulated electric field distribution at the graphene edge. (c) Schematic of an RRAM with "recessed edge" structure and (d) simulated electric field distribution at the graphene edge. (e) Schematic of an RRAM with "rounded edge" structure and (f) simulated electric field distribution at the rounded edge. For all three structures, 1 V is applied to the top electrode (or BL), and graphene (or SL) is grounded. The switching mechanism of RRAM: (g) SET process is achieved by applying positive voltage to metal electrode, and (h) inversely, RESET process is achieved by applying positive voltage to graphene.

fabricated RRAM was obtained using high-resolution TEM microscopy, clearly showing a 5 nm *h*-BN switching layer sandwiched in between monolayer graphene and Ti/Au metals [see Fig. 5(a)]**.** Through EDX analysis, it was found that elemental composition mapping within the blue dash frame distinctly conforms to the designed Gr-*h*-BN-Ti/Au RRAM configuration, corresponding to Fig. 5(b) and (c) for Au/Ti metallic leads and Fig. 5(d) for B element associated with the 5 nm *h*-BN layer, respectively. Fig. 5(e) shows the rich Al concentration attributed to the patterned Al_2O_3 substrate. After the SET process, two CFs are clearly visible in the *h*-BN near the recessed edge [see Fig. 5(f)], which proves that the graphene edge plays a dominant role in forming the CF in this *h*-BN RRAM. Through EDX analysis [Fig. 5(g)], it is found that CFs are mainly contributed by Ti ions injected from the Ti electrode. In fact, to fully exploit the edge of graphene, two test structures have been evaluated. Compared to the planar structure [Fig. 6(a)], the eventually employed recessed structure [Fig. 6(c)] is found to offer more than 100% higher near-edge electric field based on finite-element electromagnetic simulation [Fig. 6(b) and (d)]. Although the real structure [See TEM image, Fig. 5(f)] shows a rounded edge as illustrated in Fig. 6(e), instead of a perfect rectangular profile at the edge, due to the non-uniformity of etching rate around it, according to the simulation results, the electric field maxima near the rounded edge [Fig. 6(f)], ~1.2×10^9 V/m, is still much larger than that at the planar edge (0.67×10^9 V/m). Moreover, our simulation results indicate that the recessed architecture can make CFs more confined, localized, and sizeable to bridge the graphene-fringe and metal electrodes for achieving low resistance state with relative ease [Fig. 6(g)]. This is due to the stronger electric-field distribution at the recessed edge, thereby alleviating the possibility of CFs being randomly generated out of the edge region, resulting in a tightly distributed SET/RESET voltage [see the statistics shown in Fig. 7(c)]. Fig. 6(h) shows the RESET state driven by the application of inverse voltage, where Ti ions can be migrated back to the top electrode, thereby disconnecting the CFs and achieving a high-resistance state (HRS). It is also noted that the use of *h*-BN as an oxygen-free switching layer can further mitigate the oxidation at the metal-to-active layer contact to minimize variability of the operation voltage of SET/RESET and any unexpected changes in the

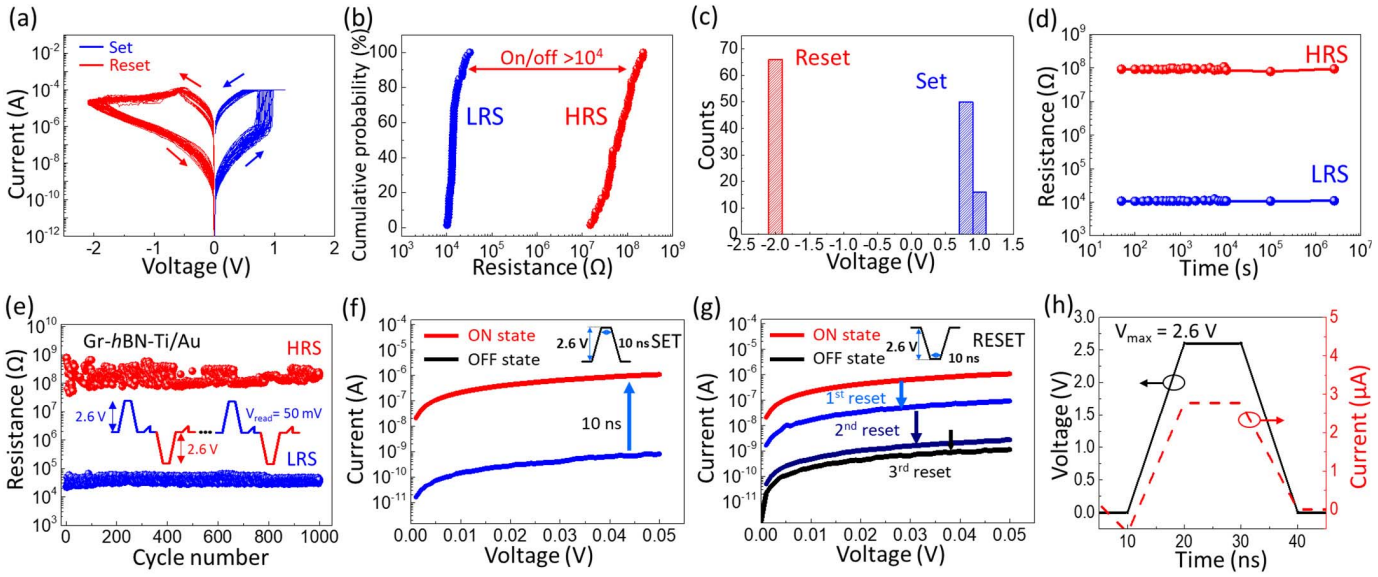

Fig. 7. (a) 65-DC switching cycles of an individual Gr-*h*-BN-Ti/Au RRAM cell. (b) Cumulative distribution of the resistance per cycle (read at 50 mV) in HRS and LRS. (c) Statistics of SET/RESET voltages for two-state switching in (a). (d) Resistive-state retention time of HRS and LRS over 30 days at ambient condition. (e) 1000 SET/RESET cycles (see the waveform in the inset) does not change the resistive states. Read current versus voltage of a Gr-*h*-BN-Ti/Au RRAM (1R) cell before and after a 10 ns wide - (f) SET pulse (see inset), and (g) RESET pulse (see inset). Three successive 10 ns wide pulses are needed to RESET the cell completely, suggesting a possible multi-bit implementation scheme. (h) Measured current versus voltage within a pulse. Switching energy of the RRAM can be obtained by integrating (current × voltage) over the entire pulse period.

resistance-dominated storage states, respectively. In contrast, oxide-based switching layers are likely to cause performance degradation due to the nonpreferred oxidization reaction in terms of oxygen migration into metallic electrode/filament over extended storage cycling [28]. These attractive electrical characteristics of the designed conductive bridge Gr-*h*-BN-Ti/Au RRAM, such as small operation voltages, relatively large ON/OFF ratio, and robust endurance cycling, are demonstrated in Fig. 7.

Fig. 7(a) shows the measured DC switching characteristics over 65 loops. As shown, the butterfly shape has barely changed post cycling, indicating that the used materials are of high quality and the fabrication process did not introduce much degradation, in agreement with [29] and [30]. Cumulative probability versus cell resistance is extracted at 50 mV and shown in Fig. 7(b). The HRS and low-resistance state (LRS) resistance ratio of over 10^4, provides a sufficiently large window for the implementation of multi-bit cell and analog memristor, and greatly reduces the static leakage power. The steep/sharp distribution of state resistance [Fig. 7(b)] and SET/RESET [Fig. 7(c)] voltages over cycling, again, confirms the low variability of resistance and operation voltage distribution, i.e., high device robustness. The LRS is obtained by the formation of CFs and the resistance varies due to the variation in the thickness and the number of CFs. The HRS is obtained when the CFs are ruptured. The mechanism of conduction in the HRS can be Poole-Frenkel emission, Schottky emission, and/or space charge limited current injection, etc. [31]. HRS shows larger variability w.r.t LRS. Possible factors that lead to the larger variability of HRS include non-ideal dielectric properties of the switching layer, fabrication process conditions, and interfacial properties of the switching layer and the electrode. Fig. 7(d) shows that both HRS and LRS resistances remain unchanged over 30 days, proving the extraordinary retention capability of this device. Fig. 7(e) shows that the SET/RESET cycling endurance can be larger than 1000. To examine how fast this RRAM can operate, we perform SET and RESET operations with a series of pulses with widths ranging from 10 μs down to 10 ns, which is the resolution of our pulse generator. It is found that the device can be SET with a single 10 ns wide pulse (2.6 V in height), with cell resistance reduced by four orders of magnitude [Fig. 7(f)]. Note that the upper bound of SET operation speed should be smaller than 10 ns. 10 ns wide pulse (with the same amplitude of 2.6 V but reversed polarity) can realize RESET as well, but end up with a smaller resistance window. Three successive RESET pulses can fully restore the cell resistance to the original status before SET [Fig. 7(g)]. This property suggests that the multi-bit implementation of this cell can be realized in the RESET period with $<$10 ns per bit. The switching time t_s can be expressed as [32]

$$t_s \propto \frac{1}{\upsilon} e^{\beta(E_a - qaE)} \tag{1}$$

where υ is the attempt frequency; β is the inverse thermal energy, i.e., $= 1/k_BT$ (k_B is Boltzmann's constant and T is the temperature); E_a is the activation energy; a is the lattice constant; q and E are the charge of the ion and the electric field, respectively. Given that the formula (1) expresses switching time as a function of E, t_s can be exponentially reduced as the E is enhanced further, leading to the diminished operation time in SET/RESET switching processes. Compared to the E of planar structure, the field intensity of recessed edge contact is 2-fold larger as shown in Fig. 6(d). Thus, smaller t_s is anticipated in the Gr-edge-contacted *h*-BN RRAM device. Fig. 7(h) shows the measured current of the RRAM cell as a function of voltage within a SET impulse. The likely sub-10 ns SET/RESET time of this device aided by its compactness makes this device competitive not only for DRAM but even for cache memory applications. Moreover,

the h-BN RRAM with such swift access latency of less than 10 ns, thereby significantly reducing storage cycle time, can meet the demands for stringent energy-efficient computation units. Such a significant reduction of access latency results in the minimal switching energy consumption of 0.07 pJ/bit, which can be calculated with the equation

$$E_{\text{switching}} = \int_0^{t_d} I \times V \, dt \tag{2}$$

where I, V, and t_d represent the current response, input voltage, and spike duration, respectively. On the other hand, the average energy consumed per switching event of a single transistor can be extracted through the power-delay product (PDP) formula [33], defined as

$$\text{PDP} = CV_{\text{DD}}^2 \tag{3}$$

where C is transistor gate capacitance, V_{DD} is the supply voltage. Compared to current 2-D transistors used in-memory technology (Table I), our GSG heterojunction transistor adopting stacked gate dielectrics (5 nm h-BN/5 nm Al_2O_3) with a capacitance of ~364 nF/cm^2 exhibits a minimal power consumption calculated at 0.096 pJ/bit, which is attributed to the optimized contact resistance [34], [35]. To analyze the device-to-device variability, 20 RRAM cells have been characterized using DC cycling measurements for extraction of electrical properties. Fig. 8(a) shows the SET/RESET $I-V$ loops of all 20 devices, in which ten have 3 nm thick h-BN (in red color) and the rest have 5 nm thick h-BN (in blue color). Compared to those 3 nm h-BN RRAM cells, a significant reduction in OFF-state current can be achieved using 5 nm thick h-BN switching layer, as expected. Moreover, the histogram plot [Fig. 8(b)] shows that RRAM cells with 5 nm thick provide lower variation of ON/OFF ratio compared to those with 3 nm h-BN. As shown in Fig. 8(c), the critical SET voltage has a distribution, which could be improved in the future through materials engineering, programming methodology, and fabrication process improvements.

Furthermore, in terms of brain-inspired neuromorphic computing, one of the pivotal characteristics of the biological synapse is synaptic plasticity. Synaptic plasticity can be mimicked by the proposed RRAM by increasing (potentiation) and decreasing (depression) the resistance (weight) value by appropriate voltage pulses [36]. To illustrate the h-BN RRAM cell's ability to compatibly serve as an artificial neural synapse [Fig. 9(a)] that can implement spike-induced weight updates, a designed 10 ns pulse train with increasing amplitude is employed to manipulate the resistance states of RRAM, indicating that the long-term potentiation and depression functionalities of a synapse can be emulated, as shown in Fig. 9(b). Notably, the quantized switching phenomenon in the RRAM RESET process has been demonstrated in a previous report [37], which is attributed to an atomic-level reaction in metallic filaments leading to gradual rupture of the CFs. Given this unique feature, multi-step resistance modulation of our RRAM-based bionic synapse can be conducted by step-wise voltage pulse input to implement the tunable weight update for neuromorphic computing.

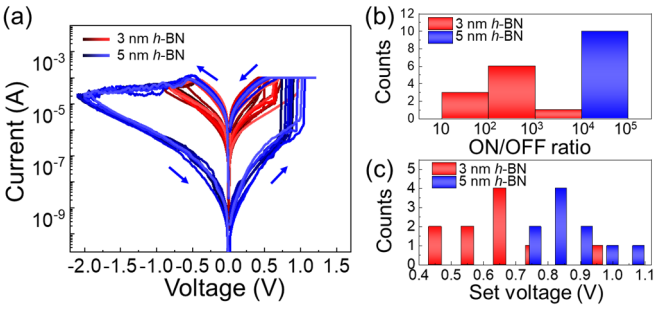

Fig. 8. (a) Comparison of median $I-V$ curves of a total of twenty RRAM cells with 3 nm/5 nm thick h-BN switching layer current. Cumulative distribution of (b) ON/OFF ratio and (c) critical set voltages of ten devices with 3 nm thick h-BN and ten devices with 5 nm thick h-BN, respectively.

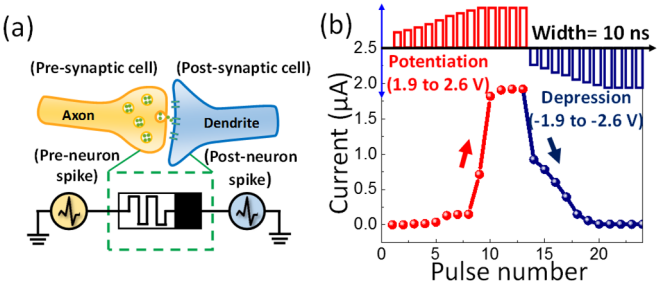

Fig. 9. (a) Schematic of synaptic transmission in neural system and illustration of an RRAM-based bionic neural cell. (b) Emulation of the long-term potentiation and depression features of a synapse, by applying a pulse train with increasing amplitude to the RRAM cell. Each pulse is only 10 ns wide, indicating the potential of this RRAM cell for implementing a high-speed synapse.

C. 0.5T0.5R Memory Cell

Fig. 10(a) shows the DC switching characteristics of a 0.5T0.5R memory cell, as shown in the optical image in the inset, over 54 loops with gate voltage set to be 2 V/3 V for SET/RESET. It can be observed that the SET/RESET voltages increase to 2.5 V w.r.t the 0.8 V of a discrete RRAM cell. Also, the ON/OFF current ratio is reduced to be 10^2, w.r.t the 10^4 of a discrete RRAM cell, which is expected. The transistor component limits the compliance current of the entire memory unit and consumes an appreciable portion of the voltage applied between SL and BL, thus degrading the SET/RESET voltage and ON/OFF ratio. This can be mitigated by enlarging the transistor or increasing the gate voltage, but with the penalty of increased footprint or affected robustness, respectively. The cell resistance variation is also degraded as shown in the widened distribution in Fig. 10(b). Although these numbers are better than previous reports on 2-D materials-based 1T1R memory cells [38], [39], more follow-up efforts are needed to derive further improvements. It is noteworthy that the transistor component introduces one more knob (the gate voltage, V_g) to control the memory cell resistance. As shown in Fig. 10(c), a series of V_g pulses of different amplitudes are applied to the transistor component along with 10 ns wide SET/RESET pulse voltage, resulting in different current (or resistance) status of the cell. During the V_g tuning, SET/RESET voltages are set to be 1.85/−1.75 V. The SET current value can be further modulated using various gate voltages, showing the maximum difference in the ON/OFF ratio of up to one order of magnitude, which provides an additional degree of freedom for multi-bit control. Moreover, this gate-controllable

TABLE I

COMPARISON OF 0.5T0.5R MEMORY CELL AGAINST PREVIOUS REPORTS ON VAN DER WAALS MATERIALS-BASED RRAM DEVICES AND CELLS

Metrics \ Reference		[38]	[39]	[40]	[41]	[42]	[43]	**This work**
Active Switching Layer		CVD *h*-BN	WO_3/ WSe_2	CVD *h*-BN	CVD *h*-BN	CVD *h*-BN	CVD *h*-BN	**CVD *h*-BN**
Layer Thickness		5 nm	6 nm	0.4 nm	8 nm	Multilayer	Multilayer	**5 nm**
Forming Voltage		Forming-free	Forming-free	Forming-free	NA	NA	3-8 V	**Forming-free**
Switching Voltage		0.5/ -0.8 V	0.8/ 0.3 V (Unipolar)	0.7/ -0.5 V	3.0/ -2.0 V	2.0/ -2.0 V	2.75/ -2.5 V	**0.9/ -2.0 V**
Switching Time		NA	700 ns	<15 ns	NA	< 5 µs	200 ns	**< 10 ns**
Retention Time		> 5000 s	> 10^4 s	> 10^5 s	> 10^6 s	> 3×10^3 s	NA	**> 10^6 s (30 days)**
Endurance		> 50 cycles	90 cycles	50 cycles	1500 cycles	70 cycles	80,000 cycles	**> 1000 cycles**
ON/OFF Ratio		1R: 5 1T1R: 4	1R: 10^3 1T1R: 10^2	1R: 10^7	1R: 10^4	1R: 10^8	1R: 10^6	**1R: 10^4 0.5T0.5R: 10^2**
Switching Energy	**Transistor Part**	2.88 pJ/bit (Back Gate)	2.15 pJ/bit (Back Gate)	NA	NA	NA	NA	**0.096 pJ/bit (Top Gate)**
	RRAM Part	NA	2.6 pJ/bit	NA	NA	NA	20 fJ/bit	**0.07 pJ/bit**
On-current (Transistor Part)		15 µA/µm (1L MoS_2)	150 µA/µm (Multilayer WSe_2)	NA	NA	NA	NA	**127 µA/µm (1L WS_2)**
Working Area		1040 µm² (1T+1R)	28 µm² (1T+1R)	1 µm² (RRAM only)	25 µm² (RRAM only)	~ 10 µm² (RRAM only)	~ 0.02 µm² (RRAM only)	**12 µm² (0.5T0.5R)**

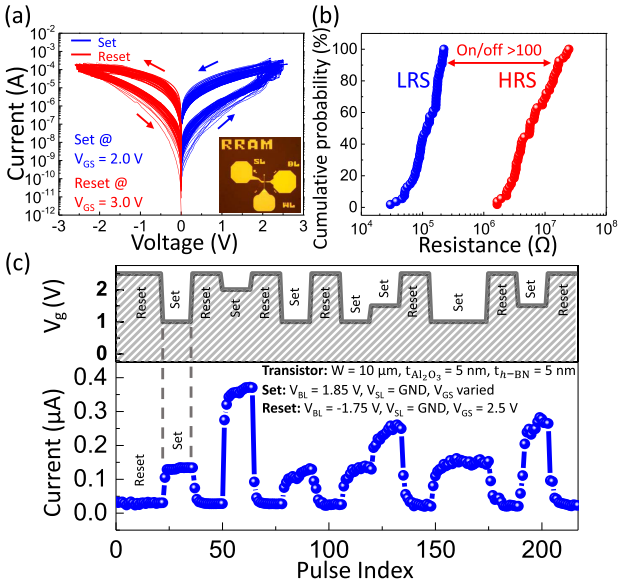

Fig. 10. (a) 54-DC switching cycles of a fabricated 0.5T0.5R cell. Inset: optical image. (b) Cumulative distribution of the resistance per cycle (read at 50 mV) in HRS and LRS. Due to the impact of the transistor component on the compliance current, ON/OFF ratio is reduced to 100. (c) Gate voltage (V_g) of the transistor component can be used to fine tune the cell current (or resistance), which represents an additional knob, beside SET or RESET voltage, for weight manipulation of neural networks.

memory cell also exhibits rapid response and stabilization during various current states' modulation as the storage-states undergo RESET-to-SET transition and vice versa, suggesting low cycling hysteresis. These aforementioned properties make our device a unique three-terminal RRAM that can record the analog weight information, suitably serving as a bionic neuron.

IV. CONCLUSION

This work demonstrated a novel compact memory cell that integrates the functionalities of both transistor and RRAM into a single "smart" device, which is uniquely enabled by vdW materials, leveraging CVD grown lateral graphene-WS_2-graphene heterostructure as well as vertical heterostructured *h*-BN/WS_2 and *h*-BN/graphene stack configurations, thereby justifying the designation of "0.5T0.5R" memory cell. This hybrid structure reduces the device count of a conventional 1T1R cell by half and is shown to exhibit extraordinary performance, energy efficiency, and compactness, as benchmarked against previous reports on vdW-material based memory cells/devices [38]–[43] in Table I. Specifically, sub-10 ns SET/RESET, 0.07 pJ/bit energy consumption, and 12 μm^2 cell footprint are achieved, which represents a great leap in advancing RRAM and "1T1R" memory technology. Additionally, the ultra scalability of any TMD-channel based FET [44], including the GSG-HFET [22], indicates that this "0.5T0.5R" memory cell has great potential to further reduce the entire memory cell's lateral footprint. Furthermore, the demonstrated 0.5T0.5R hybrid memory cell array can be monolithically 3-D-stacked [45] to build the ultimate high-density nonvolatile memory arrays and neuromorphic/in-memory computing systems, with significantly higher vertical density than the conventional 1T1R architecture (with RRAM on top of the transistor), unprecedented performance, and energy-efficiency to emulate the workings of the human brain in the near future.

ACKNOWLEDGMENT

All process steps for device fabrication and materials characterization were carried out using the Nanostructure Cleanroom Facility at the California NanoSystems Institute and the Nanofabrication Facilities at UCSB—part of the National Nanotechnology Infrastructure Network. The authors acknowledge Aidan Taylor at UCSB for TEM support and valuable discussions. The authors also thank Tanmay Chavan and Arnab Pal of the Nanoelectronics Research Lab at UCSB for useful feedback and discussions.

REFERENCES

[1] W. H. Kautz, "Cellular logic-in-memory arrays," *IEEE Trans. Comput.*, vol. C-18, no. 8, pp. 719–727, Aug. 1969, doi: 10.1109/T-C.1969.222754.

[2] T. N. Theis and H.-S.-P. Wong, "The end of Moore's law: A new beginning for information technology," *Comput. Sci. Eng.*, vol. 19, no. 2, pp. 41–50, Mar. 2017, doi: 10.1109/MCSE.2017.29.

[3] H. S. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature Nanotechnol.*, vol. 10, no. 3, pp. 191–194, Mar. 2015, doi: 10.1038/nnano.2015.29.

[4] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nature Nanotechnol.*, vol. 15, no. 7, pp. 529–544, Jul. 2020, doi: 10.1038/s41565-020-0655-z.

[5] W. W. Zhuang *et al.*, "Novel colossal magnetoresistive thin film nonvolatile resistance random access memory (RRAM)," in *IEDM Tech. Dig.*, Dec. 2002, pp. 193–196, doi: 10.1109/IEDM.2002.1175811.

[6] I. G. Baek *et al.*, "Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," in *IEDM Tech. Dig.*, Dec. 2004, pp. 587–590, doi: 10.1109/IEDM.2004.1419228.

[7] C. An *et al.*, "Non-volatile resistive switching for advanced memory applications," in *IEDM Tech. Dig.*, Dec. 2005, pp. 746–749, doi: 10.1109/IEDM.2005.1609461.

[8] H. Y. Lee *et al.*, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *IEDM Tech. Dig.*, Dec. 2008, pp. 1–4, doi: 10.1109/IEDM.2008.4796677.

[9] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990, doi: 10.1109/5.58356.

[10] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nature Nanotechnol.*, vol. 3, no. 7, pp. 429–433, Jul. 2008, doi: 10.1038/nnano.2008.160.

[11] P. A. Merolla *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 73–668, Aug. 2014, doi: 10.1126/science.1254642.

[12] M. Davies *et al.*, "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, Jan. 2018, doi: 10.1109/MM.2018.112130359.

[13] B. Chen *et al.*, "Physical mechanisms of endurance degradation in TMO-RRAM," in *IEDM Tech. Dig.*, Dec. 2011, pp. 12.3.1–12.3.4, doi: 10.1109/IEDM.2011.6131539.

[14] H.-C. Wu, Y.-S. Lin, and S.-W. Lin, "Mechanisms of visible light photocatalysis in n-doped anatase TiO_2 with oxygen vacancies from GGA+U calculations," *Int. J. Photoenergy*, vol. 2013, pp. 1–7, Jan. 2013, doi: 10.1155/2013/289328.

[15] D. Z. Gao, J. Strand, M. S. Munde, and A. L. Shluger, "Mechanisms of oxygen vacancy aggregation in SiO_2 and HfO_2," *Frontiers Phys.*, vol. 7, p. 43, Mar. 2019, doi: 10.3389/fphy.2019.00043.

[16] Z.-L. Hou, M.-S. Cao, J. Yuan, X.-Y. Fang, and X.-L. Shi, "High-temperature conductance loss dominated defect level in h-BN: Experiments and first principles calculations," *J. Appl. Phys.*, vol. 105, no. 7, Apr. 2009, Art. no. 076103, doi: 10.1063/1.3086388.

[17] W. Cao, J. Kang, W. Liu, and K. Banerjee, "A compact current-voltage model for 2D semiconductor based field-effect transistors considering interface traps, mobility degradation, and inefficient doping effect," *IEEE Trans. Electron Devices*, vol. 61, no. 12, pp. 429–4282, Dec. 2014, doi: 10.1109/TED.2014.2365028.

[18] M. Mao, Y. Cao, S. Yu, and C. Chakrabarti, "Optimizing latency, energy, and reliability of 1T1R ReRAM through appropriate voltage settings," in *Proc. 33rd IEEE Int. Conf. Comput. Design (ICCD)*, Oct. 2015, pp. 359–366, doi: 10.1109/ICCD.2015.7357125.

[19] P. Ajayan, P. Kim, and K. Banerjee, "Two-dimensional van der Waals materials," *Phys. Today*, vol. 69, no. 9, pp. 38–44, Sep. 2016, doi: 10.1063/pt.3.3297.

[20] W. Cao *et al.*, "2-D layered materials for next-generation electronics: Opportunities and challenges," *IEEE Trans. Electron Devices*, vol. 65, no. 10, pp. 4109–4121, Oct. 2018, doi: 10.1109/TED.2018.2867441.

[21] C. Yeh, D. Zhang, W. Cao, and K. Banerjee, "0.5T0.5R—Introducing an ultra-compact memory cell enabled by shared graphene edge-contact and h-BN insulator," in *IEDM Tech. Dig.*, Dec. 2020, pp. 12.3.1–12.3.4, doi: 10.1109/IEDM13553.2020.9371902.

[22] C. Yeh, W. Cao, A. Pal, K. Parto, and K. Banerjee, "Area-selective-CVD technology enabled top-gated and scalable 2D-heterojunction transistors with dynamically tunable Schottky barrier," in *IEDM Tech. Dig.*, Dec. 2019, pp. 23.4.1–23.4.4, doi: 10.1109/IEDM19573.2019.8993600.

[23] J. Kang *et al.*, "Graphene and beyond-graphene 2D crystals for next-generation green electronics," *Proc. SPIE*, vol. 9083, Jun. 2014, Art. no. 908305, doi: 10.1117/12.2051198.

[24] D. Zhang *et al.*, "Thickness-tunable growth of ultra-large, continuous and high-dielectric h-BN thin films," *J. Mater. Chem. C*, vol. 7, no. 7, pp. 1871–1879, Feb. 2019, doi: 10.1039/c8tc05345f.

[25] W. Liu, H. Li, C. Xu, Y. Khatami, and K. Banerjee, "Synthesis of high-quality monolayer and bilayer graphene on copper using chemical vapor deposition," *Carbon*, vol. 49, no. 13, pp. 4122–4130, Nov. 2011, doi: 10.1016/j.carbon.2011.05.047.

[26] H. Rokni and W. Lu, "Direct measurements of interfacial adhesion in 2D materials and van der Waals heterostructures in ambient air," *Nature Commun.*, vol. 11, no. 1, pp. 1–14, Nov. 2020, doi: 10.1038/s41467-020-19411-7.

[27] A. K. Geim, S. V. Dubonos, I. V. Grigorieva, K. S. Novoselov, A. A. Zhukov, and S. Y. Shapoval, "Microfabricated adhesive mimicking gecko foot-hair," *Nature Mater.*, vol. 2, no. 7, pp. 461–463, Jun. 2003, doi: 10.1038/nmat917.

[28] J. Robertson, O. Sharia, and A. A. Demkov, "Fermi level pinning by defects in HfO_2-metal gate stacks," *Appl. Phys. Lett.*, vol. 91, no. 13, Sep. 2007, Art. no. 132912, doi: 10.1063/1.2790479.

[29] F. M. Puglisi *et al.*, "2D h-BN based RRAM devices," in *IEDM Tech. Dig.*, Dec. 2016, pp. 34.8.1–34.8.4, doi: 10.1109/IEDM.2016.7838544.

[30] L. Sun *et al.*, "Self-selective van der Waals heterostructures for large scale memory array," *Nature Commun.*, vol. 10, no. 1, p. 3161, Jul. 2019, doi: 10.1038/s41467-019-11187-9.

[31] H.-S. P. Wong *et al.*, "Metal-oxide RRAM," *Proc. IEEE*, vol. 100, no. 6, pp. 970–1951, Jun. 2012, doi: 10.1109/.JPROC.2012.2190369.

[32] Y. Yang, P. Gao, S. Gaba, T. Chang, X. Pan, and W. Lu, "Observation of conducting filament growth in nanoscale resistive memories," *Nature Commun.*, vol. 3, no. 1, pp. 1–8, Mar. 2012, doi: 10.1038/ncomms1737.

[33] S.-C. Lin and K. Banerjee, "A design-specific and thermally-aware methodology for trading-off power and performance in leakage-dominant CMOS technologies," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 11, pp. 1488–1498, Nov. 2008, doi: 10.1109/TVLSI.2008.2001060.

[34] J. Kang, W. Liu, D. Sarkar, D. Jena, and K. Banerjee, "Computational study of metal contacts to monolayer transition-metal dichalcogenide semiconductors," *Phys. Rev. X*, vol. 4, no. 3, pp. 1–14, Jul. 2014, doi: 10.1103/PhysRevX.4.031005.

[35] A. Allain, J. Kang, K. Banerjee, and A. Kis, "Electrical contacts to two-dimensional semiconductors," *Nature Mater.*, vol. 14, no. 12, pp. 1195–1205, Nov. 2015, doi: 10.1038/nmat4452.

[36] Y. Shi *et al.*, "Electronic synapses made of layered two-dimensional materials," *Nature Electron.*, vol. 1, no. 8, pp. 458–465, Aug. 2018, doi: 10.1038/s41928-018-0118-9.

[37] Y.-E. Syu *et al.*, "Atomic-level quantized reaction of HfO_x memristor," *Appl. Phys. Lett.*, vol. 102, no. 17, Apr. 2013, Art. no. 172903, doi: 10.1063/1.4802821.

[38] C. Wang *et al.*, "3D monolithic stacked 1T1R cells using monolayer MoS_2 FET and hBN RRAM fabricated at low (150°C) temperature," in *IEDM Tech. Dig.*, Dec. 2018, pp. 22.5.1–22.5.4, doi: 10.1109/IEDM.2018.8614495.

[39] M. Sivan *et al.*, "All WSe_2 1T1R resistive RAM cell for future monolithic 3D embedded memory integration," *Nature Commun.*, vol. 10, no. 1, p. 5201, Nov. 2019, doi: 10.1038/s41467-019-13176-4.

[40] X. Wu *et al.*, "Thinnest nonvolatile memory based on monolayer h-BN," *Adv. Mater.*, vol. 31, no. 15, Apr. 2019, Art. no. 1806790, doi: 10.1002/adma.201806790.

[41] P. Zhuang *et al.*, "Nonpolar resistive switching of multilayer-hBN-based memories," *Adv. Electron. Mater.*, vol. 6, no. 1, Nov. 2019, Art. no. 1900979, doi: 10.1002/aelm.201900979.

[42] W. Lin, P. Zhuang, D. Akinwande, X.-A. Zhang, and W. Cai, "Oxygen-assisted synthesis of hBN films for resistive random access memories," *Appl. Phys. Lett.*, vol. 115, no. 7, Aug. 2019, Art. no. 073101, doi: 10.1063/1.5100495.

[43] S. Chen *et al.*, "Wafer-scale integration of two-dimensional materials in high-density memristive crossbar arrays for artificial neural networks," *Nature Electron.*, vol. 3, no. 10, pp. 638–645, Oct. 2020, doi: 10.1038/s41928-020-00473-w.

[44] W. Cao, J. Kang, D. Sarkar, W. Liu, and K. Banerjee, "2D semiconductor FETs—Projections and design for sub-10 nm VLSI," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3459–3469, Nov. 2015, doi: 10.1109/TED.2015.2443039.

[45] J. Jiang, K. Parto, W. Cao, and K. Banerjee, "Ultimate monolithic-3D integration with 2D materials: Rationale, prospects, and challenges," *IEEE J. Electron Devices Soc.*, vol. 7, pp. 878–887, Jun. 2019, doi: 10.1109/JEDS.2019.2925150.