Two-dimensional materials enabled next-generation low-energy compute and connectivity

Arnab Pal, Kunjesh Agashiwala, Junkai Jiang, Dujiao Zhang, Tanmay Chavan, Ankit Kumar, Chao-Hui Yeh, Wei Cao, and Kaustav Banerjee*

Since the invention of the metal–oxide–semiconductor field-effect transistor (MOSFET) in late 1959, the impact of electronics on human society has been increasingly pervasive, heavily regulating modern health, transport, finance, entertainment, and social media sectors through "Big Data." However, daily generation of petabytes of data from these sectors, along with their associated communication overhead, is placing an immense strain on the conventional computing and communication technologies, which were not developed exclusively for big data. Tackling these problems calls for a holistic overhaul of the current semiconductor technology, from materials to architecture, and two-dimensional (2D)-layered materials with their exotic electrical and structural properties are well positioned to accomplish just that. This perspective article aims to provide an overview of the key technological innovations in the nanoelectronics domain that have been achieved with 2D-materials thus far, and to bring forth the promise of this new materials family in developing brain-inspired ultra low-energy on-chip computing and communication techniques to usher a new era in electronics.

Introduction

Need for energy-efficient electronics: The emergence of the complementary-metal-oxide-semiconductor (CMOS) technology in the late sixties ushered in an explosive development in the field of microelectronics that has intensified in the recent decade with the emergence of "big data" and data centers. This, along with the rapid growth in machine learning (ML) enabled artificial intelligence (AI) and Internet-of-things (IoT) applications, is driving transformative societal and economic changes through various implementations in self-driving cars, avionics, and smart grid technologies. This exponential growth in the volume of data being generated and consumed, however, demands an equally high exponential increase in the capacity needed for processing and moving this data around **(Figure 1)**, thereby dissipating more energy. In fact, it has been estimated that energy dissipation of data centers will triple over the next-decade and is expected to leave a carbon footprint larger than that of the entire aviation industry. Moreover, Moore's Law,[1] which has enabled energy-efficient computing with shrinkage of transistor feature size every 2–3 years, is slowing down as transistors scale down to sub-20 nm feature sizes, thereby increasing power density while the fundamental heat dissipation mechanism in chip packaging/cooling solutions remains unchanged.[2] Therefore, to further increase chip computational efficiency and repel the imminent threat of global warming through energy-intensive inefficient computing, it is imperative that innovative solutions to low-energy computing be actively explored. **Figure 2** shows the way forward with solutions encompassing a wide range of emerging logic (both energy-efficient, and programmable that increases functionality) and interconnect technologies, that can be integrated in the vertical dimension via monolithic-3D (M3D) integration

Arnab Pal, Department of Electrical and Computer Engineering, University of California, CA, USA;
Kunjesh Agashiwala, Department of Electrical and Computer Engineering, University of California, CA, USA;
Junkai Jiang, Department of Electrical and Computer Engineering, University of California, CA, USA;
Dujiao Zhang, Department of Electrical and Computer Engineering, University of California, CA, USA;
Tanmay Chavan, Department of Electrical and Computer Engineering, University of California, CA, USA;
Ankit Kumar, Department of Electrical and Computer Engineering, University of California, CA, USA;
Chao-Hui Yeh, Department of Electrical and Computer Engineering, University of California, CA, USA;
Wei Cao, Department of Electrical and Computer Engineering, University of California, CA, USA;
Kaustav Banerjee, Department of Electrical and Computer Engineering, University of California, CA, USA; kaustav@ece.ucsb.edu
*Corresponding author
doi:10.1557/s43577-022-00270-0

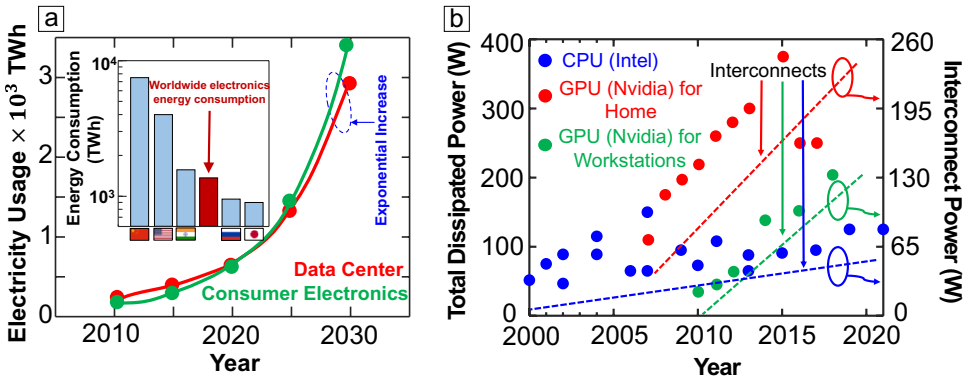

Figure 1. Overview of worldwide and on-chip electronics energy/power consumption. (a) Shows the rise in the energy consumption for data centers and consumer electronics as a function of time, and its projection beyond 2021. The exponential rise is projected to cause electronics energy consumption to dominate over those of all nations by 2030, which stands at #4 in the global rankings of nation-wise energy consumption at present. (b) On-chip power consumption for CPUs and GPUs as a function of time, showing a steady increase in the total dissipated power (TDP). Interconnect power consumption, which can be expected to contribute close to 2/3rd of the overall power consumption, also shows a steady increase, contributing to low chip energy efficiency.

and provide new technology platforms for alternate computing schemes such as in-memory and neuromorphic (NM) computing. However, these improvements need a radical departure from the conventional CMOS-Cu material platforms and can only be made possible through a significant overhaul of the current semiconductor technology.

Promise of 2D-materials in designing energy-efficient electronics: Recently, the emergence of layered 2D-materials (2D-M) (**Figure 3**) with exotic properties of inherent thinness,[3–5] defect-free interface,[6] optimal bandgap and effective mass of charge carriers,[4] strong band-tail suppression,[7] and bulk silicon comparable mobilities,[8,9] has provided sufficient promise and excitement to make this post-Si/Cu revolution a possibility. Figure 3 shows the entire family of 2D-materials[5] encompassing a broad range of electronic materials including metals, semimetals, semiconductors, insulators, superconductors, and topological insulators. Despite this wide range of conduction mechanisms, all 2D-materials are layered, held together by strong in-plane covalent bonding and by relatively weak van der Waals (vdW) forces along the out-of-plane direction, which allows one to selectively grow or peel off individual layers, each ~ 0.35 nm thick, offering precise thickness control.

Since the net energy consumption of any design is limited by two fundamental factors: energy consumed by the computation (device power) and energy consumed by the data communication i.e., interconnects, solutions to low power computing, therefore, require concurrent solutions to both (Figure 2). The ubiquitous electronic switch used for logic computing, the MOSFET, has a Boltzmann-limited minimal subthreshold swing (SS) of 60 mV/decade at room temperature, thereby imposing an upper limit on the maximum energy efficiency of such a device.[4,10] Sub-kT/q logic switches overcome the Boltzmann limit and therefore, can be operated with smaller voltages, leading to significant energy savings (**Figure 4**a). Among the several low SS (< 60 mV/decade) devices proposed, the most notable ones are tunneling-FETs (TFETs),[11,12] negative-capacitance FET (NCFET),[13] Dirac-source FETs (DSFETs),[14] impact-ionization FETs (IFETs),[15] and nano-electro-mechanical FET (NEMFET).[16,17] However, despite the theoretical promise, realistically there are several design and reliability concerns that have been detrimental toward their commercial implementation. NCFETs, for example, suffer from severely limited design space constraints that hamper its practical realization,[18] in addition to enhanced hysteresis and reliability issues. DSFETs can only yield sub-60 SS over a limited voltage range and offer SS no smaller than the contemporaries, along with a low density of states (near graphene Dirac point) limited ON-current.[19] Similarly, IFETs[20] and NEMFETs[21] suffer from severe reliability issues with hot-carrier (generated during avalanche breakdown) mediated oxide breakdown, and the mechanical failure of the conducting cantilever, respectively. TFETs, therefore, despite their quantum mechanical tunneling limited ON-current affecting applicability in conventional computing, can be beneficial for low-energy and low-frequency circuits.

However, the impact of these newly discovered low SS devices is still limited by the computing architecture that they are used in. Specifically, conventional von Neumann (VN) architecture with separate computing and memory elements consume energy not only for both the logic- and the memory-functionalities but also for data communication between them. Neuromorphic (NM) computing[22] is an alternative, which mimics certain neuro-biological architectures of the human brain, and therefore, can be designed to be highly parallel with very low-power consumption that has the potential to perform complex operations in a smaller area; and can also perform computation in-memory. Therefore, integration of both memory and data processing in NM architecture not only cuts down on their discrete power consumption, but also minimizes the energy needed for data movement. This has led to NM computing garnering much attention leading to an explosive market growth recently, particularly for ML

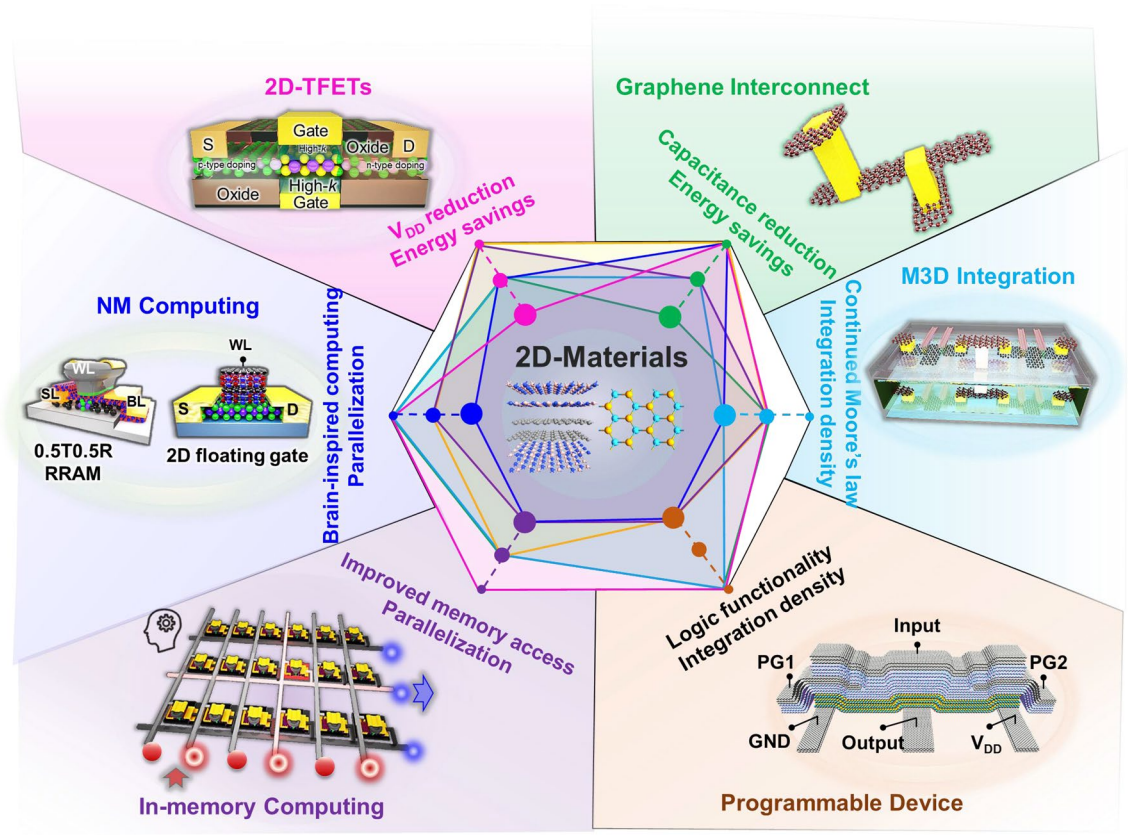

Figure 2. Low-energy computing and communication directions enabled by various 2D-materials (2D-M) w.r.t state-of-the-art von Neumann (VN) computation with CMOS and Cu interconnects. The use of 2D-M in designing low-energy low-subthreshold swing (SS) 2D-TFETs (pink), next-generation graphene (Gr) interconnects (green), monolithic-3D integration (light blue), programmable devices (orange), in-memory (purple), and neuromorphic computing (deep blue), with their respective figures of merits (FOMs), w.r.t state-of-the-art CMOS/Cu technology with VN computing (shown on the outermost hexagon at the center with smaller solid circles). The size of the circles represents increasing FOMs, with the biggest circles in the inner hexagon representing the desired corner, enabled with 2D-M. Note that the color of the lines and the size of the dots they connect correspond to the degree of benefits derived from the various approaches illustrated around the outermost hexagon. Low-SS 2D-TFETs for example, are expected to yield highest FOM for V_{DD} reduction, no benefit in cap reduction, moderate benefit (because of lower generated heat) in integration density, and no benefits in other FOMs. Gr interconnects offer cap reduction, V_{DD} scaling (due to smaller cap), and moderate memory access and integration density FOMs. M3D integration offers highest integration FOM, moderate memory access, V_{DD} and cap reduction FOMs; programmable devices offer highest functionality, moderate memory access, V_{DD} and cap FOMs; in-memory computing offers highest parallelization and memory access, and logic functionality FOMs, while NM computing offers highest FOMs overall, but with low cap (due to extensive interconnects) FOM. Overall, 2D-materials with their inherent thinness offer benefit in integration density in all application spheres.

enabled AI applications, led by commercial NM chips, including Intel's Loihi[23] and IBM's TrueNorth.[24] However, these commercial implementations use conventional MOSFETs for their digital circuitry, which as discussed above, suffer from large OFF-current and Boltzmann limited SS, further degraded due to non-optimal electrostatics offered by the 3D-material Si.[4] 2D-M, with their inherent thinness and pristine interfaces not only improve electrostatics, thereby offering SS improvements, but their low density of band-tails and trap-states help with carrier transport (Figure 4b), thereby helping improve the overall energy efficiency for both logic and memory devices.

As devices become more energy-efficient through fundamental material syntheses and device scaling approaches, interconnects however, have not seen similar improvements in energy efficiency through scaling (as discussed in the section titled "Low-energy connectivity with graphene interconnects"),[25] which implies that we are fast approaching a point in computing where the energy to move the data around in circuits dominates the energy required to perform computation on the data itself, a phenomenon which is commonly referred to as the "memory-wall."[26] In fact, in conventional ICs it is estimated that the interconnect energy consumption can contribute up to twice the energy consumption by the active devices,[27] and this problem is going to exacerbate as we further move to low-energy low-SS logic devices. Therefore, it is imperative that future interconnect technology advances, primarily on material domains, must be

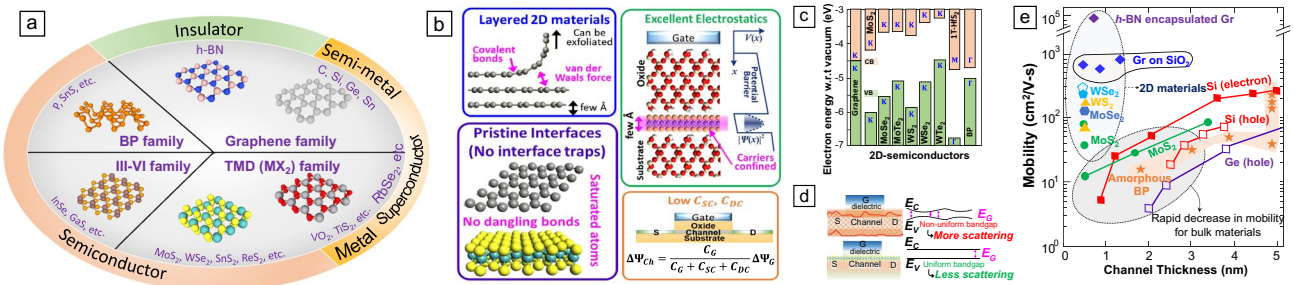

Figure 3. Overview of 2D-materials. (a) Broad 2D-materials (2D-M) family including semimetals (e.g., graphene), semiconductors (e.g., transition-metal dichalcogenides (TMDs)), and insulators (e.g., hexagonal-boron-nitride (*h*-BN)).[5] (b) Schematic illustrating the 2D-nature of these materials held together by weak vdW forces along the out-of-plane direction, and the excellent electrostatics offered by these ultra-thin materials with pristine interfaces, resulting in an overall better channel potential modulation with an applied gate bias. C_{SC} is the source-channel capacitance, C_{DC} is the drain-channel capacitance and C_G is the gate capacitance. ψ_{ch} and ψ_G are the channel surface potential and applied gate bias, respectively. (c) Band-offsets of conventional 2D-materials w.r.t vacuum reference, showing conduction band (CB) and valence band (VB) with their respective minima and maxima. (d) The pristine thickness control of 2D-TMDs over conventional bulk materials allows for low variability in the bandgap, thereby unaffecting the mobility at extreme thinness. The large bandgap variation in a thin bulk semiconductor, however, leads to enhanced scattering and degrades the mobility. (e) Mobility comparison of various 2D-M as a function of channel thickness, showing significant mobility advantages at thinner channel thickness, w.r.t bulk materials, which exhibit rapidly degrading mobility values. Data has been collected from various experimental results reported in the literature.

concurrently explored, to simultaneously allow continued scaling and low-energy computing. Graphene (Gr) is one such 2D-M, which has shown tremendous potential to be a prospective interconnect technology, primarily due to its ultra-high current carrying capacity, and ability to be engineered to exhibit resistivity and resistance values (per unit length) much lower than that of state-of-the-art metal interconnects.[28] Moreover, Gr-based interconnects can be significantly thinned down due to their high current carrying capacity,[28,29] resulting in a significant reduction in wire capacitance and a corresponding increase in the circuit performance and operating frequency.[28,30] Additionally, these 2D-enabled devices and interconnects can be uniquely engineered to create a monolithically integrated 3D stack of 2D-M (M3D), resulting in ultra-high-density integration for next-generation low-energy computing and communication[31,32] technologies.

This paper, therefore, looks at future solutions for low-energy computing enabled by 2D-M, and is organized as follows: The section titled "Low-energy computing with 2D-material enabled tunneling-FETs" discusses the recent advancements in the design of 2D-M enabled vertical- and lateral-TFETs, the next section titled "Beyond von Neumann low-energy computing enabled by 2D-based devices" discusses the recent advancements in the memory space, specifically flash devices and Resistive RAM (RRAM), and the section titled "Low-energy connectivity with graphene interconnects" discusses the prospects of Gr interconnects and 3D integration with 2D-M, for low-energy computing.

Low-energy computing with 2D-material enabled tunneling-FETs

Tunnel FETs work on the principle of band-to-band tunneling (BTBT) (Figure 4c),[12,33,34] where carriers upon the application of a suitable gate-source voltage can tunnel from the occupied source-valence band (VB) to the empty channel-conduction band (CB), thereby constituting a sharp turn-on of the drain current, especially since OFF-current is suppressed by filtering-off the high-energy carriers in the tail of the Boltzmann distribution.[35] The extreme sensitivity of the tunneling probability of carriers to the junction electric field, modulated by the applied gate-source bias, results in a highly non-linear current–voltage characteristics, where the gate voltage dependent SS is minimum at the onset of BTBT, and progressively degrades as the drain current increases.[35] For effective energy-efficient operation, therefore, it must be ensured that the average TFET-SS, at least over 4 decades of ON current swing, is around 40 mV/decade to obtain a benefit in energy efficiency of ~50% over ideal MOSFETs (detailed calculations are shown in the section titled "Low-energy connectivity with graphene interconnects"). However, TFETs fabricated from conventional 3D-materials like Si, Ge, and III-V compounds, exhibit either low ON-current, or a low-SS only at very low current value, due to presence of trap states and non-optimal electrostatics.[36–38] These problems can be significantly alleviated with 2D-M,[33–35,39] which by virtue of their ultra-thin pristine body improve electrostatics, and their comparatively thinner depletion widths with strong band-tail suppression, comparable effective masses and bandgap, and choice for a wide variety of source-channel heterojunctions, offer benefits in transport physics, that together (Figures 3, 4b) enable 2D-TFETs to achieve both ultra-low OFF current and low SS, therefore, providing promise to realize high-performance (HP) TFETs.[35] Moreover, their layered nature eases the fabrication challenges of fabricating vertical junctions, thereby making vertical-TFETs in addition to conventional lateral TFETs, easily realizable.

Vertical 2D-TFET

The first demonstration of a 2D-M-based vertical tunneling device[40] involved a field-emission type tunneling current along

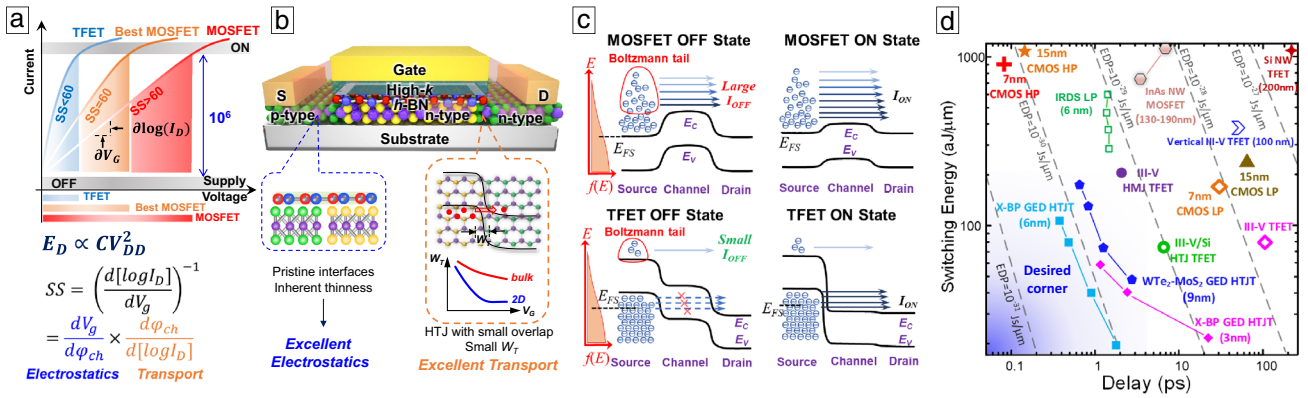

Figure 4. Low-energy computing with 2D-TFETs. (a) Motivation for low-SS devices which allow simultaneous supply voltage (V_{DD}) and energy (E_D) scaling. SS is determined by the device electrostatics and transport mechanism. Dynamic energy (E_D) is proportional to capacitance (C) and V_{DD}^2. φ_{ch} is the surface potential. (b) Schematic of a lateral single gated 2D-TFET fabricated with p-type 2D-TMD source and n-type 2D-TMD channel and drain regions, with 2D-insulator *h*-BN for the gate dielectric. The use of 2D-TMDs with pristine interface and inherent thinness allows for excellent electrostatics, while the choice for heterojunctions (HTJs) with optimum band alignment and smaller depletion widths improve the transport efficiency. (c) Operational principle of a MOSFET w.r.t a TFET. The energy-filtering of the Boltzmann tail in TFETs leads to very low OFF-current while the alignment of the source-VB and channel-CB allows for a sudden increase of the TFET ON-current, leading to low SS. (d) Atto-Joule computing enabled by lateral 2D HTJ TFETs[53] compared to other state-of-the-art TFETs and MOSFETs showing benefits in energy efficiency, also quantified in terms of the decreasing energy-delay product (EDP).

the out-of-plane direction in 2D-M between two adjacent layers of Gr separated by hexagonal-boron nitride (*h*-BN) tunnel barrier (TB). This was followed by a similar experiment,[41] where the TB was replaced by the lower bandgap few-layered WS_2. Similarly, a tunneling diode using 2D-TMDs, based on a vertical heterojunction (HTJ) of MoS_2-WSe_2 was demonstrated[42] where the bottom- and top-electrodes controlled MoS_2 and WSe_2, respectively, and negative differential resistance (NDR) behavior with a small tunneling current (~nA/μm^2) were observed. These experiments were followed by the demonstration of interband tunneling current in a vertical 2D HTJ TFET[43] where tunneling was observed between gate modulated p-type WSe_2 and a degenerate n^+-$SnSe_2$ layer. A relatively large minimum SS of 100 mV/decade with a 4-decade average SS of ~260 mV/decade, and a low ON-current of 100 nA/μm were observed, unpromising for low-energy applications. Based on a similar material combination, Yan et al.[44] demonstrated a vertical TFET and achieved a minimum SS of 37 mV/decade, but with a 4-decade average SS of 317 mV/decade and an ON current of 2 μA/μm. A different material combination of WSe_2-MoS_2 for vertical TFETs[45] yielded a 4-decade average SS of 676 mV/decade and ON current of 100 nA/μm. The average SS (over 3 decades) was subsequently improved to 414 mV/decade in Reference 46 with a 50 nm high-k HfO_2 dielectric, however, the ON-current was limited to only 3 nA/μm, too low for any practical application. Recently, a vertical TFET was demonstrated based on a WSe_2-$SnSe_2$ HTJ[47,48] with 10 nm of back gate HfO_2 dielectric and a 4-decade average SS of 174 mV/decade was observed, but with a similar low ON-current of 10 nA/μm^2.

Therefore, despite years of research into realizing high ON-current and low SS vertical TFETs based on 2D-TMDs, the search has proved to be elusive, mainly due to the lack of suitable doping techniques needed to increase the tunneling electric field. 3D-materials like Ge/Si can be degenerately doped with techniques employed in the IC industry and are therefore, good options for designing the source terminal of the TFET, while the channel can be realized using 2D-M. Therefore, such a 3D-2D HTJ TFET can theoretically yield both low SS and high ON-current as was demonstrated in Reference 49 where a Ge-MoS_2 source-channel HTJ resulted in a minimum SS of only 3.9 mV/decade and an average SS of ~31 mV/decade over 4 decades of current swing (from 0.1 pA to 1 nA), and at a low power-supply voltage of 0.1 V; all achieved with a bilayer MoS_2 channel thickness of 1.3 nm, thereby indicating the potential for ultra-low-energy operation with extreme device scalability. These low SS values have been confirmed to be achievable with rigorous band-tail analysis of 2D semiconductors, and various 2D-2D and 2D-3D HTJs.[7] However, the use of VLSI incompatible high-k solid polymer (ionic gate) dielectric and a comparatively low ON-current of 0.3 μA/μm^2 limit the viability of this specific 2D-TFET for most practical applications. Based on a similar idea, a vertical 3D (Si)—2D (MoS_2) HTJ TFET was demonstrated in Reference 50 that achieved a 4-decade averaged SS of 77 mV/decade with an ON-current of 0.013 μA/μm^2.

Therefore, although 3D-2D TFETs have achieved better performance compared to 2D-2D vertical TFETs, the presence of the naturally existing vdW gap along with high carrier effective mass[51,52] along the out-of-plane direction of any 2D-M act as a deterrent to achieving high ON-current. A lateral 2D TFET architecture[53] with well-designed source-channel HTJ can remedy this.

Lateral 2D-TFET

The design of 2D-TFETs in a lateral configuration was first proposed and theoretically evaluated in Reference 53 (Figure 4b). Such design was shown to be further improved with optimal material combinations resulting in a source-channel staggered HTJ, which reduces the effective barrier height and results in a simultaneous improvement of both ON-current and SS. Through rigorous simulations, the material combination of WTe_2-MoS_2 in a gated Esaki diode (GED) structure was found to be most optimal, that is, capable of generating both high ON current (> 728 μA/μm) and a 4-decade average SS of 18 mV/decade at a channel length of 12 nm and V_{DD} of 0.5 V. Although, GEDs are scalable down to channel length of 3 nm, the increased OFF current due to direct source-drain tunneling at channel lengths below ~ 6 nm motivates the use of a larger effective mass material, such as BP (~1.5 m_0) or SnS_2 (~2.1 m_0) for the channel. Figure 4d shows the energy-delay benchmarking of the various 2D HTJ GEDs with different V_{DD} (0.2–0.5 V) w.r.t IRDS requirements (L_g = 6 nm) as well as some state-of-the-art TFETs (from simulation and experiments) and CMOS devices, showing unmatched delay and switching energy metrics and a correspondingly smaller energy-delay product (EDP), thereby favoring lateral 2D-TFETs. Moreover, since TFETs, particularly GEDs with optimized source-channel material,[53] are immune to channel length scaling, they can be suitably scaled down to sub-10 nm dimensions, which also lowers the net device capacitance. This reduction in the device capacitance, therefore, minimizes the need for high ON-current in these scaled devices (< 10 nm channel length) to achieve same delay penalty, and they can be operated in the desired corner of Figure 4d (delay and switching energy less than 1 ps and 100 aJ/μm, respectively) provided an ON-current greater 262 μA/μm is achieved (assuming 0.5 nm effective oxide thickness).

On the experimental front, a lateral homojunction (HMJ) TFET based on BP as channel material was demonstrated in Reference 54 that achieved an ON-current of 200 μA/μm^2 (at V_{DS} = 0.8 V), but with a poor minimum SS of 170 mV/decade. Recently, a lateral reconfigurable-TFET utilizing multilayer BP for the source and monolayer BP for channel and drain, resulting in a broken source-channel HTJ, was demonstrated,[55] where a low 4-decade average SS of 24 mV/decade was obtained with a moderately low ON-current of 3 μA/μm. Although very promising, the need for a large back gate voltage of − 45 V for suitable electrostatic doping of the p-TFET renders its potential for energy-efficient computing weak. Moreover, the use of chemically unstable BP (to ambient environment) demands careful encapsulation, impairing large-scale manufacturability.

The biggest hindrance to experimentally realizing HP 2D-TFETs has been the quality and defect-free growth of these HTJs with suitably high source/drain doping. However, recently, significant progress has been made in this domain where experimentalists have devised ways to introduce substitutional impurities during the CVD (chemical vapor deposition) growth of 2D-TMDs to highly dope the material[56,57] and realize perfect edge-contacted HTJs.[58,59] Therefore, although it has not been feasible to realize a HP TFET with the characteristics predicted by Reference 53, and promising results from References 49 and 55, it is reasonable to believe that with advances in synthesis and doping of 2D-M, along with reduction in contact resistance to 2D semiconductors,[60,61] suitably designed lateral 2D HTJ-TFETs can achieve more than tenfold benefit in energy-efficiency w.r.t state-of-the-art logic devices (Figure 4d), thereby revolutionizing computing. Moreover, in addition to their merits in operation of low-energy low-frequency circuits (more details in the next section), optimally designed GED TFETs with their high ON-current of ~ 1 mA/μm, approaching that of HP-FinFETs, along with their 4-decade average SS of ~ 18 mV/dec can be equally attractive options for implementing high-frequency conventional circuits.

Beyond von Neumann low-energy computing enabled by 2D-based devices

Apart from using 2D logic devices for energy-efficient conventional VN computing, alternate computing architectures, like NM and in-memory computing (IMC) has generated immense research interest lately. While NM architectures implement spiking neural nets (SNNs),[23,24,62] where the implemented hardware emulates biological neurons and synapses (**Figure 5**a–d), IMC architectures enable fast computation of multiply and-accumulate (MAC) and matrix-vector-multiplication (MVM) operations (Figure 5e–g)[63] in artificial neural nets (ANN),[64,65] which are necessary for realizing machine learning (ML)-based AI applications through back-propagation and error correcting algorithms. Figure 5 shows the outline of both NM and IMC architectures, and the benefits in low-energy computing offered w.r.t conventional VN computing by both. NM computing architectures derive benefits in energy efficiency from both lower frequency of neuronal spikes that incrementally increase the membrane potential eventually leading to neuron firing, and massive parallelization (the human brain roughly connects 1 neuron to 10^4 synapses). In comparison, IMC architecture saves energy by minimizing the energy lost for data communication between separate computing- and memory-elements. The ultimate goal for hardware enabled NM architectures would be to reach the ultra-energy efficiency of the human brain, which in fact, consumes four orders less energy for cognitive tasks w.r.t the current state-of-the-art VN computing machines. However, it is important to note that, while in theory, IMC and NM computing can offer significant benefits to computing efficiency, this can only be achieved once the energy consumption of the accompanying control circuitry (CC) can be minimized. While this can surely be bettered by the introduction of new memory/computing/connectivity solutions as discussed further in this paper, optimization of the computing architecture is also of utmost importance. For example, it has

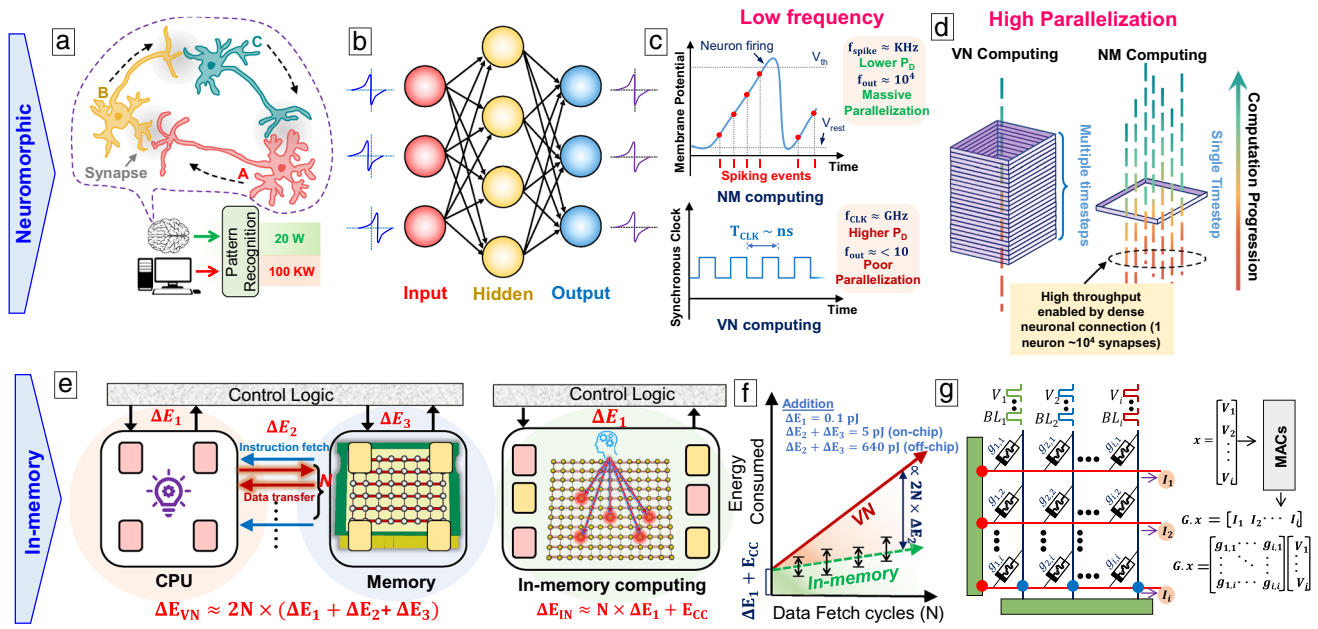

Figure 5. Overview of NM and in-memory computing and their promise in enabling low-energy computing. (a) Biological neuronal connection of neurons A, B, and C and their respective synaptic connections. The signal flows from neuron A to B to C. These dense synaptic connections enable an extremely high computational efficiency for the human brain, which in fact, consumes only ~ 20 W of power for pattern recognition, w.r.t state-of-the-art computer, which consumes ~ 100 kW. (b) Hardware implementation of the brain-inspired NM computing through a spiking-neural-net (SNN) where the inputs and outputs to the network are neuronal spikes. (c) The promise of low-power NM computing comes through two major factors: (1) reduced frequency (~ kHz) of operation as compared to ~ GHz clock frequency for conventional VN computing, thereby leading to low dynamic power (P_D); (2) the massive parallelization enabled by the dense neuronal connection in the brain where each neuron connects to ~ 10^4 synapses (fan-out (f_{out}) of 10^4), while VN computing generally has a low f_{out} (<10). Neuronal membrane potential increases with each incoming spike, which causes a generation of an output spike once this membrane potential exceeds a threshold (V_{th}), leading to resetting of the neuron membrane potential to the resting potential (V_{rest}). (d) Highly parallelized NM circuit completes multiple computational tasks (for algorithms leveraging parallelization) in a single timestep w.r.t VN computing, which needs multiple timesteps to achieve the same. (e) Schematic of a conventional VN (left) and in-memory (right) architectures. The VN computer with separate computing and memory units wastes energy for data movement (ΔE_2) among them, in addition to individual computing (ΔE_1) and memory (ΔE_3) energies. Individual computing and memory units are shown with colored rectangles. For an operation needing N cycles of data fetch, the total energy consumption for the VN computer is also shown, where the factor of 2 represents the storing- and fetching- of stored data. An in-memory computing architecture on the other hand, combines computing and memory units within (shown with distributed solid color bars), and therefore, consumes energy only for computation. E_{CC} represents the energy consumed by the peripheral control circuitry (CC) of in-memory computing architectures, and combines energy consumed by analog to digital- (ADC) and digital to analog- (DAC) converters, in addition to energy consumed by peripheral memory units in storage and retrieval of data from memory. These constitute additional energy penalties that are accounted for by error bars in (f), where comparison between the energy consumed by in-memory and VN computing architectures as a function of the data fetch cycles is sketched (the dashed line representing energy consumption of IMC architectures represent the mean). The large energy requirements for on-chip cache (~ 5 pJ) and off-chip memory (~ 640 pJ) data access as compared to a simple computing task of ~ 0.1 pJ highlights the stark difference between these two architectures as N grows large. (g) Circuit layout and electrical wiring for a hardware implementation of in-memory computing network in a single clock cycle. The output vector (I_j) can be obtained using the matrix multiplication of the row bit-line (BL) voltage inputs V_i with the conductance matrix of the individual resistive elements ($g_{i,j}$).

been estimated[66] that while energy consumption of ADC and DAC circuits can deteriorate the energy consumption of an analog-IMC architecture by 12×, an additional 10× energy efficiency is degraded due to memory access overhead from on-chip to off-chip memory buffers. These comprise significant additional overhead that need focused research effort, mostly on the architectural level,[67] that will otherwise limit the practicality of these alternate computing architectures. Nevertheless, this section discusses the hardware opportunities for implementing both NM and IMC architectures by exploiting emerging 2D-M.

In the NM/SNN architecture, which works in an event-driven update scheme, that is, computation is initiated upon receipt of a specific input, nonvolatile memory (NVM), devices have been extensively used for implementing the synaptic functionality between neurons (Figure 5a), while low-power devices, like TFET, can be used for realizing neuron spiking behavior. Learning rules, deriving their inspiration from biology, like spike time-dependent plasticity (STDP) (**Figure 6**a),[68] are based on the Hebbian learning principle[69,70] that programs the synaptic weight based on the time difference of two neuronal firing events. A

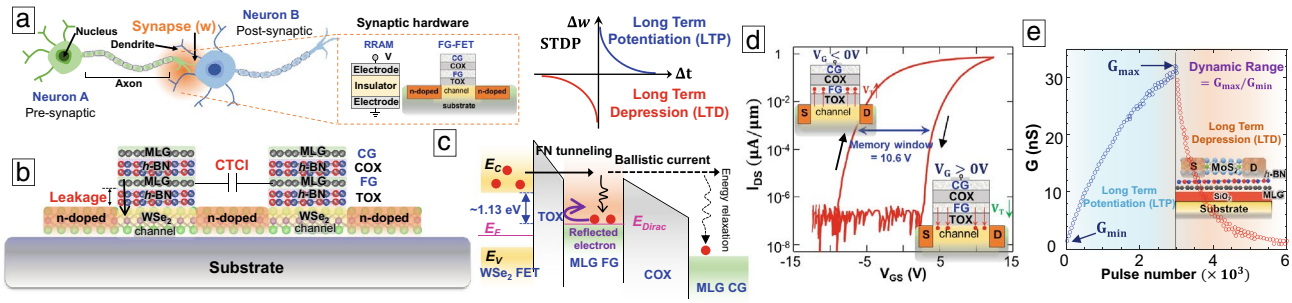

Figure 6. FG-FETs and their use as synaptic devices. (a) Biological neuronal and synaptic connections, with synaptic hardware implementations employing FG-FET and RRAM. Biological STDP curve for updating the change of synaptic weight (Δw) is also shown, Δt represents pre- and post-synaptic firing timing events, where neuron A is the pre-synaptic neuron and neuron B is the post-synaptic neuron. (b) Cross-sectional view of two WSe_2 NMOS FG-FETs showing electrical constraints of—leakage of charge from MLG FG through TOX leading to loss in retention time, and CTCI between MLG-FGs of adjacent FG-FETs. (c) Shows the mechanism of current flow across the vertical cross section of one FG-FET during programming operation. Use of 1L-WSe_2 for FG-FET channel and MLG for FG allows for a large electron barrier height of 1.13 eV between their respective conduction band energies (E_C), which inhibits tunneling of stored charges back into the FG-FET, thereby helping with retention time. The Fermi (E_F)- and Dirac cone (E_{Dirac})- energies of the FG-FET and MLG are also shown. (d) Transfer characteristics of an FG-FET demonstrating hysteresis during gate bias sweep (V_{GS}) due to trapped charges in the FG, leading to a memory window of 10.6 V[99] taken at a constant current of 1 nA/μm. A larger memory window permits better endurance and noise margin. Inset figures show operation. (e) LTP and LTD in synaptic FG-FET[96] by application of identical pulses (–15 V for LTP and +15 V for LTD) on the gate terminal. G is the conductance. This behavior of conductance increasing with pulse number corresponds to a positive Δw, as is shown with the sketch in (a).

causal firing relationship, that is, pre-synaptic neuron firing ahead of post-synaptic neuron, causes an increase in the synaptic weight in a process called long-term potentiation (LTP), while an anti-causal firing event causes a decrease in the synaptic weight through long-term depression (LTD). Various flash-based and RRAM devices have been explored which implement this functionality, through modulation of their conductivity. However, for hardware synapses to be feasible for large-scale NM circuit implementation, certain important figures of merit (FOM) such as—low (~fJ) energy consumption per conductance state change ($E_{\Delta G}$) with high linearity (for both LTP and LTD),[71] along with the presence of at least 128 distinct conductance states (7 bits) between the maximum and minimum conductance values to allow for high-fidelity operation, with a high dynamic range of more than 10 to allow operation with high noise margins, must be met;[72] all with high endurance and retention time and low device-to-device and cycle-to-cycle variation.[73] 2D-materials by virtue of their excellent properties can help satisfy most of them.

2D-M-based flash memory

Floating gate transistor (FG-FET) and charge trap flash (CTF) are the two main constituents of the flash[74] memory. Currently, vertically integrated QLC (quad-level cell) 3D-NAND from Intel,[75] BiCS (bit-cost scalable) from Toshiba,[76] and TCAT (terabit cell array transistor) from Samsung[77] are the major industry drivers of this technology.

Figure 6b shows the schematic of two such FG-FET-based flash memory devices, implemented with 2D-M that comprises a control gate (CG), multi-layer graphene (MLG)-based floating gate (FG), *h*-BN control-oxide (COX) and tunnel-oxide (TOX) and monolayer (1L) WSe_2 as FG-FET channel material, and Figure 6c shows the carrier-transport across the vertical cross section of such a device. Application of a large positive bias to the CG results in tunneling or hot carrier injection of electrons from the FG-FET channel into the FG, increasing the threshold voltage and turning the device OFF (bit 0) (for an n-type FG-FET), while a negative bias moves these deposited electrons back, programming the device back into ON-state (bit 1). Since the data retention time is determined by the retention time of tunneled electrons in the FG, hence, thicker COX and TOX are preferred which degrade FET electrostatics. However, judiciously designed 2D-FG-FETs with WSe_2 for the channel material and MLG for FG,[78] for example, can offer a distinct band-offset between the channel and FG E_c (conduction band energy) (Figure 6c), thereby, helping improve retention time even with thinner TOX. Moreover, 1L-WSe_2 with its inherent thinness help improve electrostatics and the relatively thin MLG minimizes the cell-to-cell interference (CTCI) due to lower FG-to-FG coupling capacitance,[78] thereby, enabling further scaling and compaction. Therefore, 2D-M are very promising for use in designing next-generation FG-FETs, including CTFs,[78–80] particularly for feature sizes beyond the 10-nm node.[81] In addition to these geometric properties, the large effective mass of carriers in the out-of-plane direction in 2D-M[51,52] suppresses tunneling leakage current, thereby alleviating the problems faced by conventionally used poly-Si; while their strong in-plane bonding prevents unwanted metal ion diffusion into gate-dielectric, preserving its quality and solving the problems faced by metallic FGs.[82] Also, the presence of vdW gap improves noise performance[83] leading to larger retention times. Moreover, 2D-M can be easily stacked in the out-of-plane direction, resulting in easier 3D

integration,[31,32] and therefore complement the current industrial 3D NAND technology (see subsection B of section titled "Low-energy connectivity with graphene interconnects").

Experimental 2D FG-FET demonstrations

This potential of atomically thin nanomaterials had been recognized with the first experimental effort on designing an FG-FET with the low-dimensional material, carbon nanotube (CNT),[84] where the CNT was used for designing both CG and FG to achieve a memory window of 400 mV. Subsequently, 2D-M like graphene-oxide (GO),[85] Gr,[86] and MLG[87] were utilized as the FGs for fabricating flash devices, resulting in large memory windows (Figure 6d), highlighting its prospects. The first flash memory device employing 2D-layered materials for both the FG (MLG) and the FET (MoS_2) was designed with Gr as the source/drain electrodes, and a large memory window of 8 V was achieved.[88,89] This was followed up by demonstrations with Gr/MoS_2[90] and MoS_2/BP[91] for FG/FET, respectively. In addition to these prototype demonstrations, the prospects of designing all 2D-enabled FG-FETs with Gr as the FG in flash, and graphene nanoribbon (GNR) as the FG in charge trap memory, was proposed and extensively studied in Reference 78, which is now being implemented in commercial flash cell designs albeit with conventional materials.

In addition to the discussed FG implementations, 2D-FETs on account of their large bandgap can also prevent unwanted gate-induced-drain-leakage (GIDL)[92] current that sets an upper limit on the refreshing frequency of DRAMs, and therefore, improves energy efficiency. Moreover, while alternate memory devices employing 2D-ferroelectrics have been explored,[93] where 2D-M can offer benefits in energy efficiency over conventional 3D ferroelectrics, discussion on them were skipped in favor of more mature FG-FET and RRAM technologies, which are either commercially employed or have been intensively studied for promising applications in emerging computing architectures.

Implementation of 2D FG-FET for beyond VN computing

The above discussed benefits of using 2D-M in designing FG-FETs also make it an attractive solution for designing synaptic devices and arrays, targeted toward NM and in-memory computing. Particularly, the excellent electrostatics offered by 2D-M can lead to much desired improvements in—energy-efficient synaptic potentiation and depression, increase in number of conductance states, and a linear and symmetrical conductance change, thereby leading to larger noise margins. This tremendous prospect has been suitably recognized with various experimental efforts. For example, Yi et al.[94] demonstrated synaptic behavior in a bottom-gated (BG) double FG structure with a MoS_2 channel with *h*-BN-MoS_2-*h*-BN-Au gate stack on a SiO_2–Si substrate. The device exhibits linear and symmetric conductance change with retention time $> 10^4$ s, however; the requirement of a large gate bias of 29 V restricts its potential for use in realistic low-energy circuits. This need for large programming voltage was addressed in Reference 95 where an extended Gr FG was used to enhance the electrostatics, and a programming voltage of ~ 7 V with ultra-low energy of 5 fJ were achieved, but with a larger device footprint restricting its potential for miniaturization. FG devices with MoS_2 as the channel material and Gr for the FG have been studied extensively.[95–97] The first demonstration of STDP behavior in a top-gated 2D-FG synaptic transistor, employing CVD- and metal–organic CVD (MOCVD)- grown MoS_2 (for FET) and Gr (for FG) respectively, with a low linearity factor of 2 and large number (100) of conductance states was achieved in Reference 98. While Tang et al.[96] achieved large (~ 400) conductance states (Figure 6e) with a non-linearity factor of ~ 1 with good endurance ($> 10^5$ cycles) and a large retention time of 10^5 s, its energy consumption per potentiation/depression pulse of 18 fJ was worse than that of 7.3fJ in Reference 97 and 5 fJ in Reference 95. However, both References 95 and 97 require the use of complicated pulse inputs for accomplishing the STDP behavior that need additional circuitry and dissipate energy. Therefore, it is important to note that although STDP behavior has been demonstrated in FG-FETs, they require careful waveform engineering of the input pulse trains to implement the SNN architecture specific STDP curves that inevitably introduce additional complexities. Hence, Figure 6e only shows the modulation of the FG-FET conductance with application of pulse inputs, which is the fundamental step toward achieving eventual synaptic behavior. FG-FETs have also been used for performing logic in-memory computation as shown recently,[99] where logic operations of NOR and NAND in MoS_2-based FG-FETs were demonstrated. Although a relatively large programming voltage of ~ 12 V and low voltage gain (< 2) limits energy efficiency, its transfer-free fabrication highlights the opportunity for large-scale fabrication of 2D-ANN arrays in the future.

Therefore, although significant progress has been made in the demonstration of 2D-enabled FG memories targeted for NM and IMC, its applications in mimicking a neuron spiking behavior in SNN and executing MAC operations in ANN are still missing experimentally. This is primarily due to the relatively immature 2D-fabrication process w.r.t CMOS process, restricting large-scale manufacturability needed for fabricating additional circuitry for demonstrations of SNN neurons and ANN arrays.

2D-M enabled RRAM

Although the flash memories described earlier can theoretically implement MAC operations for ANN implementation of IMC, the need for three control terminals and their relatively low endurance and long write times, in addition to large programming voltages, impose an upper limit on how trainable, fast, and energy-efficient the implementation will be. Resistive RAM (RRAMs) are an alternative,[100–102] which are two-terminal NVM elements that exhibit higher endurance and ON–OFF ratios, thereby making the operation more noise insensitive. Moreover, their resistance, like FG-FETs, can be tuned granularly by application of pulsed voltage biases making multi-level

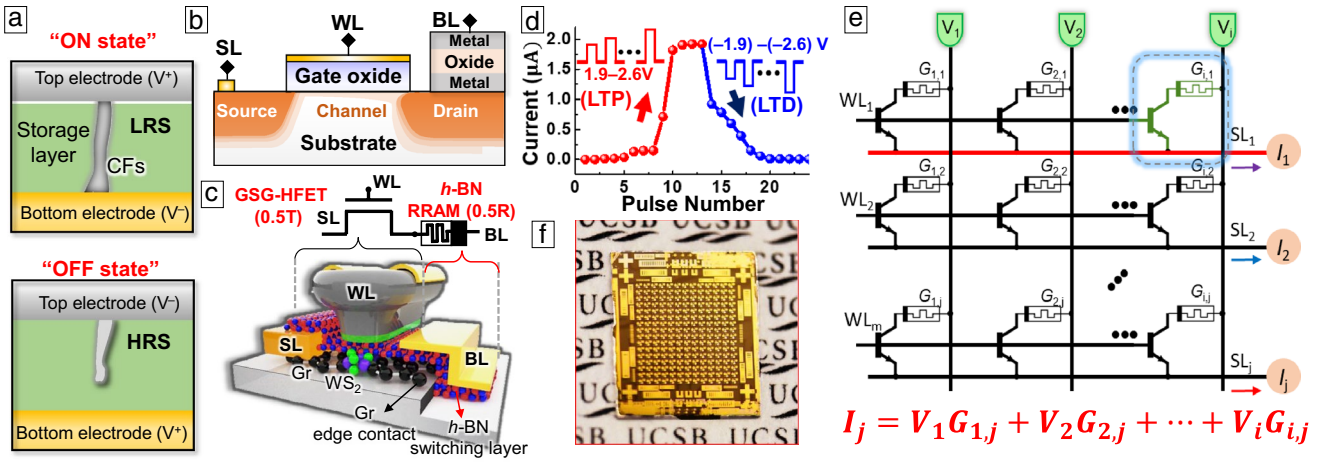

Figure 7. RRAM and its use for NM and IMC. (a) Switching mechanism of an RRAM where creation and rupture of conducting filament programs the device into a low resistance state (LRS/ON) and high resistance state (HRS/OFF) state, respectively. (b) Schematic of a conventional 1T1R-RRAM with discrete transistor and memory. (c) Schematic of the fabricated 0.5T0.5R-RRAM, which merges the transistor and memory into a single device. (d) Application of potentiation and depression pulses to the RRAM bit-line (BL) increases/decreases the current, that is, the conductance, respectively, mimicking LTP and LTD operation in NM applications.[102] (e) Illustration of N × M crossbar-network with 0.5T0.5R-RRAM memory cells for ANN based in-memory computing. The output current (I_j) through each select-line (SL) is the multiplicative sum of the applied input voltage (V_i) at BL and the cell conductance (G_i), as shown in the equation below. (f) Optical image of a large-scale memory array demonstrated at UCSB.

operation possible. The traditional RRAM (**Figure**7a,b) consists of a metal–insulator-metal (MIM) stack, where a conductive path between the two metal layers is created upon the application of an external bias leading to the soft breakdown of the insulator. This, therefore, programs the device from high resistance state (HRS) to low resistance state (LRS) (Figure 7a). Conventionally, compounds of oxygen (metal-oxides) are used for the insulators in RRAMs, which however, leads to severe variability and endurance issues,[103] mainly due to undesired redox reaction of generated oxygen ions with the electrodes, and unwanted diffusion of filament atoms. In this aspect, 2D insulators like *h*-BN without the presence of oxygen demonstrate superior chemical stability due to strong sp^2 hybridized bonds, larger formation energy of B vacancies[104] and dangling bond-free interface,[105] thereby alleviating oxidation reaction to metallic filaments. Moreover, its relatively large bandgap (~5.9 eV) results in a larger average window for SET/RESET operation,[106] and have therefore, become superior choices for designing RRAMs, as shown in recent literature.[101,106] In addition to the use of *h*-BN as the switching layer, several other 2D-M have been used by researchers, for example, Gr,[107] GO,[108] BP,[109] MoS_2[110,111] where the switching mechanisms were attributed due to the migration of intrinsic atomic species (e.g., O in GO, S in MoS_2), and in some cases, combined with penetration of metal ions. The GO[108] and BP[109] RRAMs demonstrated a retention time of 10^4 and 10^5 seconds, and ON–OFF current ratio of 10^3 and 10^5, respectively, with a very high endurance of 10^8 cycles.

An array of such RRAMs in a large-scale crossbar array (Figure 7e) can lead to much desired improvements in low-power energy-efficient IMC where the RRAMs, by virtue of their programmable resistance, can resemble MVM functionality for IMC.[112,113] However, the use of a sole RRAM can lead to increased leakage current through the array via current sneak paths, decreasing the efficiency, which can be combatted with the addition of a selector device, commonly a transistor (Figure 7b) that also maintains the operation voltage. Nevertheless, the doubled device count (RRAM and transistor) leads to the inevitable degradation of area efficiency. This limitation was recently overcome (References 101 and 102) wherein by innovatively combining the transistor (1T) and the RRAM (1R) into a single device, called the 0.5T0.5R memory cell (Figure 7c), the device count was halved leading to record performance (<10 ns switching-speed), energy- (~0.07 pJ/bit) and area-efficiency (smallest footprint among all reported 2D-M-based RRAM memory units), as well as great retention (10^6 s) and endurance (>1000 cycles). This is in fact, the first demonstration of a merged transistor-memory device,[114] and was achieved using *h*-BN as both the common RRAM active switching layer and the gate dielectric for the WS_2-based 2D-FET. Gate-tunable Gr were used for the electrical contacts in the FET for achieving an ultra-low contact resistance (~0.67 kΩ·μm) as observed in Reference 115. Moreover, since the lateral footprint of this device depends on the size of the transistor (the insulating *h*-BN can be only a monolayer), further scaling of the device therefore depends on the scaling

performance of the GSG (Gr-2D-semiconductor-Gr) FET, which has already been shown to be extremely scalable, down to channel lengths of few nm.[102]

RRAMs have also been used for demonstration of STDP[116] behavior, where a bilayer MoS_2 was used as the switching layer, but its low ON–OFF ratio of 10 and a comparatively high $E_{\Delta G}$ of 0.6 μJ were improved to 10^6 and 70 nJ, respectively, by use of *h*-BN as the switching layer.[117] However, both these demonstrations lacked multi-bit programming capability which was addressed (Figure 7d) in References 101 and 102 by co-integrating the RRAM with a transistor. Although Sangwan et al.[118] demonstrated RRAM operation in a single transistor, the lack of separate memory and logic units cannot avert the large OFF-state current leakage in an array operation. Moreover, the large switching time of 1 ms and design of the RRAM in Reference 118 on grain boundaries can lead to atom migration and a large device-to-device variability. Recently, Chen et al.[106] demonstrated the first wafer-scale integration of *h*-BN-based RRAMs and achieved a smaller $E_{\Delta G}$ of 22.5 pJ with multi-bit functionality and 200 ns switching time. Although these demonstrated properties of various implementations are inferior w.r.t the 0.5T0.5R-RRAM,[101,102] for the latter to be useful in a practical IMC circuit, certain other specific criteria should be met, already discussed in References 70 and 73, among which the RRAM can, at present, only meet the requirement for large ON–OFF conductance ratio (> 50), thereby necessitating further optimization of the device design. Such an optimized device can be expected to satisfy all the demands for emerging abundant-data computing targeted for an energy-efficient in-memory/brain-inspired learning network[102] (Figure 7e, f), and therefore, can also potentially be used as a bionic cell for implementing a low-energy and high-speed synapse. Note that although Figure 7e, f show a large-scale array, it is essentially a collection of individual 0.5T0.5R-RRAM cells, highlighting the capability for fabricating a functional large-scale crossbar network targeted for demonstrating IMC operations in the future. Notably, hardware implementations leveraging RRAM crossbar used as MACs to compute MVM multiplication through cycles in-memory, such as supervised cognition, equation solver, and cybersecurity applications,[119–121] have pointed out the potential of fast-speed and low-energy parallel processing in-memory beyond the conventional VN computing, which therefore, not only makes the developed compact-0.5T0.5R cell a viable solution for the advancement of high-density and superior-performance storage technology, but also, for enabling a fast, low-power, highly reliable and cost-effective solution for next-generation NM and in-memory computing.

2D-TFET-based NM circuit

The above sections described the device-based approach toward implementing both SNN- and ANN-based NM and in-memory computing, but as is clear, implementing the most biologically plausible SNN operation, where the same device represents both neurons and synapses, is either still left to be discovered, or requires the need for extensive additional circuitry for generation of programmable voltage pulses, inevitably degrading the area- and energy-efficiency of the implemented design. Commercial NM chips like Intel's Loihi[23] and IBM's TrueNorth,[24] circumvent this problem by implementing the SNN architecture on mature CMOS process technology, which as described earlier, has limited energy efficiency. The low-energy and low-operating (~KHz) frequencies of the NM circuits can, however, be very beneficial avenues for applications with TFETs, which with its ultra-low OFF-current and low SS characteristics can lead to many-fold increase in energy efficiency. To adequately quantify the performance benefits of using these devices in NM computing, a suitable robust and scalable compact model was developed[34] that is capable of capturing intricate device physics of both HMJ- and HTJ-TFETs down to channel lengths as small as 5 nm. This model was subsequently used to design and benchmark a fully functional leaky-integrate-fire (LIF) NM circuit, along with its Hebbian learning circuitry, and the performance was compared w.r.t implementations with commercial 7 nm PTM multi-gate (MG) low-standby power (LSTP) FinFET transistors.[122] A benefit in energy efficiency of close to two orders was observed in favor of the TFET implementation. Use of next-generation Gr interconnects (discussed in more detail in the next section) are expected to further improve this energy efficiency.

The lore of NM computing which promises to bring a paradigm shift in low-energy computing, approaching efficiencies of that of the human brain, requires a fully holistic approach to the entire computing platform, including computing devices which can function both as synapses and neurons, thereby, removing the interface cost and complexities. However, the search for this elusive device has yet proved unfruitful, due to the lack of reconfigurability and modularity in conventional bulk materials and their designs. 2D-materials by virtue of their lego-like structure[6] allowing for easier designing of heterojunctions, and their unique properties—including flexibility, enhanced optical-[123,124] and electrical-[125] responsiveness, can be engineered to design devices with both synaptic and neuronal functionalities, in addition to low-energy logic devices (TFETs and MOSFETs), thus, paving the way for novel NM architectures. In addition, to achieve the density of connections on the scale of the human brain, where each neuron connects to ~ 10^4 synapses, requires an extensive interconnect network, which is unachievable with current interconnect technology. The next section discusses the prospects offered by 2D-materials in tackling these challenges and paving the way to realize artificial brain.

Low-energy connectivity with graphene interconnects

While the low-energy 2D-switches and alternate computing architectures described in the preceding sections are promising candidates for low-power, energy-efficient computation, it is worth noting that the significant benefits to

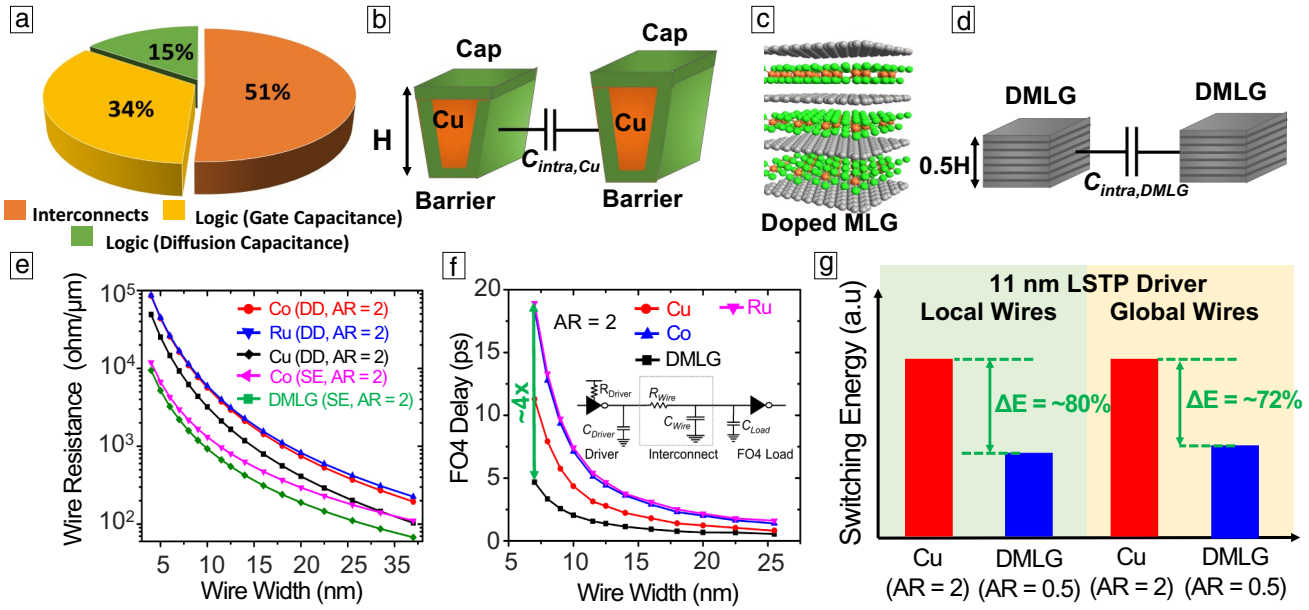

Figure 8. Next-generation low-energy computing with DGNR interconnects. (a) Power distribution in logic and interconnects for conventional computing.[27] (b) Schematic of the cross section of two-parallel copper wires of thickness H with a barrier layer. (c) Schematic cross section of an intercalation-doped MLG wire. (d) Schematic cross section of two parallel doped-MLG wires, with a thickness of 0.5H. (e) Wire resistance vs. wire width for conventional metal interconnects in comparison with DMLG wires. (f) Delay for a unit-sized inverter driving a fanout of 4 (FO4) load via an interconnect, whose length is 100x the minimum gate pitch, as a function of wire width. The simulations in (e) and (f) have been carried out using an 11 nm multi-gate LSTP driver, with an experimentally achieved doping level of – 0.6 eV[28] for doped MLG interconnects. For conventional metal interconnects, empirically derived models[129] have been used for estimating the resistance and FO4 delay. (g) Switching energy comparison between Cu and DMLG interconnects connecting an 11 nm node multi-gate LSTP driver, showing ~80% benefits in energy savings for local wires, and ~72% benefit for global wires. These simulations assume the same delay penalty of ~5% for both the local and global wires. A power-optimal repeater insertion methodology[137] has been assumed for the global wire simulations, where the optimal GNR length between two repeaters is ~0.072 mm. All the simulations in (e–g) have been carried out in HSPICE. Switching energy in (g) was evaluated by simulating a single inverter driving a FO4 circuit connected with Cu/DMLG interconnects at various frequencies and extracting the slope of the power–frequency curve. Contact resistances to DMLG have been included in the empirically derived model in Reference 28.

overall chip energy consumption cannot be made without making the "communication," that is, interconnects more efficient. This is due to interconnects in conventional ICs accounting for more than ~50% of the total circuit capacitance (**Figure 8**a), arising due to the need for connecting billions of transistors in ever increasing densities, and constitute a substantial bottleneck.[27,126] Addressing this bottleneck necessitates the need for less resistive, more reliable, thinner, and faster wiring solutions. However, traditional metallization technologies, such as copper (Cu) (Figure 8b), cobalt (Co), ruthenium (Ru), and other noble metals suffer from significant size effects as their dimensions are scaled down, mainly due to a non-linear increase in resistivity, wire and *via* resistance, which increases self-heating, degrades electromigration reliability and thereby limits their current carrying capacity and performance.[126–130] 2D-materials, such as Gr (or more specifically, doped MLG) (Figure 8c) were first proposed[131,132] as a promising solution to these existing interconnect scaling challenges, due to its superior electrical and thermal properties, and was later experimentally shown to beat the resistivity values of Cu (at sub-20 nm critical dimensions) by an appropriate level of (intercalation) doping (Figure 8c–e).[28] Intercalation doping introduces (via diffusion) foreign atoms between the layers of MLG (Figure 8c) to efficiently modulate its conductivity (via charge transfer), and the process becomes increasingly efficient for narrow MLG wires (< 20 nm). Apart from its use as a wiring technology, MLG can also be utilized as an ultra-thin capping layer for certain metal interconnects,[133] lowering resistance by up to ~15%, and also as a heat-sink, allowing for more effective thermal management.[31,32,134] Furthermore, recent advances in CMOS-compatible (i.e., below the thermal budget of 450°C and transfer free) MLG synthesis have resolved a major process integration issue, making it a strong contender for next-generation low-energy computing interconnect solution,[28,30,135,136] as further elaborated in the following section.

Implications for low-energy computing

The significantly higher (> 100-fold) current carrying capacity of doped MLG wires w.r.t conventional Cu wires (Figure 8d) allows them to be thinner, thereby lowering the inter-/intra-wire capacitances associated with MLG-based interconnects, and thus leading to significant performance boost and

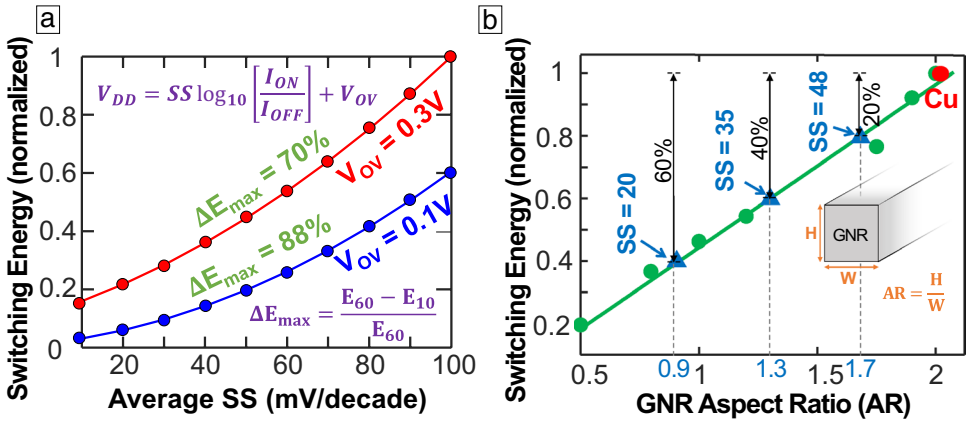

Figure 9. Energy saving benefits for a low SS switch vs. DMLG interconnects. (a) Savings in dynamic energy (normalized) w.r.t ideal MOSFET (SS = 60) as the device SS is scaled down from 100 mV/decade to 10 mV/decade for 2 different overdrive voltages (V_{OV}) and a constant I_{ON}/I_{OFF} ratio of 10^6. Equation for evaluating the operating V_{DD} is also shown. ΔE_{max} shows the maximum energy savings of a SS = 10 switch w.r.t ideal MOSFET with SS = 60. (b) Normalized switching energy as a function of the GNR aspect ratio (AR) w.r.t Cu wires (AR = 2) for a fixed wire width (W) of 20 nm. Similar driver conditions as in Figure 8 have been used in these simulations. Significant reduction in the wire capacitance, and hence, the dynamic energy consumption can be observed by moving to a lower AR DMLG interconnect, which is not possible in conventional Cu wires, due to severe resistivity and reliability degradation at lower ARs. The reduction in switching energy with AR for DMLG interconnects can be attributed to the significant reduction in intra-layer wire capacitance, which is one of the dominant factors in an ultra-dense wiring scheme for advanced technology nodes. Comparisons in energy savings possible by moving to a low SS switch with Cu interconnects, operating at an overdrive voltage of 0.3 V, are also shown where the SS values are reported in mV/decade. For example, a 48-mV low SS switch (over 6 orders of drain current swing—difficult to fabricate) offers 20% benefit in energy, similar to what is obtained by replacing Cu wires with DMLG wires with an AR of 1.7, when interconnect capacitance dominates the entire circuit capacitance.

lower switching energies.[28] This was experimentally demonstrated in Reference 28, where CVD grown high-quality DMLG showed lower resistance (Figure 8e) and delay (Figure 8f) (at sub-50 nm wire dimensions) w.r.t conventional metals, while also improving reliability by > 100-fold,[30] performance by > four-fold,[30] and energy efficiency by ~ 80/72% at the local/global wiring levels.[28] However, the CMOS-incompatible steps of transferring the MLG from the growth substrate (typically metallic) to the desired substrate along with high CVD growth temperature (~ 1000°C) had restricted the integration of MLG wires for large-scale low-energy computing. This limitation was recently overcome[30,135,136] by employing a unique solid-phase CMOS-compatible MLG growth[30] and intercalation doping process,[138] while maintaining similar DMLG performance, reliability, and energy efficiency benefits. Although the above-mentioned demonstration used a fixed thickness of 20 nm for DMLG interconnects, its significantly higher current carrying capacity w.r.t Cu wires (Figure 8d) allow for further lowering of its thickness, which would significantly reduce the (inter- and intra-) wire capacitances at sub-10 nm technology nodes, the key factor in determining the performance of aggressively scaled wires, thus providing significant energy and performance boosts, offering ~ 80% benefit in energy efficiency w.r.t Cu interconnects at the local wiring levels (Figure 8g) (for the same delay penalty of ~ 5%).

To highlight the significance of these energy efficiency gains, it is worthwhile to note that these energy savings benefits correspond to a VN architecture designed with logic devices demonstrating a 6-decade averaged SS of ~ 10 mV/decade (**Figure 9**a), w.r.t ideal CMOS devices (SS = 60), that requires revolutionary progress in the current state-of-the-art 2D-material syntheses and device architecture, and is a far cry from what has been achieved to date (Figure 4d), thereby conveying that the near-future of low-energy computing could greatly benefit from the integration of low aspect ratio (AR) DMLG interconnects in the mainstream CMOS technology, by allowing significant reduction in the dynamic switching energy (Figure 9b). These energy gains can be further enhanced with well-designed lateral 2D-TFET logic devices,[53] which can improve the computation energy efficiency by an additional ~ 60% due to their low SS (20 mV/decade over 6 decades of drain current), thereby leading to > 90% benefit in overall chip-scale energy (i.e., ~ ten-fold increase in the integration density with similar energy footprint). Monolithically integrated 2D-materials/devices stacks can further improve this integration density, as discussed in the following section.

2D-M enhanced monolithic 3D integration

Monolithic 3D (M-3D) integration, where multiple active layers, such as logic, memory, analog, RF are fabricated sequentially on top of each other, can not only enhance integration densities but can also alleviate the "memory-wall" bottleneck, by providing low latency, high-bandwidth, and energy-efficient communication channels across multiple levels.[139,140,141] Theoretical studies have demonstrated that due to their ultra-thin pristine body with excellent electrostatics, 2D-M-based 3D integration can improve integration density by more than ten-fold when compared to through-silicon-*via* (TSV)-based 3D

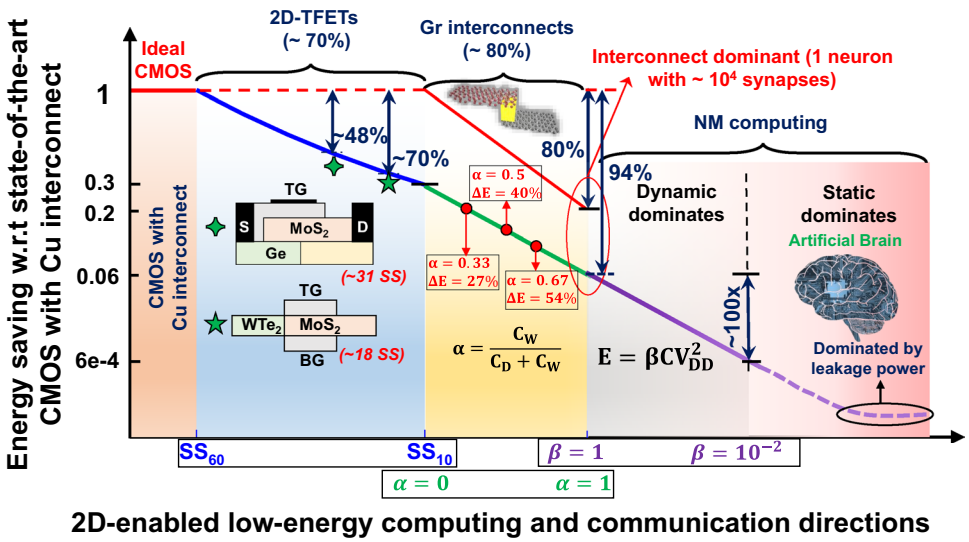

Figure 10. Energy savings from 2D-materials enabled logic and communication technologies. Assuming conventional CMOS computing, with ideal 60 mV/decade logic devices, dissipate one unit of energy for conventional VN computing with Cu interconnects, substitution of low-SS 2D-TFETs, with 6-decade averaged SS ranging from 60 to 10 mV/decade for logic devices, can lead to V_{DD} scaling with the same assumed ON/OFF ratio of 6 orders thereby leading to ~70% benefit in dynamic energy savings (V_{OV}=0.3 V). Performance projections of—demonstrated vertical TFET[49] and—proposed lateral WTe_2-MoS_2 GED,[53] with average SS of 31 and 18 mV/decade (over 4 decades) are also shown, promising energy efficiencies close to 48% and 70%, respectively. Substitution of Cu (AR=2) with Gr interconnects (AR=0.5) can reduce interconnect capacitance, thereby further improving the energy efficiency by almost ~80%, regardless of the logic device used (ideal CMOS shown with red line, or SS=10 switch shown with green line), when the entire circuit capacitance is dominated by interconnect (C_W), that is, α = 1. Cases for which C_W is 33% (α = 0.33), or 50% (α = 0.5), or 67% (α = 0.67) of the total circuit capacitance comprising device-input, parasitic, and diffusion capacitances (C_D), along with interconnect capacitances are also shown, promising energy efficiencies of 27%, 40%, and 54% over Cu interconnects, respectively. α = 1 also corresponds to interconnect dominant capacitance, expected from a full-fledged NM circuit with each neuron connecting to 10^4 synapses. Changing the computing architecture to NM computing can further lower energy because their low activity factors (β) allow for a proportionately lower dynamic switching energy. Simulations for β ranging from 1 to 10^{-2} for a SS=10 device with Gr interconnects at α = 1 are shown in purple solid line, below which the static energy due to device leakage current is expected to dominate the total energy dissipation (purple-dashed line). Overall, 2D-M with low SS devices and interconnects can lead to ~94% benefit in energy savings in conventional computing, and >100-fold higher in NM computing.

ICs, and by more than 1.5-fold when compared to conventional M-3D integration.[31,32] Recent experimental attempts of M-3D integration based on low-dimensional materials involved combining CNT-FETs with RRAM and silicon FETs on 4 different vertical layers[142] and realizing 3D-stacked 1T-1R cell based on MoS_2 FET and *h*-BN RRAM;[143] both of which although promising, necessitates the employment of a CMOS-incompatible step of 2D-M transfer, restricting the practical implementation of these technologies for low-energy computing. Additionally, the integration of 2D-M (such as doped MLG) as interconnects in the BEOL of a conventional IC can help in significantly reducing the wire thickness, thus improving inter-tier wire delay, cross-talk, and energy dissipation. Besides, the higher lateral thermal conductivity of MLG interconnects also helps in improving the lateral heat spread and remove the thermal hot-spots in 3D-ICs, which can be a major design challenge for designing ultra-scaled 3D-ICs.[32] Furthermore, the inherent thinness of 2D-M allows for significant reduction of the M-3D stack thickness, thereby reducing thermal resistance and mitigating self-heating of the upper tiers. This not only permits design of robust 3D ICs, but also minimizes temperature induced leakage power dissipation, thereby improving overall chip energy efficiency.[3,32] Finally, the lower temperature and transfer free growth techniques of 2D BEOL candidates, such as recently demonstrated MLG[30,136] can also be extended for demonstrating low temperature growth of 2D-dielectrics such as *h*-BN, as well as 2D-semiconductors (such as MoS_2, WS_2, etc.), which provides sufficient promise for integrating 2D-M in the BEOL of conventional Si ICs. Besides M-3D, while there have been several attempts at heterogenous integration of bulk Si with 2D-M via vdW epitaxy or remote epitaxy[144,145] opening up entirely new playground of materials growth and interface engineering, these techniques not only significantly increase the overall process complexity, but also involve the VLSI-incompatible step of transfer, which makes their large-scale practical implementations unfeasible.

Furthermore, apart from being an excellent interconnect material, utilizing Gr as a shielding layer reduces the inter-layer dielectric (ILD) thickness by 90%, with the added benefit of it being a good heat sink, allowing for good thermal management, thus facilitating the realization of dense 3D-ICs using 2D-M.[32] Moreover, the reduction in the inter-tier thickness using 2D-M can lead to significant improvement in the inter-tier communication speed and the energy-efficiency because of smaller interconnect parasitics.[31,32] In spite of these terrific advantages over conventional bulk materials, M-3D integration solely with 2D-M must overcome several significant process challenges

that hinder the practical realization of this technology. Even though low temperature synthesis of high-quality MLG is now achievable,[29,30,136] development of high-quality direct growth techniques of *h*-BN and 2D-TMDs at low temperatures with precise control over their thickness must still be achieved. Alternatively, heterogeneous integration of 2D-semiconductors with Si to build area-efficient vertical devices and structures, including complementary FETs, can also be explored to bring forward unique opportunities in addressing both process- and electrical-challenges of M-3D integration realized solely with a homogeneous semiconductor material system.

Summary

The article provided a comprehensive overview of the future directions for low-power energy-efficient computation and communication, specifically highlighting the potential significant benefits of employing 2D-materials in logic devices, memory, interconnects, and alternative non-von Neumann computing architectures for the same (**Figure 10**). It is shown that while well-designed 2D-TFETs for logic computing, and Gr for interconnects, can yield energy benefits of around 70% in overall computing and 80% in communication energies, respectively, they can be combined together to yield over 94% benefits in energy efficiency over the conventional CMOS-Cu computing technology. Moreover, efficient implementation of neuromorphic and in-memory computing architecture enabled by 2D-M based memory units of RRAMs and FG-FETs can lead to an additional energy benefit of more than 100-fold (Figure 10) over conventional CMOS-Cu VN computing architectures. Additionally, the seamless integration of these computing and communication blocks in a monolithically integrated heterogeneous 3D platform, all made possible with 2D-materials, can potentially deliver further benefits in energy-efficiency and integration density, paving the way for a revolutionary next-generation "brain-like" ultra-low-energy computing and connectivity platform.

Acknowledgments

The research reported in this article was supported by the ARO (Grant W911NF1810366), the AFOSR (DURIP Grant FA9550-18-1-0448), the JST CREST program (Grant SB180064), the UC MRPI Research Program (Grant MRP-17-454999), and the NSF (Grant CCF 2132820). 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.

Data availability

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

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. G.E. Moore, *Electronics* **38**, 33 (1965)
2. S.-C. Lin, K. Banerjee, *IEEE Trans. Electron Devices* **55**, 245 (2008)
3. W. Cao, J. Jiang, X. Xie, A. Pal, J.H. Chu, J. Kang, K. Banerjee, *IEEE Trans. Electron Devices* **65**, 4109 (2018)
4. W. Cao, J. Kang, D. Sarkar, W. Liu, K. Banerjee, *IEEE Trans. Electron Devices* **62**, 3459 (2015)
5. P. Ajayan, P. Kim, K. Banerjee, *Phys. Today* **69**, 38 (2016)
6. W. Cao, J.H. Chu, K. Parto, K. Banerjee, *npj 2D Mater. Appl.* **5**, 1 (2021)
7. H. Zhang, W. Cao, J. Kang, K. Banerjee, *2016 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2016), p. 30.3.1. https://ieeexplore.ieee.org/document/7838512
8. W. Liu, J. Kang, D. Sarkar, Y. Khatami, D. Jena, K. Banerjee, *Nano Lett.* **13**, 1983 (2013)
9. A. Pal, W. Cao, J. Kang, K. Banerjee, *2017 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2017), p. 31.3.1. https://ieeexplore.ieee.org/document/8268480
10. T. Sakurai, *IEICE Trans. Electron.* **E87-C**, 429 (2004)
11. T. Baba, *Jpn. J. Appl. Phys.* **31**, L455 (1992)
12. Y. Khatami, K. Banerjee, *IEEE Trans. Electron Devices* **56**, 2752 (2009)
13. S. Salahuddin, S. Datta, *Nano Lett.* **8**, 405 (2008)
14. C. Qiu, F. Liu, L. Xu, B. Deng, M. Xiao, J. Si, L. Lin, Z. Zhang, J. Wang, H. Guo, H. Peng, L. Peng, *Science* **361**, 387 (2018)
15. M.T. Björk, O. Hayden, H. Schmid, H. Riel, W. Riess, *Appl. Phys. Lett.* **90**, 142110 (2007)
16. H. Kam, Hei, D.T. Lee, R.T. Howe, T-J. King, *2005 IEEE International Electron Devices Meeting* (IEEE, Washington, DC, 2005), p. 4.6.3. https://ieeexplore.ieee.org/document/1609380
17. H. Dadgour, M.M. Hussain, K. Banerjee, *2010 ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED),* (Association for Computing Machinery, Special Interest Group on Design Automation/IEEE, Austin, TX, 2010), p. 7. https://ieeexplore.ieee.org/document/5599021
18. W. Cao, K. Banerjee, *Nat. Commun.* **11**, 196 (2020)
19. E.G. Marin, D. Marian, M. Perucchini, G. Fiori, G. Iannaccone, *ACS Nano* **14**, 1982 (2020)
20. F. Mayer, C.L. Royer, D. Blachier, L. Clavelier, S. Deleonibus, *IEEE Trans. Electron Devices* **55**, 1373 (2008)
21. H. Dadgour, M. M. Hussain, A. Cassell, N. Singh, K. Banerjee, *2011 IEEE International Reliability Physics Symposium* (IEEE, Monterey, CA, 2011), p. 3D-3. https://ieeexplore.ieee.org/abstract/document/5784489
22. C. Mead, *Proc. IEEE* **78**, 1629 (1990)
23. M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S.H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, H. Wang, *IEEE Micro* **38**, 82 (2018)
24. P.A. Merolla, J.V. Arthur, R. Alvarez-Icaza, A.S. Cassidy, J. Sawada, F. Akopyan, B.L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, *Science* **345**, 668 (2014)
25. J.D. Meindl, *Comput. Sci. Eng.* **5**, 20 (2003)
26. W.A. Wulf, S.A. McKee, *ACM SIGARCH Comput. Architect. News* **23**, 20 (1995)
27. N. Magen, A. Kolodny, U. Weiser, N. Shamir, *2004 International Workshop on System Level Interconnect Prediction (SLIP),* (Association for Computing Machinery, Paris, France, 2004), p. 7. https://doi.org/10.1145/966747.966750
28. J. Jiang, J. Kang, W. Cao, X. Xie, H. Zhang, J.H. Chu, W. Liu, K. Banerjee, *Nano Lett.* **17**, 1482 (2017)

29. K. Agashiwala, J. Jiang, C. H. Yeh, K. Parto, D. Zhang, K. Banerjee, *2020 IEEE International Electron Device Meeting* (IEEE, Virtual, 2020), p. 31.1.1. https://ieeexplore.ieee.org/abstract/document/9371986
30. J. Jiang, J. H. Chu, K. Banerjee, *2018 IEEE International Electron Devices Meeting* (IEEE, San Francisco, 2018), p. 34.5.1. https://ieeexplore.ieee.org/abstract/document/8614535
31. J. Jiang, K. Parto, W. Cao, K. Banerjee, *2018 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference* (IEEE, Burlingame, CA, 2018), p. 1. https://ieeexplore.ieee.org/abstract/document/8640131
32. J. Jiang, K. Parto, W. Cao, K. Banerjee, *IEEE J. Electron Devices Soc.* **7**, 878 (2019)
33. Y. Khatami, J. Kang, K. Banerjee, *Appl. Phys. Lett.* **102**, 043114 (2013)
34. A. Pal, W. Cao, K. Banerjee, *IEEE Trans. Electron Devices* **67**, 4473 (2020)
35. W. Cao, D. Sarkar, Y. Khatami, J. Kang, K. Banerjee, *AIP Adv.* **4**, 067141 (2014)
36. R. Gandhi, Z. Chen, N. Singh, K. Banerjee, S. Lee, *IEEE Electron Device Lett.* **32**, 437 (2011)
37. K. Tomioka, M. Yoshimura, T. Fukui, *2012 Symposium on VLSI Technology* (IEEE, Honolulu, 2012), p. 47. https://ieeexplore.ieee.org/document/6242454
38. C. Convertino, C.B. Zota, H. Schmid, A.M. Ionescu, K.E. Moselund, *J. Phys. Condens. Matter* **30**, 264005 (2018)
39. J. Kang, D. Sarkar, Y. Khatami, K. Banerjee, *Appl. Phys. Lett.* **103**, 083113 (2013)
40. L. Britnell, R.V. Gorbachev, R. Jalil, B.D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M.I. Katsnelson, L. Eaves, S.V. Morozov, N.M.R. Peres, J. Leist, A.K. Geim, K.S. Novoselov, L.A. Ponomarenko, *Science* **335**, 947 (2012)
41. T. Georgiou, R. Jalil, B.D. Belle, L. Britnell, R.V. Gorbachev, S.V. Morozov, Y.-J. Kim, A. Gholinia, S.J. Haigh, O. Makarovsky, L. Eaves, L.A. Ponomarenko, A.K. Geim, K.S. Novoselov, A. Mishchenko, *Nat. Nanotechnol.* **8**, 100 (2013)
42. T. Roy, M. Tosun, X. Cao, H. Fang, D.H. Lien, P. Zhao, Y. Chen, Y. Chueh, J. Guo, A. Javey, *ACS Nano* **9**, 2071 (2015)
43. T. Roy, M. Tosun, M. Hettick, G.H. Ahn, C. Hu, A. Javey, *Appl. Phys. Lett.* **108**, 083111 (2016)
44. X. Yan, C. Liu, C. Li, W. Bao, S. Ding, D.W. Zhang, P. Zhou, *Small* **13**, 1701478 (2017)
45. J. He, N. Fang, K. Nakamura, K. Ueno, T. Taniguchi, K. Watanabe, K. Nagashio, *Adv. Electron. Mater.* **4**, 1800207 (2018)
46. Z. Guo, Y. Chen, H. Zhang, J. Wang, W. Hu, S. Ding, D.W. Zhang, P. Zhou, W. Bao, *Adv. Sci.* **5**, 1800237 (2018)
47. N. Oliva, L. Capua, M. Cavalieri, A. M. Ionescu, *2019 IEEE International Electron Devices Meeting* (IEEE, San Francisco, 2019), p. 37.2.1. https://ieeexplore.ieee.org/abstract/document/8993643
48. N. Oliva, J. Backman, L. Capua, M. Cavalieri, M. Luisier, A.M. Ionescu, *npj 2D Mater. Appl.* **4**, 1 (2020)
49. D. Sarkar, X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, P.M. Ajayan, K. Banerjee, *Nature* **526**, 91 (2015)
50. G.H. Shin, B. Koo, H. Park, Y. Woo, J.E. Lee, S. Choi, *ACS Appl. Mater. Interfaces* **10**, 40212 (2018)
51. W. Cao, M. Huang, C.-H. Yeh, K. Parto, K. Banerjee, *IEEE Trans. Electron Devices* **67**, 1310 (2020)
52. K. Matsubara, K. Sugihara, T. Tsuzuku, *Phys. Rev. B* **41**, 969 (1990)
53. W. Cao, J. Jiang, J. Kang, D. Sarkar, W. Liu, K. Banerjee, *2015 IEEE International Electron Devices Meeting* (IEEE, Washington DC, 2015), p. 12.3.1. https://ieeexplore.ieee.org/abstract/document/7409682
54. P. Wu, T. Ameen, H. Zhang, L.A. Bendersky, H. Ilatikhameneh, G. Klimeck, R. Rahman, A.V. Davydov, J. Appenzeller, *ACS Nano* **13**, 377 (2018)
55. S. Kim, G. Myeong, W. Shin, H. Lim, B. Kim, T. Jin, S. Chang, K. Watanabe, T. Taniguchi, S. Cho, *Nat. Nanotechnol.* **15**, 203 (2020)
56. T. Zhang, K. Fujisawa, F. Zhang, M. Liu, M.C. Lucking, R.N. Gontijo, Y. Lei, H. Liu, K. Crust, T.G. Nakajima, H. Terrones, A.L. Elías, M. Terrones, *ACS Nano* **14**, 4326 (2020)
57. L. Loh, Z. Zhang, M. Bosman, G. Eda, *Nano Res.* **14**, 1668 (2021)
58. P.K. Sahoo, S. Memaran, Y. Xin, L. Balicas, H.R. Gutiérrez, *Nature* **553**, 63 (2018)
59. Z. Zhang, P. Chen, X. Duan, K. Zang, J. Luo, X. Duan, *Science* **357**, 788 (2017)
60. A. Allain, J. Kang, K. Banerjee, A. Kis, *Nat. Mater.* **14**, 1195 (2015)
61. J. Kang, W. Liu, D. Sarkar, D. Jena, K. Banerjee, *Phys. Rev. X* **4**, 031005 (2014)
62. W. Maass, *Neural Netw.* **10**, 1659 (1997)
63. Z. Wang, H. Wu, G.W. Burr, C.S. Hwang, K.L. Wang, Q. Xia, J.J. Yang, *Nat. Rev. Mater.* **5**, 173 (2020)
64. F. Rosenblatt, *The Perceptron, a Perceiving and Recognizing Automaton Project Para* (Cornell Aeronautical Laboratory, Buffalo, NY, 1957)
65. M. Prezioso, F. Merrikh-Bayat, B.D. Hoskins, G.C. Adam, K.K. Likharev, D.B. Strukov, *Nature* **521**, 61 (2015)
66. P. Houshmand, S. Cosemans, L. Mei, I. Papistas, D. Bhattacharjee, P. Debacker, A. Mallik, D. Verkest, and M. Verhelst, *2020 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2020), p. 29.1.1. https://ieeexplore.ieee.org/abstract/document/9372006
67. A. Shafiee, A. Nag, N. Muralimanohar, R. Balasubramonian, J.P. Strachan, M. Hu, R.S. Williams, V. Srikumar, *ACM SIGARCH Comput. Archit. News* **44**, 14 (2016)
68. C. Clopath, L. Büsing, E. Vasilaki, W. Gerstner, *Nat. Neurosci.* **13**, 344 (2010)
69. D.O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Psychology Press, London, UK, 2005)
70. S. Song, K.D. Miller, L.F. Abbott, *Nat. Neurosci.* **3**, 919 (2000)
71. S. Seo, J.J. Lee, H.J. Lee, H.W. Lee, S. Oh, J.J. Lee, K. Heo, J.H. Park, *ACS Appl. Electron. Mater.* **2**, 371 (2020)
72. S. Yu, *Proc. IEEE* **106**, 260 (2018)
73. P-Y. Chen, X. Peng, S. Yu, *2017 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2017), p. 6.1.1. https://ieeexplore.ieee.org/abstract/document/8268337
74. T. Hang, C. Sheng, H. Tzu, Y. Pei, W.S. Yu, H.K. Yeu, R. Liu, L.C. Yuan, presented at Int. Conf. on Solid-State and Integrated-Circuit Technology (ICSICT'08), Beijing, China, 2008, p. 807
75. K. Parat, C. Dennison, *2015 IEEE International Electron Devices Meeting* (IEEE, Washington DC, 2015), p. 3.3.1. https://ieeexplore.ieee.org/document/7409618
76. R. Katsumata, M. Kito, Y. Fukuzumi, M. Kido, H. Tanaka, Y. Komori, M. Ishiduki, J. Matsunami, T. Fujiwara, Y. Nagata, L. Zhang, Y. Iwata, R. Kirisawa, H. Aochi, A. Nitayama, *2009 Symposium on VLSI Technology* (IEEE, Kyoto, Japan, 2009), p. 136. https://ieeexplore.ieee.org/document/5200662
77. J. Jang, H-S. Kim, W. Cho, H. Cho, J. Kim, S.I. Shim, Younggoan, J.-H. Jeong, B.-K. Son, D.W. Kim, Kihyun, J.-J. Shim, J.S. Lim, K.-H. Kim, S.Y. Yi, J.-Y. Lim, D. Chung, H.-C. Moon, S. Hwang, J.-W. Lee, Y.-H. Son, U.-I. Chung, W.-S. Lee, *2009 Symposium on VLSI Technology* (IEEE, Honolulu, 2009), p. 192. https://ieeexplore.ieee.org/document/5200595
78. W. Cao, J. Kang, S. Bertolazzi, A. Kis, K. Banerjee, *IEEE Trans. Electron Devices* **61**, 3456 (2014)
79. S. Bertolazzi, P. Bondavalli, S. Roche, T. San, S.Y. Choi, L. Colombo, F. Bonaccorso, P. Samori, *Adv. Mater.* **31**, 1806663 (2019)
80. A.J. Hong, E.B. Song, H.S. Yu, M.J. Allen, J. Kim, J.D. Fowler, J.K. Wassei, Y. Park, Y. Wang, J. Zou, R.B. Kaner, B.H. Weiller, K.L. Wang, *ACS Nano* **5**, 7812 (2011)
81. S. Raghunathan, T. Krishnamohan, K. Parat, K. Saraswat, *2009 IEEE International Electron Devices Meeting* (IEEE, Baltimore, 2009), p. 1. https://ieeexplore.ieee.org/document/5424216
82. R. Degraeve, M. Aoulaiche, B. Kaczer, Ph. Roussel, T. Kauerauf, S. Sahhaf, G. Groeseneken, *2008 International Symposium on the Physical and Failure Analysis of Integrated Circuits* (IEEE, Singapore, 2008), p. 1. https://ieeexplore.ieee.org/document/4588195
83. X. Xie, D. Sarkar, W. Liu, J. Kang, O. Marinov, M.J. Deen, K. Banerjee, *ACS Nano* **8**, 5633 (2014)
84. X.B. Lu, J.Y. Dai, *Appl. Phys. Lett.* **88**, 113104 (2006)
85. S. Wang, J. Pu, D.S.H. Chan, B.J. Cho, K.P. Loh, *Appl. Phys. Lett.* **96**, 143109 (2010)
86. A.J. Hong, E.B. Song, H.S. Yu, M.J. Allen, J. Kim, J.D. Fowler, J.K. Wassei, Y. Park, Y. Wang, J. Zou, R.B. Kaner, B.H. Weiller, K.L. Wang, *ACS Nano* **5**, 7812 (2011)

87. A. Mishra, A. Janardanan, M. Khare, H. Kalita, A. Kottantharayil, *IEEE Electron Device Lett.* **34**, 1136 (2013)
88. S. Bertolazzi, D. Krasnozhon, A. Kis, *ACS Nano* **7**, 3246 (2013)
89. S. Manzeli, D. Ovchinnikov, D. Pasquier, O.V. Yazyev, A. Kis, *Nat. Rev. Mater.* **2**, 17033 (2017)
90. M.S. Choi, G.-H. Lee, Y.-J. Yu, D.-Y. Lee, S.H. Lee, P. Kim, J. Hone, W.J. Yoo, *Nat. Commun.* **4**, 1624 (2013)
91. Y. Cai, G. Zhang, Y.W. Zhang, *Sci. Rep.* **4**, 6677 (2014)
92. J. Kang, W. Cao, A. Pal, S. Pandey, S. Kramer, R. Hill, G. Sandhu, K. Banerjee, *2017 IEEE International Electron Devices Meeting* (IEEE, San Francisco, 2017), p. 31.2.1. https://ieeexplore.ieee.org/document/8268479
93. A.I. Khan, A. Keshavarzi, S. Datta, *Nat. Electron.* **3**, 588 (2020)
94. S.G. Yi, M.U. Park, S.H. Kim, C.J. Lee, J. Kwon, G.H. Lee, K.H. Yoo, *ACS Appl. Mater. Interfaces* **10**, 31480 (2018)
95. T. Paul, T. Ahmed, K.K. Tiwari, C.S. Thakur, A. Ghosh, *2D Mater.* **6**, 045008 (2019)
96. J. Tang, C. He, J. Tang, K. Yue, Q. Zhang, Y. Liu, Q. Wang, S. Wang, N. Li, C. Shen, Y. Zhao, J. Liu, J. Yuan, Z. Wei, J. Li, K. Watanabe, T. Taniguchi, D. Shang, S. Wang, W. Yang, R. Yang, D. Shi, G. Zhang, *Adv. Funct. Mater.* **31**, 2011083 (2021)
97. C. He, J. Tang, D.S. Shang, J. Tang, Y. Xi, S. Wang, N. Li et al., *ACS Appl. Mater. Interfaces* **12**, 11945 (2020)
98. E. Park, M. Kim, T.S. Kim, I.S. Kim, J. Park, J. Kim, Y. Jeong, S. Lee, I. Kim, J.-K. Park, G.T. Kim, J. Chang, K. Kang, J.Y. Kwak, *Nanoscale* **12**, 24503 (2020)
99. G.M. Marega, Y. Zhao, A. Avsar, Z. Wang, M. Tripathi, A. Radenovic, A. Kis, *Nature* **587**, 72 (2020)
100. K. Roy, A. Jaiswal, P. Panda, *Nature* **575**, 607 (2019)
101. D. Zhang, C.-H. Yeh, W. Cao, K. Banerjee, *IEEE Trans. Electron Devices* **68**, 2033 (2021)
102. C.-H. Yeh, D. Zhang, W. Cao, K. Banerjee, *2020 International Electron Devices Meeting* (IEEE, San Francisco, 2020), p. 12.3.1. https://ieeexplore.ieee.org/document/9371902
103. B. Chen, Y. Lu, B. Gao, Y.H. Fu, F.F. Zhang, P. Huang, Y.S. Chen, L.F. Liu, X.Y. Liu, J.F. Kang, Y.Y. Wang, Z. Fang, H.Y. Yu, X. Li, X.P. Wang, N. Singh, G.Q. Lo, D.L. Kwong, *2011 International Electron Devices Meeting* (IEEE, Washington DC, 2011), p. 12.3.1. https://ieeexplore.ieee.org/document/6131539
104. Z.L. Hou, M.S. Cao, J. Yuan, X.Y. Fang, X.L. Shi, *J. Appl. Phys.* **105**, 076103 (2009)
105. A. Pakdel, Y. Bando, D. Golberg, *Chem. Soc. Rev.* **43**, 934 (2014)
106. S. Chen, M.R. Mahmoodi, Y. Shi, C. Mahata, B. Yuan, X. Liang, C. Wen, F. Hui, D. Akinwande, D.B. Strukov, M. Lanza, *Nat. Electron.* **3**, 638 (2020)
107. F. Hui, E.G. Gutierrez, S. Long, Q. Liu, A.K. Ott, A.C. Ferrari, M. Lanza, *Adv. Electron. Mater.* **3**, 1600195 (2017)
108. B. Zhang, Y. Chen, Y. Ren, L. Xu, G. Liu, E. Kang, C. Wang, C. Zhu, K. Neoh, *Chemistry* **19**, 6265 (2013)
109. C. Hao, F. Wen, J. Xiang, S. Yuan, B. Yang, L. Li, W. Wang, Z. Zeng, L. Wang, Z. Liu, Y. Tian, *Adv. Funct. Mater.* **26**, 2016 (2016)
110. V.K. Sangwan, D. Jariwala, I.S. Kim, K.-S. Chen, T.J. Marks, L.J. Lauhon, M.C. Hersam, *Nat. Nanotechnol.* **10**, 403 (2015)
111. X. Feng, Y. Li, L. Wang, Z. G. Yu, S. Chen, W.-C. Tan, N. Macadam, G. Hu, X. Gong, T. Hasan, Y.-W. Zhang, A.V.-Y. Thean, K.-W. Ang, *2019 Symposium on VLSI Technology* (IEEE, Kyoto, Japan, 2019), p. T88. https://ieeexplore.ieee.org/abstract/document/8776520
112. P.M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang, W.D. Lu, *Nat. Nanotechnol.* **12**, 784 (2017)
113. M. Prezioso, I. Kataeva, F.M. Bayat, B. Hoskins, G. Adam, T. Sota, K. Likharev, D. Strukov, *2015 IEEE International Electron Devices Meeting* (IEEE, Washington DC, 2015), p. 17.4.1. https://ieeexplore.ieee.org/document/7409719
114. S. Thomas, *Nat. Electron.* **4**, 321 (2021)
115. C.-H. Yeh, W. Cao, A. Pal, K. Parto, K. Banerjee, *2019 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2019), p. 23.4.1. https://ieeexplore.ieee.org/abstract/document/8993600
116. R. Xu, H. Jang, M.-H. Lee, D. Amanov, Y. Cho, H. Kim, S. Park, H. Shin, D. Ham, *Nano Lett.* **19**, 2411 (2019)
117. Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H.-S.P. Wong, M. Lanza, *Nat. Electron.* **1**, 458 (2018)
118. V. Sangwan, H. Lee, H. Bergeron, I. Balla, M.E. Beck, K. Chen, M.C. Hersam, *Nature* **554**, 500 (2018)
119. Q. Xia, J.J. Yang, *Nat. Mater.* **18**, 309 (2019)
120. P. Yao, H. Wu, B. Gao, S.B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.S.P. Wong, H. Qian, *Nat. Commun.* **8**, 15199 (2017)
121. M.A. Zidan, Y. Jeong, J. Lee, B. Chen, S. Huang, M.J. Kushner, W.D. Lu, *Nat. Electron.* **1**, 411 (2018)
122. Predictive Technology Model. http://ptm.asu.edu/
123. X. Xie, J. Kang, W. Cao, J.H. Chu, Y. Gong, P.M. Ajayan, K. Banerjee, *Sci. Rep.* **7**, 1 (2017)
124. X. Xie, J. Kang, Y. Gong, P. M. Ajayan, K. Banerjee, *2017 IEEE International Electron Devices Meeting* (IEEE, San Francisco, CA, 2017), p. 5.3.1. https://ieeexplore.ieee.org/document/8268332
125. D. Sarkar, W. Liu, X. Xie, A.C. Anselmo, S. Mitragotri, K. Banerjee, *ACS Nano* **8**, 3992 (2014)
126. K. Banerjee, A. Mehrotra, *IEEE Circuits Devices Mag.* **17**, 16 (2001)
127. N. Srivastava, K. Banerjee, *JOM* **56**, 30 (2004)
128. S. Im, N. Srivastava, K. Banerjee, K.E. Goodson, *IEEE Trans. Electron Devices* **52**, 2710 (2005)
129. C.K. Hu, J. Kelly, H. Huang, K. Motoyama, H. Shobha, Y. Ostrovski, J.H.-C. Chen, R. Patlolla, B. Peethala, P. Adusumilli, T. Spooner, R. Quon, L.M. Gignac, C. Breslin, G. Lian, M. Ali, J. Benedict, X.S. Lin, S. Smith, V. Kamineni, X. Zhang, F. Mont, S. Siddiqui, F. Baumann, *2018 IEEE International Reliability Physics Symposium* (IEEE, Burlingame, CA, 2018), p. 4F.1.1. https://ieeexplore.ieee.org/document/8353597
130. H. Li, C. Xu, N. Srivastava, K. Banerjee, *IEEE Trans. Electron Devices* **56**, 1799 (2009)
131. C. Xu, H. Li, K. Banerjee, *2008 IEEE International Electron Devices Meeting* (IEEE, San Francisco, 2008), p. 2.0.1. https://ieeexplore.ieee.org/abstract/document/4796651
132. C. Xu, H. Li, K. Banerjee, *IEEE Trans. Electron Devices* **56**, 1567 (2009)
133. R. Chau, *2019 International Electron Device Meeting* (IEEE, San Francisco, 2019), p. 1.1.1. https://ieeexplore.ieee.org/document/8993462
134. Z. Yan, G. Liu, J.M. Khan, A.A. Balandin, *Nat. Commun.* **3**, 827 (2012)
135. S. Thomas, *Nat. Electron.* **1**, 612 (2018)
136. K. Agashiwala, J. Jiang, K. Parto, D. Zhang, C.-H. Yeh, K. Banerjee, *IEEE Trans. Electron Devices* **68**, 2083 (2021)
137. K. Banerjee, A. Mehrotra, *IEEE Trans. Electron Devices* **49**, 2001 (2002)
138. J. Kang, Y. Matsumoto, X. Li, J. Jiang, X. Xie, K. Kawamoto, M. Kenmoku, J.H. Chu, W. Liu, J. Mao, K. Ueno, K. Banerjee, *Nat. Electron.* **1**, 46 (2018)
139. J. Kang, W. Cao, X. Xie, D. Sarkar, W. Liu, K. Banerjee, *Proc. SPIE* **9083**, 908305 (2014)
140. K. Banerjee, S.J. Souri, P. Kapur, K.C. Saraswat, *Proc. IEEE* **89**, 602 (2001)
141. M. Vinet, P. Batude, C. Fenouillet-Beranger, F. Clermidy, L. Brunet, O. Rozeau, J.M. Hartmann, O. Billoint, G. Cibrario, B. Previtali, C. Tabone, B. Sklenard, O. Turkyilmaz, F. Ponthenier, N. Rambal, M.-P. Samson, F. Deprat, V. Lu, L. Pasini, S. Thuries, H. Sarhan, J.-E. Michallet, O. Faynot, *2014 SOI-3D-Subthrehold Microelectronics Technology Unified Conference (S3S)* (IEEE, Milbrae, CA, 2014), p. 1. https://ieeexplore.ieee.org/document/7028194
142. M.M. Shulaker, T. F. Wu, A. Pal, L. Zhao, Y. Nishi, K. Saraswat, H. S. Wong, S. Mitra, *2014 International Electron Devices Meeting* (IEEE, San Francisco, CA, 2014), p. 27.4.1. https://ieeexplore.ieee.org/document/7047120
143. C. H. Wang, C. McClellan, Y. Shi, X. Zheng, V. Chen, M. Lanza, E. Pop, H. S. Wong, *2018 International Electron Devices Meeting* (IEEE, San Francisco, CA, 2018), p. 22.5.1. https://ieeexplore.ieee.org/document/8614495
144. H. Kum, D. Lee, W. Kong, H. Kim, Y. Park, Y. Kim, Y. Baek, S.-H. Bae, K. Lee, J. Kim, *Nat. Electron.* **2**, 439 (2019)
145. D. Jariwala, T.J. Marks, M.C. Hersam, *Nat. Mater.* **16**, 170 (2017)

Arnab Pal is currently a PhD candidate in the Nanoelectronics Research Lab in the Electrical and Computer Engineering Department at the University of California, Santa Barbara. He received a MTech degree in microelectronics from the Indian Institute of Technology Kanpur, in 2015. His current research interests include exploring novel techniques for improving the energy efficiency and performance of nanodevices based on two-dimensional materials. Pal can be reached by email at arnab@ucsb.edu.

Kunjesh Agashiwala is currently a PhD candidate in the Nanoelectronics Research Lab in the Electrical and Computer Engineering Department at the University of California, Santa Barbara. He received a dual degree, MSc (Hons) in physics and BE (Hons) in electrical and electronics engineering from Birla Institute of Technology and Science, Pilani, India, in 2017. His current research is focused on the design and exploration of two-dimensional materials enabled interconnects and on-chip inductors for nextgeneration integrated circuits. Agashiwala can be reached by email at kunjesh@ucsb.edu.

Junkai Jiang is currently an R&D engineer at Google in Sunnyvale. He received a BSc degree in microelectronics from Peking University, Beijing, China, in 2012, and a PhD degree in electrical and computer engineering from the University of California, Santa Barbara (USCB) in 2019. Jiang's doctoral research has been recognized with the IEEE Electron Devices Society's PhD Fellowship Award (2018) and UCSB's Winifred and Louis Lancaster Dissertation Award (2020). His research interests include the modeling, design, and fabrication of emerging interconnect structures and passive devices based on two-dimensional materials. Jiang can be reached by email at junkaijiang@ucsb.edu.

Dujiao Zhang is currently a PhD candidate in the School of Electrical Engineering at Xi'an Jiaotong University, China. He received a bachelor's degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2015. In Fall 2019, he joined Prof. Banerjee's Nanoelectronics Research Lab in the Electrical and Computer Engineering Department at the University of California, Santa Barbara, as a visiting scholar where his research focused on developing novel synthesis techniques for various two-dimensional materials in electronic applications. Zhang can be reached by email at dujiao@ucsb.ed.

Tanmay Chavan is currently pursuing a PhD degree in the Nanoelectronics Research Lab in the Electrical and Computer Engineering Department at the University of California, Santa Barbara. He received a MTech degree in microelectronics from the Indian Institute of Technology Bombay, in 2018. His current research focuses on the design of novel electronic devices based on two-dimensional materials. Chavan can be reached by email at tanmaychavan@ucsb.edu.

Ankit Kumar is currently pursuing a PhD degree in the Nanoelectronics Research Lab in the Electrical and Computer Engineering Department at the University of California, Santa Barbara. He received a dual degree, BE (Hons) in electrical and electronics engineering and MSc (Hons) in physics from Birla Institute of Technology and Science (BITS), Pilani, India, in 2018. His research interests involve exploring two-dimensional materials, both theoretically and experimentally, for application in interconnect and other passive elements for next-generation integrated circuits. Kumar can be reached by email at ankitkumar@ucsb.edu.

Chao-Hui Yeh is currently working as an associate specialist in the Nanoelectronics Research Lab (NRL) in the Electrical and Computer Engineering Department at the University of California, Santa Barbara. He received a PhD degree in electronics engineering from National Tsing Hua University, Taiwan, in 2014, for his work on high-speed graphene electronics and ferromagnetic oxide materials. His research at NRL is focused on two-dimensional material-based nanotransistors and nonvolatile resistive random-access memory technology. Yeh can be reached by email at chyeh@ucsb.edu.

Wei Cao is currently a device engineer at Intel Corporation, Santa Clara. He received a PhD degree in electrical and computer engineering from the University of California, Santa Barbara (UCSB), in 2017. During 2017–2019, he worked as a postdoctoral fellow in the Nanoelectronics Research Lab at UCSB. His research interests lie in the design, fabrication, and the circuit exploration of two-dimensional material-based electron devices. Cao can be reached by email at weicao@ucsb.edu.

Kaustav Banerjee is a professor of electrical and computer engineering at the University of California, Santa Barbara, and director of the Nanoelectronics Research Lab. He received a PhD degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1999. His current research focuses on the physics, technology, and applications of two-dimensional van der Waals materials and heterostructures. Banerjee is a Fellow of IEEE, the American Physical Society, Japan Society for the Promotion of Science, and AAAS, and has been recognized with numerous awards and honors including the Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation, Germany, and the IEEE Kiyo Tomiyasu Award—one of the Institute's highest honors. In 2019, he was identified as one of the world's most influential scientific minds by Clarivate Analytics. Banerjee can be reached by email at kaustav@ece.ucsb.edu.