2023 Edition

Chapter 16: Emerging Research Devices

For updates, visit http://eps.ieee.org/hir

The HIR is devised and intended for technology assessment only and is without regard to any commercial considerations pertaining to individual products or equipment.

We acknowledge with gratitude the use of material and figures in this Roadmap that are excerpted from original sources.
Figures & tables should be re-used only with the permission of the original source.

Table of Contents

Chapter 1: Heterogeneous Integration Roadmap: Driving Force for Systems of the Future
Chapter 2: High Performance Computing and Data Centers
Chapter 3: Heterogeneous Integration for the Internet of Things (IoT)
Chapter 4: Medical, Health and Wearables
Chapter 5: Automotive
Chapter 6: Aerospace and Defense
Chapter 7: Mobile
Chapter 8: Single Chip and Multi Chip Integration
Chapter 9: Integrated Photonics
Chapter 10: Integrated Power Electronics
Chapter 11: MEMS and Sensor Integration
Chapter 12: 5G, RF and Analog Mixed Signal
Chapter 13: Co-Design
Chapter 14: Modeling and Simulation
Chapter 15: Materials and Emerging Research Materials
Chapter 16: Emerging Research Devices 1
Executive Summary 1
1. Nanoscale Vacuum Electronics 2
2. Neuromorphic Devices 7
3. Printed Electronics 14
4. Spintronic Devices 19
5. Graphene and 2D Material Electronics 25
6. Carbon Nanotube Electronics 40
7. Plasmonic Devices 45
Chapter 17: Test Technology
Chapter 18: Supply Chain
Chapter 19: Cyber Security
Chapter 20: Thermal
Chapter 21: SiP and Module
Chapter 22: Interconnects for 2D and 3D Architectures
Chapter 23: Wafer-Level Packaging, Fan-in and Fan-out
Chapter 24: Reliability

Chapter 16: Emerging Research Devices

Send corrections, comments and suggested updates to the TWG chair, using our HIR SmartSheet:

https://rebrand.ly/HIR-feedback

Executive Summary

Information technology has relied upon scaling CMOS to maintain the pace of progress since it was introduced in 1963. The approaching economic end of Moore's Law is slowing that pace of progress. Although scaling of CMOS continues, the benefits are reduced as we approach the limits of the physics for these devices. The historical benefits of lower cost, lower power, smaller size and weight, higher functional density and higher performance are reduced and, in some cases, no longer available. We are now experiencing a dramatic increase in research and development to find new devices to replace CMOS for memory and logic applications.

There have been many new candidate devices and materials introduced in recent years, but most are still in the early stages of development. The objective of this chapter is to identify new device types that hold the promise of maintaining or accelerating the pace of progress over the next 25 years, and the difficult challenges that must be overcome to realize that potential.

The increasing penetration of high-performance electronics into mobile consumer devices is resulting in a rapidly growing number of product types. These products typically have lower unit volume and shorter product life cycle than traditional computer components. Device design and engineering cost, at the same time, is rising rapidly as we approach the limit of the physics. Thus, time to market and non-recurring engineering cost will be critical challenges to overcome for emerging research devices.

We have achieved 50 years of progress through evolutionary change that has led to information technology penetrating every phase of our lives. This is no longer possible. The future must embrace revolutionary change in devices and materials if we are to maintain our historical pace of progress.

Introduction

Dimensional scaling that has been going on for more than five decades is reaching its fundamental limits. Continued gain in device speed while lowering power consumption is increasingly difficult. These factors together have created the urgent need to explore new devices for information processing and memory, new architectures, and new approaches for heterogeneous integration of existing functions for emerging applications.

The scope of this Chapter includes supporting other HIR Working Groups with new devices required to meet the difficult challenges they identify and assessing the emerging research devices and technologies for this purpose. Table 1 below lists Emerging Research Device types known today. However, the most important devices over the next 25 years are likely to be the types that we have not yet imagined.

Table 1: Candidate emerging research devices

Candidate emerging research devices
Nanoscale vacuum electronics
Neuromorphic devices
Quantum devices for information processing
Spintronic devices
Flexible electronics
3 dimensional stacked devices
Nanowire electronics
Carbon nanotube electronics
Graphene electronics
Other 2D material-based electronics
Plasmonic devices
Power electronics
Electronics for harsh environment – automotive, industrial, space
Electronics for harsh environment – extreme temperature, radiation, vibration, etc.
New MEMS and Sensors both as components and integrated into sub-systems
Phase change memory (both thin film and nanowire-based)
Resistive random-access memory (both thin film and nanowire based)
Ferroelectric memory
NEMS based memory
Molecular memory

This HIR 2023 edition includes examples for seven of the device types listed in table 1 that are already emerging as candidates to replace CMOS in specific applications. This includes a section dealing with printed electronics which can provide a cost-effective solution for low-volume product with reduced non-recurring engineering. It also shows promise for printing living tissue, relying upon cellular self-assembly to obtain bio-printed living tissue[1]:

- Nanoscale Vacuum Electronics
- Neuromorphic Devices
- Printed Electronics
- Spintronic Devices
- Graphene and 2D Material Electronics
- Carbon Nanotube Electronics
- Plasmonic Devices

1. Nanoscale Vacuum Electronics

Vacuum electronics potentially offers unique advantages over semiconductor electronics in speed, power and resistance to damage by extreme environments such as radiation and high temperature [1-4]. Reducing the critical dimensions and fabrication using modern IC manufacturing techniques including extreme UV (EUV) lithography can bring vacuum electronics into the sub-10 nm nanoscale era. These advantages, however, have not yet been realized, and much remains to be understood to enable more efficient operation at the component level and integration into circuits. The device performance (e.g. drive current vs. voltage) is governed by the following fundamental processes: emission from the cathode; transport in a nano-vacuum channel; and collection by the anode.

Preliminary efforts [1] in this area have produced nanoscale vacuum channel transistors with a source-to-drain distance of less than 50 nm and a surround gate, as shown in Figure 1. The device operated under a drain voltage of 2 V and a gate bias below 5 V, providing a drive current of 3 μA. Entirely silicon-processing steps were followed on a 200 mm wafer. Since the mean free path at atmospheric pressure is tens of micrometers, which is a few orders of magnitude larger than the channel dimension, no vacuum was used. The device performance was found to be robust against high temperature and ionizing radiation. A similar effort [2] on 150 mm SiC wafers, but for a vertical structure with the source at the bottom and drain at the top, proved equally successful. The surround gate in both cases provided excellent electrostatic control. These early demonstrations of wafer-scale fabrication appear to be promising for the future of miniaturizing vacuum devices into the nanoscale.

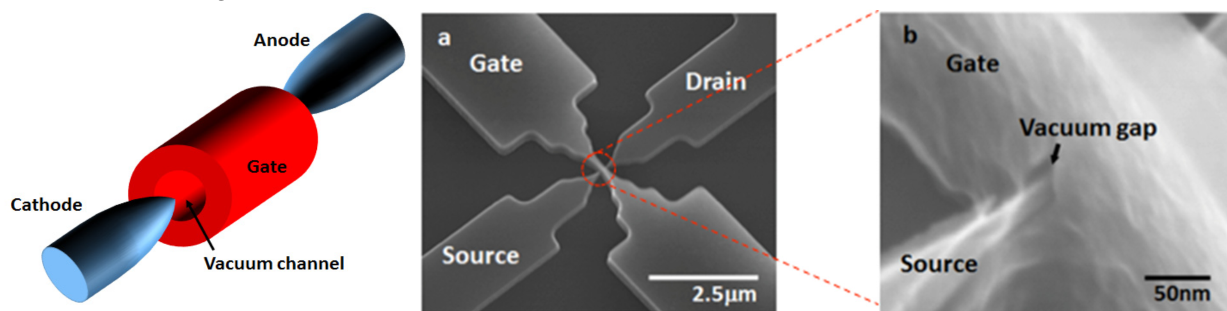


Figure 1: Schematic of a silicon nanoscale vacuum channel transistor with a surround gate, and SEM images of the fabricated device [1]

Vacuum electronics does not have complementary devices working in pairs as in the case of CMOS and thus, the low standby power consumption and high noise immunity common in CMOS technology is missing in vacuum electronics. Recently, a proposal has been made [5] to construct complementary vacuum devices by combining field emission with nanoelectromechanical (NEM) actuation of the gate.

One of the main limitations of nanoscale vacuum channel transistors (NVCTs) is their low transconductance. The small intrinsic gain is attributed to the limited current density. Despite the fact that vacuum is ideal as carrier transport medium, the tunneling resistance at the emitter-vacuum interface largely limits the emission current. The emission current is determined by the energy barrier and electric field on the cathode surface. The surface field needs to be at least ~1 GV/m for significant emission. The channel current has been significantly improved in the past decade by reducing the gap size to ~10 nm range and/or by modifying the work function of the emitter surface. Reducing the

[1] *American Institute of Physics June 4, 2019 issue of Applied Physics Reviews, from AIP Publishing*

gap size naturally results in reducing the operating voltage, and less than 5 V operation has been demonstrated. Despite these improvements, the transconductance still remains one or two orders of magnitude smaller than those of CMOS transistors.

The maximum obtainable current of a vacuum FET is ultimately limited by the Child-Langmuir (C-L) law, which dictates a $1/(gap)^2$-dependence of channel current. Theoretically, the gap size can be reduced to the direct tunneling regime (~2 nm). Extremely scaled channel length can shorten the carrier transit time. Down-scaling the gap size will continue to increase the current, but still the upper bound of the channel current is set by this C-L limit, which is caused by the space charge effect of electrons in transit. Another limitation of vacuum electronics at circuit level is the unavailability of CMOS-like circuits that will minimize static power consumption. Other challenges include finite output impedance, long-term reliability, and lack of fundamental models and simulators, especially for the sub-10 nm regime.

The ageing model needs to be understood for long-term reliability. The long-term degradation behavior of vacuum electronics devices is different from that of sold-state electronic devices due to differences in the underlying mechanism. Degradation mechanisms in vacuum devices include atomic migration near the cathode, whisker growth, gas leak into the vacuum channel, ion bombardment due to residual gas, and phase shift of material. The acceleration stress test methods need to be standardized.

For the cathode, a mechanically and chemically robust yet sharp needle-like tip is preferred. The top-down fashioned fabrication from bulk semiconductor crystal may cause variability in terms of such demand. A hybrid of top-down fabrication as baseline and bottom-up growth of 1D/2D nanotube/nanowire/nanosheet for the cathode could be a solution.

Difficult challenges and potential solutions include:

5-year horizon:

- Manufacturing of nano-gap beyond 10 nm: precise control of gap size is critically important for reproducible device characteristics.
  - Solutions: advanced nanopatterning techniques, bottom up self-assembly or epitaxial regrowth of electrode, leveraging advanced lithography such as double- and quadruple-patterning, hybridizing with 1D or 2D materials.
- Related issue: roughness of the emitter surface: the surface roughness at ~10-nm-scale gap sensitively affects emission current via an atomic scale geometric (field enhancement) and/or gap-size effect.
  - Solutions: improvements in manufacturing process, electrode surface coating technology, exploration of new materials with inherently uniform outer surface at atomic scale (eg. 2D material), exploration of new materials with high (chemical and mechanical) robustness against bombardment.
- Overcoming the Child-Langmuir's space-charge-limit of channel current.
  - Solutions: employ a space-charge neutralization concept; incorporate electron-transparent conducting (or semiconducting) 2D material (such as monolayer graphene) into the void channel toward the anode side; update Child-Langmuir model to take into account the phenomena of charge transport at sub-10 nm scale.
- Low-voltage operation (< ~1 V) with enhanced channel current: important for low-power integrated circuits.
  - Solutions: negative work function material, heterostructured superlattice, employ the 2D electron-gas (2DEG) effect at channel edge (low-voltage emission of electrons and ballistic transport in a nano-void channel).
  - Solutions: employ a plasmonic effect for photo-enhanced emission of electrons at the cathode.

15-year horizon:

- Developing CMOS-like circuits for low standby power.
  - Solutions: incorporate a tunnel FET as a conjugate form of vacuum FET; pair a p-channel V-FET with an n-channel T-FET.
- Developing THz-range vacuum transistors and circuits.
  - Solutions: Explore novel pathways for implementing circuits in the absence of complementary devices.

Reliability, Packaging and Lifetime

Gated Field Emission Arrays (GFEAs) and individual gated emitters have been considered for a range of applications over the last few decades. Numerous companies pursued GFEAs for use in Field Emission Displays (FEDs) in the 1990's to early 2000's. The first FED was demonstrated by LETI-CEA in Grenoble, France. While companies pursuing FEDs had some technical successes, costs and the substantial gains in Liquid Crystal Display

technologies ultimately prevented any meaningful applications of FEDs. Because most of this work was performed in industry, little of that technological development has been published, and so many of the successes in reliability, lifetime performance, and vacuum packaging are little known outside these groups. As gated emitters are now being considered for vacuum transistors, some of the same issues that were studied in FED development are still a concern, and therefore a summary of what we know from FED development is given below.

Reliability

One of the major concerns with GFEAs is reliability. Much of the published work comes from universities and laboratories and not from the companies that developed FEDs. The failure mechanisms in these structures occur from poor dielectrics or defects in the structure, surface breakdown, or emitter tip failure. Proper processing and design can prevent the first problem, and in company fab facilities, these types of structures can be reliably fabricated without significant defects. Such is not usually the case at university facilities; hence these devices often suffer from a number of reliability issues related to fabrication quality and not the inherent problems of the GFEAs. It is important to note that FEDs often had current requirements that were low (<100 mA/cm^2), which does make it easier to develop. Once these process problems are resolved, tip arcing and breakdown becomes the major problem.

A team with LET-CEA and Texas Instruments published some details on the use of ballast resistors to limit current to the emitter tips and prevent arcs [6]. This approach served three purposes: (1) limiting the current to the tips, (2) improving uniformity of emission across the FED, and (3) providing a small built-in lifetime with the voltage drop across the resistor. When the resistors are large enough to provide current limit, there is insufficient charge to allow the excursion to a cathodic vacuum arc [7-9]. The resistance value and placement of the resistor are critical. It is preferable that the resistor be large enough that during typical operation there is a voltage drop across the resistor that is already limiting current, e.g., a 3 V drop for a 70 V gate to emitter drive signal. This level of voltage drop provides current limit and improves uniformity. It was found that for molybdenum field emitter tips operating with an average current of 3-5 nA, a resistor value around 1 GΩ provided the needed ballast. However, the resistor must be close to the tip in an array. If the resistor is far away from an array of tips, the gate to emitter capacitance might provide enough stored charge to allow the arc to still occur even though the resistor might limit the severity [9]. The arc initiates in < 1 ns, so the stored charge available must be small enough to limit the current to << 0.1 A. Hence, a key to emitter tip reliability is resistance or a limitation in available charge to prevent the arc from initiating. In the lifetime study of 5.2" diagonal FEDs fabricated at PixTech, the use of the ballast resistors (sputtered polysilicon thin film layer) allowed for FED operations >10,000 hrs. Rarely was emitter tip arc damage found on de-packaged displays.

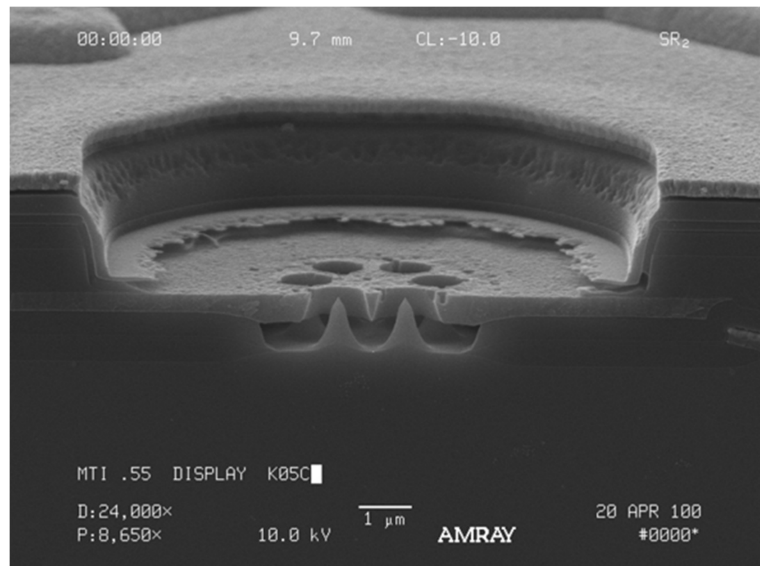
At Micron technology, 0.55" diagonal, active-matrix FEDs using silicon tips were fabricated almost entirely in a 200 mm silicon wafer DRAM memory fab. These displays eventually yielded well across the wafer and were extremely reliable. Because of the quality of the fabrication facilities and the engineering support, processes such as silicon tip etch and CMP were very stable. The active-matrix circuitry included several NMOS transistors to control emission current at the pixel (6-8 tips), a row address line, and a switched-capacitor charge control circuit. The switched-capacitor circuit then limits the current to the emitter tips. SEM images of the tip structures (cross section) and pixel layout (using false colors) are shown in Fig. 2. These devices were packaged to form FEDs and operated full-on (white screen) for ~40,000 hrs. and no arcs were identified, but it should be noted that not all of these packages were de-processed for study.

Therefore, tip reliability can be achieved by providing a mechanism that limits the current to such a level that the arc cannot initiate (<<0.1 A per emitter tip). This approach requires a process flow that can accommodate the ballast resistors or control circuitry. The problem with the resistors, of course, is that they also may limit temporal response for the application.

Packaging

One of the concerns in the renewed interest of GFEAs is vacuum packaging, if the transistors operate at anything other than atmospheric pressure. There is a concern that the packaging technology is unreliable and costly despite progress particularly with MEMS devices. The major vacuum packaging issue for FEDs was the use of glass. The thin glass needed for flat displays made the packing process very challenging. Utilizing the technologies developed in the CRT and vacuum fluorescent display areas, most FED companies pursued the standard approach of using glass frits and getters to seal and pump the displays. The processes developed depended upon the chosen glass type and the seal process. There are basically two sealing approaches: vacuum sealing and exhaust sealing.

In exhaust sealing, the glass package is assembled and fired in either vacuum or gas (usually nitrogen) to prevent oxidation. Such sealing of glass in FEDs was usually performed around 500 deg. C depending upon the glass and the type of glass frit. For example, the PixTech FED GFEA cathodes were fabricated on glass substrates, the same type of glass used by LCD manufacturers (aluminosilicate) during the 1990s. After scribing the glass to form each display cathode, the cathode and anode (phosphor screen) were aligned and fitted together. Usually, the frit is dispensed on the anode. Weights or clamps cause the frit to extrude during heating forming the seal (~500 deg. C). Spacers (glass beads or rods) separate the anode from the cathode to form the vacuum gap. PixTech offset the entire structure to leave a "getter box" that was attached, using frit, to the end of the display. This box had an opening to place the getters. After firing in the oven or belt furnace, two types of getters were placed in the box through the hole. One was a non-evaporable getter (NEG) that is activated by temperature during seal under vacuum. The other was an evaporable getter that is "flashed" using inductive coupling after seal. This type of flashed getter is often visible in the exhaust tubes in CRTs. In the final seal process, a glass cap was placed over the hole in the getter box. This cap also had frit to form a seal. During the final seal, the entire package was heated to ~400 deg. C under vacuum. This heating caused the frit on the cap to soften and extrude, but the frit that was fired during the first package fire process did not soften. One critical step, known by display and microwave vacuum device manufacturers, is that water vapor must be baked out of the package prior to final seal. Typically, a temperature around 350 to 370 deg. C is used [10] with the package or components under vacuum. This is often performed with a soak at temperature from 6-15 hrs. depending upon the process. Once the soak has desorbed the water vapor, the temperature is increased to ~400 deg. C to seal off the package with the glass cap. The seal temperature activated the NEG. The evaporable getter is activated after the package is removed from the seal oven.



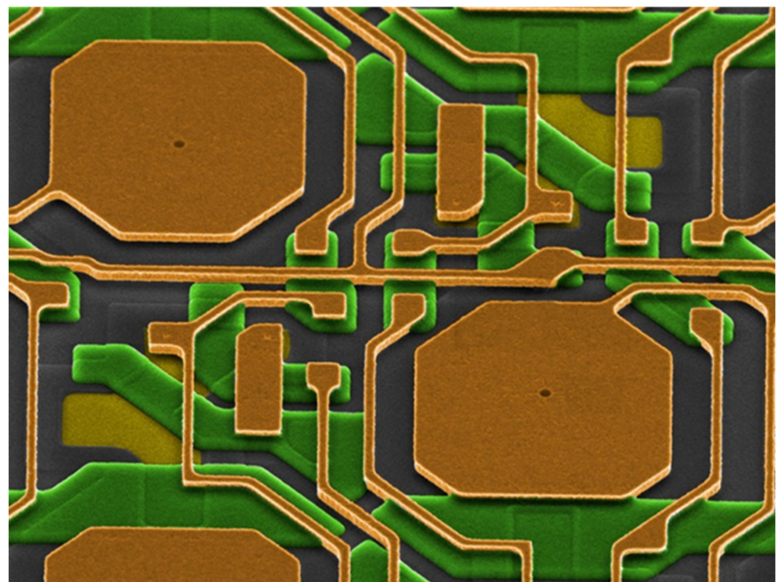


Figure 2. SEM images of the Micron 0.55" FED cathode (top) silicon tips in cross section forming 1 pixel and (bottom) false color image showing the layout of the active matrix circuitry showing several pixels but each has only 1 emitter tip.

Micron Technology used a different approach. The 0.55" FED cathode was fabricated on silicon. Hence, the package had three components: the silicon cathode, the phosphor screen on glass, and a back plate assembled from glass. A pictorial representation and photographs of a monochrome display package (front and back) are shown in Fig. 3. In this process the back plate was assembled and formed using a frit that fired at 700 deg. C. This back plate formed the holding location for the NEG. The cathode was aligned and flip-chip bonded to the phosphor screen, which had "rails" on the edges to form spacers and make electrical connections. In these displays a 60 µm anode to cathode vacuum gap was used with an anode voltage of 1400-1600 V. In this approach the getter adsorbed gases from its location on the backside of the cathode. A glass frit that fired at ~400 deg. C was dispensed on the back plate. Then the anode, with cathode attached, was assembled to the back plate using a spring clip to hold the backplate against the anode. This entire assembly was placed in a vacuum furnace. The furnace was ramped to 360 deg. C

under vacuum to cause water desorption, and then after 8-10 hrs. the temperature was raised to 400 deg. C. At this point the glass frit on the anode extruded, sealing off the package. Note that no exhaust port or opening for exhaust was needed. Gas was pumped from the package through the gaps formed by the variation in the glass frit and its contact to the anode. To ensure this, slight bumps in the frit were often purposely created at the corners of the frit bead. The seal temperature again activated the NEG. The 0.55" FEDs and 5.2" FEDs were tested nearly 20 years after being sealed and were still operational.

This entire "back-end" process for the 0.55" FED was automated. The silicon wafers were coated with photoresist to protect the emitter tips from particles during sawing. After sawing, the cathode die were picked and placed in boats for resist strip. Pick and place tools took the phosphor screens and placed them in boats. The cathodes were flip-chip bonded to the screens. Back plates were then placed on the phosphor screens in the boats. A spring clip was placed on the back plate and connected to the boat to provide the sealing force. Five FED packages were assembled per boat. Five boats were placed into a cassette, and these cassettes (4-5) were placed in the vacuum furnace for seal. After seal, the cassettes were unloaded, and each boat was placed in a test machine to turn on and check the displays. Good displays were connected with flex tape and sent to a burn-in tool on boards for 48 hrs. of testing followed by final display test. Hence, this packaging lends itself to low-cost batch processing.

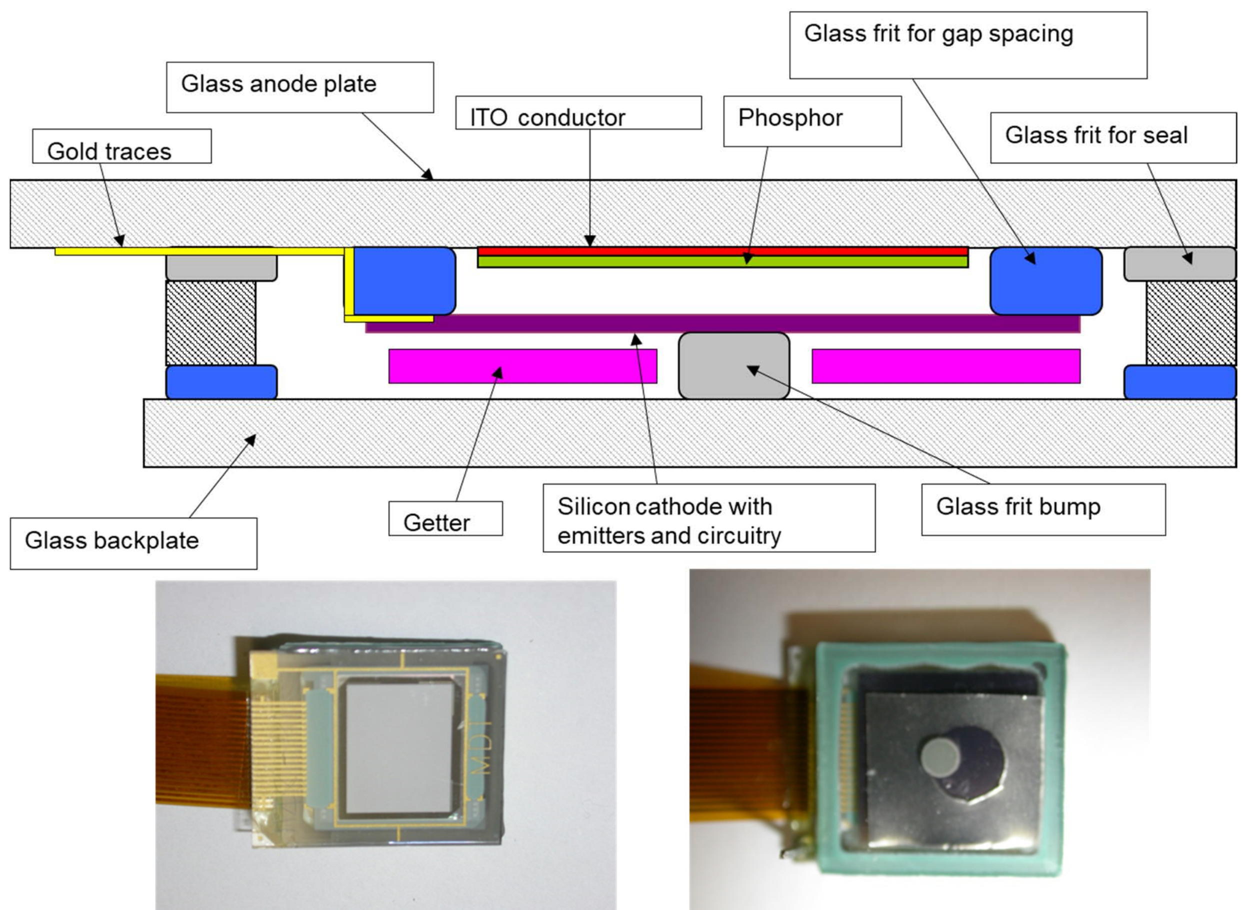


Figure 3. (Top) pictorial representation of the 0.55" FED package showing back plate, cathode, phosphor screen, and getter and (bottom) front and back photographs of the package showing the white phosphor, the flex tape, the frit seal, and the getter (metal with hole punched out).

Lifetime

Once reliable devices are achieved and packaged, lifetime testing begins. While vacuum chamber testing is a good starting point, most development requires testing of dozens to hundreds of devices in vacuum packages. The minimum testing time should be 1000 hrs. to get an understanding of the device performance. In this discussion, we are considering packaged FEDs as mentioned above. Oxidation appears to be the primary cause of emitter tip degradation for both silicon and molybdenum tips [11]. The tips are extremely sensitive to oxidation. For example, heating the silicon tips to 400 deg. C in air for 5 min caused degradation in the 0.55" packaging process discussed above. The performance could be recovered by dipping the silicon tip FED cathode in a buffered HF to remove

silicon dioxide. It was also found that operation in a vacuum test chamber with a phosphor screen caused degradation with the same effect when treated with HF. Moly tips also showed rapid oxidation from the phosphor screens. Vacuum baking could recover the tips. Hence, no permanent damage to the tips (Si or Mo) was ever identified during lifetime testing. The performance was always found to be recoverable. Even metal anodes have an oxide layer, and over hundreds of hours, these anodes caused degradations.

Two methods were implemented to reduce the oxygen content on the phosphor screen. In one case, the phosphor screens were "electron scrubbed" to reduce the surface of the phosphor, depleting the impact surface of oxygen. This approach allowed the 0.55" FEDs with silicon tips to achieve 40,000 hrs. of operation for approximately 20 displays. Electron scrubbing was also used on the 5.2" passive matrix FEDs. This approached improved the degradation (50% reduction in current) from hundreds to many thousands of hours. Dozens of displays were tested to demonstrate this effect. The second approach was applied to 7" color FEDs developed at PixTech. These FEDs worked at 3-5 kV on the phosphor screen. Here, a thin aluminum layer was used to improve efficiency, similar to the process in CRTs, but the oxide on the aluminum surface still degraded the emitters. PixTech engineers developed a very thin (<0.06 µm) aluminum nitride layer. Because this layer was low in oxygen, lifetimes improved from hundreds of hours to ~10,000 hrs.

Hence, the primary degradation mechanism for tips is oxygen liberated from the anode, but the implementation of solutions depends upon applications. FEDs have anodes that vary in bias from 200 V to 10,000 V. It is not clear how the lower anode voltage (< 50 V) in vacuum transistors may affect this oxygen liberation from the anode. It should also be noted that FEDs were not operated past 40,000 hrs. and there was some current degradation. Therefore, other degradation mechanisms are possible.

References

1. J.W. Han, et al., Nanoscale vacuum channel transistors, Nano Lett. 17, pp. 2146-2151 (2017).
2. J.W. Han et al., Nanoscale vacuum channel transistors fabricated on silicon carbide wafers, Nature Electronics, 2, pp. 405-411 (2019).
3. S. Srisonphan, Y.S. Jung and H.K. Kim, Metal-oxide-semiconductor field-effect-transistor with a vacuum channel, Nature Nanotechnology, 7, pp. 504-508 (2012).
4. S. Niirantar, T. Ahmed, G. Ren, P. Gutruf, M. Bhaskaran, S. Walia and S. Sriram, Metal-air transistors: semiconductor-free field-emission air-channel nanoelectronics, Nano Lett. 18, pp. 7478-7484 (2018).
5. R. Bhattacharya, J.W. Han et al., Complementary vacuum field emission transistor, IEEE Trans. Electron Devices, 68, pp. 5244-5249, 2021.
6. J. D. Levine, R. Meyer, R. Baptist, T. E. Felter, and A. A. Talin, Field emission from microtip test arrays using resistor stabilization, J. Vac. Sci. Technol. B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena 13, 474, 1995; https://doi.org/10.1116/1.588336
7. M. Gilmore, N. E. McGruer, J. Browning, and W. J. Bintz, Measurement of Gated Field Emitter Failures, Rev. Scientific Instruments, 64, pp. 581-582, 1993. https://doi.org/10.1063/1.1144181
8. J. Browning, S. Meassick, Z. Xia, C. Chan, and N. McGruer, Ion and electron energies in gated field emitter failures, IEEE Trans. Plasma Sci., 21, pp. 259-260, 1993. DOI: 10.1109/27.219398
9. J. Browning, N. McGruer, S. Meassick, C. Chan, W. Bintz, and M. Gilmore, Gated Field Emitter Failures: Experiment and Theory, IEEE Trans. Plasma Sci., 20, pp. 499-506, 1992. DOI: 10.1109/27.163587
10 R. Bhattacharya, N. Karaulac, W, Chern, A.I Akinwande, and J. Browning, Temperature effects on gated silicon field emission array performance, J. Vac. Sci. Technol. B 39, 023201, 2021; https://doi.org/10.1116/6.0000753
11. B. R. Chalamala, R. M. Wallace, and B. E. Gnade, Effect of O_2 on the electron emission characteristics of active molybdenum field emission cathode arrays, J. Vac. Sci. Technol. B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena 16, 2859, 1998; https://doi.org/10.1116/1.590285.

2. Neuromorphic Devices

As digital CMOS scales to dimensions in the 7-5 nm range, the cost increase per generation has outstripped the shrinkage factor, which defeats the benefits Moore's Law. The cost per unit area is increasing for each new generation. Secondly, the power consumption for massively scaled devices has become significant, which makes their usage in mobile devices unattractive. Beginning at about the 65nm node, passive power has often exceeded active power. Third, packaging technology has not proceeded at pace with chip technology. Fourth, although not directly related to devices, EDA tools – because of the increasing complexity of design rules, the scale of integration and the multiplying reliability checks required have made the cost and time to design and verify chips skyrocket. Fifth, interconnect technology continues to be a challenge, including contact resistance, as the impedance of chip-level interconnects scale in the wrong direction. Sixth, yield at the wafer level will be low, raising the cost of the

system. These problems are multiplied at the wafer scale. And from a power and yield consideration, they are prohibitive.

Background and Overview

The key concept motivating neuromorphic devices is that one can use the brain as a model for a computer. This is a pure biomimetic approach. In the human brain, higher-level information processing occurs in the neo-cortex. Neurons work together to interpret sensory data (input) to produce cognition, spatial reasoning, planning and generation of motor commands and language (output). Different parts of the neo-cortex specialize in one or more specific processing tasks but do so in concert with the other parts. The wiring of the brain consists of some 86 billion neurons, each of which is connected in a massively parallel network to tens of thousands of other neurons. If the combination of incoming signals received by a neuron from other neurons raises its input voltage past a certain threshold, voltage-gated ion channels begin to open, which then transmit an electro-chemical signal to multiple other neurons. The signal can be either excitatory or inhibitory. Thus, many millions of neurons are active (firing) at any one time, whereas many billions are quiescent. One can then pattern a heterogeneous integration scheme by building discrete devices, each mimicking the principal function of one part of the neo-cortex while interconnected to the others.

Neuroscientists used to believe that learning occurred solely in the synapses, whereby they interpret signals from axons and retransmit them via dendrites to neurons. The current wisdom is that learning comes about primarily by the creation of new synapses (synaptic plasticity). Up to 40% of synapses transmitting to a neuron changes daily.[1] This is how new things are learned without interfering with prior knowledge.

Dendrites are not just bits of wire. They also have their own apparatus for making spikes. If enough inputs are activated in the same small bit of dendrite, then the sum of those simultaneous inputs will be larger than the sum of each input acting alone. This is non-linear behavior. Each dendrite from a neuron can integrate inputs, then output a spike, which means that each branch of a dendrite acts like a little nonlinear output device, summing up and outputting a local spike if that branch gets enough inputs at roughly the same time. The neuron then collects all these non-linear inputs and responds. It only responds when a lot of the inputs are active together in time and in space (on the same bit of dendrite). It cannot detect anything that is not a local spike, which means that it ignores most of its individual inputs. (It is not understood how remote signals that trigger motor responses connect). So the neurons respond only to a small fraction of the information they receive, with the rest tucked away in their dendrites. That is, it acts as a kind of processor, which means to say that it is its own little computer, but not the traditional kind, since the inputs are not binary and the output is more a result of a weighted polling than an arithmetic summing.

Computers represent data in a brittle fashion. If one bit flips and there is no error correction, an error is produced. The brain, on the other hand, uses sparse representation of data where relatively few neurons are active at one time, even though it takes thousands of signals to stimulate a single neuron. This is what really enables understanding, thinking and planning, since individual signals can overlap, producing generalized patterns. The signals can also combine, which allows multiple ideas to exist simultaneously, permitting the brain to handle uncertainty relatively easily.

In the neo-cortex, sensory data is processed in a hierarchy of its parts. When the brain attempts to recognize a pattern, an input is passed from one part to another and as it does, it gains more complex levels of extraction as each part represents a different sensory-motor experience. At the end of the process, a pattern is recognized. That is, the recognition is not simply by sound or image or touch but by all these integrated together. A computer attempts to emulate this behavior when designed to implement deep learning schemes, but does so much less efficiently and accurately and with many more deep layers than used by the brain.

A neuron has an average electric potential of about 70mV [2], which changes when it receives signals from other neurons. It will then spike when this membrane voltage reaches a certain threshold. The spike or pulse has a width of a few ms and a height of about 40mV. It can be transformed when it reaches a synapse before being transmitted down connecting dendrites to other neurons. The energy cost of creating an electric potential in a synapse is about 10fJ [3]. The gate of a MOS transistor used in state-of-the-art CPUs is about 0.5fJ. At the cell level, a neuron matches neither the speed nor the low energy of a modern transistor switch. On the other hand, the power expended during a synapse is five orders of magnitude less than that of a switching transistor because it is two million times slower. At the chip level, when transistors are used to simulate a brain function with a similar sized portion as that of the brain, the power consumption is 10 billion times larger and is less accurate [4]. The clear point here is that the key to achieving brain-like power consumption will not be achieved by the device per se but rather by the chip architecture and design. This will be aided by changing how the device is operated.

So, what makes the brain so much more efficient than computers? What will it take to make super computers operate at 20w, as the brain does? Here are some of the key differences. The brain transmits spikes that are binary. It retains the nice feature of digital electronics in that it can transmit a signal a long way without loss of integrity and is refreshed at every synapse. On the other hand, it really is a mixed-signal organ in that inputs to neurons are weighted. If it were a transistor, many of the signals would be sub-threshold. Second, there are no clocks. Everything works asynchronously, with actions triggered locally. Some digital designs today also use asynchronous clock trees. The difference is a massive difference in parallelism and in its mode of operation. Only 1-2% of the neurons in the neo-cortex are firing at any one time [1]. Third, operations are highly local. Ninety percent of neurons only connect to a thousand or so neighbors such that the average lengths of axons and dendrites are short compared to the size of the brain. The brain does not require global interconnects. It does require a global blood supply and lymphatic system. There is a need to connect remote areas of the brain when, for example, a motor reflex is needed in response to a visual queue, but this function of the brain is not understood since there does not appear to be an observable global, physical connection. Fourth, a synapse retains and remembers its state even when not active. Fifth, a neuron is able to branch in three dimensions. Sixth, memory elements are not separate from signal processing elements. There are no CPUs or bus lines.

Difficult challenges for the technical issue

The first issue to confront is the chicken-and-egg problem of what comes first. Is it the device, or the architecture and design? Experience teaches us that it is best to develop and optimize the device in parallel and iteratively with the architecture. As explained before, the major gain to be realized in power reduction will come from the architecture, algorithms and design more so than from the device. In fact, the device will almost certainly be more than one device (neuron and synapse) and each device could be a cell made of several components rather than a single component. On the other hand, the component and cell designs will enable the architecture.

A straightforward way to resolve the chicken-and-egg question is to select an application. An application requires a design specification, which, in turn, sets requirements for the components, which, in turn, highlights the attributes that need improvement. It is a key enabler. Neuromorphic chips are not particularly good at number processing, but excel in applications which feature processing of continuous sensory inputs. A natural choice for an application might be an AI/deep learning task. Today, there does not appear to be any neuromorphic design that does the job as well as other alternatives. A first step would be to demonstrate a chip that is superior in power, cost or accuracy to an existing design, followed by its incorporation into a HI design as a second step.

Speech and facial pattern recognition are accomplished in the cloud using high-powered servers. A target application might be to replace this server function with a lower power version, one that perhaps uses far fewer hidden layers. The brain does the job with as few as four, whereas a deep learning app may require a hundred layers. A more aggressive target might be to replace the server function entirely by putting a neuromorphic chip in a mobile device.

Without HI, the application will require standard and complex I/Os as well as ESD cells which are all obtainable, since standard components are envisioned. However, in a wafer-scale HI application where one neuromorphic chip communicates with another, analog signals may be all that are necessary.

Finally, work needs to be done to improve the control of transistors in sub-threshold operation. Small changes in gate voltage result in large changes in drain current, which make chips prone to failure. The problem is a yield challenge as well as a reliability challenge, as devices burn-in with use causing sensitive parameters to shift during the device lifetime.

Requirements for key attributes over time

We need a cell design that imitates a neuron soma and one that imitates a synapse. We need a transistor that operates reliably and predictably in the sub-threshold region. We do not need an ultra-fast transistor, since the processing speed of a neuromorphic circuit will depend upon parallelism, and it is advantageous to have a low-leakage cell, which is more easily attainable if deep submicron technology is not used. Ultra-fast transistors that are under-driven can reduce active power but less so passive power. In this same regard, a sub-threshold slope below the theoretical limit for bulk Si (60mv/decade) would be an advantage. We need a cell or a component that is non-volatile. We need a cell with one input and one output (synapse) and another cell with hundreds to thousands of inputs and outputs (neuron). We need 3D technology to facilitate massive use of local interconnects. We need the technology that will interconnect the HI chips and control the signals. This includes the controller chip as well as the interconnect hardware (interposers, etc).

The Gordian Knot that must be cut is that there are no quantified targets with acceptable tolerances defined for the required components. These depend upon the "final" architecture that will be used to implement a purely biomimetic

system. Since there is no consensus system architecture, there cannot be a well-defined, quantitative list of performance metrics. Without device performance metrics, a fully neural architecture cannot be designed. Key to resolving this issue is picking a flagship application that is suitable for a neuromorphic solution and proceeding iteratively with an architecture/circuit design/PCD project. This will have the benefit of spurring model development and EDA tools.

Potential Solutions

Several groups have been pursuing neuromorphic systems for several years. The SpiNNaker project[5] at the University of Manchester attempts to imitate spikes by using massively parallel ARM9 processors to route them asynchronously. The latest version uses 57,600 processors, each with eighteen cores and each of these simulating as many as a thousand neurons. A prototype chip has been recently designed with a second generation SpiNNaker system [24] that demonstrates efficient execution of brain-inspired algorithms, specifically a reward-based synoptic sampling model that employs structural plasticity to learn a function or a task. This simulation of a synapse is clearly a brain-inspired algorithm. It is sponsored by the EU's Human Brain Project. The Human Brain Project also sponsored the BrainScaleS Project [6] out of Heidelberg University. It aimed at understanding and emulating the function and interaction of multiple spatial and temporal scales in brain information processing. Its approach was to combine neurons and synapses in a mixed-signal application using Si transistors. The dual systems created four million neurons and one billion synapses. The TrueNorth [7] chip developed by IBM in Almaden/San Jose CA does not utilize microprocessors but rather intermixes memory with computation. It features custom neural circuits which instantiate one million neuron circuits and 256 million one-bit synapses to demonstrate its own neural model. The claim made by this group is that a synaptic transmission can be executed using only 26pJ. This is a rough indicator of how far away from an ideal implementation the current state of the art is compared to the brain. We are still more than three orders of magnitude away from what may be achievable.

While none of these projects may ultimately become the ideal neuromorphic chip, they perform a very important function by demonstrating proofs of concept. The SpiNNaker Project computer demonstrates configurability that allows different neural models and architectures to be tested. The BrainScaleS Project demonstrated multi-scale functionality and a large density of neurons. The objective was to study the biology of the brain. Unlike SpiNNaker, TrueNorth was designed for commercial application and demonstrated a neuron density of one million while operating at an impressive 70mW. It demonstrated brain-emulating asynchronous event-driven computing rather than using a processor clock. However, unlike the brain, it is an all-digital design, which allowed IBM to exploit the highly automated digital design infrastructure available today. The architecture is non- von Neumann, but the chip is not purely biomimetic.

Continuing on the architecture theme, researchers at MIT [8] have proposed a new architecture for a fully optical neural network that, in principle, could offer an enhancement in computational speed and power efficiency over state-of-the-art electronics for conventional inference tasks. They experimentally demonstrated the essential part of the concept using a programmable nanophotonic processor featuring a cascaded array of 56 programmable Mach-Zehnder interferometers in a silicon photonic integrated circuit, and demonstrated its utility for vowel recognition. As of today, their network still relies on conventional electronics to simulate neuronal firing, but if that step can be implemented optically, then only the training stage will require electronics. Once trained, the network would be able to perform equivalent matrix multiplications two orders of magnitude faster than its fully electronic counterparts. More on optical devices can be found under Section 7 of this chapter.

The most recent attempt at a truly biomimetic device has been provided by the Intel Loihi test chip [9,10,11]. This test chip has self-learning capabilities, which makes it potentially beneficial for autonomous operation and continuous learning in an unstructured environment. Further, it is 1000 times more energy efficient than other ICs which feature trained algorithms, and incorporates many of the features enumerated above. Specifically, it incorporates a fully asynchronous neuromorphic many-core (128) mesh that supports a wide range of sparse, hierarchical and recurrent neural network topologies with each neuron capable of communicating with thousands of other neurons. Each neuromorphic core includes a learning engine that can be programmed to adapt network parameters during operation, supporting supervised, unsupervised, reinforcement and other learning paradigms. It was manufactured using Intel's 14 nm process technology with a total of 130,000 neurons and 130 million synapses. This is six or seven order of magnitude less than what the human brain achieves. As in TrueNorth, the architecture is non-von Neumann and asynchronously timed but the implementation is digital and hence not purely biomimetic.

Recently Intel announced its Pohoiki Springs [25,26] data center, which connects 768 Loihi research chips in a mesh and is making it available to the Intel Neuromorphic Research Community (INRC). Intel claims that it has the

computing power of 100 million spiking neurons. Loihi has inter-chip communication interfaces which allows for a wide range of mesh-based system implementations ranging from two (Kopoho Bay) to 768 (Pohoiki Springs).

To summarize, the ideal solution, if it is to imitate the brain, should demonstrate accuracy at low power with a compact form factor, sparse data representation, local actuation with no central processing, and real time configurability, all leading to a compact, low power learning machine. The most likely path to the ideal is a neural network with massive parallelism, operated in mixed-signal mode at sub-threshold voltages with a non-volatile synapse cell [12] and both cells (neuron and synapse) configurable in real time. Real-time configurability could be achieved via an FPGA approach operating on NVM cells or on simple multiplexers. Massive parallelism could be achieved by using TFTs or van der Waals devices [13], particularly TMCs. The former has the advantage of a volume manufacturing history. The latter has the advantage of theoretically very low sub-threshold slopes and ultra-low leakage, but the disadvantage of never really having been demonstrated in volume production or in the lab with the theoretical values achieved. That is, they are still laboratory devices. Both have the potential advantage of being built in 3D. More discussion on the use of 2D materials can be found in Section 5 of this chapter.

Another promising approach to achieving the desired device characteristics is to use thin tunneling FETs (TFETs) in conjunction with 2D devices [14]. The TFETs do not rely on energy barriers to suppress leakage current, but rather exploit quantum-mechanical tunneling to modulate drain current. Although the band gap of a single-layer 2D crystal semiconductor is well defined, by choosing a suitable heterojunction of dual-doped layers, it is possible to maximize the interlayer tunneling current but maintain a very low sub-threshold slope. The tunneling current can be boosted by the proper choice of heterojunction band offsets between the two 2D crystal layers. This type of tunneling transistor could be the thinnest possible manifestation of a TFET, which can also be scaled to the smallest dimensions, far below what may be feasible with traditional 3D semiconductors. A TFET with a 2D MoS_2 channel contacted with highly doped germanium source electrodes has been reported with a sub-threshold slope of ~32 mV/decade, a supply voltage of <0.1 V and low current values [15]. The goal of achieving such thin TFETs is also forcing controlled growth, doping and heterostructure design, and represent a promising path to achieving the device targets for neuromorphic designs. Further, 1T-FeFET [16] can be used to implement compact synaptic array structures that can be programmed at much lower read/write operating voltages.

With respect to non-volatility, any of the current bit-readable/writable NVM technologies could be used, but floating and split-gate technologies as well as dielectric storage devices all require comparatively high write voltages. MRAMs and PCRAMs require high write currents. This gives the advantage to memristors [17], SOT [18] MRAMs, STT MRAMs, ReRAMs, FERAMs or CBRAMs. In particular, magneto-metallic spintronic technologies (demonstrating low-threshold switching currents) can be used to implement synaptic crossbar arrays interfaced with neural devices that are potentially an order of magnitude more energy-efficient in comparison to other resistive neuro-mimetic devices [19]. An extended discussion on memristors can be found in Section 5D of this chapter.

Another novel concept exploits one of the characteristics of high-k gate dielectric transistors to create a non-volatile analog-like memory element [20]. Specifically, it is the charge-trapping phenomenon that is leveraged. Experimental data from 22-nm silicon-on-insulator devices reveal that a charge-trap transistor possesses promising characteristics for implementing synapses in neural networks, such as very fine tunability, weight-dependent plasticity, and low power consumption.

Finally, ideas have emerged for building neuromorphic components at the molecular level. With the recent synthesis of cyclocarbon [27], discussions are underway to use this carbon allotrope to design and build elements, which might operate at extremely low power via single electron transfers and perhaps at sub-nanoscale dimensions. To date, no device has been built nor a manufacturing process has been envisioned. Though the potential for achieving brain-like power levels is significant, such technology is still decades away from practical implementation. SIt is nonetheless important to mention because the concept is a research option for developing a truly biomimetic device at extraordinarily low power levels, daunting challenges notwithstanding.

One important issue must be addressed before novel devices can meaningfully impact neuromorphic computing. Although a host of emerging technologies has been envisioned recently, building compact synaptic crossbar-type dense matrices, these devices are generally fraught with several intrinsic inconsistencies including both spatial and temporal variations across multiple devices. How to perform reliable learning with unreliable devices is an important area of research. In references [21-22], new online learning algorithms were developed to demonstrate hardware-friendly learning that overcomes fabrication limitations. It was further shown in ref. [23] that robust learning can be achieved even under various device non-idealities, such as programming non-linearity, spatial and temporal variations, limited resolution in programming, and so on. Such algorithm- and architecture-level innovations are paramount for various emerging nanodevices to make commercial market inroads.

Technology Gaps and Research needs

The missing pieces are the cells described earlier, an architecture/design that accommodates massive parallelism (>1000 inputs/ neuron), algorithms which control the processing and the models that simulate them, and a method for testing the functionality. In an ideal implementation, the processing is hardwired in the design and defined totally by the inputs. There is no need for a system clock, or PLLs, or a CPU. Harvard and von Neumann architectures and Boolean logic are obsolete. There may be a need for defining a local time window in a cell when signals can be acquired, but that is not known yet. In a wafer-scale HI implementation, some means would have to be provided for controlling the inputs and outputs between chips. Therefore, there would need to be at least one chip with timing and watchdog capability. Once learning algorithms are fully understood, this system timing controller chip could potentially be implemented at the wafer level with a neuromorphic design.

Neuroscientists do not understand in detail how the brain accomplishes the processing it does that results in pattern recognition, cognition, learning and planning. Specifically, signal processing paths have not been worked out, nor the method of steerage. It is not even known whether the steerage is algorithmic. As mentioned previously, how global communication is achieved with no apparent physical link is not understood. This then is the highest priority research need. In the meantime, trial algorithms will need to be tested, which then will require one or many CPUs, and these will require system clocks.

With respect to the devices themselves, research is needed in producing 2D devices in volume, reduction in contact resistance and the ability to incorporate impurity doping into the films. All of the same also applies to heterojunction thin TFETs. This burden can be eased if reliable learning can be achieved with unreliable devices, which then highlights this need as a research objective.

With respect to NVM bits, many of the proposed bits require an electroforming step in order to realize the desired hysteresis characteristics that make them appealing. The results of the electro-formation are highly variable. The need then is to develop a bit that does not require an electroforming step since it would seem that the task of electroforming the billions of bits required is not practical, and even if achieved is likely not to be reproducible. There is some indication by researchers at the University of Texas that a novel 2D memristor structure made of unilayer TMDs might be able to be made into a bit that does not require electroforming and that may be able to switch with an energy approaching the thermodynamic limit (as described in Section 5 of this chapter) and subsequently updated with the use of h-BN [28].

With respect to circuits, spiking neural networks seem to have become the method of choice precisely because they most closely mimic the behavior of neurons. However, most of the large ICs utilizing neuromorphic devices resort to all-digital designs in order to exploit the very mature digital design eco-system. Power reductions have been achieved by migrating away from von Neumann and Harvard architectures while retaining conventional hardware components. But digital circuitry is inherently limited when aspiring to brain-like power levels. As pointed out by Landauer [29], physical processes that are not reversible increase the entropy of a system and when applied to logic gates, operations that are not logically reversible must also increase entropy. From that observation Landauer proffered the principle that the erasure of n bits of known information must always incur a cost of $nkT \ln(2)$ in thermodynamic entropy. Then, by definition, the thermal energy not available for useful work increases and thus must be dissipated in the system or its surroundings. An inverter is reversible, that is, it can be run backwards in time to retrieve the original condition. An XOR or NAND gate, for example, is not. Thus, substantial reductions in power are in principle achievable if reversible gates were to be developed and used. While there is no physical reason why this could not be done, in the six decades since Landauer's observation no one has succeeded in building a computer using reversible logic, unless one counts quantum computers whose operations are inherently reversible. This is a fundamental limitation and it would seem then that all-digital designs may never be able to achieve brain-like power levels if non-reversible gates are employed.

Summary

The challenges facing HI are formidable. Neuromorphic devices face all of these same challenges but offer a solution to one of the more significant ones, i.e. power consumption. Without a large reduction in power usage, it is difficult to conceive of a practical HI solution that effectively handles many hundreds of amperes of current drawn by a single wafer. On the other hand, neuromorphic devices have their own challenges. To date there have been no demonstrations of a neuromorphic implementation that have come close to the efficiency of the human brain. The state-of-the-art is still three orders of magnitude away. Furthermore, no neuromorphic chip application has found a market niche.

The pivotal issue is that there is an incomplete understanding of the algorithms used by the brain, if indeed its operation is algorithmic at all. Thus, there is not a clear target for chip architecture/design. All the current approaches

use algorithms that require one or more CPUs. This is sufficient for testing the efficacy of various algorithms and neuron and synapse concepts, but is not likely to result in brain-like efficiency. The brain does not use a CPU, at least not the traditional "central" kind.

The ideal implementation will appear when the firmware is wired in the hardware with no system clock and no CPUs. At present, there is no such design; hence progress on the near-term horizon will need to proceed using traditional mixed-signal design methods and components aimed at optimizing algorithms and refining component requirements. The device attributes discussed here anticipate the needs by imitating the various components of the neo-cortex in accordance with our present understanding. Ultimately these attributes will not be "final" until the chip architecture is settled as the optimization is inextricably joined to the device.

References

[1] What Intelligent Machines Need to Learn from the Neocortex, Jeff Hawkins, IEEE Spectrum June 2016, p35ff.

[2] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates; 2001. Electrical Potentials across Nerve Cell Membranes.

[3] Organic core-sheath nanowire artificial synapses with femtojoule energy consumption, Wentao Xu, Sung-Yong Min, Hyunsang Hwang and Tae-Woo Lee, Science Advances 17 Jun 2016: Vol. 2, no. 6, e1501326

[4] Supercomputers Ready for Use as Discovery Machines for Neuroscience, Moritz Helias, Susanne Kunkel, Gen Masumoto, Jun Igarashi, Jochen Martin Eppler, Shin Ishii, Tomoki Fukai, Abigail Morrison, and Markus Diesmann, Front Neuroinform. 2012; 6: 26.

[5] The SpiNNaker Project, Steve B. Furber ; Francesco Galluppi ; Steve Temple ; Luis A. Plana, Proceedings of the IEEE (Volume: 102, Issue: 5, May 2014)

[6] Neuromorphic hardware in the loop: Training a deep spiking network on the BrainScaleS wafer-scale system, Sebastian Schmitt ; Johann Klähn ; Guillaume Bellec ; Andreas Grübl ; Maurice Güttler ; Andreas Hartel ; Stephan Hartmann, Dan Husmann, Kai Husmann, Sebastian Jeltsch, Vitali Karasenko, Mitja Kleider, Christoph Koke, Alexander Kononov, Christian Mauch, Eric Müller, Paul Müller, Johannes Partzsch, Mihai A. Petrovici, Stefan Schiefer, Stefan Scholze, Vasilis Thanasoulis, Bernhard Vogginger, Robert Legenstein, Wolfgang Maass, Christian Mayr, René Schüffny, Johannes Schemmel, Karlheinz Meier , 2017 International Joint Conference on Neural Networks (IJCNN), 2017.

[7] TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip, Filipp Akopyan ; Jun Sawada ; Andrew Cassidy ; Rodrigo Alvarez-Icaza ; John Arthur ; Paul Merolla ; Nabil Imam, Yutaka Nakamura, Pallab Datta, Gi-Joon Nam, Brian Taba, Michael Beakes, Bernard Brezzo, Jente B. Kuang, Rajit Manohar, William P. Risk, Bryan Jackson, Dharmendra S. Modha, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (Volume: 34, Issue: 10, Oct. 2015)

[8] Deep learning with coherent nanophotonic circuits, Y.Shen, N.Harris, S.Skirlo, M.Prabhu, T.Baehr-Jones, M.Hochberg, X.Sun, S.Zhao, H.Larochelle, D.Englund & M.Soljacic. Nature Photonics, Vol.11, p.441(2017).

[9] Intel Announces Self-learning AI Chip Loihi, Corinne Reichert, http://www.zdnet.com/article/intel-announces-self-learning-ai-chip-loihi/ , Sept 2017.

[10] https://en.wikichip.org/wiki/intel/loihi

[11] Loihi: A Neuromorphic Manycore Processor with On-Chip Learning, Mike Davies et al., IEEE Micro (Volume: 38, Issue: 1, January/February 2018)

[12] A Learning-Enabled Neuron Array IC Based Upon Transistor Channel Models of Biological Phenomena, Stephen Brink ; Stephen Nease ; Paul Hasler ; Shubha Ramakrishnan ; Richard Wunderlich ; Arindam Basu ; Brian Degnan, IEEE Transactions on Biomedical Circuits and Systems (Volume: 7, Issue: 1, Feb. 2013)

[13] A microprocessor based on a two-dimensional semiconductor, Stefan Wachter, Dmitry K. Polyushkin, Ole Bethge, and Thomas Mueller, Nat Commun. 2017; 8:14948.

[14] Two-dimensional semiconductors for transistors, Manish Chhowalla, Debdeep Jena and Hua Zhang, 10.1038/natrevmats2016.52

[15] A sub-thermionic tunnel field effect transistor with an atomically thin channel, Sarkar, D. et al. Nature 526, 91–95 (2015).

[16] 1T Non-Volatile Memory Design Using Sub-10nm Ferroelectric FETs, A. Sharma and K. Roy IEEE Electron Device Letters, vol. 39, no. 3, pp. 359-362, March 2018.

[17] Integration of nanoscale memristor synapses in neuromorphic computing architectures, Giacomo Indiveri, Bernabé Linares-Barranco, Robert Legenstein, George Deligeorgis and Themistoklis Prodromakis, Nanotechnology, Volume 24, Number 38, Sept 2013.

[18] Surprises from the Spin Hall Effect, Jairo Sinova and Tomas Jungwirth, Physics Today, July 2017, p.39ff.

[19] Encoding neural and synaptic functionalities in electron spin: A pathway to efficient neuromorphic computing, Abhronil Sengupta, and Kaushik Roy, Applied Physics Reviews 4, no. 4 (2017): 041105.

[20] Unsupervised Learning Using Charge-Trap Transistors, Xuefeng Gu, S. S. Iyer, IEEE Electron Device Letters, vol. 38, no. 9, pp. 1204-1207, Sept. 2017.

[21] N. Zheng and P. Mazumder, "Online Supervised Learning for Hardware-Based Multilayer Spiking Neural Networks Through the Modulation of Weight-Dependent Spike-Timing-Dependent Plasticity," in IEEE Transactions on Neural Networks and Learning Systems, vol. PP, no. 99, pp. 1-16.
[22] N. Zheng and P. Mazumder, "Hardware-Friendly Actor-Critic Reinforcement Learning Through Modulation of Spiking-Timing Dependent Plasticity," IEEE Transactions on Computers, Vol. 66, No. 2, Feb. 1 2017.
[23] N. Zheng and P. Mazumder, "Learning in Memristor Crossbar-Based Spiking Neural Networks through Modulation of Weight Dependent Spike-Timing-Dependent Plasticity," in IEEE Transactions on Nanotechnology, to appear.
[24] Yexin Yan, David Kappel, Felix Neumaerker, Johannes Partzsch, Bernhard Vogginger, Sebastian Hoeppner, Steve Furber, Wolfgang Maass, Robert Legenstein, Christian Mayr, "Efficient Reward-Based Structural Plasticity on a SpiNNaker 2 Prototype," in IEEE Transactions on Biomedical Circuits and Systems. 13 (3): 579–591. arXiv:1903.08500[cs.NE] 2019
[25] https://newsroom.intel.com/news/intel-scales-neuromorphic-research-system-100-million-neurons/#gs.8i3xgc
[26] E. Paxon Frady, Garrick Orchard, David Florey, Nabil Imam, Ruokun Liu, JOyesh Mishra, Johathan Tse, Andreas Wild, Friedrich T. Sommer, Mike Davies, "Neuromorphic Nearest-Neighbor Search Using Intel's Pohoiki Springs," arXiv:2004.12691 [cs.NE] 2020.
[27] Katherine Kaiser, Lorel M. Scriven, Fabian Schulz, Przemyslaw Gawel, Leo Gross, Harry L. Anderson, "An sp-hybridized Molecular Carbon Allotrope, Cyclo(18)carbon," in Science 20 Sep 2019: Vol.365, Issue 6459, pp1299-1301 DOI:10.1126/science.asy1914.
[28] Samuel K. Moore, "Atom Thin Switches Could Route 5G and 6G Radio Signals," in IEEE Spectrum. June 1, 2020.
[29] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," in IBM Journal of Research and Development. 5(3):183-191. Doi:10.1147/rd.53.0183 (July 1961).

3. Printed Electronics

This section describes a range of printed devices and challenges related to materials, inks, processes and printing tools. For specific applications related to medical, health and wearable applications and systems, the reader is referred to Chapter 4 of this Roadmap. In addition, 3D printed electronics and additive manufacturing techniques in the context of integrating silicon ICs into systems are covered in Section 10 of Chapter 8.

Using printing techniques for the manufacture of functional electronic devices has the potential to provide a wide range of benefits not easily achieved through conventional processing. Examples include very large area devices, use and integration of an extensive set of functional materials (including polymers, small molecules, nanoparticles, biomaterials, etc.), facile integration with a wide range of substrates (including those that are mechanically flexible/stretchable or contoured), mass customization, ability to rapidly prototype and iterate, as well as low-mass and low-cost devices. Additive manufacturing of electronic systems leads to devices which find applications in many areas such as energy generation and storage, distributed sensing, medical devices, wearable electronics, environmental monitoring systems, information displays, interactive systems, robotics, and transient devices. Moreover, using printing to make electronic systems potentially enables new, distributed, custom manufacturing models that are more inclusive, allowing broader participation in the fabrication of goods with very wide design freedom.

There has been enormous progress in the development of electronic materials for solution-based printing as well as in functional devices and systems made using additive manufacturing techniques. Nevertheless, many multidisciplinary challenges remain for materials, devices and processing in order to achieve the desired potential for printed electronics. For example, as printing relaxes area constraints, some high-performance devices can be readily fabricated such as sensors and electrochemical energy storage devices. However, the typically low-resolution print techniques commonly used now limit our ability to build complex, high performance circuits, due to large feature sizes, low yields, lack of precise control and the electronic properties of solution-processable semiconductors. As such, hybrid approaches are often followed, where printed devices are combined with the pre-fabricated silicon integrated circuits; this adds complexity and reduces design freedom, but enables complex functionality such as high-resolution analog-to-digital conversion and wireless communication. Direct-write print systems that enable resolution of 1 micron or better have been demonstrated, but are typically very slow and not easily parallelized, and may limit the materials set that can be used.

Another example of a significant issue is true process integration for the printing of electronic systems. Complete devices and systems often require some non-print processes such as vacuum deposition or photolithography, as well as removal from the printer for certain steps, such as heating or ancillary chemical treatments. As such, for complex devices, we are generally not able to move from a design directly to a completed system without significant human intervention, thereby limiting the ability of this manufacturing technique and not realizing some potential benefits that printing, particularly digital printing, might provide including layer-by-layer printing for integration with

structural additive manufacturing, and closed-loop feedback for correction or optimization of functional devices. Developing appropriate materials and inks as well as print process and device structures will be key to overcoming these issues. Efforts to produce "all-printed" devices are at their early stage.

Printed Electronics for Heterogeneous Integration

For a baseline, there were essentially three major areas of printed electronics in commercial practice in 2018. First is the oldest usage of polymer thick film conductors principally used for interconnection and things like membrane touch switches. This is complemented by thin film processes that make use of lithography patterning to create everything from thin-film transistors to interconnects. Both approaches converge at touch display and active surface applications that make use of a variety of techniques focused on patterning conductive materials such as Indium Tin Oxide (ITO) or silver, carbon or copper nanowires. The third major area of flexible hybrid electronics is where elements of each of these techniques are blended together with additive processing and packaging methods to integrate silicon ICs into systems (see section 10 of Chapter 8). These combined techniques open a gateway for higher performance integration and more aggressive form factor benefits. As of this writing, the state of the art is the integration of 30 micron pad pitch with conductors either screen printed or aerosol jet printed at a resolution of 20 microns and an aspect ratio of 1:3 with a resistivity of 2.5-3x bulk copper. In the immediate future, there are several areas where we can expect these additive patterning techniques to bring us performance that, while not besting traditional methods on a purely performance standpoint, bring a unique combination of structure, design, materials and temperature that can enable devices not possible today.

Some specific challenges and potential solutions include the following:

5-year horizon:

- Overcoming the low performance of printed circuits.
  Potential solutions:
  - Custom, thinned, silicon ICs that provide the correct functionality and integrate readily with printing processes.
  - New materials, device types and designs for improved circuit performance.
- Many target applications for printed electronics, such as distributed sensors or wearable electronics, require stand-alone functionality. A reliable printed power source that can harvest energy from the environment and store it on board is needed.
  Potential solutions:
  - Integration of printed devices that harvest energy from light, heat, motion, etc. with printed energy storage devices such as capacitors or electrochemical cells.
- In-printer layer-by-layer simultaneous printing of multiple device types is currently a challenge due to a variety of off-line patterning and curing steps that are often needed.
  Potential solutions:
  - Develop materials/inks that can be rapidly cured within the print tool without the need for off-line processing and additional human interaction.
  - Integrated print tools that can process a wide range of inks and feature sizes from a single design.
- Simple printed chemical/physical/bio sensors are often not sufficiently selective, thereby limiting their utility.
  Potential solutions:
  - Develop arrays of similar sensors for improved selectivity; employ machine learning.
  - Develop sensor materials and/or membranes for improved selectivity.
- Integration of electronics with 3D structural additive manufacturing would be beneficial, to make complex, complete electromechanical systems.
  Potential solutions:
  - Integrated print tools, as well as tools designed to print onto curved surfaces.
  - Printed electronic skins that can be integrated with 3D structures.
  - New print techniques, perhaps with variable voxel size, that are better suited to integration of multiple materials.
  - New materials and print techniques to make printed structures compatible with printed electronics.
- Lack of good printed actuators limits ability to make electromechanical systems.
  Potential solutions:
  - Translate printed electrically-actuated, muscle-like actuators to layer-by-layer print techniques.
  - Develop new, print compatible, electrically actuated actuators.

Some specific near-term developments are summarized below:

Replacement of wire bonding – In the near term, printed electronics will be used in its flexible hybrid application to integrate ICs. It is already being used in relatively low-complexity devices to replace traditional electronics, and a great deal of work is underway advancing the reliability, conductivity and printing resolution of these devices. As this work improves and approaches "as deposited and cured" interconnects approaching 2x the bulk resistivity of metallic copper, more and more packaging applications will make use of additive techniques to replace wire bonding or to enable 2.5D integration. This will be especially common in cases where devices such as mixed signal, RF and MEMS elements will be integrated together (see chapters 11 and 12). The gating elements include the resolution, aspect ratio and resistivity of the resultant materials.

Replacement of Solder – A longer term application for printed electronics is to enable integration of devices and systems using conductive adhesives, and plasma or plasma-spray conductors, to replace solder assembly. These approaches will lose the self-centering benefits of solder, but will gain a number of advantages including lower assembly thermal stress as the lower temperature excursion will not create as significant a CTE mismatch. Also, the low-temperature integration (which in some cases can be done at room temperature) enables the integration of non-traditional materials, which are especially interesting in bio applications where certain proteins, biomarkers, or chemicals will not survive a 230°C reflow process. This facilitates the creation of a whole new landscape of biological or chemical sensing devices.

3D Surface Electronics – One area where Printed Electronics is supplanting traditional methods is the ability to deposit electronics directly onto curved surfaces. Additive techniques already are replacing lithography-based processes for antenna patterning, and we continue to see use cases where electronics are added to helmets, wing panels, or other elements that either do not have the space or cannot accommodate planar rigid circuit integration. This area continues to see growth, and as the integration of components begins to scale with the printing techniques, more and more niche applications will make use of these processes.

Embedded Electronics for 3D Printing – The vision since the late 1980s has been to integrate 3D printing and printed electronics into a system that can deposit embedded electronics into 3-dimensional parts. There are several hobby platforms that can do this integration in a crude fashion today. However, they are constrained by the layered deposition materials used for the 3D printing. Materials like PLA or ABS filaments or SLA resins do not ensure the best electrical properties. However, there is no intrinsic physics or materials science limitation in this domain. It only remains a matter of multi material deposition, integrated software, and process refinement. These will all be solved via market demands in a 5-year time horizon.

Stretchable Electronics – Perhaps one of the most compelling areas of printed electronics is that of stretchable conductors. Given the comfort requirements of wearable applications, there has been a lot of work in this area. The conductors are currently printed to silicon or TPU films, which are in some cases laminated to textiles. The areas of research and anticipated breakthrough include stretchable conductors that can maintain their resistivity through 30% and high strain. Current materials have a strain-dependent resistivity. One solution to this is to use room temperature liquid metals encapsulated in elastomers as the conductors. The Air Force Research Laboratory (AFRL) has demonstrated a tool that can deposit these together. The current material of choice is an indium gallium liquid metal.

Invisible Electronics – The combination of printing materials with index-matching ability and thin film depositions allow designers to manufacture electronics that are on the surface or embedded into products that cannot be detected easily by the human eye. A great example of this is the use in anti-counterfeit for electronic parts. Another example is the use of tracking tags applied to currency for either reflectometry tracking or in extreme cases for GNSS location in order to track (for example) a black-market transaction.

Self Assembly – Another exciting area of activity is in the area of self-assembling electronics. Researchers at Palo Alto Research Center (PARC) have shown videos of "chiplets" that can be controlled in X, Y and Theta orientations to a resolution of sub-microns in some cases. Such examples have the power to enable reconfigurable building blocks of simple elements that can be assembled and then additively interconnected without the need to mechanically grip any element of the system. Such advances will usher in a whole new era of complex aggregated systems that can handle elements either too small or too fragile to be assembled and integrated with traditional means.

Self Destruction – The corollary to self-assembly is self-destructing or vanishing electronics. Numerous applications currently look at the ability to destroy the electronics after use. Whether it is a security feature for a

defense application or an environmental feature for agricultural sensors or the ability to dissolve an implanted device after it is used in the body, thus eliminating the need for surgical removal, these unique use cases have high-value niches that will drive the development of materials in this space. Current approaches rely heavily on non-aqueous solvents. Future development will see growth in well-controlled water-based dissolution curves.

Printed Batteries – The primary issue with printed batteries relates (like all battery applications) to the barrier film deposition. In particular, the gating element relates to the availability of electrolyte chemistries that have high energy density but good atmospheric stability. Lithium polymer and zinc polymer are two of the most promising approaches, with many groups looking into the requirements for moisture and gas permeability. A common response is that the barrier layers need to be comparable to the WVTR of organic display cases, but no experimental data has been collected in order to create a standard. Multiwall CNT structures have been demonstrated with some very compelling results and the team that developed this within a government lab has arranged for it to be licensed to a commercial partner.

Printed Supercapacitors – Because of their durability and high power density, supercapacitors complement or even replace batteries in applications requiring operational stability and fast charge delivery. Several groups around the world have demonstrated printing of both electrochemical double layer capacitors (EDLC) and pseudocapacitors that feature faradaic reactions (like batteries). Common inks available today include CNTs, graphene, activated carbon, and Ag and Cu for electrodes. Additional development is needed for metal oxide inks for the pseudocapacitors. Cyclic performance in excess of 100,000 cycles has been demonstrated for all-printed EDLCs. The cycle life of pseudocapacitors is necessarily shorter due to the faradaic reactions but they generally provide higher power and energy densities compared to EDLCs. Just as in the case of printed batteries, passivation techniques are needed to protect the device from ambient permeability and ensure longevity.

Printed Active Materials – Despite uninspiring mobility at operating conditions anywhere approaching room temperature, continuous improvements in the costs and consistency of printed active materials has made steady progress, with metal oxide and carbon nanotube structures being available in 2018 for even basic labs to formulate an ink and print active matrix or array sensors. Many materials are now available with both P and N type, allowing for the active deposition of complementary logic. Current CNT approaches can harness the power of graphene's electrical properties yielding mobilities that are reasonably compelling. However, major issues relate to the degradation in air. When exposed to atmosphere, CNT transistors will only survive for a few days. The next step is for work to improve the passivation of the CNT materials. Additive techniques are already in place to print barriers for organic LED structures that can achieve a WVTR of 10^{-6}. There are also clever design approaches that can be used to mitigate issues like the threshold voltage shift that is common in printed transistor devices.

Chem FET – The next frontier for printed electronics is to move out of the focus area of voltage-based field effect transistors and move to chemical or biological triggered FET devices. These developments will allow for the device to interact directly with the stimulus material without being processed through a sensor that is interrogated on a time cycle by a purely electronics logic circuit. These types of systems have the ability to open an entirely new domain of low power, environmentally responsive systems. Bio-FETs are typically classified by the agent that triggers a transistor response. The major categories include En-FET which is an enzyme-modified FET; Immuno-FET which is an immunologically modified FET; DNA-FET which is a DNA-modified FET; and CPFET which is a cell-potential FET. The Chemical FET could be considered a superset of Bio FETs and is essentially a doped material wherein there is a preferential binding for a particular target analyte in a solution separating the source and gate electrodes. A concentration gradient between the gate electrode and the solution exists as a function of the semi-permeable membrane on the FET surface. Improvements of both semi-permeable membrane material as well as new analyte binding chemistries that produce the triggering ion concentrations will expand this field and have the potential to usher in a new era of human-machine interfaces that go well beyond our current reliance on manual data input and visual display.

In gas sensing applications, chemiresistors are more common than CHEMFETs, as they are easier to fabricate, regardless of conventional wafer fabrication or by printing. A large amount of work has been done on printed gas sensors for the detection of a wide variety of indoor and outdoor air pollutants, toxic gases and vapors and biomarkers in human breath. Common sensor materials include carbon nanotubes, graphene and other 2D materials (see Section 5 of this Chapter) and metal oxides. The printed sensors are simple, highly sensitive (easily meeting OSHA guidelines for various gases and vapors) and readily amenable for integration with signal processing and other communication electronics. The biggest unsolved challenge is selective discrimination of the target of interest among the vast

background of gases and vapors, which requires machine learning in most cases, as it is difficult to come up with a material or a scheme that responds only to the target of interest and nothing else.

Printing tools: Screen printing, inkjet printing, aerosol jet printing and others have been widely used for printing metallic, semiconducting and dielectric layers needed in a range of applications. The common shortcoming is that they all need a follow-on sintering step to obtain the required consolidated thin film in further device processing. Sintering techniques include thermal, plasma, microwave and laser based approaches that yield resistivities 5-100 X of the bulk value at the conclusion of the sintering step. Thermal sintering is done at temperatures of up to 300 ^{0}C lasting an hour or two whereas laser sintering is faster. Recently, an alternative has emerged in the form of plasma printing, which is aerosol printing aided by plasma operating at atmospheric pressure. The energetic environment of the plasma results in good adhesion of the printed material onto the substrate with nanoparticles consolidated into a thin film upon printing. This "self-sintering" enables a simpler approach without the need for additional mechanisms for curing. Turning off the ink supply and running just the inert gas plasma upon completion of the printing step could further help to improve the film morphology and conductivity.

Altogether, the next phases of process developments will certainly favor additive techniques that can interface electronics more holistically with the natural world. Printed electronics is well positioned to deliver in many areas of this using the techniques discussed.

15-to-25-year horizon:

- Overcoming the low performance of printed circuits.
  Potential solutions:
  - Parallel digital print processes that are able to provide high resolution (better than 1 micron) over large areas at high speed using a wide range of materials with precise thickness control and low variability.
  - New materials and/or device types and designs that achieve vastly improved performance when compared with current devices.
  - Print techniques that enable precise control of both position and orientation of individual nanoparticles.
- Integration of devices directly onto complex or dynamic substrates, including printing directly onto human tissue.
  Potential solutions:
  - Vision systems plus closed loop feedback plus new materials.
- Digital printing should enable autonomous correction and prototyping of printed systems, which could facilitate automatic correction, optimization and perhaps discovery of devices.
  - Potential solutions:
  - Integrated print techniques plus materials that allow for rapid layer-by-layer processing plus AI.

Further Reading

1. J. Zhu and H. Cheng, Recent development of flexible and stretchable antennas for bio-integrated electronics, Sensors, 18, 4364 2018).
2. G.L. Goh et al, Inkjet-printed patch antenna emitter for wireless communication application, Virtual and Physical Prototyping, 11, pp. 289-294 (2016).
3. S. Jang et al, Inkjet printed carbon nanotubes in short channel field effect transistors: influence of nanotube distortion and gate insulator interface modification, Flex. Print. Electron. 1, 035001 (2016).
4. Y. Chu et al, Printed diodes: materials, processing, fabrication and applications, Adv. Sci., 6, 1801653 (2019).
5. B. Kim, Inkjet-printed indium oxide/carbon nanotube heterojunctions for gate-tunable diodes. Adv. Electron. Mater. 6, 1901068 (2020).
6. Y. Li et al, All inkjet-printed metal-oxide thin-film transistor array with good stability and uniformity using surface-energy patterns, ACS Appl. Mater. Interfaces, 9, pp. 8194-8200 (2017).
7. J. Semple et al, Flexible diodes for radio frequency (RF) electronics: a materials perspective, Semicond. Sci. Technol. 32, 123002 (2017).
8. K.J. Yoon et al, Electrically-generated memristor based on inkjet printed silver nanoparticles, Nanoscale Advances, 1, pp. 2990 - 2998 (2019).
9. B. Huber et al, Fully inkjet printed flexible resistive memory, Appl. Phys. Lett. 110, 143503 (2017).
10. S.K. Eshkalak et al, R review on inkjet printing of CNt composites for smart applications, Appl. Mater. Today, 9, pp. 372-386 (2017).
11. K. Rajan et al, Wearable intrinsically soft, stretchable, flexible devices for memory and computing, Sensors, 18, 367 (2018).

12. M.L. Seol et al, All-printed in-plane supercapacitors by sequential additive manufacturing process, ACS Applied Energy Materials, 3, pp. 4965-4973 (2020).
13. A. Rallanmaa et al, Highly flexible and non-toxic natural polymer gel electrolyte for printed supercapacitors for IoT, Appl. Phys. A, 125, 168 (2019).
14. A.M. Gaikwad et al, Recent progress on printed flexible batteries: mechanical challenges, printing technologies, and future prospects, Energy Technology, 3, pp. 305-328 (2015).
15. S.H. Kim et al, Printable solid-state lithium-ion batteries: a new route toward shape-conformable power sources with aesthetic versatility for flexible electronics, Nano. Lett. 15, 5168-5177 (2015).
16. X. Tian et al, Emerging 3D-printed electrochemical energy storage devices: a critical review, Adv. Energy Mater. 7, 1700127 (2017).
17. H. Ragones et al, Towards smart free form-factor 3D printable batteries, Sustainable Energy & Fuels, 2, 1542-1549 (2018).
18. M.L. Seol et al, All-printed triboelectric nanogenerator, Nano Energy, 44, pp. 82-88 (2018).
19. M.L. Seol et al, All 3D printed energy harvester for autonomous and sustainable resource utilization, Nano Energy, 52, pp. 271- 278 (2018).
20. T. Varghese et al, Flexible thermoelectric devices of ultrahigh power factor by scalble printing and interface engineering, Adv. Funct. Mater. 30, 1905796 (2019).
21. S. Shin et al, High-performance screen-printed thermoelectric films on fabrics, Sci. Rep. 7, 7317 (2017).
22. M. Orrill and S. LeBlanc, Printed thermoelectric materials and devices: fabrication techniques, advantages, and challenges, J. Appl. Polymer Sci. 134, 44256 (2016).
23. J. Kim et al, Advanced Materials for printed wearable electrochemical devices: a review, Adv. Electron. Mater. 3, 1600260 (2017).
24. D.I. Moon et al, On-demand printing of wearable thermotherapy pad, Adv. Healthcare Mater. 9, 1901575 (2020).
25. T. Han et al, 3D printed sensors for biomedical applications: a review, Sensors, 19, 1706 (2019).
26. D.I. Moon et al, A single input mutliple output (SIMO) variation-tolerant nanosensor, ACS Sensors, 3, pp. 1782-1788 (2018).
27. B. Kim et al, Carboxylated single-walled carbon nanotube sensors with varying pH for the detection of ammonia and carbon dioxide using an artificial neural network, ACS Applied Nano Materials, 2, pp. 6445-6451 (2019).
28. B. Hu et al, High performance flexible sensor based in inorganic materials, Sens. Actuators B, 176, pp. 522-533 (2013).
29. D.I. Lee et al, All 3D-printed flexible ZnO UV photodetector on an ultraflat substrate, ACS Sensors, 5, pp. 1028-1032 (2020).
30. R. Ramamurti et al., Atmospheric pressure plasma printing of nanomaterials for *IoT* applications, IEEE Open J. Nanotechnol. 1, pp. 47-56, 2020.
31. D.H. Gutierrez et al., Plasma jet printing of metallic patterns in zero gravity, *Flex. Print. Electron*. 7, 025016 (2022).

4. Spintronic Devices

The need for increasingly powerful computing hardware has spawned many ideas stipulating, primarily, the replacement of traditional transistors with alternate "switches" that dissipate miniscule amounts of energy when they switch and provide additional functionality that are beneficial for information processing. To this end, "spintronic" devices have carved out a niche. They utilize the quantum mechanical spin degree of freedom of an electron (or hole), as opposed to the charge degree of freedom, to store, process, sense and communicate information. They often have two attributes that are beneficial: low energy dissipation during the switching process and non-volatility. The latter allows a switch to remain in its final state indefinitely after power has been turned off, and that can spawn new circuitry and architectures, such as non-von-Neumann processors, Bayesian inference engines [1], ternary content addressable memory with reduced active device count and superior energy-delay product [2], instant-on computers with no boot delay, etc.

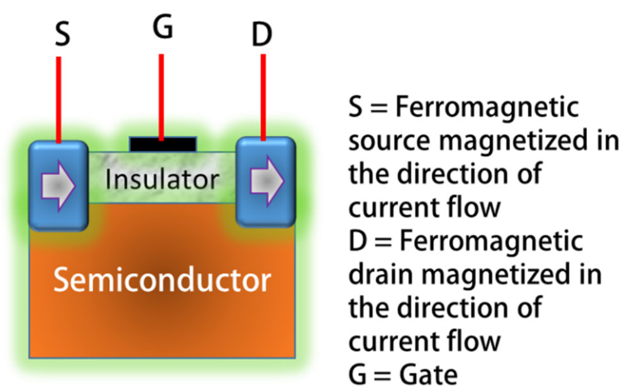


Fig. 1: Basic structure of a SPINFET

Spin Field Effect Transistors (SPINFET)

Spintronic devices have a long history, starting from the early 1990s. The iconic device, also one of the earliest, that spurred interest in device applications of an electron's spin is the famed Spin Field Effect Transistor (SPINFET) that utilizes either the Rashba spin-orbit interaction or the Dresselhaus spin-orbit interaction to elicit transistor activity [3, 4]. Its structure is identical to that of a MOSFET, except the source and the drain contacts are ferromagnetic and

act as “spin polarizer” and “spin analyzer”. The source injects spin-polarized carriers into the channel and the drain transmits only carriers of a particular spin polarization (see Fig. 1). Unlike traditional MOSFETs, which switch on and off by moving charges into or driving charges out of the active device region (the transistor’s channel), the SPINFET operates by modulating the spin polarization of the charge carriers in the channel with a gate potential without having to populate or depopulate the channel with mobile charge carriers at every switching event. However, this modality of switching does not necessarily reduce power dissipation since it takes a considerable amount of gate voltage to modulate the spin polarization of the channel carriers [5]. Furthermore, the ratio of the on-to-off conductance of a SPINFET (which is a very important metric for any switch) is low because spin injection into a SPINFET’s channel from the source, and spin filtering at the drain, are usually inefficient [6]. Any spin relaxation in the channel further reduces the on/off ratio. The highest demonstrated spin injection efficiency at a ferromagnet/semiconductor interface at room temperature is ~70%. With that level of efficiency, the on/off ratio of a SPINFET would be a mere ~3:1, while the on/off ratio of a MOSFET may be 105:1. Thus, SPINFETs may not have much of a role as a digital switch. It may have other applications (e.g. analog applications like frequency multiplier) because its transfer characteristic (drain current versus gate voltage at a fixed drain bias) is oscillatory.

Near Term Outlook and Challenges

Despite the fact that the original SPINFET was proposed nearly three decades ago, there has not been a single convincing experimental demonstration of this device, even at cryogenic temperatures, let alone room temperature. The primary impediments are inefficient spin injection from the source, inefficient spin filtering at the drain, and rapid spin relaxation in the channel. These challenges are not likely to be mitigated in the next five years, and the SPINFET will probably remain a theoretical curiosity rather than become a practical device in the short term.

Table 1: Comparison of MOSFETs and SPINFETs. The MOSFET figures pertain to the 14-nm FINFET used in the Intel® Core™ i7-6700K processor released in 2015. It works at 4 GHz with a power dissipation of 91 W and operates off a 1.2 V supply. The SPINFET figures are theoretically estimated.

	MOSFETs	SPINFETs
Switching energy	100 aJ	1 fJ
Switching speed	100 ps	100 ps
Conductance on/off ratio	10^5	<10
Unusual analog applications	No	Yes
Non-volatile	No	No

Magneto-tunneling junction devices (MTJ)

Another spintronic device that acts as a digital switch is a magneto-tunneling junction device (MTJ). It has three layers – a “hard” ferromagnetic layer whose magnetization remains permanently oriented in a chosen direction, a spacer layer, and a “soft” ferromagnetic layer whose magnetization can be re-oriented with an external agent. The external agent could be a magnetic field, a spin-polarized current exerting a spin transfer torque on the soft layer’s magnetization [7, 8] or inducing domain wall motion [9], voltage-controlled magnetic anisotropy [10, 11] or mechanical strain in the soft layer [12, 13].

The soft layer is usually shaped like an elliptical disk, and that makes the magnetization “bistable”, meaning it can point only along the major axis of the ellipse – either to the right or to the left. One of these orientations is parallel to the magnetization of the hard layer and the other is anti-parallel. When the magnetizations of the hard and soft layers are mutually parallel, the resistance of the device measured between the two layers is lower than what it would be if the magnetizations of the two layers were anti-parallel. These two resistance states – high and low – can encode the two binary bits 0 and 1.

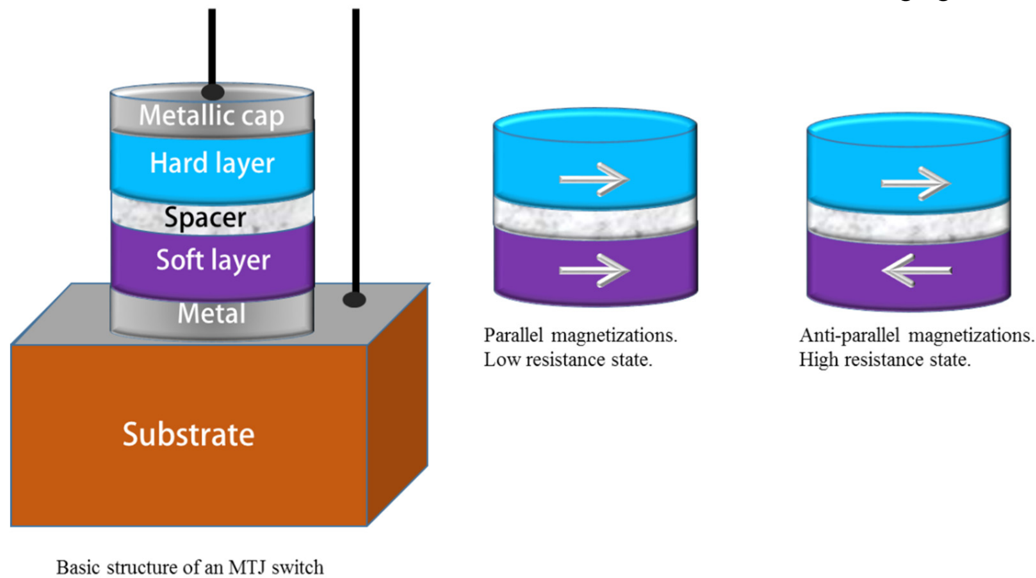


Fig. 2: An MTJ switch and the two resistance states

The MTJ can be the backbone of magnetic Boolean logic gates [14-16], memory cells [17-19], and many other devices and subsystems such as bit comparators [20]. Unfortunately, it suffers from the same drawback as SPINFETs: its resistance off/on ratio is relatively small. At the time of this writing, the highest ratio that has been demonstrated at room temperature is only about 5:1 [21], which makes it unsuitable for application as a switch or in Boolean logic gates. The switching error probability depends on the amount of energy dissipated to switch; if the energy is kept below ~1 fJ, the switching error probability can be very high at room temperature, on the order of 10^{-8}. That too makes it unsuitable for logic, but still suitable for memory, which is much more forgiving of errors than logic.

One way to switch the magnetization of the soft layer, and hence switch the resistance of the MTJ, is to apply a voltage between the hard and soft layers. If the negative terminal of the voltage source is connected to the hard layer and the positive terminal to the soft layer, then spin-polarized electrons will be injected from the hard into the soft layer with their spins aligned along the magnetization of the hard layer. These electrons will exert a spin-transfer torque on the electrons in the soft layer and orient their spins in the direction of the hard layer's magnetization, thereby making the magnetizations of the two layers mutually parallel (low resistance state). If the voltage polarity is reversed, then electrons in the soft layer whose spins are aligned along the magnetization of the hard layer will be able to flow out of the soft layer. This will deplete their supplies in the soft layer, making the soft layer's magnetization anti-parallel to that of the hard layer's (high resistance state). Instead of exerting a spin-transfer torque, a spin-polarized current can also move domain walls in the soft layer and make its magnetization flip. This is called domain wall switching and sometimes consumes less energy than spin-transfer torque switching.

The switching of the soft layer allows one to write a bit into an MTJ. When the magnetizations of the hard and soft layers are made parallel and the resistance is low, one bit, say bit 1, is written and stored in the resistance state of the MTJ. When the two layers are anti-parallel and the resistance is high, the bit 0 is written. Spin transfer torque assisted random access memory (STT-RAM) is becoming the staple of non-volatile memory cells. They can result in high density, excellent endurance and high reliability, but the energy consumption is relatively high.

There are other ways of generating a spin-polarized current to switch the magnetization of the soft layer. One employs the giant spin Hall effect (GSHE) [22, 23]. Here, the MTJ is fabricated on a heavy metal (e.g. β-tantalum) slab as shown in Fig. 3. Sending a charge current Jc through the heavy metal results in a spin current Js flowing into the soft layer and changing its magnetization. This modality results in a reduction of the switching power dissipation by a factor that depends on the thickness of the heavy metal slab. By decreasing the thickness, we can reduce the power dissipation.

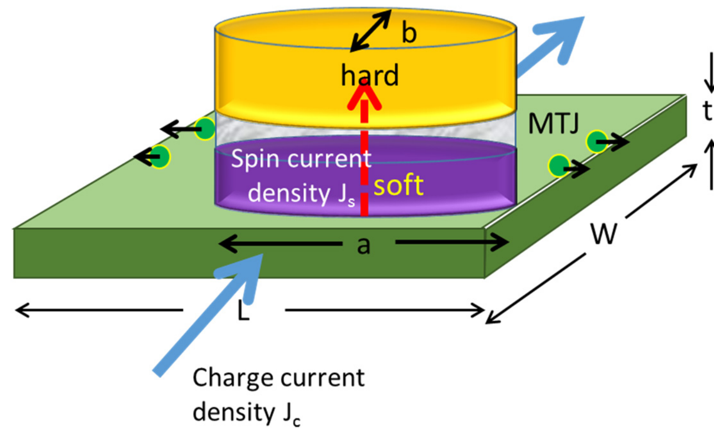


Figure 3: Switching the resistance of an MTJ with the giant spin Hall effect (GSHE)

Another way to switch the MTJ is to apply a voltage between the hard and soft layer that will inject electrons into the spacer layer and modify the magnetic anisotropy of the soft layer. The hard and soft layers can be fashioned out of perpendicular magnetic anisotropy materials such that the stable magnetization orientations of both layers are out of plane, instead of in-plane. By changing the magnetic anisotropy of the soft layer from out-of-plane to in-plane with the applied voltage, the magnetizations of the two layers can be switched between parallel and perpendicular. If an in-plane magnetic field is present, then precessional switching may switch the magnetization from parallel to anti-parallel, instead of perpendicular [24]. Thus, the applied voltage toggles the resistance state of the MTJ. This results in a "toggle" memory. Whenever a bit is to be written, the previously stored bit is first read. If it is already the desired bit, no action is taken. Otherwise, the bit is flipped (toggled) with voltage controlled magnetic anisotropy (VCMA) to write the desired bit.

Writing with spin-polarized current results in a "non-toggle" memory. Whenever we wish to write one bit, we apply voltage of a certain polarity across the MTJ and whenever we wish to write the other bit, we reverse the voltage polarity. No knowledge of the previously stored bit is required.

A third way of switching the resistance of an MTJ is by applying mechanical strain on the soft layer. If the soft layer is magnetostrictive, then strain will re-orient the magnetization via the Villari effect. By applying uniaxial strain along two different directions, the magnetization of the soft layer can be flipped [25, 26]. The strain can be generated by fabricating the soft layer on a piezoelectric thin film and then activating two sets of electrodes delineated on the film, as shown in Fig. 4. The voltages at the electrodes generate biaxial strain in the piezoelectric (compressive along the line joining the electrodes and tensile in the perpendicular direction, or vice versa, depending on the voltage polarity). This strain is partially or completely transferred to the soft layer and re-orients its magnetization. This modality of switching the magnetization of a magnetic layer is known as "straintronics" and has attracted attention due to the possibility of extremely low energy switching [27-29]. A detailed description of this is given in the next section.

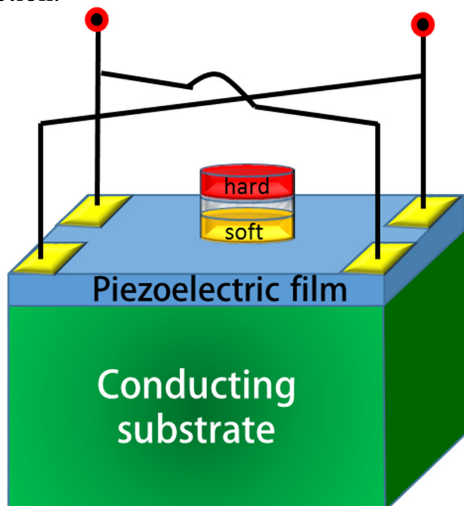


Figure 4: Straintronic switching of an MTJ

Magnetic field generated by a current	40 fJ
Spin transfer torque (with GSHE)	40-400 aJ
Domain wall motion	40 aJ
Electrically generated mechanical strain in a multiferroic ("Straintronics")	4 aJ

Table 2: Energy dissipated in switching a nanomagnet (or the soft layer of an MTJ). These figures are taken from various reported experiments [30-33].

Although straintronic switching is alluring because of the possibility of very low energy consumption, it is beset with many challenges. An elliptical nanomagnetic element is non-volatile because there is an internal energy barrier in the nanomagnet that prevents the magnetization from randomly and spontaneously flipping between the two stable orientations due to thermal perturbations. This energy barrier is due to shape anisotropy of the nanomagnet (the elliptical cross-section). In order to have adequate thermal stability, the energy barrier needs to be at least 1 eV (= 40 kT at room temperature; k = Boltzmann constant and T = absolute temperature), so the probability of random switching is $\exp[-E_b/kT] = e^{-40}$, where E_b is the energy barrier. In fact, nanomagnets in MTJ memory cells (both hard and soft layers) are actually designed with energy barriers of 60 kT or more.

For stress to be able to switch the magnetization of a magnetostrictive nanomagnet, the stress anisotropy energy has to exceed the barrier energy, i.e., $(3/2)\lambda\omega\sigma > E_b$, where λ is the magnetostriction coefficient of the magnetostrictive nanomagnet, ω is its volume and σ is the stress generated in it. We cannot make the nanomagnet volume too large without adversely affecting the scaling. We cannot generate too much stress either, and even if we could, we would prefer not to since generating a larger stress consumes larger energy. Therefore, our only option is to seek out materials that have very large magnetostriction. Elemental magnets like Co and Ni have magnetostriction coefficient λ of the order of 30 ppm while alloys like GaFe or Terfenol-D have much higher magnetostriction (Terfenol-D: 600 ppm). It is very difficult to switch the magnetization of Co or Ni nanomagnets with stress because of their low magnetostriction. In fact, stress anisotropy can be viewed as an effective magnetic field for switching and this effective field in Co or Ni nanomagnets will be only ~30 Oe, which may not be enough to flip the magnetization by overcoming an energy barrier of 1 eV or more [29]. Terfenol-D would seem to be a better choice, but there are two spoilers. First, Terfenol-D will have multiple phases and not all are highly magnetostrictive; getting the right phase is a materials challenge. Second, it has been claimed recently that when Terfenol-D nanomagnets fabricated on a piezoelectric layer are stressed and their magnetizations begin to rotate, the direct magnetostriction effect produces a back action on the piezoelectric which reduces the stress generated [30]. This effect is negligible for low magnetostriction materials like Co and Ni, but prominent for high magnetostriction materials like FeGa and Terfenol-D. All of this is discouraging since it tells us that straintronics, no matter how alluring, is a difficult feat. One has to overcome material challenges and innovate strategies to ameliorate the back action in highly magntostrictive materials in order to be able to switch reliably with straintronics. Most experiments on straintronic switching (of Co nanomagnets) have shown poor switching statistics (small fraction of an ensemble of nanomagnets switching successfully under stress) [29]. The situation does not improve much with FeGa [31] either because of the back action or because of the poor material quality that spawns pinning sites for the magnetization and hinders it from rotating. These materials challenges are daunting and may take many years to handle.

Near Term Outlook and Challenges

MTJs are now the mainstay of many memory technologies, but their prospects for Boolean logic look bleak because of the low resistance off/on ratio. Switching of MTJs is also error-prone, which makes them inappropriate for logic. The main challenge facing MTJs are:

- magnetoresistance ratio, TMR. This would require choice of appropriate hard layer and spacer materials since the TMR is determined mainly by spin-dependent quantum tunneling between the two layers through the spacer. Currently, the materials of choice are CoFeB for the two ferromagnetic layers (the hard layer is usually thicker than the soft layer) and MgO for the spacer. Higher TMR would require ferromagnetic materials with a higher degree of spin polarization than CoFeB (e.g. LSMO) and an appropriate spacer layer material to support spin-dependent tunneling.
- reducing the dynamic switching errors while keeping the switching energy dissipation low. This is particularly critical for straintronic switching where the switching error rate is high because of materials-related issues.

On a more optimistic note, MTJs may have much better prospects for non-Boolean logic. Skewed MTJs (whose hard and soft layers have non-collinear magnetic easy axes) can produce unusual device characteristics that are useful for ternary content addressable memory [2] and deep and wide learning networks. They may also have bright prospects in probabilistic computing and belief networks [1]. A number of MTJ-based designs for neurons and synapses have also been proposed and show promise for better performance compared to CMOS-based renditions [35-37]

References

1. Khasanvis S., Li M. Y., Rahman M., Salehi-Fashami M., Biswas A. K., Atulasimha J., Bandyopadhyay S. and Moritz C. A. 2015 Self-similar magneto-electric nanocircuit technology for probabilistic inference engines IEEE Trans. Nanotechnol. 14 980.
2. Dey Manasi S., Al-Rashid M. M., Atulasimha J., Bandyopadhyay S. and Trivedi A. R. 2017 Straintronic magneto-tunneling junction based ternary content-addressable memory (Parts I and II), IEEE Trans. Elec. Dev. 64 2835.
3. Datta S. and Das B. 1990 Electronic analog of the electro-optic modulator, Appl. Phys. Lett. 56 665.
4. Bandyopadhyay S. and Cahay M. 2004 Alternate spintronic analog of the electro-optic modulator, Appl. Phys. Lett. 85 1814.
5. Bandyopadhyay S. and Cahay M. 2004 Reexamination of some spintronic field-effect device concepts Appl. Phys. Lett. 85 1433.
6. Trivedi A. R. and Bandyopadhyay S. 2007 Switching voltage, dynamic power dissipation and on-to-off conductance ratio of a spin field effect transistor IET Circuits Devices and Systems 1 395.
7. Slonczewski J. C. 1996 Current-driven excitation of magnetic multilayers. J. Magn. Magn. Mater. 159, L1.
8. Berger L. 1996 Emission of spin waves by a magnetic multilayer traversed by a current Phys. Rev. B 54 9353.
9. Yamanouchi M., Chiba D., Matsukura F. and Ohno H. 2004 Current-induced domain-wall switching in a ferromagnetic semiconductor structure Nature 428 539.
10. Amiri P. K. and Wang K. L. 2012 Voltage-controlled magnetic anisotropy in spintronic devices Spin 2 1240002.
11. Shiota Y., Nozaki T., Bonell F., Murakami S., Shinjo T. and Suzuki Y. 2012 Induction of coherent magnetization switching in few atomic layers of FeCo using voltage pulses Nat. Mater. 11 39.
12. Cui J., Hockel J. L., Nordeen P. K., Pisani D. M., Liang C. Y., Carman G. P. and Lynch C. S. 2013 A method to control magnetism in individual strain-mediated magnetoelectric islands Appl. Phys. Lett. 103 232905.
13. Ahmad, H., Atulasimha, J. and Bandyopadhyay, S. 2015 Electric field control of magnetic states in isolated and dipole coupled FeGa nanomagnets delineated on a PMN-PT substrate Nanotechnology 26 401001.
14. Ney, A., Pampuch, C., Koch, R. and Ploog, K. H. 2003 Programmable computing with a single magnetoresistive element. Nature 425, 485.
15. Behin-Aein, B., Datta, D., Salahuddin, S. & Datta S. 2010 Proposal for an all-spin logic device with built-in memeory. Nature Nanotech. 5, 266.
16. Biswas, A. K., Atulasimha, J. and Bandyopadhyay, S. 2014 An error-resilient non-volatile magneto-elastic universal logic gate with ultralow energy-delay product Sci. Rep. 4, 7553
17. Tehrani S., Slaughter J. M., Chen E., Durlam M., Shi J. and DeHerren M. 1999 Progress and outlook for MRAM technology IEEE Trans. Magn. 35 2814
18. Chun K. C., Zhao H., Harms J. D., Kim T.-H., Wang J.-P, and Kim C. H. 2013 A scaling roadmap and performance evaluation of in-plane and perpendicaulr MTJ based STT-MRAMS for high-density cache memory IEEE J. Solid State Cir. 48 598.
19. Oh S., Zheng G. and Ju K 2015 Perpendicular spin transfer torque (STT) memory cell with double MgO interface and CoFeB layer for enhancement of perpendicular magnetic anisotropy US Patent 9337415B1.
20. Biswas A. K., Atulasimha, J. and Bandyopadhyay, S. 2017 Energy-efficient hybrid spintronic-straintronic nonvolatile reconfigurable equality bit comparator SPIN 7 1750004.
21. Ikeda, S., Hayakawa, J., Ashizawa, Y., Lee, Y. M., Miura, K., Hasegawa, H., Tsunoda, M., Matsukura, F. and Ohno, H. 2008 Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo spin valves annealed at high temperature Appl. Phys. Lett. 93 082508.
22. Liu L., Pai C-F., Li Y., Tseng H. W., Ralph D. C. and Buhrman R. A. 2012 Spin-torque switching with the giant spin Hall effect of tantalum Science 336 555.
23. Pai C-F. Liu L., Li Y., Tseng H. W., Ralph D. C. and Buhrman R. A. 2012 Spin transfer torque devices utilizing the giant spin Hall effect of tungsten Appl. Phys. Lett. 101 122404.
24. Kanai S., Yamanouchi M., Ikeda S., Nakatani Y., Matsukura F. and Ohno H. 2012 Electric field induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction Appl. Phys. Lett. 101 122403.
25. Biswas A. K., Bandyopadhyay S. and Atulasimha J. 2014 Complete magnetization reversal in a magnetostrictive nanomagnet with voltage-generated stress: A reliable energy-efficient non-volatile magneto-elastic memory Appl. Phys. Lett. 105 072408.
26. Biswas A. K., Ahmad H., Atulasimha J. and Bandyopadhyay S. 2017 Experimental demonstration of complete 1800 reversal of magnetization in isolated Co nanomagnets on a PMN-PT substrate with voltage generated strain Nano Letters, 17, 3478.
27. Roy K., Bandyopadhyay S. and Atulasimha J. 2012 Energy dissipation and switching delay in stress-induced switching of multiferroic devices in the presence of thermal fluctuations J. Appl. Phys. 112 023914.
28. Roy K., Bandyopadhyay S. and Atulasimha J. 2011 Hybrid spintronics and straintronics: A magnetic technology for ultra-low energy computing and signal processing Appl. Phys. Lett. 99 063108.
29. D'Souza N., Fashami M. S., Bandyopadhyay S. and Atulasimha J. 2016 Experimental clocking of nanomagnets with strain for ultralow power Boolean logic Nano Lett. 16 1069.

30. Ahmad H., Atulasimha J. and Bandyopadhyay S. 2015 Reversible strain induced magnetization switching in FeGa nanomagnets: Pathway to rewritable, non-volatile, non-toggle, extremely low energy straintronic memory Sci. Rep. 5 18264.
31. G. P. Carman, private communication.
32. Alam M. T., Siddiq M. J., Bernstein G. H. Niemier M., Porod W. and Hu X. S. 2010 On-chip clocking for nanomagnetc logic devices IEEE Trans. Nanotechnol. 9 348.
33. Fukami S., et al. 2009 Low-current perpendicular domain wall motion cell for scalable high-speed MRAM 2009 VLSI Technology Digest of Technical Papers p.230.
34. Bhowmik D., You L. and Salahuddin S. 2014 Spin Hall effect clocking of nanomagnetic logic without a magnetic field Nat. Nanotech. 9 59.
35. Sengupta A., Panda P., Wijesinghe P., Kim Y. and Roy K. 2016 Magnetic tunnel junction mimics stochastic cortical spiking neurons Sci. Rep. 6 30039.
36. Sharad M., Fan D. and Roy K. 2013 Spin-neurons: A possible path to energy-efficient neuromorphic computers Appl. Phys. Lett. 114 234906
37. Biswas A. K., Atulasimha J. and Bandyopadhyay S. 2015 The straintronic spin neuron Nanotechnology 26 285201.

5. Graphene and 2D Material Electronics

A. 2D Materials for CMOS (Logic and Memories)

Background

During the past ten years, significant efforts have been spent on using 2D materials (graphene and beyond) for CMOS. The huge interest in 2D materials for electronics has been ignited by the work of the Geim-Novoselov group from Manchester University on graphene [1] and the observation of extraordinarily high carrier mobilities [2-3]. The motivation for using high-mobility graphene for MOSFET channels has been the same as that for introducing strained Si (in the past) as well as pure-Ge or SiGe (expected in the near future), namely enhancing the speed and lowering the power consumption of CMOS. Another motivation for introducing 2D materials into CMOS is that their ultimate thinness leads to ultra-short scale lengths suggesting an excellent suppression of short-channel effects.

Difficult Challenges

Table I. Challenges for introducing 2D materials into CMOS.

	Challenge	Comment
Graphene	• Opening a gap while maintaining a high mobility	Mobility-bandgap tradeoff, fundamental problem (most likely there is no solution).
	• Exact placement of large numbers of GNR channels at predefined positions	Similar problem as for carbon nanotubes.
	• Perfect alignment of GNR channels	Similar problem as for carbon nanotubes
Beyond graphene	• Deposition at the wafer scale (i) High crystallographic quality, low density of point and line defects, low density of grain boundaries, controlled grain size, defined and constant layer number	Currently CVD growth most popular. Here, grain boundaries could not be avoided yet.
	(ii) High-quality interfaces between 2D material and underlying substrate and top-gate dielectrics with low trap densities.	First progress made.
	• Realization of complementary n- and p-channel MOSFETs with a single 2D channel material	
	• Low resistance ohmic contacts	

Current Status, Prospects, and Challenges

Early work on graphene transistors already revealed that the missing bandgap of pristine graphene prevents proper switch-off of graphene MOSFETs, which is mandatory for CMOS transistors. A sizable gap can nevertheless be opened in GNRs (graphene nanoribbons) and intensive research on developing GNR MOSFETs for logic has been conducted. This work, however, revealed two major problems that so far could not be overcome. The first problem

relates to carrier transport. It has been shown both experimentally and theoretically that the gap opening in GNRs comes at the expense of a dramatic mobility reduction. Figure 1 shows the electron mobility in different 2D materials, GNRs, and conventional semiconductors (note that the hole mobility in semiconductors shows a similar trend in general, although the mobility drop for increasing bandgap is less pronounced, in particular for 2D materials beyond graphene).

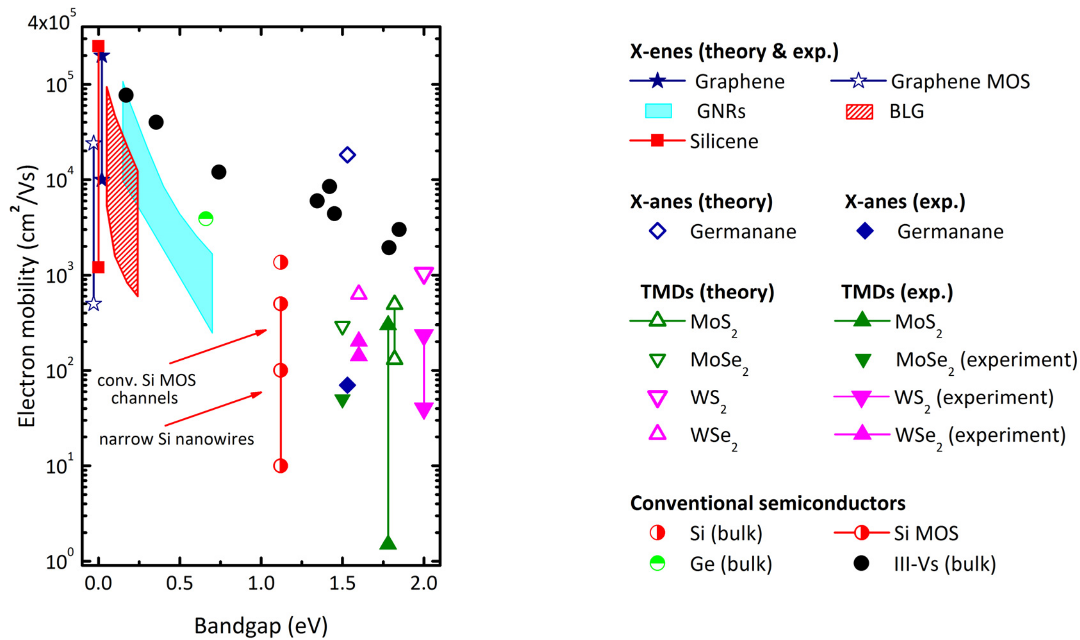


Figure 1. Room-temperature electron mobility vs. bandgap for different materials. The data for conventional 3D bulk semiconductors relates to undoped material. III-V materials (black solid circles) from left to right: InSb, InAs, $In_{0.53}Ga_{0.47}As$, InP, GaAs, $In_{0.52}Al_{0.48}As$, $Al_{0.3}Ga_{0.7}As$, $Ga_{0.5}1In_{0.49}P$. After [4], updated.

It can be seen that, for a given bandgap, GNRs do not offer a distinct advantage over conventional semiconductors in terms of mobility. This leads to the fact that the main motivation for the introduction of graphene into CMOS, i.e. the exploitation of its high carrier mobility, is lost when a gap is opened.

The second problem is related to the fabrication of GNRs. GNRs have been realized by patterning [5-6], i.e., lithography and etching, and by chemical synthesis [7-9]. Patterned GNRs suffer from rough non-ideal edges (varying edge configuration) and an altering width along their length, which deteriorates carrier transport. Moreover, the patterned GNRs reported so far are still too wide to open a gap sufficient for good switch-off. Chemically synthesized GNRs, in contrast, offer smooth edges with well-defined edge configuration (armchair or zigzag), constant width, and sufficiently wide gaps, see, e.g., Figure 6 in [10]. MOSFETs with synthesized GNR channels show good switch-off and on-off ratios around 10^5 but suffer from poor source/drain contacts. It is important to recognize that even if the contacts can be improved, the fundamental problem of the degraded mobility due to the gap opening remains. Moreover, for practical applications, hundreds of millions or even billions of GNR channels have to be placed exactly at predefined positions and must be perfectly aligned on the chip surface, which so far could not be demonstrated.

Just when the interest of the transistor community in graphene began to subside, the demonstration of single-layer MoS_2 FETs [11] gave new momentum to the research on 2D materials. Over a surprisingly short period of time, entire classes of new 2D materials have been discovered. Meanwhile almost 1,000 different 2D materials have been prepared or have been predicted to exist [12]. Many of them offer a sizeable bandgap and therefore are potentially suitable for CMOS. Figure 2 shows an (incomplete) overview on the wide field of 2D materials. Currently many groups are fabricating 2D MOSFETs, and the number of papers published on this topic has become literally unmanageable.

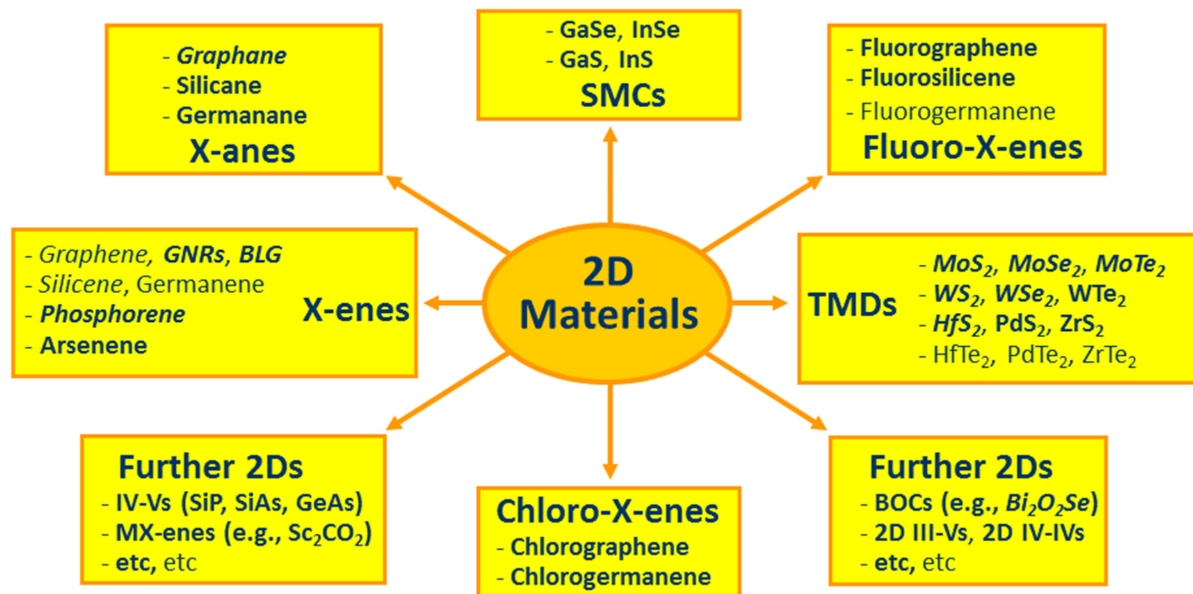


Figure 2. Classes and representative examples of 2D materials. SMC: Semimetal chalcogenide. BLG: Bilayer graphene. TMD: Transition metal dichalcogenide. BOC: Bismuth oxychalcogenide. Bold: 2D materials with sizable bandgap. Italic: 2D materials already used in experimental MOSFETs. After [4], updated.

While research on 2D MOSFETs has made substantial progress, we note that all semiconducting 2D materials follow the trend of a decreasing mobility for increasing bandgap as shown in Figure 1 and that, in terms of mobility, none of them has experimentally shown a real advantage over the conventional semiconductors yet. Moreover, the high-field transport properties of many 2D materials beyond graphene seem to be worse compared to Si. For example, electron saturation velocities are in the range of 0.1 to 0.6 × 10^7 cm/s [13-15] for MoS_2 and WS_2 compared to 10^7 cm/s for Si. Due to the low mobility and the low saturation velocity, MOSFETs with channels of these 2D materials will suffer from poor on-currents and will most likely not be able to compete with MOSFETs with Si, strained Si, SiGe, or Ge channels.

Several 2D materials beyond graphene have been successfully grown by CVD. However, the grown layers contain grains with pronounced grain boundaries (this is particularly true for CVD-grown TMD layers) and show unacceptably large trap densities. Thus, growing monocrystalline high-quality 2D materials at the wafer scale is still an open issue. Moreover, the source/drain contact resistances of 2D MOSFETs (beyond graphene) are still much too high. Finally, most as-grown 2D materials beyond graphene show a certain conductivity type (e.g., MoS_2 and WS_2 are intrinsically n-type conducting while $MoTe_2$ and phosphorene are intrinsically p-type conducting). For CMOS, however, both n-channel and p-channel MOSFETs are needed. Thus, either two different channel materials have to be used (definitely not desirable from the processing point of view) or practicable approaches to realize both n-channel and p-channel MOSFETs using a single 2D material have to be elaborated (work in this direction is underway and some progress has already been achieved [16-17]). For CMOS in particular, the TMDs offer the desirable feature of almost symmetric effective masses for electrons and holes, which suggests similar electron and hole mobilities.

Most transistor-related experimental work on 2D materials beyond graphene so far has been focused on TMDs. MoS_2 MOSFETs with very short gate lengths in the 1-10 nm range have been demonstrated. Particularly interesting is the 1-nm transistor reported in [18], which uses a 1-nm diameter carbon nanotube as gate and shows reasonably good switch-off and astonishingly small subthreshold swing. While the approach of using nanotube gates is certainly not suitable for mass production, this transistor has proven that 1 nm does not represent the physical limit for gate length scaling. Simulations of MoS_2 MOSFETs with 1 and 2 nm gate lengths support this statement and show that, due to the heavy carrier effective mass, ultra-scaled MoS_2 MOSFETs do not suffer from direct source-drain tunneling to an unacceptable extent [19-20]. While this is good news in general, currently it seems to be unlikely that 1-2 nm gate MOSFETs will ever reach the production stage since a paradigm change has been taking place in the industry. The consensus is that gate length scaling will level off around 10 nm, that the industry will favor FinFETs, thereby exploiting pitch scaling [21-22], and put more emphasis on 3D integration [23-24] instead of continuing aggressive gate length scaling at least until 2030 [25]. It should be noted that the designations for future technologies (e.g., 5 nm, 7 nm, 10 nm) do not reflect continued gate length scaling. The FinFETs for a 7 nm CMOS platform reported in [21], for example, have a gate length of 15 nm.

Summary

While research on 2D MOSFETs (graphene and beyond graphene) has created a lot of progress, the potential of these transistors for CMOS seems to be rather limited. It is not likely that 2D CMOS will enter mass production, at least not until 2030. Only if sub-5 nm gate MOSFETs return to the agenda, 2D materials with heavy carrier effective mass will become attractive since they very effectively suppress direct source-drain tunneling and guarantee reasonably good MOSFET performance down to gate length levels where other materials fail.

References

1 K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Electric field effect in atomically thin carbon films, Science, vol. 306, 666-669 (2004).
2 A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanabe, T. Taniguchi, and A. K. Geim, Micrometer-scale ballistic transport in encapsulated graphene at room temperature, Nano Lett., vol. 11, 2396-2399 (2011).
3 E. V. Castro, H. Ochoa, M. I. Katsnelson, R. V. Gorbachev, D. C. Elias, K. S. Novoselov, A. K. Geim, and F. Guinea, Limits on charge carrier mobility in suspended graphene due to flexural phonons, Phys. Rev. Lett., vol. 105, 266601 (2010).
4 F. Schwierz, J. Pezoldt, and R. Granzner, Two-dimensional materials and their prospects in transistor electronics, Nanoscale 7, 8261-8283 (2015).
5 G. Liu, Y. Wu, Y.-M. Lin, D. B. Farmer, J. A. Ott, J. Bruley, A. Grill, P. Avouris, D. Pfeiffer, A. A. Balandin, and C. Dimitrakopoulos, Epitaxial graphene nanoribbon array fabrication using BCP-assisted nanolithography, ACS Nano 6, 6786–6792 (2012).
6 W. S. Hwang, P. Zhao, K. Tahy, L. O. Nyakiti, V. D. Wheeler, R. L. Myers-Ward, C. R. E.ddy Jr., D. K. Gaskill, J. A. Robinson, W. Haensch, H. Xing, A. Seabaugh, and D. Jena, Graphene nanoribbon field-effect transistors on wafer-scale epitaxial graphene on SiC substrates, APL Mat. 3, 011101 (2015).
7 P. Ruffieux, S. Wang, B. Yang, C. Sanchez-Sanchez, J. Liu, T. Dienel, L. Talirz, P. Shinde, C. A. Pignedoli, D. Passerone, T. Dumslaff, X. Feng, K. Müllen, and R. Fasel, On-surface synthesis of graphene nanoribbons with zigzag edge topology, Nature 531, 489-493 (2016).
8 L. Talirz, H. So?de, T. Dumslaff, S. Wang, J. R. Sanchez-Valencia, J. Liu, P. Shinde, C. A. Pignedoli, L. Liang, V. Meunier, N. C. Plumb, M. Shi, X. Feng, A. Narita, K. Mu?llen, R. Fasel, and P. Ruffieux, On-surface synthesis and characterization of 9?atom wide armchair graphene nanoribbons, ACS Nano 11, 1380?1388 (2017).
9 J. P. Llinas, A. Fairbrother, G. B. Barin, W. Shi, K. Lee, S. Wu, B. Y. Choi, R. Braganza, J. Lear, N. Kau, W. Choi, C. Chen, Z. Pedramrazi, T. Dumslaff, A. Narita, X. Feng, K. Müllen, F. Fischer, A. Zettl, P. Ruffieux, E. Yablonovitch, M. Crommie, R. Fasel, and J. Bokor, Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons, Nat. Commun. 8, 633 (2017).
10 Z. Geng, B. Hähnlein, R. Granzner, M. Auge, A. A. Lebedev, V. Y. Davydov, M. Kittler, J. Pezoldt, and F. Schwierz, Graphene nanoribbons for electronic devices, Ann. Phys. 529, 1700033 (2017).
11 B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, Single-layer MoS2 transistors, Nature Nanotechnol., vol. 6, 147-150 (2011).
12 K. Bourzac, A Conversation with Toma?s Palacios, ACS Cent. Sci. 3, 677?678 (2017).
13 A. Y. Serov, V. E. Dorgan, C. D. English, and E. Pop, Multi-valley high-field transport in 2 dimensional MoS2 transistors, Proc. DRC, 183-184 (2014).
14 D. K. Ferry, Electron transport in some transition metal di-chalcogenides: MoS2 and WS2, Semicond. Sci. Technol. 32, 085003 (2017).
15 G. He, J. Nathawat, C.-P. Kwan, H. Ramamoorthy, R. Somphonsane, M. Zhao, K. Ghosh, U. Singisetti, N. Perea-López, C. Zhou, A. L. Elías, M. Terrones, Y. Gong, X. Zhang, R. Vajtai, P. M. Ajayan, D. K. Ferry, and J. P. Bird, Negative differential conductance and hot-carrier avalanching in monolayer WS2 FETs, Sci. Rep 7, 11256 (2017).
16 M. Tosun, S. Chuang, H. Fang, A. B. Sachid, M. Hettick, Y. Lin, Y. Zeng, and A. Javey, High-gain inverters based on WSe2 complementary field-effect transistors, ACS Nano 8, 4948–4953 (2014).
17 W. Sik Hwang, M. Remskar, R. Yan, V. Protasenko, K. Tahy, S. D. Chae, P. Zhao, A. Konar, H. Xing, A. Seabaugh, and D. Jena, Transistors with chemically synthesized layered semiconductor WS2 exhibiting 105 room temperature modulation and ambipolar behavior, Appl. Phys. Lett. 101, 013107 (2012).
18 S. B. Desai, S. R. Madhvapathy, A. B. Sachid, J. P. Llinas, Q. Wang, G. H. Ahn, G. Pitner, M. J. Kim, J. Bokor, C. Hu, H.-S. P. Wong, and A. Javey, MoS2 transistors with 1-nanometer gate lengths, Science 354, 99-102 (2016).
19 Z. Dong and J. Guo, Assessment of 2-D transition metal dichalcogenide FETs at sub-5-nm gate length scale, IEEE Trans. Electron Devices 64, 622-628 (2017).
20 S. Thiele, W. Kinberger, R. Granzner, G. Fiori, and F. Schwierz, The prospects of transition metal dichalcogenides for ultimately scaled CMOS, Solid-State Electron., in press (2018). https://doi.org/10.1016/j.sse.2017.11.004.
21 R. Xie, P. Montanini, K. Akarvardar, N. Tripathi, B. Haran, S. Johnson, T. Hook, B. Hamieh, D. Corliss, J. Wang, X. Miao, J. Sporre, J. Fronheiser, N. Loubet, M. Sung, S. Sieg, S. Mochizuki, C. Prindle, S. Seo, A. Greene, J. Shearer, A. Labonte, S. Fan, L. Liebmann, R. Chao, A. Arceo, K. Chung, K. Cheon, P. Adusumilli, H.P. Amanapu, Z. Bi, J. Cha, H.-

C. Chen, R. Conti, R. Galatage, O. Gluschenkov, V. Kamineni, K. Kim, C. Lee, F. Lie, Z. Liu, S. Mehta, E. Miller, H. Niimi, C. Niu, C. Park, D. Park, M. Raymond, B. Sahu, M. Sankarapandian, S. Siddiqui, R. Southwick, L. Sun, C. Surisetty, S. Tsai, S. Whang, P. Xu, Y. Xu, C. Yeh, P. Zeitzoff, J. Zhang, J. Li, J. Demarest, J. Arnold, D. Canaperi, D. Dunn, N. Felix, D. Gupta, H. Jagannathan, S. Kanakasabapathy, W. Kleemeier, C. Labelle, M. Mottura, P. Oldiges, S. Skordas, T. Standaert, T. Yamashita, M. Colburn, M. Na, V. Paruchuri, S. Lian, R. Divakaruni, T. Gow, S. Lee, A. Knorr, H. Bu, and M. Khare, A 7nm FinFET technology featuring EUV patterning and dual strained high mobility channels, Tech. Dig. IEDM, 47-50 (2016).

22 S.-Y. Wu, C. Y. Lin, M. C. Chiang, J. J. Liaw, J. Y. Cheng, S. H. Yang, C. H. Tsai, P. N. Chen, T. Miyashita, C. H. Chang, V. S. Chang, K. H. Pan, J. H. Chen, Y. S. Mor, K. T. Lai, C. S. Liang, H. F. Chen, S. Y. Chang, C. J. Lin, C. H. Hsieh, R. F. Tsui, C. H. Yao, C. C. Chen, R. Chen, C. H. Lee, H. J. Lin, C. W. Chang, K. W. Chen, M. H. Tsai, K. S. Chen, Y. Ku, S. M. Jang, A 7nm CMOS platform technology featuring 4th generation FinFET transistors with a 0.027um2 high density 6-T SRAM cell for mobile SoC applications, Tech. Dig. IEDM, 43-46 (2016).

23 A. Mallik, A. Vandooren, L. Witters, A. Walke, J. Franco, Y. Sherazi, P. Weckx, D. Yakimets, M. Bardon, B. Parvais, P. Debacker, B. W. Ku, S. K. Lim, A. Mocuta, D. Mocuta, J. Ryckaert, N. Collaert, and P. Raghavan, The impact of sequential-3D integration on semiconductor scaling roadmap, Tech. Dig. IEDM, 717-720 (2017).

24 N. Collaert, More Moore: From device scaling to 3D integration and system-technology co-optimization, Proc. Si Nanoelectron. Workshop, 123-124 (2017).

25 The International Roadmap for Semiconductors 2015, available at www. semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/.

B. 2D Materials for RF Electronics

Background

RF transistors do not necessarily need to be switched off. This fact and the high mobilities in pristine graphene have spurred expectations that graphene RF MOSFETs could supersede the conventional RF transistors, i.e., III-V HEMTs (high electron mobility transistor), Si RF MOSFETs, as well as III-V and SiGe HBTs (heterojunction bipolar transistor), and enable useful power amplification at THz frequencies [1-2].

Difficult Challenges

Table I. Challenges for using 2D materials in RF transistors.

	Challenge	Comment
Graphene RF MOSFETs	• Improving the saturation of the output characteristics • Opening a gap while maintaining the high mobility of pristine graphene • Perfect alignment of GNR channels	Difficult with pristine graphene channels (fundamental problem for gapless channels) Gap opening possible in bilayer graphene and GNRs, but: Mobility – bandgap tradeoff, fundamental problem (most likely there is no solution)
Beyond graphene RF MOSFETs	• Improving the mobility • Low resistance ohmic contacts	Mobility – bandgap tradeoff (fundamental problem).

Current Status, Prospects, and Challenges

Many groups have performed intensive research on graphene RF MOSFETs with impressive performance, e.g., a record cutoff frequency f_t of 427 GHz for a 67-nm gate graphene MOSFET [3]. It turned out, however, that MOSFETs with pristine gapless channels suffer from relatively low power gains and, consequently, low maximum frequencies of oscillation f_{max}. Note that while a high f_t (f_t is the frequency at which the small-signal current gain has dropped to unity, i.e., 0 dB) is certainly desirable, f_{max}, i.e., the frequency where the small-signal power gain drops to unity, is the much more important figure of merit for RF transistors. The reason for the low power gain and f_{max} of graphene MOSFETs is the poor saturation of the output characteristics of these transistors, which is caused by the gapless nature of graphene. Thus, the missing gap is an extremely serious problem for both logic MOSFETs with pristine graphene channels and for RF graphene MOSFETs. Figure 1 showing the f_{max}-f_t performance of graphene RF MOSFETs together with that of competing RF FET types indicates that in particular state-of-the-art III-V RF HEMTs (InP and GaAs mHEMTs, m stands for metamorphic) perform much better than their graphene counterparts.

GNR MOSFETs are expected to show a better saturation of the output characteristics (due to the gap opening). So far, however, GNR RF MOSFETs have not been demonstrated.

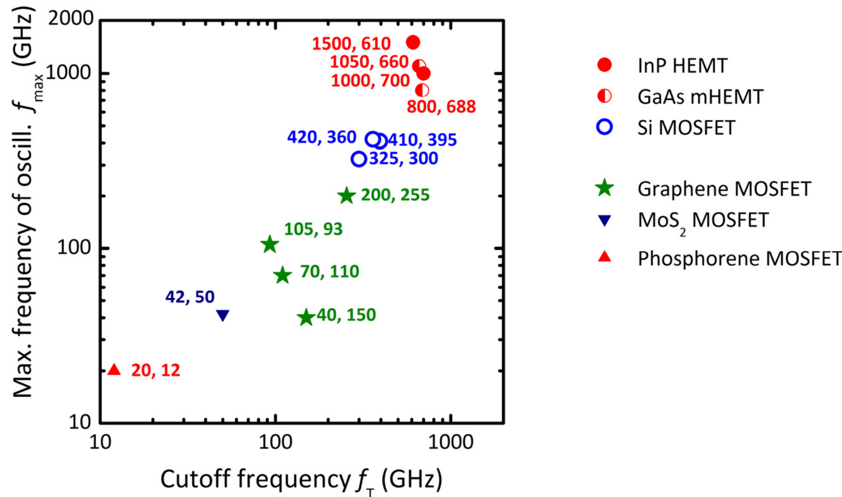


Figure 1. Maximum frequency of oscillation fmax versus cutoff frequency fT of graphene MoS_2, and phosphorene RF MOSFETs together with the corresponding data for competing RF FETs (InP HEMTs, GaAs mHEMT, and Si RF MOSFETs). After [5], updated.

RF MOSFETs with channels of 2D materials beyond graphene, in particular MoS_2 and phosphorene, have also been reported. As indicated by the representative data points in Figure 1 above, these transistors perform even worse compared to graphene RF MOSFETs. The main reason for this unsatisfying behavior is the relatively low carrier mobility in MoS_2 and phosphorene, see Figure 1 in Section A. Note that a high carrier mobility, accompanied by a good saturation of the output characteristics and a short gate length, is a precondition for a good RF FET.

Summary

The above discussion leads to the conclusion that 2D MOSFETs in general are not suitable for applications at ultra-high operating frequencies and for high-performance RF applications. It may be argued that 2D MOSFETs might be useful for RF applications at lower operating frequencies. For such purposes, however, a wide range of established and matured RF transistor technologies is available, e.g., Si bipolar RF transistors, Si LDMOSFETs (Laterally Diffused MOSFET), GaAs MESFETs (Metal-Semiconductor FET), conventional AlGaAs/GaAs HEMTs, pseudomorphic GaAs HEMTs, etc. [6-7]. This will make it extremely difficult for 2D RF MOSFETs to make inroads into the RF market. A single exception could be flexible RF circuits, see Section C next.

References

1 A. K. Geim and K. S. Novoselov, The rise of graphene, Nature Mat. 6, 183-191 (2007).
2 A. K. Geim, Graphene: Status and prospects, Science 324, 1530-1534 (2009).
3 R. Cheng, J. Bai, L. Liao, H. Zhou, Y. Chen, L. Liu, Y.-C. Lin, S. Jiang, Y. Huang, and X. Duan, High-frequency self-aligned graphene transistors with transferred gate stacks, PNAS 109, 11588-11592 (2012).
4 F. Schwierz, Graphene transistors: Status, prospects, and problems, Proc. IEEE 101, 1567-1584 (2013).
5 M. C. Lemme, L. Li, T. Palacios, and F. Schwierz, Two-dimensional materials for electronic applications, MRS Bulletin 39, 711-718 (2014).
6 F. Schwierz and J. J. Liou, RF transistors: Recent developments and roadmap toward terahertz applications, Solid-State Electron. 51, 1079-1091 (2007).
7 F. Schwierz and J. J. Liou, Modern Microwave Transistors - Theory, Design, and Performance, John Wiley & Sons, New York 2003.

C. 2D Materials for Flexible and Stretchable Electronics

2D materials are inherently bendable and stretchable and therefore natural candidates for flexible/stretchable electronics. Indeed, a lot of work on flexible 2D transistors and circuits has been done and remarkable results have been achieved [1]. Graphene MOSFETs show some potential for flexible RF applications. Although their RF performance is worse compared to high-performance III-V HEMTs (see Section B), they perform much better and show significantly higher cutoff frequencies f_t and maximum frequencies of oscillation f_{max} than OFETs (organic FETs), which are popular for flexible electronics. The fastest OFETs show f_t's below 30 MHz [2] (f_{max} data for OFETs has not been reported so far at all) compared to tens of GHz for flexible graphene MOSFETs. Moreover, semiconducting 2D materials (in particular MoS_2 and phosphorene) are attractive as channel materials for 2D

MOSFETs for both digital and RF flexible applications [3-5]. Figure 1 summarizes the reported f_t-f_{max} performance of flexible 2D MOSFETs.

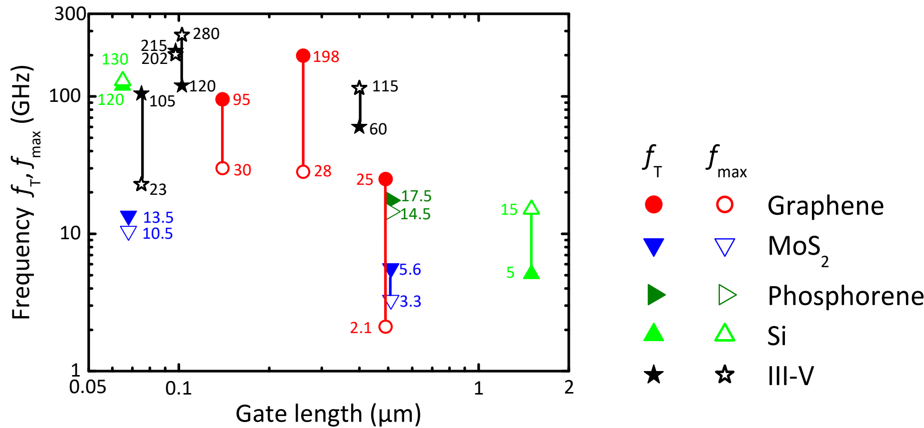


Figure 1. RF performance of flexible FETs in terms of f_T and f_{max} vs. gate length. After [6], updated.

As can be seen, Figure 1 also contains data for Si and III-V FETs, which points to the fact that 2D MOSFETs and OFETs are not the only options for flexible electronics. The flexible Si and III-V transistors included in Figure 1 have been fabricated on rigid substrates, which after transistor processing have been thinned to an extent that they become bendable, e.g. [7-9]. Such extremely thin Si and III-V substrates are frequently called nanomembranes and can easily be transferred to arbitrary flexible substrates such as polyimide.

It should be noted that the f_t-f_{max} data of the flexible 2D transistors from Figure 1 should be treated with caution since the de-embedding procedure for RF 2D transistors is frequently performed not in the same way as for Si and III-V RF transistors. De-embedding is a common practice in RF electronics to eliminate the effects of the parasitics of the measurement environment from the measured RF data. Usually all parasitics down to the large pads (needed for the RF probes) are de-embedded while the metal lines from the pads to the transistor are not de-embedded. In the RF characterization of 2D MOSFETs, however, frequently these metal lines are de-embedded as well. This full de-embedding procedure provides the RF parameters of the intrinsic device, which are difficult to compare to those obtained by the common pad de-embedding approach and leads to a very optimistic picture of the transistor's RF performance. This issue has been discussed in [10] and its relevance becomes evident from Table I comparing the f_t and f_{max} data of a 260-nm gate flexible graphene RF MOSFET obtained by different de-embedding procedures.

Table I. f_T and f_{max} of a 260-nm gate graphene MOSFETs obtained by different de-embedding procedures [10].

	As measured	Pad de-embedding	Full de-embedding
F_t (GHz)	23.6	38.7	198
f_{max} (GHz)	6.5	7.6	28.2

References

1 D. Akinwande, N. Petrone, and J. Hone, Two-dimensional flexible nanoelectronics, Nat. Commun. 5, 5678 (2014).
2 X. Guo, Y. Xu, S. Ogier, T. N. Ng, M. Caironi, A. Perinot, L. Li, J. Zhao, W. Tang, R. A. Sporea, A. Nejim, J. Carrabina, P. Cain, and F. Yan, Current status and opportunities of organic thin-film transistor technologies, IEEE Trans. Electron Devices 64, 1906-1921 (2017).
3 H.-Y. Chang, S. Yang, J. Lee, L. Tao, W.-S. Hwang, D. Jena, N. Lu, and D. Akinwande, High-performance, highly bendable MoS2 Transistors with high-K dielectrics for flexible low-power systems, ACS Nano 7, 5446-5452 (2013).
4 R. Cheng, S. Jiang, Y. Chen, Y. Liu, N. Weiss, H.-C. Cheng, H. Wu, Y. Huang, and X. Duan, Few-layer molybdenum disulfide transistors and circuits for high-speed flexible electronics, Nat. Commun. 5, 5143 (2014).
5 W. Zhu, S. Park, M. N. Yogeesh, K. M. McNicholas, S. R. Bank, and D. Akinwande, Black phosphorus flexible thin film transistors at gighertz frequencies, Nano Lett. 16, 2301-2306 (2016).
6 Z. Geng, W. Kinberger, R. Granzner, J. Pezoldt, and F. Schwierz, 2D electronics – opportunities and limitations, Proc. ESSDERC, 230-235 (2016).
7 H. Sharifi, K. Shinohara, W. Ha, J. Kang, M. Monte, C. McGuire, J. May and H. Kazem, Microwave and millimeter-wave flexible electronics, Proc. IEEE MTT-S, 1-4 (2014).
8 T.-H. Chang, K. Xiong, S. H. Park, H. Mi, H. Zhang, S. Mikaell, Y. H. Jung, J. Han, and Z. Ma, High power fast flexible electronics: Transparent RF AIGaN/GaN HEMTs on plastic substrates, Proc. IEEE MTT-S, 1-4 (2015).

9 J. Philippe, A. Lecavelier, M. Berthome, J.-F. Robillard, C. Gaquiere1, F. Danneville, D. Gloria, C. Raynaud, and E. Dubois, Application-oriented performance of RF CMOS technologies on flexible substrates, Tech. Dig. IEDM, 406-409 (2015).
10 N. Petrone, I. Meric, T. Chari, K. L. Shepard, and J. Hone, Graphene field-effect transistors for radio-frequency flexible electronics, J. Electron Device Soc., vol. 3, 44-48 (2015).

D. 2D Materials for Neuromorphic Devices

Background on Neuromorphic Computing (CMOS-based von Neumann vs Neuromorphic)

A general discussion on neuromorphic devices, architectures and computing along with current status can be found under Section 2 of this chapter. The focus here is only on 2D materials.

Over decades, computing performance has evolved at an enormous rate. The huge progress achieved, however, has been purely evolutionary. Since the introduction of the first CMOS CPU, the CDP 1801 from RCA [1], CMOS with the Si MOSFET as its basic component has been and still is the one and only technology used successfully and industry-wide for computing. The continuously enhanced computing power has been achieved by increasing the circuit complexity according to Moore's Law, accompanied (and enabled) by MOSFET scaling. Furthermore, computers are still based on the von Neumann architecture. A characteristic feature of von Neumann computers is the spatial separation of information processing in the CPU from information storage in the memory [2]. This separation makes a permanent intensive communication between the fast CPU and the slow memory inevitable. It results in significant power consumption (part of the community considers the power consumption of data centers and supercomputers already as inadmissibly high) and causes a limitation of the system speed. Therefore, it is called the von Neumann bottleneck. Another problem of conventional computers is that further CMOS scaling is getting increasingly challenging. Meanwhile there is a consensus that CMOS scaling will irrevocably come to an end within a few years. As a consequence, a worldwide search for post-CMOS devices and beyond-von-Neumann computer architectures is underway. One promising approach that has attracted significant attention recently is neuromorphic computing. The term neuromorphic was introduced by Mead almost 30 years ago [3] and means information processing inspired by biology (i.e., by the animal or human brain). Neuromorphic computing is based on entirely different principles compared to von Neumann computing and therefore really represents new territory. CMOS logic gates performing Boolean logic operations are the basis of von Neumann computers, while neurons and synapses form the basic elements of neuromorphic systems. Von Neumann computers operate essentially sequentially, need precisely specified input data, and are ideally suited for solving structured and well-defined mathematical problems [4]. Neuromorphic systems, on the other hand, inherently comprise a high degree of parallelism, can handle imprecisely specified data, and perform nondeterministic operations [2].

Neuromorphic systems can be realized in fundamentally different ways. One approach is to emulate the operation and interaction of neurons and synapses by conventional CMOS circuits. Work in this direction is already at an advanced stage. Examples for CMOS-based neuromorphic systems are the TrueNorth processor from IBM [5] and the SpiNNaker computer [6]. The more elegant and revolutionary approach, however, is to use elements whose characteristics and operation resemble that of neurons and synapses, i.e., memristors.

?

Difficult Challenges – Short Term (5 Years)

Table I. Challenges (short term) for 2D memristors.

Challenge	Comment
Elaboration of a comprehensive list of relevant performance parameters (targets)	Examples for relevant targets are Ron/Roff ratio, endurance, power consumption
Specification of the targets (providing numbers, ranges)	
Development of memristive device concepts (including device structures, material combinations, processes) showing resistance switching without electroforming	
Elaboration of a deep understanding of the operation of 2D memristors and of the origins of resistance switching	Understanding the physics, elaboration of physics-based models

Difficult Challenges – Long Term (10-25 Years)

Table II. Challenges (long term) for 2D memristors.

Challenge	Comment
Providing clear evidence (prove) that 2D memristors are a serious and highly competitive device option for neuromorphic computing	Fair comparison with CMOS-based neuromorphic systems, 2D memristors should demonstrate significantly improved performance. Devices: in terms of electrical parameters (endurance, power consumption, etc.), yield, and scalability Circuits/systems: in terms of circuit complexity, cost
Prove that 2D memristors are able to meet the performance targets	
Critical benchmarking of 2D memristors with competing emerging concepts and architectures	Benchmarking at the device, circuit, and system levels
Development of physics-based device simulators, compact models, circuit simulators, EDA tools	Comparable to CMOS tools, e.g., ATLAS/Sentaurus (device simulators), BSIM/EKV (compact models), Spice (circuit simulator), Cadence (EDA tool)
Development of technologies for realizing ultra-scaled ultra-dense 2D memristor circuits and systems	Large arrays, preferably crossbar arrays, of memristor structures showing acceptably small variability of device parameters.

Background on Memristors

Per definition, a memristor (memory-resistor) is a two-terminal device showing the following characteristic features [7-8]:

- Its resistance depends on the magnitude and polarity of the voltage applied to it AND on the length of the time that voltage has been applied.
- When the voltage is turned off, the device keeps its resistance (i.e., it remembers the resistance it had immediately before the voltage has been turned off) until the next time the voltage is turned on again.
- It has a hysteretic I-V (current-voltage) characteristics, a so-called pinched hysteresis loop with zero crossing, whose shape is frequency dependent (for increasing frequency, the hysteresis gets weaker and finally disappears), and shows resistance switching. This means that the I-V characteristics have a high-resistance portion (off-state with high off-resistance R_{off}) and a low-resistance portion (on-state with small on-resistance R_{on}).

Memristors have been used successfully to realize artificial non-biological neurons [9-12] and synapses [13-18]. It should be noted that devices showing the features mentioned are frequently not called memristors but are designated as atomic switches [19], ReRAMs (Resistance Random Access Memory) [20], or phase-change memories [21]. Moreover, a lot of work on memristive devices is not related to neuromorphic computing but to memory elements like ReRAMs and logic gates performing Boolean logic operations such as IMPLY, MAGIC, and MAD gates [22-25]. This makes the field of memristor applications quite diverse as shown in Figure 1 – see also the discussion in [26].

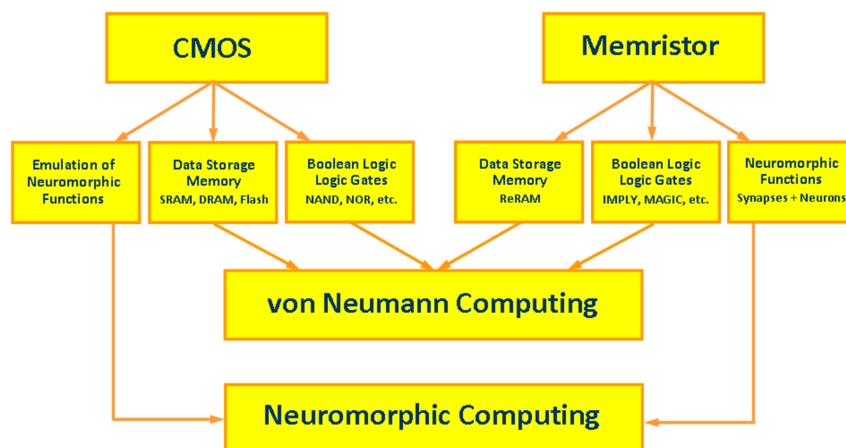
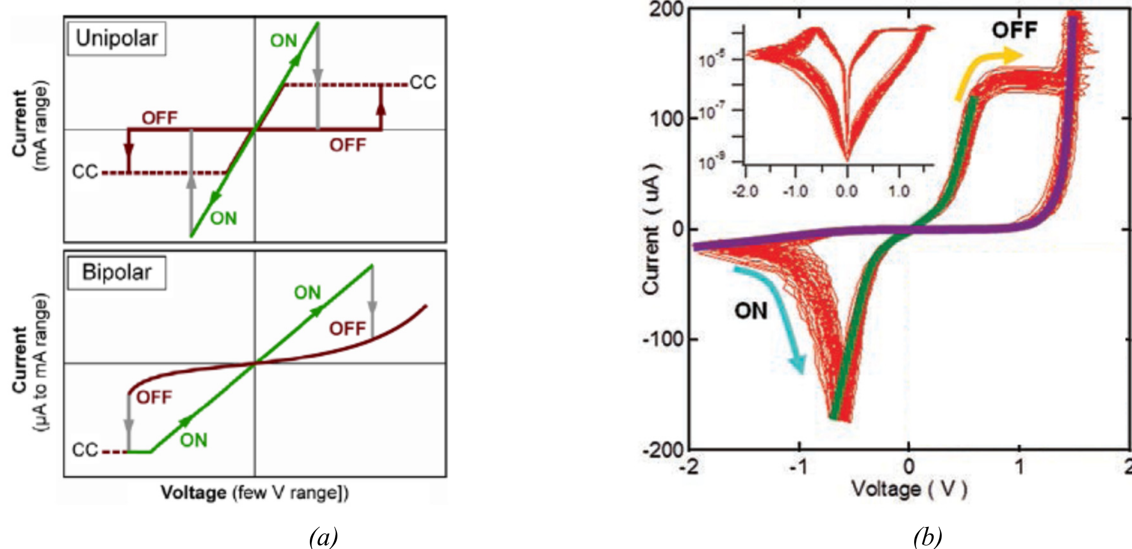


Fig. 1. Application of CMOS devices and memristors for von Neumann and neuromorphic computing.

Memristors based on a variety of different material systems have been demonstrated. They typically consist of a vertical metal/insulator/metal layer stack and show either unipolar or bipolar switching behavior as depicted in Figure 2. The schematic I-V characteristics in Figure 2(a) show that a device operating in the bipolar switching mode needs a certain voltage with one polarity to be switched into the on-state and a different voltage with opposite polarity to be switched off (bottom), while for devices operating in the unipolar mode, switching does not depend on the polarity of the voltage (top). Figure 2(b) shows the I-V characteristics of 50 experimental memristors with bipolar switching behavior.



(a) *(b)*

Fig. 2. Memristor IV characteristics. (a) Schematic IV characteristics of memristors showing unipolar (top) and bipolar switching (bottom). For both types of memristors the slopes of the IV curves are large in the on-state indicating a small Ron and much smaller in the off-state, i.e., large Roff [27]. (b) IV characteristics of 50 experimental Pt/TiO_2-x/TiO_2/Pt memristor devices [28].

Memristor switching is caused by slight structural changes inside the device. The mechanisms leading to these changes, however, are still unclear and heavily under debate. Different mechanisms and models have been suggested [28] and it is not clear whether a single mechanism is responsible for switching or if several mechanisms coexist. Resolving this issue is particularly difficult since the active regions of the memristors are commonly covered by a metal electrode so that the changes cannot be observed. Another problem is the fact that almost all reported memristor devices right after processing do not show the switching behavior discussed above unless an electroforming procedure is carried out. During electroforming, a high voltage is applied once which causes the formation of a new material phase in the device and makes it capable of switching. It should be noted the actual switching properties of the memristor after electroforming are neither well defined not predictable and vary over a wide range depending on the details of the electroforming process. Therefore, the electroforming process is considered as the least understood and most problematic step in the realization of memristors [28].

Current Status, Prospects, and Challenges – General Remarks

During the last few years, research on memristors in the three areas shown in Figure 1, i.e., (i) memories for von Neumann computing, (ii) logic gates performing Boolean operations for von Neumann computing, and (iii) neurons and synapses for neuromorphic computing, has created a lot of progress. Most notably, ReRAMs can be considered as ready for commercialization [29-31]. Memristors for Boolean logic and neuromorphic computing, on the other hand, are still far from real-world applications and many open questions do exist.

Researchers working on memristors for neuromorphic systems face the following quite specific problem. So far, the requirements on memristors for this purpose are vaguely formulated at best. This situation is unsatisfying and much different from that in CMOS. In the CMOS field, as a result of long-term close collaborations between device engineers, circuit designers, and system people, a set of widely accepted performance targets (such as required on-currents, tolerable off-currents, supply voltages, delay times, etc.) has been elaborated and published, e.g., in the subsequent ITRS editions [32]. Thus, CMOS device engineers knew exactly the requirements on new generations of CMOS transistors they were developing. Moreover, the prospects of novel alternative device concepts could be assessed in a qualified manner by benchmarking their expected performance against that of future generations of Si CMOS [33-35]. The situation for memristors to be used in neuromorphic systems is completely different. Neither

are the relevant FOMs (figure of merit) well established nor are quantified targets (e.g., numbers, ranges) for these FOMs specified. Thus, literally speaking, researchers developing memristors do not know how good these devices should perform to be useful for applications.

Current Status, Prospects, and Challenges – Memristors Using 2D Materials

Experimental devices based on 2D graphene and TMDs showing memristive behavior have been demonstrated recently. Posa et al. have reported graphene-SiO_x-graphene memristors consisting of two graphene electrodes located on SiO_2, which are laterally separated by a narrow gap only a few nm wide [36]. Within the gap, microscopically distinct SiO_x amorphous and crystalline phases are formed and constitute the active region of the devices [36]. A vertical graphene memristor consisting of Pd top and bottom electrodes with a Ta/graphene/Ta_2O_5 stack in between has been reported in [37]. Here, the graphene layer has engineered nanopores and the resistive switching is caused by the motion of oxygen vacancies through the nanopores (while outside the pores the vacancies are blocked by the graphene).

Bessonov et al. [38] demonstrated MoO_x/MoS_2 and WO_x/WS_2 stacks sandwiched between two silver electrodes. Interestingly, these vertical devices have been realized on flexible polymer instead of rigid oxidized Si substrates and show large and tunable resistances range from 10^2 to 10^8 Ω (i.e., R_{on}/R_{off} ratio 10^6), combined with low switching voltages of 0.1-0.2 V. Another vertical memristor structure with a 1T-MoS_2 layer between two silver electrodes has been reported in [39].

The Hersam group from Northwestern University has demonstrated lateral single-layer MoS_2 memristors located on an oxidized Si substrate showing R_{on}/R_{off} ratios up to 10^6 [40-41], whose operation relies on the existence of grain boundaries in the active MoS_2 layer located between the two Au contacts. Three configurations of such MoS_2 memristors with different types (shapes) of grain boundaries (intersecting, bridge, and bisecting) causing three different types of switching characteristics have been fabricated. The authors suggest that switching in their MoS_2 memristors is related to the modulation of the density of sulphur vacancies in regions close to the grain boundaries. Two further aspects of these lateral memristors are worth mentioning. First, by using the doped Si substrate as a back-gate, the behavior of the memristors can be controlled by the back-gate voltage. This represents an additional degree of flexibility to adjust the device's operation that does not exist in vertical memristors. Second, the active memristor region, i.e., the MoS_2 layer, is located at the surface and thus accessible to a variety of surface analysis techniques. This may be of great help in gaining deeper insights into the so far not well understood electroforming and switching processes. A drawback of such lateral memristors is the fact that crossbar structures offering densely packed memristor arrays will be difficult to realize.

Wang et al. have demonstrated memristors with multi-layer $MoS_{2-x}O_x$ in the active region and graphene top and bottom electrodes [42]. These devices show a high endurance of 10^7 and operate well at temperatures up to 340°C. Finally, the demonstration by the Akinwande group from the University of Texas at Austin consists of a single layer of a TMD (MoS_2, $MoSe_2$, WS_2, WSe_2) sandwiched between top and bottom Au contacts [42]. An intriguing feature of these devices is that an electroforming process is not required. Instead, they show resistance switching immediately after processing.

References

1 http://www.antiquetech.com/?page_id=682

2 R. A. Nawrocki, R. M. Voyles, and S. E. Shaheen, A mini review of neuromorphic architectures and implementations, IEEE Trans. Electron Devices 63, 3819-3829 (2016).

3 C. Mead, Neuromorphic electronic systems, Proc. IEEE 78, 1629-1636 (1990).

4 D. Kuzum, S. Yu, and H.-S. P. Wong, Synaptic electronics: materials, devices and applications, Nanotechnology 24, 382001 (2013).

5 J. Sawada, F. Akopyan, A. S. Cassidy, B. Taba, M. V. Debole, P. Datta, R. Alvarez-Icaza, A. Amir, J. V. Arthur, A. Andreopoulos, R. Appuswamy, H. Baier, D. Barch, D. J. Berg, C. di Nolfo, S. K. Esser, M. Flickner, T. A. Horvath, B. L. Jackson, J. Kusnitz, S. Lekuch, M. Mastro, T. Melano, P. A. Merolla, S. E. Millman, T. K. Nayak, N.Pass, H. E. Penner, W. P. Risk, K. Schleupen, B. Shaw, H. Wu, B. Giera, A. T. Moody, N. Mundhenk, B. C. Van Essen, E. X. Wang, D. P. Widemann, Q. Wu, W. E. Murphy, J. K. Infantolino, J. A. Ross, D. R. Shires, M. M. Vindiola, R. Namburu, and D. S. Modha, TrueNorth ecosystem for brain-inspired computing: Scalable systems, software, and applications, Proc. SC16, 130-141 (2016).

6 S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, The SpiNNaker Project, Proc. IEEE 102, 652-665 (2014).

7 R. S. Williams, How we found the missing memristor, IEEE Spectrum 45, iss. 12, 29-35 (2008).

8 J. J. Yang, D. B. Strukov, and D. R. Stewart, Memristive devices for computing, Nat. Nanotech. 8, 13-24 (2013).

9 D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, A scalable neuristor built with Mott memristors, Nat. Mater. 12, 114-117 (2013).

10 J. Lin, A. Annadi, S. Sonde, C. Chen, L. Stan, K. V. L. V. Achari, S. Ramanathan, and S. Guha, Low-voltage artificial neuron using feedback engineered insulator-to-metal-transition devices, Tech. Dig. IEDM, 862-865 (2016).
11 T. Tuma, A. Pantazi, M- Le Gallo, A. Sebastian, and E. Eleftheriou, Stochastic phase-change neurons, Nat. Nanotech. 11, 693 (2016).
12 X. Zhang, W. Wang, Q. Liu, X. Zhao, J. Wei, R. Cao, Z. Yao, X. Zhu, F. Zhang, H. Lv, S. Long, and M. Liu, An artificial neuron based on a threshold switching memristor, IEEE Electron Device Lett. 39, in press (2018).
13 C. Zamarreno-Ramos, L. A. Camunas-Mesa, J. A. Perez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, and B. Linares-Barranco, On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual corte, Front. Neurosci. 5, 26 (2011).
14 A. Demming, J. K. Gimzewski, and D. Vuillaume, Synaptic electronics, Nanotechnol. 24, 380201 (2014).
15 T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco, STDP and STDP variations with memristors for spiking neuromorphic learning systems," Front. Neurosci. 7, 2 (2013).
16 M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. DeSalvo, Bio-inspired stochastic computing using binary CBRAM synapses, IEEE Trans. Electron Devices 60, 2402–2409 (2013).
17 S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetti, and D. Ielmini, Spiketiming dependent plasticity in a transistor-selected resistive switching memory, Nanotechnology 38, 384012 (2013).
18 S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, Nanoscale memristor device as synapse in neuromorphic systems, Nano Lett. 10, 1297–1301 (2010).
19 Z. Wang, T. Gu, T. Tada and S. Watanabe, Excess-silver-induced bridge formation in a silver sulfide atomic switch, Appl. Phys. Lett. 93, 152106 (2008).
20 H. Akinaga and H. Shima, Resistive Random Access Memory (ReRAM) Based on Metal Oxides, Proc. IEEE 98, 2237-xx (2010).
21 H.-S. P. Wong, S. Raoux, S. B. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, Phase Change Memory, Proc. IEEE 98, 2201-2227 (2010).
22 S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, Memristor-based material implication (IMPLY) logic: Design principles and methodologies, IEEE Trans. VLSI Syst. 22, 2054-2066 (2014).
23 S. Kvatinsky, D. Belousov, S. Liman, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, MAGIC—Memristor-aided logic, IEEE Trans. Circuits Syst. II, Exp. Briefs 895-899 (2014).
24 L. Guckert, and E. E. Swartzlander, Optimized memristor-based multipliers, IEEE Trans. Circ. Syst. – I, 64, 373-385 (2017).
25 Y.-F. Chang, F. Zhou, B. W. Fowler, Y.-C. Chen, C.-C. Hsieh, L. Guckert, E. E. Swartzlander, and J. C. Lee, Memcomputing (memristor + computing) in intrinsic SiOx-based resistive switching memory: Arithmetic operations for logic applications, IEEE Trans. Electron Devices 64, 2977-2983 (2017).
26 I. Vourkas and G. Ch. Sirakoulis, Emerging memristor-based logic circuit design approaches: A review, IEEE Circuits Syst. Mag., 16, Iss. 3, 15-30 (2016).
27 R. Waser, Resistive non-volatile memory devices, Microelectron. Eng. 86, 1925–1928 (2009).
28 J. J. Yang and R. S. Williams, Memristive devices in computing system: Promises and challenges, ACM J. Emerg. Technol. Comput. Syst. 9, 11 (2013).
29 https://www.eetimes.com/document.asp?doc_id=1325521
30 http://www.memristor.org/news/475/hp-and-hynix-semiconductor-reram-memristor-set-to-market
31 https://www.eetimes.com/document.asp?doc_id=1327307
32 http://www.itrs2.net/itrs-reports.html and https://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/
33 K. Bernstein, R. K. Cavin, III, W. Porod, A. Seabaugh, and J. Welser, Device and architecture outlook for beyond CMOS switches, Proc. IEEE 98, 2169-2184 (2010).
34 D. E. Nikonov, and I. A. Young, Overview of beyond-CMOS devices and a uniform methodology for their benchmarking, Proc. IEEE 101, 2498-2533 (2013).
35 D. E. Nikonov and I. A. Young, Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits, IEEE J. Explor. Solid-State Computat. Devices Circuits 1, 3-11 (2015).
36 L. Posa, M. El Abbassi, P. Makk, B. Santa, C. Nef, M. Csontos, M. Calame, and A. Halbritter, Multiple physical time scales and dead time rule in few-nanometers sized graphene-SiOx-graphene memristors, Nano Lett. 17, 6783?6789 (2017).
37 J. Lee, C. Du, K. Sun, E. Kioupakis, and W. D. Lu, Tuning ionic transport in memristive devices by graphene with engineered nanopores, ACS Nano 16, 3571?3579 (2016).
38 A. A. Bessonov, M. N. Kirikova, D. I. Petukhov, M. Allen, T. Ryhänen, and M. J. A. Bailey, Layered memristive and memcapacitive switches for printable electronics, Nat. Mater 14, 199-204 (2015).
39 P. Cheng, K. Sun, and Y. H. Hu, Memristive behavior and ideal memristor of 1T Phase MoS2 nanosheets, Nano Lett. 16, 572?576 (2016).
40 V. K. Sangwan, D. Jariwala, I. S. Kim, K.-S. Chen, T. J. Marks, L. J. Lauhon, and M. C. Hersam, Gate-tunable memristive phenomena mediated by grain boundaries in single-layer MoS2, Nat. Nanotechnol. 10, 403-406 (2015).

41 V. K. Sangwan, H.-S. Lee, and M. C. Hersam, Gate-tunable memristors from monolayer MoS2, Tech. Dig. IEDM, 107-110 (2017).
42 M. Wang, S. Cai, C. Pan, C. Wang, X. Lian, Y. Zhuo, K. Xu, T. Cao, X. Pan, B. Wang, S. Liang, J. J. Yang, P. Wang, and F. Miao, Robust memristors based on layered two-dimensional materials, arXiv:1801.00530.
43 R. Ge, X. Wu, M. Kim, J. Shi, S. Sonde, L. Tao, Y. Zhang, J. C. Lee, and D. Akinwande, Atomristor: Nonvolatile resistance switching in atomic sheets of transition metal dichalcogenides, Nano Lett. 18, 434?441 (2018).

E. 2D Materials for Sensors

Background

2D materials offer much larger surface-to-volume ratios than bulk materials. Therefore, they can intuitively be expected to be perfectly suited for senor devices with high-sensitivity, and this is certainly part of the motivation for the intensive theoretical and experimental work on developing sensors from 2D materials. A second reason that 2D materials are particularly promising for gas sensors is the fact that the very popular metal oxide gas sensors show sufficiently high sensitivity only at elevated temperatures while 2D gas sensors show high sensitivity already at room temperature [1].

Difficult Challenges

Table I. Challenges for 2D gas sensors.

Challenge	Comment
Elaboration of a list of relevant performance parameters for 2D gas sensors (targets) and specification of numbers or ranges for the targets needed for certain applications	Experiences from conventional gas sensors will serve as a valuable guideline
Lowering the detection limit, improvement of response and recovery times	
Improvement of sensitivity and selectivity	

Current Status, Prospects, and Challenges

A variety of 2D materials are under investigation (theoretical studies and experiments) for application in different types of sensors, and research in this area is extremely intensive, e.g., [1-10]. So far, most work on 2D sensors has been focused on graphene and MoS_2 gas sensors. Therefore, and since gas sensors represent an important branch of sensor technology, the following discussion is limited to graphene and MoS_2 gas sensors.

A lot of theoretical work, most notably ab initio calculations, has been done to study the adsorption of different gases on 2D materials such as graphene [11-14] and MoS_2 [15-18]. Moreover, a great many experimental studies on gas adsorption on 2D materials has been performed and a wealth of experimental data has been published. On the other hand, a comprehensive comparison of the performance of 2D gas sensors and conventional gas sensors in terms of static sensor figures of merit such as detection limit, sensitivity, and selectivity as well as dynamic figures of merit such as response and recovery times is still missing. This, however, is urgently needed to seriously assess the merits and drawbacks of 2D gas sensors compared to conventional gas sensors. Also, none of these materials are inherently selective to any gas or vapor, so selective discrimination must be enabled by other means such as artificial intelligence or machine learning. Tables II and III summarize the experimental efforts and show the types of gases whose absorption has been studied.

Table II. Overview of experiments on graphene gas sensors. The X's indicate which gases have been sensed.

Gas / Ref.	[19]	[20]	[21]	[22]	[23]	[24	[25]	[26]	[27]	[28]
NO	X									
NO_2	X		X	X	X					
NH_3	X		X	X	X		X			
N_2O	X									X
O_2	X								X	
SO_2	X									
CO_2	X					X				
H_2O	X	X			X					
$C_9H_{18}O$		X								
$C_8H_{16}O_2$		X								
$C_6H_{15}N$		X								
$C_7H_6N_2O_4$			X							
CO					X					
CH_3OH								X		
C_2H_5OH								X		
CH_3CN								X		
C_4H_8O								X		

Table III. Overview of experiments on MoS_2 gas sensors. The X's indicate which gases have been sensed.

Gas / Ref.	[29]	[30]	[31]	[32]	[15]	[33,34]	[35]	[36]	[37]
NO	X								X
NH_3	X				X				X
NO_2		X		X	X				
$C_6H_{15}N$		X					X		
O_2			X					X	X
CO_2			X						
CH_3CN						X			
C_2H_5OH		X				X		X	
CH_3OH		X				X			
C_6H_5-CH_3		X				X			
$CHCl_3$		X				X			
C_3H_6O							X		
H_2O								X	
H_2									X

References

1 S. Yang, C. Jiang, and S.-H. Wei, Gas sensing in 2D materials, Appl. Phys. Rev. 4, 021304 (2017).

2 S. K. Tuteja, T. Duffield, and S. Neethirajan, Suresh, Graphene-based multiplexed disposable electrochemical biosensor for rapid on-farm monitoring of NEFA and beta HBA dairy biomarkers, J. Mater. Chem. B 5, 6930-6940 (2017).

3 J.E. Kim, Y. H No, J. N. Kim, Y. S. Shin, W. T. Kang, Y. R. Kim, K. N. Kim, Y. H. Kim, and W. J. Yu, Highly sensitive graphene biosensor by monomolecular self-assembly of receptors on graphene surface, Appl. Phys. Lett. 110, 203702 (2017).

4 S. Cinti and F. Arduini, Graphene-based screen-printed electrochemical (bio)sensors and their applications: Efforts and criticisms, Biosens. Bioelectron. 89, 107-122 (2017).

5 P. Salvo, B. Melai, N. Calisi, C. Paoletti, F. Bellagambi, A. Kirchhain, M. G. Trivella, R. Fuoco, and F. Di Francesco, Graphene-based devices for measuring pH, Sens. Actuator B Chem. 256, 976-991 (2018).
6 D. Zappa, Molybdenum dichalcogenides for environmental chemical sensing, Materials 10, 1418 (2017).
7 S. Mao, J. B. Chang, H. H. Pu, G. H. Lu, Q. Y. He, H. Zhang, and J. H. Chen, Two-dimensional nanomaterial-based field-effect transistors for chemical and biological sensing, Chem. Soc. Rev. 46, 6872-6904 (2017).
8 V. Shukla, J. Warna, N. K. Jena, A. Grigoriev, and R. Ahuja, Toward the realization of 2D borophene based gas sensor, ?J. Phys. Chem. C 121, 26869-26876 (2017).
9 M. Pumera, Phosphorene and black phosphorus for sensing and biosensing, Trends Analyt. Chem. 93, 1-6 (2017).
10 C. Zhu, D. Du, and Y. Lin, Graphene-like 2D nanomaterial-based biointerfaces for biosensing applications, Biosens. Bioelectron. 89, 43-55 (2017).
11 O. Leenaerts, B. Partoens, and F. M. Peeters, Adsorption of H2O, NH3, CO, NO2, and NO on graphene: A first-principles study, Phys. Rev. B 77, 125416 (2008).
12 X. Lin, J. Ni, and C. Fang, Adsorption capacity of H2O, NH3, CO, and NO2 on the pristine graphene, J. Appl. Phys. 113, 034306 (2013).
13 W. Wang, Y. Zhang, C. Shen, and Y. Chai, Adsorption of CO molecules on doped graphene: A first-principles study, AIP Adv. 6, 025317 (2016).
14 L. Maa, J.-M. Zhanga, K.-W. Xu, and V. Jida, A first-principles study on gas sensing properties of graphene and Pd-doped graphene, Appl. Surf. Sci. 343, 121-127 (2015).
15 D. J. Late, Y.-K. Huang, B. Liu, J. Acharya, S. N. Shirodkar, J. Luo, A. Yan, D. Charles, U. V. Waghmare, V. P. Dravid, and C. N. R. Rao, Sensing behavior of atomically thin-Layered MoS2 Transistors, ACS Nano 7, 4879–4891 (2013).
16 N. Yu, L. Wang, M. Li, X. Sun, T. Hou, and Y. Li, Molybdenum disulfide as a highly efficient adsorbent for non-polar gases, Phys. Chem. Chem. Phys. 17, 11700-11704 (2015).
17 M. P. K Sahoo, J. Wang, Y. Zhang, T. Shimada, and T. Kitamura, Modulation of gas adsorption and magnetic properties of monolayer-MoS2 by antisite defect and strain, J. Phys. Chem. C 120, 14113?14121 (2016).
18 H. Li, M. Huang, and G. Cao, Markedly different adsorption behaviors of gas molecules on defective monolayer MoS2: A first-principles study, Phys. Chem. Chem. Phys. 18, 15110 (2016).
19 G. Chen, T. M. Paronyan, and A. R. Harutyunyan, Sub-ppt gas detection with pristine graphene, Appl. Phys. Lett. 101, 053119 (2012).
20 Y. Dan, Y. Lu, N. J. Kybert, Z. Luo, and A. T. C. Johnson, Intrinsic response of graphene vapor sensors, Nano Lett. 9, 1472-1475 (2009).
21 J. D. Fowler, M. J. Allen, V. C. Tung, Y. Yang, R. B. Kaner, and B. H. Weiller, Practical chemical sensors from chemically derived graphene, ACS Nano 3, 301–306 (2009).
22 N. Harada, K. Hayashi, M. Kataoka, J. Yamaguchi, M. Ohtomo, M. Ohfuchi, I. Soga, D. Kondo, T. Iwai, and S. Sato, Graphene-gate transistors for gas sensing and threshold control, Tech. Dig. IEDM, 476-479 (2016).
23 F. Schedin, A. K. Geim, S. V. Morozov, E. W. Hill, P. Blake, M. I. Katsnelson, and K. S. Novoselov, Detection of individual gas molecules adsorbed on graphene, Nat. Mater. 6, 652-655 (2007).
24 H. J. Yoon, D. H. Jun, J. H. Yang, Z. Zhou, S. S. Yang, and M. M.-C. Cheng, Carbon dioxide gas sensor using a graphene sheet, Sens. Actuators B 157, 310?313 (2011).
25 E. R. Hugo, J. Prasoon, K. G. Awnish, R. G. Humberto, W. C. Milton, A. T. Srinivas, and C. E. Peter, Adsorption of ammonia on graphene. Nanotechnol. 20, 245501 (2009).
26 S. Rumyantsev, G. Liu, M. S. Shur, R. A. Potyrailo, and A. A. Balandin, Selective gas sensing with a single pristine graphene transistor, Nano Lett. 12, 2294?2298 (2012).
27 G. Chen, T. M. Paronyan, and A. R. Harutyunyan, Sub-ppt gas detection with pristine graphene, Appl. Phys. Lett. 101, 243502 (2012).
28 R. K. Paul, S. Badhulika, N. M. Saucedo, and A. Mulchandani, Graphene nanomesh as highly sensitive chemiresistor gas sensor, Anal. Chem. 84, 8171?8178 (2012).
29 D. Burman, R. Ghosh, S. Santra, S. K. Ray, and P. K. Guha, Role of vacancy sites and UV-ozone treatment on few layered MoS2 nanoflakes for toxic gas detection, Nanotechnol. 28 435502 (2017).
30 S.-Y. Cho, S. J. Kim, Y. Lee, J.-S. Kim, W.-B. Jung, H.-W. Yoo, J. Kim, and H.-T. Jung, Highly enhanced gas adsorption properties in vertically aligned MoS2 layers, ACS Nano 9, 9314–9321 (2015).
31 H. ul Hassan, J. Mun, B. S. Kang, J. Y. Song, T. Kima, and S.-W. Kang, Sensor based on chemical vapour deposition grown molybdenum disulphide for gas sensing application, RSC Adv. 6, 75839-75843 (2016).
32 R. Kumar, N. Goel, and M. Kumar, UV-activated MoS2 based fast and reversible NO2 sensor at room temperature, ACS Sens. 2, 1744-1752 (2017).
33 G. Liu, S. L. Rumyantsev, C. Jiang, M. S. Shur, and A. A. Balandin, Selective gas sensing with h-BN capped MoS2 heterostructure thin-film transistors, IEEE Electron Device Lett. 36, 1202-1204 (2015).
34 F. K. Perkins, A. L. Friedman, E. Cobas, P. M. Campbell, G. G. Jernigan, and B. T. Jonker, Chemical vapor sensing with monolayer MoS2, Nano Lett. 13, 668?673 (2013).
35 R. Samnakay, C. Jiang, S. L. Rumyantsev, M. S. Shur, and A. A. Balandin, Selective chemical vapor sensing with few-layer MoS2 thin-film transistors: Comparison with graphene devices, Appl. Phys. Lett. 106, 023115 (2015).

36 Y. Tong, Y. Liu, Y. Zhao, J. Thong, D. S. H. Chan, and C. Zhu, Selectivity of MoS2 gas sensors based on a time constant spectrum method, Sensors and Actuators A 255, 28–33 (2017).
37 A. Venkatesan, S. Rathi, I.-Y. Lee, J. Park, D. Lim, M. Kang, H.-I. Joh, G.-H. Kim, and E. S. Kannan, Molybdenum disulfide nanoparticles decorated reduced graphene oxide: Highly sensitive and selective hydrogen sensor, Nanotechnology 28 365501 (2017).
38 E. Singh et al, Flexible garphene-based wearable gas and chemical sensors, ACS Appl. Mater. Interfaces 9, 34544-34586 (2017).

6. Carbon Nanotube Electronics

In the past decade, tremendous progress has been made in solving various challenges in carbon nanotube (CNT) field-effect transistors (CNTFETs). There have been significant demonstrations and analysis of complete digital systems using CNTFETs, indicating that CNTs can bring speed and energy-efficiency benefits that no other device technology can offer.

A. Performance Projection

a. Device-Level Benefits

To quantify key device-level CNTFET benefits that enable energy-efficient digital very-large-scale integration (VLSI) logic circuits, a useful metric is the electrostatic scale length (λ), which quantifies how susceptible a FET is to short-channel effects [1]; it should be small to enable shorter L_G (thus improving gate capacitance: C_G) without degrading sub-threshold slope (SS). Two approaches for reducing λ are: 1) improve FET geometry (e.g., from top-gate to gate-all-around (GAA)), and 2) reduce the semiconductor body thickness (T_{BODY}). While evolving from planar Si FET to 3D FinFET to GAA nanowire FET (NWFET) reduces λ [2], continued significant λ reduction requires reducing T_{BODY}. Unfortunately, for bulk materials (e.g., all Si-, Ge-, and III-V-based semiconductors), carrier transport severely degrades as T_{BODY} scales to sub-10 nm dimensions [3]–[8] as surface roughness scattering lowers FET effective drive current (I_{EFF}).

Here is the key advantage of CNTFET: CNTs inherently maintain high mobility and carrier velocity even at very thin (~1-2 nm) T_{BODY} (experimental mobility: >2,500 cm^2/V-s [Zhou 2005], and injection velocity = 4.1×10^7 cm/s [Lee 2015a], for DCNT < 2 nm). In contrast, experimental Si-based FinFET demonstrations with T_{BODY} < 3 nm exhibit mobility < 300 cm^2/V-s [3]-[8]. This leads to major energy efficiency benefits for CNTFETs. At the 7-nm node technology, CNFETs are projected to offer 9.0× EDP benefits compared to Si/SiGe FinFETs for the same I_{OFF} (100 nA/µm) and power density (~65 W/cm^2). Si/SiGe NWFETs offer <30% EDP benefits compared to Si/SiGe FinFETs [9].

Similar to 1D CNTFETs, transistors with 2D layered materials have low parasitic capacitances and thin T_{BODY} for short gate length. Yet, the difference in benefits of 2D FET vs. CNFET originates from the difference in carrier transport. As an example, comparing 2D FETs vs. Si FETs, black phosphorous (BP) FETs can provide 2.2× EDP benefits, and molybdenum disulfide (MoS_2) FETs can provide 1.7× EDP benefits compared to Si FinFETs at 0.6 V, due to the lower middle-end-of-line (MEOL) parasitic capacitances, where the BSIM-CMG (common multigate) model [10] is used to model the FETs with carrier mobility and velocity extracted from experimental and simulation data [11].

b. System-Level Benefits

CNTFETs present a unique opportunity to revolutionize computing and achieve the next 1,000X in energy efficiency [12]. Specifically, they enable monolithic 3D nanosystems, with multiple layers of computation and memory densely integrated over the same starting substrate, truly embodying computation immersed in memory. The key is that CNTFETs can be fabricated at very low processing temperatures (< 250°C) [13], and so multiple layers of CNTFET circuits, along with layers of emerging memory technologies such as Resistive RAM (RRAM) [14] or Spin-Transfer Torque magnetic RAM (STT-MRAM) [15] which can also be fabricated at low processing temperatures, can be vertically interleaved and densely connected with conventional back-end-of-line inter-layer vias. In stark contrast, such monolithic 3D systems are very challenging to fabricate using today's silicon-based technologies, as high-performance silicon FETs can require processing temperatures greater than 1,000°C, thus damaging or destroying FETs and wires on the bottom layers. Furthermore, monolithic 3D nanosystem prototypes have already been experimentally demonstrated, including a 4-layer monolithic 3D system integrating >2 million CNTFETs and 1 megabit of RRAM, all fabricated directly on top of >1 million silicon FETs [16]; by combining the benefits of new nanotechnologies, such as CNFETs, along with the new 3D architectures that they naturally enable, digital system energy efficiency can be improved by 1,000× compared to computing systems today [12].

B. Progress in Device-Level Technology Development

a. Wafer-Scale-Aligned CNT Growth and Transfer

Wafer-scale growth of aligned CNTs with Fe catalyst using chemical vapor deposition has been demonstrated to enable arbitrary logic definition without any layout customization. The degree of alignment is >99%. The average length of the CNTs is ~400μm with an average density of 5~10 CNTs/μm [17].

For CNTFETs to show the promised benefits of high performance and energy efficiency, the speed (corresponding to current-drive per unit layout width, I_{ON}) of CNFET technology must surpass that of silicon-based digital systems. Improving upon the transfer technique developed in 2009 [18], density of 100 CNTs/μm has been achieved by using a multiple transfer technique, called Controlled IDC Density Enhancement by Repeated transfers (CIDER). The transfer process can be applied to any target substrate for any arbitrary CNT density (target density is ~200 – 250 CNTs/μm) while maintaining the alignment of CNTs. CNFETs with the highest current drive (per unit layout width) of 100 mA/μm have been demonstrated using CIDER [19].

b. Purification of CNTs

Numerous techniques have been developed to sort the semiconducting CNTs (s-CNTs) from the metallic CNTs (m-CNTs), as the presence of m-CNTs results in increased off-state leakage currents (IOFF), leading to decrease in noise margin and incorrect logic functionality. It is important to note that the band gap of the CNTs is mainly determined by the CNT diameter rather than the chirality itself, as multiple chirality can map to the same diameter that are semiconducting [20], [21]. Therefore, precise chirality control is not required, but rather the CNT diameter and the removal of m-CNTs need to be controlled for the purification of CNTs.

After the wafer-scale-aligned CNT growth and transfer technique as described in section 3a, an m-CNT removal technique is electrically implemented which demonstrates high selectivity (≥ 99.99% of m-CNT removal) and high scalability (any arbitrary CPP and high CNT density ≥ 200 CNTs/μm) [22]. Though it has been challenging to simultaneously achieve high selectivity and scalability with previous m-CNT removal techniques [18], [23]–[25], the various techniques can be combined to achieve further selectivity and scalability.

An alternative approach to create pure s-CNTs is solution-based purification. Several successful approaches have been demonstrated with >99% semiconducting purity including polymer-based techniques [26] and column chromatography [27]. The challenge of re-assembling these solution-dispersed semiconducting CNTs with alignment and density on the wafer surface has been overcome through techniques such as floating evaporative self-assembly [28]. Other advances in CNT growth include growing enhanced purity using carbon source engineering [29], achieving >100 CNTs/μm in a single growth step [30], and >95% single-chirality CNTs with uniform diameters using crystalline catalysts [31].

c. Controlled CNT Pitch

The largest source of variation in a highly-scaled CNT logic technology is the imprecise pitch between adjacent CNTs that varies the number of CNTs in minimum-width transistors [32]. One approach that achieves controlled CNT pitch uses patterned selective surface treatments that align a single-CNT per patterned feature [33]. Though the density of such demonstrations is not yet sufficient for scaled technology nodes, there has been significant progress made on increasing the CNT density. Aligned CVD growth achieves a pitch variability (σ^2/μ^2) of approximately 0.5 while the target is below 0.25.

d. CNT-NFET and -PFET

Fabrication of CNT-NFETs has been challenging, as CNTs are inherently p-type due to doping of oxygen present in air. The robustness of NFET fabrication using low-work-function metals as source and drain metal contacts has been studied, in which CNT-NFETs with erbium contacts showed statistically comparable performance to CNT-PFETs with palladium contacts [34]. In addition, CNT-NFET and CNT-PFET of 10-nm gate length show symmetric performance, both with SS of 70 mV/decade [35].

e. Short-Channel CNTFETs

While Dennard scaling is becoming increasingly challenging in Si-based transistors, there have been various studies showing the potential of continued scaling in CNFETs without suffering short-channel effects: gate length of 5 nm with subthreshold slope of 73 mV/decade [35], 14 nm node combined with the m-CNT removal technique, achieving ION/IOFF > 10^4 [22], and 32 nm channel length of CNFETs for demonstration of integrated infra-red sensor [36].

f. Reduction of Hysteresis

Hysteresis in CNTFETs originates from the traps surrounding an isolated 1D channel [37], and there have been efforts to reduce the density of interfacial traps near the CNTs to eliminate hysteresis. However, interface engineering

alone may not be sufficient to eliminate hysteresis completely. An effective method is to reduce the influence of such traps by simply scaling the effective oxide thickness (EOT) for better electrostatic control from the gate. Hysteresis of less than 0.5% of the gate-source voltage range has been demonstrated with 1.6 nm of EOT [38].

C. Progress in System-Level Studies

a. CNT Computer

Through a combination of design and processing techniques, known as the imperfection-immune paradigm [39], several system-level demonstrations have already been accomplished. First, it enabled the first CNTFET computer, built entirely using CNTFETs [40]. This CNTFET computer demonstrated the ability to run programs, run a basic operating system that performs multitasking, and to execute MIPS instructions, and was the first system demonstration using entirely CNTFETs.

b. Nanosystem

Large-scale monolithic 3D nanosystems have also been demonstrated using a combination of CNTFETs, Resistive RAM (RRAM), and Silicon CMOS. A 3D nanosystem consisting of over 2 million CNTFETs and 1 megabit of RRAM has been demonstrated [16]. By combining low-temperature processing techniques, critical for monolithic 3D integration, RRAM arrays, silicon and CNTFET computation units, memory access circuitry, and CNTFET gas sensors were integrated into a single system. This nanosystem is capable of capturing massive amounts of data directly into on-chip memory, performing in-situ processing of the captured data, and producing highly-processed information.

c. Hyperdimensional Computing

The most recent large-scale system demonstration consists of a brain-inspired, error-resilient computation model suitable for cognitive tasks realized using monolithic 3D integration of CNTFETs and RRAM [41]. This 3D nanosystem consists of CNTFETs integrated directly on top of RRAM, while exploiting the inherent properties of each of the nanotechnologies to perform language classification. This hardware prototype is able to perform continuous online learning and inference with high accuracy.

d. Ring Oscillator

Two teams have recently published demonstrations of CNTFET-based ring oscillator speed, with IBM achieving 355 ps stage delay for $L_{channel}$ = 100 nm at VDD = 1.9 V [42] and Peking University achieving 18 ps stage delay for L_{gate} = 115 nm at VDD ≈ 2.8 V [43]. The results are highly encouraging, as the stage delays of this CNTFET ring oscillator are comparable to the delay (~11 ps) of commercial Si CMOS ICs at L_{gate} = 130 nm at VDD = 1.5 V. Both groups claim that the remaining challenge is to integrate techniques for high CNT density with controlled pitch to further improve the speed.

D. Outstanding Challenges

a. Threshold Voltage Variations

One important challenge that needs to be solved before realizing the potential benefits of CNTFETs is the imprecise threshold voltage (V_T) control, which causes device-to-device variability [44]. The large variations in V_T result in the variations of the off-state leakage current and increase of the standby power consumption. The circuit integrity and performance are also compromised, as a poor control in V_T leads to variations in on-state current as well. Overall, without a precise control of V_T, CNTFETs cannot achieve a desirable uniformity, yield, and performance at the circuit level. The origin of V_T variation is likely due to residual traps and other instabilities from inadequate process control in academic fabrication facilities rather than something fundamental to the 1D channel material itself.

b. Contact Resistance

Contact resistance degrades the drive current in all scaled transistors with small contacts, and CNT transistors are no exception. The best contact resistances achieved for both P-type [45] and N-type [35] CNTFETs for long contacts approach 5 kΩ and 7.5 kΩ per CNT, respectively. In a multi-CNT CNTFET with 150 CNT/µm, this translates to a contact resistivity of ≈ 30-50 Ω-µm. However, as the contact dimension shrinks below 50 nm, the contact resistance increases proportionally to $1/L_{contact}$. Recently, researchers at IBM have demonstrated a novel end-bonded contact geometry that features contact-length invariation contact resistance for low resistance scaled contacts to CNT. This is a very important advance that enables highly scaled CNT contacts with contact resistance values as low as 18 kΩ per CNT for contacts down to 9 nm [46]. The end-contact is capable of integration with complementary CNTFETs [47], and dense CNT arrays for high performance scaled CNFETs [48].

c. Control of Doping Strength

A doping mechanism of CNT relies on charge transfer at the interface between the surrounding dielectric and the CNT (charge transfer doping) [49]. Although various doping techniques have been explored [50]–[53], a method to control or quantify the doping strength has not yet been established. Stability of doping over time and at elevated temperatures has yet to be demonstrated.

d. Low-Parasitic Device Structure

Various CNTFET structures have been used for device-level studies, such as global and local back-gate, top-gate, and gate-all-around structures. However, to realize the expected benefits at the system level, it is important to reduce the effects of parasitics for higher speed. For instance, a device structure to lower the parasitic capacitance is to minimize the overlap between the gate and the source (or drain) metal contacts [32].

e. High-Performance Ring Oscillator

A key performance benchmark for a logic technology is ring oscillator speed and energy efficiency. As such, a frontier of future CNT research is to integrate the recent advances into functional demonstrations of circuit performance and efficiency using a ring oscillator. The demonstrations of ring oscillator speed [42], [43] have yet to achieve rail-to-rail oscillation and demonstrate energy efficiency (and low voltage operation). The eventual goal of these efforts will be to demonstrate superior performance at lower supply voltage than the state-of-the-art Si CMOS, and realize the promise of energy-efficient CNT-based electronic systems.

References

[1] D. J. Frank, Y. Taur, and H. S. P. Wong, “Generalized scale length for two-dimensional effects in MOSFET’s,” IEEE Electron Device Lett., vol. 19, no. 10, pp. 385–387, 1998.

[2] K. J. Kuhn, “Considerations for ultimate CMOS scaling,” IEEE Transactions on Electron Devices, vol. 59, no. 7. pp. 1813–1828, 2012.

[3] K. Uchida, J. K. J. Koga, and S. T. S. Takagi, “Experimental study on carrier transport mechanisms in double- and single-gate ultrathin-body MOSFETs - Coulomb scattering, volume inversion, and δTSOI-induced scattering,” IEEE Int. Electron Devices Meet. 2003, vol. I, no. c, pp. 1–4, 2003.

[4] L. Gomez, I. Aberg, and J. L. Hoyt, “Electron transport in strained-silicon directly on insulator ultrathin-body n-MOSFETs with body thickness ranging from 2 to 25 nm,” IEEE Electron Device Lett., vol. 28, no. 4, pp. 285–287, 2007.

[5] D. S. Sung et al., “Investigation of nanowire size dependency on TSNWFET,” in Technical Digest - International Electron Devices Meeting, IEDM, 2007, pp. 891–894.

[6] P. Hashemi et al., “Strained Si1-xGex-on-insulator PMOS FinFETs with excellent sub-threshold leakage, extremely-high short-channel performance and source injection velocity for 10nm node and beyond,” in Digest of Technical Papers - Symposium on VLSI Technology, 2014.

[7] P. Hashemi et al., “First demonstration of high-Ge-content strained-Si1-xGex (x=0.5) on insulator PMOS FinFETs with high hole mobility and aggressively scaled fin dimensions and gate lengths for high-performance applications,” in Technical Digest - International Electron Devices Meeting, IEDM, 2015, vol. 2015–Febru, no. February, p. 16.1.1-16.1.4.

[8] Y. Sasaki et al., “Novel junction design for NMOS Si Bulk-FinFETs with extension doping by PEALD phosphorus doped silicate glass,” in Technical Digest - International Electron Devices Meeting, IEDM, 2015, vol. 2016–Febru, p. 21.8.1-21.8.4.

[9] G. Hills, “Energy-efficient Digital VLSI using Carbon Nanotube Field-Effect Transistors,” Stanford University, 2018.

[10] N. Paydavosi et al., “BSIM - SPICE models enable FinFET and UTB IC designs,” IEEE Access, vol. 1, pp. 201–215, 2013.

[11] C. S. Lee, B. Cline, S. Sinha, G. Yeric, and H. S. P. Wong, “32-bit Processor core at 5-nm technology: Analysis of transistor and interconnect impact on VLSI system performance,” in Technical Digest - International Electron Devices Meeting, IEDM, 2017, p. 28.3.1-28.3.4.

[12] M. M. S. Aly et al., “Energy-efficient abundant-data computing: The N3XT 1,000,” Computer (Long. Beach. Calif)., vol. 48, no. 12, pp. 24–33, 2015.

[13] H. Wei, M. Shulaker, H. S. P. Wong, and S. Mitra, “Monolithic three-dimensional integration of carbon nanotube FET complementary logic circuits,” in Technical Digest - International Electron Devices Meeting, IEDM, 2013.

[14] M. J. Lee et al., “A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta2O5-x/TaO2-xbilayer structures,” Nat. Mater., vol. 10, no. 8, pp. 625–630, 2011.

[15] H. Noguchi et al., “7.2 4Mb STT-MRAM-Based Cache with Memory- Access-Aware Power Optimization and Write-Verify-Write / Read-Modify-Write Scheme,” IEEE Int. Solid-State Circuits Conf., pp. 132–134, 2016.

[16] M. M. Shulaker et al., “Three-dimensional integration of nanotechnologies for computing and data storage on a single chip,” Nature, vol. 547, no. 7661, pp. 74–78, 2017.

[17] N. Patil et al., “Wafer-scale growth and transfer of aligned single-walled carbon nanotubes,” IEEE Trans. Nanotechnol., vol. 8, no. 4, pp. 498–504, 2009.

[18] N. Patil et al., “VMR: VLSI-compatible metallic carbon nanotube removal for imperfection-immune cascaded multi-stage digital logic circuits using carbon nanotube FETs,” in Technical Digest - International Electron Devices Meeting, IEDM, 2009.
[19] M. M. Shulaker, G. Pitner, G. Hills, M. Giachino, H.-S. P. Wong, and S. Mitra, “High-performance carbon nanotube field-effect transistors,” 2014 IEEE Int. Electron Devices Meet., vol. 4, no. 5, p. 33.6.1-33.6.4, 2014.
[20] H.-S. P. Wong and D. Akinwande, Carbon nanotube and graphene device physics, vol. 9780521519. 2010.
[21] Y. Matsuda, J. Tahir-kheli, W. A. Goddard, W. A. G. Iii, and W. A. Goddard, “Definitive Band Gaps for Single-Wall Carbon Nanotubes,” J. Phys. Chem. Lett., vol. 1, no. 19, pp. 2946–2950, 2010.
[22] M. M. Shulaker, G. Hills, T. F. Wu, Z. Bao, H. S. P. Wong, and S. Mitra, “Efficient metallic carbon nanotube removal for highly-scaled technologies,” Tech. Dig. - Int. Electron Devices Meet. IEDM, vol. 2016–Febru, p. 32.4.1-32.4.4, 2016.
[23] Q. Cao, S. J. Han, G. S. Tulevski, Y. Zhu, D. D. Lu, and W. Haensch, “Arrays of single-walled carbon nanotubes with full surface coverage for high-performance electronics,” Nat. Nanotechnol., vol. 8, no. 3, pp. 180–186, 2013.
[24] G. Zhang et al., “Selective etching of metallic carbon nanotubes by gas-phase reaction,” Science (80-.)., vol. 314, no. 5801, pp. 974–977, 2006.
[25] S. H. Jin et al., “Using nanoscale thermocapillary flows to create arrays of purely semiconducting single-walled carbon nanotubes,” Nat. Nanotechnol., vol. 8, no. 5, pp. 347–355, 2013.
[26] T. Lei, X. Chen, G. Pitner, H. S. P. Wong, and Z. Bao, “Removable and Recyclable Conjugated Polymers for Highly Selective and High-Yield Dispersion and Release of Low-Cost Carbon Nanotubes,” J. Am. Chem. Soc., vol. 138, no. 3, pp. 802–805, 2016.
[27] G. S. Tulevski, A. D. Franklin, and A. Afzali, “High purity isolation and quantification of semiconducting carbon nanotubes via column chromatography,” ACS Nano, vol. 7, no. 4, pp. 2971–2976, 2013.
[28] Y. Joo, G. J. Brady, M. S. Arnold, and P. Gopalan, “Dose-controlled, floating evaporative self-Assembly and alignment of semiconducting carbon nanotubes from organic solvents,” Langmuir, vol. 30, no. 12, pp. 3460–3466, 2014.
[29] J. Parker, C. Beasley, A. Lin, H. Y. Chen, and H. S. Philip Wong, “Increasing the semiconducting fraction in ensembles of single-walled carbon nanotubes,” Carbon., vol. 50, no. 14, pp. 5093–5098, 2012.
[30] Y. Hu et al., “Growth of high-density horizontally aligned SWNT arrays using Trojan catalysts,” Nat. Commun., vol. 6, 2015.
[31] F. Yang et al., “Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts,” Nature, vol. 510, no. 7506, pp. 522–524, 2014.
[32] G. Hills et al., “Rapid exploration of processing and design guidelines to overcome carbon nanotube variations,” in Proceedings of the 50th Annual Design Automation Conference on - DAC ’13, 2013, p. 1.
[33] H. Park et al., “High-density integration of carbon nanotubes via chemical self-assembly,” Nat. Nanotechnol., vol. 7, no. 12, pp. 787–791, 2012.
[34] D. Shahrjerdi et al., “High device yield carbon nanotube NFETs for high-performance logic applications,” Tech. Dig. - Int. Electron Devices Meet. IEDM, pp. 509–512, 2011.
[35] C. Qiu, Z. Zhang, M. Xiao, Y. Yang, D. Zhong, and L. M. Peng, “Scaling carbon nanotube complementary transistors to 5-nm gate lengths,” Science (80-.)., vol. 355, no. 6322, pp. 271–276, 2017.
[36] M. M. Shulaker et al., “Carbon nanotube circuit integration up to sub-20 nm channel lengths,” ACS Nano, vol. 8, no. 4, pp. 3434–3443, 2014.
[37] R. S. Park et al., “Hysteresis in Carbon Nanotube Transistors: Measurement and Analysis of Trap Density, Energy Level, and Spatial Distribution,” ACS Nano, vol. 10, no. 4, pp. 4599–4608, 2016.
[38] R. S. Park, G. Hills, J. Sohn, S. Mitra, M. M. Shulaker, and H. S. P. Wong, “Hysteresis-Free Carbon Nanotube Field-Effect Transistors,” ACS Nano, vol. 11, no. 5, pp. 4785–4791, 2017.
[39] J. Zhang et al., “Carbon Nanotube Robust Digital VLSI,” IEEE Trans. Comput. Des. Integr. Circuits Syst., vol. 31, no. 4, pp. 453–471, 2012.
[40] M. M. Shulaker et al., “Carbon nanotube computer,” Nature, vol. 501, no. 7468, pp. 526–530, 2013.
[41] T. F. Wu et al., “Brain-Inspired Computing Exploiting Carbon Nanotube FETs and Resistive RAM: Hyperdimensional,” IEEE Int. Solid-State Circuits Conf., pp. 492–494, 2018.
[42] S. J. Han et al., “High-speed logic integrated circuits with solution-processed self-assembled carbon nanotubes,” Nat. Nanotechnol., vol. 12, no. 9, pp. 861–865, 2017.
[43] D. Zhong et al., “Gigahertz integrated circuits based on carbon nanotube films,” Nat. Electron., vol. 1, pp. 40–45, 2017.
[44] Q. Cao et al., “Origins and Characteristics of the Threshold Voltage Variability of Quasiballistic Single-Walled Carbon Nanotube Field-E ff ect Transistors,” no. 2, pp. 1936–1944, 2015.
[45] A. D. Franklin and W. Haensch, “Defining and overcoming the contact resistance challenge in scaled carbon nanotube transistors,” in Device Research Conference - Conference Digest, DRC, 2014, pp. 191–192.
[46] Q. Cao et al., “End-bonded contacts for carbon nanotube transistors with low, size-independent resistance,” Science (80-.)., vol. 350, no. 6256, pp. 68–72, 2015.
[47] J. Tang, Q. Cao, D. B. Farmer, G. Tulevski, and S. J. Han, “Carbon nanotube complementary logic with low-temperature processed end-bonded metal contacts,” in Technical Digest - International Electron Devices Meeting, IEDM, 2017, p. 5.1.1-5.1.4.

[48] Q. Cao, J. Tersoff, D. B. Farmer, Y. Zhu, and S. J. Han, “Carbon nanotube transistors scaled to a 40-nanometer footprint,” Science (80-.)., vol. 356, no. 6345, pp. 1369–1372, 2017.
[49] L. Suriyasena Liyanage, X. Xu, G. Pitner, Z. Bao, and H. S. P. Wong, “VLSI-compatible carbon nanotube doping technique with low work-function metal oxides,” Nano Lett., vol. 14, no. 4, pp. 1884–1890, 2014.
[50] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, “Controlling doping and carrier injection in carbon nanotube transistors,” Appl. Phys. Lett., vol. 80, no. 15, pp. 2773–2775, 2002.
[51] A. Javey, R. Tu, D. B. Farmer, J. Guo, R. G. Gordon, and H. Dai, “High performance n-type carbon nanotube field-effect transistors with chemically doped contacts,” Nano Lett., vol. 5, no. 2, pp. 345–348, 2005.
[52] A. D. Franklin et al., “Carbon nanotube complementary wrap-gate transistors,” Nano Lett., vol. 13, no. 6, pp. 2490–2495, 2013.
[53] F. Xia et al., “Surface Charge Transfer Doping via Transition Metal Oxides for Efficient p-Type Doping of II-VI Nanostructures,” ACS Nano, vol. 10, no. 11, pp. 10283–10293, 2016.

7. Plasmonic Devices

Developing large-bandwidth technologies that are low-cost, energy-efficient, and compact is an essential and yet challenging requirement in scaling the data-driven IT infrastructure from microprocessors to data centers and telecommunication networks. In conventional electronics, the power dissipation in microchips increases super-linearly as clock frequency increases. This trade-off between bandwidth and energy efficiency has severely limited the development of modern computing systems. In addition, the communication bottleneck between the electronic logic and memory devices is another limiting factor in improving the computing performance in terms of bandwidth and energy.

Photonics technology potentially offers a promising solution to this problem, owing to its inherent advantages in large bandwidth, low transportation losses, and small (yet predictable) latency: an optical signal, once generated, propagates by itself without the need to apply a voltage. Silicon-based nano/microphotonic devices and circuits have been extensively studied for integration on CMOS chips. More recently plasmonics has emerged as a new contender with its unique capability of confining light into deep-subwavelength scale dimensions. Plasmonic circuits can deliver both optical and electrical signals through the same metal-circuitry. The surface-plasmon fields at a metal/dielectric interface mostly distribute to the dielectric side. In a metal/dielectric/metal (MIM) structure (so-called gap-plasmon waveguide), a strong electric field can be induced in the dielectric layer by applying low-voltage signals across the metal layers, while plasmon fields remain narrowly confined in the same dielectric layer carrying optical signals. This MIM structure, with incorporation of proper transduction material in the dielectric layer part, enables ultracompact, high speed, energy-efficient modulation of plasmonic waves.

Despite these promising aspects, the efforts to develop plasmonic circuits have been hampered by a challenging issue. The major hurdle is the large plasmon losses in metals: plasmonic signals do not propagate more than 10-100 μm distance, prohibiting chip-scale integration of plasmonic circuits. This high loss of plasmon propagation is attributed to various scattering mechanisms of electrons in metals and has been viewed to be insurmountable.

5-year horizon:

- Develop low-loss plasmonic materials/waveguides:
  - **Solutions**: develop low-loss plasmonic metamaterial structure by employing a hyperbolic metamaterial; maximally suppress the electric fields in the metal layers part of the metamaterial structure. We can expect ~10mm propagation length of surface plasmons.

15-to-25-year horizon:

- Develop chip-scale multifunctional plasmonic circuits:
  - **Solutions**: integrate with various transduction materials (such as electro-optic, phase-change materials) for high-speed energy-efficient modulation/switching.
  -

Spoof Surface Plasmon Polariton (SSPP) or Spoof Plasmonic Technology

In the electromagnetic (EM) spectrum, the terahertz (THz) domain (300 GHz - 10 THz) remains elusive, underdeveloped and underexplored for computing, communications and signal processing due to the lack of enabling electronic and photonic technologies. On the one hand, conventional silicon semiconductor devices fall short of normal operation at THz frequencies, while, on the other hand, optical techniques cannot be readily scaled down to design THz circuits and systems.

Spoof surface plasmon polariton (SSPP), also known as spoof plasmonics or artificial plasmonics, are the states of the electromagnetic field bound to regularly patterned conducting surface with sub-wavelength device structures. SSPP surface can be qualitatively represented as an effective medium with the dielectric function having the Drude form with the effective (spoof) plasma frequency determined by the SSPP structures. The principal difference between SSPP and conventional surface plasmon polariton (SPP) is that in the case of SSPP, the field does not penetrate into the material, and the plasmonic form of the dispersion phenomenon is due to leaky resonances inside the corrugations. Consequently, the spoof plasma frequency is set up by the geometry of the grooves rather than by electrons in the conductor. Thus, SSPP not only provides means for subwavelength confinement of the EM field but also to engineer metamaterials with prescribed spoof plasma frequency. Notably, the SPP effect is pronounced in 100's of THz frequencies as opposed to SSPP, which is effective in terahertz domain (0.3 – 10 THz).

THz research is expected to transcend the confines of traditional electronic systems and will impact multiple disciplines. It will establish SSPP as a special kind of excitation, which not only mimics the dispersion of SPP, but also reproduces plasmon features within the context of interactions with matter. SSPP structures can sense as well as perform complex Boolean-type functions, thereby enabling a fusion of sensing and processing of bio-data in tag-free and non-destructive bio-sensing architectures. Despite the numerous potentials SSPP components hold, the design and development of SSPP devices encounter several challenges stemming from a host of reasons, namely, the lack of a unified theoretical framework for hybrid electronic-SSPP circuits and devices and multifunctional SSPP designs.

5-to-10-year horizon

- Development of a theoretical foundation of SSPP integrated architectures.
- Development of fabrication techniques for basic SSPP passive components such as:

1. Passive filter using cascade structures
2. THz power divider with phase coherence
3. THz dispersion compensator
4. THz multiplexer and de-multiplexer
5. THz circulator.

- Development of fabrication techniques for basic SSPP active components such as:

1. THz source with high extraction efficiency using heterogeneous SSPP
2. THz buffer/ repeater using near field SSPP transceiver
3. THz modulator using pre-engineered defect incorporated SSPP
4. THz active filter with unity power gain.

15-to-25-year horizon

- Development of SSPP-based information processing and communications circuits.
- Development of SSPP-based bio-sensing and chemical detection architectures.
- Development of SSPP-based on-chip solutions for THz imaging and spectroscopy.
- Development of testing and measurements equipment for THz components and systems.

TWG Members

Chair: Meyya Meyyappan

Supriyo Bandyopadhyay	Bill Bottoms	Jin-Woo Han
Amr S. Helmy	Muhammad Hussein	Cheol Seong Hwang
Hong Koo Kim	Jason Marsh	Pinaki Mazumder
Oussama Moutanbbir	Arokia Nathan	Kaushik Roy
Frank Schwierz	Doug Verret	Greg Whiting
Philip Wong	Jim Browning	