INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS™

2018 UPDATE

BEYOND CMOS

Table of Contents

List of Figures

List of Tables

Link to Beyond CMOS Tables File

ACKNOWLEDGMENTS

Sapan Agarwal
Hiro Akinaga
Mustafa Badaroglu
Christian Binek
Geoffrey Burr
Leonid Butov
Kerem Camsari
Robin Cantor
Gert Cauwenberghs
An Chen
Winston Chern
Supriyo Datta
John Dallesasse
Shamik Das
Eric Dauler
Erik DeBenedictis
Peter Dowben
Tetsuo Endoh
Bob Fagaly
Pascal Febvre
Cathy Foley
Akira Fujiwara
Michael Frank
Paul Franzon
Mike Garner
Chakku Goplan
Bogdan Govoreanu
Deep Gupta
Masami Hane
Jennifer Hasler
Yoshihiro Hayashi
Anna Herr
Mutsuo Hidaka
Toshiro Hiramoto
D. Scott Holmes
Sharon Hu
Engin Ipek
Satoshi Kamiyama
Kiyoshi Kawabata
Asif Khan
Ilya Krivorotov
Anna Leese de Escobar
Xiuling Li
Tsu-Jae King Liu
Matthew Marinella
Tzvetan Metodi
Rivu Midya
Nancy Missert
Oleg Mukhanov
Johannes Muller
Azad Naeemi
Mike Niemier
Dmitri Nikonov
Yutaka Ohno
Chenyun Pan
Satyavolu Papa Rao
Ferdinand Peper
Shriram Ramanathan
Mingyi Rao
Shashi Paul
Titash Rakshit
Arijit Raychowdhury
Horst Rogalla
Sayeef Salahuddin
Shintaro Sato
Marc Sherwin
Takahiro Shinada
John-Paul Strachan
Tarek Taha
Shinichi Takagi
Norikatsu Takaura
Naoki Takeuchi
Tsutomu Teduka
Sergey Tolpygo
Yasuhide Tomioka
Wilman Tsai
Tohru Tsuruoka
Robert Voigt
Zhongrui Wang
R. Stanley Williams
Dirk Wouters
Kojiro Yagami
J. Joshua Yang
Noboyuki Yoshikawa
Victor Zhirnov

BEYOND CMOS

1. INTRODUCTION

1.1. SCOPE OF BEYOND-CMOS FOCUS TEAM

Dimensional and functional scaling[1] of CMOS is driving information processing[2] technology into a broadening spectrum of new applications. Scaling has enabled many of these applications through increased performance and complexity. As dimensional scaling of CMOS will eventually approach fundamental limits, several new information processing devices and microarchitectures for both existing and new functions are being explored to extend the historical integrated circuit scaling cadence. This is driving interest in new devices for information processing and memory, new technologies for heterogeneous integration of multiple functions, and new paradigms for system architecture. This chapter, therefore, provides an IRDS perspective on emerging research device technologies and serves as a bridge between conventional CMOS and the realm of nanoelectronics beyond the end of CMOS scaling. (Material challenges related to emerging research devices are addressed in a complementary 2017 chapter entitled *Emerging Research Materials*.)

An overarching goal of this chapter is to survey, assess, and catalog viable emerging devices and novel architectures for their long-range potential and technological maturity and to identify the scientific/technological challenges gating their acceptance by the semiconductor industry as having acceptable risk for further development. This chapter also surveys new applications of emerging technologies, e.g., hardware security, cryogenic electronics.

This goal is accomplished by addressing two technology-defining domains: 1) extending the functionality of the CMOS platform via heterogeneous integration of new technologies ("More Moore"), and 2) stimulating invention of new information processing paradigms ("Beyond CMOS"). The relationship between these domains is schematically illustrated in Figure BC1.1. Novel computing paradigms and application pulls (e.g., big data, IoT, artificial intelligence, autonomous systems, exascale computing) introduce higher performance and efficiency requirements, which is increasingly difficult for the saturating More Moore technologies to fulfill. Beyond-CMOS technologies may provide the required devices, processes, and architectures for the new era of computing.

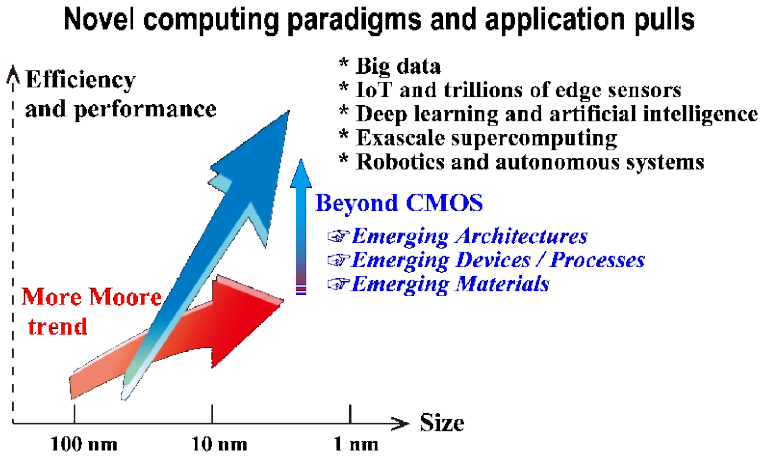


Figure BC1.1 Relationship of More Moore, Beyond CMOS, and Novel Computing Paradigms and Applications (Courtesy of Japan beyond-CMOS group)

The chapter is intended to provide an objective, informative resource for the constituent nanoelectronics communities pursuing: 1) research, 2) tool development, 3) funding support, and 4) investment. These communities include universities, research

[1] *Functional Scaling: Suppose that a system has been realized to execute a specific function in a given, currently available, technology. We say that system has been functionally scaled if the system is realized in an alternate technology such that it performs the identical function as the original system and offers improvements in at least one of size, power, speed, or cost, and does not degrade in any of the other metrics.*

[2] *Information processing refers to the input, transmission, storage, manipulation or processing, and output of data. The scope of the BC Chapter is restricted to data or information manipulation, transmission, and storage.*

institutes, and industrial research laboratories; tool suppliers; research funding agencies; and the semiconductor industry. The potential and maturity of each emerging research device and architecture technology are reviewed and assessed to identify the most important scientific and technological challenges that must be overcome for a candidate device or architecture to become a viable approach.

The chapter is divided into five sections: 1) memory devices, 2) information processing or logic devices, 3) emerging application areas, 4) emerging device-architecture interaction, and 5) assessment. Some detail is provided for each entry regarding operation principles, advantages, technical challenges, maturity, and current and projected performance. The chapter also discusses applications and architectural focus combining emerging research devices offering specialized, unique functions as heterogeneous core processors integrated with a CMOS platform technology. This represents the nearer term focus of the chapter, with the longer-term focus remaining on discovery of an alternate information processing technology beyond digital CMOS.

The technology entries in this chapter adopt the main entries of the International Technology Roadmap of Semiconductors (ITRS) Emerging Research Devices (ERD) chapter. The memory device section is expanded to include a new technology entry: “massive storage devices”. The previous “novel STTRAM devices” entry has been modified to “novel magnetic memory” to cover a broader range of devices. The “macromolecular (polymer) memory”, which was previously included in the category of ReRAM, is listed as a separate entry. “ReRAM” continues to be extensively tracked with simplified categorization. The logic device section categorizes technology entries as CMOS extension, charge-based beyond-CMOS devices, and non-charge-based beyond-CMOS devices.

1.2. Difficult Challenges

1.2.1. Introduction

The semiconductor industry is facing some difficult challenges related to extending integrated circuit technology to new applications and to beyond the end of CMOS dimensional scaling. One class relates to propelling CMOS beyond its ultimate density and functionality by integrating a new high-speed, high-density, and low-power memory technology onto the CMOS platform. Another class is to extend CMOS scaling with alternative channel materials. The third class is information processing technologies substantially beyond those attainable by CMOS using an innovative combination of new devices, interconnect, and architectural approaches for extending CMOS and eventually inventing a new information processing platform technology. The fourth class is to extend ultimately scaled CMOS as a platform technology into new domains of functionalities and application, also known as “more than Moore”. The fifth class is to bridge the gap between novel devices and unconventional architectures and computing paradigms. These difficult challenges are summarized in Table BC1.1.

1.2.2. Device Technologies

Difficult challenges gating development of beyond-CMOS devices include those related to memory technologies, information processing or logic devices, and heterogeneous integration of multi-functional components, a.k.a. More-than-Moore (MtM) or Functional Diversification.

One challenge is the need of a new memory technology that combines the best features of current memories in a fabrication technology compatible with CMOS process flow and that can be scaled beyond the present limits of SRAM and FLASH. This would provide a memory device fabrication technology required for both stand-alone and embedded memory applications. The ability of an MPU to execute programs is limited by interaction between the processor and the memory, and scaling does not automatically solve this problem. The current evolutionary solution is to increase MPU cache memory, thereby increasing the floor space that SRAM occupies on an MPU chip. This trend eventually leads to a decrease of the net information throughput. In addition to auxiliary circuitry to maintain stored data, volatility of semiconductor memory requires external storage media with slow access (e.g., magnetic hard drives, optical CD, etc.). Therefore, development of electrically accessible non-volatile memory with high speed and high density would initiate a revolution in computer architecture. This development would provide a significant increase in information throughput beyond the traditional benefits of scaling when fully realized for nanoscale CMOS devices.

Table BC1.1 Beyond CMOS Difficult Challenges

Difficult Challenges	*Summary of Issues and Opportunities*
Scale high-speed, dense, embeddable, volatile/non-volatile memory technologies to replace SRAM and FLASH in appropriate applications.	The scaling limits of SRAM and FLASH in 2D are driving the need for new memory technologies to replace SRAM and FLASH memories. Identify the most promising technical approach(es) to obtain electrically accessible, high-speed, high-density, low-power, (preferably) embeddable volatile and non-volatile memories. The desired material/device properties must be maintained through and after high temperature and corrosive chemical processing. Reliability issues should be identified and addressed early in the technology development.
Extend CMOS scaling	Develop new materials to replace silicon (or III-V, Ge) as alternate channel and source/drain to increase the saturation velocity and to further reduce V_{dd} and power dissipation in MOSFETs while minimizing leakage currents Develop means to control the variability of critical dimensions and statistical distributions (e.g., gate length, channel thickness, S/D doping concentrations, etc.) Accommodate the heterogeneous integration of dissimilar materials. The desired material/device properties must be maintained through and after high temperature and corrosive chemical processing. Reliability issues should be identified and addressed early in this development.
Continue functional scaling of information processing technology substantially beyond that attainable by ultimately scaled CMOS.	Invent and reduce to practice a new information processing technology to replace CMOS as the performance driver. Ensure that a new information processing technology has compatible memory technologies and interconnect solutions. A new information processing technology must be compatible with a system architecture that can fully utilize the new device. Non-binary data representations or non-Boolean logic may be required to employ a new device for information processing, which will drive the need for new system architectures. Bridge the gap that exists between materials behaviors and device functions. Accommodate the heterogeneous integration of dissimilar materials. Reliability issues should be identified and addressed early in the technology development.
Extend ultimately scaled CMOS as a platform technology into new domains of functionalities and application (“more than Moore, MtM”).	Discover and reduce to practice new device technologies and primitive-level architecture to provide special purpose optimized functional cores (e.g., accelerator functions) heterogeneously integrable with CMOS. Provide added value by incorporating functionalities that do not necessarily scale according to “Moore’s Law”. Heterogeneous integration of digital *and* non-digital functionalities into compact systems that will be the key driver for a wide variety of application fields, such as communication, automotive, environmental control, healthcare, security, and entertainment.
Bridge the gap between novel devices and unconventional architectures and computing paradigms.	Identify suitable opportunities in unconventional architectures and computing paradigms that can utilize unique characteristics of novel devices. Identify emerging devices that can implement computing functions and architectures more efficiently than CMOS and Boolean logic.

A related challenge is to sustain scaling of CMOS logic technology. One approach to continuing performance gains as CMOS scaling matures in the next decade is to replace the strained silicon MOSFET channel (and the source/drain) with an alternate material offering a higher potential quasi-ballistic-carrier velocity and higher mobility than strained silicon. Candidate materials include strained Ge, SiGe, a variety of III-V compound semiconductors, and carbon materials. Introduction of non-silicon materials into the channel and source/drain regions of an otherwise silicon MOSFET (i.e., onto a silicon substrate) is fraught with several very difficult challenges. These challenges include heterogeneous fabrication of high-quality (i.e., defect free) channel and source/drain materials on non-lattice matched silicon, minimization of band-to-band tunneling in narrow bandgap channel materials, elimination of Fermi level pinning in the channel/gate dielectric interface, and fabrication of high-κ gate dielectrics on the passivated channel materials. Additional challenges are to sustain the required reduction in leakage currents and power dissipation in these ultimately scaled CMOS gates and to introduce these new materials into the MOSFET while simultaneously minimizing the increasing variations in critical dimensions and statistical fluctuations in the channel (source/drain) doping concentrations.

The industry is now addressing the increasing importance of a new trend of functional diversification, where added value to devices is provided by incorporating functionalities that do not necessarily scale according to "Moore's Law". In this chapter, an "Emerging Application Areas" section covers unconventional applications of existing and novel technologies. Two topics, emerging devices for hardware security and cryogenic electronics, are discussed in this section.

A longer-term challenge is invention and reduction to practice of a manufacturable information processing technology addressing "beyond CMOS" applications. For example, emerging research devices might be used to realize special purpose processor cores that could be integrated with multiple CMOS CPU cores to obtain performance advantages. These new special purpose cores may provide a particular system function much more efficiently than a digital CMOS block, or they may offer a uniquely new function not available in a CMOS-based approach. Solutions to this challenge beyond the end of CMOS scaling may also lead to new opportunities for such an emerging research device technology to eventually replace the CMOS gate as a new information processing primitive element. A new information processing technology must also be compatible with a system architecture that can fully utilize the new device. A non-binary data representation and non-Boolean logic may be required to employ a new device for information processing. These requirements will drive the need for new system architectures. The requirements and opportunities correlating emerging devices and architectures are discussed in the "Emerging Device-Architecture Interaction" section.

1.2.3. MATERIALS TECHNOLOGIES

The most difficult challenge for Beyond CMOS is to deliver materials with controlled properties that will enable operation of emerging research devices in high density at the nanometer scale. To improve control of material properties for high-density devices, research on materials synthesis must be integrated with work on new and improved metrology and modeling. These important objectives are addressed in the "2017 Emerging Research Materials" chapter.

1.3. NANO-INFORMATION PROCESSING TAXONOMY

Information processing systems to accomplish a specific function, in general, require several different interactive layers of technology. One comprehensive top-down list of these layers begins with the required application or system function, leading to system architecture, micro- or nano-architecture, circuits, devices, and materials. A different bottom-up representation of this hierarchy begins with the lowest physical layer represented by a computational state variable and ends with the highest layer represented by the architecture. In this representation focused on generic information processing at the device/circuit level, a fundamental unit of information (e.g., a bit) is represented by a computational state variable, for example, the position of a bead in the ancient abacus calculator or the charge (or voltage) state of a node capacitance in CMOS logic. The electronic charge as a binary computational state variable serves as the foundation for the von Neumann computational system architecture. A device provides the physical means of representing and manipulating a computational state variable among its two or more allowed discrete states. Eventually, device concepts may transition from simple binary switches to devices with more complex information processing functionality, perhaps with multiple fan-in and fan-out. The device is a physical structure resulting from the assemblage of a variety of materials possessing certain desired properties obtained through exercising a set of fabrication processes. An important layer, therefore, encompasses the various materials and processes necessary to fabricate the required device structure, which is a focus of the "Beyond CMOS (BC)" chapter. The data representation is how the computational state variable is encoded by the assemblage of devices to process the bits or data. Two of the most common examples of data representation are binary digital and continuous or analog signal. This layer is within the scope of the BC chapter. The architecture layer encompasses three subclasses of this taxonomy: 1) nano-architecture or the physical arrangement or assemblage of devices to form higher level functional primitives to represent and execute a computational model, 2) the computational model that describes the algorithm by which information is processed using the primitives, e.g., logic, arithmetic, memory, cellular nonlinear network (CNN), and 3) the system-level architecture that describes the conceptual structure and functional behavior of the system exercising the computational model.

2. EMERGING MEMORY DEVICES

The emerging research memory technologies tabulated in this section are a representative sample of published research efforts (circa 2015 – 2017) describing alternative approaches to established memory technologies.[3] The scope of this section also includes updated subsections addressing the "Select Device" required for a crossbar memory application and an updated treatment of "Storage Class Memory" (including Solid State Disks).

[3] *Including a particular approach in this section does not in any way constitute advocacy or endorsement. Conversely, not including a particular concept in this section does not in any way constitute rejection of that approach. This listing does point out that existing research efforts are exploring a variety of basic memory mechanisms.*

Figure BC2.1 is a taxonomy of the prototypical and emerging memory technologies. An overarching theme is the need to monolithically integrate each of these memory options onto a CMOS technology platform in a seamless manner. Fabrication technologies are sought that are modifications of or additions to established CMOS platform technologies. A goal is to provide the end user with a device that behaves similarly to the familiar silicon memory chip.

This memory portion of this section is organized around a set of eight technology entries shown in the column headers of Table BC2.1. These entries were selected using a systematic survey of the literature to determine areas of greatest worldwide research activity. Each technology entry listed has several sub-categories of devices that are grouped together to simplify the discussion. Key parameters associated with the technologies are listed in the table. For each parameter, two values for performance are given: 1) theoretically predicted performance values based on calculations and early experimental demonstrations, 2) up-to-date experimental values of these performance parameters reported in the cited technical references.

The tables have been extensively footnoted, and details may be found in the indicated references. The text associated with the table gives a brief summary of the operating principles of each device and significant scientific and technological issues, not captured in the table, but which must be resolved to demonstrate feasibility.

The purpose of many memory systems is to store massive amounts of data, and therefore memory capacity (or memory density) is one of the most important system parameters. In a typical memory system, the memory cells are connected to form a two-dimensional array, and it is essential to consider the performance of memory cells in the context of this array architecture. A memory cell in such an array can be viewed as being composed of two fundamental components: the 'storage node' and the 'select device', the latter of which allows a given memory cell in an array to be addressed for read or write. Both components impact scaling limits for memory. For several emerging resistance-based memories, the storage node can, in principle, be scaled down below 10 nm,[1] and the memory density will be limited by the select device. Thus, the select device represents a serious bottleneck for ReRAM scaling to 10 nm and beyond. Planar transistors (e.g. FET or BJT) are typically used as select devices. In a two-dimensional layout using in-plane select FETs the cell layout area is Acell=(6-8)F^2. In order to reach the highest possible 2-D memory density of $4F^2$, a vertical select transistor can be used. Table BC2.3 shows several examples of vertical transistor approaches currently being pursued for select devices. Another approach to obtaining a select device with a small footprint is a two-terminal nonlinear device, e.g. a diode, either as a separate device or a strong nonlinearity intrinsic to a resistive memory element. Table BC2.4 displays benchmark parameters required for a 2-terminal select device, and Table BC2.5 summarizes the operating parameters for several candidate 2-terminal select devices.

Storage-class memory (SCM) describes a device category that combines the benefits of solid-state memory, such as high performance and robustness, with the archival capabilities and low cost per bit of conventional hard-disk magnetic storage. Such a device requires a non-volatile memory technology that can be manufactured at a very low cost per bit. Table BC2.6 lists a representative set of target specifications for SCM devices and systems, which are compared against benchmark parameters offered by existing technologies (HDD, NAND Flash, and DRAM). Two columns are shown, one for the slower S-class Storage Class Memory, and one for fast M-class SCM, as described in Section 2.4. These numbers describe the performance characteristics that will likely be required from one or more emerging memory devices in order to enable the emerging application space of Storage Class Memory. Table BC2.7 illustrates the potential for storage-class memory applications of a number of prototypical memory technologies (Table BC2.1) and emerging research memory candidates (Table BC2.2). The table shows qualitative assessments across a variety of device characteristics, based on the target system parameters from Table BC2.6. These tables are discussed in more detail in Section 2.4.

Table BC2.1 Emerging Research Memory Devices—Demonstrated and Projected Parameters

Table BC2.2 Emerging Research Memory Devices—Redox RAM Demonstrated and Projected Parameters

2.1. MEMORY TAXONOMY

Figure BC2.1 provides a simple visual method of categorizing memory technologies. At the highest level, memory technologies are separated by the ability to retain data without power. Non-volatile memory offers essential use advantages, and the degree to which non-volatility exists is measured in terms of the length of time that data can be expected to be retained. Volatile memories also have a characteristic retention time that can vary from milliseconds to (for practical purposes) the length of time that power remains on. Non-volatile memory technologies are further categorized by their maturity. Flash memory is considered the baseline non-volatile memory because it is highly mature, well optimized, and has a significant commercial presence. Flash memory is the benchmark against which prototypical and emerging non-volatile memory technologies are measured. Prototypical memory technologies are at a point of maturity where they are commercially available (generally for niche applications), and have a large

scientific, technological, and systems knowledge base available in the literature. The focus of this section is Emerging Memory Technologies. These are the least mature memory technologies in Figure BC2.1, but they have been shown to offer significant potential benefits if various scientific and technological hurdles can be overcome. This section provides an overview of these emerging technologies, their potential benefits, and the key research challenges that will allow them to become viable commercial technologies.

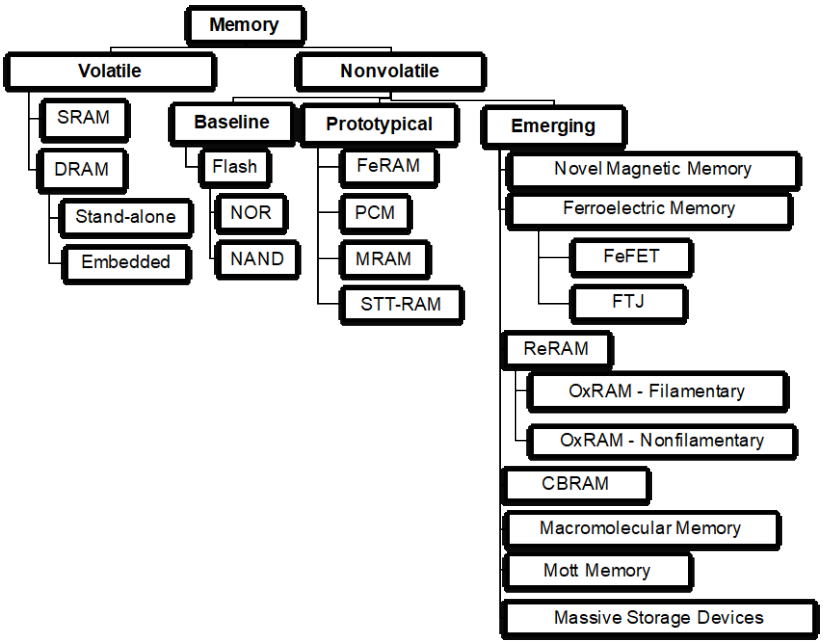


Figure BC2.1 Taxonomy of Emerging Memory Devices

2.2. EMERGING MEMORY DEVICES

2.2.1. NOVEL MAGNETIC MEMORY

Spin-transfer torque RAM (STT-RAM) is covered in the More Moore Chapter of IRDS since this memory had matured and become commercially available. However, there has been recent progress in emerging memory devices which utilize novel magnetic mechanisms, and in response a completely updated section covering this topic has been added to Beyond CMOS. There has been an evolution in thinking, as to what should be the target magneto-electric devices for CMOS 'plug-in' replacement logic and memory. The earliest magneto-electric devices were based on a magnetic tunnel junction structure[2,3] which consists of two ferromagnetic (FM) layers separated by a non-magnetic insulator where the device resistance is determined by the relative orientation of the magnetization of the two FM layers. The magnetization of the free layer is exchange coupled to the Cr_2O_3 (or other magneto-electric) interface magnetization. A bias voltage applied across a magneto-electric layer, like chromia Cr_2O_3, reverses the interface magnetization, which in turn switches the magnetization of the free layer[4, 5, 6]. A non-magnetic Hall bar on top of magneto-electric chromia has now been demonstrated as a readout mechanism for a memory state via the anomalous Hall effect.[7,8]

More recent attention[6,9,10,11,12] has shifted to the magneto-electric transistor (the ME-spinFET). Magneto-electric transistor schemes are based on polarization of the semiconductor channel, by the boundary polarization of the magneto-electric gate. The advantage to the magneto-electric field effect transistor is that such schemes avoid the complexity and detrimental switching energetics associated with exchange-coupled ferromagnets. Spintronic devices based solely on the switching of a magneto-electric transistor, will have a switching speed that will be limited only by the switching dynamics of that magneto-electric material and above all are voltage controlled spintronic devices. Moreover, these magneto-electric devices promise to provide a unique field effect spin transistor (spin-FET)-based interface for input/output of other novel computational devices. These devices offer spintronics without a ferromagnet, with faster write speeds (<20 ps/full adder), at a lower cost in energy (<20 aJ/full adder), greater temperature stability (operational to 400 K or more), and scalability, and require far fewer device elements (transistor equivalents) than CMOS.

Magneto-electric field effect transistor device concepts that emphasize the value of using a narrow channel conductor, with strong spin-orbit coupling (SOC), will have, however, enhanced ON/OFF ratios and even greater functionality. To some extent the device has been "proven" as the anomalous Hall effect [7,8] devices demonstrate feasibility.

2.2.1.1. THE ANTI-FERROMAGNET SPIN-ORBIT READ (AFSOR) DEVICE

The anti-ferromagnet spin-orbit read (AFSOR) device structure has interesting advantages: the potential for high and sharp voltage 'turn-on'; inherent non-volatility of magnetic state variables; absence of switching currents; large ON/OFF ratios; and multistate logic and memory applications. The design will provide reliable room-temperature operation with large ON/OFF ratios (>10^7) well beyond what can be achieved using magnetic tunnel junctions.[13,14] Again, the core idea is the use of the boundary polarization of the magneto-electric to spin polarize or partly spin polarize a very thin semiconductor, ideally a 2D material, with very large spin orbit coupling.

If the semiconductor channel retains large spin orbit coupling, then the spin current, mediated by the gate boundary polarization, may be enhanced and, to some extent, topologically protected. The latter implies that each spin current has a preferred direction.

2.2.1.2. THE MAGNETO-ELECTRIC SPIN-FET MULTIPLEXER

There is a variant where inversion symmetry is not as strictly broken, which leads to a non-volatile spintronics version of multiplexer logic (MUX). The magneto-electric spin-FET multiplexer also exploits the modulation of the spin-orbit splitting of the electronic bands of the semiconductor channel through a "proximity" magnetic field derived from a voltage-controlled magneto-electric material. By using semiconductor channels with large spin-orbit coupling, we expect to obtain a transverse spin Hall current, as well as a spin current overall. Depending on the magnitude of the effective magnetic field in the narrow channel, we anticipate two different operational regimes. Like the AFSOR magneto-electric spin FET, the magneto-electric spin-FET multiplexer uses spin-orbit coupling in the channel to modulate spin polarization and hence the conductance (by spin) of the device. There is a source-drain voltage and current difference between the two FM source contacts due to the spin-Hall effect when spin-orbit coupling is present. This output voltage can be modulated by the gate or gates, which influences the spin-orbit interaction in the channel especially when it is both top and bottom gated. The spin-Hall voltage in the device can be increased by using different FMs in the source and drain. To increase the spin fidelity of current injection at the source end, one could add a suitable tunnel junction layer (basically a 1nm oxide layer) between the magnetic source and the 2D semiconductor channel, though this latter modification would result in diminished source-drain currents.

The challenges in pushing forward these technologies extends not only to the fabrication and characterization of a new generation of non-volatile magneto-electric devices, but also to ascertaining the optimal implementation of CMOS plug in replacement circuits. That said, the magneto-electric transistor has far fewer challenges to implementation than the magneto-electric magnetic tunnel junction, so, not surprisingly, there is a shifting of development effort toward these and related devices for both memory and logic.

2.2.2. OXIDE-BASED RESISTIVE MEMORY (OXRAM)

The redox-based nanoionic memory operation is based on a *change in resistance* of a MIM structure caused by ion (cation or anion) migration combined with redox processes involving the electrode material or the insulator material, or both.[15,16,17] Three classes of electrically induced phenomena have been identified that involve chemical effects, i.e., effects which relate to redox processes in the MIM cell. In these three ReRAM classes, there is a competition between thermal and electrochemical driving forces involved in the switching mechanism. Two major types of ReRAM exist: i) those based on metal oxide (OxRAM), which involve oxygen vacancy motion, and ii) conducting bridge-based RAM (CBRAM), which involves metal cation motion. This section covers the three categories of OxRAM, and conducting bridge-based RAM (CBRAM) is covered in the following section. Beyond CMOS has sub-categorized oxide-based ReRAM (OxRAM) based on the electrical switching type (bipolar versus unipolar) and whether the device forms a filament. Most of the literature fits into the three categories: bipolar filamentary, unipolar filamentary, and nonfilamentary.[18]

In most cases, the conduction is of a filamentary nature, and hence a one-time formation process is required before the bi-stable switching can be started. If this effect can be controlled, memories based on this bi-stable switching process can be scaled to very small feature sizes. The switching speed is limited by the ion transport. If the active distance over which the anions or cations move is small (in the <10 nm regime) the switching time can be as low as a few nanoseconds. Many of the finer details of the ReRAM switching mechanisms are still unknown. Developing an understanding of the physical mechanisms governing switching of the redox memory is a key challenge for this technology. Nevertheless, recent experimental demonstrations of scalability,[19] retention,[20] and endurance[21] are encouraging.

2.2.2.1. BIPOLAR-FILAMENTARY OXRAM

Bipolar filamentary OxRAM is the most common form of oxide-based ReRAM. As the name suggests these devices are able to maintain their resistance after the application of an electrical stimulus. The mechanism of operation of these devices often involves valance change of the dielectrics (e.g. oxides) or redox reactions of the dielectrics between the electrodes.[22,23] The dielectrics are mostly comprised of one or a few layers of insulating materials[24] (e.g., oxide AlO_x, HfO_x, TaO_x, TiO_x, WO_x, ZrO_x, oxynitrides AlO_xN_y, or nitrides including AlN_x and CuN_x) sandwiched between two electrodes. During the time of fabrication or the electroforming process, the electrode with a relatively active metal interacts with the dielectric material to generate anion

vacancies (Oxygen or Nitrogen).[25] The counter electrode is typically inert to provide an electrical contact and form a relatively high electronic barrier at the interface (e.g. Schottky-like barrier)[26]. TaOx and HfOx are the leading candidates among the aforementioned dielectrics, due to their superior performance (e.g. endurance) and CMOS compatibility.

A one-time electroforming process is required for most OxRAM switches to create a conduction filament(s)/channel(s) across the dielectric layer linking the electrodes. All subsequent switching events which are operated at significantly smaller voltages are believed to rupture and recover a part of the filament in the vicinity of the inert metal electrodes. Electroforming is an electrochemical process where the reduction of the oxide and nitride dielectrics takes place [27]. The electroforming process can be avoided by employing a thinner dielectric layer with an anion scavenging layer, thereby leading to electroforming free devices.

Since the demonstration of a single crosspoint HfOx device with a 10 nm dimension in 2011[28], scaling to a smaller size had been achieved by employing a sidewall electrode in a 1×3 nm^2 cross-sectional HfO_x-based OxRAM device with reasonable performance in terms of both endurance and retention.[29] Regarding repeatability, up to 10^{12} cycles has been demonstrated with $Zr:SiO_x$ sandwiched by graphene oxide layers recently.[30] This is comparable to the previous endurance record demonstrated using a Ta_2O_{5-x}/TaO_{2-x} structure developed by Samsung in 2011.[31] The fastest OxRAM was demonstrated on a TaO_x device with a programing duration about 100 ps in 2011.[32] Recently, 85ps switching was demonstrated in nitride based devices.[33] Remarkable progress has been made on improvement of the OxRAM reliability, especially for TaO_x. Extrapolated retention at 85°C by stressing TaO_x in the temperature range from 300°C to 360°C is estimated to be years with an activation energy of 1.6 eV.[34] Reliable switching operations have been demonstrated at 340°C with devices based on 2D layered heterostructures (e.g., graphene/$MoS_{2-x}O_x$/graphene).[35]

Unconventional electrodes like graphene have been paired with HfO_x dielectrics and yield a low power consumption, a write/erase energy about 230 fJ per bit for a single programming transition.[36] Recently a Pt/BMO((Bi, Mn)O_x)/Pt structured OxRAM device was used to demonstrate an even lower write/erase energy per transition of the order of 3.8 pJ/bit for read and 20 pJ/bit for write operation.[37]

Large scale integration of OxRAM switching based on 1T1R schemes has been carried out by Toshiba, Panasonic and IMEC. In 2013, Toshiba announced the 32 Gb RRAM chip integrated with 24 nm CMOS.[38] In 2014, Panasonic and IMEC demonstrated the encapsulated cell structure with an $Ir/Ta_2O_5/TaO_x/TaN$ stack on a 2-Mbit chip with 40 nm node. In addition, passive integration of 1S1R scheme has been reported by Crossbar on a 4-Mbit chip, but the material stack of the OxRAM switch has not been revealed.[39]

A number of technical challenges remain despite the significant advancements made in the field of OxRAM. These technical challenges are hampering the commercialization of OxRAM. One of the main challenges is the fact that the switching currents for devices based on the currently most mature materials (e.g., HfO_x and TaO_x) are still too high (above tens of μA) for large arrays. Apart from that, the filament formation and rupture processes are stochastic in nature, which leads to variation in switching parameters like the voltage and resistance distribution of the switching. This is especially detrimental to certain applications such as multilevel cell memory. Passive arrays of OxRAM switches will require either select devices (selectors) or intrinsic IV nonlinearity. However, an external select device that meets all the requirements, including great endurance, large selectivity (ON/OFF ratio), good scalability, fast speed, low variability, suitable operation current and voltage, remains to be demonstrated.

2.2.2.2. Bipolar Non-Filamentary OxRAM

The *Bipolar Non-Filamentary* OxRAM is a non-volatile bipolar resistive switching device composed of one or more oxide layers. One layer is a conductive metal oxide (CMO), which is usually a perovskite such as PrCaMnO3 or $Nb:SrTiO_3$.[40] In contrast to *Unipolar* and *Bipolar Filamentary* OxRAM devices – typically based on binary oxides such as TiO_x, NiO_x, HfO_x, TaO_x or combinations thereof—the resistance change effect of the *Bipolar Non-Filamentary* OxRAM is *uniform*. Depending on the materials choice and structure the current is conducted across the entire electrode area, or at least across the majority of this area. A forming step to create a conductive filament is *not* needed. Non-volatile memory functionality is achieved by the field-driven redistribution of oxygen vacancies close to the contact resulting in a change of the electronic transport properties of the interface (e.g. by modifying the Schottky barrier height). Oxygen can be exchanged between layers due to the exponential increase in ion mobility at high fields. Low current densities, uniform conduction, and bipolar switching imply that substantial self-heating is not involved. Typical R_{OFF} to R_{ON} ratios are on the order of 10.

One class of the Bipolar Non-Filamentary OxRAM includes a deposited ion conductive tunnel layer (Tunnel ReRAM), e.g. ZrO_2. Here, a redistribution of oxygen vacancies causes a change of the electronic transport properties of the tunnel barrier. Low current densities and area scaling of device currents enable ultra-high-density memory applications. Set, reset, and read currents scale with device area. In addition, set, reset, and write currents are controlled by the tunnel oxide and hence, can be adjusted by changing the tunnel barrier thickness. Both set and reset IV characteristics are highly nonlinear enabling true 1R cross-point architectures *without* the need for an additional selector device for asymmetric arrays up to 512×4096 bit. No external circuitry

is needed for current control during set operation. A continuous transition between on and off states allow straightforward multi-level programming without the need for precise current control.

The typical thickness of the CMO is greater than 5 nm and the tunnel barrier is typically 2–3 nm. If a tunnel barrier is present, the adjacent electrode needs to be an inert metal such as Pt to prevent oxidation during operation. For the case of PCMO cell, low deposition temperatures less than 425°C of all layers enables back end integration schemes.

Currently the technology is in the research and development stage. Depending on material system and structure cycling endurance over 10,000 cycles and up to a billion cycles as well as data retention from days to months at 70°C has been achieved on single devices.[41, 42, 43] Within the Bipolar Non-Filamentary OxRAM device family the Tunnel OxRAM is probably the furthest developed technology. Single device functionality is demonstrated down to 30 nm. Set, reset, and read currents scale with area and tunnel oxide thickness facilitating sub µA switching currents with read currents in the order of a few nA to a few 100 nA. BEOL integration schemes and CMOS/OxRAM functionality are verified for 200 nm devices on 200 mm CMOS wafers. True cross-point array (1R) functionality utilizing the self-selecting non-linear device IV characteristics and transistor-less array operation is demonstrated on fully decoded 4kb true cross-point arrays (1R) build on top of CMOS base wafers. SLC and MLC operations are demonstrated within 4kb arrays.

Major challenges to be resolved towards the commercialization of Bipolar Non-filamentary OxRAM are, in order of priority, a) improvement of data retention, b) the integration of conductive metal oxide layer (perosvkites) via ALD or the replacement of CMO by more process friendly materials, and c) the replacement of Pt electrodes by a non-reactive, more process-friendly electrode material.

The most important issue is the improvement of retention and the "voltage-time dilemma." This dilemma hypothesizes physical reasons as to why it is difficult in a particular device and material system to simultaneously obtain a long retention, with short low read voltages, and fast switching at moderate write voltages.[44] Even though the exact mechanism is still under investigation there is a common agreement that oxygen vacancies are moved by the external electric field resulting in different resistance states of the memory cell. Vacancy drift at room temperature is possible due to a field dependent mobility, which increases exponentially with field at fields of 1 MV/cm and larger. However, current models based on a field dependent mobility underestimate the experimentally observed ratio between set/reset times and data retention indicating that the mechanism is only partly understood. More theoretical work is needed to understand the kinetics of programming and retention mechanisms. Once understood, materials need to be chosen to maximize the ratio between set/reset and retention times. The goal is to set/reset devices at low temperatures and meet retention requirement of 10 years at 70°C, 85°C, and 125°C, depending on the application. A multi-layer ReRAM structure (HfO_2/A_2O_3) was shown to improve retention by suppressing tail bit failure due to decreased oxygen ion diffusivity.[45]

Memory cells using conductive perovskite material as an electrode have proven to show excellent device-to-device and wafer-to-wafer reproducibility with yields close to 100%. One of the reasons might be that perovskites display high oxygen vacancy mobilities and tolerate large variations in the oxygen content while maintaining its crystal structure. From an integration perspective, ALD is the method of choice for advanced technology nodes and future 3D integration schemes. Key issues are the control of the metals ratio (perovskites are ternary or quaternary oxides), the control of the oxygen stoichiometry in the cell, oxygen loss in the presence of reducing atmospheres like H_2, as well as high temperatures required for crystallization. Eventually a migration to binary oxides with comparable properties might be required to resolve the integration challenges.

Platinum or other noble electrodes display superior device performance over fab friendly electrodes like TiN. On the one hand it was observed that the oxidation resistance of TiN is not sufficient to prevent oxidation and the formation of TiO_2 during operation. On the other hand, inert electrodes such as Pt or Pt-like metals are difficult to integrate. New oxidation resistant electrodes and Pt alternatives are required to reduce integration challenges and enable 3D integration schemes.

2.2.2.3. UNIPOLAR FILAMENTARY OXRAM

Note that unipolar filamentary OxRAM has been removed from the memory tracking tables, due to lack of research over the period covered by this Beyond CMOS chapter. However, this text section has been maintained to provide background on earlier unipolar OxRAM work, due to the close relationship and key differences with bipolar OxRAM.

Unipolar OxRAM is another resistive switching device, also referred to in the literature as thermochemical memory (TCM)[15] due to its primary switching mechanism. The device structure consists of a top electrode metal/insulator/bottom electrode metal (MIM) structure. Typical insulator materials are metal-oxides such as NiO_x, HfO_x, etc., and common metal electrodes include TiN, Pt, Ni, and W. In general, the device can be asymmetric (i.e. top electrode material differs from bottom electrode material), but unlike other types of ReRAM, asymmetry is not required.

The first reported resistive switching in these MIM structures after 2000 was unipolar in nature (see reference[46] for the first integrated device work that put metal oxide ReRAM in the spotlight). Unipolar is defined as switching where the same polarity

of voltage needs to be applied for changing the resistance from high to low (SET) or from low to high (RESET). Note that in the general case, polarity is still important (e.g. repeatable SET/RESET switching only occurs for one polarity of voltage with respect to one of the electrodes[47]). Only in symmetric structures (e.g. Pt/HfO_2/Pt), nonpolar behavior can be obtained, where SET and RESET are occurring irrespective of voltage polarity.[48]

The switching process is generally understood as being filamentary, where conduction is caused by a filamentary arrangement of defects (e.g. oxygen vacancies) throughout the thickness of the insulator film. As with other filamentary OxRAM devices, an initial high voltage "electroforming" step is required to form the conduction filament, while subsequent RESET/SET switching is thought to occur through local breaking/restoration of this conduction path.

The unipolar character of the switching indicates that drift (of charged defects) in an electric field plays a less important role (than it does in bipolar switching resistive memory), but that thermal effects probably dominate.[49,50] On the other hand, polarity effects indicate anodic oxidation (e.g. at Ni or Pt electrodes) is responsible for RESET.[47] These findings suggest a thermo-chemical "fuse" model for describing this unipolar switching. It has been shown for different MIM structures that both unipolar and bipolar switching mechanisms can be induced, depending on the operation conditions.[51,52,53,54] An interesting work reporting on the Scaling Effect on Unipolar and Bipolar Resistive Switching of Metal Oxides was published.[55]

A unipolar switching device is seen as advantageous for making scaled memory arrays, as it only requires a selector device as simple as a diode that can be stacked vertically with the memory device in a dense crossbar array.[46] In addition, the use of a single program voltage polarity greatly simplifies the circuitry.

On the other hand, as has been exemplified in mixed mode (unipolar/bipolar) operation of memory cells, there are important trade-offs between the unipolar and bipolar switching modes. On the positive side, unipolar switching mode typically shows a higher ON/OFF resistance ratio. On the negative side, unipolar switching is typically obtained at higher switching power (higher currents) than the bipolar mode, and also endurance is much more limited. As a result, major research and development work on resistive memories has shifted towards bipolar switching mechanisms. Yet, some interesting recent development work has been reported.[56,57,58,59,60,61] One paper[56] shows an endurance of over 10^6 cycles with a resistive window of over 5 orders of magnitude (and a reset current ~1mA). Others[57,58,59] demonstrate how unipolar RRAM elements can be integrated in a very simple way in an existing CMOS process (known as Contact ReRAM technology). This may provide a very inexpensive embedded ReRAM technology. Recently, integration unipolar ReRAM with a 29 nm CMOS process was reported.[59] The key attributes were a small cell size (0.03 μm^2), switching voltage of less than 3V, RESET current of less than 60 µA, endurance > 10^6 cycles, and short SET and RESET times of 500 ns and 100 µs, respectively. One paper[60] shows 4Mb array data using this same Contact-RRAM technology, fabricated using a 65 nm CMOS process. To accommodate for the low logic VDD process, on-chip charge pump was applied. Set and reset voltages are less than 2V. Another paper[61] reports on a novel approach using thermal assisted switching to lower the switching current.

As stated above, large OFF to ON resistance ratio is an attribute of unipolar switching. The low resistance window and large intrinsic variability of bipolar switching OxRAM may require complex and time-consuming switching operation schemes (e.g. the so-called verify scheme). Further study of the stability and control of the large resistance window (at low current levels), are required to determine if unipolar OxRAM variability can be improved, potentially even allowing for multi-level cell operation.

Major challenges to be resolved are the high switching current that seems inherent to the unipolar operation mode. Reset currents less than 100 µA are achieved but need further reduction to less than 10 µA.[57,58,59] Recently, a possible solution incorporating thermally assisted switching has been presented.[61]

2.2.3. CONDUCTING BRIDGE MEMORY

Conductive Bridge RAM (CBRAM), also referred to as, Programmable Metallization Cell (PMC), and electrochemical metallization cells, is a device which utilizes electrochemical control of nano-scale quantities of metal in thin dielectric films or solid electrolytes to perform the resistive switching operation.[62] The basic CBRAM cell is a metal–ion conductor–metal (MIM) system consisting of an electrode made of an electrochemically active material such as Ag, Cu or Ni, an electrochemically inert electrode such as W, Ta, or Pt, and a thin film of solid electrolyte sandwiched between both electrodes.[63] Large, non-volatile resistance changes are caused by the oxidation and reduction of the metal ions by the application of low bias voltages. Key attributes are low voltage, low current, rapid write and erase, good retention and endurance, and the ability for the storage cells to be physically scaled to a few tens of nm. The material class for the dielectric film or the solid electrolyte is comprised of oxides, higher chalcogenides (including glasses), semiconductors, as well as organic compounds including polymers.

CBRAM is a strong emerging memory candidate primarily due to scalability (~10 nm),[64] ultra-low energy operations due to fast read, write and erase times, and low voltage requirements.[65] Maturity of the CBRAM technology development can be assessed by the fact that many companies are either shipping products based on CBRAM or are in advanced stages of commercialization. Recent publications show CBRAM technology application in various markets including SSDs,[66] embedded NVM,[67] and serial interface non-volatile memory replacement.[68] In 2012, a CBRAM-based serial NVM replacement product became commercially

available.[69] In 2014, a 16 GB CBRAM array based on a CuTe CBRAM cell was demonstrated.[70,71] Such efforts are critical to identify core technology challenges[72] and fundamental materials and mechanisms.[73] Novel applications such as reconfigurable switch[74] and synaptic elements in Neuromorphic systems[75] based on CBRAM are also gaining prominence and are expected to expand the application base for this technology.

As with other filamentary ReRAM technologies, CBRAM is challenged by bit level variability,[72] the random nature of reliability failure such as retention or endurance, and random telegraph noise potentially contributing to read disturbs.[76] Such issues require large populations of bits to be studied, which suggests collaboration between universities and industry may be beneficial. Focus on fundamental understanding and simultaneously addressing some mitigation path such as error correction schemes, redundancy and algorithm development would enable closing the technology gap.

Engineering hurdles include the availability and integration of new materials used in CBRAM at advanced process nodes especially when there could be issues with compatibility of thermal budgets and process tooling. The availability of integrated array level information suggests that some of these challenges are being resolved in the recent years.[68,74] Active participation from semiconductor equipment vendors and material suppliers would assist in overcoming manufacturing hurdles rapidly.

2.2.4. MACROMOLECULAR (POLYMER) MEMORY

Macromolecular memory is a category of memory which focuses structures incorporating a layer of polymer, and the polymer may contain nano-particles, small molecules and nanoparticles that are sandwiched between two metal electrodes. This structure allows two different stable electrical states controllable through an external electrical voltage. These two stable electrical states, which are often called ON and OFF states (or 0 and 1), exhibit resistive, ferroelectric or capacitive natures according to the physical properties of the sandwich. The first fully-organic memory devices, based on nano-composite (a blend of poly-vinyl-phenol (PVP) and Bucky-ball (C60)) was presented in Materials Research Society.[77] Around the same time, memory devices using gold nano-particles and 8-hydroxyquinoline, dispersed in a polystyrene matrix, were also demonstrated.[78] Since then, the interest to use an admixture of nano-particles, small molecules and polymers in the manufacture of electronic memory devices, is on the rise. Non-volatile memory effects with a non-destructive read have been reported for a surprisingly large variety of polymeric/organic materials and blends of polymers with nanoparticles and molecules. Unlike the other four categories, this category is based on the material used in the switching layer(s) of the cell, but the mechanism in not specified. Both bipolar and unipolar (all pulses of the same polarity) switching have been demonstrated. Macromolecular ReRAM may have a mechanism placing it in one of the four main ReRAM categories listed above. However, the other mechanisms behind the electrical bistability, such as capacitive and ferroelectric, have also been reported.

Depending on the structure of the polymer, a variety of mechanisms can be operative. For polymers supporting transport of inorganic ions, formation of metallic filaments is reported. In semiconducting polymers supporting ion transport, dynamic doping due to migration of inorganic ions occurs. Ferroelectric polymers in blends with semiconducting material give rise to a memory effect-based modification of charge injection barriers by the ferroelectric polarization. However, for many polymeric materials, the origin of the resistive switching is not well understood. To date no specific design criteria for the polymer are known, although clear correlations between memory effect and electronic properties of the polymer have been demonstrated.

Stability of the memory states at high temperatures (85°C, 2×10^4 s) has been demonstrated.[79,80] Programming at very low power (70 nW) has been realized.[80] Assuming a 15 ns switching time determined for the same system, one might achieve a write energy of 6×10^{-15} J/bit. Furthermore, low programming voltages have been realized: +1.4 and -1.3 V for the two states with good retention time (>10^4 s).[81] Downscaling of polymer resistive memory cell to the 100 nm length scale has been reported.[82] At this length scale, integration of memory cells into an 8×8 array could be shown. Polymer memory cells on a flexible substrate have been shown.[83] For amorphous carbon, downscaling to nanometer sized cells has been published (1×10^3 nm^2).[84] Using carbon nanotubes as macromolecular electrodes and aluminum oxide as interlayer, isolated, non-volatile, rewriteable memory cells with an active area of essentially 36 nm^2 have been achieved, requiring a switching power less than 100 nW, with estimated switching energies below 10 fJ per bit.[85] With regards to the mechanism of operation, extensive work on the class of polyimide polymers has shown clear correlations between electronic structure of the polymer and memory effects, although a comprehensive picture for the operation has not yet emerged. A number of studies have indicated an active role of the interface between macromolecular material and (native) oxide layers in the operation of the memory involving charge trapping.[86,87] The recent and past studies show resistive,[88] capacitive (charge storage, based on electric dipole formation)[89,90] and ferroelectric behavior[91] of such devices. Thus, there is a need to open up a further discussion on the right pathway to realize such memory.

In macromolecular memory, a large variety of operation mechanisms can be operative. A key research question concerns distinguishing different mechanisms and evaluating the potential and possibilities of each mechanism. A second subsequent step would be to identify model systems for each mechanism. Having such a model system then provides a possibility to benchmark the operation of the macromolecular materials. These research steps would be crucial for establishing and securing the collaboration of the chemical industry; for design, synthesis and development of the next generation macromolecular materials

for memory applications, clear guidelines on the required structural and electronic properties of the macromolecular material are needed. For instance, memory effect originating for metallization and formation of metallic filaments requires macromolecular materials that support transport of ions and have appropriate internal free volume for ion conduction. Here the field could benefit from interaction with the field of polymer batteries. Ferroelectric polymers have been shown to give rise to resistive memory[92] and could benefit enormously from development of new macromolecular polymeric materials with combined ferroelectric switching and semiconducting structural units. Finally, a number of macromolecular memories involve oxide layers. Here mutually beneficial interaction with the (research) community on metal oxide ReRAM switching could spring, because at the macromolecular / oxide interface trap states can be engineered by tuning the electron levels of the macromolecular material.

In a nutshell, this area certainly needs an attention from theoretical physicists, materials scientists, chemists and device engineers. There are a number of issues that need to be addressed before we can embark on extending these devices to the real world. Such issues involve understanding of the electrical bistability mechanism in nano-composite (there are a number of contradicting theories), maintaining the difference between low and high conduction states for a longer period time by ensuring the stability of the high and low states, selecting environmentally friendly materials required for fabrication of nano-composite/polymer materials, and developing a cost-effective methodology for the fabrication of devices.

It is not possible to replace silicon-based memory devices in the foreseeable future. However, there are a number of other applications where "cheap" electronic memory devices can play a vital role. For example, nano-composite based memory devices can be directly printed on medicine bottles/packages and the information about the patient and schedule of taking medicine can be stored on the printed device.

2.2.5. FERROELECTRIC MEMORY

Coding digital memory states by the electrically alterable polarization direction of ferroelectrics has been successfully implemented and commercialized in capacitor-based Ferroelectric Random Access Memory (covered in Table BC2.1). However, in this technology the identification of the memory state requires a destructive read operation and largely depends on the total polarization charge on a ferroelectric capacitor, which in terms of lateral dimensions is expected to shrink with every new technology node. In contrast to that, alternative device concepts, such as the ferroelectric field effect transistor (FeFET) and the ferroelectric tunnel junction (FTJ), allow for a non-destructive detection of the memory state and promise improved scalability of the memory cell. The current status of and key challenges for these emerging ferroelectric memories will be assessed within this section.

2.2.5.1. FERROELECTRIC FET

The FeFET is best described as a conventional MISFET that contains a ferroelectric oxide in addition to or instead of the commonly utilized SiO_x, SiON or HfO_2 insulators. The former case requires the direct and preferably epitaxial contact of the ferroelectric to the semiconductor channel (metal-ferroelectric-semiconductor-FET, MFSFET), whereas the latter and commonly applied case maintains a buffer layer between the channel material and the ferroelectric (metal-ferroelectric-insulator-semiconductor-FET, MFISFET). When additionally introducing a floating gate in-between the buffer layer and the ferroelectric, a metal-ferroelectric-metal-insulator-semiconductor structure (MFMISFET) may be obtained that shares its equivalent circuit representation with the MFISFET approach. By applying a sufficiently high voltage pulse to the gate of the FeFET (i.e. voltage drop across the ferroelectric layer larger than its coercive voltage V_c), the polarization direction of the ferroelectric can be set to either assist in the inversion of the channel or to enhance its accumulation state. This results in a polarization dependent shift of the threshold voltage V_T, which allows for a non-destructive read operation and a 1T memory operation comparable to that of FLASH devices.

In order to assess the material and device requirements for a reliable and scalable FeFET technology the following two intrinsic relations in a ferroelectric gate stack need to be considered. First it is important to note that the extent of the aforementioned V_T-shift (memory window) in FeFET devices is primarily determined by the V_C of the implemented ferroelectric rather than by its remnant polarization P_r.[93] This results in a scaling versus memory window trade-off as V_c is proportional to the coercive field E_c and thickness d_{FE} of the ferroelectric. The inability of the commonly utilized perovskite-based FeFETs to laterally scale beyond the 180 nm node is therewith not solely based on the insufficient thickness scaling of perovskite ferroelectrics,[94,95] but rather due to their low E_c (SBT: 10-100 kV/cm, PZT: ~50 kV/cm, summarized in[96]) that in order to maintain a reasonable memory window requires compensation by a large d_{FE}. A solution to this scaling retardation is provided by the high coercive field (1-2 MV/cm) and thickness scalable FE-HfO_2.[97] This CMOS-compatible material innovation enabled the demonstration of a FeFET technology scaled to the 28 nm node utilizing a conventional HKMG technology and is already used in high volume production.[98] The close resemblance of the HKMG transistor and the FE-HfO_2-based memory transistor proves especially useful for the realization of an embedded memory solution with greatly reduced mask counts as compared to embedded FLASH.

The second noteworthy and important characteristic of the FeFET gate stack is related to its intrinsic capacitive voltage divider, which causes a significant gate voltage drop and buildup of electric field not only across the ferroelectric, but also across the non-

ferroelectric insulator in the gate stack. When additionally considering the incapability of the linear insulator to fully compensate the polarization charge of the ferroelectric layer, it becomes apparent that even in the case of no external biasing the capacitive voltage divider leads to a buildup of a permanent electric field. The so-called depolarization field building up in the ferroelectric is opposed to the polarization direction of the ferroelectric and to the electric field induced in the insulator[99]. The capacitive voltage divider is therefore directly responsible for the retention loss during stand-by as well as for the gate voltage distribution and the corresponding charge injection during write operations. This retention and endurance critical distribution of the electric field within the gate stack may be optimized by choosing the insulator capacitance as high as possible and the ferroelectric capacitance as low as possible. In the perovskite-based FeFET this is achieved by utilizing high-k buffer layers and is additionally fostered by the unavoidably large physical thickness of the perovskite ferroelectrics.[4, 100]. In the case of the aggressively scaled FE-HfO_2-based FeFET, the small thickness of the ferroelectric is compensated by the comparably low permittivity of HfO_2, the possibility to use ultra-thin interfacial layers, and by the depolarization resilience of the high E_c.[96,101] This leads to the situation that despite the markedly different stack dimensions and materials used, the electrically obtained characteristics are quite similar. Fast switching speed (≤100 ns), switching voltages in the range of 4-6 V, and 10-year data retention and endurance in the range of 10^{12} switching cycles have been demonstrated for FE-HfO_2-[97,98,102,103] as well as for perovskite-based FeFETs.[93,104,105] In the case of cycling endurance, however, the high E_c of FE-HfO_2 and the correspondingly large electric field in the insulator facilitates charge trapping during write operation, which was identified as the root cause for the limited endurance of 10^5 cycles observed in FE-HfO_2-based FeFETs with ultra-thin interfacial layer enabling excellent data retention.[106] Nevertheless, in an alternative approach utilizing a thicker insulator and sub-loop operation it was demonstrated that at the cost of retention a cycling endurance $>10^{12}$ may still be obtained.[102] In the current stage of development this endurance versus retention trade-off may be tailored, spanning the application range from embedded NOR-FLASH replacement with high retention requirements to low refresh rate 1T DRAM requiring high cycling endurance.

Entirely overcoming this endurance versus retention trade-off will require an improved stack design that may include a tailored polarization hysteresis (low P_r and high P_r/P_s ratio)[93], a reduced trap density at the interfaces,[106] an optimized capacitive voltage divider by area scaling in the MFMISFET approach[107] or the realization of a MFSFET device by implementing recent breakthroughs in the epitaxial growth of FE-HfO_2.[108] Despite promising results obtained for perovskite-based FeFET devices implemented into 64Kb NAND-Arrays at a feature size of 5 µm[105], little is known about the variability and array characteristics of FeFET devices scaled to technology nodes approaching the grain or domain size of the implemented ferroelectrics. Initial investigations on phase and grain distribution in doped HfO_2 based ferroelectric thin films and the effects of such granularity on device level characteristics of scaled FeFETs (such as on the statistical nature of switching) have recently been reported in Refs. [109,110,111]. Recently, 64 kb and 32 Mb FeFET arrays were demonstrated in the 28 nm[112] and the 22 nm FD-SOI CMOS platform,[113] respectively—in each case, a clear low and high V_T separation at the array level was demonstrated. Nevertheless, in order to fully judge the variability of ferroelectric phase stability at the nanoscale and to guide material optimization and fundamental understanding of the phenomenon, larger array statistics in the kB to Mb range and high-resolution PFM data will be required. Besides, recent demonstration of non-volatile memory operation based on antiferroelectricity—a phenomenon closely related to ferroelectricity—in work-function engineered ZrO_2 thin film capacitors may allude to new way of addressing and potentially solving some of these challenges in FeFETs.[114]

2.2.5.2. FERROELECTRIC TUNNEL JUNCTION

The ferroelectric tunnel junction, a ferroelectric ultra-thin film commonly sandwiched by asymmetric electrodes and/or interfaces, exhibits ferroelectric polarization induced resistive switching by a non-volatile modulation of barrier height. With the tunneling current depending exponentially on the barrier height, the ferroelectric dipole orientation either codes for a high or a low resistance state in the FTJ, which can be read out non-destructively. The resulting tunneling electroresistance (TER) effect of FTJs, the ratio between HRS and LRS, is usually in the range of 10 to 100 ([115] and references therein). However, giant TER of > 10^4 has most recently been reported in a super-tetragonal $BiFeO_3$ based FTJ by Yamada et al.[116] A similarly high TER was demonstrated by Wen and co-workers[117] for a $BaTiO_3$ tunnel barrier by replacing one metal electrode of the FTJ with a semiconducting electrode. With this new junction design, the modulation of tunneling current does not only rely on barrier height, but due to a variable space charge region in the semiconductor, also on a barrier width modification. With these most recent findings, two strategies to achieve giant TER have been identified: either use a ferroelectric barrier with a large polarization such as $BiFeO_3$ or use a semiconductor as electrode material to modulate the barrier width by field-induced carrier depletion.

The MFM-based structure of FTJs may be able to enable a retention time (> 10 years) and very high cycling endurance (> 10^{14}) properties of conventional FRAM. Nonetheless, in order to have a significant tunneling current, ferroelectric films in FTJs usually have a thickness ranging from several unit cells to ~5 nm, which is much thinner than in commercialized 1T-1C FRAM (> 50 nm). Due to larger interface contributions and increased leakage currents at reduced thickness, experimental data of these material systems might strongly deviate from their thick film behavior and need to be assessed separately.[118] However, even though only limited data are available up to this point, promising single cell characteristics have already been demonstrated, such as the most recent demonstration of $4x10^6$ endurance cycles and extrapolated data retention of 10 years at room temperature for a $BiFeO_3$-

based FTJ.[119] In the context of retention, it should be noted that despite improved TER, the newly proposed MFS-FTJ structure will give rise to a depolarization field, which will most likely degrade memory retention in a similar manner as described for the FeFET in Section 2.2.5.1. The highly energy efficient electric field switching, common to all ferroelectric memories, enables fast (10 ns[120]) and low voltage (1.4 V[119]) switching in FTJ devices and results in a minimal power consumption during write operation (1.4 fJ/bit, calculation based on the device characteristics given[121]). Due to the availability of non-destructive read-out and the further reduced ferroelectric thickness in FTJ devices as compared to conventional FRAM, improved voltage scaling and total energy consumption may be expected from this technology.

Similar to most other two-terminal resistor-based memories with insufficient self-rectification, the elimination of sneak currents in large crossbar arrays is most efficiently suppressed utilizing 1T-1R or 1D-1R cell architecture. In terms of scaling, this two-element memory cell, as well as the scalability of the selector device itself, has to be considered.[122] Simply based on the lateral dimensions of the FTJ element (assuming unlimited scalability of the selector), scaling below 50 nm^2,[121]based on PFM data,[123] most likely appears possible. However, with further scaling a simultaneous enhancement of the LRS current density is required to maintain readability in massively parallel memory architectures. A recent breakthrough of $1.4x10^5$ A/cm^2 current density at 300 nm feature size has been achieved by Bruno et al.[124] utilizing low resistivity nickelate electrodes. Based on these results maintaining 10 µA read current for feature sizes <100 nm appears possible. New FTJ concepts are also emerging; for example, engineered domain walls within the ferroelectric layer in an FTJ structure can lead to exotic quantum phenomenon such as resonant electron tunneling and quantum oscillations in the electrical conductance albeit at low temperatures.[125]

FTJ based memories are currently at a very early development stage, and most of the research activity is focused on perovskite-based ferroelectrics. Further investigations reaching beyond single device characterization will be needed to fully judge the scalability of FTJ as well as its MLC capability suggested in Ref. [126]. So far, no conclusions can be drawn on retention and statistical distribution of the polarization induced resistance states in large arrays. However, when considering the collective phenomenon of ferroelectricity with multiple dipoles contributing to a resistance change as opposed to filament-type resistive switching, advanced scalability may be expected. First results have shown that the FTJ is very similar to ReRAM in terms of electrical behavior and memory design, albeit distinct physical mechanisms. It should be noted that current prototypes could actually have both FTJ and ReRAM traits, as resistive switching is common among oxides including ferroelectric perovskites ([127] and references therein). For future development, the ferroelectric film in an ideal FTJ should be as thin as possible to allow scalability (while maintaining sufficient read current) and much less defective than that in ReRAM (e.g., with less oxygen vacancies), so that the mechanism of ferroelectric switching can dominate electrical behavior with little influences from mechanisms related to conducting filaments. The manufacturability of the rather complex electrode-perovskite ferroelectric-system of the FTJ concept will largely rely on the availability of high throughput and CMOS-compatible epitaxial growth techniques for large substrates or alternatively on the unrestricted feasibility demonstration of a polycrystalline FTJ. In this context it is worth noting that the CMOS-compatibility and advanced thickness scalability of ferroelectrics based on HfO_2 and its doped variant[103] as well as recent breakthroughs in its epitaxial growth[108] might yield great potential for the manufacturability of competitive FTJs. Experimental demonstrations of FTJs based on doped variant of HfO_2 were recently reported in Refs. [128,129,130].

2.2.6. MASSIVE STORAGE DEVICES

Device scaling has become a matter of strategic importance for modern and future information storage technologies, which motivates an exploration of unconventional materials with competitive performance attributes. By 2040 the conservative estimate of the worldwide amount of stored data is 10^24 bits, and the high estimate is ~10^29 bits[131] (these estimates are based on research by Hilbert and Lopez[132]). In nature, much of the data about the structure and operation of a living cell is stored in the molecule of deoxyribonucleic acid (DNA) and using nucleic acids molecules, such as DNA, for memory storage has been proposed. DNA has an information storage density that is several orders of magnitude higher than any other known storage technology: 1 kg of DNA stores 10^24 bits, for which $>10^9$ kg of silicon Flash memory would be needed.[131] Thus, a few tens of kilograms of DNA could meet all of the world's storage needs for centuries to come.

A number of recent studies have shown that DNA can support scalable, random-access and error-free information storage.[133,134,135] A state-of-the-art operating system developed by at the University of Washington with an industry partner is a DNA-based archival storage framework that supports random access from a DNA key-value store.[136] The DNA-stored files are compatible with mainstream digital format, and large-scale DNA storage up to 200 MB has been demonstrated.[137] There are still many unknowns regarding both DNA operations in cell and with regard to the potential of DNA technology for massive storage applications. DNA volumetric memory density far exceeds (10^3–10^7×) projected ultimate electronic memory densities. Also, in the living cell, the memory read/write operations occur at high speed (<100 µs/bit) and require very low energy of ~10^{-17} J/bit or 10^{-11} W/GB.[138] DNA can store information stably at room temperature for hundreds of years with zero power requirements, making it an excellent candidate for large-scale archival storage.[138] Also, DNA is an extremely abundant and totally recyclable material. Recently, a method for efficient encoding of information—including a full computer operating system—into DNA was presented, which approaches the theoretical maximum for information stored per nucleotide.[139] One of the goals for research

efforts is to demonstrate miniaturized, on-chip integrated DNA storage. New methods for DNA synthesis and sequencing are key components for these developments.

Two major categories of technical challenges remain:

- Physical Media: Improving scale, speed, cost of synthesis and sequencing technologies.
- Operating System: Creating scalable indexing, random access and search capabilities.

The key technological and scientific challenges are in improving performances beyond the life sciences industry. In the life science industry applications require perfect synthesis and perfect sequencing, while scale, throughput and cost are secondary considerations. For data storage, high read and write error rates can be tolerated, and information encoding schemes can be used. In this application, scale and throughput and cost are primary considerations. Current DNA storage workflows can take several days to write and then read data, due to reliance on life sciences technologies that were not designed for use in the same system. The demonstrated DNA write-read cycle is too slow and costly to support exascale archival data storage. Solving this problem will require: 1) Substantial reductions in the cost of DNA synthesis and sequencing, and 2) Deployment of these technologies in a fully automated end-to-end workflow.

2.2.7. MOTT MEMORY

Mott memory is a metal/insulator/metal capacitor structure consisting of a correlated electron insulator (or Mott insulator). Correlated electron insulators often show the electronic phase transition accompanied by a drastic change in their resistivity under external stimuli such as temperature, magnetic field, electric field, and light. Mott memory exploits this electronic phase transition (called Mott metal-to-insulator transition or Mott transition[140]) induced by an electric field. A mechanism of the Mott memory has been theoretically proposed in terms of the interfacial Mott transition induced by the carrier accumulation at a Schottky-like interface between a metal electrode and a correlated electron insulator.[141] The theory also predicted that the resistive switching due to the interfacial Mott transition has a non-volatile-memory functionality, because the Mott transition is a first-order phase transition due to its nature.[140] In addition, Mott memory based on the Mott transition involving a large number of carriers (more than 10^{22} cm^{-3}) has in principle an advantage in device scaling, because there are a sufficient number of carries for the Mott transition even in a nanoscale device. In an ideal Mott transition, the electrons localized due to the strong electron-electron correlation come to be itinerant, via the stimuli, such as application of an electric field, and so forth. It needs no dopants, and the mechanism withstands the miniaturization of the (silicon) devices.

The Mott transition induced by an electric field or carrier injection has been experimentally demonstrated in a correlated electron material of $Pr_{1-x}Ca_xMnO_3$.[142] After this demonstration, two-terminal devices such as switches and memories have been intensively studied using such correlated electron oxides as $Pr_{1-x}Ca_xMnO_3$,[143,144] VO_2,[145,146,147] $SmNiO_3$,[148] NiO,[149,150] Ca_2RuO_4,[151] and NbO_2,[152,153] and using Mott-insulator chalcogenides of AM_4X_8 (A=Ga, Ge; M=V, Nb, Ta; X=S, Se).[154,155,156,157] In addition to these inorganic materials, recently, reversible and non-volatile resistive switching based on the electronic phase change between charge-crystalline state and quenched charge glass has been demonstrated in the organic correlated materials of θ-(BEDT-TTF)$_2X$ (where X denotes an anion).[158]

$SmNiO_3$ exhibits a colossal (8 orders in magnitude) resistance jump by hydrogenation. The $SmNiO_3$ channel with the solid state proton gate has demonstrated the electric base gated large ON/OFF switching.[148] The trigger for switching is based on the proton intercalation by electric field, and the DFT calculation explains the large gap opening by additional electron doping via protonation and is the origin for colossal resistance jump phenomena.[159] These results indicate that the device using the metal – insulator (Mott) transition driven by the strong electron-electron correlation is powerful as well as appropriate for the switching devices.

Scalability has been demonstrated down 110 × 110 nm^2 in Mott memristors consisting of NbO_2 that shows the temperature-driven Mott transition from a low-temperature insulator phase to a high-temperature metal phase. The switching speed, energies, and endurance of the NbO_2-Mott memristors have been evaluated to be less than 2.3 ns, of the order of 100 fJ, and $>10^9$, respectively.[152,153] The programing and read voltages reported so far are <2 V and <0.2 V, respectively.[150] The non-volatile resistive switching of AM_4X_8 single crystals was induced by the electric field of less than 10 kV/cm.[154,155,156,157] This suggests that if the device consisting of a 10-nm-thick AM_4X_8 film is fabricated, the switching voltage will be less than 0.01 V.

Although non-volatile switching has been reported in the devices based on AM_4X_8 and θ-(BEDT-TTF)$_2X$, their retention characteristics are not elucidated in detail.[154,155,156,157,158] In addition, the NbO_2-Mott memristors and VO_2-based devices are volatile switch.[145,146,147,152,153] The retention is thus a major concern of Mott memory. In principle, the Mott transition can be driven even by a small amount of carrier doping to the integer-filling or half-filling valence states of the transition element.[140] However, because of disorders, defects, and spatial variation of chemical composition, a rather large amount of carriers of more than 10^{22} cm^{-3} are required to drive the Mott transition in actual correlated electron materials, resulting in a relatively large switching voltage required in the Mott memory. Therefore, one of the key challenges is the control of crystallinity and chemical

composition in the thin films of correlated electron materials, including the integration of the correlated electron materials onto Si platform. There are some theoretical mechanisms proposed for Mott memories such as the interfacial Mott transition[141] and the formation of conductive filament generated by local Mott transition.[154,156,157] However, a thorough understanding of the mechanism has not been achieved yet. Therefore, the elucidation of detailed mechanism is also a major research challenge.

2.3. MEMORY SELECT DEVICE

The *capacity* (or *density*) is one of the most important parameters for memory systems. In a typical memory system, memory devices (cells) are connected to form an array. A memory cell in an array can be viewed as being composed of two components: the '*storage node'*, which is usually characterized by an element with switchable states, and the '*selector'*, which allows the storage node to be selectively addressed for read and write. Both components impact scaling limits of memory. It should be noted that for several advanced concepts of resistance-based memories, the storage node could in principle be scaled down below 10 nm,[160] and the memory density is often limited by the selector devices. Thus, the selector device represents a serious bottleneck for emerging memory scaling to 10 nm and beyond.

The most commonly used memory selector devices are transistors (e.g., FET or BJT), as in DRAM, FRAM, *etc*. Flash memory is an example of a storage node (floating gate) and a selector (transistor) combined in one device. Planar transistors typically have the footprint around (6-8)F^2. In order to reach the highest possible 2D memory density of $4F^2$, a vertical transistor selector needs to be used. However, transistors as selector devices are generally unsuitable for 3D memory architectures. Two-terminal memory selector devices are preferred for scalability and can be used in crossbar memory arrays to achieve $4F^2$ footprint.[161,162] The function of selector devices is essentially to minimize leakage through unselected paths ("sneak paths"). Two-terminal selector devices can achieve this through asymmetry (e.g., rectifying diodes) or nonlinearity (e.g., nonlinear devices).[163] Volatile switches can also be used as selector devices. Figure BC2.2 shows a taxonomy of memory selector devices. In addition to external selector devices, some storage elements may have inherent self-selecting properties (e.g., intrinsic nonlinearity or self-rectification), which may enable functional crossbar arrays without external selectors.

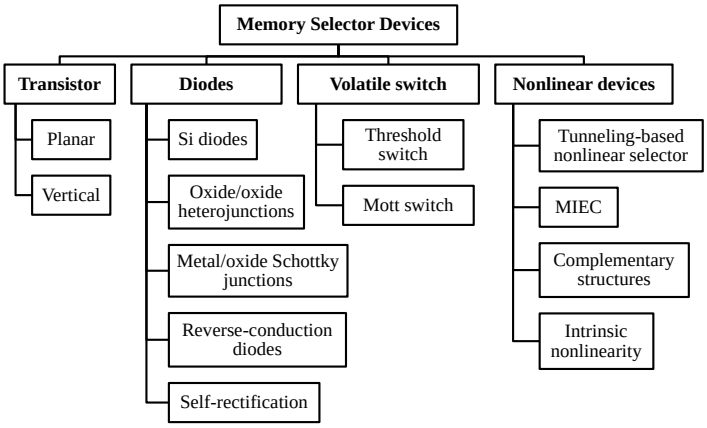


Figure BC2.2. Taxonomy of Memory Select Devices

2.3.1. VERTICAL TRANSISTORS

Several examples of experimental demonstrations of vertical transistors used as selectors in memory arrays are presented in Table BC2.3. While a vertical transistor selector allows for the highest planar array density ($4F^2$), it is challenging to integrate a transistor selector into stacked 3D memory. For example, to avoid thermal stress on the memory elements on the existing layers, the processing temperature of the vertical transistor in 3D stacks must be low. Also, making contact to the third terminal (gate) of vertical transistors constitutes an additional integration challenge, which usually results in cells size larger than ($4F^2$);[164] although true $4F^2$ arrays can, in principle, be implemented with 3-terminal selector devices.[165]

Table BC2.3 Experimental Demonstrations of Vertical Transistors in Memory Arrays

2.3.2. DIODE SELECTOR DEVICES

General requirements for two-terminal selector devices are sufficient ON currents at proper bias to support read and write operations and sufficient ON/OFF ratio to enable selection. The minimum ON current required for fast read operation is ~1 μA (Table BC2.4). The required ON/OFF ratio depends on the size of the memory block, $m \times m$: for example using a standard scheme

of array biasing the required ON/OFF ratio should be in the range of 10^7–10^8 for m=10^3–10^4, in order to minimize the 'sneak' currents.[166] These specifications are quite challenging, and the experimentally demonstrated selector devices can rarely meet them. Thus, selector devices are becoming a critical challenge of emerging memory. It should be noted that different application targets of resistance-based memories also impact selector device requirements.

Table BC2.4 Benchmark Select Device Parameters

The simplest realization of diode selectors uses semiconductor diodes, such as a *pn*-junction diode, Schottky diode, or heterojunction diode. Such devices are suitable for a unipolar memory cell. For bipolar memory cells, selectors with bi-directional switching are needed. Proposed examples include Zener diodes,[167] BARITT diodes,[168] and reverse breakdown Schottky diodes.[169]

2.3.2.1. Si diode Selector Devices

Both single-crystal Si[170] and poly-Si[171,172,173] diodes have been developed as selector devices for PCM arrays. To provide high ON current, the contact resistivity needs to be reduced to $<10^{-7}$ Ω cm^2, which was achieved by engineering the metal electrodes and electrode-Si interface.[171] A short-time annealing technique helps to reduce the OFF current and enlarge the ON/OFF ratio. Poly-Si technology can achieve ON current density of 10^7 A/cm^2 (at ~ 1.8V) and an ON/OFF ratio of 10^8. It is believed that Si diodes can be scaled beyond 20 nm or 10 nm. Poly-Si diode selector devices have been integrated in PCM crossbar arrays, 3D vertical chain-cell type PCM,[172] and a 1 Gb PCM test chip.[173] A major challenge of Si diodes is the high processing temperature (above 1000°C) required to crystallize Si to reduce contact resistivity and OFF current.

2.3.2.2. Oxide Diode Selector Devices

Oxide-based heterojunction[174,175,176,177] or Schottky junction[180,181,182,183] diodes may be fabricated at lower temperatures and used as selector devices. They are particularly suitable for oxide-based RRAM devices. A p-NiO_x/n-TiO_x diode has demonstrated a rectification ratio of 10^5 at ±3V and ON current density of 5×10^3 A/cm^2 (at ~ 2.5V)[174]. A p-CuO_x/n-$InZnO_x$ diode achieved higher ON current density of 10^4 A/cm^2 (at ~ 1.3V) and was integrated with NiO_x RRAM in a 2-layer 8×8 crossbar array[175,176] and with Al_2O_3 antifuse in a one-time-programmable (OTP) memory.[177] Si substrates can be used as a part of heterojunction diodes as demonstrated in n-ZnO/p-Si[178] and n-Ge-nanowire/p-Si diodes.[179] In TiO_x-based diodes with Pt electrodes, temperature-dependent current-voltage (I-V) characteristics confirm a Schottky barrier of ~0.55 eV at the TiO_x/Pt interface.[180] The rectification ratio is ~ 1.6×10^4 at ±1V but ON current density is low (~ 13A/cm^2) due to large size. A Pt/TiO_2/Ti diode with Pt as the Schottky contact and Ti the ohmic contact achieved higher rectification ratio of 10^7 – 10^9 at ±1V.[181] Another demonstration of Pt/TiO_2/Ti Schottky diodes improved ON current density to ~ 3×10^5 A/cm^2 at 2 V on a 4 μm^2 area.[182] Measurement showed that current is not uniform across the diode area, probably due to edge leakage. Therefore, current density is higher at smaller diode size. An Ag/n-ZnO Schottky diode with non-alloyed Ti/Au ohmic contact demonstrated a rectification ratio of 10^5 and forward current density over 10^4 A/cm^2 at 2 V.[183] In addition to oxide Schottky diodes, Si Schottky diodes are also utilized as selector devices, e.g., Al/p-Si.[184] The ON current of oxide-based heterojunction diodes is often limited by both contact resistance and density of states of the oxide materials.

2.3.3. Volatile Switch as Selector Devices

Volatile resistive switching devices can also be utilized as selector devices. They provide access to a selected memory element in their ON state and block sneak paths in OFF state. The device structure and physics of operation of these devices are sometimes similar to those of the storage nodes. The main difference is that nonvolatility is required for the storage node, while for select devices volatile switching characteristics allow them to be switched quickly between ON and OFF states.

2.3.3.1. Mott Switch

This device is based on metal-insulator transition (i.e., Mott transition) and exhibits a low resistance above a critical electric field (threshold voltage, V_{th}). It recovers to a high-resistance state if the voltage is below a hold voltage (V_{hold}). If the electronic conditions that triggered Mott transition can relax within the memory device operation time scale, the Mott transition device is essentially a volatile resistive switch and can be utilized as a selector device. A VO_2-based device has been demonstrated as a selector device for NiO_x RRAM element.[185] However, the feasibility of Mott-transition switches as selector devices still needs further research. It should be noted that VO_2 undergoes a phase transition to the metallic state at temperature around 68°C, which restricts its operation temperatures and limits practical applications of a VO_2 selector as current specifications require operational temperature of 85°C. Suitable Mott materials with higher transition temperatures need to be investigated. Metal insulator transitions at ~ 130°C, and electrically driven switching were observed in thin films of $SmNiO_3$ [186].

2.3.3.2. THRESHOLD SWITCH

This type of device is based on the threshold-switching effect observed in thin-film MIM structures caused by electronic charge injection. Significant resistance reduction occurs at V_{th}, and this low-resistance state quickly recovers to the original high-resistance state when the applied voltage falls below V_{hold}. It was reported that chalcogenide-based threshold switches could be used as access devices in PCM arrays.[187] Niobium oxide is found to possess both memory switching and threshold switching properties at different compositions, based on which hybrid memory (W/bi-layer-NbO_x/Pt) was demonstrated in a 1kb array.[188] NbO_x-based selectors have also been integrated with TiO_x/TaO_x based RRAM in crossbar arrays at 5xnm node.[189] In Si-As-Te ternary alloy, the composition (controlled by the sputtering power during deposition) determines the emergence of threshold switching.[190] Both V_{th} and V_{hold} vary with composition, which may provide a method to optimize the selector device operation window. Another threshold switch device based on chalcogenide AsTeGeSiN was shown to be scalable to 30nm with current density exceeding 10MA/cm^2 and endurance over 10^8 cycles.[191,192] It was integrated with TaO_x-based RRAM devices. An unavoidable finite delay time was found in this selector device due to intrinsic properties of the chalcogenide material, which may limit the selector speed. Another doped-chalcogenide (material undisclosed) based selector demonstrates low V_{hold} (0.2 V), large ON/OFF ratio (>10^7), fast speed (<10 ns), long endurance (>10^9 cycles) and good thermal stability (180°C).[193] A so-called "FAST" (Field Assisted Superliner Threshold) selector was recently reported, with abrupt switching (<5 mV/dec), high ON/OFF ratio (10^7), and long endurance (10^8 cycles).[194] Unlike other threshold switch selectors, this device does not exhibit recovery to OFF-state at certain V_{hold}, which appears more like a nonlinear selector. A 4Mb 1S1R crossbar RRAM array has been demonstrated based on this selector.

2.3.4. NONLINEAR SELECTOR DEVICES

Similar to volatile switches, nonlinear selector devices can be used with bipolar memory elements, which is an advantage over rectifying diode selectors.

2.3.4.1. NONLINEAR SELECTOR DEVICES

Nonlinearity in device characteristics can be introduced with non-ohmic transport mechanisms, e.g., tunneling. A Ni/TiO_2/Ni nonlinear selector device is integrated with HfO_2-RRAM to demonstrate a 1S1R memory structure.[195] Another selector device with Pt/TiO_2/TiN structure is combined with a bi-layer Pt/TiO_{2-x}/TiO_2/W RRAM for a functional memory device.[196] A so-called "varistor" selector device is based on a sandwiched TaO_x/TiO_2/TaO_x structure.[197] It was found that the substitution of Ti^{4+} in TiO_2 by Ta^{5+} ions increases the conductivity of the initially insulating TiO_2 layer. The ON-current of nonlinear selector devices can be modulated by oxide thickness and oxidation conditions. Another multi-oxide stack (Ta_2O_5/TaO_x/TiO_x) based selector also leverages various interface engineering techniques to obtain high ON-current density (>10^7 A/cm^2), high nonlinearity ratio (~10^4), and low OFF-current (<100 nA).[198] It is integrated with CBRAM in a 1 kb crossbar array. A back-to-back diode structure, $n^+/p/n^+$ poly-Si, is also suggested as a selector device, where the middle p-layer is fully depleted and a drain induced barrier lowering (DIBL) effect causes exponential current increase with applied voltage.[199]

2.3.4.2. MIEC SWITCH

The device is made from Cu-containing "mixed ionic and electronic conduction" (MIEC) materials sandwiched between an inert top electrode (TE) (e.g., TiN, W) and a bottom electrode (BE). Negative voltage applied on TE pulls Cu^+ in MIEC away from the BE and creates vacancies near the BE. The hole and vacancy concentrations depend exponentially on the applied voltage. Symmetrical diode-like I-V characteristics are achieved with two inert electrodes. Large fraction of mobile Cu^+ enables high current density (> tens of MA/cm^2).[200] Endurance above 10^8 cycles has been demonstrated on MIEC devices in small arrays.[201] The MIEC selector devices were also integrated with PCM in a 512 kb testing array using 180nm CMOS process.[202] The scalability of MIEC select devices was tested to below 30nm in diameter and below 12nm in thickness.[203]

2.3.4.3. COMPLEMENTARY RESISTIVE SWITCHES

A complementary resistive switch (CRS) provides a self-selecting memory by connecting two bipolar RRAM devices anti-serially.[204] It may be considered as "constructed nonlinearity". Both states "0" and "1" have high resistance in CRS, which helps to minimize leakage through sneak paths. In either state, one of the two RRAMs is in LRS and the other in HRS. When reading a "1" state, the HRS device is switched to LRS and both devices end up in LRS. When reading a "0" state, no switching occurs and CRS remains in HRS. Notice that the reading operation is destructive, although non-destructive readout method was also proposed.[205] CRS has been demonstrated in different resistive switching devices, e.g., Cu/SiO_2/Pt bipolar resistive switches,[206] amorphous carbon-based RRAM,[207] TaO_x–based RRAM,[208] multi-layer TiO_x device,[209] HfO_x RRAM,[210] ZrO_x/HfO_x bi-layer RRAM,[211] Cu/TaO_2 atomic switch,[212] Nb_2O_{5-x}/NbO_y RRAM,[213] *etc.*

Table BC2.5a summarizes experimentally demonstrated parameters of some two-terminal select devices, including diodes, volatile switches, and nonlinear devices. Table BC2.5b summarizes the parameters of some reported self-rectifying memories. It should be emphasized that these summary tables can only capture a snapshot of selector device characteristics; however, the functionality of these devices depends on their actual voltage in arrays with random data patterns and the balance between

selectors and storage elements. Therefore, these parameter tables should only be used for illustration purpose, not for rigorous benchmark or assessment.

It remains a great challenge for the demonstrated selector devices to meet all the requirements in Table BC2.4. For scaled two-terminal select devices, two fundamental challenges are *contact resistance*[171] and *lateral depletion effects.*[214,215] Very high doping concentration is needed to minimize both effects. However, high doping concentrations result in increased reverse bias currents in classical diode structures and therefore reduced I_{on}/I_{off} ratio. For switch-type selector devices the main challenges are identifying the right material and the switching mechanism to achieve the required drive current density, I_{on}/I_{off} ratio, and reliability.

Table BC2.5a Experimentally Demonstrated Two-terminal Memory Select Devices

Table BC2.5b Experimentally Demonstrated Self-selecting Memory Devices (self-rectifying)

2.4. STORAGE CLASS MEMORY

2.4.1. STORAGE CLASS MEMORY DEVICES

2.4.1.1. TRADITIONAL STORAGE: HDD AND FLASH SOLID-STATE DRIVES

Conventionally, magnetic hard-disk drives are used for non-volatile data storage. The cost of HDD storage in $/GB is extremely low and continues to decrease. Although the bandwidth with which contiguous data can be streamed is high, the poor random-access time of HDDs limits the maximum number of I/O requests per second (IOPs). In addition, HDDs have relatively high energy consumption, a large form factor, and are subject to mechanical reliability failures in ways that solid state technologies are not. Despite these issues, the sheer number and growth in HDD shipments per year (380,000 Petabytes in 2012, growing at 32% per year) means that magnetic disk storage is highly unlikely to be "replaced" by solid-state drives at any time in the foreseeable future.[216]

Non-volatile semiconductor memory in the form of NAND Flash has become a widely used alternative storage technology, offering faster access times, smaller size and lower energy consumption when compared to HDD. However, there are several serious limitations of NAND Flash for storage applications, such as poor endurance (10^3–10^5 erase cycles), only modest retention (typically 10 years on a new device, but only 1 year at the end of rated endurance lifetime), long erase time (~ms), and high operating voltage (~15 V). Another difficult challenge of NAND Flash SSD is posed by its page/block-based architecture. By not allowing for direct overwrite of data, sophisticated procedures for garbage collection, wear-leveling and bulk erase are required. This in turn requires additional computation – which reduces performance and increases cost and power because of the need for a local processor, RAM, and logic – as well as over-provisioning of the SSD which further increases cost per effective user-bit of data.[217]

Although Flash memory technology continues to project for further density scaling, inherent performance characteristics such as read, write and erase latencies have been nearly constant for more than a decade.[218] While the introduction of multi-level cell (MLC) Flash devices extended Flash memory capacities by a small integral factor (2–4), the combination of scaling and MLC have resulted in the degradation of both retention time and endurance, two parameters critical for storage applications. The migration of NAND Flash into the vertical dimension above the silicon has continued this trend of improving bit density (and thus cost-per-bit) while maintaining or in some cases, even slightly degrading the latency, retention, and endurance characteristics of present-day NAND Flash.

This outlook for existing technologies has opened interesting opportunities for prototypical and emerging research memory technologies to enter the non-volatile solid-state-memory space.

2.4.1.2. WHAT IS STORAGE CLASS MEMORY?

Storage-class memory (SCM) describes a device category that combines the benefits of solid-state memory, such as high performance and robustness, with the archival capabilities and low cost of conventional hard-disk magnetic storage.[219,220] Such a device requires a non-volatile memory (NVM) technology that could be manufactured at a very low cost per bit.

A number of suitable NVM candidate technologies have long received research attention, originally under the motivation of readying a "replacement" for NAND Flash, should that prove necessary. Yet the scaling roadmap for NAND Flash has progressed steadily so far, without needing any replacement by such technologies. So long as the established commodity continues to scale successfully, there would seem to be little need to gamble on implementing an unproven replacement technology instead.

However, while these NVM candidate technologies are still relatively unproven compared to Flash, there is a strong opportunity for one or more of them to find success in applications that do not involve simply "replacing" NAND Flash. Storage Class Memory can be thought of as the realization that many of these emerging alternative non-volatile memory technologies can potentially offer significantly *more* than Flash, in terms of higher endurance, significantly faster performance, and direct-byte access capabilities. In principle, Storage Class Memory could engender two entirely new and distinct levels within the memory and storage hierarchy. These levels would be differentiated from each other by access time, with both levels located within more than two orders of magnitude between the latencies of off-chip DRAM (~80 ns) and NAND Flash (20 µs).

2.4.1.2.1. STORAGE-TYPE SCM

The first new level, identified as S-type storage-class memory (S-SCM), serves as a high-performance solid-state drive, accessed by the system I/O controller much like an HDD. S-SCM must provide at least the same data retention as Flash, allowing S-SCM modules to be stored offline, while offering new direct overwrite and random-access capabilities (which can lead to improved performance and simpler systems) that NAND Flash devices cannot provide. However, because of the modest (perhaps 10x) advantage in read latency over NAND Flash, it is critical that the eventual cost-per-bit for S-SCM be no worse than 3-10x higher than NAND Flash. While such costs need not be realized immediately at first introduction, it would need to be very clear early on that costs could steadily approach such a level relative to Flash.

Note however that such system cost reduction can come from other sources than the raw cost of the device technology: a slightly-higher-cost NVM technology that enabled a simple, low-cost SSD by eliminating or simplifying costly and /or performance-degrading overhead components would achieve the same overall goal. If the cost per bit could be driven low enough through ultrahigh memory density, ultimately such an S-SCM device could potentially replace magnetic hard-disk drives in enterprise storage server systems as well as in mobile computers (subject to the same issues mentioned above in terms of needing numerous IC fabs to ship the many petabytes of HDD delivered to those markets[216]).

2.4.1.2.2. MEMORY-TYPE SCM

The second new level within the memory and storage hierarchy, termed M-type storage-class memory (M-SCM), should offer a read/write latency of less than ~200 ns. These specifications would allow it to remain synchronous with a memory system, allowing direct connection from a memory controller and bypassing the inefficiencies of access through the I/O controller. The role of M-SCM would be to augment a small amount of DRAM to provide the same overall system performance as a DRAM-only system, while providing moderate retention, lower power-per-GB and lower cost-per-GB than DRAM. Again, as with S-SCM, the cost target is critical. It would be desirable to have cross-use of the same technology in either embedded applications or as a standalone S-SCM, in order to spread out the development risk of an M-SCM technology. The retention requirements for M-SCM are less stringent, since the role of non-volatility might be primarily to provide full recovery from crashes or short-term power outages, requiring non-volatility over a period of perhaps 7–21 days.

Particularly critical for M-SCM will be device endurance, since the time available for wear-leveling, error-correction, and other similar techniques is limited. The volatile portion of the memory hierarchy will have effectively infinite endurance compared to any of the non-volatile memory candidates that could become an M-SCM. Even if device endurance can be pushed well over 10^9 cycles, it is quite likely that the role of M-SCM will need to be carefully engineered within a cascaded-cache or other Hybrid Memory approach.[221] That said, M-SCM offers a host of new opportunities to system designers, opening up the possibility of programming with truly persistent data, committing critical transactions to M-SCM rather than to HDD, and performing commit-in-place database operations.

2.4.1.3. TARGET SPECIFICATIONS FOR SCM

Since the density and cost requirements of SCM transcend the straightforward scaling application of Moore's Law, additional techniques will be needed to achieve the ultrahigh memory densities and extremely low cost demanded by SCM, such as 1) 3-D integration of multiple layers of memory, currently implemented commercially for write-once solid-state memory,[222] and/or 2) Multiple level cell (MLC) techniques.

Table BC2.6 lists a representative set of *target* specifications for SCM devices and systems compared with benchmark parameters of existing technologies (HDD and NAND Flash). As described above, SCM applications can be expected to naturally separate based on latency. Although S-class SCM is the slower of these two targeted specifications, read and write latencies should be in the 1–5 µsec regime in order to provide sufficient performance advantage over NAND Flash. Similarly, endurance of S-class SCM should offer at least 1 million program-erase cycles, offering a distinct advantage over NAND Flash. In order to support off-line storage, 10-year retention at 85°C should be available.

In order to make overall system power usage (as shown in Table BC2.6) competitive with NAND Flash and HDD, and since faster I/O interfaces can be expected to consume considerable power, the device-level power requirements must be extremely minimal. This is particularly important since low latency is necessary but not sufficient for enabling high bandwidth – high parallelism is also required. This in turn mandates a sufficiently low power per bit access, both in terms of peripheral circuitry

and device-level write and read power requirements. Finally, standby power should be made extremely low, offering opportunities for significant system power savings without loss of performance through rapid switching between active and standby states.

In order to achieve the desired cost target of within 3–10x of the cost of NAND Flash, the effective areal density will similarly need to be quite similar to 1X-node planar NAND Flash. This low cost structure would then need to be maintained by subsequent SCM generations, through some combination of further scaling in lateral dimension, by increasing the number of multiple layers, or by increasing the number of bits per cell.

Also shown in Table BC2.6 are the target specifications for M-type SCM devices. Given the faster latency target (which enables coherent access through a memory controller), the program-erase cycle endurance must be higher, so that the overall non-volatile memory system can offer a sufficiently large lifetime before needing replacement or upgrade. Although some studies have shown that a device endurance of 1e7 cycles is sufficient to enable device lifetimes on the order of 3–10 years,[223] we anticipate that the need for sufficient engineering margin would suggest a minimum cycle endurance of 1e9 cycles. While such endurance levels support the use of M-class SCM in memory support roles, significantly higher endurance values (e.g., 1e12 or 1e14 cycles) would allow M-class SCM to be used in more varied memory applications, where the total number of memory accesses may become very large.

Table BC2.6 Target Device and System Specifications for SCM

Table BC2.7 Potential of Current Prototypical and Emerging Research Memory Candidates for SCM Applications

2.4.1.4. First SCM Products Reach the Market

In July 2015, a new non-volatile memory technology called "3D-Xpoint" was announced. This technology is said to offer 1000x lower latency and 1000x higher endurance than NAND Flash, at a density that is 10x higher than DRAM.[224,225] (Note that it is most likely that the latency referred to here is write latency rather than read latency, since NAND write latency is much slower than its read latency.) 3D-Xpoint technology, said to have been implemented at the 128Gbit chip level, is based on a two-layer stacked crossbar array, with each intersection point containing a non-volatile memory device and a nonlinear access device.[226,227] The particular non-volatile memory device was not specified, other than that it depends on bulk changes of resistance,[228] nor was the nonlinear access device described. Speculation based on patent searches and job solicitations suggest that the technology may be a combination of some variant of phase change memory and an Ovonic Threshold Switching access device.[228]

While the details of the devices involved may still be uncertain, the projected array specifications and the target applications are, for all intents and purposes, indistinguishable from those described above for S-type Storage Class Memory. Thus we can consider 3D-Xpoint as the first commercial implementation of the Storage Class Memory concept first described in 2008.[219,220] Furthermore, in a later presentation, a second, "Performance-focused" form of 3D-Xpoint memory was described as being under active development.[229] Compared to the initial "Cost-focused" form of 3D-Xpoint memory, this variant is said to offer even faster latencies and higher endurance (Figure BC2.3). Thus, this second variant of 3D-Xpoint is somewhat similar to M-type SCM as described above, both in terms of its specifications and in terms of potential applications. The only major difference, as observed in Figure BC2.3, is the strong similarity between the expected volatility of Performance-focused 3D-Xpoint and the known volatility of DRAM. In contrast, one of the benefits of M-type SCM was supposed to be its non-volatility. Ideally, retention of data for perhaps 1-3 weeks would permit successful recovery of server data, even after a power outage due to a natural disaster or other major event.

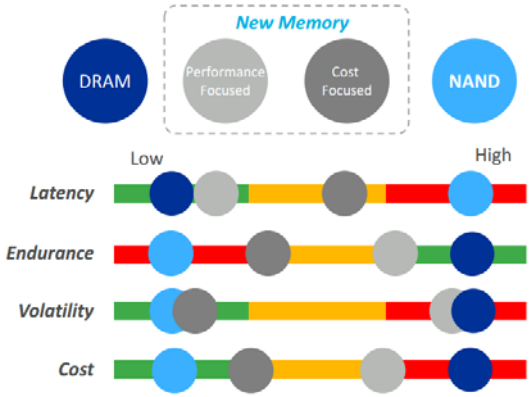


Figure BC2.3 Comparison of Performance of Different Memory Technologies

2.4.2. STORAGE CLASS MEMORY ARCHITECTURES

2.4.2.1. INTRODUCTION

In traditional computing, SRAM is used as a series of caches, which DRAM tries to refill as fast as possible. The entire system image is stored in a non-volatile medium, traditionally a hard drive, which is then swapped to and from memory as needed. However, this situation has been changing rapidly. Application needs are both scaling in size and evolving in scope, and have rapidly exhausted the capabilities of the traditional memory hierarchy.

By combining the reliability, fast access, and endurance of a solid-state memory together with the low-cost archival capabilities and vast capacity of a magnetic hard disk drive, Storage Class Memory (SCM) offers several interesting opportunities for creating new levels in the memory hierarchies that could help with these problems. SCM offers compact yet robust non-volatile memory systems with greatly improved cost/performance ratios relative to other technologies. S-class SCM represents ultra-fast long-term storage, similar to an SSD but with higher endurance, lower latencies, and byte-addressable access. M-class represents dense and low-power non-volatile memory at speeds close to DRAM.

In order to implement SCM, both the emerging memory technologies discussed in Section 2.4.1 as well as new interfaces and architectures will be needed, in order to fully use the potential and to compensate for the weaknesses of various new memory technologies. In this section, we explore the Emerging Research Architecture implications and challenges associated with Storage Class Memory.

2.4.2.2. CHALLENGES IN MEMORY SYSTEMS

Current memory systems range in size from Gigabytes (low-volume ASIC systems, FPGAs and mobile systems) through Terabytes (multicore systems that manage execution of many threads for personal or departmental computing), to Petabytes (for database, Big Data, cloud computing, and other data analytics applications), and up to Exabytes (next-generation, exascale scientific computing). In all cases, speed (both in terms of latency of data reads and writes as well as bandwidth), power consumption, and cost are absolutely critical. However, the importance of other system aspects can vary across these different application spaces.

Historically, roughly one-third of the power in a large computer system is consumed in the memory sub-system.[230] Some portion of this is refresh power, required by the volatile nature of DRAM. As a result, modern data servers consume considerable power even when operating at low utilization rates. For example, it has been reported that servers are typically operating at over 50% of their peak power consumption even at very low utilization rates.[231] The requirement for rapid transition to full operation precludes using a hibernate mode. As a result, a persistent memory that did not require constant refresh would be valuable.

Many computer systems are not running at peak load continuously. Such systems (including mobile or data analytics) become much more efficient if power can be turned off rapidly while maintaining persistent stored data, since power usage can then become proportional to the instantaneous computational load. This provides additional incentive for the non-volatile storage aspect of SCM.

Some applications such as data analytics and ASIC systems can benefit from having associative memories or content addressability, while other applications might gain little. Mobile systems can become even more compact if many different memory tiers can be combined on the same chip or package, including non-volatile M-class or even S-class Storage Class Memory.

Total cost of ownership is influenced by cost-to-purchase, cost-to-maintain, and system lifetime. Current cost-to-purchase trends are that Hard Disk Drives (HDD) cost roughly an order of magnitude less per bit than Flash memory, which in turn costs almost an order of magnitude less per bit than DRAM. However, cost-to-purchase is not the only consideration. It is anticipated that S-class SCM will consume considerably less power than HDD (both directly and in terms of required cooling), and will take up considerably less floorspace. One early projection what that by 2020, if the main storage system of a data center is still built solely from HDD, the target performance of 8.4 G-SIO/s could consume as much as 93 MW and require 98,568 square feet of floor space.[232] In contrast, the improved performance of emerging memories could supply this performance with only 4 kW and 12 square feet. Given the cost of energy, this differential can easily shift the total cost advantage to emerging memory, away from HDD, even if a cost per bit differential still exists.

These requirements have led to considerable early investigation into new memory architectures, exploiting emerging memory devices, often in conjunction with DRAM and HDDs in novel architectures. These new Storage Class Memories (SCM) are differentiated as whether they are intended to be close to the CPU (M-class) or to largely supplement the hard-drives and SSDs (S-class).

The emergence of SCM leads to the need to resolve issues beyond the device level, including software organization, wear leveling management, and error management. Because of the inherent speed in SCMs, software can easily limit the system performance. Some of these changes have already been initiated by the advent of SSDs. All types of IO software – from the filesystem, through the operating system and up to applications – had to be redesigned in order to best leverage both SSDs and then SCMs. The number of software interactions was reduced, and disk-centric features were removed. Inefficiencies buried deep within conventional software were accounting for anywhere from 70% to 94% of the total IO latency.[233] It is likely to be valuable to give application software direct access to the SCM interface, although this can then require additional considerations to protect the SCM device from malicious software. This direct access has not occurred for SSDs, with manufacturers applying numerous undocumented operations between the input data and the raw storage. (One example is the inversion of entire data blocks, based on the number of 0s or 1s to be stored.) However, this approach is not typically used in current operating systems that use some form of File Address Table as an intermediate index mechanism.

Access patterns in data-intensive computing can vary substantially. While some companies continue to use relational databases, others have switched to flat databases that must be separately indexed to create connections amongst entries. In general, database accesses tend to be fairly atomic (as small as a few bytes) and can be widely distributed across the entire database. This is true for both reads and writes, and since the relative ratio of reads and writes varies widely by application, the optimality of any one design can depend strongly on the particular workload.

A specific issue that arises with SCMs is wear leveling. While DRAMs and HDDs can support a large number of writes to the same location without failure, most of the emerging non-volatile memory device technologies cannot. Thus, there is a need for low-overhead mechanisms to "spread" the writes around uniformly, generally referred to as "wear leveling". An important issue in any file system is that certain data (such as metadata) is written to quite frequently. It is important to make sure that such storage locations are not subject to fast wearout, e.g. by using a more robust technology for such portions of the file system.

Error management is a broader problem than just wear leveling. While DRAM has traditionally benefited from simple methods such as ECC and EDCs, Flash with its large page sizes and slow accesses can afford more sophisticated algorithms such as LDPC. Unfortunately, SCM will need more error correction than DRAM but will need faster error correction than Flash, especially for M-class SCMs. This is an open area for research. Some possible options include exploring codes that exploit specifics of error patterns, such as Tensor codes, and the use of in-situ scrub,[234] where accumulated errors are periodically eliminated so that one or two-bit error correction can remain sufficient.

2.4.2.3. Emerging Memory Architectures for M-Class SCM

Storage Class Memory architectures that are intended to replace, merge with, or support DRAM, and be close to the CPU, are referred to as M-type or Memory-type SCM (M-SCM). The required properties of this memory have many similarities to DRAM, including its interfaces, architecture, endurance, and read and write speed. Since write endurance of an emerging research memory is likely to be inferior to DRAM, considerable scope exists for architectural innovation. It will be necessary to choose how to integrate multiple memory technologies to optimize performance and power while maximizing lifetime. In addition, advanced load leveling that preserves the word level interface and suitable error correction will be needed.

The interface is likely to be a word-addressable bus, treating the entire memory system as one flat address space. Since the cost of adapting to new memory interfaces is sizeable, an interface standard that could support multiple generations of M-SCM devices would be highly preferred. Many systems (such as in automobiles) might be deployed for a long time, so any new standard should be backward-compatible. Such a standard should compatible to DRAM interfaces (though with simpler control commands) and should reuse existing controllers and PHY (physical layers), as well as power supplies, as much as possible. It should be power efficient, e.g. supporting small page sizes, and should support future directions, such as 3D Master/slave configurations. The M-

SCM device should indicate when writes have been completed successfully. Finally, an M-SCM standard should support multiple data rates, such as a DDR-like speed for the DRAM and a slower rate for the NVRAM.[235]

While wear-leveling in a block-based architecture requires significant overhead to track the number of writes to each block, simple techniques such as "Start-Gap" Wear-Leveling are available for direct-byte-access memories such as PCM (Phase Change Memory).[236] In this technique, a pair of registers are used to identify the location of the start point and an empty gap within a region of memory. After some threshold number of write accesses, the gap register is moved through the region, with the start register incrementing each time the gap register passes through the entire region. Additional considerations can be added to defend against detrimental attacks intended to intentionally wear out the memory.

With such techniques, even an M-class SCM that is markedly slower than DRAM can offer improved performance by increasing available capacity and by reducing the occurrence of costly cache misses.[237] With proper caching, a carefully-designed M-SCM system could potentially even match DRAM performance despite its lower device latency.[238] The presence of a small DRAM cache helps keep the slower speed of the M-class SCM from affecting overall system performance in many common workloads. Even with an endurance of 1e7 cycles, the system lifetime has been shown to be on the order of 3 years. Techniques for reducing the write traffic back to the SCM device can help improve this by as much as a factor of 3 under realistic workloads.

Direct replacement of DRAM with a slightly slower M-class SCM has also been considered, for the particular example of STT-MRAM.[239] Since individual byte-level writes to STT-MRAM consume more power than in DRAM, a direct replacement is not competitive in terms of energy or performance. However, by re-architecting the interaction between the output buffer and the STT-MRAM, unnecessary writes back to the NVM can be eliminated, producing a sizeable energy improvement at almost no loss in performance. However, the use of write buffers means that the device must be able to complete all writes back to non-volatile memory in the event of power loss. Integrating PCM into the mobile environment, together with a redesigned memory management controller, is predicted to deliver a six times improvement in speed and also extends the memory lifetime six times.[240]

Caches are intended to ensure that frequently needed data is located near to the processor, in nearby, low-latency memory. In storage architectures, "hot" or frequently accessed data is identified and then moved to faster tiers of storage. However, as the number of tiers or caches increases, a significant amount of time and energy is being spent moving data. An alternative approach is to completely rethink the hardware/software interface. By organizing the computational system around the data, data is not brought to the processor but instead processing is performed in proximity to the stored data. One such emerging data-centric chip architectures termed "Nanostores"[241] was predicted to offer 10–60x improvements in energy efficiency.[242]

Given the slower than expected deployment of scaled emerging devices for M-class SCM, several projects have employed large amounts of DRAM as a surrogate for and M-class SCM. Though Bresniker et.al. describe a computer enabled by a large non-volatile "Universal memory,"[243] press reports indicate that early commercial machines will be built with large amounts of DRAM. They point out that an NVM version allows "occasionally-on computing" but could have the downside that new types of bugs might appear as OSs and programs effectively run indefinitely and cannot "re-create their memory state representations each time they start". New applications and algorithms might emerge as a result of keeping data in perpetuity, or at least for long periods of time.

2.4.2.4. Emerging Memory Architectures for S-Class SCM

S (Storage) type SCMs are intended to replace or supplement hard-disk drives as main storage, much like current Flash-based SSDs, but with even more IOPs (I/O operations per second). Their main advantage will be speed, avoiding the seek time penalty of main drives. However, to succeed, their total cost of ownership needs to approach that of HDDs. Research issues include whether the SCM serves as a disk cache or is directly managed, how load leveling is implemented while retaining a sufficiently fast and flexible interface, how error correction is implemented, and identifying the optimal mix of fast-yet-expensive and slow-yet-inexpensive storage technologies. The effective performance of Flash SSD, itself slower than S-SCM, has been strongly affected by interface performance. For instance, the SATA (Serial Advanced Technology Attachment) interface was originally designed for HDD, and was commonly used for early SSD devices despite not being optimized for Flash SSD.[244]

One possible introduction of these new memory devices to the market would be as *hybrid* solid-state discs, where the new memory technology complements the traditional Flash memory to boost the SSD performance. Experimental implementations of FeRAM/Flash[245]and PCRAM/Flash[246] have been explored. It was shown that the PCRAM/Flash hybrid improves SSD operations by decreasing the energy consumption and increasing the lifetime of Flash memory.

Additional open questions for S-SCM include storage management, interface, and architectural integration, such as whether such a system should be treated like a fast disk drive or as a managed extension of main memory. To date, disk-like systems built using non-volatile memories have had disk-like interfaces, with fixed-sized blocks and a translation layer used to obtain block addresses.

However, since the file system also performs a table lookup, some portion of SCM performance is sacrificed. In addition, non-NAND-Flash SCMs have randomly accessible bits and do not need to be organized as fixed-size blocks.[247]

While preserving this two-table structure means that no changes to the operating system are required to use or to switch between new S-SCM technologies, the full advantages of such fast storage devices cannot be realized. There are two alternative approaches to eliminate one of these lookup tables. In the Direct Access mode, the translation table is removed, so that the operating system must then understand how to address the SCM devices. However, any change in how table entries are calculated (such as improvements in garbage collection or wearleveling) would then require changes in the operating system.

In contrast, in an Object-Based access model, the file system is organized as a series of (key, value) objects. While this requires a one-time change to operating systems, all specific details of the SCM could be implemented at a low level. This model leads to greater efficiency in terms of both speed and effective "file" density and also offers the potential for enhanced reliability.

Another issue for SCM-based systems will be addressing the asymmetry between read and write in devices such as PCM or other emerging non-volatile memories.[248] Such asymmetry can affect the ordering and atomicity of writes and needs to considered in system or algorithm design.[249] Atomicity is critical for operations such as database transactions properties, so that either all of a series of related database operations occur, or none of them occur.

Longer write latencies, in technologies such as PCM, can be compensated by techniques such as data comparison writes,[250] partial writes,[251] or specialized algorithms/structures that trade writes for reads.[252,253] (These last set of techniques can also help reduce endurance problems.) Write ordering and atomicity problems can be finessed by hardware primitives. These can either be existing hardware primitives – cache modes (e.g., write-back, write-combining), memory barriers, cache line flush[254,255,256] – or newly-proposed hardware primitives, such as atomic 8-byte writes and epoch barriers.[257,258]

Even first-generation PCM chips, although implemented without a DRAM cache, compare favorably with state-of-the-art SSDs implemented with NAND Flash, particularly for small (<2 KB) writes and for reads of all sizes.[259] The CPU overhead per input-output operation is also greatly reduced. Another observation for even first-generation PCM chips is that while the average read latency is similar to NAND Flash, the worst-case latency outliers for NAND Flash can be many orders of magnitude slower than the worst-case PCM access. This is particularly important considering that such S-class SCM systems will typically be used to increase system performance by improving the delivery of urgently needed "hot" data.

Another new software consideration for both S- and M-class SCM is the increased importance of avoiding memory corruption, either through memory leaks, pointer errors, or other issues related to memory allocation and deallocation.[260] Since part of the memory system is now non-volatile, such issues are now pervasive and may be difficult to detect and remove without affecting stored user data.

General libraries and programming interfaces – such as NV-heaps,[261] Mnemosyne,[262] NVMalloc,[263] and recovery and durable structures[264] – have been proposed to expose SCM as a persistent heap and thus ease its adoption. Schemes for filesystem support have been developed to transparently utilize as byte-addressable persistent storage, including Intel's PMFS,[265] BPFS, FRASH,[266] ConquestFS,[267] and SCMFS.[268]

Table BC2.8. *Likely Desirable Properties of M (Memory) Type and S (Storage) Type Storage Class Memories*

3. EMERGING LOGIC AND ALTERNATIVE INFORMATION PROCESSING DEVICES

3.1. TAXONOMY

One of the central objectives of this chapter is to review recent research in devices beyond silicon transistors and to forecast the development of novel logic switches that might replace the silicon transistor as the device driving technological development within the semiconductor industry. Such a replacement is thought to be potentially viable if one or more of the following capabilities is afforded by a novel device: 1) an increase in device density (and corresponding decrease in cost) beyond that achievable by ultimately scaled CMOS; 2) an increase beyond CMOS in switching speed, *e.g.*, through improvements in the normalized drive current or reduction in switched capacitance; 3) a reduction beyond CMOS in switching energy, associated with a reduction in overall circuit energy consumption; or 4) the enabling of novel information processing functions that cannot be performed as efficiently using conventional CMOS.

With the evolution of the Roadmap from the ITRS to the IRDS, an expanded focus on systems requires also that one evaluate the applicability of these novel devices for new types of systems and architectures, above and beyond the hierarchical influence of devices via small core circuits or functional building blocks. Historically, this evaluation has been conducted in this section in terms of possible applications for alternative information processing devices (i.e., those very unlike CMOS transistors). Beginning with the 2017 Edition of the IRDS, the focus pivots further toward architectures and applications that depart significantly from the device→circuit→logic gate→functional block→system paradigm. Primarily, in this Edition, the focus for these departures is detailed in Sections 4 and 5 of this chapter. As future editions emerge, more details will be codified in this section to provide stronger linkages to alternative systems and architectures.

For 2017, the organization of this section is intended to reflect a progression of options that might enable an orderly transition from CMOS to devices that depart increasingly from CMOS in terms of structure, materials, or operation. That organization is depicted in Figure BC3.1.

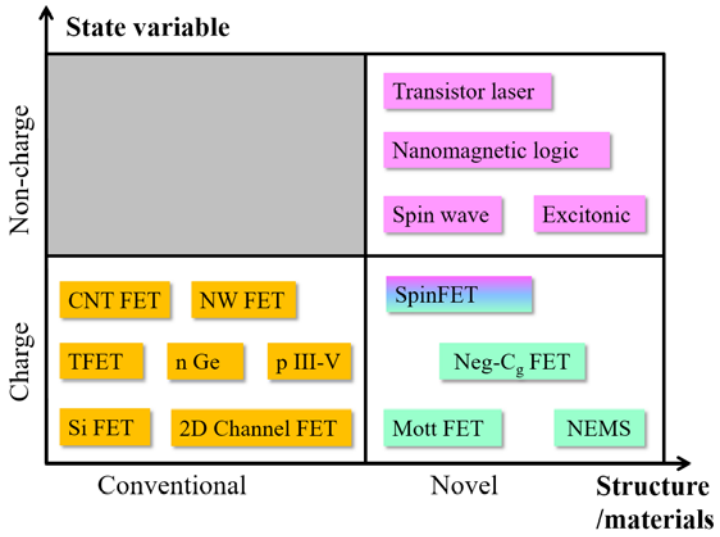


Figure BC3.1 Taxonomy of Options for Emerging Logic Devices

Note: The devices examined in this chapter are differentiated according to 1) whether the structure and/or materials are conventional or novel, and 2) whether the information carrier is electron charge or some non-charge entity. Since a conventional FET structure and material imply a charge-based device, this classification results in a three-part taxonomy.

The resulting taxonomy of emerging logic devices is conveyed in the three tables associated with this section. Table BC3.1a is titled "Extending MOSFETs to the End of the Roadmap" and tabulates characteristics of the devices in the lower-left quadrant of Figure BC3.1. These devices are reviewed below in Section 3.2. Table BC3.1b, titled "Charge-Based Beyond CMOS: Non-Conventional FETs and Other Charge-Based Information Carrier Devices," consists of those devices in the lower-right quadrant of the figure, which are reviewed in Section 3.3. Finally, Table BC3.1c details the "Alternative Information Processing Devices" that are listed in the upper-right quadrant of Figure BC3.1 and that are reviewed below in Section 3.4.

Table BC3.1a MOSFETS: Extending MOSFETs to End of Roadmap

Table BC3.1b Charge-based Beyond CMOS: Non-conventional FETs and Other Charge-based Information Carrier Devices

Table BC3.1c: Alternative Information Processing Devices

3.2. Devices for CMOS Extension

3.2.1. Carbon Nanotube FETs

For many researchers, the search for an ideal semiconductor to be used in FETs succeeded when single-walled carbon nanotubes (CNTs) were first shown to yield promising devices more than 15 years ago. Owing to their naturally ultrathin body (~1 nm diameter cylinders of hexagonally bonded carbon atoms), superb electron and hole transport properties, and reasonable energy

gap of ~0.6 – 0.8 eV, CNTs offer solutions in most of the areas that other semiconductors fundamentally fail when scaled to the sub-10 nm dimensional scale. CNT FETs operate as Schottky barrier transistors with nearly transparent barriers to carrier injection achieved for both n- and p-type transport. They are intrinsic semiconductors and cannot be doped in the traditional sense; hence, no inversion layers of charge form to allow current flow. Rather, the gate field lowers the energy barrier in the CNT channel to allow for carriers to be injected from the metal contacts. The most prominent advantages of CNT FETs over other options for aggressively scaled devices are the room temperature ballistic transport of charge carriers, the reasonable energy gap, the demonstrated potential to yield high performance at low operating voltage, and scalability to sub-10 nm dimensions with minimal short channel effects.

In the past several years, significant advances have been made in understanding and enhancing device performance in CNT FETs. These include 1) realizing end-bonded contacts having an effective contact length of 0 nm with reasonable performance,[269] 2) detailing the impact of contact scalability in CNT FETs,[270] 3) maintaining performance as the channel length is scaled down to 9 nm without observing short channel effects,[271] 4) fabricating complementary gate-all-around FETs,[272] 5) fabricating an FET with an intrinsic fT of 153 GHz,[273] 6) fabricating CMOS inverters and pass-transistor logic operating at 0.4 V with a non-doped CNT,[274] 7) fabricating a carbon nanotube computer composed of 178 FETs,[275] 8) progress towards reducing the variability in CNT FETs,[276] 9) understanding origins of hysteresis,[277] and (10) fabricating CNT FETs with ON-current of 0.5 mA/μm.[278]

In addition to improvements at the device level, continuous progress has been achieved toward overcoming the dominant material challenges,[279] including the need to achieve purified and sorted semiconducting CNTs with a relatively uniform diameter distribution and then position the CNTs into aligned, closely packed arrays with consistent pitch. With a target purity of 99.9999% semiconducting CNTs and placement density of >125 CNTs/μm (<8 nm pitch), much work still remains. However, it is important to note that progress continues to be steady and without fundamental obstacles barring these goals from being realized. There remains a need for further research toward improving other device-level aspects, including further reduction of contact effects at small contact lengths, demonstrated reduction in variability, improved control of gate dielectric interfaces and properties, and the experimental study of devices and circuits fabricated using the most scaled and relevant device structures and materials. In short, much work remains for CNT FETs, but they have some of the most substantial (and already demonstrated) potential in high-performance, low-voltage, sub-10 nm scaled transistor applications.

3.2.2. NANOWIRE FETS

Nanowire field-effect transistors are structures in which the conventional planar MOSFET channel is replaced with a semiconducting nanowire. Such nanowires have been demonstrated with diameters as small as 0.5 nm.[280] They may be composed of a wide variety of materials, including silicon, germanium, various III-V compound semiconductors (III-As, III-Sb, III-P, and III-N), II-VI materials (CdSe, ZnSe, CdS, ZnS), as well as semiconducting oxides (In_2O_3, ZnO, TiO_2), etc.[281] Importantly, at low diameters, these nanowires exhibit quantum confinement behavior, i.e., 1-D ballistic conduction,[282] and match well with the gate-all-around structure that permits the reduction of short channel effects and other limitations to the scaling of planar MOSFETs.

Important progress has been made in the fabrication of semiconducting nanowires for use as FET channels, for which there are two principal formation methods. The first method is top-down, by which semiconducting channels are formed through lithography and etch. The second approach is bottom-up growth including catalyzed vapor-liquid-solid (VLS) growth[283,284] and template-assisted selective epitaxy (TASE).[285] These mechanisms have been used to realize a variety of nanowire geometries, including core-shell and core-multishell heterostructures.[286,287] Although the top–down methods have been prevailing in the industrial manufacturing, the bottom–up methods have been developed to a point that is worth serious consideration for practical use, given their advantage in heterogeneous integration and convenience for in situ passivation through heterostructures. Nanowire gate-all-around geometry is of interest primarily due to its superb gate electrostatics that can allow further gate-length scaling.

Vertical nanowire transistors have been fabricated in this manner using Si,[288] InAs,[289,290] and ZnO.[291] Core-shell gate-all-around configurations display excellent gate control and few short channel effects. Circuit and system functionality of nanowire devices has been demonstrated, including vertical InAs MOSFETs with 103 GHz switching speed,[292] a down-conversion mixer based on vertical InAs transistors that showed cut-off frequency of 2 GHz,[293] and extended, programmable arrays (“tiles”) of non-volatile nanowire-based Flash memory that are used to build circuits such as full-adder, full-subtractor, multiplexer, demultiplexer, clocked D-latch and finite state machines.[294,295] Planar VLS III-As nanowires perfectly aligned in-plane have enabled multigate HEMTs to enhance high frequency performance.[296,297]

Despite the promising results that are mostly obtained from academic research labs, bottom-up nanowire transistors still face significant challenges before commercialization is feasible. There is an urgent need to improve device yield and uniformity, as well as position registry if the nanowires are to be transferred to a different substrate.[298] On the other hand, the potential of bottom-up nanowire transistors for monolithic core-shell architecture and the inherently relaxed lattice matching requirements for heterogeneous integration have not been completely exploited for low power and high speed applications.[299] For top-down

fabricated nanowires, with the demonstration of standalone transistors, CMOS integration, and a functional ring oscillator, vertically stacked planar Si nanowires hold enormous potential to succeed FinFETs in sub-5nm technology nodes.[300,301,302]

3.2.3. 2D MATERIAL CHANNEL FETS

Two-dimensional (2D) materials transition metal dichalcogenides (TMDCs), including graphene, are promising candidates for future channel materials in LSIs. Since they are layered materials, they can be as thin as one-layer (one-atom) thick without degrading their properties, which cannot be realized with conventional three-dimensional materials. Electrostatic control of such a thin channel is easier than a conventional bulk channel, suppressing "short channel effects" often encountered by scaled transistors. Carrier mobility of such 2D materials can be very high. Furthermore, novel devices based on principles different from that of CMOS can be realized using 2D materials, as discussed later.

Graphene has especially attracted attention as a channel material due to its extremely high mobility since the year of 2004, when a report on monolayer graphene prepared by exfoliation of graphite crystals was published.[303] The report showed that the field-effect mobility of graphene on SiO_2 was as high as ~10,000 cm^2/Vs. It was then predicted that the room-temperature mobility of graphene on SiO_2 would be limited to ~40,000 cm^2/Vs due to scattering by surface phonons of the SiO_2 substrate.[304] In fact, much higher field effect mobility was obtained using suspended graphene. Values as high as 120,000 cm^2/Vs and 1,000,000 cm^2/Vs at 240 K and liquid-helium temperature were obtained.[305,306] Recently, hexagonal boron nitride (hBN), an inert and flat material, has been used as a substrate for graphene-channel transistors, and it has been shown that the field effect mobility of such devices can exceed 100,000 cm^2/Vs at room temperature near the charge neutrality point.[307,308] Furthermore, a device in which graphene was sandwiched between two hBN flakes and edge-contacted with two metal electrodes exhibited mobility even higher than the above.[309] The results indicate that hBN can be excellent passivation film for graphene devices.

Graphene can also be obtained on SiC surface by annealing a SiC crystal at high temperatures (often called "epitaxial graphene"). The annealing temperatures range from 1200ºC to 2000ºC depending on the annealing environment.[310,311] Chemical vapor deposition of graphene on metal foil or film has also been demonstrated.[312,313,314,315] Typical growth temperatures are around 1000ºC. It has been shown that monolayer graphene is preferentially formed on Cu foil or film,[312,314,316] while multi-layer graphene can be formed on Ni, Co, and Fe catalyst.[313,315,317] The quality of epitaxial graphene and CVD graphene is now as good as that of exfoliated graphene.[318,319]

Efforts have been made to fabricate graphene-channel transistors, where fabrication processes such as doping and contacting to graphene channels have been developed.[320] However, since graphene does not have a bandgap, such transistors cannot have an ON/OFF ratio high enough for digital applications. Several approaches have been proposed to open a bandgap of graphene. One of the two major approaches is to apply an electric field perpendicular to AB-stacked bilayer graphene.[321,322,323,324,325] Experimentally, a transport gap of 130 meV was obtained at an electrical displacement of 2.2 V/nm, providing an ON/OFF ratio of ~100 at room temperature.[324] This ON/OFF ratio is probably the largest for this approach so far, which is, however, not high enough for logic applications. In fact, it has been pointed out that a small stacking fault of AB-stacked bilayer graphene can increase the off current,[326] which is a serious problem.

A promising approach to form a bandgap in graphene is to make it narrow, that is, to form a graphene nanoribbon (GNR).[327,328,329,330,331,332,333,334] In fact, simulations using a first-principles many-electron Green's function approach within the GW approximation have predicted that bandgaps can be as large as ~5 eV depending on their widths for armchair-edged GNRs (AGNRs).[334,335] Formation of GNRs was first attempted by using electron beam lithography and etching.[328] Carrier transport through such a GNR was also investigated. An energy gap of ~200 meV was obtained for a GNR with a width of 15 nm. Devices with multiple GNRs with a sub-10 nm half-pitch were fabricated using patterning with directed self-assembly of block copolymers.[336] The transport characteristics of such top-down GNRs were poor, however. This is mainly because the edges of such GNRs were not well controlled, probably with a lot of defects.[337,338] Recently, however, attempts to form GNRs with controlled edges have been made using bottom-up approaches. In fact, Cai et al. demonstrated the growth of armchair-edged GNRs (AGNRs) from 10,10'-dibromo-9,9'-bianthryl precursors.[339] In their approach, precursor molecules are deposited onto a clean Au(111) surface by vacuum evaporation in ultra-high vacuum. The substrate is then heated to 200ºC to remove Br from the precursors and to connect them with each other at the Br-removed points, forming polymers. By further heating the substrate to 400ºC, the polymers were cyclodehydrogenated to form AGNRs with a uniform width. The AGNR formed is referred to as 7AGNR, because it has seven dimer lines in the width direction. The band gap of 7AGNR is 3.7-3.8 eV according to the simulations above[334,335] and agrees with an experimentally obtained bandgap (~2.3 eV) considering image-charge corrections by Au substrate.[340] Now several types of GNRs have been formed using similar approaches with different precursors.[341] As for AGNRs with a smaller bandgap, 9AGRs with a theoretical bandgap of about 2.2-2.3 eV have been obtained.[305] Formation of 13AGNRs with a theoretical bandgap of about 2.3-2.5 eV was also demonstrated although the GNRs were rather short, typically less than 10 nm in this case. The successful formation of atomically precise AGNRs paves a way for their application to transistor channels.

Performance of GNR-channel transistors have been predicted by numerical simulations.[342,343,344] It was shown that a transistor with multiple-GNR channels (width: 1.47 nm, pitch: 3.47 nm) with a channel length of 15 nm exhibited an on-current exceeding 1 mA/μm with ON/OFF ratio larger than 10^5 and a subthreshold swing of 64 mV at a drain voltage of 0.1 V[40]. Transistors using 7AGNRs as channels were fabricated and evaluated experimentally. The performance of transistors was, however, poor with a very low on-current and an ON/OFF ratio of 3.6 x 10^3 at a drain voltage of 1V.[345] The small on-current was attributed to large Schottky barriers at the source and drain contacts caused by the large bandgap of 7AGNR. Use of 9AGNRs and 13AGNRs with smaller bandgaps actually improved the transistor performance. In fact, transistors using 9GNRs as channel exhibited ON/OFF ratios as high as 105 and on-current of 1 μA at a drain voltage of 1V, although the number of GNRs in each transistor is unclear.[346] The performance is not yet as good as a counterpart using carbon nanotubes (CNTs)[347] but expected to improve further by, for example, covering GNRs with hBN and realizing better contacts between GNRs and source/drain electrodes.

New principle devices using GNRs have also been proposed. One is a tunneling field-effect transistor (TFET). A higher on-current than that of Si TFET has been predicted.[348] Use of strained graphene as a channel can also realize tunneling-like transport, according to simulations.[349] A Klein-tunneling-based device has also been proposed.[350] Graphene can offer possibilities for employing novel switching mechanism for future electronics.

Transition metal dichalcogenides (TMDCs) are another 2D material attracting attention. TMDCs have the chemical formula of MX_2, where M is a transition metal element and X is a chalcogen. They can be metallic, half-metallic, semiconducting, or superconducting depending on their compositions. Molybdenum disulfide, MoS_2, is probably the most popular semiconducting TMDC, whose single layer was isolated for electrical measurements in 2005.[351] Electrical properties of semiconducting TMDCs depend on the number of layers due to quantum confinement effects and changes in symmetry. For example, single-layer MoS_2 has a direct band gap of 1.9 eV, while bulk MoS_2 has an indirect bandgap of 1.2 eV.[352] The bandgaps also vary with the compositions. For example, first principles calculations performed using the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional show that MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2 have bandgaps of 2.02 eV, 1,72 eV, 1.28 eV, 1.98 eV, 1.63 eV, and 1.03 eV, respectively.[353]

Transistors with TMDCs as a channel material have been demonstrated. Kis et al. fabricated a top-gate MoS_2-channel transistor, demonstrating a large ON/OFF ratio (~10^8) and good subthreshold swing (74 mV/decade).[354] Exfoliated single-layer MoS_2 was used in this experiment. However, field-effect mobility was relatively low (60-70 cm2/Vs).[355] In fact, transport in single-layer TMDC is severely affected by the environment, often degrading its electrical property. It has actually been demonstrated that mobility of TMDC increases with thickness.[356,357] This is because influences of charged impurities in substrate decrease as the thickness increases. As for MoS_2, field-effect mobility as high as 480 cm^2/Vs was obtained for a 47-nm thick MoS_2 channel.[358] MoS_2 is naturally electron-doped, which is possibly caused by chalcogen vacancies, while WSe2 is hole-doped.[359] It is actually still difficult to control doping level of TMDCs, which is a major challenge to realize TMDC-based electronics. Incidentally, Desai et al. fabricated MoS_2-channel transistor with 1-nm gate lengths by using a CNT as a gate.[360] The ON/OFF ratio and subthreshold swing were ~10^6 and 65 mV/decade, respectively. This is probably the shortest channel transistor ever made.

New principles devices using TMDCs have also been demonstrated. Sarkar et al. fabricated a tunneling FET (TFET) by placing n-type MoS_2 on top of p-type Ge, forming a p-n junction.[361] The TFET was gated by using a solid polymer electrolyte. The subthreshold swing had an average of 31.1mV/decade for four decades of drain current at room temperature. Britnell et al. prepared stacking structures consisting of graphene/MoS_2/graphene and graphene/hBN/graphene, demonstrating a switching device based on tunneling phenomena.[362] Although the tunneling devices introduced here can offer steep-slope switching properties, on-current is relatively low, which is an important issue to address for real-world applications. Fabrication of a lateral heterojunction of TMDCs, such as $MoSe_2$ and WSe_2, have also been demonstrated,[363,364] and a p-n junction by such a heterojunction has been made.[365,366] Valleytronics based on carriers located in two inequivalent valleys in the wave number space also attract attention.[367,368] New principles devices using TMDCs, as introduced here, may have good opportunities in future electronics.

Growth technology of TMDCs is also in progress. In fact, the synthesis of TMDC film dates back to the 1980's, when the growth was performed with van der Waals epitaxy.[369] More recently, Lee et al. demonstrated CVD growth of MoS_2 using MoO_3 and S powder as precursors.[370,371] The growth temperature was 650˚C. Single-crystal monolayer MoS_2 flakes were successfully obtained. This method was further improved and single-crystal MoS_2 flakes as large as 120 μm in lateral size were obtained.[372] There have been a quite a few reports using similar methods for synthesizing TMDCs, including $MoSe_2$,[373] WS_2,[370,374] and WSe_2.[375] However, uniform growth over a large area is not easy using this powder-based CVD technique. In 2015, Kang et al. succeeded in large area growth of MoS_2 by MOCVD using $Mo(CO)_6$ and $(C_2H_5)_2S$ as precursors.[376] The electron mobility of MoS_2 in this case was 30 cm^2/Vs at room temperature. In general, mobility of CVD MoS_2 is lower than that of exfoliated MoS_2, which is still an important issue to address. Furthermore, MoS_2 deposition by using sputtering is also in progress.[377,378,379] The sputtering method is scalable, but it is still difficult to obtain film with a quality as high as that by MOCVD.

3.2.4. P-TYPE III-V CHANNEL REPLACEMENT DEVICES

III-V compound semiconductors as replacements for n-type channel materials have attracted considerable attention because of their excellent bulk electron mobilities.[380] To create high performance CMOS circuits, high mobility p-channel materials are also required. Among III-V compound semiconductors, Sb-based semiconductors exhibit high hole mobilities when used as bulk materials; for example, InSb and GaSb show mobilities of 850 $cm^2 \cdot V^{-1} \cdot s^{-1}$ and 800 $cm^2 \cdot V^{-1} \cdot s^{-1}$, respectively,[381] which are significantly greater than the bulk Si hole mobility of 500 $cm^2 \cdot V^{-1} \cdot s^{-1}$. Moreover, hole mobility can be further increased by introducing biaxial compressive strain. This can be accomplished using pseudomorphic growth on a material with a smaller lattice constant.[381, 382, 383, 384] Another method of improving hole mobility is the application of uniaxial strain[383, 385,] which can be used in a similar manner as used in Si MOSFETs. The piezoresistance coefficient of p-InGaSb is 1.5 times greater than that of Si when uniaxial strain is applied.[386] Another advantage of InGaSb is its superior characteristics when used as an n-channel material. InGaSb has an electron mobility of over 4000 $cm^2 \cdot V^{-1} \cdot s^{-1}$. Thus, CMOS circuits can be fabricated using a single channel material system.[387]

The highest hole mobilities when using compressive strain were found in GaSb/AlAsSb heterostructures; they exhibit a mobility[384] of 1500 $cm^2 \cdot V^{-1} \cdot s^{-1}$. In the case of an InGaSb/AlGaSb system, 1500 $cm^2 \cdot V^{-1} \cdot s^{-1}$ was also observed[382]. In InSb, the highest mobility[386] is 1230 $cm^2 \cdot V^{-1} \cdot s^{-1}$. However, the thickness of the quantum well was 5 nm; a wider well might provide higher mobility.[384] Regarding device performance, compressively strained InSb quantum well pFETs with a gate length of 40 nm have an f_T value of 140 GHz, a g_m of 510 mS/mm, and an I_{on} of 150 mA/mm at a power supply voltage of 0.5 V. When the gate length was 125 nm, the sub-threshold slope[386] was 90 mV/dec.

A device on a Si substrate is essential as a replacement material for Si channels in MOSFETs. To realize III-V p-FETs on a Si substrate, some methods are reported. The first method is transfer of nano-ribbon,[388], nanowire[389] or film.[390] In the case of nano-ribbon, the peak effective mobility of the device is 820 $cm^2 \cdot V^{-1} \cdot s^{-1}$, achieved using an InAs cladding layer for surface passivation and hole confinement. By employing high-κ materials (10-nm-thick ZrO_2), the obtained sub-threshold slope was 130 mV/dec, even at an interface surface density of 1.4×10^{13} $cm^{-2} \cdot eV^{-1}$. In combination with an InAs channel using the same technique, CMOS gates were fabricated and the logic operations of NOT and NAND were demonstrated.[391] In the case of this trial, the peak mobility was reduced to 370 $cm^2 \cdot V^{-1} \cdot s^{-1}$ because of the thin well (5 nm) in the enhancement mode. The obtained sub-threshold slope was 156 mV/dec. In the case of nanowire, the inverter circuit was fabricated from single nanowire with InAs and GaSb.[389] As another possibility of InAs/GaSb nanowire, it can apply tunnel FET.[392] Rapid melt growth (RMG) method[393] can also supply GaSb thin film on Si, and transistor operation by RMG was reported.[394]

To realize a p-FET using an Sb-based channel, one notable problem is the interface property of metal-insulator-semiconductor structures. At present, the best I-V properties have been reported for compressively strained InSb quantum well p-FETs (with a gate length of 40 nm) using an HEMT structure to create a Schottky gate;[386] a MOS structure is a preferred choice to achieve low gate leakage currents. In MOSFETs, the lowest sub-threshold slope is limited to 120 mV/dec when the gate length is 5 mm.[395] Although the interface state density is approximately 3×10^{11} $cm^{-2} \cdot eV^{-1}$ in the mid-bandgap region, it rises to 1×10^{13} $cm^{-2} \cdot eV^{-1}$ near the conduction band edge. Moreover, the insulator thickness (Al_2O_3) in this trial was limited to 10 nm. Thus, improvements to the method of obtaining a low interface state density using a thinner insulator are required because the interface state density strongly depends on the fabrication process.[396] The requirement of having a high on-current is another problem. For a 5-mm-long MOSFET, an on-current of 70 mA/mm was reported;[395] however, shorter gate lengths were not studied in this report. The shortest reported gate length in a p-MOSFET is 750 nm,[396] and the channel dependence shows an inverse relationship between the gate length and the on-current. In the 750 nm-long device, the highest on-current was 70 mA/mm. Moreover, the estimated on-current in the InGaSb HEMT structure was 200 mA/mm, even when the gate length was 20 nm.[397] Taking both the significant degradation of mobility at a high sheet carrier concentration and carrier starvation due to insufficient carrier concentration due to poor ion implantation in a III-V semiconductor into account, the feasibility of high on currents must be verified.

3.2.5. N-TYPE GE CHANNEL REPLACEMENT DEVICES

High electron mobility in bulk Ge and (111)-inversion layer, which are 2.4 times and 3.1 times the counterpart of Si, respectively,[398] is considered a fundamental advantage of Ge-nMOSFETs. In addition, comparable or even higher ballistic current has been estimated for (111)Ge-nMOSFETs than for GaAs, InAs and InSb nMOSFETs due to the higher density of states,[399] although the electron mobility of Ge is less than those of III-V compound semiconductors. Not only the performance but also the lower cost of forming CMOS circuits may be another advantage of introducing Ge-nMOSFETs than forming III-V nFET/Ge pFET dual-channel CMOS configuration. Sub-100-nm FinFET-type Ge-nMOSFETs have already been demonstrated[400,401] for the counterpart of Ge-pMOSFETs in 7-nm technology node and beyond. Strain technologies to enhance the electron mobility using source-drain stressors[402] and stress liners[403] have been demonstrated and simulated, respectively. Simple CMOS circuits such as invertors and ring oscillators composed of Ge p- and nMOSFETs have been demonstrated on Ge-on-insulator substrates[404] and on poly-Ge layers on SiO_2/Si substrates.[405] A circuit simulation based on TCAD results for p- and n-channel Ge-on-insulator (GOI) MOSFETs predicted that the GOI-CMOS outperforms Si-on-insulator CMOS in various elemental circuits and

conditions.[406] As a new material, GeSn semiconductor alloy, in which direct band gap was predicted for a Sn fraction over ~10%, has been examined for CMOS and Si-photonics devices.[407,408] Theoretical calculations for a lattice-relaxed GeSn-nMOSFET with a Sn fraction around 10% predicted slightly higher carrier injection velocity than in a Ge-nMOSFET thanks to the higher electron population to the G valley with a lower effective mass.

The best reported inversion electron mobility at a high inversion electron density (N_s) is 429 cm^2/Vs at $N_s=1\times10^{13}$ cm^{-2} or 488 cm^2/Vs at $N_s=8\times10^{12}$ cm^{-2}.[409,410] It is significant that these high-mobility values were observed for devices with low EOT values of around 1 nm in terms of implementing Ge-nMOSFETs in scaled CMOS-LSIs. The scaling of gate-length and EOT has been demonstrated independently down to 35 nm and 0.39 nm, respectively.[400,411] The sub-threshold swing and off-state leakage current of Ge-nMOSFETs are getting better but still slightly worse than those of Ge-pMOSFETs[412], suggesting that the interface state density near the conduction-band edge should be improved. The drain current for scaled Ge nMOSFETs has been greatly improved during the latest 2 years and has reached ~250 mA/mm for $L_g=40$ nm device at $V_{dd}=1V$.[400] However, the current drivability is still lower than that of state-of-the-art Si-nMOSFETs of the similar gate-lengths. Reducing the high parasitic resistance in the Ge-nMOSFETs is a key technology to obtain acceptable current drivability as a counterpart of high-performance Ge-pMOSFETs. One of the major problems is the low activation of n-type dopants in Ge, which is usually saturated up to around 1×10^{19} cm^{-3} by means of conventional ion-implantation and RTA techniques. Theoretical calculation predicted the specific contact resistivity required in the ITRS ($<1\times10^{-8}$ ohm.cm^2) can be satisfied for a metal/0.7-nm-ZnO/n^+-Ge contact if an electron density of over 1×10^{20} cm^{-3} is attained in the Ge layer.[411] Experimentally, enhancing the activation of n-type dopants up to over 2×10^{20} cm^{-3} by Sb ion implantation and laser annealing processes has been achieved.[413] Reduction in a specific contact resistivity (r_c) down to ~3×10^{-8} ohm.cm^2 for NiGe/n^+-Ge contact has been also reported by the 2-step ion-implantation technique.[414]

The most significant issue for the implementation of Ge-nMOSFETs in future high-performance CMOS technologies is the demonstration of competing short-channel performances with conventional Si-nMOSFETs. Integration of the contact formation processes to reduce parasitic resistance with the gate-stack formation processes to realize sub-nm-EOT with low interface state density is a key technology. These should be consistent with integration processes for technology generations of 7-nm node and beyond. A common gate stack process for p- and n-MOSFETs is also desirable in terms of the cost merit against III-V/Ge dual channel devices.

3.2.6. TUNNEL FETS

Tunneling Field Effect Transistors (TFETs) have the potential to achieve a low operating voltage by overcoming the thermally limited subthreshold swing voltage of 60 mV/decade by utilizing tunneling as a switching mechanism.[415,416,417] In its simplest form, a TFET is a gated, reverse-biased p-i-n diode with a gate controlled intrinsic channel. There are two mechanisms that can be used to achieve a low voltage turn on. The gate voltage can be used to modulate the thickness of the tunneling barrier at the source channel junction and thus modulate the tunneling probability.[418,419,420,421] The thickness of the tunneling barrier is controlled by changing the electric field in the tunneling junction. Alternatively, it is possible use energy filtering or density of states switching. If the conduction and valence band do not overlap at the tunneling junction, no current can flow. Once they do overlap, current can flow. Simulations have predicted arbitrarily steep subthreshold swings when relying on density of states switching as the current is abruptly cutoff when the conduction band and valence band no longer overlap.[417] If phonons or short channel lengths are accounted for, simulated subthreshold swings on the order of 20–30 mV/decade are typical.[422] It is possible to use the tunneling switching mechanisms in series with the standard MOSFET thermal switching mechanism to get an overall subthreshold swing that is steeper than 60 mV/decade when no individual mechanism is steeper than 60 mV/decade.[423] The best experimental results to date have relied on a combination of thermal switching and density of states switching.[424].So far, the experimental results are far worse than simulated device characteristics. The review by Lu and Seabaugh[415] shows a comprehensive benchmarking of published experimental results prior to 2014. The benchmarking shows two problems with TFETs as a MOSFET replacement: 1) Devices are unable to achieve SS <60 mV/decade over a large range or at useful current levels and 2) The on-state current is too low for reasonable performance.

The review shows 14 reports of subthreshold swings below 60 mV/decade, and a few additional results have been published since. Most of the results are for group-IV materials such as Si,[420,425,426,427,428,429] strained SiGe,[430] Si/Ge,[431] and strained Ge.[432] Nanowire III-V TFETs have shown even steeper swings. A InP/GaAs heterojunction[433] has shown 30 mV/decade at 1 pA/μm. The steepest result ever reported is in a Si/InAs heterojunction[434] of approximately 20 mV/decade at 0.1 pA/μm. However, there are only a few data points defining this result. Even for low power applications, at least 1–10 μA/μm is needed.[435] Recently, a promising InAs/GaAsSb/GaSb nanowire heterostructure TFET with a 48 mV/decade SS at 67 nA/μm and an I_{60} (current at 60 mV/decade) of 0.31 μA/μm was demonstrated.[424].

Researchers attempting to achieve higher on-current TFETs have traditionally relied on reducing the effective mass by using III-V's and reducing effective bandgap by using a heterostructure. While this has increased the current, the subthreshold swings and off currents have gotten worse. The increase in off-state current and subthreshold swing needs to be decoupled from the increase

in on-state current. Unfortunately, this is a fundamental tradeoff when modulating the thickness of the tunneling barrier:[416] barrier thickness modulation only gives a step subthreshold swing at low current densities.

An ideal density of states (DOS) switch would switch abruptly from zero-conductance to the desired on-conductance thus displaying zero subthreshold swing voltage.[436] Practically, the band edges are not perfectly sharp, so there is a finite density of states extending into the band gap. Optical measurements of intrinsic GaAs imply a band edge steepness of 17 meV/decade.[437] However, the electrically measured joint DOS in diodes has generally indicated a steepness >90 mV/decade.[438] This broadening is likely due to the spatial inhomogeneity and on heavy doping that appears in real devices. Effectively, there are many distinct channel thresholds in a macroscopic device, leading to threshold broadening. Fortunately, it's been experimentally demonstrated that a band edge worse than 60 mV/decade can be combined with a thermal switching mechanism to give an overall subthreshold swing better than 60 mV/decade.[423]

TFETs are reverse-biased diodes hence are subject to generation in the depletion region. These generation events include but are not limited to bulk and interface trap-assisted Shockley-Read-Hall (SRH)[439,440,441] and spontaneous and Auger generation.[442] Calculations based upon these mechanisms show that these significantly degrade the subthreshold swing and increase the leakage currents but do not prevent TFETs from achieving sub-60 mV/decade subthreshold swing. Material defects and gate interface traps make these effects worse and result in worse band edges.

To overcome these challenges, better material perfection than ever before is needed. Every defect or dangling bond can create a trap that ruins the band edge or creates a parallel conduction path. The defects due to doping can be eliminated by electrostatically inducing carriers. Proof of concept devices can be made by making the device a few nanometers large so that there is a low probability of having a trap within the device. 2D transition metal dichalcogenides (TMD) heterostructures potentially have better electrostatic control and lower defects as there are ideally no dangling bonds at the semiconductor oxide interface.

3.3. Beyond-CMOS devices: Charge-Based

3.3.1. Spin FET and Spin MOSFET Transistors

Spin-transistors are classified as "non-conventional charge-based extended CMOS devices,"[443] and can be further divided into two categories: the spin-FETs proposed by Datta and Das[444] and spin-MOSFETs proposed by Sugahara and Tanaka.[445] The structures of both types of spin transistors consist of a ferromagnetic source and a ferromagnetic drain which act as a spin injector and a detector, respectively. Although the devices have similar structures, they have quite different operating principles.[443,446] In spin-MOSFETs, the gate has the same current switching function as in ordinary transistors, whereas in the spin-FETs, the gate acts to control the spin direction by utilizing the Rashba spin-orbit interaction. Both types of devices behave as a transistor and function as a magnetoresistive device. The important features of spin transistors are that they allow a variable current to be controlled by the magnetization configuration of the ferromagnetic electrodes (spin-MOSFETs) or the spin direction of the carriers (spin-FETs), and they offer the capability for non-volatile information storage using the magnetization configurations. These features are very useful for energy-efficient, low-power circuit architectures that cannot be achieved by ordinary CMOS circuits. Non-volatile logic and reconfigurable logic circuits have been proposed using the spin-MOSFET and the pseudo-spin-MOSFETs, which are suitable for power-gating systems with low static energy.[446,447,448,449,450,451,452]

The read and write operations of spin-FETs[446,453] and spin-MOSFETs[446,454,455,456,457] have not yet been fully experimentally verified. However, there has been important progress in the underlying technologies. Spin injection, detection and manipulation are essential [455] for realizing fully functional spin transistors. One key development has been half-metallic ferromagnetic materials for highly efficient spin injector/detectors. Theories[458,459,460,461] also predict that the insertion of a tunnel barrier between the ferromagnet (FM) and semiconductor (SC) in order to optimize the interface resistance between FM, and SC is a promising method for producing highly efficient spin injection and detection.

Spin injection and detection in Si have been achieved using hot-electron transport[462] and spin-polarized tunneling.[463,464,465,466,467,468,469,470] The electrical creation and detection of spin accumulation in n-type and p-type Si were demonstrated using FM/tunnel contacts[463,464,466,467,468,469,470] and FM/Schottky-tunnel-barrier contacts[471] up to room temperature (RT). Electrical spin injection into Si channels has also been demonstrated up to 500 K[465] by using a spin relaxation measurement method (Hanle measurement). A relatively long spin lifetime has been observed in heavily doped Si at RT.[464,467,468,470] Spin transport in SC has been studied intensively in various kinds of tunnel and channel materials. Observation of spin signals in heavily doped Ge[472,473,474] and GaAs[475,476] at RT has also been reported. Another study[477, 478] reported low-resistance spin injection and detection into Si using graphene and boron nitride tunnel barriers. Local magnetoresistance signal up to RT have been observed in long channel devices with FM/MgO tunnel barrier/Si lateral spin valve structures.[479,480] Recently, a basic read operation was demonstrated[457] at room temperature by using spin-MOSFETs with a back-gated structure and an interface technology in which the local magnetoresistance signal is observed at RT.

Since the reported[457] difference in the drain current between parallel and antiparallel spin configurations is very small at room temperature, the development of half-metallic ferromagnetic materials on SC is expected to be very important. Half-metallic Heusler alloys[481,482,483,484,485] are one of the most promising half-metallic ferromagnetic materials, because they are relatively easy to prepare. In addition to the progresses shown in the 2013 edition of the ITRS and references, spin injection and detection has been reported that is more efficient in SCs using a Heusler compound than in the ordinal CoFe electrodes.[486,487,488,489] More efforts to improve the qualities of Heusler compounds and FM/SC interfaces are required in order to observe large differences in drain current between parallel and antiparallel spin configurations. It should be noted that half-metallic materials are a required technique for spin-FETs because the current switching function needs to be controlled by spin precession induced by the action of the gate voltage. In spin-MOSFETs, however, since the relative magnetization configurations of the source and drain are used to modify the output current, half-metallic materials might not be required for spin-MOSFET.

Alternative approaches for realizing spin-MOSFETs have been proposed.[446,451,454,490, 491] Pseudo-spin-MOSFETs are circuits that reproduce the functions of spin-MOSFETs using an ordinary MOSFET and a magnetic tunnel junction (*MTJ*) which is connected to the MOSFET in a negative feedback configuration. Although pseudo-spin-MOSFETs offer the same functionality as spin transistors, such as the ability to drive variable current, pseudo-spin-MOSFETs have larger resistance than spin-FETs or spin-MOSFETs.

In terms of spin manipulation, spin precession of polarized carriers in the channel controlled by a gate voltage has been observed experimentally at low temperature.[492] This result confirms that spin-orbit interaction can be controlled by gate voltage. Materials with a strong spin-orbit interaction, such as InGaAs, InAs and InSb, are required[446] in order for the channel to sufficiently induce the Rashba spin-orbit interaction. However, materials with a strong spin-orbit interaction decrease spin lifetime. The use of a narrow wire channel structure[493,494,495] and the so-called persistent spin helix condition[496,497,498,499,500] has been proposed in order to increase spin lifetime. The experimental proof of spin injection, detection and manipulation at RT is needed in order to create spin-FETs with a channel material having strong spin-orbit interaction. Spin-MOSFET[454,455] and the pseudo-spin-MOSFET[446,450,491] use spin-transfer-torque (*STT*) switching, similar to the *STT*-MRAM that was commercialized at the end of 2012.[501]

To date, long channel devices with center-to-center spacing between the FM source/drain electrodes larger than $l_{FM\text{-}FM}$ >1000 nm have been investigated. To realize full read and write operations in spin-MOSFETs and spin-FETs at RT, experimental studies using short channel devices with $l_{FM\text{-}FMl}$ <100 nm are needed.

3.3.2. NEGATIVE GATE CAPACITANCE FET

Salahuddin and Datta originally proposed[502] that, based upon the energy landscape of ferroelectric capacitors, it should be possible to implement a step-up voltage transformer that will amplify the gate voltage of a MOSFET. This would be accomplished by replacing the standard insulator in the gate stack with a ferroelectric insulator of appropriate thickness. The resulting device is called a negative gate capacitance FET or NCFET. The gate operation in this device would lead to subthreshold swing (STS) lower than 60 mV/decade and might enable low voltage/low power operation. The main advantage of such a device[503] is that it is a relatively straightforward replacement of conventional FETs. Thus, high I_{on} levels similar to advanced CMOS would be achievable with lower voltages. An early experimental attempt to demonstrate a low-STS NCFET, based on a P(VDF-TrFE)/SiO_2 organic ferroelectric gate stack, was reported[504] in 2008, and subsequently[505] in a more controlled structure in 2010. These experiments established the proof of concept of sub-60 mV/decade operation using the principle of negative capacitance.

In addition to these experiments using polymer-based ferroelectrics, negative differential capacitance was demonstrated in a crystalline capacitor stack.[506] Essentially, it was demonstrated that in a bi-layer of dielectric Strontium Titanate ($SrTiO_3$: STO) and Lead Zirconate Titanate ($Pb_xZr_{1-x}TiO_3$: PZT), the total capacitance is larger than what it would be for just the STO of the same thickness as used in the bi-layer. This necessarily demonstrates the stabilization of PZT at a state of negative differential capacitance. More recently, in a single PZT capacitor, a direct measurement of negative capacitance was demonstrated.[507] That work determined that when a ferroelectric capacitor is pulsed with an input voltage, it shows an 'inductance-like' discharging in addition to a capacitive charging.

As a recent significant result, it is now possible to grow ferroelectric materials using the atomic layer deposition process (ALD) by doping the frequently used gate dielectric HfO_2 by constituents such as Zr, Al or Si.[508] Using this doped Hf based ALD ferroelectric, a number of experiments have demonstrated the negative capacitance effect.[509,510,511] For example, by using $HfZrO_2$ as a gate dielectric, sub-60 mV/decade STS was demonstrated[511] in finFETs with L_g=30 nm for both nFET and pFET structures. In the last two years, multiple papers have reported this effect for various material systems and channel lengths. Significant among them is the demonstration by GlobalFoundries of NCFETs in their 14 nm finFET technology, with improved subthreshold swing, lower OFF current and lower active power and ring oscillators running at GHz speed.[512]

3.3.3. NEMS SWITCH

Nano-Electro-Mechanical (NEM) switches are devices that employ electrostatic force to actuate a movable structure to make or break physical contact between two conductive electrodes. Mechanical switches feature two fundamental properties which are unattainable in MOSFETs: *zero off-state leakage* and *zero subthreshold swing*.[513] The first property provides for zero standby power dissipation, while the second property suggests the potential to operate at very low voltages for low dynamic power dissipation as well. Another advantage is that a mechanical switch can be operated with either positive or negative voltage polarity due to the ambipolar nature of the electrostatic force, so that an electrostatically actuated relay can be configured as a pull-down or pull-up device, respectively.[514] Additional advantages of mechanical devices include robust operation across a wide temperature range,[515] immunity to ionizing radiation, and compatibility with inexpensive substrates such as glass or even plastic. Since NEM switches comprise only conductive electrodes and air gaps, they can be fabricated at relatively low process temperatures in a modular fashion over CMOS circuitry. Indeed, since the most advanced CMOS technology in production today incorporates air-gapped interconnects,[516] it is conceivable for NEM switches to be implemented using multiple interconnect layers in an advanced back-end-of-line (BEOL) process.[517] Potential applications for mechanical switches in a hybrid NEMS-CMOS technology include CMOS power gating,[518] configuration of CMOS FPGAs,[519] non-volatile storage of information in SRAM and CAM cells,517 and energy-efficient, fast and reconfigurable look-up tables.[520]

NEM switches can be fabricated using conventional planar processing techniques (thin-film deposition, lithography and etch steps), with a final release step in which a sacrificial material such as silicon dioxide, photoresist, polyimide or silicon is selectively removed to form the actuation and contact air-gaps. (The gap in the contact region(s) can be made smaller than the actuation gap to reduce the switching delay and energy as well as the contact velocity for reduced wear and contact bounce.) The smallest actuation and contact gap demonstrated to date for a functional NEM structure fabricated using a top-down approach is 4 nm.[521] This device is a vertically actuated switch featuring a 30 nm-thick, 300 nm-wide, 1.4 μm-long doubly clamped TiW beam with a pull-in voltage of approximately 0.4 V. As expected, the off-state current is immeasurably low and the sub-threshold swing is practically zero. However, it is a 2-terminal device, not suitable for logic switch application.

In a digital logic circuit, it is necessary to connect multiple switches in series to implement various logic functions. To avoid having the state of a switch depend undesirably on the state of other switches in the series stack, a reference electrode is needed, such that the voltage applied between the control ("gate") electrode and the reference ("body") electrode determines the state of the switch, *i.e.* whether current can flow between the output ("source" and "drain") electrodes.[522] By biasing the reference electrode, the operating voltage of a relay can be minimized. Similarly, as for MOSFETs, a constant-field scaling methodology can be applied to scale NEM switches for improved device density, switching delay and switching energy.[523] The ultimate device density achievable may be comparable to that for MOSFETs, in principle. However, the switching delay of a mechanical logic switch (relay) is much longer than that of a MOSFET because it is dominated by the mechanical (motional) delay, ~1 ns,[513] rather than the electrical (charging/discharging) delay. Because of the large ratio between the mechanical and electrical delays of a relay, an optimized IC design should arrange for all mechanical movement to happen simultaneously, *i.e.* relay-based digital logic circuits should be comprised of single-stage complex gates so that the delay per operation is essentially one mechanical delay.[514] This generally results in significantly lower device counts as compared to the optimal CMOS implementations, especially since a relay can pass both low and high logic levels, and the body electrode can also be connected to a logic signal. By incorporating multiple pairs of source and drain electrodes into each gated structure[524] and/or by partitioning the gate electrode into multiple input electrodes,[525] the device count and hence the area required can be further reduced. A variety of relay-based computational and memory building blocks have been experimentally demonstrated to date.[526,527]

Since an optimized relay-based logic circuit has a topology that is very different from that of an optimized CMOS logic circuit, an assessment of the prospective benefits of NEM switch technology must be made at the level of complete circuit blocks. Such circuit-level assessments (*e.g.* of full adder and multiplier circuits) indicate that relays can provide for more than 10x reduction in energy per operation as compared with MOSFETs, and can reach clock speeds in the GHz regime.[528] Thus, a major potential advantage of NEM switch technology is *improved energy efficiency*. Moreover, by engineering the contact adhesive force and structural stiffness, bi-stable operation can be achieved, making scaled mechanical switches attractive for embedded non-volatile memory applications.[529]

Reliable operation is necessary for practical application of M/NEM switches in electronic circuits. Due to their extremely small mass (less than 1 ng), mechanical vibration/shock is not an issue. Structural fatigue is easily avoided by designing the movable electrode such that the maximum induced strain is well below the yield strength. Mechanical wear and Joule heating at the contacting points lead to increased real contact area and eventually stiction-induced device failure. This issue can be mitigated by using a refractory material to minimize wear and material transfer and by reducing the device operating voltage. (It should be noted that the higher associated contact resistance would not significantly compromise performance since the speed of an optimally designed relay-based logic circuit is limited by the mechanical delay rather than the electrical delay. The on-state resistance of a logic relay can be as high as ~10 kΩ.[514]) NEM switches with tungsten contacts have been demonstrated to have endurance up to 1 billion on/off cycles at 2.5 Volts for a relatively large load capacitance of 300 pF (*i.e.* exaggerated electrical

delay).[530] Endurance exceeding 10^{16} on/off cycles is projected for an operating voltage below 1 Volt and load capacitance below 1 pF. A gradual increase in contact resistance caused by surface oxidation or the formation of friction polymers during the course of device operation can lead to circuit failure and is the primary reliability challenge for M/NEM logic switches today. Stable conductive oxide contact materials and/or hermetic packaging are potential solutions to this issue.

Contact adhesive force sets the minimum spring restoring force of the movable electrode (to ensure that the switch turns off), which in turn sets a lower limit on the electrostatic actuation force (to ensure that the switch turns on) and hence the switching energy. For ultimately scaled contacts, the surface adhesion energy will be set either by metallic bonding or by van der Waals force (for oxidized contacting surfaces). The minimum switching energy for a nanoscale relay is anticipated to be on the order of 10 aJ, which compares well against the switching energy for an ultimately scaled MOSFET.[531,532]

In conclusion, NEM switches are intriguing candidates for ultra-low-power applications because they have negligible off-state leakage currents and abrupt switching behavior which, in principle, enables a very low operating voltage. Practical challenges remain to be solved, to achieve stable on-state resistance ((R_{ON}) and to minimize contact adhesive force within R_{ON} limits. A circuit-level assessment of energy *vs.* delay performance indicates that nanoscale relays should provide for more than 10× improvement in energy efficiency as compared with CMOS transistors, for applications requiring clock speeds below 100 MHz. Thus, they are poised to lead a resurgence in mechanical computing for the Internet of Things.

3.3.4. MOTT FET

Mott field-effect transistor (Mott FET) utilizes a phase change in a correlated electron system induced by a gate as the fundamental switching paradigm.[533,534] Mott FETs could have a similar structure as conventional semiconductor FET, with the semiconductor channel materials being replaced by correlated electron materials. Correlated electron materials can undergo Mott insulator-to-metal phase transitions under an applied electric field due to electrostatically doped carriers.[535,536] Besides electric field excitation, the Mott phase transition can also be triggered by photo- and thermal-excitations for optical and thermal switches. Defects created by environmental exposure to chemicals or electrochemical reactions can also induce Mott transition via carrier doping. The critical threshold for inducing phase change can be tuned via stress.

Mott FET structure has been explored with different oxide channel materials.[534] Among several correlated materials that could be explored as channel materials for Mott FET, vanadium dioxide (VO_2) has attracted much attention due the sharp metal-insulator transition near room temperature (nearly five orders in single crystals).[537] The phase transition time constant in VO_2 materials is in sub-picosecond range determined by optical pump-probe methods.[538] Device modeling indicates that the VO_2-channel-based Mott FET lower bound switching time is of the order of 0.5 ps at a power dissipation of 0.1 μW.[539] VO_2 Mott channels have been experimentally studied with thin film devices and the field effect has been demonstrated in preliminary device structures.[540, 541, 542] Recently, prototypes of VO_2 transistors using ionic liquid gating have shown larger ON/OFF ratio at room temperature than with solid gate dielectrics like hafnia.[543, 544] The conductance modulation happens at a slow speed, however, due to the large charging time constants.[545] The possibility of electrochemical reactions must also be carefully examined in these proof-of-principle devices due to the instability of the liquid-oxide interfaces and the ease of cations in such complex oxides to change valence state.[545, 546, 547] On the other hand, unlike traditional CMOS that is volatile and digital, electrochemically gated transistors exhibit non-volatile and analog behaviors, which can be utilized to demonstrate synaptic transistors[548] and circuits[549] that mimic neural activities in the animal brains. Voltage induced phase transitions in two-terminal Mott switches have also been implemented to realize neuron-like devices[550] and steep-slope transistors.[551] As a Mott device with purely electrostatic modulation in the solid-state base, the VO_2-FET with a high-k oxide/organic hybrid dielectric gate has been proposed.[552,553] Their reversible as well as quick resistance switching upon an application of gate bias and the maximum resistance modulation at the Metal-Insulator transition temperature indicate the possibility of purely electrostatic field-induced metal-insulator (Mott) transition.

Experimental challenges with correlated electron oxide Mott FETs include fundamental understanding of gate oxide-functional oxide interfaces and local band structure changes in the presence of electric fields. Methods to extract quantitative properties (such as defect density) of the interfaces are an important topic that have not been explored much to date. The relatively large intrinsic carrier density in many of the Mott insulators requires the growth of ultra-thin channel materials and smooth gate oxide-functional oxide interfaces for optimized device performance. It is also important to understand the origin of low room-temperature carrier mobility in these materials.[535] Theoretical studies on the channel/dielectric interfacial electronic band structure are needed for the modeling of subthreshold behaviors of Mott FETs. Understanding the electronic transition mechanisms while de-coupling from structural Peierls (lattice) distortions is also of interest and important in the context of energy dissipation for switching.

While the electric field-induced transitions are typically explored with Mott FET, nanoscale thermal switches with Mott materials could also be of substantial interest. Recent simulation studies of "ON and "OFF" times for nanoscale two-terminal VO_2 switches indicate the possibility of sub-ns switching speeds in ultra-thin device elements in the vicinity of room temperature.[554,555] Such devices could also be of interest to Mott memory.[556] One can, in a broader sense, visualize such correlated electron systems as

'threshold materials' wherein the conducting state can be rapidly switched by a slight external perturbation, and hence lead to applications in electron devices. Electronically driven transitions in perovksite-structured oxides such as rare-earth nickelates[557,558] or cobaltates[559] with minimal lattice distortions would also be relevant in this regard. Three-terminal devices are being investigated using these materials and will likely be an area of growth.[560,561,562, 563,564] $SmNiO_3$, with its metal-insulator transition temperature near 130ºC and nearly hysteresis-free transition, is particular interesting due to the possibility of direct integration onto CMOS platforms. Floating gate transistors have recently been demonstrated on silicon.[565] It has been found that non-thermal electron doping in $SmNiO_3$ can lead to a colossal increase in its resistivity, which has been utilized to demonstrate a solid-state proton-gated transistor with large ON/OFF ratio.[566] Clearly, these preliminary results suggest the promise of correlated oxide semiconductors for logic devices, while the doping process indicates slower dynamics than possible with purely electrostatic carrier density modulation. The non-volatile nature of the Mott transition in 3-terminal devices suggest combining memory operations into a single device and could be explored further. Architectural innovations that can create new computing modalities with slower switches but at lower power consumption can benefit in the near term with results to date while in the longer term transistor gate stacks need to be studied further for these classes of emerging semiconductors.

3.4. Beyond-CMOS devices: Alternative Information Processing

3.4.1. Spin Wave Device

A Spin Wave Device (SWD) is a type of magnetic logic device exploiting collective spin oscillation (spin waves) for information transmission and processing.[567,568,569,570] The basic elements of the SWD include: 1) magneto-electric cells (e.g. multiferroic elements) aimed to convert voltage pulses into the spin waves and vice versa;[571,572] 2) magnetic waveguides – spin wave buses for spin wave signal propagation between the magneto-electric cells,[573] 3) magnetic junctions to couple two or several waveguides, and 4) phase shifters to control the phase of the propagating spin waves. A SWD converts input voltage signals into the spin waves, computes with spin waves, and converts the output spin waves into the voltage signals. Computing with spin waves utilizes spin wave interference, which enables functional nanometer scale logic devices. Since the first proposal on spin wave logic,[567] the SWD concept has evolved in different ways encompassing volatile[574] and non-volatile,[575] Boolean[575,576] and non-Boolean,[577] single-frequency and multi-frequency circuits.[578] The primary expected advantage of SWD over Si CMOS are the following: 1) the ability to utilize phase in addition to amplitude for building logic devices with a fewer number of elements than required for transistor-based approach, 2) power consumption minimization by exploiting built-in non-volatile magnetic memory, and 3) parallel data processing on multiple frequencies in a single core structure by exploiting each frequency as a distinct information channel.

Micrometer scale SWD MAJ gate has been experimentally demonstrated.[579] It is based on $Ni_{81}Fe_{19}$ structure, operates within 1–3 GHz frequency range, and exhibits signal-to-noise ratio of approximately 10 at room temperature.[579] The internal delay of SWD is defined by the spin wave group velocity (e.g. 3×10^6 cm/s in $Ni_{81}Fe_{19}$ waveguides). Power dissipation in SWD is mainly defined by the efficiency of the spin wave excitation. Recent experiments with synthetic multiferroics comprising piezoelectric (lead magnesium niobate-lead titanate PMN-PT) and magnetostrictive (Ni) materials have demonstrated spin wave generation by relatively low electric field (e.g. 0.6 MV/m for PMN-PT/Ni).[580] The later translates in ultra-low power consumption (e.g. 1aJ per multiferroic switching). Recent experimental demonstration[572] of parametric spin wave excitation by voltage-controlled magnetic anisotropy (VCMA) is another promising route towards energy-efficient generation of spin waves.

Recently, a new type of SWD magnonic holographic memory (MHM), was proposed.[577] The principle of operation of MHM is similar to optical holographic memory, while spin waves are utilized instead of light waves. The first 2-bit MHM prototype based on yttrium iron garnet structure has been demonstrated.[581] MHM also possesses unique capabilities for pattern recognition by exploiting correlation between the phases of the input waves and the output interference pattern. Pattern recognition using MHM has been recently demonstrated.[582] The potential advantage of spin wave utilization includes the possibility of on-chip integration with the conventional electronic devices via multiferroic elements. In addition, magnonic holograms can show very high information density (about $1Tb/cm^2$) due to the nanometer scale wavelength of spin waves. According to estimates, the functional throughput of magnonic holographic devices may exceed 10^{18} $bits/s/cm^2$.[577]

There are several important milestones to be achieved for further SWD development: 1) nanomagnet switching by spin waves (e.g. by the combined effect of the voltage-assisted anisotropy change and a magnetic field produced by the incoming spin wave), and2) integration of several magneto-electric cells on a single spin wave bus. In order to have an advantage over CMOS in functional throughput, the operational wavelength of SWDs should be scaled down below 100 nm.[575] The success of the SWD will also depend on the ability to restore/amplify spin waves (e.g. by multiferroic elements or spin torque oscillators).

3.4.2. Nanomagnetic Logic

Nanomagnetic Logic (NML) uses fringing field interactions between magnetic islands to perform Boolean[583] and non-Boolean[584,585,586] logic operations. Binary information is represented via magnetization state. Most work with NML has focused on devices that couple in-plane (iNML). Recent work has also employed devices with out-of-plane, perpendicular magnetic

anisotropy (PMA). Perpendicular magnetic logic (pNML) devices typically switch through nucleation and domain wall propagation.[587,588] This reversal behavior is markedly different from in-plane permalloy nanomagnets, which remain nearly single-domain during switching and rotate coherently. That said, Bhowmick *et al.*[589] report using the spin Hall effect (SHE) to clock devices with PMA. With this approach, devices are placed in an in-plane, metastable state.

In all implementations of NML, externally supplied switching energy is generally needed to re-evaluate (i.e., "clock") a magnet ensemble with new inputs.[590] A clock modulates the energy barriers between magnetization states of the devices that comprise an NML circuit. With iNML, subgroups of devices are usually placed in a metastable state when clocked, and an input signal at one end can propagate through a part of said subgroup.[591] The other "end" of the subgroup is held in a metastable state to prevent back propagation of data. While pNML devices could be clocked in a similar manner (e.g., via the SHE[589]), it is also possible to globally clock a large ensemble of pNML devices (e.g., with an oscillating, out-of-plane magnetic field[586]). With either approach fine grain pipelining (and consequently improved throughput) could be achieved. (Subgroup size in iNML circuits will likely be limited by the granularity of the clock structure.)

Notably, either NML implementation can retain state without power and could be radiation hard. Moreover, it is possible that NML devices would dissipate less than 40 *kT* per switching event for a gate operation.[583] When clock overhead is considered, projections suggest that NML ensembles could still best low power CMOS equivalents in terms of metrics such as energy delay product, etc. Finally, NML appears to be scalable to ultimate limits using individual atomic spins.[592]

Present status for iNML: Fanout structures,[593] a 1-bit full adder,[594] etc. have been experimentally demonstrated and successfully re-evaluated with new inputs. Both field-coupled[595,596] and spin transfer torque (STT)[597,598,599] electrical inputs have been realized. For electrical output, multiple NML-magnetic tunnel junction hybrids[600] have been proposed and simulated. *Non-Boolean* hardware based on nanomagnetic discs has also been studied.[601] This work exploits the ability of nanomagnet systems to settle into energy minima to solve quadratic optimization problems (i.e., for computer vision applications).

Field based, CMOS compatible line clock structures have been used to simultaneously switch the states of multiple magnetic islands,[602] as well as to re-evaluate NML lines and gates with new inputs.[603] Additionally, recent experiments have considered materials-based solutions to further reduce field/energy requirements for line clock structures. Notably, results from Li *et al.*[604,605] suggest that component energy metrics could be further reduced by as much as 16X.

Voltage controlled clocking – e.g., multiferroics[606] and magnetostriction[607] – have also been proposed as potential clocking mechanisms for iNML. With respect to magnetostriction, this work suggests that stress-based clocking would lead to clock energy dissipation of just ~200 kT per device. Other recent work[608] has demonstrated strain-induced, 180-degree switching in anti-ferromagnetically coupled, single-domain magnets and short inverter chains. The SHE could be employed as a path to reduced energy clocking (i.e., 10-100X).[609]

Present status for pNML: Experimental results suggest that appropriately irradiated pNML structures[610] (to define dataflow directionality) can be controlled by a uniform, homogeneous, oscillating, global clock field.[611] Majority gates,[612] AF-lines,[611] multi-plane signal crossings,[613] and full adders[586,614] have all been experimentally demonstrated with this approach. Notably, Irina *et al.*[615] report the experimental demonstration of a three-dimensional majority gate (where input magnets reside in different planes). Each input combination was tested 50 times at room temperature. In each instance – and given the same clock window – the gate produced the correct output value. This could enable other compact and robust circuits (such as the five-magnet full adder discussed by Perricone *et al.*[616]). When considering I/O and clocking, field coupled electrical inputs have been demonstrated.[617] Also, large arrays of pNML devices could be controlled with on-chip inductor structures[618] that could be coupled with a capacitance in an LC oscillator – which opens the door to adiabatic energy recycling. Devices with PMA are also amenable to voltage-controlled clocking.[619]

Indeed, as noted above, small lines of pNML devices have been clocked via the SHE.[589] If 100 ps pulses are achievable, that work reports that the energy-delay product (EDP) of a pNML-based 32-bit adder would best that of low power CMOS. Individual pNML devices have also been clocked with on-chip inductor structures at 10 MHz. Simulation-based studies[620] suggest that a NAND/NOR operation would require just 2.8 aJ (assuming 200 nm × 200 nm devices and 50 MHz frequency and 10 functional layers of devices). Monolithic 3D NML co-processors have also been proposed.[621]

Active research on reliable switching work from 2008 suggests that in a soliton operating mode, dipole-to-dipole coupling could be insufficient to prevent thermal noise from inducing premature/random switching.[622] In that work, at the device-level, magnetocrystalline biaxial anisotropy could further promote hard axis stability of an iNML ensemble. Alternatively, adiabatic switching[623] and/or field gradients could potentially mitigate the effects of premature switching. Simulations suggest that NML circuits could tolerate some field misalignment.[624] Whether or not a circuit ultimately exhibits reliable and deterministic switching is very much a function of how it is clocked and requires additional study. More recently, nanomagnetic structures with spatially uniform clock fields,[625] leveraging shorter range exchange coupling between adjacent nanomagnets[626] (i.e., when compared to larger range parasitics that may be associated with dipole coupling), and nanoscale magnetic ratchets[627] (i.e., where specially

shaped magnets and strain are employed to achieve deterministic, 180-degree rotations) have all been explored and could potentially help to achieve more reliable magnetic switching.

Active research on fault tolerant architectures should be explored. As one example, stochastic computing (SC) could be an effective architectural strategy. Here, serial bit streams or parallel "bundles" of wires are used to encode probability values to represent and process information.[628] While bit streams/wire bundles are digital, information is conveyed through the statistical distribution of the logical values. With physical uncertainty, the fractional numbers correspond to the probability of occurrence of a logical one versus a logical zero. Computations in the deterministic Boolean domain are transformed into probabilistic computations in the real domain. Complex functions with simple logic are possible.

Bit flips afflict all the bits in the stream with equal probability. The result of bit flips is a mere fluctuation in the statistics of the stream, not a catastrophic error. Thus, SC enables meaningful computation even with high device error rates.[628] Furthermore, the SC architecture is pipelineable by nature. As the length of the stochastic bit stream increases, the precision of the value represented by it also increases. This allows a system to tradeoff precision with computation time. Thus, higher error rates that NML might experience can be tolerated by SC architectures. Pipelined SC architectures are a natural fit for inherently pipelined NML. Venkatesan *et al.*[629] and Perricone *et al.*[630] report examples of initial circuit design and benchmarking efforts in this area.

Active research on energy-efficient clocking including voltage controlled clocking should continue to be pursued – not only for benefits with respect to energy (e.g., as articulated by Nikonov and Young[631]), but also with respect to the fine-grained control it provides for an NML ensemble (which is useful in reducing error rates[632] as well as architecturally[618]). Domain-wall based switching could be another viable alternative for clocking.[633]

3.4.3. EXCITONIC DEVICES

Excitonic devices are based on excitons as computational state variables. Excitonic devices are suited to the development of an advanced energy-efficient alternative to electronics due to the specific properties of excitons: 1) Excitons are bosons and can form a coherent condensate with vanishing resistance for exciton currents and low switching voltage for excitonic transistors. This allows creating energy-efficient computation with power consumption per switch significantly smaller than in electronic circuits. 2) Excitonic signal processing can be directly coupled to optical communication in exciton optical interconnects. 3) The sizes of excitonic devices are scaled by the exciton radius and de Broglie wavelength that are much smaller than the photon wavelength. Furthermore, excitons can be efficiently controlled by voltage. This gives the opportunity to realize excitonic circuits at scales much smaller than for photonic devices.

The advantages listed above are realized using specially designed indirect excitons, IXs. An IX is a bound pair of an electron and a hole in separated layers. The properties of IXs make them different from conventional excitons and suitable for the development of energy-efficient computing: 1) IXs have oriented electrical dipole moments. As a result, the IX energy and IX currents are controlled by voltage allowing the realization of a field effect transistor operating with IXs in place of electrons. 2) The IX emission rate can be tuned over many orders of magnitude. Turning the emission off allows the realization of multi-element IX circuits with suppressed losses while turning it on allows fast write and readout. 3) The low overlaps between electrons and holes in IXs allow the realization of a coherent condensate with the suppressed thermal tails and dissipationless IX currents enabling energy-efficient computation.

Experimental proof-of-principle for excitonic devices including IX transistors,[634] diodes,[635] and CCD[636] was demonstrated. IX condensate and coherent IX currents[637] and, recently, long range coherent IX spin currents[638] were observed at low temperatures. New heterostructures based on single-atomic-layers of transition-metal dichalcogenide (TMD) for room-temperature IX circuits were proposed.[639] Recent progress includes the realization of IXs at room temperature in TMD heterostructures,[640] discovery of IX coherent spin currents in GaAs heterostructures[641], development of first split-gate device for IXs, the basic mesoscopic device[642], development of advanced platform for high-mobility excitonic devices[643], development of TMD heterostructures with small IX linewidth and finding charged IXs, indirect trions. At present, the studies are focused on the development of basic concepts of excitonic devices at low temperatures using GaAs heterostructures and development of room-temperature excitonic devices using TMD heterostructures.

3.4.4. TRANSISTOR LASER

The Light-Emitting Transistor (LET)[644] and Transistor Laser (TL)[645,646] utilize a fundamental characteristic of bipolar transistors - that electron-hole recombination in the base is an essential feature of the transistor and that the resulting photon signal in a direct-bandgap base is correlated to the electrical signal driving and being driven by the device. The TL can be thought of as a 5 terminal heterojunction bipolar transistor (HBT) with 3 conventional electrical terminals (emitter, base, and collector) and 2 optical terminals (input-generation and output-recombination).[647] Very-high-speed transmission and processing are enabled by the projected capability to achieve over 200 GHz bandwidths in the GaAs- or InP-based devices.[648] An advantage of the TL is that a single epitaxial layer structure can be used for devices that generate photons, detect photons, and perform electronic functions. The layer structure of the TL resembles a heterojunction bipolar transistor with features added to enhance base

recombination and control base transit time.[646,649] When used to realize conventional logic architectures, for example NOR gates,[650] the key advantage is speed. With processing-intensive operations and using the energy-delay product as a metric for comparison, a 30–100 times improvement is expected over conventional CMOS, leading to both faster processing and improved energy efficiency. An even greater benefit might be achieved through the use of architectures that perform electronic-photonic processing in the analog domain.

The first demonstration of lasing in the TL occurred in 2004. Since that time, progress has been made on understanding the device physics and on using the TL for discrete optical interconnects. A key initial objective has been examining factors affecting device bandwidth. Edge-emitting TLs with large active regions (200 μm × 1 μm) have been modulated to 22 Gbps and have shown a measured bandwidth of 10.4 GHz.[651] Relative intensity noise (RIN) as low as -151 dB/Hz has also been demonstrated, showing an approximately 28 dB improvement over diode lasers.[652] To improve speed and enable integration, reducing the size of the active region is critical. For that reason, Vertical-Cavity Transistor Lasers (VCTL) have been examined and demonstrated.[653] Initial VCTLs had limited temperature range due to misalignment of the cavity reflectivity and gain peaks. More recently, work has been underway to show that electronic-photonic logic can be made using the transistor laser. The initial target is a TL-based NOR gate.

The ultimate performance in both power and speed will be achieved as the device is size is scaled, as projected performance to bandwidths in excess of 100 GHz has a sound rationale but has yet to be realized. The use of vertical-cavity structures to reduce the device footprint has been a first step in the scaling process but further work is needed on microcavity vertical-cavity transistor lasers (VCTLs). Scaling beyond micron-scale devices such as this is possible, but the key will be the design of optical structures such as photonic crystal cavities that will enable small numbers of photons to be captured and directed to act on other TL structures. As scaling advances, further examination of device physics to reduce the effective minority carrier radiative lifetime will be key, along with the examination of effects that might impact the modulation response. Further work is also needed on InP-based devices (1310 and 1550 nm emission) to facilitate the use of silicon waveguides for optical signal routing. Additional questions at the architecture level involve the best way to use the TL in computer systems. What is enabled by having very high speed optical links? What architectures make sense for electronic-photonic NOR gates? Are other approaches to computation enabled, such as analog methods?

3.4.5. MAGNETOELECTRIC LOGIC

The earliest magneto-electric device concepts were based on a magnetic tunnel junction structure [1,2], which consists of two ferromagnetic (FM) layers separated by a non-magnetic insulator where the device resistance is determined by the relative orientation of the magnetization of the two FM layers. The magnetization of the free layer is exchange coupled to the Cr2O3 (or other magneto-electric) interface magnetization. A bias voltage applied across a magneto-electric layer, like chromia Cr2O3, layer reverses the interface magnetization, which in turn switches the magnetization of the free layer [3-5]. A lot of thought has been given to how this might be implements as logic elements [6-12], and to some extent benchmarked against CMOS [8,9,13]. The disadvantage is that while much faster than many spintronics device, there is long delay time in device operation, due to the slow speed of switching a ferromagnet.

Other magneto-electric devices, like the composite–input magnetoelectric–based logic technology (CoMET) [14] are limited by the switching speed of the ferroelectric and domain wall motion. The basis of these devices is an input switches a ferroelectric material, in contact with a ferromagnet with in-plane magnetic anisotropy placed on top of an intra-gate ferromagnet interconnect with perpendicular magnetic anisotropy. The input voltage nucleates a domain wall while a current is used to drive the domain wall to the output end of the device. A similar magneto-electric device structure, but now using spin-orbit coupling, has also been proposed [15]. Again, both devices will have long delay times, due to the slow switching speed of the ferromagnetic layer, and in the case of the CoMET device, the slow speed of domain wall propagation - where using higher currents to drive the domain wall at faster velocities comes at an energy cost. Also using spin orbit coupling, but now explicity also using a magneto-electric layer for electrical control of exchange bias of a laterally scaled spin valve is the nonvolatile magneto-electric spin-orbit (MESO) logic [16], but the delay time is limited again by the switching speed of the ferromagnetic layer.

More recently attention [5,17-23] has shifted to the magneto-electric spin field effect transistor (the ME-spinFET), where the most basic structure is outlined in Figure 1. Magneto-electric transistor schemes are based on polarization of the semiconductor channel, by the boundary polarization of the magneto-electric gate. The advantage to the magneto-electric field effect transistor is that such schemes avoid the complexity and detrimental switching energetics associated with magneto-electric exchange-coupled ferromagnetic devices. Spintronic devices based solely on the switching of a magneto-electric, will have a switching speed will be limited only by the switching dynamics of that magneto-electric material and above all are voltage controlled spintronic devices. Moreover, these magneto-electric devices promise to provide a unique field effect spin transistor (spin-FET)-based interface for input/output of other novel computational devices. This is spintronics without a ferromagnet, with faster write speeds (<20 ps/full adder), at a lower cost in energy (<20 aJ/full adder) [23], greater temperature stability

(operational to 400 K or more), and scalability, requiring far fewer device elements (transistor equivalents) than CMOS. These do differ from the conventional field transistor in that the ME-FET must be both top and bottom gated, son the result is that this is a 4, not 3 terminal transistor [17-22].

The anti-ferromagnet spin-orbit read (AFSOR) logic device structure (Figure 2) has interesting advantages: the potential for high and sharp voltage 'turn-on'; inherent non-volatility of magnetic state variables; absence of switching currents; large on/off ratios; and multistate logic and memory applications. The design will provide reliable room-temperature operation with large on/off ratios (>107) well beyond what can be achieved using magnetic tunnel junctions [21,22]. Again, the core idea is the use of the boundary polarization of the magneto-electric to spin polarize or partly spin polarize a very thin semiconductor, ideally a 2D material, with very large spin orbit coupling.

If the semiconductor channel retains large spin orbit coupling, then the spin current, mediated by the gate boundary polarization, may be enhanced and, to some extent, topologically protected. The latter implies that each spin current has a preferred direction, as indicated in Figure 2.

The silicon CMOS majority gate (left) requires 13 components. The ME-spinFET majority gate requires, including clocking, of 6 components. This represents an area improvement of over 50%, assuming similar size transistors. This is equivalent to greater than one process node. If we split the magnetoelectric side of the gate, so that the channel can independently be spin polarized up or down, this results in a component reduction from the previous best for the MEFET of six, down to four components, a further 50% reduction over the standard MEFET circuit, and a reduction to less than 30% in area compared to CMOS [24].

There is a variant of Figure 3, where inversion symmetry is not as strictly broken, that leads to a nonvolatile spintronics version of multiplexer logic (MUX). The magneto-electric spin-FET multiplexer (Figure 3) also exploits the modulation of the spin-orbit splitting of the electronic bands of the semiconductor channel through a "proximity" magnetic field derived from a voltage-controlled magneto-electric material. Here, by using semiconductor channels with large spin-orbit coupling, we expect to obtain a transverse spin Hall current, as well as a spin current overall. Depending on the magnitude of the effective magnetic field in the narrow channel, we anticipate two different operational regimes. Like the AFSOR magneto-electric spin FET, the magneto-electric spin-FET multiplexer in Figure 3 uses spin-orbit coupling in the channel to modulate spin polarization and hence the conductance (by spin) of the device. There is a source-drain voltage and current difference, between the two FM source contacts, due to the spin-Hall effect when spin-orbit coupling is present. This output voltage can be modulated by the gate or gates, which influences the spin-orbit interaction in the channel especially when it is both top and bottom gated especially. The spin-Hall voltage in the device can be increased by using different FMs in the source and drain. To increase the spin fidelity of current injection at the source end, one could add a suitable tunnel junction layer (basically a 1nm oxide layer) between the magnetic source and the 2D semiconductor channel. This latter modification would result in diminished source-drain currents though. Again, there is a reduction is delay time and energy cost because these devices are nonvolatile, there is no magnetization reversal of a ferromagnet involved and the implementation of this device concept would require only 5 components for a majority gate compared to the 13 components required of a silicon CMOS majority gate.

The challenges in pushing forward these technologies extends not only to the fabrication and characterization of a new generation of nonvolatile magneto-electric devices, but also to ascertaining the optimal implementation of CMOS plug in replacement circuits. Questions that need to be resolved include demonstration that the magneto-electric devices can be scaled to less than 10 nm, and this includes finding a suitable 2D channel material that can be polarized by exchange coupling with the boundary polarization of the magneto-electric and yet does not suffer from large scale edge scattering. Experimental demonstration of limits to the switching speeds of any antiferromagnetic magnetic electric remain absent. That said, the magneto-electric transistor has far fewer challenges to implementation than the magneto-electric magnetic tunnel junction, so, not surprisingly, there is a shifting of development effort toward those and related devices for both memory and logic.

4. Emerging Application Areas

4.1. Emerging Devices for Security Applications

4.1.1. Introduction

Like performance, power, and reliability, hardware security is becoming a critical design consideration. Hardware security threats in the IC supply chain, include 1) counterfeiting of semiconductor components, 2) side-channel attacks, 3) invasive/semi-invasive reverse engineering, and 4) IP piracy. A rapid growth in the "Internet of Things" (IoT) only exacerbates problems. While hardware security enhancements and circuit protection methods can mitigate security threats in protected components, they often incur a high cost with respect to performance, power and/or cost.

Advances in emerging, post-CMOS technologies may provide hardware security researchers with new opportunities to change the passive role that CMOS technology currently plays in security applications. While many emerging technologies aim to sustain

Moore's Law-based performance scaling and/or to improve energy efficiency,[654,655,656] emerging technologies also demonstrate unique features that could drastically simplify circuit structures for protection against hardware security threats. Security applications could not only benefit from the non-traditional I-V characteristics of some emerging devices, but also help shape research at the device level by raising security measures to the level of other design metrics.

At present, many emerging technologies being studied in the context of hardware security applications are related to designing physically unclonable functions (PUFs). Many post-CMOS devices[657,658,659] have been suggested as a pathway to a PUF design. (More detailed reviews are also available.[660]) With a PUF, challenge/response pairs are mapped (typically in a trusted environment). Responses are derived from natural/random variations and disorders in an integrated circuit that cannot be copied (or *cloned*) by an adversary. PUFs have been employed for tasks such as device authentication,[660] to securely extract software,[661] in trusted Field Programmable Gate Arrays (FPGAs),[662] and for encrypted storage.[661] Post-CMOS devices also find utility as random number generators (RNGs) that may be employed for secure communication channels (e.g., to generate session keys[660]). That said, while intriguing, PUFs and RNGs may only cover a small part of the hardware security landscape. (Furthermore, one must be careful that PUF designs based on emerging technologies do not depend on device characteristics that a designer would like to *eliminate* when considering utility for logic or memory.)

Given the many emerging devices being studied[654] and that few if any devices were proposed with hardware security as a "killer application," this document also reports initial efforts as to how the unique I-V characteristics of emerging transistors that are not found in traditional MOSFETs could benefit hardware security applications.

Below, we review the efforts described above, beginning with efforts to design PUFs and RNGs with emerging technologies. How device characteristics can enable novel circuits to achieve hardware security-centric ends such as IP protection, logic locking, and the prevention of side channel attacks are also discussed.

4.1.2. PHYSICALLY UNCLONABLE FUNCTIONS (PUFS) AND EMERGING TECHNOLOGIES

A variety of different emerging logic and memory technologies have been considered in the context of PUFs. As has been reviewed,[660] variations in the required write time in spin torque transfer random access memory (STT-RAM) was proposed to create a domain wall memory PUF.[657] Other structures based on magnetic tunnel junctions have also been proposed.[663,664] Variations in write times have also been exploited to produce unique responses in phase change memory (PCM) arrays.[665] The variability of ReRAM presents a natural opportunity for PUF implementation, and array demonstration has been reported.[666, 667] At the array-level, variations in diode resistivity have also been used to derive challenge/response pairs from crossbar structures.[668] PUFs based on graphene[669] and carbon nanotubes have also been proposed/considered.[670]

As a more representative case study, prior work[660] considers an array structure based on process variation in memristors[671,672] to create a PUF structure (referred to as *NanoPUF*[660]). NanoPUF is based on 1) a crossbar with memristors. 2) A challenge is applied to the memristor array by using a row decoder to apply a voltage amplitude (V_{dd}) to a given row that can vary in duration; a column decoder connects a given column to a resistance R_{load}. All other rows and columns remain floating. 3) A response circuit (to collect outputs to different challenges) would consists of R_{load} and a current comparator that compares I_{out} from a given column to a reference current I_{ref}. A logic 1 might be recorded if $I_{out} > I_{ref}$, while a logic 0 might be recorded if $I_{out} < I_{ref}$. With respect to PUF functionality, when a write pulse is applied, natural process variations will cause some memristors to turn on (leading to a logic 1), and others to remain off (leading to a logic 0). While the time of the right pulse serves as one variable,[671] the pulse's duration and amplitude may also be varied.[672]

4.1.3. RANDOM NUMBER GENERATORS (RNGS) AND EMERGING TECHNOLOGIES

The inherent randomness in emerging devices can also be used to generate random numbers.[660] As a representative case study, prior work[660] explores an approach based on contact-resistive random access memory (CRRAM).[673] (Note that a CRRAM device may be based on a layer of silicon dioxide that is sandwiched between two electrodes; the bottom electrode could simply be the drain of a CMOS transistor – which in turn suggests that RNGs based on emerging technologies can be CMOS compatible.[674])

During operation, the current flowing in a filament channel will be (randomly) impacted by any electrons trapped in the insulating layer. If a high voltage is applied to a device, the current in the filament channel will be large and not impacted by trapped electrons. However, with the application of a lower voltage, the width of a filament will shrink, and the trapped electrons *will* (randomly) influence output current.[674] Indeed, RNGs based on emerging devices[674] can successfully pass randomness tests such as those provided by the National Institute of Standards and Technology (NIST).

As random number are derived from current passing through filaments, memristors, PCM, and RRAM devices can also be leveraged to build similar RNGs.[660] Additional device options for random number generation are discussed in section 5.3.1.

4.1.4. OTHER HARDWARE SECURITY PRIMITIVES BASED ON EMERGING TECHNOLOGIES

Below, other security-centric primitives (non-PUFs and non-RNGs) based on emerging technologies are also discussed. How new devices might be employed for IP protection and to prevent side channel attacks are considered. In each section, device characteristics of interest are discussed first. Subsequent discussions then consider how device characteristics can be employed to achieve a security centric end.

4.1.4.1. EMERGING TECHNOLOGIES FOR IP PROTECTION

Tunable Polarity: In many nanoscale FETs (45nm and below), the superposition of n-type and p-type carriers is observable under normal bias conditions. The ambipolarity phenomenon exists in various materials such as silicon,[675] carbon nanotubes[676] and graphene.[677] By controlling ambipolarity, device polarity can be adjusted/tuned post-deployment. Transistors with a configurable polarity – e.g., carbon nanotubes,[678] graphene,[679] silicon nanowires (SiNWs),[680] and transition metal dichalcogenides (TMDs)[681] – have already been experimentally demonstrated.

As more detailed examples, SiNW FETs have an ultra-thin body structure and lightly doped channel which provides the ability to change the carrier type in the channel by means of a gate. FET operation is enabled by the regulation of Schottky barriers at the source/drain junctions. The control gate (CG) acts conventionally by turning the device on and off via a gate voltage. The polarity gate (PG) acts on the side regions of the device, in proximity to the source/drain (S/D) Schottky junctions, switching the device polarity dynamically between n- and p-type. The input and output voltage levels are compatible, enabling directly-cascadable logic gates.[682]

Ambipolarity is an inherent property of TFETs due to the use of different doping types for drain and source if an n/i/p doping profile is employed.[683] By properly biasing the n-doped and p-doped regions as well as the gate, a TFET can function either as an n- or p-type device, and no polarity gate is needed. As the magnitude of ambipolar current can be tuned (i.e., reduced) via doping or by increasing the drain extension length,[683] one can envision fabricating devices that could be better suited for logic as well as security-related applications. Given that the screening length in TMD devices scales with their body thickness, one can achieve substantial tunneling currents.

Polymorphic logic gates: The ability to dynamically change the polarity of a transistor opens the door to define the functionality of a layout or a netlist post fabrication. Though one may use field programmable gate arrays (FPGAs) to achieve the same goal, FPGAs cannot compete with ASICs in terms of performance and power, and an FPGA's reliance on configuration bits being stored in memory introduces another vulnerability. Security primitives to be discussed can serve as building blocks for IP protection, IP piracy prevention, and to counter hardware Trojan attacks.

Polymorphic logic circuits provide an effective way for logic encryption such that attackers cannot easily identify circuit functionality even though the entire netlist/layout is available. However, polymorphic logic gates have never been widely used in CMOS circuits mainly due to the difficulties in designing such circuits using CMOS technology.

SiNW FET based polymorphic gates to prevent IP piracy have been introduced.[684,685] If the control gate (CG) of a SiNW FET is connected to a normal input while the polarity gate (PG) is treated as the polymorphic control input, we can easily change the circuit functionality through different configurations on the polymorphic control inputs without a performance penalty. For example, a SiNW FET based NAND gate can be converted to a NOR gate, whereas a CMOS-based NAND cannot be converted to a fully functioning NOR by switching power and ground.

TFET-based polymorphic logic circuits have also recently been developed.[686] By properly biasing the gate, the n-doped region, and the p-doped region, a TFET device can function either as an n-type transistor or p-type transistor. If the n-doped region of the two parallel TFETs is connected to V_{DD}, and the p-doped region of the bottom TFET is connected to GND, the circuit behaves like a NAND gate. If the n-doped region of the two parallel TFETs is connected to GND and the p-doped region of the bottom TFET is connected to V_{DD}, the circuit behaves as a NOR gate. By using two MUXes (one at the top and the other at the bottom) to select between the two types of connections, the circuit then functions as a polymorphic gate where the control to the MUXes forms a 1-bit key.[686]

One can readily design polymorphic functional modules using the low-cost polymorphic logic gates built from either SiNW FETs or TFETs that only perform a desired computation if properly configured. If some key components (e.g., the datapath) in an ASIC are designed in this manner, the chip is thus encrypted such that a key, i.e., the correct circuit configuration, is required to unlock the circuit functionality. Without the key, invalid users or attackers cannot use the circuit. Thus, IP cloning and IP piracy can be prevented with extremely low performance overhead. A 32-bit polymorphic adder using SiNW FETs has been designed and simulated. Two pairs of configuration bits (with up to 32-bits in length) are introduced and the adder can only perform addition functionality if the correct configuration bits are provided.

Camouflaging Layout: Split manufacturing and IC camouflaging are used to secure the CMOS fabrication process, albeit with high overhead and decreased circuit reliability. With CMOS camouflaging layouts, both power and area would increase

significantly in order to achieve high levels of protection.[687] A CMOS camouflaging layout that can function either as an XOR, NAND or NOR gate requires at least 12 transistors. Emerging technologies help reduce the area overhead. Recent work has demonstrated that only 4 SiNW FETs with tunable polarity are required to build a camouflaging layout that can perform NAND, NOR, XOR or XNOR functionality.[688,689] Again, the SiNW FET based camouflaging layout has more functionality and requires less area than CMOS counterparts and could offer higher levels of protection to circuit designs.

Security Analysis: Logic obfuscation is subject to brute-force attacks. If there are *N* polymorphic gates incorporated in the design, it would take 2^N trials for an attacker to determine the exact functionality of the circuit. As the value of *N* increases, the probability of successfully mounting a brute-force attack becomes extremely low. In a preliminary implementation of 32-bit adder, the incorporated key size is 32 bit.[686] The probability that an attacker can retrieve the correct key becomes $1/2^{32}$ (2.33×10^{-10}). Obviously, polymorphic based logic obfuscation techniques are resistant to a conventional brute-force attack. With respect to camouflaging layouts, given that our proposed SiNW based camouflaging layout can perform four different functions, the probability that an attacker can retrieve the correct layout is 25%. Therefore, if *N* SiNW FET camouflaging layouts are incorporated in a design, the attacker has to compute up to 4^N times to resolve the correct layout design. Compared to polymorphic gate-based logic obfuscation, camouflaging layout embraces higher security level but with larger area overhead.

4.1.4.2. EMERGING TECHNOLOGIES TO PREVENT SIDE-CHANNEL ATTACKS

Many post-CMOS transistors aim to achieve steeper subthreshold swing, which in turn enables lower operating voltage and power. Many devices in this space also exhibit I-V characteristics that that are not representative of a conventional MOSFET. An example of how to exploit said characteristics for designing hardware security primitives is discussed.

Steep slope transistors: TFETs have been exploited to design current mode logic (CML) style light-weight ciphers.[690,691] The high energy carriers in TFETs can be filtered by the gate-voltage-controlled tunneling such that a sub-60 mV/decade subthreshold swing is achievable at room temperature.[692] With improved steep slope and high on-current at a low supply voltage, TFETs could enable supply voltage scaling to address challenges such as undesirable leakage currents, threshold voltage reduction, etc. Different types of TFETs have been developed and fabricated.[692,693]

Bell-Shaped I-Vs: Emerging transistor technologies may also exhibit bell-shaped I-V curves. Symmetric graphene FETs (SymFETs) and ThinTFETs are representatives of this group. In a SymFET, tunneling occurs between two, 2-D materials separated by a thin insulator. The I_{DS}-V_{GS} relationship exhibits a strong, negative differential resistance (NDR) region. The I-V characteristics of the device are "bell-shaped," and the device can remain off even at higher values of V_{DS}. The magnitude of the current peak and the position of the peak are tunable via the top gate (V_{TG}) and back gate (V_{BG}) voltages of the device.[684] Such behavior has been observed experimentally.[694,695] More specifically, V_{TG} and V_{BG} change the carrier type/density of the drain and source graphene layers by the electrostatic field, which can modulate I_{DS}. ITFETs or ThinTFETs may exhibit similar I-V characteristics.[696]

Preventing fault injection: Side-channel analysis, such as fault injection, power, and timing, allows attackers to learn about internal circuit signals without destroying the fabricated chips. Countermeasures have been proposed to balance the delay and power consumption when performing encryption/decryption at either the algorithm or circuit levels.[697] These methods often cause higher power consumption and longer computation time in order to balance the side-channel signals under different conditions. Thus, an important goal is to prevent fault injection and to counter side-channel analysis by introducing low-cost, on-chip voltage/current monitors and protectors. Graphene SymFETs, which have a voltage-controlled unique peak current can be used to build low-cost, high-sensitivity circuit protectors through supply voltage monitoring.

Recent work has developed a SymFET-based power supply protector.[684,685] With only two SymFETs, the power supply protector can easily monitor the supply voltage to ensure that the supply voltage to the circuit-under-protection is within a predefined range. In the event of a fault injection, the decreased supply voltage will power down the circuit rather than injecting a single-bit fault,[685] and can thus protect the circuit from fault injection attacks. If one uses V_{out} as the power supply to a circuit under protection (e.g., an adder), due to the bell-shaped I-V characteristic of the SymFET, an intentional lowering of V_{DD} cuts off the power supply. Thus, the sum and carry-out of the full adder output is '0', and no delay related faults are induced. A similar CMOS power supply protector would require op-amps for voltage comparison. As a result of the voltage/current monitors developed thus far, voltage/current-based fault injections can be largely prevented. By inserting the protectors in the critical components of a given circuit design, the power supply to these components can be monitored and protected.[686] (SymFET-based Boolean logic is also possible.[698])

Preventing differential power analysis (DPA): As an advanced side-channel attack scheme, DPA employs analysis of statistic power consumption measurements from a crypto system to obtain secret keys. Since the introduction of DPA[700], there has been many efforts to develop low-cost and efficient countermeasures. Countermeasures are generally classified into two categories: 1) algorithm-level solutions and 2) hardware-level solutions.

Algorithm-level solutions aim to design cryptographic algorithms that can withstand a certain amount of information leakage,[699] e.g., frequently changing the keys to prevent the attacker from collecting enough power traces[700] or using masking bits during the internal stages to limit information leakage.[701]

A more practical circuit-level method for preventing DPA attack leverages a *sense amplifier-based logic* (SABL) or *current mode logic* (CML) for cryptographic algorithm implementations.[702] A CML gate includes a tail current source, a current steering core and a differential load. A CML gate will switch the constant current through the differential network of input transistors, utilizing the reduced voltage swing on the two load devices as the output. Although CML is not widely used in mainstream circuit design, its unique features, namely low latency and stable power consumption, can be leveraged to serve as a countermeasure against a DPA attack.

The strength of the CML-based approach is the constant power consumption of differential logic which can counter power-based attacks as operation power is independent of processed data. The drawback with these (mostly CMOS-based) logic designs, is their large area and power consumption when compared to static single ended logic. When considering hardware for the IoT (where the systems can be severely power constrained), system designers are presented with a dilemma in which they need to choose either high security or low power consumption. Emerging transistor technologies could help mitigate risks of DPA attacks while maintaining low power consumption.

Recent work has implemented a standard cell library of TFET CML gates and conducted a detailed study of their performance, power and area with respect to CMOS equivalents.[703] Standard cells were used to implement and evaluate TFET-based CML on a 32-bit KATAN cipher (a light-weight block cipher). All KATAN ciphers share the same key schedule with the key size of 80 bits as well as the 254-round iteration with the same non-linear function units.[704]

The two CML implementations consume less gate equivalents and area compared to the two static counterparts given that the majority of KATAN32 is made up by the D flip flops. The area of TFET CML KATAN32 is 1.441 μm^2, which is about 60% less than the Static TFET KATAN32. The power consumption of TFET CML (9.76 μW) is slightly lower than *static* CMOS (9.96 μW). It also outperforms CMOS CML.

Moreover, the correlation coefficient of a TFET static KATAN32 reaches its highest when the correct keys are applied. By comparison, the correlation coefficient of TFET CML KATAN32 is much more scattered, and all four hypothetical keys are equally distributed. Thus, the TFET CML KATAN32 implementation can successfully counteract CPA. Because the power consumption is mainly determined by AND/XOR logic gates of two nonlinear functions – and the effect of CPA is maximized – the correlation coefficients for KATAN32 are higher on average than other block ciphers, e.g., CPA on S-box[705].

4.2. Cryogenic Electronics

With the 2018 Edition of the IRDS Roadmap, findings on cryogenic electronics and quantum information systems are captured in a standalone chapter, accessed at this link.

5. Emerging Device-Architecture Interaction

5.1. Introduction

Emerging Beyond-CMOS devices will likely have new issues at the interfaces between the device level and higher levels (*e.g.*, circuit, architecture and application) of computer design. Many of the emerging device classes that are currently being explored are not intended simply as “drop-in” replacements for CMOS devices, but will require new types of circuit designs, new functional module architectures, and even new software to best utilize the new devices’ capabilities.

Novel design issues that span the device and architecture levels need to be considered when adopting new low-level computing paradigms. Devices may be organized in radically new ways to carry out computation in a very different style from what we may consider the most “conventional” computing paradigm, which has relied on standard combinational and sequential irreversible Boolean logic. Examples of such unconventional or alternative computing paradigms include:

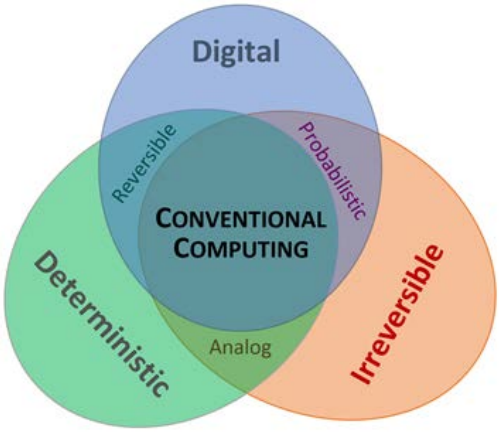


Figure BC5.1. Categorization of Conventional vs. Alternative Computing Paradigms Addressed in This Section

Note: The conventional computing paradigm is designed to be digital, deterministic, and irreversible. Typical analog computing schemes relax the digital requirement while maintaining (at least approximate) determinism and are often physically irreversible. In probabilistic computing, we abandon the requirement for determinism, while preserving the irreversible, digital nature of the computation. And typical reversible computing techniques maintain the digital, deterministic nature of computation while attempting to avoid irreversibility. More generally, alternative computing paradigms may respect only one (or none) of the three illustrated constraints.

Analog computing (§5.2) – This broad category of non-digital computing paradigms includes some of the neural models (discussed below) but also spans a much broader range of other analog approaches such as coupled oscillators (§5.2.3.1),[706] energy-minimization/annealing systems (§5.2.4), and chaotic circuits (§5.2.3.2).

Neural (including neuromorphic or neural-inspired) computing (§5.2.2) – Many of the new devices that have been considered for use in this sub-paradigm of analog computing are envisioned to play a role analogous to biological synapses. At a higher level, a variety of overall neural computation schemes (analog, spiking, etc.) can be envisioned. While neural models (like any analog model) can be simulated in conventional digital CMOS, a more direct circuit-level implementation potentially offers substantial performance improvements (per unit power consumption and/or per unit cost). Two-terminal synapse-like devices are frequently considered for organization in crossbar arrays (§5.2.1) which then offers the potential to perform very efficient analog matrix operations (§5.2.1.1, §5.2.1.2). A variety of different physical state variables (*e.g.*, capacitive, resistive, magnetic) have been explored for encoding the synapse weights.

Probabilistic/stochastic circuits (§5.3) – In relation to §4.1.3, devices and circuits that produce truly nondeterministic or random outputs at the hardware level may be useful for accelerating probabilistic algorithms such as Monte Carlo or simulated annealing, for generating secure cryptographic keys, and for other applications.

Reversible (adiabatic and/or ballistic) computing (§5.4) – Computing paradigms that approach logical and physical reversibility offer the potential to greatly exceed the energy efficiency of all other approaches to general-purpose digital computation. Primitive devices for reversible computing may include devices having fairly conventional functions (such as switches or oscillators). These devices would need to be optimized differently to use quasi-reversible physical processes such as near-adiabatic state transitions, near-ballistic signal propagation, highly elastic interactions, and highly underdamped oscillations. Reversible devices also must be organized into circuits and architectures in tightly constrained ways, for reversibility at the logical as well as physical level.[896] Careful fine-tuning and optimization of analog circuit characteristics (*e.g.*, resonator quality factors or elasticity of ballistic interactions) remains a difficult and crucially important engineering challenge that must be met in order for this paradigm to realize its promise.

Quantum computing – Quantum computing[707] offers the potential to carry out exponentially more efficient algorithms for a variety of specialized problem classes.[708] Devices for quantum computing are very different from conventional devices, and fine-tuning their characteristics to avoid decoherence while organizing them effectively into scalable architectures has, to date, proved to be a formidable engineering challenge.[709] IRDS is beginning to address quantum computing this year in the new Cryogenic Electronics & Quantum Information Processing chapter (CEQIP). We will not address quantum computing further in the present section.

This categorization can be understood by reference to Figure BC5.1. However, the reader should please note that the material in this section does not comprise an exhaustive list of *all* possible new computing paradigms, new devices, new circuits, or new architectures. It is only intended to serve as a representative sample of several new general computing paradigms and specific technology concepts and as a starting point for further discussion of this overall subject area in future workshops.

5.2. ANALOG COMPUTATION

Analog computing attempts to let "physics do the computation" by using physical processes more directly (as opposed to through the traditional digital abstraction barrier) to compute complex functions. Historically, this required inefficient analog circuitry for all elements, and expensive analog to digital conversion, resulting in limited applications, specifically those requiring analog signal processing. Recently, new analog devices have enabled a new generation of efficient analog architectures. This is especially true for hybrid analog and digital systems where efficient designs may exploit analog preprocessing and computation prior to digitization. Analog preprocessing can reduce the required A/D precision and therefore reduce the system energy consumption by orders of magnitude.[710] Additionally, new architectures can be used for ultra-low-power co-processors for conventional CMOS designs. A key challenge is that analog signals are typically low precision, with energy and latencies increasing exponentially with higher bit precision. Fortunately, many machine learning and other applications are being developed that can tolerate such lower precision computation.

In this subsection, we review a number of recent examples of analog computational technologies, starting with some currently popular neural-inspired architectures that also have broader applications in linear algebra.

5.2.1. CROSSBAR ARCHITECTURES

Analog crossbars or memory arrays can perform low-precision matrix operations in parallel, by processing analog data directly at each memory element. Thus, n 1990, Carver Mead projected that custom analog matrix vector multiplications would be thousands of times more energy efficient than custom digital computation.[711] Because a digital memory must individually access each memory cell and move the data to a separate computation unit, digital systems consume more energy and incur longer latencies. Computing on larger crossbars/matrices allows for any analog overhead to be averaged out over many matrix elements. Any two- or three-terminal device that features a modifiable internal physical state variable (which might be, for example, a variable resistance, a variable capacitance, a stored charge, or a stored magnetic field) that modifies the device's behavior can be used as a building block for analog operations. Several different types of crossbar architectures are summarized in the following sections.

5.2.1.1. MATRIX VECTOR MULTIPLICATION (MVM) AND VECTOR MATRIX MULTIPLICATION (VMM)

MVM and VMM are key computational kernels underlying many different algorithms. There are several approaches to accelerating these kernels. Any programmable resistor such as a two-terminal resistive memory or a three-terminal floating gate cell can be used. Alternatively, a capacitive MVM can be designed by adding charge from capacitive memory elements.[712]

For many algorithms such as neural network inference (of an already-trained network), accelerating MVM accelerates the bulk of the computation, allowing for large system level gains. An $N \times N$ crossbar accelerates $O(N^2)$ operations, leaving only $O(N)$ inputs and outputs that need to be processed and communicated. This allows each unit of communication and processing costs to be amortized over N memory elements.

Analog VMMs have been used for experimental demonstration of threshold logic,[713] compressed sensing initial filtering,[714] robotic navigation and control,[715] adaptive filtering,[716] Fourier transforms[717] and more. Additionally, Analog VMM techniques have been used for ultra-low power classification and neural networks.[718]

5.2.1.1.1. RESISTIVE MVM AND VMM

Resistive MVMs are based on using Ohms law, $V = I \times R$, to perform multiplication, and Kirchhoff's current law on of to perform addition by summing currents. Programmable resistors are used to program the weights. Arranging the memory elements in an array allows for the entire operation to be performed in a single parallel step, giving a fundamental $O(N)$ energy and latency advantage over a standard digital memory that, at best, must access a memory array one row at a time.[719] An MVM and the transpose VMM can be performed on the same memory array, by changing whether the rows of an array are driven and the columns are read, or vice versa.

The key metrics for a resistive MVM are 1) the energy per multiply and accumulate, 2) latency per MVM, 3) crossbar and supporting circuitry area per matrix element, 4) crossbar dimensions, 5) input/output bit precision for digitally driven MVMs, and 5) the standard deviation of the noise or error per weight when programmed as a percentage of the weight range.

If high-resistance memory elements ($R_{on} = 100$ MΩ) with good analog properties are developed, one ReRAM based crossbar design projects that each multiply and accumulate operation will require 12 fJ when using 8-bit A/Ds and only 0.4 fJ when using 2-bit A/Ds.[720] The latency for a 1024×1024 MVM will only be 384 ns or 11 ns for 8-bit or 2-bit A/Ds respectively. This is over $100 \times$ better than an optimized SRAM based accelerator, which would require 2,700 fJ and 4,000 ns for 8 bits. The area per weight for the 8-bit A/D ReRAM accelerator is 0.05 μm^2, $16 \times$ better than the 0.8 μm^2 needed for an SRAM accelerator. The energy and latency are dominated by the A/D circuitry and not by the crossbar itself, with the A/D converters and digital circuitry occupying $10 \times$ the area of the ReRAM array itself.

To allow for large arrays and minimize parasitic resistance drops, high resistance (~100 MΩ) memory elements are needed. The higher the resistance, the larger the array possible, and the more any A/D costs and system level communication costs are amortized out. However, such high resistances would prolong and potentially complicate the process of programming each conductance value accurately to encode already-trained neural network weights or matrix element values.

A key design choice is the bit precision of the inputs and outputs to the crossbar. The fewer the bits needed by an algorithm, the more efficient the crossbar is. If analog or binary inputs/outputs can be used, the A/D costs can also be avoided. The inputs to the crossbar can be encoded in voltage, time, or digitally. Voltage encoding applies different voltages to represent different analog input values. This requires circuitry to create different input voltages, and it requires that the memory elements have a linear I-V relationship, greatly complicating the use of nonlinear access devices.[721] Encoding inputs in variable length pulses requires longer reads and an integrator to sum the resulting current. Digital encoding applies each bit of the input sequentially and then combines the result digitally. For digital encoding, the usefulness of lower order bits in the input is limited by the noise/errors on the highest order bit.

The precision with which each resistor needs to be programmed depends on the particular application. For neural network inference, the algorithm can be robust to read noise of up to 5% of the weight range or more.[722]

Many different memory elements can be used for the programmable resistor:

***ReRAM*:** ReRAM memory is very dense and can be integrated in the back end of the line, avoiding the use of chip area for the memory. A key challenge is maintaining good analog properties and high resistance at the same time. A thermal confinement method was proposed to realize analog ReRAM[723] and demonstrated 1Kb ReRAM array for facial recognition.[724] Several ReRAM crossbars have been demonstrated for inference.[725, 726]

Phase Change Memory: Phase-change memory (PCM) offers a wide range of analog memory states due to the large contrast between the amorphous and crystalline phases.[727] Key challenges for inference include reducing resistance drift due to amorphous relaxation and ensuring long retention at high operating temperatures. Cell designs that can suppress resistance drift have been proposed.[728, 729] Most VMM results using PCM to date have focused on in-situ training (see section on Outer-Product Update below).[752]

***LISTA/ENODe*:** New three terminal battery inspired devices including the Li-Ion Synaptic Transistor (LISTA)[730] and the electrochemical neuromorphic organic device (ENODe)[731] were recently proposed as an analog memory element. They are based on charging/discharging a battery. As a battery is charged, the resistance of the cathode changes. These devices have shown very good analog control but are currently too slow for training accelerators, with write pulses on the order of a millisecond. Since these devices are effectively tiny batteries, an access device is needed on the third "gate" terminal, to prevent the device from discharging when the state is not being written in an array configuration.

Floating Gate: Floating gate-based synapses were first developed in 1994.[732] They are modified EEPROM devices and can be fabricated in a standard CMOS process. They can be programmed to within 0.2% accuracy,[733] and full systems have been developed to allow arbitrary devices in an array to be programmed.[734] The FG VMM is a slightly modified EEPROM array with the associated density from such a structure. EEPROM devices already store 4 bits (16 levels) in a single transistor of 100 nm × 100 nm area in 32nm process.[735] Commercial EEPROM manufacturers have shown devices at 15 nm and 19nm[736,737] and [738]). EEPROM devices are found on every CMOS IC node, including 7nm and 11nm nodes. At these nodes, we still expect very small capacitors to retain 100s of quantization levels (7-10 bits) limited by electron resolution; in practice, larger capacitors are used, resulting in sufficiently high potential resolution. One expects EEPROM linear scaling down to 10nm process to result in a 30 nm × 30 nm or smaller array pitch area.

Capacitor-on-Gate: A recent proposal was to tie a small capacitor that can be programmed with standard CMOS devices to the gate of a read transistor[739]. In contrast to DRAM, where the charge on the capacitor is transferred through a select transistor onto a bitline for readout, here the voltage on the capacitor modulates the conductance of a read transistor by direct attachment to its gate terminal. Although the charge leaks away with a time-constant of milliseconds, the training process can succeed if the time-per-example is at least 100,000x smaller than the decay constant (e.g., 20ms decay constant and 200ns-per-training-example).[740] In order to have good update linearity so that the amount of charge added and subtracted are balanced, additional large transistors can be used in each unit cell.[739,740] Recently, Ambrogio et al. introduced a combined PCM+capacitor-on-gate unit-cell, in which the PCM provided the non-volatile storage, the capacitor-on-gate devices provided high update linearity with a small number of transistors, and variability between charge-addition and charge-subtraction was compensated upon weight transfer from capacitor-on-gate cell to the PCM devices using a "polarity inversion" technique[741]. This unit-cell was shown to allow GPU-equivalent training accuracies, despite the known imperfections of PCM devices and typical fab-level CMOS variability in the capacitor-on-gate devices.[741]

5.2.1.1.2. CHARGE-BASED ANALOG ARRAYS FOR VMM

Charge-based analog arrays are amenable to very low energy and high-density parallel implementations of vector-matrix multiplication (VMM). Efficient charge storage and weighting in array-based analog computing are achieved through the use of capacitive reactive elements or other charge-based linear weighting elements such as charge-coupled devices (CCDs) and charge-injection devices (CID). Their efficiency stems from inherent charge conservation throughout the computational cycle.

Charge injection device (CID) arrays store each bit of the matrix element in a DRAM storage element. The charge for each bit in a weight is stored in one of two locations. If the input bit that is multiplying the weight bit is 1, the charge is non-destructively shifted between locations during readout causing a charge to be capacitively induced on the bit line. The charge induced by multiple weights can be summed and sensed allowing the entire matrix to be read out in a single operation. Multi-bit inputs are processed serially. Furthermore, charge is recycled during the computation, and so adiabatic techniques can be used to further lower the energy (at the cost of speed).

Table BC5.1 Metrics for Analog Capacitive Vector-Matrix Multiply (VMM) ICs

High-density mixed-signal adiabatic processors[742,743,744] using CIDs have been developed using these principles. To optimize for resonant adiabatic energy recovery a stochastic encoding and decoding scheme can be used to ensure a constant capacitive load of the CID array. This has resulted in better than 1.1 TMACS/mW efficiency excluding on-chip digitization.[744]

Alternatively, several approaches have combined a capacitive charge based VMM with analog to digital conversion to maintain high overall system efficiencies. Many analog multiply-and-accumulate operations can be performed for each digitization. High precision implementations of capacitive charge based VMM have achieved low-pJ/MAC energy efficiencies,[745] while low-precision versions have achieved efficiencies at the level of 100 fJ/MAC.[746] Comparison of key metrics with the state-of-the-art in analog capacitive VMM ICs[747,748,749,750,745] is given in Table BC5.1 above.

5.2.1.2. OUTER PRODUCT UPDATE (OPU)

Analog resistive memory crossbars can also perform a parallel write or an outer product, rank 1 update where all the weights are incremented by the outer product of a vector applied to the rows and a different vector applied to the columns. This is a key kernel for many learning algorithms such as backpropagation[722] and sparse coding.[719,751] Row inputs are encoded in time and column inputs are encoded in either time[751] or voltage.[722]

When a VMM and MVM are combined on the same crossbar, extremely efficient learning accelerators can be designed,[720,752,753] with the potential to be 100–1000× more energy efficient and faster than an optimized digital ReRAM or SRAM based accelerator.[720]

The same figures of merit and design considerations for a VMM apply to the OPU. Additionally, the 1) write noise and 2) asymmetric write nonlinearity are important for determining how well a learning algorithm will perform. The 3) ability to withstand failures, and 4) endurance are also important for training systems. To have an efficient learning accelerator, parallel blind updates are needed where weights are updated without knowing the previous value and without verifying that the correct value is written. To obtain ideal accuracies, a low write noise is needed, less than 0.4% of the weight range. Even more important is having low asymmetries in the write process. The change in conductance for a positive pulse should be the same as that for a negative pulse for all starting states.[722] Often the conductance will saturate near a maximum where a positive pulse will not change the conductance, while a single negative pulse will cause a large decrease in conductance. This significantly lowers accuracy as it only takes a single negative write pulse to cancel many positive pulses.

Several devices have been examined for neural network training, including phase change memory,[752] resistive memory[720,754] and novel lithium-based devices.[730,731] Currently no devices meet all the ideal requirements for training (high resistance >10 MΩ, low write noise <0.4%, low write asymmetry[722]). Nevertheless, algorithmic approaches such as periodic carry,[755] Local Gains,[756] or the inclusion of semi-volatile capacitor-on-gate devices[741] can be used to help compensate and achieve ideal accuracies. Several co-design tools have been developed to model the impact of device level properties on algorithmic performance[752,757,758] which have allowed for the algorithmic development needed. Additionally, lower resistance devices can be used to give smaller near-term gains in performance.

The need for high on-state resistance and good analog characteristics means that filamentary resistive memories may not work as well as non-filamentary devices. A resistance higher than a quantum of conductance, 13 kΩ, requires current to tunnel through a barrier. This presents a fundamental problem for a filamentary device: a single atom can halve that tunneling barrier, resulting in huge variability and poor analog characteristics.

5.2.1.3. LARGE-SCALE FIELD PROGRAMMABLE ANALOG ARRAYS (FPAAS)

A field programmable analog array has configurable analog blocks and configurable interconnects between those blocks.[759,760,761] FPAAs allow users to build analog applications without having expertise in IC design. The fundamental breakthrough was recognizing that a switch matrix of single floating gate elements could be used for analog computation. The routing crossbar networks were, in fact, crossbar networks that could support VMM computation as well as a number of other computations.[762] Routing was no longer dead weight, as perceived for FPGA architectures. The floating gate cells could also allow for mismatch calibration at the mismatch source.[763,764] The density for VMM in FPAA architectures is nearly the level of custom IC design. These analog computations can be made robust to temperature fluctuations.[765] These techniques have been utilized by a number of students in university courses.[766,767] FPAA based VMMs can be scaled to small geometry (*e.g.* 40 nm and smaller) and operated at RF frequencies.[768,769]

5.2.1.4. RESISTIVE MEMORY CROSSBAR SOLVER

Resistive memory crossbars can be used to solve matrix problems, $Ax = b$ for x as illustrated in Figure BC5.2.[770] Each resistive memory is a programmable resistor that represents a matrix element in A. (Alternatively, any programmable analog element can also be used.) The equation $Ax = b$ can be mapped to Ohm's law, $\sum G_{ij}(V_j - V_i) = I_i$. The V_i are set to zero by an op-amp. Currents, I_i, are applied to the crossbar and the resulting voltages, V_j, are measured. The op-amps provide feedback allowing the V_j to be determined.

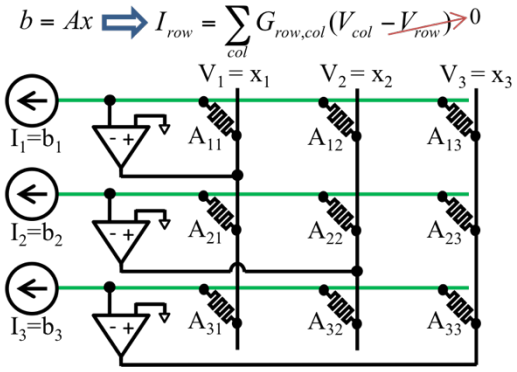


Figure BC5.2 A Resistive Memory Crossbar $Ax = b$ Solver is Illustrated

Note: The op-amps are used to provide feedback to find the solution and force the row voltages to zero.

The biggest challenge in taking advantage of analog solvers for HPC is that analog operations only offer low precision, ~8 bits fixed point, while HPC applications often demand 32 or more bits of floating point precision. This can be potentially overcome by hybrid analog/digital systems where the computationally intense parts of a calculation can be done in analog, while the required precision can be achieved by refining the solution in digital using a method with lower computational complexity[771]. This allows for some digital computation, while still getting a reduction in the overall computational complexity. The precision can potentially be improved by using iterative refinement or by using the crossbar to initialize a digital solver. The analog solution can also be used as a preconditioner within a Krylov method like CG or GMRES. Large matrices can be broken down into smaller blocks compatible with the accelerator and scaling can be used to compensate for finite on/off ranges. Work is still needed to show how noisy crossbar solvers can be used with ill-conditioned matrices. In general, iterative refinement will only converge if the noise is less than $1/\kappa$ where κ is the condition number of a matrix.

5.2.1.5. TERNARY CONTENT ADDRESSABLE MEMORY (TCAM)

Content-addressable memory (CAM), or associative memory, compares input search data against a table of stored data and returns the address of matching data. Ternary CAM (TCAM) allows for partial matches or "don't care" values. Conventional TCAM cells require multiple transistors per cell, resulting in a low-density, high cost cell, which could only be used in applications where speed is critical. Resistive memory crossbars can be used to build a high density TCAM that do not require any transistors per memory cell.[772,773,774]

The TCAM array accesses the entire memory in parallel, similar to an MVM, allowing for extremely efficient search operations. To realize a TCAM the resistive memories must be able to reliably switch between two different resistance states that differ by at least three orders of magnitude. Second, to maintain competitive energy dissipation, the resistive memories should have a high resistance >500 kΩ. The higher the ON/OFF ratio and the higher the resistance, the larger the array that can be built. A key challenge in the practical realization of TCAM arrays is the need for highly reliable devices and/or error correction that is compatible with a TCAM array.

5.2.2. NEURAL COMPUTING

Broadly speaking, neural (or neuromorphic or neural-inspired) computing models consist of a network of localized elements or "neurons," and connections or "synapses." As discussed in §5.2.1, analog crossbars can be used as building blocks for conventional neural networks to give significant improvements in energy, latency and area over digital accelerators. For accelerators specialized to a particular algorithm, various analog neurons can be used to process the crossbar outputs and avoid the need for and high cost of analog to digital conversion.

There is also a lot of work on more biologically inspired neural hardware. For biologically inspired neurons, connections are often designed to be persistent on short timescales, but (depending on the model) may exhibit mutability/plasticity in their strength and/or topology on longer timescales to facilitate, for example, adaptive in-situ learning. Connections between neurons can be modeled as discrete events or spikes (often implemented using an address event representation) or continuous-valued analog signals such as voltage or current. A key challenge for more biologically inspired architectures is the need for algorithmic co-design. For many neuro-inspired computational models, further research is needed to demonstrate state of the art machine-learning performance.

5.2.2.1. ANALOG NEURON FUNCTIONAL BLOCK

Depending on the algorithm being accelerated, analog neurons can accelerate a variety of functions such as:

1. spatial and temporal integration of input signals
2. stochastic firing of an action potential after a threshold is reached
3. relaxation to the resting potential if the time delay between excitations is too long.[775,776,777]
4. Non-linear activation functions. In many neural networks, after a linear multiply accumulate is performed on a set of inputs, or in a hidden layer, a non-linear activation or transfer function is applied. The presence of this function prevents the network from mathematically collapsing into a single linear equation, which helps improve the computational capabilities as the number of layers increases. Common functions include sigmoid, tanh and rectified linear (ReLu). In an analog network it is desirable that this function be performed by a single device.

These analog blocks are often with other building blocks such as crossbars, §5.2.1, to accelerate neural networks, to build dynamical systems, §5.2.3, create energy minimization circuits, §5.2.4 and enable probabilistic circuits, §5.3. Stochastic neurons/devices are discussed in §5.3 rather than here. Using an analog neuron allows analog data to be processed directly, obviating the need for time- and energy-consuming ADC/DAC in the circuits.

Recently, Mott memristors,[778] phase change based memristive switches,[779] and chalcogenide threshold switches[780] have all been reported to be capable of performing temporal voltage signal integration in which the effects of non-simultaneous unitary postsynaptic potentials add in time. Device candidates to perform the transfer function include STT-MRAM.[781]

Ionic diffusion dynamics or electrical instabilities enable a single memristive device to perform analog functions like resistance tuning or pulse generation after accumulating input,[779,782] which typically requires a large number of CMOS transistors. These analog functions are critical in emulating synaptic and neuronal behavior. Taking into account the simple structure and scalability of a two terminal memristor, both the complexity of a circuit and the area consumed to build a neuron network will be much less compared to a CMOS circuit. Being a passive device, the memristor offers stimulation dependent electric conductance. However, the passive nature results in a lack of power to maintain sustainable neural signal propagation in a network if passive synaptic blocks are employed. Additional power sources are mandatory for neural networks with multiple layers.

The benchmarks used to evaluate electronic spiking neurons generally measure the energy per operation, the fabrication cost, or the chip area of the integrated functional block, and the fidelity to the desired neuron function (e.g. integrate and fire). High reliability and low variation of devices are two key factors for the viability of a neuron technology. Device failure will cost a lot more circuitry for error detection and correction.[783] Large variation increases the difficulty for designing peripheral circuits and degrades the adaptability of the block.

To achieve unsupervised learning in a network, particularly those based on spike-timing-dependent plasticity, neuron blocks should be capable of programming the synaptic blocks. A key design challenge will be engineering the forward and back-propagating action potentials from the analog neuron blocks to potentiate or depress synapses for real time and *in situ* learning.

5.2.2.2. CELLULAR NEURAL NETWORK

The cellular neural network (CeNN) is a non-Boolean computing architecture that contains an array of computing cells that are connected to nearby cells. Since interconnects are major limitations in modern VLSI systems, CeNN systems take advantage of the local communication and encounter fewer constraints imposed by interconnects. The CeNN is a brain-inspired computing architecture that relies on neurons to integrate the incoming currents. The accumulated and activated output signal drives nearby neurons through weighted synapses. CeNNs can be used to create associative memories for voice and image recognition.

CMOS based CeNNs can be implemented by analog circuits using operational amplifiers and operational transconductance amplifiers (OTAs) as neurons and synapses, respectively.[784,785] Some recent work has also investigated CeNN using beyond-CMOS charge-based devices, such as TFETs, to potentially improve energy efficiency[786,787] thanks to their steep subthreshold slope and low operating voltage.

Using novel devices whose dynamics match the dynamics state of cells in CeNN can be far more efficient than op-amp and OTA based CeNNs. For instance, magnet switching dynamics that follow the Landau-Lifshitz-Gilbert (LLG) equation with a spin-transfer-torque term[788,789] are quite similar to the cell dynamics in CeNN. CeNNs have been designed based on spin diffusion using all-spin logic (ASL) with PMA magnets as the basic building block.[790] A CeNN cell can also be implemented by using MTJs as synapses and using spin Hall effect or domain wall propagation based devices as the neuron.[791] In all these cases, the read-out circuit consists of read and reference MTJs and an inverter that amplifies the voltage division between the two MTJs.

The associative memory application is widely used in the tasks of voice and image recognition, which can be efficiently performed in the CeNN architecture.[792,793] Five numbers, '1' – '5', are associated with the other five numbers, '6' – '0'. Hebbian learning is used for storing patterns.[794] Patterns with noisy input pixels can still be recalled. The delay per CeNN operation is dependent on the input pattern, input noise, and thermal noise.

In Fig. BC5.3, CeNN architectures based on the following devices are benchmarked: ferroelectric negative capacitance FET (NCFET),[795,796] tunneling FET (TFET),[797,798,799,800] piezoelectric FET (PiezoFET),[801] graphene pn junction switch (GpnJ),[802] various FETs based on two-dimensional materials,[803,804] all-spin logic (ASL),[805] charge-coupled spin logic (CSL),[806] and domain wall logic (mLogic).[807] In addition, the benefit of adding an extra layer of the copper plate between the SHE material and the free magnet is investigated to enhance spin injection through lateral diffusion.[808]

Fig. BC5.3 shows the performance comparison among three types of CeNN implementations using 4-bit weight synapses to achieve 90% recall accuracy for a given input noise of 10%. For the charge-based CeNN implementation, CMOS HP and LV devices are employed to quantify the performance of the digital CeNN and to compare against their analog counterparts. It is shown that the digital CeNNs are quite power hungry and slow. This is because multiple cycles are required to read out the weights from the register and perform the summation in the adder, which is energy and time consuming. In general, the analog CeNNs implemented by TFETs dissipate less energy thanks to their steep subthreshold slope and lower supply voltage. For optimal circuit-level performance, the preferred device properties for charge-based devices are steep threshold swing, large bias current, low supply voltage, and low capacitance.

In contrast to Boolean circuits, spintronic devices are more competitive compared to charge-based devices. This is because a single magnet can mimic the functionality of a neuron, and these spintronic devices operate at a low supply voltage. The domain wall device provides the best performance, in terms of Energy-Delay Product, thanks to its low critical current requirement. The spin diffusion based CeNN with IMA magnets consumes more energy due to the large critical current required to switch the magnet. For optimal circuit-level performance using spintronic devices, several properties are desired including: MTJs with a large TMR and a moderate resistance-area product, large spin injection coefficient β, large perpendicular anisotropy Ku for PMA magnets, large spin Hall angle θ for SHE materials, and small critical depinning current for domain wall magnets.

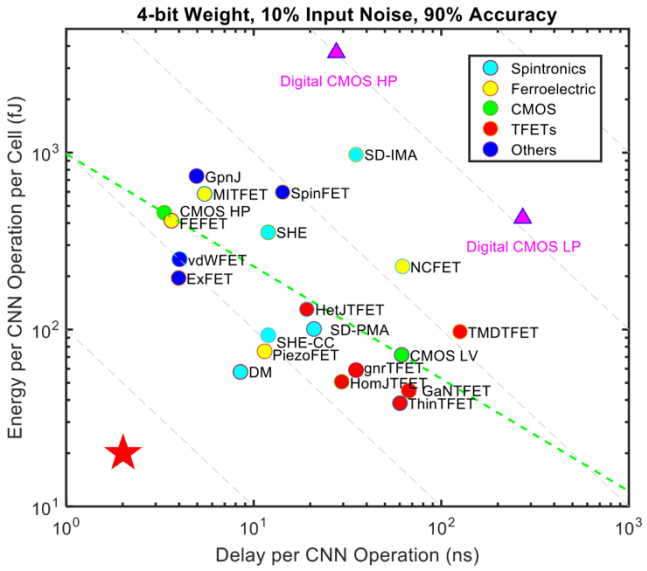


Figure BC5.3 Comparison of Energy and Delay per Operation Among Various Beyond-CMOS Technologies for CeNN Based on Analog, Digital, and Spintronic Implementations

Note: For the text labels, SD, SHE, and DW stand for spin diffusion, spin Hall effect, and domain wall motion, respectively, and CC represents the copper collector.

5.2.3. COMPUTING WITH DYNAMICAL SYSTEMS

In computing with dynamical systems, the built-in dynamical behavior of a physical system exhibiting continuous degrees of freedom is used to compute. The entire computational process can be analog, with only the results being digitized. The following subsections give a few examples of different types of dynamical systems-based approaches.

5.2.3.1. COMPUTING WITH COUPLED OSCILLATORS

A popular class of computing models utilizing continuous-time dynamical systems are the *coupled oscillator models*.

5.2.3.1.1. DEVICE TECHNOLOGIES FOR COUPLED OSCILLATOR SYSTEMS

It is challenging to build compact, low-power oscillators that can also be coupled together to give predictable phase or frequency dynamics. Standard digital ring oscillators based on CMOS inverter feedback loops, as well as the typical transistor-driven LC oscillators used in RF designs, are both less than ideal, due to device nonlinearities in the digital regime and the high biasing currents used in linear-regime analog small-signal oscillators, as well as relatively large oscillator sizes in both cases. Thus, various "Beyond CMOS" oscillator technologies have been explored. Such technologies can use and manipulate the charge, spin, or quantum properties of electrons, or use photons. Important examples include spin-torque,[809,810,811] insulator-metal-transition,[812] optical,[813,814] and quantum.[815] Memristor based oscillators have also been constructed.[816,817,818]

Non-silicon electrical oscillators include two important kinds which are currently being developed. One prominent effort is the use of spin torque oscillators (STOs) coupled with using spin diffusion currents, or electrical signals, for providing a computational platform for machine learning, spiking neural networks, and others.[809,810,811,813,819,820,821] However, the high current densities of STOs and the limited range of spin diffusion currents continue to pose serious challenges in creating coupled networks of such oscillators. Optical oscillators have been studied[814, 822] and used for computing,[813] but challenges include bulky components, difficult interfacing between the electrical and optical domains, and lack of programmability to enable an optical computing apparatus. Another promising non-silicon technology for very compact oscillators is the IMT (insulator-metal transition) material-based oscillator technology.[823,824] As the oscillation mechanism is completely electrical, the coupling of oscillators can be done easily using electrical components. There have been other implementation efforts for electrical oscillators[825,826,827,828,829] but the focus has been to build high frequency and low power individual oscillators, as opposed to the demonstration of coupled systems of oscillators, or the generation of interesting dynamics for computing. A comparison of some computing-focused electrical oscillators is shown in Table BC5.2 below:

Table BC5.2 Comparison of Some Electrical Oscillators for Computing

5.2.3.1.2. THEORETICAL MODELS OF COUPLED OSCILLATOR SYSTEMS

Coupled oscillator models can explain many natural, chemical and biological synchronization phenomena like the synchronized flashing of fireflies, pacemaker cells in the human heart, chemical oscillations, neural oscillations, and laser arrays, to name a few.[830] The simplest theoretical models of oscillators start with sinusoidal oscillators which have been extensively studied,[831,832,833] and their application in the computational paradigm has been well demonstrated.[834,835] A generalized description of oscillators in these models is usually a canonical phase model,[830,836] where the coupling mechanisms are generally assumed to be weak and composed of simple periodic functions which explicitly depend on phases. This model calls for simple harmonic oscillators with coupling that is assumed to affect each other's phases linearly. If the oscillators with phases θ_1 and θ_2 have frequencies ω_1 and ω_2, then in Kuramoto models the coupling of two oscillators will result in the phase equations:

$$\dot{\theta}_1 = \omega_1 + k(\theta_1 - \theta_2)$$
$$\dot{\theta}_2 = \omega_2 + k(\theta_2 - \theta_1)$$

where k is the coupling constant. A *Kuramoto* system of N oscillators is described by

$$\dot{\theta}_i = \dot{\omega}_i + \frac{K}{N}\sum_{j=1}^{N} \sin(\theta_j - \theta_i), \quad i = 1, \ldots, N$$

where θ_i and ω_i are the phase and frequency respectively of the i^{th} oscillator. Several studies on more general periodic coupling functions have also been studied.[837]

Along with sinusoidal oscillators, non-linear Van der Pol oscillators[838] and several of its variants, like the Morris-lecar neuron model,[839] have also been studied, and the applicability of such models in neurobiological and chemical oscillators has been demonstrated.[840,841,842] A single Van der Pol oscillator is defined by adding a non-linearity in the simple harmonic oscillator model

$$u'' + \epsilon(u^2 - 1)u' + u = 0$$

which results in relaxation behavior. Hence these are also called *relaxation* type oscillators. These analyses also assume non-realistic coupling dynamics like weak or pulse coupling, and do not focus on engineering aspects of building such coupled oscillators.

Such analytic models of coupled oscillatory systems almost always require a canonical phase description of the oscillators and a periodic phase dependent additive coupling that can be classified as weak. Strong coupling for relaxation type oscillators built using electrical circuits lack good explanations. Some theoretical studies have focused on pulse coupling,[843,844,845] and injection locking,[846,847] but these models are not suitable for understanding coupled relaxation oscillators that are based on repeated charging and discharging of a capacitor.[848, 849,850] These oscillators show piecewise linear dynamics instead of continuous dynamics as in the previous models, and the analysis of coupling is rather difficult as the limit cycle spans different "pieces" of the dynamics.

5.2.3.1.3. Computing Models Using Coupled Oscillators

Computing models using coupled oscillators can be categorized into pairwise coupling and multi-oscillator coupling. Pairwise coupling, *i.e.* coupling of just two oscillators, has simpler dynamics and hence can be used for simple operations like analog subtraction in a discriminant circuit.[851] Multi-oscillator coupling can be used for computing complex operations. Coupled oscillator based associative memories[835,852,853] have been proposed, where the dynamics can retrieve stored patterns and the steady state corresponds to the memorized pattern closest to the input, but successful implementations have yet to come. Another important application for combinatorial optimization, specifically graph coloring, was described[849] and developed theoretically with support from experimental demonstrations using relaxation oscillators based on phase-change IMT materials. Another model using multi-oscillator coupling is the Ising Machine; an implementation of this model using Optical Parametric Oscillators (OPO) was shown.[813]

5.2.3.2. Sub kT Computing Using Chaotic Logic

Shannon's noisy channel coding theorem[854] shows that one can reliably communicate information on a channel subject to noise (at a sufficiently low bitrate) even when the transmitted signal power is below the noise floor (*i.e.*, at a signal-to-noise ratio of less than 1). Moreover, any computational process can be viewed as just a special case of a communication channel, namely, one that simply happens to transform the encoded data in transit—since the derivation of Shannon's theorem relies solely on counting distinguishable signals, and nothing about how the signals are being counted in Shannon's argument precludes the encoded data from being transformed as it passes through the channel. This observation suggests that performing reliable computation utilizing signal energies (that is, energies associated with the dynamical variables in the system) at average levels $\ll kT$ (*i.e.*, well below the thermal noise floor) should theoretically also be possible—although the output bit rate (per unit signal bandwidth) will scale down with the average signal energy.

In 2016, Frank and DeBenedictis investigated a theoretical approach for implementing digital computation using chaotic dynamical systems[855,856,857] which provided evidence that the above theoretical observation is correct. In that approach, the long-term average value of a chaotically evolving dynamical degree of freedom encodes a digital bit. The interactions between degrees of freedom are tailored such that the bit-values represented by different degrees of freedom correspond to the results that would be computed in an ordinary Boolean circuit. This method can also be considered an example of the more general category of analog energy-minimization-based approaches (§5.2.3.3 and §5.2.4.2). However, this method does not require cooling the system to low noise temperatures, as is frequently done in energy-minimization approaches. Instead, the dynamical network uses a variation on adiabatic/reversible computing principles (§5.4) to adiabatically cause the system to transition between different warm, chaotic "strange attractors" that represent different computational states reversibly, without energy loss. The dynamical energy of the signal variables is itself conserved within the (Hamiltonian) dynamical system, and so the total energy dissipated per result computed can approach zero in this model as the rate of transformation decreases.

One disadvantage of the particular approach explored in that work is that it exhibits an apparent exponential increase in the real time required for convergence of the results as the complexity of the computation (number of logic gates) increases. However, as far as is known at this time, it is possible that faster variations on this or similar techniques may be found with further investigation.

5.2.3.3. Other dynamical systems for computing

Apart from coupled oscillators and chaotic logic, a number of other dynamical system models have been studied for various applications, but few have been implemented in some kind of hardware with proven advantages over corresponding digital

implementations of algorithms. Hopfield networks are attractor networks proposed for associative memories[858] where the fixed points (or stable states) of the system correspond to memories, and the dynamics of the network is such that the system settles to the fixed point which is closest to the initial state the system starts from. Hopfield style models have also been used for optimization.[859]

Cellular neural networks[860] (addressed in more depth in §5.2.2.2 above) consist of interconnected nodes where each node has linear or non-linear dynamics and the connections specify the coupling between their corresponding differential equations. Their CMOS implementations have been proposed with applications like pattern matching. Ising Machine[813,861,862] models have been proposed based on coupled spin glasses. The energy minima of such networks correspond to the solutions of an NP-hard combinatorial optimization problem and hence can model other NP-hard problems as well.[861] Another dynamical system for constraint satisfaction[863] is built on similar principles. An architecture based on non-repeating phase relations[864] between fabricated CMOS oscillators tries to emulate stochastic local search (SLS) for constraint satisfaction problems. An interesting approach based on memory co-processors was introduced as Memcomputing.[865] Interesting insights can also be obtained by looking into dynamical systems like iterated maps,[866] cellular automata,[867,868,869] and 0-1 continuous reformulations of discrete optimization problems.[870]

Although some of the above-mentioned methods target NP-hard problems, it's important to note that, to date, no general physical computing method (including analog and quantum approaches) has yet been clearly demonstrated to be capable of solving NP-hard problems without requiring exponential physical resources (energy and/or time) to be invested in the physical process performing the computation. The prevailing belief among computational complexity theorists[871] is that solving NP-hard problems efficiently would require uncovering new physics (*i.e.*, beyond standard quantum mechanics).

5.2.4. ANALOG ARCHITECTURE & ENERGY MINIMIZATION

Table BC5.3 Some Useful Analog Computing Design Patterns

In the past, IRDS principally roadmapped digital logic and memory, but the expansion to analog involves moving beyond the digital design paradigm. Digital systems comprise gates and DRAM that use analog devices internally yet create a digital interface to wiring. A wiring network then defines an architecture, possibly hierarchical, that is largely independent of the analog properties of the wires.

The digital design paradigm can be adapted to analog, but the resulting design paradigm is more complex. An analog system has a variety of signal types, such as voltage, current, or frequency, that are transported over the wiring. Modules based on nanodevices may be passive or too small to include a powerful signal drive, often making the analog properties of the wire an important part of the circuit's behavior.

General properties of computing systems have been studied with results such as Landauer's implementation-independent minimum energy dissipation for irreversible digital operations.[872] However, "analog" and "digital" are both implementation approaches for computing systems, so one may expect that there should be an equivalent to Landauer's minimum energy for irreversible *analog* systems—and indeed, at least in some cases, there is,[873] but such analog limits are not as well known.

5.2.4.1. BUILDING BLOCKS OF ANALOG SYSTEMS

The analog systems of interest today can be expressed as the hierarchical composition of a set of building blocks, just like digital electronics and software. There is no universally accepted set of building blocks at the moment, but Table BC5.3 (above) lists some commonly-encountered analog building blocks or design patterns that can be seen to apply to many of the analog architecture concepts discussed in this section.

5.2.4.2. ENERGY MINIMIZATION FOR THE TRAVELING SALESMAN PROBLEM

Some principles of analog system design will now be illustrated using a system that solves the Traveling Salesman Problem.

The Traveling Salesman Problem is to find the best route for a salesman that needs to visit a series of cities, each pair separated by some distance. The salesman seeks the shortest route that visits each city exactly once and then returns to the starting city. Exact solutions to this problem are known in computational complexity theory to be NP-hard, meaning that it is too hard for any computer, analog or digital, to solve exactly, in general, and at large scale. However, there are many ways to compute approximate solutions, meaning finding a short route but not necessarily the shortest route.

An example solution uses energy minimization using multiple runs on a Hopfield network. A Hopfield network is essentially one of the popular neural network arrays with its outputs fed back to its inputs. The synapse weights define an energy landscape based on binary values emerging from the neurons. If a Hopfield network is initialized or driven to a particular combination neuron values and then the drive is removed, the network will follow a trajectory that will take it to a minimum, or local minimum, of the energy landscape.

To use a Hopfield network for the Traveling Salesman Problem, the synapse weights are set to encode the $n \times n$ intercity distances. The system drives the neurons to a starting point for the salesman's route and removes drive. The system will settle into a candidate salesman's route in an amount of time equal to a few time constants of the feedback loop.

The method described above may find the ideal solution or just a better but suboptimal solution. To improve the odds of finding the best solution, the circuit can include either a true random noise generator or a chaotic pseudo-random noise generator. Under control of an external digital computer, the Hopfield network is driven to a random starting point and released many times in a cycle. The randomness causes many of the starting points to be different, making it more likely that the system will find the global minimum. The digital computer collects all the results, checking each to see which is best.

Such a system leverages multiple new circuit blocks including both a crossbar to encode the $n \times n$ intercity distances, an analog neuron to run the Hopfield network and either a chaos or noise generator to get randomness. As the Traveling Salesman problem is NP-hard, no solution method can solve it exactly at scale. Nevertheless, the analog solution seems to be at least comparable in efficiency with some software algorithms. It is also one of the more sophisticated analog computing examples, including all the aspects from Table BC5.3.

5.2.5. ANALOG COMPUTING—CLOSING REMARKS

Analog computation, as surveyed above, presents us with an intriguing and varied array of options for transcending the limitations that apply to the present-day digital approaches to computing. However, more work is needed to better characterize the range of applications for which analog computation can provide significant advantages over digital.

Enormously complex digital information-processing systems have been constructed by leveraging hardware description languages and programming languages that enable encapsulation, composability, and hierarchical design. In order to enable complex analog systems, more flexible, powerful languages (graphical and/or textual) for representing general classes of analog circuits and architectures are needed, to allow for a similar "modular" design approach.

Even the most advanced and sophisticated analog computational structures reviewed above in this section are only just the beginning. It appears that analog computing represents a vast field of future study, one that would likely benefit from a much more intensive level of exploration than it has received to date.

5.3. PROBABILISTIC CIRCUITS

Traditionally, conventional computational processes are designed to be deterministic, with computational results determined by the machine's initial state and inputs. Nevertheless, computations that are *intentionally* designed to behave randomly or stochastically, even at the level of individual bit-operations, are of interest and can have many useful applications such as simulated annealing, Monte Carlo simulation, machine learning or Boltzmann machines, randomized algorithms[874] as studied in computational complexity theory, and cryptographically secure random number generation for generating secure private keys. Noise in biological neurons is beneficial for information processing in nonlinear systems, and is essential for computation and learning in cortical microcircuits.[875,876,877,878]

Obtaining randomness in traditional CMOS is difficult and typically relies on a pseudo-random number generator. This requires a large circuit block and significant computational effort to obtain high quality random numbers. Several new devices have been proposed to obtain true randomness as discussed in §5.3.1. These allow for a random bit to be generated with a single device. Chaotic devices can be used to turn poor quality randomness into high quality random numbers. New architectures such as probabilistic p-logic (§5.3.2) or a travelling salesman solver (§5.2.4.2) can be designed using the new true random number generators.

5.3.1. DEVICE TECHNOLOGIES FOR RANDOM BIT GENERATION

New devices based on memristors, avalanche breakdown, and magnetic tunnel junctions and other technologies have been proposed for generating random bits. A key enabling functionality for some architectures like probabilistic (p)-logic is the ability to tuneably control the probability of a zero or one based on an input current or voltage. Several proposed devices are listed below. (See also §4.1.3.)

Magnetic Tunnel Junctions (MTJ): Existing Embedded MRAM technology can be used to create a tuneable random bit, provided that the Magnetic Tunnel Junctions are engineered to be thermally unstable. Such thermally unstable magnets have been experimentally observed. As MTJ dimensions are scaled, keeping them thermally stable becomes a hard challenge for memories, therefore destabilizing them in a controllable manner should be feasible in current technology.

Low-barrier MTJs can convert ambient thermal noise on nanomagnets into a fluctuating resistance, which is then used to build a device with tunable randomness when integrated with minimal CMOS periphery. The fluctuating resistance change due to thermal magnetic noise in MTJs can be measured by Tunneling Magneto resistance (TMR). State-of-the-art TMR values range from upwards of 100% to 600% demonstrated by the Tohoku Group,[879] and commercial STT-MRAM devices

exhibit >100% TMR. A large TMR would enable a robust functional unit for controllable randomness. The theoretical limit for TMR in MgO-based MTJs has been reported[880] to be 1,000% and can presumably be larger. There is currently intense research activity in half-metallic ferromagnets to increase TMR.

Single-Electron Bipolar Avalanche Transistor (SEBAT): The single-electron bipolar avalanche transistor (SEBAT) is a novel Geiger-mode avalanche bipolar transistor structure.[881,882] The device generates Poisson-distributed digital output pulses at rates between 1kHz and 20MHz. The pulse rate is linearly proportional to the emitted current. A MOS transistor is also formed within the base region of the device, allowing for voltage control of the pulse rate. The device is fully compatible with low-voltage CMOS circuits and standard digital process steps.

Memristors/Resistive RAM: The intrinsic variability of memristive switching, particularly the switching delay time of memristors, can be a good source of stochasticity.[660,883,884] Such stochasticity originates from the ionic dynamics within the memristors.[885]

Contact-resistive random access memory (CRRAM): CRRAM can be used for random number generation.[660,673] A CRRAM device may be based on a layer of silicon dioxide that is sandwiched between two electrodes; the bottom electrode could simply be the drain of a CMOS transistor.[674] During operation, the current flowing in a filament channel will be (randomly) impacted by any electrons trapped in the insulating layer. If a high voltage is applied to a device, the current in the filament channel will be large and not impacted by trapped electrons. However, with the application of a lower voltage, the width of a filament will shrink, and the trapped electrons *will* (randomly) influence output current.

CMOS: There are different ways to obtain random number generators (RNG) in CMOS using different physical noise sources, one being "jitter" in ring oscillators.[886] These TRNGs can be tuned into tunable random number generators as required but require significant amounts of area when compared to single device alternatives.

5.3.2. PROBABILISTIC (P)- LOGIC

In a series of recent papers, Camsari and Datta proposed a device-to-systems framework for a probabilistic (*p*)-logic[887,888,889,890,891,892,893] introducing the concept of *p*-bits and *p*-circuits. The authors explored how *p*-bits can be enabled by existing technology and showed different applications of them including image recognition,[894] combinatorial optimization,[890] Bayesian inference[887] and an enhanced type of Boolean logic including "invertible" operation.[888,889]

The main function of a *p*-bit is to provide a *tunable* randomness at its output terminal controlled by an input terminal. The tunability of randomness allows a network of *p*-bits (*p*-circuits) to be able to get correlated with one another when appropriately connected. This enables the *p*-bits to behave as a *hardware* building block for a range of systems level applications. By externally fixing any of the bits in a *p*-circuit, the other bits will stochastically oscillate between values that are allowed by the circuit. This allows circuits to be operated in "reverse." For example, a *p*-circuit implemented as a standard digital multiplier can function as a divider (or factorizer) operating in "reverse" as a consequence of the input/output equivalence of *p*-circuits.[888] Such "invertible" units are envisioned as part of a larger system that can be used as efficient hardware accelerators for specific problems that may not come as naturally to standard CMOS.

5.4. REVERSIBLE COMPUTING

Besides analog computing (§5.2) and probabilistic computing (§5.3), a third dimension along which we may explore departures from the conventional computing paradigm, as illustrated in Fig. BC5.1, is *reversible computing*.[895] When a computation is *reversible* in this context, we mean that the lowest-level computational processes should be arranged to approach a condition of being both *logically reversible* and *thermodynamically reversible*. To say that a computation is *logically reversible* means that known or deterministically computed information is not discarded from the digital state of the machine and ejected to a randomizing thermal environment. To say that it approaches being *thermodynamically reversible* here means that the total increase in physical entropy incurred by the machine's operation per useful computational operation performed is extremely small, with the vision that it can be brought closer to zero asymptotically as the technology continues to be improved.

In 1961, Rolf Landauer of IBM argued[872] that there is a fundamental physical limit on the energy efficiency of conventional *irreversible* digital operations, meaning those that carry out a many-to-one transformation of the space of computational states that is used. Landauer's limit states that an amount $kT \ln 2$ of energy (where k is Boltzmann's constant and T is the temperature of the heat bath) must be (irreversibly) dissipated to heat per bit's worth of (known or correlated) information that is lost from the computational state. Landauer's argument can be rigorously derived from fundamental physical considerations.[896,897,898]

An important caveat is that Landauer's limit only applies to computational information that is correlated with other available information, as opposed to independent random information.[897,899] A computational bit that includes *no* correlations with other available bits is, in effect, *already* entropy, and thus it can be transferred back and forth between a stable, digital form in a computer and a rapidly-fluctuating physical form in a thermal environment with asymptotically zero net increase in total entropy, by, for example, adiabatically raising and lowering a potential energy barrier separating two degenerate states.

However, most bits in a digital computer are correlated bits, having been computed deterministically from other available bits. Performing a many-to-one transformation such as destructively overwriting or erasing such a bit *obliviously* (*i.e.*, without regards to its existing correlations) therefore increases total entropy by one bit's worth ($k \ln 2$) and thus implies at least $kT \ln 2$ energy consumption (decrease in free energy). Fundamentally, then, the only way to avoid Landauer's limit, in a deterministic computational process, is to avoid many-to-one transformations of the computational state. Bennett[900] showed that this is always possible; that is, any desired irreversible computation can always be embedded into an equivalent reversible one. Such an embedding generally appears to incur some algorithmic overheads,[901] but if reversible devices become cheaper and more energy-efficient over time, in principle, these resource overheads can be outweighed by the achievable energy savings. In the long run, reversible computing is the *only* physically possible path by which the amount of general digital computation that can be performed per unit energy and cost might continue to be increased indefinitely, without any known fundamental limit.

In typical adiabatic implementations of reversible computing in specific device technologies, there is frequently a linear device-level tradeoff between the energy dissipation and the time required to carry out a given primitive digital operation in the adiabatic limit. (*Energy dissipation*) × (*Time*) = *const*. However, there is no proof that this same tradeoff relation extends to all possible implementation technologies, and in fact, recent results suggest that an exponential downscaling of energy dissipation with delay may sometimes be possible when quantum effects are leveraged.[902] Further, even among cases where the relation applies, there are no known technology-independent lower bounds on the value of the energy-delay constant. Although there are indeed firm quantum lower limits on the product of energy *invested* times delay,[903] there are no known fundamental limits (greater than zero) on energy *dissipated* for a given delay value. Further, even given a fixed value of the constant, thermally limited parallel processors can still benefit from reversible computing in terms of their aggregate performance. For example, in cooling-limited stacked 3D logic scenarios, the per-area performance advantage of asymptotically adiabatic technologies increases with the square root of 3D processor thickness.[904] And in loosely-coupled, arbitrarily-massively parallelizable applications with fixed power budgets, the aggregate performance gain from reversible computing scales up with the energy efficiency, that is, arbitrarily.

Unfortunately, as of today, experimentally realizing reversible computing's promise to far exceed the system-level energy efficiency of all conventional computers remains an extremely difficult engineering challenge. Although a variety of different adiabatic[905,906,907,908,909,910,911,912,913,914,902] and ballistic[915,916,917,918,919,920,921,922,923,924,925] schemes for the realization of reversible logic have been proposed, it is challenging in practice to actually achieve large energy efficiency gains at the system level when accounting for a wide variety of real-world parasitic energy dissipation mechanisms that exist, which would all need to be systematically eliminated or reduced. There is not yet any "magic bullet" method that automatically addresses all of the many possible energy-loss mechanisms in a complete system at once. Logical reversibility (when suitably generalized[896]) is a *necessary* condition for approaching physical reversibility in deterministic digital computations, but by no means a *sufficient* one.

However, while approaching the ideal of physically reversible computing is by no means an *easy* path forward, it is at least *a* way to move forward. Potentially, even in CMOS, adiabatic circuits might be able to demonstrate useful energy efficiency gains for some applications even in the relatively near term if sufficiently high-*Q* resonators can be developed.[926] And in principle, even some of the existing reversible superconducting logic styles (such as RQFP[927,928,929,930,931,932] and nSQUID[933,934] logic) already appear to be capable of achieving energy dissipation below the Landauer limit in principle, although the available analyses of these logic styles do not yet account for energy dissipation associated with supplying the power-clock driving signals.

Reversible computing can also, potentially usefully, be combined with probabilistic computing (§5.3); if random digital bits are obtained by taking in entropy from the thermal environment and capturing it in a stable form, this can actually reduce environmental entropy (albeit without reducing total entropy, since the entropy of the digital state has then increased). Once a randomized reversible computation utilizing such bits of "true" entropy has completed, those random bits can later be returned to the thermal environment with no net thermodynamic cost.[897] Thus, the requirement for a nondeterministic computation to be thermodynamically reversible is somewhat looser than is the case for a deterministic computation; many-to-one (irreversible) transformations can be allowed together with one-to-many (nondeterministic) transformations in a computation,[935] as long as, overall over the course of the computation, previously-established correlations are not becoming lost.

Reversible computing is normally conceived of as a strategy for making *digital* computation more energy efficient. More generally, can a broad variety of *analog* computing schemes be developed that are also thermodynamically reversible? Record energy efficiencies for charge-based analog vector-matrix multiplication have been demonstrated using adiabatic principles as discussed in §5.2.1.1.2.[744] Further, fundamental physics is reversible at the microscale, which suggests that a sufficiently carefully-engineered analog computer might be made to approach macroscopic reversibility, and that its energy efficiency might be increased without limit as its technology is further refined. The degrees of freedom utilized for the analog physical computation would likely have to be very well-isolated from the system's thermal degrees of freedom, and the usual tendency for complex dynamical systems to devolve towards chaotic behavior would have to be suppressed in some way, or else made into a useful feature of the computational process. The previously mentioned work on chaotic logic (§5.2.3.2) suggests one potential technique

for harnessing the chaotic analog behavior of conservative dynamical systems usefully for computational purposes, but many other, more sophisticated methods may be possible.

5.5. Device-Architecture Interaction: Conclusions/Recommendations

In this section, we have surveyed a variety of concepts and R&D directions for the development of novel Beyond CMOS computing technologies that represent an effort to think "outside the box," in the sense of looking beyond just developing simple drop-in replacements for traditional logic and memory cells. More broadly, new hardware designs spanning multiple levels from the devices up through circuits and architectures must be considered, and the interactions between the various levels explored. More specifically, we expand the scope of future computing technologies beyond traditional irreversible, deterministic digital logic to include a broad range of alternative, unconventional computational paradigms, such as analog, probabilistic, and (classical) reversible computing paradigms. Quantum computing[707] is another paradigm that may become increasingly important for certain applications; it is being addressed in a new IRDS chapter (Cryogenic Electronics and Quantum Information Processing).

This section's focus on analog computing and, in particular, analog implementations of neural computational models that also facilitate efficiently performing complex computational functions such as matrix operations, reflects a set of presently popular research directions. Neural computing is an extremely "hot" area right now, due to the blossoming commercial success of neural computing techniques such as Deep Learning, Convolutional Neural Nets, and machine learning technologies. These techniques are widely used in analyzing and mining massive datasets in this present age of Big Data, and for developing "smart" IoT and mobile/wearable products.

Apart from analog/neural computing, the other alternative computing paradigms addressed in this section (probabilistic and reversible) may be covered more extensively in the future, to the extent that they prove necessary and useful for transcending the energy efficiency limits of deterministic, irreversible styles of computation (which includes many of the analog and neural styles). Of course, the broadest possible class of computational mechanisms consists of those that are not necessarily exclusively digital, deterministic, *or* irreversible (*i.e.*, the entirety of the Venn diagram in Figure BC5.1, including its exterior), and so certainly, the most efficient future machines will be found somewhere within that broadest class. However, at this time, the exact range of computational applications for which loosening each of these traditional constraints confers large benefits, as well as the exact extent of those benefits, remains very unclear, and will need to be proved out over time by ongoing R&D efforts.

Recommendations. In general, computing paradigms outside of the traditional irreversible, deterministic, digital paradigm are still very under-developed, compared to the conventional paradigm. This is not surprising, considering that the conventional paradigm historically facilitated the development of a design abstraction hierarchy that permitted enormously complex systems to be constructed. As a result, the complexity and efficiency of those systems increased exponentially as Moore's Law made the underlying devices cheaper and more efficient. However, Dennard scaling has now ended, the end of the CMOS roadmap appears to be in sight, with no clear successor having been identified, and fundamental thermodynamic limits are also coming into view. Thus, today there is an increasing level of interest in expanding the scope of our investigations to include unconventional computing paradigms that may transcend the limits of the traditional computing paradigm.

Analog computing offers many promising potential directions for technology development, some of which were outlined in §5.2, and it is likely that many other analog computing concepts also exist that are less well-developed at this time. Given the somewhat immature status today of analog computing technologies in general, it still seems too early in the game to "choose winners," in terms of specific analog technologies to focus exclusively on developing commercially. Thus, at this time, the ongoing exploration of an even broader range of alternatives at the research level may be called for. However, design hierarchy remains a useful concept even in the analog realm, so it seems likely that an extensive range of abstractions and reusable building blocks facilitating some (more or less) "general-purpose" style of analog computing, such as that envisioned in §5.2.4, which allows heterogeneously mixing analog computing primitives of various types, can be developed over time.

That being said, in the nearer term, it's quite possible that special-purpose analog accelerators for certain applications (such as the VMM and other matrix operations, and other neural-related applications addressed in §5.2.2) may appear commercially relatively soon, since that line of research has already progressed significantly towards applications, and the demand for these kinds of applications from the Big Data analytics and Machine Learning market sectors remains strong.

Beyond this, the other technologies addressed in this section (including the broad spectrum of analog computing technologies, as well as the probabilistic and reversible technologies) are mostly at a relatively low technology readiness level, and so major commercial applications for these are probably a bit further out. However, looking forwards, there should probably be an increased level of investment in terms of basic research and early-stage technology development in such areas. Neural computing is certainly extremely valuable right now, for a wide range of machine learning, data analytics, and pattern recognition applications, but it is still somewhat special purpose, compared to conventional computing, and so it is not necessarily directly highly beneficial for *all* computing applications of interest. In contrast, if, for example, reversible computing can be developed

successfully, including favorably engineered cost/performance tradeoffs,[936] it offers the potential to make *all* computing applications (including general-purpose digital computing) much more energy-efficient (and, ultimately, cost-efficient) than is possible in the conventional paradigm. Probabilistic logic may have its own broad range of applications as well. However, since these kinds of alternative technology directions are much less well-developed than the neural approach at present, they will require an increased level of long-horizon research investments in the near to medium term, if they are to make substantial progress towards realizing their promise by the time progress stalls in the presently more well-developed areas (conventional and neural computing).

Overall, the potential utility of new styles of "Beyond CMOS" computing that rethink computation — not just at the device level, but also in terms of the entire computing paradigm, with changes to the machine design also at the circuit level, the architecture level, and higher levels — is vast. It is our recommendation, based on the results of a 2017 workshop and our overall efforts in surveying this field, that these alternative computing styles deserve a greatly increasing level of attention and investment as the apparent end of the CMOS roadmap draws closer.

6. ASSESSMENT

6.1. INTRODUCTION

The purpose of this section is to assess beyond-CMOS memory and logic devices considered in this chapter benchmarked against current memory or CMOS technologies. Emerging architectures are not compared as each architecture often accelerates a different computational kernel. Assessments for individual architectures are given in the respective sections. Two methods of assessments were reported previously in the ITRS ERD chapter: a "quantitative emerging device benchmarking" conducted by research consortia, e.g., the Nanoelectronics Research Initiative (NRI), and a "survey-based assessment".

In the "quantitative emerging device benchmarking", each emerging device is evaluated by its operation in conventional Boolean Logic circuits, *e.g.*, a unity gain inverter, a 2-input NAND gate, and a 32-bit shift register. Metrics evaluated include speed, areal footprint, power dissipation, *etc*. Each parameter is compared with the performance projected for high performance and low power 5nm CMOS applications. This "beyond CMOS" chapter will update the quantitative benchmarking section with the latest NRI results.

Up to the 2013 ERD Chapter, a survey-based critical review was conducted based on eight criteria to compare emerging devices against their CMOS benchmark. Spider chart has been used to visualize the perceived potential of these technology entries. However, the limited number of survey results sometimes raises questions of the accuracy of this survey. The most recent "survey-based assessment" was conducted in the 2014 ERD Emerging Memory and Logic Device Assessment Workshops (Albuquerque, NM). The survey collects voting on emerging technologies evaluated in the workshops in the categories of the "most promising" and the "most need of resources" to assess the potential of these technology entries perceived by ERD experts. Although the survey-based assessment has not been conducted recently, a summary of previous assessments is included in this chapter for reference.

An important issue regarding emerging charge-based nanoelectronic switch elements is related to the fundamental limits to the scaling of these new devices, and how they compare with CMOS technology at its projected end of scaling. An analysis[937] concludes that the fundamental limit of scaling an electronic charge-based switch is only a factor of 3 times smaller than the physical gate length of a silicon MOSFET in 2024. Furthermore, the density of these switches is limited by a maximum allowable power dissipation of approximately 100W/cm^2, and not by their size. The conclusion of this work is that MOSFET technology scaled to its practical limit in terms of size and power density will asymptotically reach the theoretical limits of scaling for charge-based devices.

Most of the proposed beyond-CMOS devices are very different from their CMOS counterparts, and often pass computational state variables (or tokens) other than charge. Alternative state variables include collective or single spins, excitons, plasmons, photons, magnetic domains, qubits, and even material domains (*e.g.*, ferromagnetic). With the multiplicity of programs characterizing the physics of proposed new structures, it is necessary to find ways to benchmark the technologies effectively. This requires a combination of existing benchmarks used for CMOS and new benchmarks which take into account the idiosyncrasies of the new device behavior. Even more challenging is to extend this process to consider new circuits and architectures beyond the Boolean architecture used by CMOS today, which may enable these devices to complete transactions more effectively.

6.1.1. ARCHITECTURAL REQUIREMENTS FOR A COMPETITIVE LOGIC DEVICE

The circuit designer and architect depend on the logic switch to exhibit specific desired characteristics in order to insure successful realization of a wide range of applications. These characteristics,[938] which have since been supplemented in the literature, include:

- Inversion and flexibility (can form an infinite number of logic functions)
- Isolation (output does not affect input)
- Logic gain (output may drive more than one following gate and provides a high I_{on}/I_{off} ratio)
- Logical completeness (the device is capable of realizing any arbitrary logic function)
- Self restoring / stable (signal quality restored in each gate)
- Low cost manufacturability (acceptable process tolerance)
- Reliability (aging, wear-out, radiation immunity)
- Performance (transaction throughput improvement)
- Span of control (measures number of devices that may be reached within a characteristic delay of the switch[939])

Devices with intrinsic properties supporting the above features will be adopted more readily by the industry. Moreover, devices which enable architectures that address emerging concerns such as computational efficiency, complexity management, self-organized reliability and serviceability, and intrinsic cyber-security[940] are particularly valuable.

6.2. QUANTITATIVE EMERGING DEVICE BENCHMARKING

The Nanoelectronics Research Initiative (https://www.src.org/program/nri/) has been benchmarking several diverse beyond-CMOS technologies, trying to balance the need for quantitative metrics to assess a new device concept's potential with the need to allow device research to progress in new directions which might not lend themselves to existing metrics.[941, 942,943] Several of the more promising NRI devices have been described in detail in the Logic and Emerging Information Processing Device Section. Some results of this benchmark study were included in 2015 ITRS ERD chapter.

While all these efforts are still very much a work in progress – and no concrete decisions have been made on which devices should be chosen or eliminated as candidates for significantly extending or augmenting the roadmap as CMOS scaling slows – this section summarizes some of the data and insights gained from these studies. Further benchmarks may alter some of the conclusions here and the outlook on some of these devices, but the overall message on the challenge of finding a beyond CMOS device which can compete well across the full spectrum of benchmarks of interest remains.

6.2.1. QUANTITATIVE RESULTS

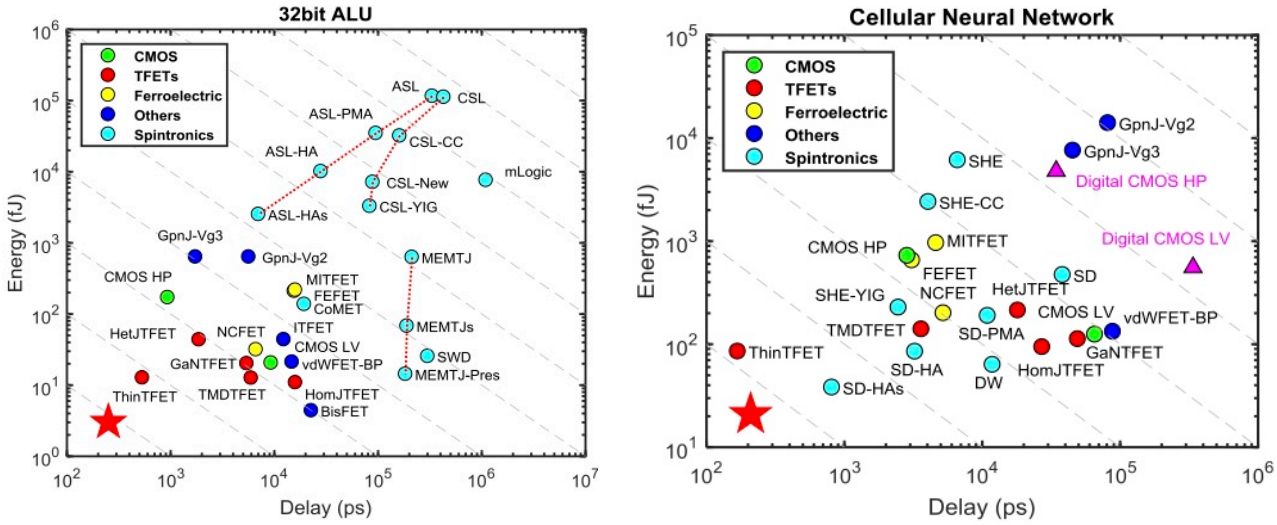
NRI benchmarking analyzes the potential of major emerging switches using a variety of information tokens and communication transport mechanisms. Specifically, the projected effectiveness of these devices used in a number of logic gate configurations was evaluated and normalized to CMOS at the 5nm generation (projection). The initial work has focused on "standard" Boolean logic architecture, since the CMOS equivalent is readily available for comparison. It should be noted that the majority of devices are evaluated via simulations since many of them have not yet been built, so it should be considered only a "snapshot in time" of the potential of any given device. Data on all of them are still evolving.

Device name	acronym	input(s)	control	int. state	output	material
Si MOSFET high perf.	CMOS HP	V	Vg	Q	V	silicon
Si MOSFET low voltage	CMOS LV	V	Vg	Q	V	InAs
van der Walls FET	vdWFET	V	Vg	Q	V	MoS_2
Homojunction III-V TFET	HomJTFET	V	Vg	R	V	InAs
Heterojunction III-V TFET	HetJTFET	V	Vg	R	V	GaSb/InAs
Graphene nanoribbon TFET	gnrFTET	V	Vg	R	V	graphene
Interlayer tunneling FET	ITFET	V	Vg	R	V	graphene
Two D Heterojunction Interlayer TFET	ThinFET	V	Vg	R	V	$WTe_2/SnSe_2$
GaN TFET	GaNFET	V	Vg	R	V	GaN
Transition Metal Dichalchogenide TFET	TMDTFET	V	Vg	R	V	WTe_2
Graphene pn-junction	GpnJ	V	Vg	R	V	graphene
Ferroelectric FET	FEFET	V	Vg	P	V	PZT
Negative capacitance FET	NCFET	V	Vg	P	V	PZT
Piezoelectric FET	PiezoFET	V	V	σ	V	AlN
Bilayer pseudospin FET	BisFET	V	Vg	BC	V	graphene
Excitonic FET	ExFET	V	Vg	BC	V	$MoS_2/MoSe_2$
Metal-insulator transistor	MITFET	V	Vg	Orb	V	$NdNiO_3$
SpinFET (Sughara-Tanaka)	SpinFET	V	Vg, Vm	Q, M	V	CoFeB
All-spin logic	ASL	M	V	M	M	CoPtCrB
Charge-spin logic	CSL	I	V	M	I	CoPtCrB
Spin torque domain wall	STT/DW	I	V	M	I	CoFeB
Spin majority gate	SMG	M	V	M	M	PMN-PT
Spin torque oscillator	STO	I	V	M	I	CoPtCrB
Spin wave device	SWD	M	I or V	M	M	PMN-PT
Nanomagnetic logic	NML	M	B or V	M	M	PMN-PT

Figure BC6.1 List of Devices Considered in NRI Benchmarking with Their Computational Variables and Classification[943]

At a high level, the data from these studies corroborates qualitative insights from earlier works, suggesting that many new logic switch structures may have some advantages over CMOS in terms of power or energy, but they are also inferior to CMOS in delay. This is perhaps not surprising; the primary goal for nanoelectronics and NRI is to find a lower power device[944] since power density is a primary concern for future CMOS scaling. The power-speed tradeoffs commonly observed in CMOS are also extended into the emerging devices. It is also important to understand the impact of transport delay for the different information tokens these devices employ. Communication with many non-charge tokens can be significantly slower than moving charge, although this may be balanced in some cases with lower energy for transport. The combination of the new balance between switch speed, switch area, and interconnect speed can lead to advantages in the span of control for a given technology. For some of the technologies (*e.g.*, nanomagnetic logic), there is no strong distinction between the switch and the interconnect, indicating the need for novel architecture to exploit unique attributes of these technologies. Figure BC6.1 lists the devices studied in the NRI benchmarking[943].

A simplified 32-bit arithmetic logic unit (ALU) was built from these devices to evaluate their performance and the result is summarized in Figure BC6.2(a) [945]. While tunneling devices (e.g., TFET) show limited advantages over CMOS in terms of energy-delay product, most beyond-CMOS devices are inferior to CMOS in energy and/or delay. For example, the majority of spintronic devices are slower than CMOS and also show no energy advantage.



(a) (b)

Figure BC6.2 (a) Energy versus Delay of a 32-bit ALU for a Variety of Charge- and Spin-based Devices; (b) Energy versus Delay per Memory Association Operation Using Cellular Neural Network (CNN) for a Variety of Charge- and Spin-based Devices[945]

At the architecture level, the ability to speculate on how these devices will perform is still in its infancy. While the ultimate goal is to compare at a very high level – *e.g.*, how many MIPS can be produced for 100 mW in 1 mm^2? – the current work must extrapolate from only very primitive gate structures. One initial attempt to start this process has been to look at the relative "logical effort"[946] for these technologies, a figure of merit that ties fundamental technology to a resulting logic transaction. Several of the devices appear to offer advantage over CMOS in logical effort, particular for more complex functions, which increases the urgency of doing more joint device-architecture co-design for these emerging technologies.

The direction of device-architecture co-optimization has driven NRI benchmarking to explore non-Boolean applications of beyond-CMOS devices. Cellular Neural Network (CNN) has been utilized as a benchmarking model that has been implemented with various novel devices.[945] The energy and delay of CNN based on beyond-CMOS devices are compared with CMOS-based CNN in Figure BC6.2(b). Tunneling devices have significant performance improvement because of their steep subthreshold slopes and large driving current at ultra-low supply voltage. Interestingly, spintronic devices are much closer to the preferred corner in CNN implementation in comparison with 32-bit ALU. This is because some characteristics of spintronic devices (e.g., spin diffusion, domain wall motion) may mimic the functionality of a neuron (e.g. integration) more naturally in a single device.

6.2.2. OBSERVATIONS

A number of common themes have emerged from these benchmark studies and in the observations made during recent studies of beyond-CMOS replacement switches[947]. A few noteworthy concepts:

1. The low voltage energy-delay tradeoff conundrum will continue to be a challenge for all devices. Getting to low voltage must remain a priority for achieving low power, but new approaches to getting throughput with 'slow' devices must be developed.
2. Most of the architectures that have been considered to date in the context of new devices utilize binary logic to implement von Neumann computing structures. In this area, CMOS implementations are difficult to supplant because they are very competitive across the spectrum of energy, delay and area – not surprising since these architectures have evolved over several decades to exploit the properties of CMOS most effectively. Novel electron-based devices – which can include devices that take advantage of collective and non-equilibrium effects – appear to be the best candidates as a drop-in replacement for CMOS for binary logic applications.
3. As the behavior of other emerging research devices becomes better understood, work on novel architectures that leverage these features will be increasingly important. A device that may not be competitive at doing a simple NAND function may have advantages in doing a complex adder or multiplier instead. Understanding the right building blocks for each device to maximize throughput of the system will be critical. This may be best accomplished by thinking about the high-level metric a system or core is designed to achieve (*e.g.*, computation, pattern recognition, FFT, *etc.*) and finding the best match between the device and circuit for maximizing this metric.
4. Increasing functional integration and on-chip switch count will continue to grow. To that end, in any logic architectural alternative, both flexible rich logic circuit libraries and reconfigurability will be required for new switch implementations.
5. Patterning, precision layer deposition, material purity, dopant placement, and alignment precision critical to CMOS will continue to be important in the realization of architectures using these new switches.
6. Assessment of novel architectures using new switches must also include the transport mechanism for the information tokens. Fundamental relationships connecting information generation with information communication spatially and temporally will dictate CMOS' successor.

Based on the current data and observations, it is clear that CMOS will remain the primary basis for IC chips for the coming years. While it is unlikely that any of the current emerging devices could entirely replace CMOS, several do seem to offer advantages, such as ultra-low power or nonvolatility, which could be utilized to augment CMOS or to enable better performance in specific application spaces. One potential area for entry is that of special purpose cores or accelerators that could off-load specific computations from the primary general purpose processor and provide overall improvement in system performance. If scaling slows in delivering the historically expected performance improvements in future generations, heterogeneous multi-core chips may be a more attractive option. These would include specific, custom-designed cores dedicated to accelerate high-value functions, such as accelerators already widely used today in CMOS (e.g. Encryption/Decryption, Compression/Decompression, Floating Point Units, Digital Signal Processors, etc.), as well as potentially new, higher-level functions (e.g. voice recognition). While integrating dissimilar technologies and materials is a big challenge, advances in packaging and 3D integration may make this more feasible over time, but the performance improvement would need to be large to balance this effort.

As a general rule, an accelerator is considered as an adjunct to the core processors if replacing its software implementation improves overall core processor throughput by approximately ten percent; an accelerator using a non-CMOS technology would likely need to offer an order of magnitude performance improvement relative to its CMOS implementation to be considered worthwhile. That is a high bar, but there may be instances where the unique characteristics of emerging devices, combined with a complementary architecture, could be used as an advantage in implementing a particular function. At the same time, the changing landscape of electronics (moving from uniform, general purpose computing devices to a spectrum of devices with varying purposes, power constraints, and environments spanning servers in data centers to smart phones to embedded sensors) and the changing landscape of workloads and processing needs (Big Data, unstructured information, real-time computing, 3D rich graphics) are increasing the need for new computing solutions. One of the primary goals then for future beyond-CMOS work should be to focus on specific emerging functions and optimize between the device and architecture to achieve solutions that can break through the current power/performance limits.

6.3. Survey-Based Benchmarking of Beyond CMOS Memory & Logic Technologies

Although survey-based emerging device assessment has not been continued after the 2015 ITRS ERD chapter, previous survey-based assessments in ERD chapters are summarized here for references.

6.3.1. Emerging Device Assessment in 2014 ERD Workshops

In August 2014, ERD organized an "Emerging Memory Device Assessment Workshop" and an "Emerging Logic Device Assessment Workshop", where nine memory devices and fourteen logic devices were evaluated. A survey was conducted in the workshops for the experts to vote on the "most promising" devices and devices "needing more resources". Figure BC6.3 shows

the relative number of votes received by emerging devices in these two categories, ranked from high to low in the "most promising" category (red color bars).

In the "most promising" memory device category, the vote clearly accumulated to a few well-known memory devices: STTRAM, ReRAM (including CBRAM and oxide-based ReRAM), and PCM, ranked from high to low. Some memory devices received few vote, due to lack of progress. Results in this category reflect consensus among experts based on R&D status of these devices. The "need more resources" category reflects perceived value of these devices in the view of the experts and also experts' consideration of R&D resource allocation based on existing investment (or lack of investment) for each device. For example, with heavy R&D investment on STTRAM that is considered most promising, it is not surprising that it ranks low in the need of resources. The strong interest in emerging FeFET memory is closely linked to the discovery of ferroelectricity in doped HfO_x. Among emerging logic devices, "carbon nanomaterial device" (mainly carbon nanotube FET), tunnel FET, and nanowire FET were ranked as one of the most promising emerging logic devices. Notice that they are all charge-based devices, but involve novel materials, structures, and mechanisms. "Piezotronic transistors", "negative-capacitance FET", and "2D channel FET" were considered top choices for enhanced research investment.

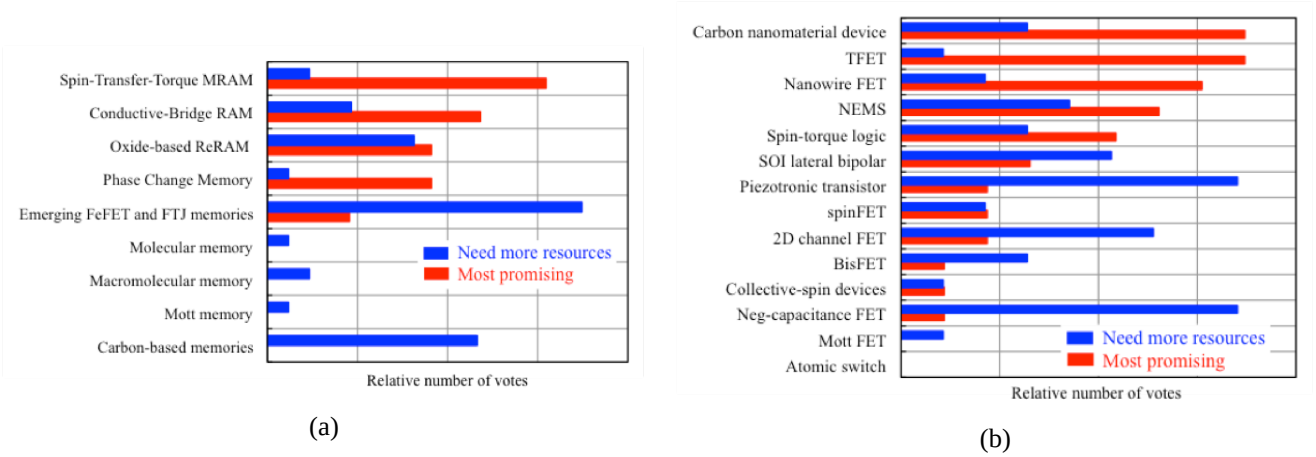


Figure BC6.3 (a) Survey of Emerging Memory Devices and (b) Survey of Emerging Logic Devices in 2014 ERD Emerging Logic Workshop (Albuquerque, NM)

6.3.2. ERD Survey Criteria, Methodology, and Results

In the traditional survey-based assessment conducted by ERD, a set of relevance or evaluation criteria, defined below, are used to parameterize the extent to which "CMOS Extension" and "Beyond CMOS" technologies are applicable to memory or information processing applications. The relevance criteria are: 1) Scalability, 2) Speed, 3) Energy Efficiency, 4) Gain (Logic) or ON/OFF Ratio (Memory), 5) Operational Reliability, 6) Operational Temperature, 7) CMOS Technological Compatibility, and 8) CMOS Architectural Compatibility. Description of each criterion can be found in 2013 ERD chapter.

Each CMOS extension and beyond-CMOS emerging memory and logic device technology is evaluated against these criteria according to a single factor. For logic, this factor relates to the *projected potential performance* of a nanoscale device technology, assuming its successful development to maturity, *compared to that for silicon CMOS scaled to the end of the Roadmap*. For memory, this factor relates the *projected potential performance* of each nanoscale memory device technology, assuming its successful development to maturity, *compared to that for ultimately scaled current silicon memory technology which the new memory would displace*. Performance potential for each criterion is assigned a value from 1–3, with "3" substantially exceeding ultimately-scaled CMOS, and "1" substantially inferior to CMOS or, again, a comparable existing memory technology. This evaluation is determined by a survey of the ERD Working Group members composed of individuals representing a broad range of technical backgrounds and expertise. Details of the assessment values are also included in 2013 ERD chapter.

Although this survey-based critical review has been conducted in ERD for several versions and has been widely cited in literatures, the decreasing number of votes of some less popular devices has raised concerns about the accuracy of some of the results. Figures BC6.4 and EBC6.5 summarize the last critical review conducted in 2013 for emerging memory devices and emerging logic devices, respectively. Notice that the technology entries in these figures are based on the 2013 ERD chapter, while some of them have been removed in this chapter (*e.g.*, molecular memory, atomic switch, *etc.*) and several new technologies are added in this chapter (*e.g.*, novel magnetic memory, transistor laser, *etc.*)

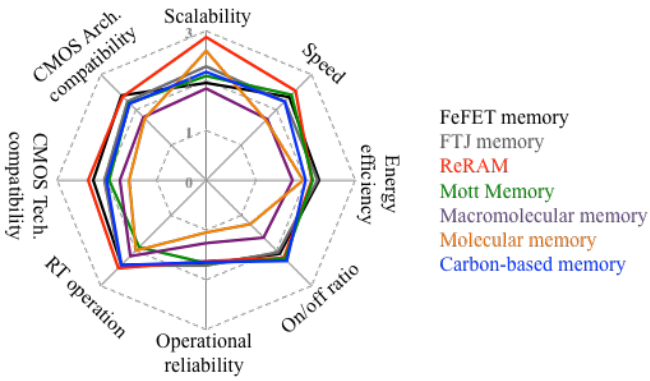
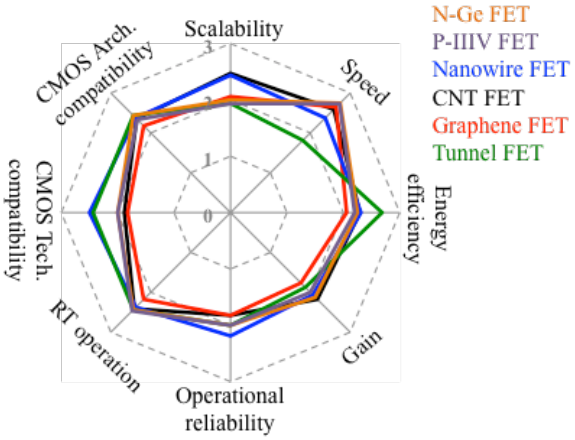
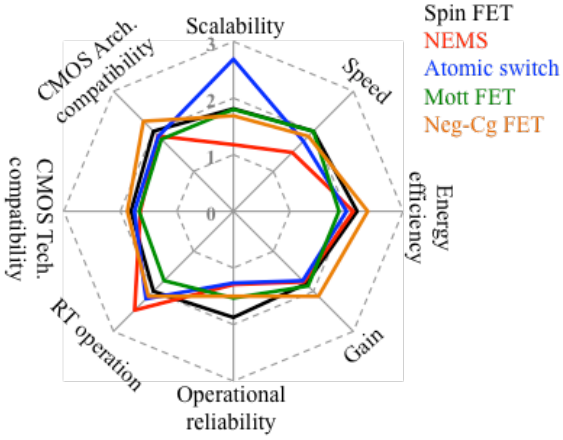


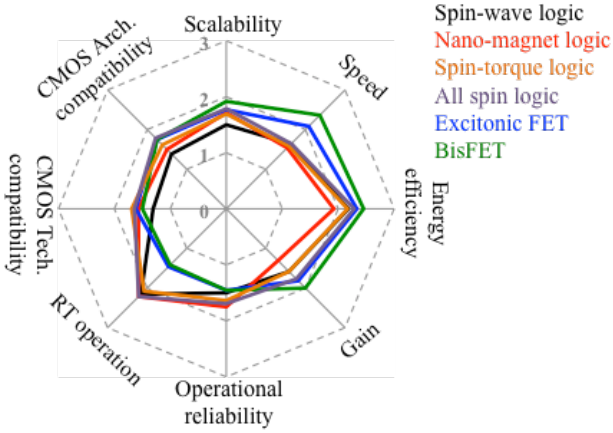
Figure BC6.4 Comparison of Emerging Memory Devices Based on 2013 Critical Review



(a) CMOS extension devices



(b) Charge-based beyond-CMOS devices



(c) Non-charge-based beyond-CMOS devices

Figure BC6.5 Comparison of Emerging Logic Devices Based on 2013 ITRS ERD Critical Review: (a) CMOS Extension Devices; (b) Charge-based Beyond-CMOS Devices; (c) Non-charge-based Beyond-CMOS Devices

Since "3" represents the best result and "1" the worst in the spider chart, devices with larger circle area represent more promising devices. In Figure BC6.5 for emerging logic devices, the perceived potential of "beyond-CMOS devices" is generally poorer than "CMOS-extension devices". Within "beyond-CMOS devices", "non-charge-based devices" area also perceived slightly less promising than "charge-based devices". The general trend is consistent with the quantitative assessment in section 6.2. Multiple factors contribute to this result, including the strength of CMOS as a platform technology, the challenges of beyond-CMOS

devices in materials, fabrication, and even mechanisms, the lack of memory and interconnect solutions for beyond-CMOS devices, *etc*.

7. SUMMARY

The "beyond-CMOS" chapter systematically surveys emerging memory and logic devices (sections 2 and 3), novel technologies (section 4), and alternative architectures and computing paradigms (section 5), to explore potential solutions beyond the conventional scaling of CMOS technologies. Although high performance at low power consumption has been a primary objective of beyond-CMOS devices, novel functionalities and applications have become increasingly important. The recent emergence of energy-efficient data-intensive cognitive applications is also shifting the emphasis from high-precision computing solutions to novel computing paradigms with massive parallelism and bio-inspired mechanisms. Research opportunities exist in the co-optimization of beyond-CMOS devices and architectures to explore unique device characteristics and architectural designs.

Although a beyond-CMOS device competitive against CMOS FET has not been identified, beyond-CMOS devices with dramatically enhanced scalability and performance while simultaneously reducing the energy dissipation per functional operation would still be fundamentally important and a worthwhile research objective. In considering the many disparate new approaches proposed to provide order of magnitude scaling of information processing beyond that attainable with ultimately scaled CMOS, the following set of guiding principles are proposed to provide a useful structure for directing research on "Beyond CMOS" information processing technology.

- ***COMPUTATIONAL STATE VARIABLE(S) OTHER THAN SOLELY ELECTRON CHARGE***

These include spin, phase, multipole orientation, mechanical position, polarity, orbital symmetry, magnetic flux quanta, molecular configuration, and other quantum states. The estimated performance comparison of alternative state variable devices to ultimately scaled CMOS should be made as early in a program as possible to down-select and identify key trade-offs.

- ***NON-THERMAL EQUILIBRIUM SYSTEMS***

These are systems that are out of equilibrium with the ambient thermal environment for some period of their operation, thereby reducing the perturbations of stored information energy in the system caused by thermal interactions with the environment. The purpose is to allow lower energy computational processing while maintaining information integrity.

- ***NOVEL ENERGY TRANSFER INTERACTIONS***

These interactions would provide the interconnect function between communicating information processing elements. Energy transfer mechanisms for device interconnection could be based on short range interactions, including, for example, quantum exchange and double exchange interactions, electron hopping, Förster coupling (dipole–dipole coupling), tunneling and coherent phonons.

- ***NANOSCALE THERMAL MANAGEMENT***

This could be accomplished by manipulating lattice phonons for constructive energy transport and heat removal.

- ***SUB-LITHOGRAPHIC MANUFACTURING PROCESS***

One example of this principle is directed self-assembly of complex structures composed of nanoscale building blocks. These self-assembly approaches should address non-regular, hierarchically organized structures, be tied to specific device ideas, and be consistent with high volume manufacturing processes.

- ***ALTERNATIVE ARCHITECTURES***

In this case, architecture is the functional arrangement on a single chip of interconnected devices that includes embedded computational components. These architectures could utilize, for special purposes, novel devices other than CMOS to perform unique functions.

8. ENDNOTES/REFERENCES

1 V. V. Zhirnov, R. K. Cavin, S. Menzel, E. Linn, S. Schmelzer, D. Bräuhaus, C. Schindler and R. Waser, “Memory Devices: Energy-Space-Time Trade-offs,” *Proc. IEEE*, vol. 98, pp. 2185-2200, Dec. 2010.

2 Ch. Binek, B. Doudin, “Magnetoelectronics with magnetoelectrics,” *J. Phys. Condens. Matter* 17, L39-L44 (2005).

3 M. Bibes, A. Barthélémy, “Multiferroics: Towards a magnetoelectric memory,” *Nature Mat.* 7, 425-426 (2008).

4 X. He, Y. Wang, N. Wu, A. N. Caruso, E. Vescovo, K. D. Belashchenko, P. A. Dowben, Ch. Binek, “Robust isothermal electric control of exchange bias at room temperature,” *Nature Materials* 9, 579 (2010).

5 N. Wu, Xi He, A. Wysocki, U. Lanke, T. Komesu, K.D. Belashchenko, Ch. Binek, P.A. Dowben, “Imaging and Control of Surface Magnetization Domains in a Magnetoelectric Antiferromagnet,” *Phys. Rev. Lett.* 106, 087202 (2011).

6 P. A. Dowben, C. Binek, and D. E. Nikonov, “Potential of Nonvolatile Magnetoelectric Devices for Spintronic Applications,” *Nanoscale Silicon Devices*; edited by Shuni Oda and David Ferry; Taylor and Francis (London) 2016, Chapter 11, pp 255–278.

7 T. Kosub, M. Kopte, F. Radu, O. G. Schmidt, and D. Makarov, ““All-Electric Access to the Magnetic-Field-Invariant Magnetization of Antiferromagnets,” *Phys. Rev. Lett.* 115, 097201 (2015)

8 T. Kosub et al. “Purely antiferromagnetic magnetoelectric random access memory,” *Nat. Commun.* 8, 13985 (2017)

9 R. Choudhary, P. Kumar, P. Manchanda, D. J. Sellmyer, P. A. Dowben, A. Kashyap, *et al.*, “Interface-Induced Spin Polarization in Graphene on Chromia,”*IEEE Magnetics Lett.* 7, 3101604 (2016).

10 J. A. Kelber, C. Binek, P. A. Dowben, and K. Belashchenko, "Magneto-electric voltage controlled spin transistors," U.S. Patent 9,718,700, August 1, 2017.

11 Y. G. Semenov, K. W. Kim, and J. M. Zavada, “Spin field effect transistor with a graphene channel,” *Appl. Phys. Lett.* 91, 153105 (2007).

12 S. C. Stuart, B. Gray, D. Nevola, L. Su, E. Sachet, M. Ulrich, *et al.*, *Physica Status Solidi-Rapid Res. Lett.* 10, 242-247 (2016).

13 P. A. Dowben, C. Binek, K. Zhang, L. Wang, W.-N. Mei, J. P. Bird, U. Singisetti, X. Hong, K. L. Wang, D. Nikonov, "Exploiting strong spin-orbit coupling in a magneto-electric transistor,” *IEEE J. Exploratory Solid-State Computational Devices and Circuits* (2018), in press.

14 D. E. Nikonov, C. Binek, X. Hong, J. P. Bird, L. Wang, P. A. Dowben, “Anti-Ferromagnetic Magneto-electric Spin-Orbit Read Logic”. U.S. Patent Application No.: 62/460,164; filed February 17, 2017

15 R. Waser, R. Dittman, G. Staikov, and K. Szot, “Redox-based resistive switching memories – nanoionic mechanisms, prospects, and challenges,” *Adv. Mat.* 21 (2009) 2632-2663

16 H. Akinaga and H. Shima, “Resistive random access memory (ReRAM) based on metal oxides,” *Proc. IEEE* vol. 98, pp. 2237–2251, 2010.

17 R. Waser, R. Bruchhaus, and S. Menzel, "Redox-based resistive switching memories," *Nanoelectronics and Information Technology*, R. Waser, Ed., ed Weinheim, Germany: Wiley-VCH, 2013.

18 M. J. Marinella, "Emerging resistive switching memory technologies: Overview and current status," *2014 IEEE Intl Symp on Circuits and Systems (ISCAS)*, Melbourne VIC, 2014, pp. 830–833.

19 B. Govoreanu, *et al.*, "10x10nm^2 Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation," *IEDM Tech. Dig.*, 2011, pp. 31.6.1–31.6.4.

20 Z. Wei, *et al.*, "Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism," *IEDM*, 2008, pp. 1–4.

21 M-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y-B. Kim, C-J. Kim, D. H. Seo, S. Seo, U-I. Chung, I-K. Yoo, K. Kim, “A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta2O5-x/TaO2-x bilayer structures,” *Nature Mat.* July 2011.

22 R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-Based Resistive Switching Memories - Nanoionic Mechanisms, Prospects, and Challenges," *Advanced Materials*, vol. 21, pp. 2632–2663, 2009.

23 J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive devices for computing," *Nature Nanotech.*, vol. 8, pp. 13–24, 2013.

24 F. Pan, S. Gao, C. Chen, C. Song, and F. Zeng, "Recent progress in resistive random access memories: Materials, switching mechanisms, and performance," *Materials Science and Engineering: R: Reports*, vol. 83, pp. 1–59, Sept. 2014.

25 J. J. Yang, F. Miao, M. D. Pickett, D. A. A. Ohlberg, D. R. Stewart, C. N. Lau, *et al.*, "The mechanism of electroforming of metal oxide memristive switches," *Nanotechnology*, vol. 20, p. 215201, May 2009.

26 J. J. Yang, M. D. Pickett, X. Li, D. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nature nanotechnology*, vol. 3, pp. 429–433, 2008.

27 D. S. Jeong, H. Schroeder, U. Breuer, and R. Waser, "Characteristic electroforming behavior in Pt/TiO2/Pt resistive switching cells depending on atmosphere " *Journal of Applied Physics*, vol. 104, p. 123716, 2008.

28 B. Govoreanu, G. S. Kar, Y. Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, *et al.*, "10x10nm^{2}Hf/HfOxcrossbar resistive RAM with excellent performance, reliability and low-energy operation," *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, pp. 31.6.1–31.6.4.

[29] L. Kai-Shin, C. Ho, L. Ming-Taou, C. Min-Cheng, H. Cho-Lun, J. M. Lu, *et al.*, "Utilizing Sub-5 nm sidewall electrode technology for atomic-scale resistive memory fabrication," *VLSI Technology (VLSI-Technology): Digest of Technical Papers, 2014 Symposium on*, 2014, pp. 1–2.

[30] Chang, K.C., Zhang, R., Chang, T.C., Tsai, T.M., Chu, T.J., Chen, H.L., Shih, C.C., Pan, C.H., Su, Y.T., Wu, P.J. and Sze, S.M., , December.,"High performance, excellent reliability multifunctional graphene oxide doped memristor achieved by self-protect ive compliance current structure," in *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014,(pp. 33-3), IEEE.

[31] Lee, M.J., Lee, C.B., Lee, D., Lee, S.R., Chang, M., Hur, J.H., Kim, Y.B., Kim, C.J., Seo, D.H., Seo, S. and Chung, U.I.,"A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta 2 O 5− x/TaO 2− x bilayer structures," in *Nature Materials*, *10*(8), 2011, p.625.

[32] C. T. Antonio, S. John Paul, M.-R. Gilberto, and R. S. Williams, "Sub-nanosecond switching of a tantalum oxide memristor," *Nanotechnology*, vol. 22, p. 485203, 2011.

[33] B. J. Choi, A. C. Torrezan, J. P. Strachan, P. G. Kotula, A. J. Lohn, M. J. Marinella, R. S. Williams and J. Joshua Yang, "High-speed and low-energy nitride memristors" , *Advanced Functional Materials* 26, 5290 (2016)

[34] S. Choi, J. Lee, S. Kim, and W. D. Lu, "Retention failure analysis of metal-oxide based resistive memory," *Applied Physics Letters*, vol. 105, p. 113510, 2014.

[35] M. Wang, S. Cai, C. Pan, C. Wang, X. Lian, K. Xu, Y. Zhuo, J. Joshua Yang, P. Wang and F. Miao, "Robust memristors based on fully layered two-dimensional materials" , *Nature Electronics 1*, (2018). doi:10.1038/s41928-018-0021-4

[36] S. Lee, J. Sohn, Z. Jiang, H.-Y. Chen, and H. S. Philip Wong, "Metal oxide-resistive memory using graphene-edge electrodes," *Nat Commun*, vol. 6, Sept. 2015.

[37] Kang, Chen-Fang, et al. "Self-formed conductive nanofilaments in (Bi, Mn) Ox for ultralow-power memory devices." *Nano Energy* 13 (2015): 283–290.

[38] T. Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, *et al.*, "A 130.7 mm^2 2-layer 32Gb ReRAM memory device in 24 nm technology," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 210–211.

[39] J. Sung Hyun, T. Kumar, S. Narayanan, W. D. Lu, and H. Nazarian, "3D-stackable crossbar resistive memory based on Field Assisted Superlinear Threshold (FAST) selector," *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014, pp. 6.7.1-6.7.4.

[40] A. Sawa et al., , "Hysteretic current–voltage characteristics and resistance switching at a rectifying Ti/Pr0.7Ca0.3MnO3 interface," *APL*, 85, p. 4073, 2004.

[41] Hyejung Choi et al., "The Effect of Tunnel Barrier at Resistive Switching Device for Low Power Memory Applications," *IEEE International Memory Workshop*, 2011.

[42] R. Meyer et al., "Oxide Dual-Layer Memory Element for Scalable Non-Volatile Cross-Point Memory Technology," *IEEE NVMTS*, 2008.

[43] C.J. Chevallier et al., "A 0.13 µm 64 Mb multi-layered conductive metal-oxide memory," *IEEE Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2010.

[44] H. Schroeder, V. V. Zhirnov, R. K. Cavin, and R. Waser, "Voltage-time dilemma of pure electronic mechanisms in resistive switching memory cells," *Journal of Applied Physics*, vol. 107, p. 54517, Mar. 2010.

[45] X. Huang, H. Wu, B. Gao, D.C. Sekar, L. Dai, M. Kellam, G. Bronner, N. Deng, and H. Qian, "HfO2/Al2O3 multilayer for RRAM arrays: a technique to improve tail-bit retention," *Nanotech.* **27**, 395201 (2016).

[46] I. G. Baek, *et al.*, "Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses," *2004 IEDM Tech. Dig.*, 2004, pp. 587–590.

[47] L. Goux, *et al.*, "Field-driven ultrafast sub-ns programming in WAl2O3TiCuTe-based 1T1R CBRAM system," *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 69–70.

[48] C. Cagli, *et al.*, "Experimental and theoretical study of electrode effects in HfO2 based RRAM," *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, pp. 28.7.1–28.7.4.

[49] C. Cagli, D. Ielmini, F. Nardi, and A. L. Lacaita, "Evidence for threshold switching in the set process of NiO-based RRAM and physical modeling for set, reset, retention and disturb prediction," *IEDM Tech. Dig.*, 2008, pp. 1–4.

[50] U. Russo, *et al.*, "Conductive-filament switching analysis and self-accelerated thermal dissolution model for reset in NiO-based RRAM," *IEDM Tech. Dig.*, 2007, pp. 775–778.

[51] L. Goux, *et al.*, "Coexistence of the bipolar and unipolar resistive-switching modes in NiO cells made by thermal oxidation of Ni layers," *Journal of Applied Physics*, vol. 107, p. 024512, Jan. 2010.

[52] D. S. Jeong, H. Schroeder, and R. Waser, "Coexistence of Bipolar and Unipolar Resistive Switching Behaviors in a Pt⁄TiO2⁄Pt Stack," *Electrochemical and Solid-State Letters*, vol. 10, pp. G51-G53, Aug. 2007.

[53] L. Goux, *et al.*, "Roles and Effects of TiN and Pt Electrodes in Resistive-Switching HfO2 Systems," *Electrochemical and Solid-State Letters*, vol. 14, pp. H244-H246, June 2011.

[54] YY. Chen *et al*. "Insights into Ni-filament formation in unipolar-switching Ni/HfO2/TiN resistive random access memory device," *Appl. Phys. Lett.* 100, 113513 (2012)

[55] T. Yanagida, *et al.*, "Scaling Effect on Unipolar and Bipolar Resistive Switching of Metal Oxides," *Sci. Rep.*, vol. 3, p. 1657, Apr. 2013.

[56] X. A. Tran, *et al.*, "High performance unipolar AlO_yHfO_xNi based RRAM compatible with Si diodes for 3D application," *VLSI Technology (VLSIT), 2011 Symposium on*, 2011, pp. 44–45.

[57] Y.-H. Tseng, *et al.*, "High density and ultra small cell size of Contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits," *IEDM Tech. Dig.*, 2009, pp. 1–4.

[58] Y.-H. Tseng, *et al.*, "Electron trapping effect on the switching behavior of contact RRAM devices through random telegraph noise analysis," *IEDM Tech. Dig.*, 2010, pp. 28.5.1–28.5.4.

[59] W. C. Shen, *et al.*, "High-κ metal gate contact RRAM (CRRAM) in pure 28 nm CMOS logic process," *IEDM Tech. Dig.*, 2012, pp. 31.6.1–31.6.4.

[60] M.-F. Chang, *et al.*, "A 0.5V 4Mb logic-process compatible embedded resistive RAM (ReRAM) in 65 nm CMOS using low-voltage current-mode sensing scheme with 45 ns random read time," *ISSCC Tech. Dig.*, 2012, pp. 434–436.

[61] W.-C. Chien, *et al.*, "A novel high performance WOx ReRAM based on thermally-induced SET operation," *VLSI Technology*, 2013, pp. T100–T101.

[62] Kozicki, M. N., et al. "Information storage using nanoscale electrodeposition of metal in solid electrolytes." *Superlattices and microstructures*, vol. 34.3, pp. 459-465 2003.

[63] Valov, Ilia, et al. "Electrochemical metallization memories—fundamentals, applications, prospects." *Nanotechnology* 22.25 (2011): 254003.

[64] Otsuka, Wataru, et al. "A 4mb conductive-bridge resistive memory with 2.3 gb/s read-throughput and 216mb/s program-throughput." *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2011 IEEE

[65] Gilbert, Nad, et al. "A 0.6 V 8 pJ/write Non-Volatile CBRAM Macro Embedded in a Body Sensor Node for Ultra Low Energy Applications." *VLSI Circuits (VLSIC), 2013 Symposium on. IEEE*, 2013.

[66] K. Tsutsui, "ReRAM for Fast Storage Application," Presented at August 2012 Flash Memory Summit Santa Clara, CA available at http://www.flashmemorysummit.com/English/Collaterals/Proceedings/2012/20120822_S203C_Tsutusui.pdf

[67] Altis Semiconductor Press release on embedded CBRAM available at http://www.altissemiconductor.com/en/index.php/about-altis/media-center/press-releases-menu/174-altis-ecbram

[68] Gopalan, C., et al. "Demonstration of conductive bridging random access memory (CBRAM) in logic CMOS process." *Solid-State Electronics* 58.1 (2011): 54-61.

[69] Adesto Technologies, http://www.adestotech.com/cbram

[70] R. Fackenthal *et al.*, "19.7 A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology," *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, San Francisco, CA, 2014, pp. 338–339.

[71] J. Zahurak *et al.*, "Process integration of a 27 nm, 16Gb Cu ReRAM," *2014 IEEE International Electron Devices Meeting*, San Francisco, CA, 2014, pp. 6.2.1–6.2.4.

[72] Prall, Kirk, et al. "An Update on Emerging Memory: Progress to 2Xnm." Memory Workshop (IMW), 2012 4th IEEE International. IEEE, 2012.

[73] Sankaran, Kiroubanand, et al. "Modeling of Copper Diffusion in Amorphous Aluminum Oxide in CBRAM Memory Stack." ECS Transactions 45.3 (2012): 317-330.

[74] Miyamura, Makoto, et al. "Programmable cell array using rewritable solid-electrolyte switch integrated in 90 nm CMOS." *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International.* IEEE, 2011.

[75] Suri, M., et al. "CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (Cochlea) and visual (Retina) cognitive processing applications." *Electron Devices Meeting (IEDM), 2012 IEEE International.* IEEE, 2012.

[76] Choi, S., et al. "Resistance drift model for conductive-bridge (CB) RAM by filament surface relaxation." *Memory Workshop (IMW), 2012 4th IEEE International.* IEEE, 2012.

[77] A Kanwal, S. Paul, M. Chhowalla, "Organic Memory Devices Using C60 and Insulator Polymer," *MRS Proceedings on Materials and Processes for Nonvolatile Memories*, 2005, 830, D7.2.1.

[78] J.Y Ouyang, C.W Chu CW, C.R Szmanda, L.P Ma, and Y Yang, "Programmable polymer thin film and non-volatile memory device," *Nature Materials*, 3, 918 (2004)

[79] S.-H. Liu ; W.-L. Yang ; C.-C. Wu ; T.-S. Chao ; M.-R. Ye ; Y.-Y. Su ; P.-Y. Wang ; M.-J. Tsai, "High-Performance Polyimide-Based ReRAM for Nonvolatile Memory Application " *IEEE Electron Dev. Lett.* 34, 123 – 125, 2013.

[80] W. Bai, R. Huang ; Y. Cai ; Y. Tang ; X. Zhang ; Y. Wang "Record Low-Power Organic RRAM With Sub-20-nA Reset Current," *IEEE Eelectron. Dev. Lett.* 34, 223-225, 2013.

[81] Y.-C. Lai, D.Y. Wang, I-S. Huang, Y.T. Chen, Y.-H. Hsu, T.-Y. Lin, H.-F. Meng, T.C. Chang, Y.J. Yang, C.C. Chen, F.-C. Hsu, Y.-F. Chen J. , "Low operation voltage macromolecular composite memory assisted by graphene nanoflakes," *Mater. Chem.* C 1,552-559, 2013.

[82] J.J. Kim, B. Cho; K.S. Kim; T. Lee, G.Y. Jung, "Electrical Characterization of Unipolar Organic Resistive Memory Devices Scaled Down by a Direct Metal-Transfer Method," *Adv. Mater.* 23, 2104-2107, 2011.

83 S. Song,;,J. Jang, Y. Ji, S. Park, T.W. Kim, Y. Song, M.H. Yoon, H.C. Ko, G.Y. Jung, T. Lee, "Twistable nonvolatile organic resistive memory devices," *Org. Electron.* 14, 2087-2092, 2013.

84 Y. Chai, Y. Wu, K. Takei, H.Y. Chen, S.M. Yu, P.C.H. Chan, A. Javey, H.S.P. Wong, "Nanoscale Bipolar and Complementary Resistive Switching Memory Based on Amorphous Carbon," *IEEE Trans. Electron. Dev*. 2011, 58, 3933-3939.

85 C.-L. Tsai, F. Xiong, E. Pop, S. Moonsub, "Resistive Random Access Memory Enabled by Carbon Nanotube Crossbar Electrodes," *ACS Nano* 7, 5360-5366, 2013.

86 P. Siebeneicher, H. Kleemann, K. Leo, and B. Lüssem, "Non-volatile organic memory devices comprising SiO_2 and C_{60} showing 10^4 switching cycles," *Appl. Phys. Lett*. 100, 193301 2012.

87 Q. Chen, B. F. Bory, A. Kiazadeh, P. R. F. Rocha, H. L. Gomes, F. Verbakel, D. M. De Leeuw, S. C. J. Meskers, "Opto-electronic characterization of electron traps upon forming polymer oxide memory diodes," *Appl. Phys. Lett.* 99, 083305, 2011.

88 Lin, W.-P., Liu, S.-J., Gong, T., Zhao, Q. and Huang, W., "Polymer-Based Resistive Memory Materials and Devices," *Adv. Mater.*, 26: 570–606, 2014.

89 I. Salaoru, S. Alotaibi (a1), Z. Al Halafi (a1) and S Paul, "Creating Electrical Bistability Using Nano-bits – Application in 2-Terminal Memory Devices," *MRS Advances*, 2, 195-208, 2017.

90 Paul, S., "Realization of nonvolatile memory devices using small organic molecules and polymer," *IEEE Transactions on Nanotechnology*, vol. 6, no. 2, pp. 191-195, 2007.

91 K. Asadi, D.M. de Leeuw, B. de Boer, B. P.W.M. Blom, "Organic non-volatile memories from ferroelectric phase-separated blends," *Nat. Mater*. 7, 547-550, 2008.

92 N. Tsutsumi, X. Bai, and W. Sakai, "Towards nonvolatile memory devices based on ferroelectric polymers," *AIP Advances* 2, 012154, 2012.

93 S.L. Miller and P.J. McWhorter, "Physics of the ferroelectric nonvolatile memory field effect transistor," *J. Appl. Phys*. 72, 5999 1992.

94 T. Oikawa, H. Morioka, A. Nagai, H. Funakubo, and K. Saito, "Thickness scaling of polycrystalline Pb(Zr,Ti)O3 films down to 35nm prepared by metalorganic chemical vapor deposition having good ferroelectric properties," *Appl. Phys. Lett*. 85, 1754 (2004).

95 J. Celinska, V. Joshi, S. Narayan, L. McMillan, and Paz de Araujo, C., "Effects of scaling the film thickness on the ferroelectric properties of SrBi2Ta2O9 ultra thin films," *Appl. Phys. Lett.* 82, 3937 (2003).

96 J. Müller, P. Polakowski, S. Mueller, and T. Mikolajick, "Ferroelectric hafnium oxide based materials and devices: Assessment of current status and future prospects," *ECS J. Solid State Sci. Technol. (ECS Journal of Solid State Science and Technology)* 4, N30-N35 (2015).

97 T.S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors," *IEEE International Electron Devices Meeting (IEDM)*, Washington D.C., USA, 5-7 December (2011.), pp. 547–550.

98 J. Müller, E. Yurchuk, T. Schlosser, J. Paul, R. Hoffmann, S. Müller, D. Martin, S. Slesazeck, P. Polakowski, J. Sundqvist, M. Czernohorsky, K. Seidel, P. Kucher, R. Boschke, M. Trentzsch, K. Gebauer, U. Schroder, and T. Mikolajick, "Ferroelectricity in HfO2 enables nonvolatile data storage in 28 nm HKMG, in Symposium on VLSI Technology (VLSIT)," (2012.), pp. 25–26.

99 T.P. Ma and J.-P. Han, "Why is nonvolatile ferroelectric memory field-effect transistor still elusive?" *IEEE Electron Device Lett*. 23, 386 (2002).

100 S. Sakai, R. Ilangovan, and M. Takahashi, "Pt/SrBi2Ta2O9/Hf-Al-O/Si field-effect-transistor with long retention using unsaturated ferroelectric polarization switching," *Jap. J. Apl. Phys*. 43, 7876 (2004).

101 J. Müller, S. Müller, P. Polakowski, and T. Mikolajick, "Ferroelectric hafnium oxide: a game changer to FRAM?," *14th Non-Volatile Memory Technology Symposium (NVMTS)*, Jeju, South Korea (2014.).

102 C.-H. Cheng and A. Chin, "Low-Leakage-Current DRAM-Like Memory Using a One-Transistor Ferroelectric MOSFET with a Hf-Based Gate Dielectric," *IEEE Electron Device Lett.* 35, 138 (2014).

103 J. Müller, T.S. Böscke, S. Müller, E. Yurchuk, P. Polakowski, J. Paul, D. Martin, T. Schenk, K. Khullar, A. Kersch, W. Weinreich, S. Riedel, K. Seidel, A. Kumar, T.M. Arruda, S.V. Kalinin, T. Schlösser, R. Boschke, R. van Bentum, U. Schröder, and T. Mikolajick, "Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories," *IEEE International Electron Devices Meeting (IEDM)*, Washington (2013.), pp. 10.8.1 - 10.8.4.

104 W. Zhang, M. Takahashi, and S. Sakai, "Electrical properties of CaxSr1− xBi2Ta2O9 ferroelectric-gate field-effect transistors," *Semicond. Sci. Technol*. 28, 85003 (2013).

105 X. Zhang, M. Takahashi, K. Takeuchi, and S. Sakai, "64 kbit ferroelectric-gate-transistor-integrated NAND flash memory with 7.5 V program and long data retention," *Jap. J. Apl. Phys.* 51, 04DD01 (2012).

106 E. Yurchuk, S. Mueller, D. Martin, S. Slesazeck, U. Schroeder, T. Mikolajick, J. Müller, J. Paul, R. Hoffmann, J. Sundqvist, T. Schlosser, R. Boschke, R. van Bentum, and M. Trentzsch, "Origin of the endurance degradation in the novel HfO2-based 1T ferroelectric non-volatile memories," *IEEE International Reliability Physics Symposium (IRPS)* (2014.), pp. 2E.5.1 - 2E.5.5.

107 H.-T. Lue, C.-J. Wu, and T.-Y. Tseng, "Device modeling of ferroelectric memory field-effect transistor (FeMFET)," *IEEE T. Electron Dev*. 49, 1790 (2002).

[108] P.D. Lomenzo, Q. Takmeel, C.M. Fancher, C. Zhou, N.G. Rudawski, S. Moghaddam, J.L. Jones, and T. Nishida, "Ferroelectric Si-Doped HfO2 Device Properties on Highly Doped Germanium," *Electron Device Letters*, IEEE 36, 766 (2015).

[109] E. D. Grimley, T. Schenk, T. Mikolajick, U. Schroeder, and J. M. LeBeau, "Atomic Structure of Domain and Interphase Boundaries in Ferroelectric HfO2. Advanced Materials Interfaces," 1701258 (2018). Doi: 10.1002/admi.201701258

[110] M. Pešić, F. P. G. Fengler, L/. Larcher, A. Padovani, T. Schenk, E. D. Grimley, X. Sang, J. M. LeBeau, S. Slesazeck, U. Schroeder, and T. Mikolajick, "Physical Mechanisms behind the Field-Cycling Behavior of HfO2-Based Ferroelectric Capacitors," *Advanced Functional Materials* 26, 4601 (2016).

[111] H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, and S. Slesazeck, "Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors," *ACS applied materials & interfaces* 9, 3792 (2017).

[112] M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Müller, S. Slesazeck, J. A. Ocker. "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," in *Electron Devices Meeting (IEDM), 2016 IEEE International*, 2016 Dec 3 (pp. 11-5). IEEE.

[113] S. Dünkel S, M. Trentzsch, R. Richter, P. Moll, C. Fuchs, O. Gehring, M. Majer, S. Wittek, B. Müller, T. Melde, H. Mulaosmanovic, "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond,"in *Electron Devices Meeting (IEDM), 2017 IEEE International*, 2017 Dec 2 (pp. 19-7). IEEE.

[114] M. Pesic, S. Knebel, M. Hoffmann, C. Richter, T. Mikolajick, U. Schroeder, "How to make DRAM non-volatile? Anti-ferroelectrics: A new paradigm for universal memories," *Electron Devices Meeting (IEDM), 2016 IEEE International* 2016 Dec 3 (pp. 11-6). IEEE.

[115] V. Garcia and M. Bibes, "Ferroelectric tunnel junctions for information storage and processing," *Nature Communications* 5 (2014).

[116] H. Yamada, V. Garcia, S. Fusil, S. Boyn, M. Marinova, A. Gloter, S. Xavier, J. Grollier, E. Jacquet, C. Carrétéro, C. Deranlot, M. Bibes, and A. Barthélémy, "Giant Electroresistance of Super-tetragonal BiFeO3-Based Ferroelectric Tunnel Junctions," *ACS Nano* 7, 5385 (2013).

[117] Z. Wen, C. Li, D. Wu, A. Li, and N. Ming, "Ferroelectric-field-effect-enhanced electroresistance in metal/ferroelectric/semiconductor tunnel junctions," *Nature Materials* 12, 617 (2013).

[118] C.H. Ahn, K.M. Rabe, and J.-M. Triscone, "Ferroelectricity at the nanoscale: local polarization in oxide thin films and heterostructures," *Science* 303, 488 (2004).

[119] S. Boyn, S. Girod, V. Garcia, S. Fusil, S. Xavier, C. Deranlot, H. Yamada, C. Carrétéro, E. Jacquet, M. Bibes, A. Barthélémy, and J. Grollier, "High-performance ferroelectric memory based on fully patterned tunnel junctions," *Appl. Phys. Lett*. 104, 52909 (2014).

[120] A. Chanthbouala, A. Crassous, V. Garcia, K. Bouzehouane, S. Fusil, X. Moya, J. Allibe, B. Dlubak, J. Grollier, S. Xavier, C. Deranlot, A. Moshar, R. Proksch, N.D. Mathur, M. Bibes, and A. Barthelemy, "Solid-state memories based on ferroelectric tunnel junctions," *Nat Nano* 7, 101 (2012).

[121] X.S. Gao, J.M. Liu, K. Au, and J.Y. Dai, "Nanoscale ferroelectric tunnel junctions based on ultrathin BaTiO3 film and Ag nanoelectrodes," *Applied Physics Letters* 101, 142905 (2012).

[122] A. Chen, "Emerging memory selector devices," *13th Non-Volatile Memory Technology Symposium (NVMTS)*, Singapore (2013.), pp. 1–5.

[123] Y. Kim, Y. Kim, H. Han, S. Jesse, S. Hyun, W. Lee, S.V. Kalinin, and J.K. Kim, "Towards the limit of ferroelectric nanostructures: switchable sub-10 nm nanoisland arrays," *J. Mater. Chem*. C 1, 5299 (2013).

[124] F.Y. Bruno, S. Boyn, V. Garcia, S. Fusil, S. Girod, C. Carrétéro, M. Marinova, A. Gloter, S. Xavier, C. Deranlot, M. Bibes, and A. Barthélémy, "Million-fold Resistance Change in Ferroelectric Tunnel Junctions Based on Nickelate Electrodes," *Advanced Electronic Materials* 2 (2016). DOI: 10.1002/aelm.201500245

[125] G. Sanchez-Santolino, J. Tornos, D. Hernandez-Martin, J. I. Beltran, C. Munuera, M. Cabero, A. Perez-Muñoz, J. Ricote, F. Mompean, M. Garcia-Hernandez, Z. Sefrioui, "Resonant electron tunnelling assisted by charged domain walls in multiferroic tunnel junctions," *Nature Nanotechnology* 12, 655 (2017).

[126] A. Chanthbouala, V. Garcia, R.O. Cherifi, K. Bouzehouane, S. Fusil, X. Moya, S. Xavier, H. Yamada, C. Deranlot, N.D. Mathur, M. Bibes, A. Barthélémy, and J. Grollier, "A ferroelectric memristor," *Nature Materials* 11, 860 (2012).

[127] H. Kohlstedt, A. Petraru, K. Szot, A. Rüdiger, P. Meuffels, H. Haselier, R. Waser, and V. Nagarajan, "Method to distinguish ferroelectric from nonferroelectric origin in case of resistive switching in ferroelectric capacitors," *Applied Physics Letters* 92, 62907 (2008).

[128] X. Tian, and A. Toriumi, "New opportunity of ferroelectric tunnel junction memory with ultrathin HfO_2-based oxides," *Electron Devices Technology and Manufacturing Conference (EDTM), 2017 IEEE* (pp. 36-64). IEEE.

[129] F. Ambriz-Vargas, G. Kolhatkar, M. Broyer, A. Hadj-Youssef, R. Nouar, A. Sarkissian, R. Thomas, C. Gomez-Yáñez, M. A. Gauthier, A. Ruediger, "A Complementary Metal Oxide Semiconductor Process-Compatible Ferroelectric Tunnel Junction," *ACS applied materials & interfaces* 9, 13262 (2017).

130 S. Fujii, Y. Kamimuta, T. Ino, Y. Nakasaki, R. Takaishi, M. Saitoh, "First demonstration and performance improvement of ferroelectric HfO_2-based resistive switch with low operation current and intrinsic diode property," in VLSI *Technology, 2016 IEEE Symposium on*, 2016 Jun 14 (pp. 1-2). IEEE.

131 Rebooting the IT Revolution: A Call for Action. SIA-SRC Report (2015): https://www.semiconductors.org/clientuploads/Resources/RITR%20WEB%20version%20FINAL.pdf

132 M. Hilbert and P. Lopez, "The World's Technological Capacity to Store, Communicate, and Compute Information," *Science* 332 (2011) 60

133 G. M. Church, Y. Gao, K. Yuan, S. Kosuri, "Next-generation digital information storage in DNA," *Science* 337 (2012) 1628

134 N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature* 494 (2013) 77

135 S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, "CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria," *Nature* 547 (2017) 345

136 J. Bornholt, L. Ceze, R. Lopez, G. Seelig, D. M. Carmean, K. Strauss, "A DNA-Based Archival Storage System," *OPERATING SYSTEMS REVIEW* 50 (2017) 637-649

137 J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, K. Strauss, "Toward a DNA-based archival storage system," *IEEE* Micro 37 (2017)

138 V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, "Nucleic acid memory," *Nature Materials* 15 (2016) 366

139 Y. Erlich and D. Zielinski, "DNA Fountain enables a robust and efficient storage architecture," *Science* 355 (2017) 950

140 N. F. Mott, *Metal-Insulator Transitions*, 2nd ed. (Taylor & Francis, London, 1990)

141 T. Oka and N. Nagaosa, "Interfaces of Correlated Electron Systems: Proposed Mechanism for Colossal Electroresistance," *Phys. Rev. Lett.* 95 (2005) 266403

142 A. Asamitsu, Y. Tomioka, H. Kuwahara, and Y. Tokura, "Current switching of resistive states in magnetoresistive manganites," *Nature* 388 (1997) 50

143 S. Q. Liu, N. J. Wu, and A. Ignatiev, "Electric-pulse-induced reversible resistance change effect in magnetoresistive films,," *Appl. Phys. Lett.* 76 (2000) 2749

144 A. Sawa, T. Fujii. M. Kawasaki, and Y. Tokura, "Hysteretic current-voltage characteristics and resistance switching at a rectifying Ti/Pr0.7Ca0.3MnO3 interface,," *Appl. Phys. Lett.* 85 (2004) 4073

145 D. Ruzmetov, G. Gopalakrishnan, J. Deng, V. Narayanamurti, S. Ramanathan, "Electrical triggering of metal-insulator transition in nanoscale vanadium oxide junctions,," *J. Appl. Phys.* 106 (2009) 083702

146 Y. Zhou, X. Chen, C. Ko, Z. Yang, and S. Ramanathan, "Voltage-Triggered Ultrafast Phase Transition in Vanadium Dioxide Switches,," *IEEE Electron Device Letters* 34 (2013) 220

147 M. Son, J. Lee, J. Park, J. Shin, G. Choi, S. Jung, W. Lee, S. Kim, S. Park, and H. Hwang, "Excellent Selector Characteristics of Nanoscale VO2 for High-Density Bipolar ReRAM Applications,," *IEEE Electron Device Letters* 32 (2011) 1579

148 S. D. Ha, G. H. Aydogdu, and S. Ramanathan, "Metal-insulator transition and electrically driven memristive characteristics of SmNiO3 thin films,," *Appl. Phys Lett.* 98 (2011) 012105

149 K-H. Xue, C. A. Paz de Araujo, J. Celinska, C. McWilliams, "A non-filamentary model for unipolar switching transition metal oxide resistance random acess memories,," *J. Appl. Phys*. 109 (2011) 091602

150 C. R. McWilliams, J. Celinska, C. A. Paz de Araujo, K-H. Xue, "Device characterization of correlated electron random access memories,," *J. Appl. Phys*.109 (2011) 091608

151 F. Nakamura, M. Sakaki, Y. Yamanaka, S. Tamaru, T. Suzuki, and Y. Maeno, "Electric-field-induced metal maintained by current of the Mott insulator Ca2RuO,4," *Scientific Reports* 3 (2013) 2536

152 M. D. Pickett and R. S. Williams, "Sub-100 fJ and sub-nanosecond thermally driven threshold switching in niobium oxide crosspoint nanodevices," *Nanotechnology* 23 (2012) 215202

153 M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors,," *Nature Materials* 12 (2013) 114

154 P. Stoliar , L. Cario , E. Janod , B. Corraze , C. Guillot-Deudon ,S. Salmon-Bourmand , V. Guiot , J. Tranchant , and M. Rozenberg, "Universal Electric-Field-Driven Resistive Transition in Narrow-Gap Mott Insulators,," *Advanced Materials* 25 (2013) 3222

155 V. Guiot, L. Cario, E. Janod, B. Corraze, V. Ta Phuoc, M. Rozenberg, P. Stoliar, T. Cren, and D. Roditchev, "Avalanche breakdown in GaTa4Se8-xTex narrow-gap Mott insulators,," *Nature Communications* 4 (2013) 1722

156 P. Stoliar, M. Rozenberg, E. Janod, B. Corraze, J. Tranchant, and L. Cario, "Nonthermal and purely electronic resistive switching in a Mott memory,," *Physical Review* B 90 (2014) 045146

157 E. Janod, J. Tranchant, B. Corraze, M. Querré, P. Stoliar, M. Rozenberg, T. Cren, D. Roditchev, V. T. Phuoc, M.-P. Besland, and L. Cario, "Resistive Switching in Mott Insulators and Correlated Systems,," *Advanced Functional Materials* (2015); DOI: 10.1002/adfm.201500823

158 H. Oike, F. Kagawa, N. Ogawa, A. Ueda, H. Mori, M. Kawasaki, and Y. Tokura, "Phase-change memory function of correlated electrons in organic conductors,," *Physical Review* B 91 (2015) 041101(R)

159 Fan Zuo, Priyadarshini Panda, Michele Kotiuga, Jiarui Li, Mingu Kang, Claudio Mazzoli, Hua Zhou, Andi Barbour, Stuart Wilkins, Badri Narayanan, Mathew Cherukara, Zhen Zhang, Subramanian K. R. S. Sankaranarayanan, Riccardo Comin, Karin M. Rabe, Kaushik Roy & Shriram Ramanathan, "Habituation based synaptic plasticity and organismic learning in a quantum perovskite,," *Nature Communications* 8, 240 (2017).

160 V. V. Zhirnov, R. K. Cavin, S. Menzel, E. Linn, S. Schmelzer, D. Bräuhaus, C. Schindler and R. Waser, "Memory Devices: Energy-Space-Time Trade-offs,," *Proc. IEEE* 98 (2010) 2185–2200

161 G. W. Burr, R. S. Shenoy, P. Narayanan, K. Virwani, A. Padilla B. Kurdi, and H. Hwang, "Selection devices for 3-D crosspoint memory," *Journal of Vacuum Science & Technology* B, 32(4), 040802 (2014).

162 C. Kügeler, R. Rosezin, E. Linn, R. Bruchhaus, R. Waser, "Materials, technologies, and circuit concepts for nanocrossbar-based bipolar RRAM," *Appl. Phys.* A (2011) 791-809

163 A. Chen, "Nonlinearity and Asymmetry for Device Selection in Cross-bar Memory Arrays," *IEEE Trans. Electron Dev*., 62(9), 2857-2864, (2015)

164 L. Li, K. Lu, B. Rajendran, T. D. Happ, H-L. Lung, C. Lam, and M. Chan, "Driving Device Comparison for Phase-Change Memory," *IEEE Trans. Electron. Dev*. 58 (2011) 664-671

165 U. Gruening-von Schwerin, *Patent* DE 10 2006 040238 A1; *US Patent Application* "Integrated circuit having memory cells and method of manufacture," US 2009/012758

166 G. H. Kim, K. M. Kim, J. Y. Seok, H. J. Lee, D-Y. Cho, J. H. Han, and C. S. Hwang, "A theoretical model for Schottky diodes for excluding the sneak current in cross bar array resistive memory,," *Nanotechnology* 21 (2010) 385202

167 H. Toda, "Three-dimensional programmable resistance memory device with a read/write circuit stacked under a memory cell array,," Patent (2009). US7606059

168 P. Woerlee et al. "Electrical device and method of manufacturing therefore," *Patent Application* (2005). WO 2005/124787 A2

169 S. C. Puthentheradam, D. K. Schroder, M. N. Kozicki. "Inherent diode isolation in programmable metallization cell resistive memory elements,," *Appl. Phys.* A 102 (2011) 817-826

170 J.H. Oh, et al, "Full integration of highly manufacturable 512Mb PRAM based on 90nm technology," *IEDM Tech. Dig*., pp. 515–518, Dec. 2006.

171 Y. Sasago, et al, "Cross-point phase change memory with 4F2 cell size driven by low-contact-resistivity poly-Si diode," *Symposium VLSI Tech*., pp. 24–25, Jun. 2009.

172 M. Kinoshita, et al, "Scalable 3-D vertical chain-cell-type phase-change memory with 4F2 poly-Si diodes," *Symposium VLSI Tech.*, pp. 35-36, Jun. 2012.

173 S.H. Lee, et al, "Highly Productive PCRAM Technology Platform and Full Chip Operation: Based on 4F2 (84nm Pitch) Cell Scheme for 1 Gb and Beyond," *IEDM Tech. Dig*., pp. 47-50, Dec. 2011.

174 M.J. Lee, et al, "A low-temperature-grown oxide diode as a new switch element for high-density, nonvolatile memories," *Adv. Mater.*, vol. 19, no. 1, pp. 73–76, Jan. 2007.

175 M.J. Lee, et al, "2-stack 1D-1R cross-point structure with oxide diodes as switch elements for high density resistance RAM applications," *IEDM Tech. Dig*., pp. 771–774, Dec. 2007.

176 M.J. Lee, et al, "Stack friendly all-oxide 3D RRAM using GaInZnO peripheral TFT realized over glass substrates," *IEDM Tech. Dig*., Dec. 2008.

177 S.E. Ahn, et al, "Stackable all-oxide-based nonvolatile memory with Al2O3 antifuse and p-CuOx/n-InZnOx diode," *IEEE Electron Dev. Lett.*, vol. 30, no. 5, pp. 550–552, May 2009.

178 Y. Choi, et al, "High current fast switching n-ZnO/p-Si diode," *J. Phys. D: Appl. Phys.*, vol. 43, pp. 345101-1-4, Aug. 2010.

179 S. Kim, Y. Zhang, J.P. McVittie, H. Jagannathan, Y. Nishi, and H.S.P. Wong, "Integrating phase-change memory cell with Ge nanowire diode for crosspoint memory—experimental demonstration and analysis," *IEEE Trans. Electron Dev*., vol. 55, no. 9, pp. 2307–2313, Sep. 2008.

180 Y.C. Shin, et al, "(In,Sn)2O3 /TiO2 /Pt Schottky-type diode switch for the TiO2 resistive switching memory array," *Appl. Phys. Lett.*, vol. 92, no. 16, pp. 162904-1-3, Apr. 2008.

181 W.Y. Park, et al, "A Pt/TiO2/Ti Schottky-type selection diode for alleviating the sneak current in resistance switching memory arrays," *Nanotechnology*, vol. 21, no. 19, pp. 195201-1-4, May 2010.

182 G.H. Kim, et al, "Schottky diode with excellent performance for large integration density of crossbar resistive memory," *Appl. Phys. Lett*., vol. 100, no. 21, pp. 213508-1-3, May 2012.

183 N. Huby, et al, "New selector based on zinc oxide grown by low temperature atomic layer deposition for vertically stacked non-volatile memory devices," *Microelectronic Eng*., vol. 85, no. 12, pp. 2442-2444, Dec. 2008.

184 B. Cho, et al, "Rewritable Switching of One Diode–One Resistor Nonvolatile Organic Memory Devices," *Adv. Mater.*, vol. 22, no. 11, pp. 1228–1232, Mar. 2010.

185 M. J. Lee et al., "Two Series Oxide Resistors Applicable to High Speed and High Density Nonvolatile Memory," *Adv. Mater*., 19, 3919 (2007)

186 S. D. Ha, G. H. Aydogdu, and S. Ramanathan, "Metal-insulator transition and electrically driven memristive characteristics of $SmNiO_3$ thin films," *Appl. Phys Lett.* 98 (2011) 012105

187 D. Kau, et al, "A stackable cross point phase change memory," *2009 IEDM*, 617

188 S. Kim, et al, "Ultrathin (<10 nm) Nb2O5/NbO2 hybrid memory with both memory and selector characteristics for high density 3D vertically stackable RRAM applications," *Symposium VLSI Tech*., pp. 155-156, Jun. 2012.

189 W.G. Kim, et al, "NbO_2-based low power and cost effective 1S1R switching for high density cross point ReRAM application," *VLSI Tech. Sym.*, 138 (2014).

190 J.H. Lee, et al, "Threshold switching in Si-As-Te thin film for the selector device of crossbar resistive memory," *Appl. Phys. Lett.*, vol. 100, no. 12, pp. 123505-1-4, Mar. 2012.

191 M.J. Lee, et al, "Highly-Scalable Threshold Switching Select Device based on Chalcogenide Glasses for 3D Nanoscaled Memory Arrays," *IEDM Tech. Dig*., pp. 33–35, Dec. 2012.

192 S. Kim, et al, "Performance of threshold switching in chalcogenide glass for 3D stackable selector," *VLSI Tech. Sym*., 240 (2013).

193 H. Yang, et al, "Novel selector for high density non-volatile memory with ultra-low holding voltage and 10^7 on/off ratio," *VLSI Tech. Sym.*, 130 (2015).

194 S.H. Jo, et al, "3D-stackable crossbar resistive memory based on field assisted superliner threshold (FAST) selector," *IEDM*, 160 (2014).

195 J.J. Huang, Y.M. Tseng, C.W. Hsu, and T.H. Hou, "Bipolar nonlinear Ni/TiO2/Ni selector for 1S1R crossbar array applications," *IEEE Electron Dev. Lett.*, vol. 32, no. 10, pp. 1427–1429, Oct. 2011.

196 J. Shin, et al, "TiO2-based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application," *J. Appl. Phys*., vol. 109, no. 3, pp. 033712-1-4, Feb. 2011.

197 W. Lee, et al, "Varistor-type bidirectional switch (JMAX>107A/cm2, selectivity~104) for 3D bipolar resistive memory arrays," *Symposium VLSI Tech*., pp. 37–38, Jun. 2012.

198 J. Woo, et al, "Multi-layer tunnel barrier ($Ta_2O_5/TaO_x/TiO_2$) engineering for bipolar RRAM selector applications," *VLSI Tech. Sym*., 168 (2013).

199 Y.H. Song, S.Y. Park, J.M. Lee, H.J. Yang, G.H. Kil, "Bidirectional two-terminal switching device for crossbar array architecture," *IEEE Electron Dev. Lett*., vol. 32, no. 8, pp. 1023–1025, Aug. 2011.

200 K. Gopalakrishnan, et al, "Highly Scalable Novel Access Device based on Mixed Ionic Electronic Conduction (MIEC) Materials for High Density Phase Change Memory (PCM) Arrays," *2010 VLSI Symp.*, 205

201 R.S. Shenoy, et al, "Endurance and scaling trends of novel access-devices for multi-layer crosspoint memory based on mixed ionic electronic conduction (MIEC) materials," *Symposium VLSI Tech*., T5B-1, Jun. 2011.

202 G.W. Burr, et al, "Large-scale (512kbit) integration of multilayer-ready access-devices based on mixed-ionic-electronic-conduction (MIEC) at 100% yield," *Symposium VLSI Tech*., T5.4, Jun. 2012.

203 K. Virwani, et al, "Sub-30 nm scaling and high-speed operation of fully-confined access-devices for 3D crosspoint memory based on mixed-ionic-electronic-conduction (MIEC) materials," *IEDM Tech. Dig*., pp. 36–39, Dec. 2012.

204 E. Linn, R. Rosezin, C. Kugeler, and R. Waser, Complementary resistive switches for passive nanocrossbar memories, *NATURE MATERIALS* 9 (2010) 403–406

205 S. Tappertzhofen, et al., "Capacity based nondestructive readout for complementary resistive switches," *Nanotech*., vol. 22, no. 39, pp. 395203-1-7, Sep. 2011.

206 R. Rosezin, et al., "Integrated complementary resistive switches for passive high-density nanocrossbar arrays," *IEEE Electron Dev. Lett.*, vol. 32, no. 2, pp. 191–193, Feb. 2011.

207 Y. Chai, et al., "Nanoscale bipolar and complementary resistive switching memory based on amorphous carbon," *IEEE Trans. Electron Dev*., vol. 58, no. 11, pp. 3933–3939, Nov. 2011.

208 S. Schmelzer, E. Linn, U. Bottger, and R. Waser, "Uniform complementary resistive switching in tantalum oxide using current sweeps," *IEEE Electron Dev. Lett.*, vol. 34, no. 1, pp. 114–116, Jan. 2013.

209 Y. C. Bae, et al., "Oxygen ion drift-induced complementary resistive switching in homo TiOx/TiOy/TiOx and hetero TiOx/TiON/TiOx triple multilayer frameworks," *Adv. Funct. Materi*., vol. 22, no. 4, pp. 709–716, Feb. 2012.

210 F. Nardi, S. Balatti, S. Larentis, and D. Ielmini, "Complementary switching in metal oxides: toward diode-less crossbar RRAMs," *IEDM Tech. Dig*., pp. 709–712, Dec. 2011.

211 J. Lee, et al, "Diode-less nano-scale ZrOx/HfOx RRAM device with excellent switching uniformity and reliability for high-density cross-point memory applications," *IEDM Tech. Dig*., pp. 452–455, Dec. 2010.

212 N. Banno, et al, "Nonvolatile 32 x 32 crossbar atom switch block integrated on a 65-nm CMOS platform," *Symposium VLSI Tech*., pp. 39–40, Jun. 2012.

213 X. Liu, et al, "Complementary resistive switching in niobium oxide-based resistive memory devices," *IEEE Electron Dev. Lett*., vol. 34, no. 2, pp. 235-237, Feb. 2013.

214 B. S. Simpkins, M. A. Mastro, C. R. Eddy, R. E. Pehrsson, "Surface depletion effects in semiconducting nanowires," *J. Appl. Phys*. 103 (2008) 104313

215 V. V. Zhirnov, R. Meade, R. K. Cavin, G. Sandhu, "Scaling limits of resistive memories ," *Nanotechnology* 22 (2011) 254027

216 R. E. Fontana, Jr. G. M. Decad, and S. R. Hetzler, "The Impact of Areal Density and Millions of Square Inches (MSI) of Produced Memory on Petabyte Shipments of TAPE, NAND Flash, and HDD," *MSS&T 2013 Conference Proceedings*, May 2013.

217 Y. Deng and J. Zhou, "Architectures and optimization methods of flash memory based storage systems," *J. Syst. Arch.* 57 (2011) 214–227.

218 L. M. Grupp, A. M. Caulfield, J. Coburn, S. Swanson, E. Yaakobi, P. H. Siegel, J. K. Wolf "Characterizing Flash Memory: Anomalies, Observations, and Applications," *MICRO'*09, Dec. 12-16, 2009, New York, NY, USA, pp. 24–33

219 G. W. Burr, B. N. Kurdi, J. C. Scott, C. H. Lam, K. Gopalakrishnan, and R. S. Shenoy, "Overview of Candidate Device Technologies for Storage-Class Memory," *IBM J. Res. & Dev*. 52, No. 4/5, 449–464 (2008).

220 R. F. Freitas and W. W. Wilcke, "Storage-Class Memory: The Next Storage System Technology," *IBM J. Res. & Dev*. 52, No. 4/5, 439–447 (2008).

221 M. Franceschini, M. Qureshi, J. Karidis, L. Lastras, A. Bivens, P. Dube, and B. Abali, "Architectural Solutions for Storage-Class Memory in Main Memory," *CMRR Non-volatile Memories Workshop*, April 2010, http://nvmw.eng.ucsd.edu/2010/documents/Franceschini_Michael.pdf

222 M. Johnson, A. Al-Shamma, D. Bosch, M. Crowley, M. Farmwald, L. Fasoli, A. Ilkbahar, et al., "512-Mb PROM with a Three-Dimensional Array of Diode/Antifuse Memory Cells," *IEEE J. Solid-State Circ*. 38, No. 11, 1920–1928 (2003).

223 M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable high performance main memory system using phase-change memory technology," *ISCA '09 – Proceedings of the 36th annual International Symposium on Computer Architecture*, pages 24–33, ACM, (2009).

224 www.micron.com/about/innovations/3d-xpoint-technology

225 www.computerworld.com/article/2951869/computer-hardware/intel-and-micron-unveil-3d-xpoint-a-new-class-of-memory.html

226 https://www.micron.com/products/advanced-solutions/3d-xpoint-technology

227 https://www.intel.com/content/www/us/en/architecture-and-technology/intel-micron-3d-xpoint-webcast.html

228 www.eetimes.com/author.asp?section_id=36&doc_id=1327313

229 www.hpcwire.com/2015/08/27/micron-steers-roadmap-through-memory-scaling-course/

230 "Final Report, Exascale Study Group: Technology Challenges in Advanced Exascale Systems" (DARPA), 2007.

231 L.A. Barroso, and U. Holzle, "The case for Energy-Proportional Computing," *IEEE Computer*, Vol. 40(12), Dec. 2007, pp. 33–37.

232 R. F. Freitas and W. W. Wilcke, "Storage-Class Memory: The Next Storage System Technology," *IBM J. Res. & Dev*. 52, No. 4/5, 439–447 (2008).

233 S. Swanson, "System architecture implications for M/S-class SCMs, *ITRS SCM workshop*, July 2012.

234 M. Awasthi, M. Shevgoor, K. Sudan, B. Rajendran, R. Balasubramonian, V. Srinivasan, "Efficient scrub mechanims for error-prone emerging memories," *HPCA 2012*.

235 K. H. Kim, "Memory Interfaces for M-Class SCMs," *ITRS SCM workshop*, July 2012.

236. K. Qureshi, J. Karidis, M. Franceschini, V. Srinivasan, L. Lastras, and B. Abali, "Enhancing lifetime and security of pcm-based main memory with start-gap wear leveling," *MICRO 42: Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 14–23, ACM, (2009).

237 M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable high performance main memory system using phase-change memory technology," *ISCA '09 - Proceedings of the 36th annual International Symposium on Computer Architecture*, pages 24–33, ACM, (2009).

238 A. M. Caulfield, J. Coburn, T. Mollov, A. De, A. Akel, J. He, A. Jagatheesan, "Understanding the impact of emerging non-volatile memories on high-performance, io-intensive computing," *Proc. ACM/IEEE Int. Conf. High Perform. Comput.,Netw., Storage Anal*., 2010, pp. 1–11.

239 E. Kultursay, M. Kandemir, A. Sivasubramaniam, and O. Mutlu, "Evaluating STT-RAM as an energy-efficient main memory alternative," *Proceedings of the 2013 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, (2013).

240 H. Lee, "High-Performance NAND and PRAM Hybrid Storage Design for Consumer Electronics," *IEEETrans.Consumer Electronics*, Vol. 56(1), 112–118 (2010).

241 P. Ranganathan, "From Microprocessors to Nanostores: Rethinking Data-Centric Systems," *COMPUTER* 44 (2011) 39–48.

242 J. Chang, "Data-centric computing and Nanostores," *ITRS SCM workshop*, July 2012.

243 K.M. Bresniker, S. Singhal and R.S. Williams, "Adapting to thrive in a new economy of memory abundance," *IEEE Computer*, Dec. 2015, pp. 44–53.

244 D. Kim, K. Bang, S-H. Ha, S. Yoon, and E-Y. Chung, "Architecture exploration of high-performance PCs with a solid-state disk," *IEEE Trans. Comp*. 59 (2010) 879–890

[245] J. H. Yoon, E. H. Nam, Y. J. Seong, H. Kim, B. S. Kim, S. L. Min, Y. Cho, "Chameleon: A high performance Flash/FRAM hybrid solid state disk architecture," *IEEE Comp. Arch. Lett.* 7 (2008) 17-20

[246] H. G. Lee, "High-Performance NAND and PRAM Hybrid Storage Design for Consumer Electronics," IEEE *Trans.Consumer Electronics*, Vol. 56(1), 112–118 (2010).

[247] E. L. Miller, "Object-based interfaces for efficient and portable access to S-class SCMs," *ITRS SCM workshop*, July 2012.

[248] H. Zhang, G. Chen, B. C. Ooi, K.-L. Tan, and M. Zhang, "In-Memory Big Data Management and Processing: A Survey," *IEEE Trans. Knowledge Data. Engr*., 27(7), 1920–1948 (2015).

[249] S. Chen, P. B. Gibbons, and S. Nath, "Rethinking database algorithms for phase change memory," *Proc. CIDR*, 2011, pp. 21–31.

[250] B.-D. Yang, J.-E. Lee, J.-S. Kim, J. Cho, S.-Y. Lee, and B.-G. Yu, "A low power phase-change random access memory using a data comparison write scheme," *Proc. IEEE Int. Symp. Circuits Syst*., 2007, pp. 3014–3017.

[251] B. C. Lee, E. Ipek, O. Mutlu, and D. Burger, "Architecting phase change memory as a scalable dram alternative," *Proc. 36th Annu. Int. Symp. Comput. Archit.*, 2009, pp. 2–3.

[252] S. D. Viglas, "Write-limited sorts and joins for persistent memory," *Proc. VLDB Endowment*, vol. 7, pp. 413–424, 2014.

[253][253] S. Chen and Q. Jin, "Persistent b+-trees in non-volatile main memory," *Proc. VLDB Endowment*, vol. 8, pp. 786–797, 2015.J. Huang, K. Schwan, and M. K. Qureshi, "NVRAM-aware logging in transaction systems," *Proc. VLDB Endowment*, vol. 8, pp. 389–400, 2014.

[254] T. Wang and R. Johnson, "Scalable logging through emerging non-volatile memory," *Proc. VLDB Endowment*, vol. 7, pp. 865–876, 2014.

[255] J. Huang, K. Schwan, and M. K. Qureshi, "NVRAM-aware logging in transaction systems," *Proc. VLDB Endowment*, vol. 8, pp. 389–400, 2014.

[256] A. Chatzistergiou, M. Cintra, and S. D. Viglas, "REWIND: Recovery write-ahead system for in-memory non-volatitledatastructures," *Proc. VLDB Endowment*, vol. 8, pp. 497–508, 2015.

[257] R. Fang, H.-I. Hsiao, B. He, C. Mohan, and Y. Wang, "High performance database logging using storage class memory," *Proc. IEEE 27th Int. Conf. Data Eng.*, 2011, pp. 1221–1231.

[258] J. Condit, E. B. Nightingale, C. Frost, E. Ipek, B. Lee, D. Burger, and D. Coetzee, "Better I/O through byte-addressable, persistent memory," *Proc. ACM SIGOPS 22nd Symp. Operating Syst. Principles*, 2009, pp. 133–146.

[259] A. Akel, A. M. Caulfield, T. I. Mollov, R. K. Gupta, and S. Swanson, "Onyx: a protoype phase change memory storage array," *Hot Storage*, (2011).

[260] Coburn, A. M. Caulfield, A. Akel, L. M. Grupp, R. K. Gupta, R. Jhala, and S. Swanson, "Nv-Heaps: Making Persistent Objects Fast and Safe with Next-Generation, Non-Volatile Memories," *ACM Sigplan Notices*, **47**(4), 105-117, (2012).

[261] J. Coburn, A. M. Caulfield, A. Akel, L. M. Grupp, R. K. Gupta, R. Jhala, and S. Swanson, "Nv-Heaps: Making Persistent Objects Fast and Safe with Next-Generation, Non-Volatile Memories," *ACM Sigplan Notices*, **47**(4), 105-117, (2012).

[262] H. Volos, A. J. Tack, and M. M. Swift, "Mnemosyne: Lightweight persistent memory," *Proc. 16th Int. Conf. Archit. Support Program. Languages Operating Syst*., 2011, pp. 91–104.

[263] I. Moraru, D. G. Andersen, M. Kaminsky, N. Tolia, P. Ranganathan, and N. Binkert, "Consistent, durable, and safe memory management for byte-addressable non volatile main memory," presented at the *Conf. Timely Results Operating Syst. Held in Conjunction with SOSP*, Farmington, PA, USA, 2013.

[264] S. Venkataraman, N. Tolia, P. Ranganathan, and R. H. Campbell, "Consistent and durable data structures for non-volatile byte addressable memory," *Proc. 9th USENIX Conf. File Storage Technol*., 2011, p. 5.

[265] S. R. Dulloor, S. Kumar, A. Keshavamurthy, P. Lantz, D. Reddy, R. Sankaran, and J. Jackson, "System software for persistent memory," *Proc. 9th Eur. Conf. Comput. Syst*., 2014, pp. 15:1–15:15.

[266] J. Jung, Y. Won, E. Kim, H. Shin, and B. Jeon, "FRASH: Exploiting storage class memory in hybrid file system for hierarchical storage," *ACM Trans. Storage*, vol. 6, pp. 3:1–3:25, 2010.

[267] A.-I. A. Wang, G. Kuenning, P. Reiher, and G. Popek, "The conquest file system: Better performance through a disk/persistent ram hybrid design," *ACM Trans. Storage*, vol. 2, pp. 309–348, 2006.

[268] X. Wu and A. L. N. Reddy, "SCMFS: A file system for storage class memory," P*roc. Int. Conf. High Perform. Comput., Netw., Storage Anal*., 2011, pp. 39:1–39:11.

[269] Q. Cao, S.-J. Han, J. Tersoff, A. D. Franklin, Y. Zhu, Z. Zhang, G. S. Tulevski, J. Tang, and W. Haensch, "End-bonded contacts for carbon nanotube transistors with low, size-independent resistance," *Science* 350, 68 (2015).

[270] A. D. Franklin, D. B. Farmer, and W. Haensch, "Defining and Overcoming the Contact Resistance Challenge in Scaled Carbon Nanotube Transistors," *ACS Nano* 8, 7333 (2014).

[271] A. D. Franklin, M. Luisier, S. J. Han, G. Tulevski, C. M. Breslin, L. Gignac, M. S. Lundstrom, and W. Haensch, "Sub-10 nm Carbon Nanotube Transistor," *Nano Lett.* 12, 758 (2012).

[272] A. D. Franklin, S. O. Koswatta, D. B. Farmer, J. T. Smith, L. Gignac, C. M. Breslin, S. J. Han, G. S. Tulevski, H. Miyazoe, W. Haensch, and J. Tersoff, "Carbon Nanotube Complementary Wrap-Gate Transistors," *Nano Lett.* 13, 2490 (2013).

[273] M. Steiner, M. Engel, Y. M. Lin, Y. Q. Wu, K. Jenkins, D. B. Farmer, J. J. Humes, N. L. Yoder, J. W. T. Seo, A. A. Green, M. C. Hersam, R. Krupke, and P. Avouris, "High-frequency performance of scaled carbon nanotube array field-effect transistors," *Appl. Phys. Lett.* 101, 053123 (2012).

[274] L. Ding, Z. Y. Zhang, S. B. Liang, T. Pei, S. Wang, Y. Li, W. W. Zhou, J. Liu, and L. M. Peng, "CMOS-based carbon nanotube pass-transistor logic integrated circuits," *Nat. Commun*. 677 (2012).

[275] M. M. Shulaker, G. Hills, N. Patil, H. Wei, H.-Yu Chen, H.-S. Philip Wong, and S. Mitra, "Carbon nanotube computer," *Nature* 501, 526 (2013).

[276] Q. Cao, J. Tersoff, S. J. Han, and A. V. Penumatcha, "Scaling of Device Variability and Subthreshold Swing in Ballistic Carbon Nanotube Transistors," *Phys. Rev.* Appl. 4, 024022 (2015).

[277] R. S. Park, M. M. Shulaker, G. Hills, L. S. Liyanage, S. Lee, A. Tang, S. Mitra, and H. -S. P. Wong, "Hysteresis in Carbon Nanotube Transistors: Measurement and Analysis of Trap Density, Energy Level, and Spatial Distribution," *ACS Nano* 10, 4599 (2016).

[278] G. J. Brady, Y. Joo, M. -Y. Wu, M. J. Shea, P. Gopalan, and M. S. Arnold, "Polyfluorene-Sorted, Carbon Nanotube Array Field-Effect Transistors with Increased Current Density and High On/Off Ratio," *ACS Nano* 8, 11614 (2014).

[279] G. S. Tulevski, A. D. Franklin, D. Frank, J. M. Lobez, Q. Cao, H. Park, A. Afzali, S. -J. Han, J. B. Hannon, and W. Haensch, "Toward High-Performance Digital Logic Technology with Carbon Nanotubes," *ACS Nano* 8, 8730 (2014).

[280] Ma, D. D., Lee, C. S., Au, F. C., Tong, S. Y., & Lee, S. T. (2003, Mar. 21), "Small-diameter silicon nanowire surfaces," *Science*, 299, 1874-1877.

[281] Yan, H., & Yang, P, "Semiconductor nanowires: functional building blocks for nanotechnology," in P. Yang (Ed.), *The Chemistry of Nanostructured Materials.* River Edge, NJ: World Scientific (2004).

[282] Beckman, R., Johnston-Halperin, E., Luo, Y., Green, J. E., & Hearh, J. R., "Bridging dimensions: demultiplexing ultrahigh-density nanowire circuits," *Science*, 310(5747), 465-468. (2005, Oct. 21).

[283] Chuang, S. et al., "Ballistic InAs Nanowire Transistors," *Nano Lett.* (2012). doi:10.1021/nl3040674

[284] Cui, Y., Lauhon, L. J., Gudiksen, M. S., Wang, J., & Lieber, C. M., "Diameter-controlled synthesis of single-crystal silicon nanowires," *Appl. Phys. Lett.*, 78(15), 2214-2216. (2001, Apr. 9).

[285] Schmid, H.; Borg, M., Moselund, K., Gignac, L., Breslin, C. M., Bruley, J., Cutaia, D., Riel, H., *Appl. Phys. Lett*. 2015, 106, 233101.

[286] Wu, Y., & Yang, P., "Germanium nanowire growth via simple vapor transport," *Chem. Mater.*, 12, 605-607. (2000).

[287] Xiang, J., Lu, W., Hu, Y., Wu, Y., Yan, H., & Lieber, C. M., "Ge/Si nanowire heterostructures as high-performance field-effect transistors," *Nature*, 441, 489-493. (2006).

[288] Yang, B., Buddharaju, K. D., Teo, S. H., Singh, N., Lo, G. Q., & Kwong, D. L., "Vertical silicon-nanowire formation and gate-all-around MOSFET," *IEEE Elect. Dev. Lett.*, 29(7), 791-794. (2008, Jul.).

[289] Wernersson, L.-E., Thelander, C., Lind, E., & Samuelson, L., "III-V nanowires--extending a narrowing road," *Proc. IEEE*, 98(12), 2047-2060. (2010, Dec. 12).

[290] Autran, J.-L., & Munteanu, D., "Beyond silicon bulk MOS transistor: new materials, emerging structures and ultimate devices," *Revue de l'Electricite et de l'Electronique*, 4, 25-37. (2007, Apr.).

[291] Ng, H. T., Han, J., Yamada, T., Nguyen, P., Chen, Y. P., & Meyyappan, M., "Single crystal nanowire vertical surround-gate field-effect transistor," *Nano Lett.*, 4(7), 1247-1252. (2004).

[292] Johansson, S., Memisevic, E., Wernersson, L.-E. & Lind, E., "High-Frequency Gate-All-Around Vertical InAs Nanowire MOSFETs on Si Substrates," *Electron Device Lett. IEEE* 35, 518 - 520 (2014).

[293] Berg, M. et al., "InAs nanowire MOSFETs in three-transistor configurations: single balanced RF down-conversion mixers," *Nanotechnology* 25, 485203 (2014).

[294] Yan, H., Choe, H. S., Nam, S., Hu, Y., Das, S., Klemic, J. F., et al., "Programmable nanowire circuits for nanoprocessors," *Nature*, 470, 240-244. (2011, Feb. 10).

[295] Yao, J. et al., "Nanowire nanocomputer as a finite-state machine," *Proc. Natl. Acad. Sci.* 111, 2431–2435 (2014). doi:10.1073/pnas.1323818111

[296] Fortuna, S. A., Wen, J., Chun, I., & Li, X., "Planar GaAs Nanowires on GaAs (100) Substrates: Self-Aligned, Nearly Twin-Defect Free, and Transfer-Printable," *Nano Lett.* 8, 4421-4427. (2008).

[297] Miao, X. Chabak, K., Zhang, C., Mohseni, P. K., Walker Jr, D., & Li, X., "High-speed planar GaAs nanowire arrays with fmax> 75 GHz by wafer-scale bottom-up growth," *Nano Lett.* 15 (5), pp 2780-2786. (2015)

[298] Lu, W., & Lieber, C. M., "Nanoelectronics from the bottom up," *Nature Materials*, 6, 841-850. (2007, Nov.).

[299] Zhang, C. & Li, X., "III-V Nanowire Transistors for Low-Power Logic Applications: a Review and Outlook," *IEEE Transactions on Electron Devices*, 63(1), 223. (2016)

[300] Mertens, H. et al., "Gate-All-Around MOSFETs based on Vertically Stacked Horizontal Si Nanowires in a Replacement Metal Gate Process on Bulk Si Substrates," *2016 Symposium on VLSI Technology Digest of Technical Papers*, pp. 158-159 (2016).

[301] Mertens, H., Ritzenthaler, R., Chasin, A., et al., "Vertically stacked gate-all-around Si nanowire CMOS transistors with dual work function metal gates," *2016 IEEE International Electron Devices Meeting (IEDM)*. doi:10.1109/IEDM.2016.7838456 (2016)

[302] Mertens, H. et al., "Vertically stacked gate-all-around Si nanowire transistors: Key Process Optimizations and Ring Oscillator Demonstration," *IEDM* 37.4.1 - 37.4.4 (2017)

[303] Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D.; Zhang, Y., Dubonos, S. V., Grigorieva, I. V., Firsov, A. A., "Electric field effect in atomically thin carbon films," *Science* 2004**,** *306* (5696), 666-669.

[304] Chen, J.-H., Jang, C., Xiao, S., Ishigami, M., Fuhrer, M. S., "Intrinsic and extrinsic performance limits of graphene devices on SiO2," *Nat Nano* 2008**,** *3* (4), 206-209.

[305] Bolotin, K. I., Sikes, K. J., Jiang, Z., Klima, M., Fudenberg, G., Hone, J., Kim, P., Stormer, H. L., "Ultrahigh electron mobility in suspended graphene," *Solid State Commun.* 2008**,** *146* (9-10), 351-355.

[306] Castro, E. V., Ochoa, H., Katsnelson, M. I., Gorbachev, R. V., Elias, D. C., Novoselov, K. S., Geim, A. K., Guinea, F., "Limits on Charge Carrier Mobility in Suspended Graphene due to Flexural Phonons," *Physical Review Letters* 2010**,** *105* (26), 266601.

[307] Dean, C. R., Young, A. F., Meric, I., Lee, C., Wang, L., Sorgenfrei, S., Watanabe, K., Taniguchi, T., Kim, P., Shepard, K. L., Hone, J., "Boron nitride substrates for high-quality graphene electronics," *Nature Nanotechnology* 2010**,** *5* (10), 722-726.

[308] Mayorov, A. S., Gorbachev, R. V., Morozov, S. V., Britnell, L., Jalil, R., Ponomarenko, L. A., Blake, P., Novoselov, K. S., Watanabe, K., Taniguchi, T., Geim, A. K., "Micrometer-Scale Ballistic Transport in Encapsulated Graphene at Room Temperature" *Nano Letters* 2011**,** *11* (6), 2396-2399.

[309] Wang, L., Meric, I., Huang, P. Y., Gao, Q., Gao, Y., Tran, H., Taniguchi, T., Watanabe, K., Campos, L. M., Muller, D. A., Guo, J., Kim, P., Hone, J., Shepard, K. L., Dean, C. R., "One-Dimensional Electrical Contact to a Two-Dimensional Material," *Science* 2013**,** *342* (6158), 614-617.

[310] Berger, C., Song, Z., Li, T., Li, X., Ogbazghi, A. Y., Feng, R., Dai, Z., Marchenkov, A. N., Conrad, E. H., First, P. N., de Heer, W. A., "Ultrathin Epitaxial Graphite: 2D Electron Gas Properties and a Route toward Graphene-based Nanoelectronics," *The Journal of Physical Chemistry B* 2004**,** *108* (52), 19912-19916.

[311] Emtsev, K. V., Bostwick, A., Horn, K., Jobst, J., Kellogg, G. L., Ley, L., McChesney, J. L., Ohta, T., Reshanov, S. A., Rohrl, J., Rotenberg, E., Schmid, A. K., Waldmann, D., Weber, H. B., Seyller, T., "Towards wafer-size graphene layers by atmospheric pressure graphitization of silicon carbide," *Nat Mater* 2009**,** *8* (3), 203-207.

[312] Bae, S., Kim, H., Lee, Y., Xu, X., Park, J.-S., Zheng, Y., Balakrishnan, J., Lei, T., Ri Kim, H., Song, Y. I., Kim, Y.-J., Kim, K. S., Özyilmaz, B., Ahn, J.-H., Hong, B. H., Iijima, S., "Roll-to-roll production of 30-inch graphene films for transparent electrodes," *Nature Nanotechnology* 2010**,** *5*, 574.

[313] Kondo, D., Sato, S., Yagi, K., Harada, N., Sato, M., Nihei, M., Yokoyama, N., "Low-Temperature Synthesis of Graphene and Fabrication of Top-Gated Field Effect Transistors without Using Transfer Processes," *Applied Physics Express* 2010**,** *3* (2), 025102.

[314] Li, X. S., Cai, W. W., An, J. H., Kim, S., Nah, J., Yang, D. X., Piner, R., Velamakanni, A., Jung, I., Tutuc, E., Banerjee, S. K., Colombo, L., Ruoff, R. S., "Large-Area Synthesis of High-Quality and Uniform Graphene Films on Copper Foils," *Science* 2009**,** *324* (5932), 1312-1314.

[315] Yu, Q., Lian, J., Siriponglert, S., Li, H., Chen, Y. P., Pei, S.-S., "Graphene segregated on Ni surfaces and transferred to insulators," *Applied Physics Letters* 2008**,** *93* (11), 113103.

[316] Hu, B., Ago, H., Ito, Y., Kawahara, K., Tsuji, M., Magome, E., Sumitani, K., Mizuta, N., Ikeda, K.-i., Mizuno, S., "Epitaxial growth of large-area single-layer graphene over Cu(111)/sapphire by atmospheric pressure CVD," *Carbon* 2012**,** *50* (1), 57-65.

[317] Kondo, D., Nakano, H., Zhou, B., Kubota, I., Hayashi, K., Yagi, K., Takahashi, M., Sato, M., Sato, S., Yokoyama, N., "Intercalated multi-layer graphene grown by CVD for LSI interconnects," *2013 IEEE International Interconnect Technology Conference (IITC2013)*, Kyoto, Japan, Kyoto, Japan, 2013, p 190.

[318] Banszerus, L., Schmitz, M., Engels, S., Dauber, J., Oellers, M., Haupt, F., Watanabe, K., Taniguchi, T., Beschoten, B., Stampfer, C., "Ultrahigh-mobility graphene devices from chemical vapor deposition on reusable copper," *Science Advances* 2015**,** *1* (6).

[319] Baringhaus, J., Ruan, M., Edler, F., Tejeda, A., Sicot, M., Taleb-Ibrahimi, A., Li, A.-P., Jiang, Z., Conrad, E. H., Berger, C., Tegenkamp, C., de Heer, W. A., "Exceptional ballistic transport in epitaxial graphene nanoribbons," *Nature* 2014**,** *506*, 349.

[320] Shintaro, S., "Graphene for nanoelectronics," *Japanese Journal of Applied Physics* 2015**,** *54* (4), 040102.

[321] McCann, E., "Asymmetry gap in the electronic band structure of bilayer graphene," *Physical Review B* 2006**,** *74* (16).

[322] McCann, E., Fal'ko, V. I., "Landau-Level Degeneracy and Quantum Hall Effect in a Graphite Bilayer," *Physical Review Letters* 2006**,** *96* (8), 086805.

[323] Oostinga, J. B., Heersche, H. B., Liu, X. L., Morpurgo, A. F., Vandersypen, L. M. K., "Gate-induced insulating state in bilayer graphene devices," *Nature Materials* 2008**,** *7* (2), 151-157.

[324] Xia, F. N., Farmer, D. B., Lin, Y. M., Avouris, P., "Graphene Field-Effect Transistors with High On/Off Current Ratio and Large Transport Band Gap at Room Temperature," *Nano Letters* 2010**,** *10* (2), 715-718.

[325] Zhang, Y. B., Tang, T. T., Girit, C., Hao, Z., Martin, M. C., Zettl, A., Crommie, M. F., Shen, Y. R., Wang, F., "Direct observation of a widely tunable bandgap in bilayer graphene," *Nature* 2009**,** *459* (7248), 820-823.

[326] Kim, K. S., Walter, A. L., Moreschini, L., Seyller, T., Horn, K., Rotenberg, E., Bostwick, A., "Coexisting massive and massless Dirac fermions in symmetry-broken bilayer graphene," *Nature Materials* 2013**,** *12*, 887.

[327] Fujita, M., Wakabayashi, K., Nakada, K., Kusakabe, K., "Peculiar localized state at zigzag graphite edge," *Journal of the Physical Society of Japan* 1996**,** *65* (7), 1920-1923.

[328] Han, M. Y., Özyilmaz, B., Zhang, Y., Kim, P., "Energy Band-Gap Engineering of Graphene Nanoribbons," *Physical Review Letters* 2007**,** *98* (20), 206805.

[329] Jiao, L. Y., Zhang, L., Wang, X. R., Diankov, G., Dai, H. J., "Narrow graphene nanoribbons from carbon nanotubes," *Nature* 2009**,** *458* (7240), 877-880.

[330] Kosynkin, D. V., Higginbotham, A. L., Sinitskii, A., Lomeda, J. R., Dimiev, A., Price, B. K., Tour, J. M., "Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons," *Nature* 2009**,** *458* (7240), 872-U5.

[331] Li, X. L., Wang, X. R., Zhang, L., Lee, S. W., Dai, H. J., "Chemically derived, ultrasmooth graphene nanoribbon semiconductors," *Science* 2008**,** *319* (5867), 1229-1232.

[332] Nakada, K., Fujita, M., Dresselhaus, G., Dresselhaus, M. S., "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Physical Review B* 1996**,** *54* (24), 17954-17961.

[333] Son, Y.-W., Cohen, M. L., Louie, S. G., "Energy Gaps in Graphene Nanoribbons," *Physical Review Letters* 2006**,** *97* (21), 216803.

[334] Yang, L., Park, C.-H., Son, Y.-W., Cohen, M. L., Louie, S. G., "Quasiparticle Energies and Band Gaps in Graphene Nanoribbons," *Physical Review Letters* 2007**,** *99* (18), 186801.

[335] Zhu, X., Su, H., "Scaling of Excitons in Graphene Nanoribbons with Armchair Shaped Edges," *The Journal of Physical Chemistry A* 2011**,** *115* (43), 11998-12003.

[336] Liang, X., Wi, S., "Transport Characteristics of Multichannel Transistors Made from Densely Aligned Sub-10 nm Half-Pitch Graphene Nanoribbons," *ACS Nano* 2012**,** *6* (11), 9700-9710.

[337] Gallagher, P., Todd, K., "Goldhaber-Gordon, D., Disorder-induced gap behavior in graphene nanoribbons," *Physical Review B* 2010**,** *81* (11), 115409.

[338] Han, M. Y., Brant, J. C., Kim, P., "Electron Transport in Disordered Graphene Nanoribbons," *Physical Review Letters* 2010**,** *104* (5), 056801.

[339] Cai, J. M., Ruffieux, P., Jaafar, R., Bieri, M., Braun, T., Blankenburg, S., Muoth, M., Seitsonen, A. P., Saleh, M., Feng, X. L., Mullen, K., Fasel, R., "Atomically precise bottom-up fabrication of graphene nanoribbons," *Nature* 2010**,** *466* (7305), 470-473.

[340] Ruffieux, P., Cai, J., Plumb, N. C., Patthey, L., Prezzi, D., Ferretti, A., Molinari, E., Feng, X., Müllen, K., Pignedoli, C. A., Fasel, R., "Electronic Structure of Atomically Precise Graphene Nanoribbons," *Acs Nano* 2012**,** *6* (8), 6930-6935.

[341] Talirz, L., Ruffieux, P., Fasel, R., "On-Surface Synthesis of Atomically Precise Graphene Nanoribbons," *Advanced Materials* 2016**,** *28* (29), 6222-6231.

[342] Fiori, G., Iannaccone, G., "Simulation of graphene nanoribbon field-effect transistors," *IEEE Electron Device Letters* 2007**,** *28* (8), 760-762.

[343] Harada, N., Sato, S., "Yokoyama, N., Theoretical Investigation of Graphene Nanoribbon Field-Effect Transistors Designed for Digital Applications," *Japanese Journal of Applied Physics* 2013**,** *52*, 094301.

[344] Tsuchiya, H., Ando, H., Sawamoto, S., Maegawa, T., Hara, T., Yao, H., Ogawa, M., "Comparisons of Performance Potentials of Silicon Nanowire and Graphene Nanoribbon MOSFETs Considering First-Principles Bandstructure Effects," *Electron Devices, IEEE Transactions on* 2010**,** *57* (2), 406-414.

[345] Bennett, P. B., Pedramrazi, Z., Madani, A., Chen, Y.-C., Oteyza, D. G. d., Chen, C., Fischer, F. R., Crommie, M. F., Bokor, J., "Bottom-up graphene nanoribbon field-effect transistors," *Applied Physics Letters* 2013**,** *103* (25), 253114.

[346] Llinas, J. P., Fairbrother, A., Borin Barin, G., Shi, W., Lee, K., Wu, S., Yong Choi, B., Braganza, R., Lear, J., Kau, N., Choi, W., Chen, C., Pedramrazi, Z., Dumslaff, T., Narita, A., Feng, X., Müllen, K., Fischer, F., Zettl, A., Ruffieux, P., Yablonovitch, E., Crommie, M., Fasel, R., Bokor, J., "Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons," *Nature Communications* 2017**,** *8* (1), 633.

[347] Qiu, C., Zhang, Z., Peng, L. M., "Scaling carbon nanotube CMOS FETs towards quantum limit," *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017, 2017, pp 5.5.1-5.5.4.

[348] Zhao, P., Chauhan, J., Guo, J., "Computational Study of Tunneling Transistor Based on Graphene Nanoribbon," *Nano Letters* 2009**,** *9* (2), 684-688.

[349] Souma, S., Ueyama, M., Ogawa, M., "Simulation-based design of a strained graphene field effect transistor incorporating the pseudo magnetic field effect," *Applied Physics Letters* 2014**,** *104* (21), 213505.

[350] Quentin, W., Salim, B., David, T., Nguyen, V. H., Gwendal, F., Jean-Marc, B., Philippe, D., Bernard, P., "A Klein-tunneling transistor with ballistic graphene," *2D Materials* 2014**,** *1* (1), 011006.

[351] Novoselov, K. S., Jiang, D., Schedin, F., Booth, T. J., Khotkevich, V. V., Morozov, S. V., Geim, A. K., "Two-dimensional atomic crystals," *Proceedings of the National Academy of Sciences of the United States of America* 2005**,** *102* (30), 10451-10453.

[352] Kuc, A., Zibouche, N., Heine, T., "Influence of quantum confinement on the electronic structure of the transition metal sulfide TS_2," *Physical Review B* 2011**,** *83* (24), 245213.

353 Kang, J., Tongay, S., Zhou, J., Li, J., Wu, J., "Band offsets and heterostructures of two-dimensional semiconductors," *Applied Physics Letters* 2013**,** *102* (1), 012111.

354 Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., Kis, A., "Single-layer MoS2 transistors," *Nature Nanotechnology* 2011**,** *6*, 147.

355 Radisavljevic, B., Kis, A., "Mobility engineering and a metal–insulator transition in monolayer MoS2," *Nature Materials* 2013**,** *12*, 815.

356 Das, S., Appenzeller, J., "Where Does the Current Flow in Two-Dimensional Layered Systems?" *Nano Letters* 2013**,** *13* (7), 3396-3402.

357 Li, S.-L., Wakabayashi, K., Xu, Y., Nakaharai, S., Komatsu, K., Li, W.-W., Lin, Y.-F., Aparecido-Ferreira, A., Tsukagoshi, K., "Thickness-Dependent Interfacial Coulomb Scattering in Atomically Thin Field-Effect Transistors," *Nano Letters* 2013**,** *13* (8), 3546-3552.

358 Bao, W., Cai, X., Kim, D., Sridhara, K., Fuhrer, M. S., "High mobility ambipolar MoS2 field-effect transistors: Substrate and dielectric effects," *Applied Physics Letters* 2013**,** *102* (4), 042104.

359 Fang, H., Chuang, S., Chang, T. C., Takei, K., Takahashi, T., Javey, A., "High-Performance Single Layered WSe2 p-FETs with Chemically Doped Contacts," *Nano Letters* 2012**,** *12* (7), 3788-3792.

360 Desai, S. B., Madhvapathy, S. R., Sachid, A. B., Llinas, J. P., Wang, Q., Ahn, G. H., Pitner, G., Kim, M. J., Bokor, J., Hu, C., Wong, H.-S. P., Javey, A., "MoS_2transistors with 1-nanometer gate lengths," *Science 2016***,** *354* (6308), 99-102.

361 Sarkar, D., Xie, X., Liu, W., Cao, W., Kang, J., Gong, Y., Kraemer, S., Ajayan, P. M., Banerjee, K., "A subthermionic tunnel field-effect transistor with an atomically thin channel," *Nature* 2015**,** *526*, 91.

362 Britnell, L., Gorbachev, R. V., Jalil, R., Belle, B. D., Schedin, F., Mishchenko, A., Georgiou, T., Katsnelson, M. I., Eaves, L., Morozov, S. V., Peres, N. M. R., Leist, J., Geim, A. K., Novoselov, K. S., Ponomarenko, L. A., "Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures," *Science* 2012**,** *335* (6071), 947-950.

363 Huang, C., Wu, S., Sanchez, A. M., Peters, J. J. P., Beanland, R., Ross, J. S., Rivera, P., Yao, W., Cobden, D. H., Xu, X., "Lateral heterojunctions within monolayer MoSe2–WSe2 semiconductors. *Nature Materials* 2014**,** *13*, 1096.

364 Yoshida, S., Kobayashi, Y., Sakurada, R., Mori, S., Miyata, Y., Mogi, H., Koyama, T., Takeuchi, O., Shigekawa, H., Microscopic basis for the band engineering of Mo1−xWxS2-based heterojunction," *Scientific Reports* 2015**,** *5*, 14808.

365 Gong, Y., Lin, J., Wang, X., Shi, G., Lei, S., Lin, Z., Zou, X., Ye, G., Vajtai, R., Yakobson, B. I., Terrones, H., Terrones, M., Tay, Beng K., Lou, J., Pantelides, S. T., Liu, Z., Zhou, W., Ajayan, P. M., "Vertical and in-plane heterostructures from WS2/MoS2 monolayers," *Nature Materials* 2014**,** *13*, 1135.

366 Li, M.-Y., Shi, Y., Cheng, C.-C., Lu, L.-S., Lin, Y.-C., Tang, H.-L., Tsai, M.-L., Chu, C.-W., Wei, K.-H., He, J.-H., Chang, W.-H., Suenaga, K., Li, L.-J., "Epitaxial growth of a monolayer WSe2-MoS2 lateral p-n junction with an atomically sharp interface," *Science* 2015**,** *349* (6247), 524-528.

367 Mak, K. F., He, K., Shan, J., Heinz, T. F., "Control of valley polarization in monolayer MoS2 by optical helicity," *Nature Nanotechnology* 2012**,** *7*, 494.

368 Xiao, D., Liu, G.-B., Feng, W., Xu, X., Yao, W., "Coupled Spin and Valley Physics in Monolayers of MoS2 and Other Group-VI Dichalcogenides," *Physical Review Letters* 2012**,** *108* (19), 196802.

369 Koma, A., Sunouchi, K., Miyajima, T., "Fabrication and characterization of heterostructures with subnanometer thickness," *Microelectronic Engineering* 1984**,** *2* (1), 129-136.

370 Lee, Y.-H., Yu, L., Wang, H., Fang, W., Ling, X., Shi, Y., Lin, C.-T., Huang, J.-K., Chang, M.-T., Chang, C.-S., Dresselhaus, M., Palacios, T., Li, L.-J., Kong, J., "Synthesis and Transfer of Single-Layer Transition Metal Disulfides on Diverse Surfaces," *Nano Letters* 2013**,** *13* (4), 1852-1857.

371 Lee, Y.-H., Zhang, X.-Q., Zhang, W., Chang, M.-T., Lin, C.-T., Chang, K.-D., Yu, Y.-C., Wang, J. T.-W., Chang, C.-S., Li, L.-J., Lin, T.-W., "Synthesis of Large-Area MoS2 Atomic Layers with Chemical Vapor Deposition," *Advanced Materials* 2012**,** *24* (17), 2320-2325.

372 van der Zande, A. M., Huang, P. Y., Chenet, D. A., Berkelbach, T. C., You, Y., Lee, G.-H., Heinz, T. F., Reichman, D. R., Muller, D. A., Hone, J. C., "Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide," *Nature Materials* 2013**,** *12*, 554.

373 Shaw, J. C., Zhou, H., Chen, Y., Weiss, N. O., Liu, Y., Huang, Y., Duan, X., "Chemical vapor deposition growth of monolayer MoSe2 nanosheets," *Nano Research* 2014**,** *7* (4), 511-517.

374 Cong, C., Shang, J., Wu, X., Cao, B., Peimyoo, N., Qiu, C., Sun, L., Yu, T., "Synthesis and Optical Properties of Large-Area Single-Crystalline 2D Semiconductor WS2 Monolayer from Chemical Vapor Deposition," *Advanced Optical Materials* 2014**,** *2* (2), 131-136.

375 Lin, Y.-C., Lu, N., Perea-Lopez, N., Li, J., Lin, Z., Peng, X., Lee, C. H., Sun, C., Calderin, L., Browning, P. N., Bresnehan, M. S., Kim, M. J., Mayer, T. S., Terrones, M., Robinson, J. A., "Direct Synthesis of van der Waals Solids," *ACS Nano* 2014**,** *8* (4), 3715-3723.

376 Kang, K., Xie, S., Huang, L., Han, Y., Huang, P. Y., Mak, K. F., Kim, C.-J., Muller, D., Park, J., "High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity," *Nature* 2015**,** *520*, 656.

377 Zeng, L., Tao, L., Tang, C., Zhou, B., Long, H., Chai, Y., Lau, S. P., Tsang, Y. H., "High-responsivity UV-Vis Photodetector Based on Transferable WS2 Film Deposited by Magnetron Sputtering," *Scientific Reports* 2016**,** *6*, 20343.

378 Takumi, O., Kohei, S., Seiya, I., Naomi, S., Shimpei, Y., Kentaro, M., Kuniyuki, K., Nobuyuki, S., Akira, N., Yoshinori, K., Kenji, N., Kazuo, T., Hiroshi, I., Atsushi, O., Hitoshi, W., "Multi-layered MoS 2 film formed by high-temperature sputtering for enhancement-mode nMOSFETs," *Japanese Journal of Applied Physics* 2015**,** *54* (4S), 04DN08.

379 Hayashi, K., Kataoka, M., Jippo, H., Ohfuchi, M., Iwai, T., Sato, S., "Two-dimensional SnS_2 for detecting gases causing Sick Building Syndrome," *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017, 2017, pp 18.6.1-18.6.4.

380 J. del Alamo, "Nanometre-Scale Electronics with III-V Compound Semiconductors," *Nature*, vol. 479, 317, 2011.

381 A. Nainani, D. Kim, T. Krishnamohan, and K. Saraswat, "Hole Mobility and Its Enhancement with Strain for Technologically Relevant III-V Semiconductors," *SISPAD*, 2009.

382 B. Bennett, M. Ancona, J. Boos, and B. Shanabrook, "Mobility Enhancement in Strained p-InGaSb Quantum Wells," *Appl. Phys. Lett.*, vol. 91, 042104, 2007.

383 A. Nainai, S. Raghunathan, D. Witte, M. Kobayashi, T. Irisawa, T. Krishnamohan, and K. Saraswat, "Engineering of Strained III-V Heterostructures for High Hole Mobility," *IEDM Tech Dig*., 857, 2009.,

384 B. Bennett, T. Chick, M. Ancona, and J. Boos, "Enhanced Hole Mobility and Density in GaSb Quantum Wells," *Solid State Electron*., 79, 274, 2013.

385 L. Xia, J. B. Boos, B. R. Bennett, M. G. Ancona, and J. A. del Alamo, "Hole Mobility Enhancement in InGaSb Quantum-Well Field-Effect Transistors," *Appl. Phys. Lett.*, vol. 98, 053505, 2011.

386 M. Radosavljevic, *et al*, "High-Performance 40-nm Gate Length InSb P-Channel Compressively Strained Quantum Well Field Effect Transistors for Low-Power (Vcc=0.5V) Logic Applications," *IEDM Tech. Dig*., 727, 2008.

387 Z. Yuan, A. Nainani, B. Bennett, J. Boos, M. Ancona, and K. Saraswat, "Heterostructure Design and Demonstration of InGaSb Channel III-V CMOS Transistors," *ISDRS* 2011, 2011.

388 K. Takei, *et al*. "Nanoscale InGaSb Heterostructure Membranes on Si Substrates for High Hole Mobility Transistors," *Nano Lett.*, 12, 2060, 2012.

389 A. W. Dey, J. Svensson, B. M. Borg, M. Ek, and L.-E. Wernersson, "Single InAs/GaSb nanowire low-power CMOS inverter," *Nano Lett*. 12, 5593, 2012.

390 M. Yokoyama, H. Yokoyama, M. Takenaka, and S. Takagi, "Ultrathin body GaSb-on-insulator p-channel metal-oxide-semiconductor field-effect transistors on Si fabricated by direct wafer bonding," *Appl.Phys. Lett*.,vol. 106, 073503 2015.

391 J. Nah, H. Fang, C. Wang, K. Takei, M. Lee, E. Pils, S. Krishna, and A. Javey, "III-V Complementary Metal-Oxide-Semiconductor Electronics on Silicon Substrates," *Nano Lett.*, 12, 3592, 2012.

392 A. W. Dey, B. M. Borg, B. Ganjipour, M. Ek, K. Dic k, E. Lind, C. Thelander, and L.-E. Wernersson,"High-Current GaSb/InAs(Sb) Nanowire Tunnel Field-Effect Transistors,," *IEEE Electron Dev. Lett.*, vol. 34, p.211, 2013

393 S.-L. Chen, P. B. Griffin, and J. D. Plummer, "Single-Crystal GaAs and GaSb on Insulator on Bulk Si Substrates Based on Rapid Melt Growth," *IEEE Electron Dev. Lett*., vol. 31 p.597, 2010.

394 Z. Yuan, A. Kumar, C.-Y. Chen, A. Nainani, P. Griffin, A. Wang, W. Wang, M.-H. Wong, R.Droopad, R. Contreras-Guerrero, P. Kirsch, R. Jammy, J. Plummer and K. C. Saraswat, "Optimal device architecture and hetero-integration scheme for III-V CMOS," *Symp. VLSI Tech. Dig*., 2013, T54, 2013.

395 A. Nainani, *et al*, "Development of High-κ Dielectric for Antimonides and a Sub 350ºC III-V pMOSFET Outperforming Germanium," *IEDM Tech. Dig*., 138, 2010.

396 M. Xu, R. Wang, and P. Ye, "GaSb Inversion-Mode PMOSFETs With Atomic-Layer-Deposited Al_2O_3 as Gate Dielectric," *IEEE Electron Dev. Lett.*, vol. 32, p. 883, 2011

397 M. Ancona, B. Bennet, and J. Boos, "Scaling projections for Sb-based p-channel FETs," *Solid State Electron.*, vol. 54, 1349, 2010.

398 Y.-J. Yang, W. S. Ho, C.-F. Huang, S. T. Chang and C. W. Liu, "Electron mobility enhancement in strained-germanium n-channel metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 91, 102103, 2007.

399 Shin-ichi Takagi and Satoshi Sugahara, "Comparative Study on Influence of Subband Structures on Electrical Characteristics of III-V Semiconductor, Ge and Si Channel n-MISFETs," *Extended Abstracts of SSDM*, 2006, p.1056.

400 M.J.H. van Dal, B. Duriez, G. Vellianitis, G. Doornbos, R. Oxland, M. Holland, A. Afzalian, Y.C. See, M. Passlack, C.H. Diaz, "Ge n-channel FinFET with optimized gate stack and contacts," *IEDM Tech. Dig.* , p.235, 2014.

401 J. Mitard, L. Witters, H. Arimura, Y. Sasaki, A.P. Milenin, R. Loo, A. Hikavyy, G. Eneman, P. Lagrain, H. Mertens, S. Sioncke, C. Vrancken, H. Bender, K. Barla, N. Horiguchi, A. Mocuta, N. Collaert, A.V-Y. Thean, "First Demonstration of 15nm-WFIN Inversion-Mode Relaxed-Germanium n-FinFETs with Si-cap Free RMG and NiSiGe Source/Drain," *IEDM Tech. Dig.*, p.418, 2014.

402 Yuuichi Kamimuta, Yoshihiko Moriyama, Keiji Ikeda, Minoru Oda and Tsutomu Tezuka, "Current Drive Enhancement of Strained Ge nMISFET with SiGe Stressors by Uniaxial Tensile Stress," *Extended Abstracts of SSDM*, 2011, p.835.

403 G. Eneman, D.P. Brunco, L. Witters, B. Vincent, P. Favia, A. Hikavyy, A. De Keersgieter, J. Mitard, R. Loo, A. Veloso, O. Richard, H. Bender, S.H. Lee, M. Van Dal, N. Kabir, W. Vandervorst, M. Caymax, N. Horiguchi, N. Collaert, A. Thean, "Stress Simulations for Optimal Mobility Group IV p- and nMOS FinFETs for the 14 nm Node and Beyond," *IEDM Tech. Dig.*, 2012, p.131.

404 Heng Wu, Nathan Conrad, Wei Luo, and Peide D. Ye, "First Experimental Demonstration of Ge CMOS Circuits," *IEDM Tech. Dig.* p.227, 2014.

405 Y. Kamata, M. Koike, E. Kurosawa, M. Kurosawa, H. Ota, O. Nakatsuka, S. Zaima and T. Tezuka, "Operations of CMOS Inverter and Ring Oscillator Composed of Ultra-Thin Body Poly-Ge p- and n-MISFETs for Stacked Channel 3D-IC," *Extended Abstracts of the 2014 International Conference on Solid State Devices and Materials, Tsukuba*, pp668-669, 2014.

406 Vita Pi-Ho Hu, Ming-Long Fan, Pin Su and Ching-Te Chuang, "Comprehensive Analysis of UTB GeOI Logic Circuits and 6T SRAM Cells considering Variability and Temperature Sensitivity," *IEDM Tech. Dig.*, 2011, p.753.

407 S. Gupta, et al, "Towards High Mobility GeSn Channel nMOSFETs: Improved Surface Passivation Using Novel Ozone Oxidation Method," *IEDM Tech. Dig.*, 2012, p.375.

408 S. Gupta, et al, "GeSn Channel nMOSFETs: Material Potential and Technological Outlook," *Symposium on VLSI Technology Digest of Technical Papers*, 2012, p.95.

409 C. H. Lee, C. Lu, T. Tabata, T. Nishimura, K. Nagashio and A. Toriumi, "Enhancement of High-Ns Electron Mobility in Sub-nm EOT Ge n-MOSFETs," *Symposium on VLSI Technology Digest of Technical Papers*, 2013, p.T28.

410 R. Zhang, J-C. Lin, X. Yu, M. Takenaka and S. Takagi, Examination of Physical Origins Limiting Effective Mobility of Ge MOSFETs and the Improvement by Atomic Deuterium Annealing, *Symposium on VLSI Technology Digest of Technical Papers*, 2013, T26.

411 Cheng-Ming Lin, et al, "Interfacial layer-free ZrO2 on Ge with 0.39-nm EOT, κ~43,~2×10-3 A/cm2 gate leakage, SS =85 mV/dec, Ion/Ioff =6×105, and high strain response," *IEDM Tech. Dig.*, 2012, p.509.

412 P. Paramahans, S. Gupta, R. K. Mishra, N. Agarwal, A. Nainani, Y. Huang, M.C. Abraham, S. Kapadia, U. Ganguly and S. Lodha, ZnO: an attractive option for n-type metal-interfacial layer-semiconductor (Si, Ge, SiC) contacts, *Symposium on VLSI Technology Digest of Technical Papers*, 2012, p.83.

413 G. Thareja, S. Chopra, B. Adams, Y. Kim, S. Moffatt and K. Saraswat, "High n-Type Antimony Doping Activation in Germanium Using Laser Annealing for n+/p Junction Diode," *IEEE Electron Device Lett.*, vol. 32, 2011, p.838.

414 Masahiro Koike, Yuuichi Kamimuta, Etsuo Kurosawa, and Tsutomu Tezuka, "NiGe/n+-Ge junctions with ultralow contact resistivity formed by two-step P-ion implantation," *Applied Physics Express* **7**, 051302 (2014).

415 H. Lu and A. Seabaugh, "Tunnel Field-Effect Transistors: State-of-the-Art," *Electron Devices Society, IEEE Journal of the*, vol. 2, pp. 44-49, 2014.

416 S. Agarwal and E. Yablonovitch, "Designing a Low Voltage, High Current Tunneling Transistor," *CMOS and Beyond Logic Switches for Terascale Integrated Circuits*, T.-J. King-Liu and K. Kuhn, Eds., ed Cambridge University Press, 2015.

417 A. C. Seabaugh and Z. Qin, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, pp. 2095-2110, 2010.

418 S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-Source Tunnel Field Effect Transistors with Record High I_{ON}/I_{OFF}," presented at the *2009 Symposium on VLSI Technology*, Kyoto, Japan, 2009.

419 W. Y. Choi, B. G. Park, J. D. Lee, and T. J. K. Liu, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Letters*, vol. 28, pp. 743-745, Aug 2007.

420 K. Jeon, W.-Y. Loh, P. Patel, C. Y. Kang, O. Jungwoo, A. Bowonder, *et al.*, "Si Tunnel Transistors With a Novel Silicided Source and 46mV/dec Swing," presented at the *2010 IEEE Symposium on VLSI Technology*, Honolulu, Hawaii, 2010.

421 T. Krishnamohan, K. Donghyun, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) with Record High Drive Currents and < 60 mV/dec Subthreshold Slope," presented at the *IEDM 2008. IEEE International Electron Devices Meeting*, San Francisco, CA,, 2008.

422 U. E. Avci and I. A. Young, "Heterojunction TFET Scaling and resonant-TFET for steep subthreshold slope at sub-9nm gate-length," *Electron Devices Meeting (IEDM), 2013 IEEE International*, 2013, pp. 4.3.1–4.3.4.

423 E. Memisevic, E. Lind, M. Hellenbrand, J. Svensson and L. E. Wernersson, "Impact of Band-Tails on the Subthreshold Swing of III-V Tunnel Field-Effect Transistor," *IEEE Electron Device Letters*, vol. 38, no. 12, pp. 1661-1664, Dec. 2017.

424 E. Memisevic, J. Svensson, M. Hellenbrand, E. Lind, and L. E. Wernersson, "Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with S = 48 mV/decade and Ion = 10 μA/μm for Ioff = 1 nA/μm at Vds = 0.3 V," *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 19.1.1–19.1.4.

425 D. Leonelli, A. Vandooren, R. Rooyackers, A. S. Verhulst, S. D. Gendt, M. M. Heyns, *et al.*, "Performance Enhancement in Multi Gate Tunneling Field Effect Transistors by Scaling the Fin-Width," *Japanese Journal of Applied Physics*, vol. 49, p. 04DC10, 2010.

426 R. Gandhi, Z. Chen, N. Singh, K. Banerjee, and S. Lee, "Vertical Si-Nanowire-Type Tunneling FETs With Low Subthreshold Swing (<50mV/decade) at Room Temperature," *Electron Device Letters, IEEE*, vol. 32, pp. 437–439, 2011.

427 R. Gandhi, C. Zhixian, N. Singh, K. Banerjee, and L. Sungjoo, "CMOS-Compatible Vertical-Silicon-Nanowire Gate-All-Around p-Type Tunneling FETs With <50 mV/decade Subthreshold Swing," *Electron Device Letters, IEEE*, vol. 32, pp. 1504–1506, 2011.

428 H. Qianqian, H. Ru, Z. Zhan, Q. Yingxin, J. Wenzhe, W. Chunlei, *et al.*, "A novel Si tunnel FET with 36mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration," *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 8.5.1–8.5.4.

429 L. Knoll, Z. Qing-Tai, A. Nichau, S. Trellenkamp, S. Richter, A. Schafer, *et al.*, "Inverters With Strained Si Nanowire Complementary Tunnel Field-Effect Transistors," *Electron Device Letters, IEEE*, vol. 34, pp. 813-815, 2013.

430 A. Villalon, C. Le Royer, M. Casse, D. Cooper, B. Previtali, C. Tabone, et al., "Strained tunnel FETs with record Ion: first demonstration of ETSOI TFETs with SiGe channel and RSD," *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 49–50.

431 K. Sung Hwan, K. Hei, H. Chenming, and L. Tsu-Jae King, "Germanium-source tunnel field effect transistors with record high Ion/Ioff," *VLSI Technology, 2009 Symposium on, 2009*, pp. 178–179.

432 T. Krishnamohan, K. Donghyun, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and <<60mV/dec subthreshold slope," *Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 2008*, pp. 1–3.

433 B. Ganjipour, J. Wallentin, M. T. Borgström, L. Samuelson, and C. Thelander, "Tunnel Field-Effect Transistors Based on InP-GaAs Heterostructure Nanowires," *ACS Nano*, vol. 6, pp. 3109–3113, 2012/04/24 2012.

434 K. Tomioka, M. Yoshimura, and T. Fukui, "Steep-slope tunnel field-effect transistors using III-V nanowire/Si heterojunction," *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 47–48.

435 W. G. Vandenberghe, A. S. Verhulst, B. Sorée, W. Magnus, G. Groeseneken, Q. Smets, *et al.*, "Figure of merit for and identification of sub-60 mV/decade devices," *Applied Physics Letters*, vol. 102, p. 013510, 2013.

436 J. Knoch, S. Mantl, and J. Appenzeller, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics*, vol. 51, pp. 572–578, 4// 2007.

437 S. Johnson and T. Tiedje, "Temperature dependence of the Urbach edge in GaAs," *Journal of Applied Physics*, vol. 78, pp. 5609-5613, 1995.

438 S. Agarwal and E. Yablonovitch, "Band-Edge Steepness Obtained From Esaki/Backward Diode Current-Voltage Characteristics," *Electron Devices, IEEE Transactions on*, vol. 61, pp. 1488–1493, 2014.

439 U. E. Avci *et al.*, "Study of TFET non-ideality effects for determination of geometry and defect density requirements for sub-60mV/dec Ge TFET," *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, p. 34.5.1–34.5.4.

440 R. N. Sajjad, W. Chern, J. L. Hoyt, and D. A. Antoniadis, "Trap Assisted Tunneling and Its Effect on Subthreshold Swing of Tunnel FETs," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4380–4387, Nov. 2016.

441 M. Pala and D. Esseni, "(Invited) Modeling the Influence of Interface Traps on the Transfer Characteristics of InAs Tunnel-FETs and MOSFETs," *Meet. Abstr.*, vol. MA2014-01, no. 36, pp. 1372–1372, Apr. 2014.

442 J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, "Auger generation as an intrinsic limit to tunneling field-effect transistor performance," *J. Appl. Phys.*, vol. 120, no. 8, p. 084507, Aug. 2016.

443 ITRS 2013 edition.

444 S. Datta and B. Das, "Electronic analog of the electro-optic modulator," *Appl. Phys. Lett.* vol. 56, pp.665–667, 1990.

445 S. Sugahara and M. Tanaka, "A spin metal-oxide-semiconductor field-effect transistor using half-metallic-ferromagnet contacts for the source and drain," *Appl. Phys. Lett.* vol. 84, pp.2307–2309, 2004.

446 S. Sugahara and J. Nitta, "Spin-Transistor Electronics: An Overview and Outlook," *Proc. IEEE*, vol. 98, pp. 2124–2154, 2010.

447 T. Tanamoto, H.Sugiyama, T.Inokuchi, T. Marukame, M. Ishikawa, K. Ikegami, Y. Saito, "Scalability of spin field programmable gate arrary: A reconfigurable architecture based on spin metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.* vol. 109, pp. 07C312/1-3, 2011.

448 S. Shuto, S. Yamamoto, and S. Sugahara, "Nonvolatile Static Random Access memory based on spin-transistor architecture," *J. Appl. Phys.* vol. 105, pp. 07C933/1-3, 2009.

449 S. Yamamoto, and S. Sughara, "Nonvolatile Delay Flip-Flop Based on Spin-Transistor Architecture and Its Power-Gating Applications," *Jpn. J. Appl. Phys.* vol. 49, pp. 090204/1-3, 2010.

450 S. Sugahara, Y. Shuto, and S. Yamamoto, "Nonvolatile Logic Systems Based on CMOS/Spintronics Hybrid Technology: An Overview," *Magnetics Japan*, vol. 6, pp. 5–15, 2011.

451 Y. Shuto, S. Yamamoto, H. Sukegawa, Z.C. Wen, R. Nakane, S. Mitani, M. Tanaka, K. Inomata, S. Sugahara, "Design and performance of pseudo-spin-MOSFETs using nano-CMOS devices,," *2012 IEEE International Electron Devices Meeting*, pp. 29.6.1–29.6.4, 2012.

452 S. Yamamoto, Y. Shuto, S. Sugahara, "Nonvolatile flip-flop based on pseud-spin transistor architecture and its nonvolatile power-gating applications for low-power CMOS logic," *EPJ Appl. Phys.* vol. 63, pp. 14403, 2013.

453 D. Osintev, V. Sverdlov, Z. Stanojevic, A. Makarov, S. Selberherr, "Temperature dependence of the transport properties of spin field-effect transistors built with InAs and Si channels,," *Solid-state Electronics* vol. 71, pp. 25–29, 2012.

454 Y. Saito, T. Marukame, T. Inokuchi, M. Ishikawa, H. Sugiyama, T. Tanamoto, "Spin injection, transport and read/write operation in spin-based MOSFET," *Thin Solid Films* vol. 519, pp. 8266 – 8273, 2011.

455 Y. Saito, T. Inokuchi, M. Ishikawa, H. Sugiyama, T. Marukame, T. Tanamoto, "Spin-based MOSFET and Its Applications," *J. Elec. chem. Soc.* vol. 158, pp. H1068 – H1076, 2011.

456 T. Sasaki, Y. Ando, M. Kameno, T. Tahara, H. Koike, T. Oikawa, T. Suzuki, M. Shiraishi, "Spin transport in nondegenerate Si with a spin MOSFET structure at room temperature," *Phys. Rev. Appl.* Vol. 2, pp.034005/1-6, 2014.

[457] T. Tahara, H. Koike, M. Kameno, T. Sasaki, Y. Ando, K. Tanaka, S. Miwa, Y. Suzuki, M. Shiraishi, "Room-temperature operation od Si spin MOSFET with high on/off spin signal ratio," *Appl. Phys. Express* vol. 8, pp. 113004/1-3, 2015.

[458] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, B. J. van Wees, "Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor," *Phys. Rev.* B vol. 62, pp. R4790–R4793, 2000.

[459] E. I. Rashba, "Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem," *Phys. Rev.* B vol. 62, pp. R16267–R16270, 2000.

[460] A. Fert, H. Jaffr`es, "Conditions for efficient spin injection from a ferromagnetic metal into a semiconductor," *Phys. Rev.* B vol. 64, pp. 184420–1844209, 2001.

[461] T. Tanamoto, H. Sugiyama, T. Inokuchi, M. Ishikawa, Y. Saito, "Effects of interface resistance asymmetry on local and non-local magnetoresistance structures," *Jap. J. Appl. Phys.* Vol. 52, pp. 04CM03/1-4, 2013.

[462] I. Appelbaum, B. Huang, and D. J. Monsma, "Electronic measurement and control of spin transport in silicon," *Nature*, vol. 447, pp. 295–298, 2007.

[463] S. P. Dash, S. Sharma, R. S. Patel, M. P. Jong, and R. Jansen, "Electrical creation of spin polarization in silicon at room temperature," *Nature*, vol. 462, pp. 491–494, 2009.

[464] T. Suzuki, T. Sasaki, T. Oikawa, M. Shiraishi, Y. Suzuki, and K. Noguchi, "Room-temperature electron spin transport in a highly doped Si channel," *Appl. Phys. Express* vol. 4, pp. 023003–023005, 2011.

[465] C. H. Li, O. M. J. van't Erve, and B. T. Jonker, "Electrical injection and detection of spin accumulation in silicon at 500 K with magnetic metal/silicon dioxide contacts," *Nature Commun.* vol. 2, pp. 245, 2011.

[466] K. R. Jeon, B. C. Min, I. J. Shin, C. Y. Park, H. S. Lee, Y. H. Jo, S. C. Shin, "Electrical spin accumulation with improved bias voltage dependence in crystalline CoFe/MgO/Si system," *Appl. Phys. Lett.* vol. 98, pp. 262102–262104, 2011.

[467] M. Ishikawa, H. Sugiyama, T. Inokuchi, K. Hamaya, Y. Saito, "Effect of the interface resistance of CoFe/MgO contacts on spin accumulation in silicon," *Appl. Phys. Lett.* vol. 100, pp. 252404–252406, 2012.

[468] M. Ishikawa, H. Sugiyama, T. Inokuchi, T. Tanamoto, K. Hamaya, N. Tezuka, and Y. Saito, "Maximum magnitude in bias-dependent spin accumulation signals of CoFe/MgO/Si on insulator devices," *J. Appl. Phys.* vol. 114, pp. 243904/1–6, 2013.

[469] T. Inokuchi, T. Marukame, M. Ishikawa, H. Sugiyama, Y. Saito, "Electrical spin injection into n-GaAs channels and detection through MgO/CoFeB electrodes," *Appl. Phys. Express* vol. 2, pp. 023006–023008, 2009.

[470] Y. Saito, M. Ishikawa, H. Sugiyama, T. Inokuchi, K. Hamaya, N. Tezuka, "Correlation between amplitude of spin accumulation signals investigated by Hanle effect measurement and effective junction barrier height in CoFe/MgO/n+-Si junctions," *J. Appl. Phys.* vol. 117, pp. 17C707/1-4, 2015.

[471] Y. Ando, K. Kasahara, S. Yamada, Y. Maeda, K. Masaki, Y. Hoshi, K. Sawano, M. Miyao, K. Hamaya, "Temperature evolution of spin accumulation detected electrically in a nondegenerated silicon channel," *Phys. Rev.* B vol. 85, pp. 035320/1-5, 2012.

[472] S Iba, H. Saito, A. Spiesser, S. Watanabe, R. Jansen, S. Yuasa, K. Ando, "Spin accumulation and spin lifetime in p-type germanium at room temperature," *Applied Physics Express* vol. 5, pp. 053004–053006, 2012.

[473] K.-R. Jeon, B.-C. Min, Y.-H. Park, S.-Y. Park, S.-C. Shin, "Electrical investigation of the oblique Hanle effect in ferromagnet/oxide/semiconductor contacts," *Phys. Rev.* B vol. 87, pp. 195311/1-10, 2013.

[474] K. Hamaya, G. Takemoto, Y. Baba, K. Kasahara, S. Yamada, K. Sawano, M. Miyao, "Room temperature electrical creation of spin accumulation in n-Ge using highly resistive Fe3Si/n+-Ge Schottky-tunnel contacts," *Thin Solid Films* vol. 557, pp. 382–385, 2014.

[475] G. Salis, A. Fuhrer, R. R. Schlittler, L. Gross, S. F. Alvarado, "Temperature dependence of the nonlocal voltage in an Fe/GaAs electrical spin-injection device," *Phys. Rev.* B vol. 81, pp. 205323–205327, 2010.

[476] T. Uemura, T. Akiho, M. Harada, K-i. Matsuda, M. Yamamoto, "Non-local detection of spin-polarized electrons at room temperature in Co50Fe50/GaAs Schottky tunnel junctions," *Appl. Phys. Lett.* vol. 99, pp. 082108–082110, 2011.

[477] O.M.J. Van't Erve, A.L. Friedman, E. Cobas, C.H. Li, J.T. Robinson, B.T. Jonker, "Low-resistance spin injection into silicon using graphene tunnel barriers," *Nature Nanotechnology*, vol. 7, pp. 737–742, 2012.

[478] O.M.J. Van't Erve, A.L. Friedman, C.H. Li, J.T. Robinson, J. Connell, L. J. Lauhon, B.T. Jonker, "Spin transport and Hanle effect in silicon nanowires using graphene tunnel barriers," *Nature Communications* vol. 6, pp. 7541/1-8, 2015.

[479] Y. Saito, T. Tanamoto, M. Ishikawa, H. Sugiyama, T. Inokuchi, K. Hamaya, N. Tezuka, "Local magnetoresistance through Si and its bias voltage dependence in ferromagnet/MgO/silicon-on-insulator lateral spin valves," *J. Appl. Phys.* vol. 115, pp. 17C514/1-3, 2014.

[480] T. Sasaki, T. Suzuki, Y. Ando, H. Koike, T. Oikawa, Y. Suzuki, M. Shiraishi, "Local magnetoresistance in Fe/MgO/Si lateral spin valve at room temperature," *Appl. Phys. Lett.* vol. 104, pp.052404/1-4, 2014.

[481] N. Tezuka, N. Ikeda, S. Sugimoto, K. Inomata, "175% tunnel magnetoresistance at room temperature and high thermal stability using $Co_2FeAl_{0.5}Si_{0.5}$ full-Heusler alloy electrodes," *Appl. Phys. Lett.* vol. 89, pp. 252508/1-3, 2006.

[482] K. Inomata, M. Wojcik, E. Jedryka, N. Tezuka, "Site disorder in Co_2 Fe (Al,Si) Heusler alloys and its influence on junction tunnel magnetoresistance," *Phys. Rev.* B vol. 77, pp. 214425/1-9 , 2008.

[483] M. Yamamoto, T. Ishikawa, T. Taira, G.-F. Li, K. -I. Matsuda, T. Uemura, "Effect of defects in Heusler alloy thin films on **spin**-dependent tunneling characteristics of Co_2MnSi/MgO/Co_2MnSi and Co_2MnGe/MgO/Co_2MnGe magnetic tunnel junctions," *Journal of Physics: Condensed Matter* vol. 22, pp. 164212/1-9, 2010.

484 Y. Sakuraba, M. Hattori, M. Oogane, Y. Ando, H. Kato, A. Sakuma, T. Miyazaki, H. Kubota, “Giant tunneling magnetoresistance in Co_2MnSi/Al–O/Co_2MnSi magnetic tunnel junctions,” *Appli. Phys. Lett.* vol. 88, pp. 192508/1-3, 2008.

485 G. H. Fecher, C. Felser, “High energy, high resolution photoelectron spectroscopy of $Co_2Mn_{1-x}Fe_xSi$,” *J. Phys. D: Appl. Phys.* vol. 40, pp. 1576–1581, 2007.

486 T. Saito, N. Tezuka, M. Matsuura, S. Sugimoto, “Spin injection, transport, and detection at room temperature in a lateral spin transport device with $Co_2FeAl_{0.5}Si_{0.5}$/n-GaAs Schottky tunnel junctions,” *Appl. Phys. Express*, vol. 6, pp. 103006/1-4, 2013.

487 K. Kasahara, Y. Fujita, S. Yamada, K. Sawano, M. Miyao, K. Hamaya, “Greatly enhanced generation efficiency of pure spin currents in Ge using Heusler compound Co_2FeSi electrodes,” *Appl. Phys. Express*, vol. 7, pp. 033002/1-4, 2014.

488 Y. Ebina, T. Akiho, H. –X. Liu, M. Uemura, “Effect of CoFe insertion in Co_2MnSi/CoFe/n-GaAs junctions on spin injection properties,” *Appl. Phys. Lett.*, vol. 104, pp. 172405/1-4, 2014.

489 M. Ishikawa, H. Sugiyama, T. Inokuchi, K. Hamaya, Y. Saito, “Spin transport and accumulation in n+-Si using Heusler compound Co_2FeSi/MgO tunnel contacts,” *Appl. Phys. Lett.*, vol. 107, pp. 092402/1-5, 2015.

490 Y. Shuto, R. Nakane, W. Wang, H. Sukegawa, S. Yamamoto, M. Tanaka, K. Inomata, and S. Sugahara, “A New Spin-Functional Metal–Oxide–Semiconductor Field-Effect Transistor Based on Magnetic Tunnel Junction Technology: Pseudo-Spin-MOSFET,” *Appl. Phys. Express*, vol. 3, pp. 013003/1-3, 2010.

491 R. Nakane, Y. Shuto, H. Sukegawa, Z. C. Wen, S. Yamamoto, S. Mitani, M. Tanaka, K. Inomata, S. Sugahara, “Fabrication of pseudo-spin-MOSFETs using a multi-project wafer CMOS chip,” *Solid-State Elec.*, vol. 102, pp. 52–58, 2014.

492 T. Koga, J. Nitta, T. Akazaki, H. Takayanagi, “Rashba Spin-Orbit Coupling Probed by the Weak Antilocalization Analysis in InAlAs/InGaAs/InAlAs Quantum Wells as a Function of Quantum Well Asymmetry,” *Phys. Rev. Lett.* vol. 89, pp. 046801/1-4, 2002.

493 A. G. Mal’shukov, K. A. Chao, “Waveguide diffusion model and showdown of D’yakonov-Perel’ spin relaxation in narrow two dimensional semiconductor channels,” *Phys. Rev.* B vol. 61, pp. R2413–R2416, 2000.

494 S. Kettemann, “Dimensional control of antilocalozation and spin relaxation in quantum wires,” *Phys. Rev. Lett.* vol. 98, pp. 176808/1–4, 2007.

495 A. W. Holleitner, V. Sih, R. C. Myers, A. C. Gossard, D. D. Awschalom, “Suppression of spin relaxation in submicron InGaAs wires,” *Phys. Rev. Lett.* vol. 97, pp. 036805/1–4, 2006.

496 J. Schliemann, J. C. Egues, D. Loss, “Nonballistic spin-field-effect transistor,” *Phys. Rev. Lett.* vol. 90, pp. 1468011–1468014, 2003.

497 B. A. Bernevig, J. Orenstein, S. -C. Zhang, “Extra SU(2) symmetry and persistent spin helix in a spin-orbit coupled system,” *Phys. Rev. Lett.* vol. 97, pp. 236601/1–4, 2006.

498 D. Koralek, C. P. Weber, J. Orenstein, B. A. Bernevig, S. –C. Zhang, S. Mack, D. D. Awschalom, “Emergence of the persistent spin helix in semiconductor quantum wells,” *Nature*, vol. 458, pp. 610–613, 2009.

499 M. Kohda, *et al*, “Gate-controlled persistent spin helix state in (In, Ga)As quantum wells,” *Phys. Rev.* B vol. 86, pp. 081306(R)/1-5, 2012.

500 A. Sasaki, S. Nonaka, Y. Kunihashi, M. Kohda, T. Bauernfeind, T. Dollinger, K. Richter, J. Nitta, “Direct determination of spin-orbit interaction coefficients and realization of the persistent spin helix symmetry,” *Nature Nanotechnology*, vol. 9, pp. 703–709, 2014.

501 N.D. Rizzo, D. Houssameddine, J. Janesky, R. Whig, F.B. Mancoff, M.L. Schneider, M. Deherrera, J.J. Sun, K. Nagel, S. Deshpande, H.-J. Chia, S.M. Alam, T. Andre, S. Aggarwal, J. M. Slaughter, “A fully functional 64 Mb DDR3 ST-MRAM built on 90 nm CMOS technology,” IEEE Trans. Magn. vol. 49, pp. 4441–4446, 2013.

502 S. Salahuddin and S. Datta, “Use of negative capacitance to provide a subthreshold slope lower than 60 mV/decade,” *Nanoletters*, vol. 8, No. 2, 2008.

503 S. Salahuddin and S. Datta, “Can the subthreshold swing in a classical FET be lowered below 60 mV/decade?,” *Proceedings of IEEE Electron Devices Meeting (IEDM)*, 2008.

504 G. A. Salvatore, D. Bouvet, A. M. Ionescu, “Demonstration of Subthrehold Swing Smaller Than 60 mV/decade in Fe-FET with P(VDF-TrFE)/SiO2 Gate Stack,” *IEDM 2008*, San Francisco, USA, 15-17 December 2008.

505 A. Rusu, G. A. Salvatore, D. Jiménez, A. M. Ionescu, "Metal-Ferroelectric-Metal- Oxide-Semiconductor Field Effect Transistor with Sub-60 mV/decade Subthreshold Swing and Internal Voltage Amplification,” *IEDM 2010* , San Francisco, USA, 06-08 December 2010.

506 Asif Islam Khan, Debanjan Bhowmik, Pu Yu, Sung Joo Kim, Xiaoqing Pan, Ramamoorthy Ramesh, Sayeef Salahuddin, “Experimental Evidence of Ferroelectric Negative Capacitance in Nanoscale Heterostructures,” *Applied Physics Letters* 99 (11), 113501-113501-3,2011.

507 A. I. Khan, K. Chatterjee, B. Wang, S. Drapcho, L. You, C. Serrao, S. R. Bakaul, R. Ramesh, and S. Salahuddin, "Negative capacitance in a ferroelectric capacitor," *Nature Materials*, vol. 14, no. 2, pp. 182–186, Jan. 2015.

508 Muller, J., Boscke, T. S., Muller, S., Yurchuk, E., Polakowski, P., Paul, J., ... & Weinreich, W.,“Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories,” *2013 IEEE International Electron Devices Meeting*, December 2013.

509 Cheng, C. H., & Chin, A. Low-Voltage Steep Turn-On pMOSFET Using Ferroelectric High-Gate Dielectric. *Electron Device Letters, IEEE*,*35*(2), 274–276, 2014.

510 M-H Lee et al, “Prospects for Ferroelectric HfZrOx FETs with Experimentally CET=0.98 nm, SSfor=42 mV/dec, SSrev=28 mV/dec, Switch-OFF <0.2V, and Hysteresis-Free Strategies,” *Proceedings of IEDM*, 2015.

511 Kai-Shin Li et al, “Sub-60 mV-Swing Negative-Capacitance FinFET without Hysteresis,” *Proceedings of IEDM*, 2015.

512 Krivokapic et al., “14nm Ferroelectric FinFET Technology with Steep Subthreshold Slope for Ultra Low Power Applications,” *Proc. IEDM*, 2017.

513 K. Akarvardar et al., “Design considerations for complementary nanoelectromechanical logic gates,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 299–302, 2007.

514 F. Chen et al., “Integrated circuit design with NEM relays,” *IEEE/ACM International Conference on Computer-Aided Design*, pp. 750–757, 2008.

515 H. Kam et al., “Design and reliability of a micro-relay technology for zero-standby-power digital logic applications,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 809–811, 2009.

516 S. Natarajan et al., “A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 3.7.1–3.7.3, 2014.

517 N. Xu et al., “Hybrid CMOS/BEOL-NEMS technology for ultra-low-power IC applications,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 28.8.1–28.8.4, 2014.

518 H. Fariborzi et al., "Analysis and demonstration of MEM-relay power gating," *Custom Integrated Circuits Conference*, 2010.

519 C. Chen et al., “Nano-electro-mechanical relays for FPGA routing: experimental demonstration and a design technique,” *Design, Automation & Test in Europe Conference & Exhibition*, 2012.

520 K. Kato, V. Stojanović and T.-J. K. Liu, “Embedded nano-electro-mechanical memory for energy-efficient reconfigurable logic,” *IEEE Electron Device Letters*, vol. 37, no. 12, pp. 1563–1565, 2016.

521 J. O. Lee et al., “A sub-1-volt nanoelectromechanical switching device,” *Nature Nanotechnology*, Vol. 8, pp. 36–40, 2013.

522 R. Nathanael et al., “4-terminal relay technology for complementary logic,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 223–226, 2009.

523 T.-J. K. Liu et al., “Prospects for MEM_relay logic switch technology,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 424–427, 2010.

524 R. Nathanael et al., “Multi-input/multi-output relay design for more compact and versatile implementation of digital logic with zero leakage,” *19th International Symposium on VLSI Technology, Systems and Applications*, 2012.

525 J. Jeon et al., “Multi-input relay design for more compact implementation of digital logic circuits,” *IEEE Electron Device Letters*, Vol. 33, No. 2, pp. 281–283, 2012.

526 F. Chen et al., "Demonstration of integrated micro-electro-mechanical (MEM) switch circuits for VLSI applications," *International Solid State Circuits Conference*, pp. 150–151, 2010.

527 H. Fariborzi et al., “Design and demonstration of micro-electro-mechanical relay multipliers,” *IEEE Asian Solid-State Circuits Conference*, 2011.

528 M. Spencer et al., “Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications,” *IEEE Journal of Solid-State Circuits*, Vol. 46, No. 1, pp. 308–320, 2011.

529 L. Hutin et al., “Electromechanical diode cell scaling for high-density nonvolatile memory,” *IEEE Transactions on Electron Devices*, Vol. 61, pp. 1382–1387, 2014.

530 Y. Chen et al., "Reliability of MEM relays for zero leakage logic," *SPIE Conference Proceedings* Vol. 8614, 2013.

531 H. Kam et al., "Design, optimization and scaling of MEM relays for ultra-low-power digital logic," *IEEE Transactions on Electron Devices*, Vol. 58, No. 1, pp. 236–250, 2011.

532 C. Qian et al., “Energy-delay performance optimization of NEM logic relay,” *IEEE International Electron Devices Meeting Technical Digest*, pp. 18.1.1–18.1.4, 2015.

533 C. Zhou, D. M. Newns, J. A. Misewich and P. C. Pattnaik, “A field effect transistor based on the Mott transition in a molecular layer,” *Appl Phys Lett* 70 (5), 598–600 (1997)

534 D. M. Newns, J. A. Misewich, C. C. Tsuei, A. Gupta, B. A. Scott and A. Schrott, “Mott transition field effect transistor,” *Appl Phys Lett* 73 (6), 780–782 (1998)

535 Y. Zhou and S. Ramanathan, “Correlated electron materials and field effect transistors for logic: a review,” *Crit Rev Solid State Mater Sci*, 38(4), 286–317, 2013.

536 H. Akinaga, “Recent advances and future prospects in functional-oxide nanoelectronics: the emerging materials and novel functionalities that are accelerating semiconductor device research and development,” *Japanese J. Appl. Phys.*, 52(100001), 2013.

537 Z. Yang, C. Ko and S. Ramanathan, “Oxide Electronics Utilizing Ultrafast Metal-Insulator Transitions,” *Annual Review of Materials Research* 41, 8.1 (2011).

538 A. Cavalleri, Cs. Toth, C. W. Siders, J. A. Squier, F. Raksi, P. Forget and J. C. Keiffer, “Femtosecond Structural Dynamics in VO2 during an Ultrafast Solid-Solid Phase Transition,” *Phy. Rev. Lett*. 87, 237401 (2001)

539 S. Hormoz and S. Ramanathan, "Limits on vanadium oxide Mott metal–insulator transition field-effect transistors," *Solid State Electron* 54 (6), 654–659 (2010)

540 H. T. Kim, B. G. Chae, D. H. Youn, S. L. Maeng, G. Kim, K. Y. Kang and Y. S. Lim, "Mechanism and observation of Mott transition in VO2-based two- and three-terminal devices," *New J Phys* 6, 52 (2004)

541 G. Stefanovich, A. Pergament and D. Stefanovich, J., "Electrical switching and Mott transition in VO2," *Phys: Cond Mat*, 12, 8837 (2000)

542 D. Ruzmetov, G. Gopalakrishnan, C. Ko, V. Narayanamurti and S. Ramanathan, "Three-terminal field effect devices utilizing thin film vanadium oxide as the channel layer," *J Appl Phys* 107 (11), 114516 (2010)

543 Z. Yang, Y. Zhou, and S. Ramanathan, J., "Studies on room-temperature electric-field effect in ionic-liquid gated VO2 three-terminal devices," *Appl. Phys*. 111, 014506 (2012)

544 M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa and Y. Tokura, "Collective bulk carrier delocalization driven by electrostatic surface charge accumulation," *Nature* 487, 459–462 (2012)

545 Y. Zhou and S. Ramanathan, "Relaxation dynamics of ionic liquid—VO2 interfaces and influence in electric double-layer transistors," *J. Appl. Phys*. 111, 084508 (2012)

546 H. Ji , Jiang Wei , and D. Natelson, "Modulation of the Electrical Properties of VO2 Nanobeams Using an Ionic Liquid as a Gating Medium," *Nano Lett*., 12, 2988–2992 (2012)

547 J Jeong, N Aetukuri, T Graf, TD Schladt, MG Samant and S. S. Parkin, "Suppression of Metal-Insulator Transition in VO2 by Electric Field–Induced Oxygen Vacancy Formation," *Science*, 339(6126), 1402–1405 (2013).

548 J. Shi, S. D. Ha, Y. Zhou, F. Schoofs and S. Ramanathan, "A correlated nickelate synaptic transistor," *Nature Communications*, 4, 2676 (2013)

549 S. D. Ha, J. Shi, Y. Meroz, L. Mahadevan and S. Ramanathan, "Neuromimetic Circuits with Synaptic Devices Based on Strongly Correlated Electron Systems," *Physical Review Applied*, 2, 064003 (2014),

550 M. D. Pickett, G. Medeiros-Ribeiro and R. Stanley Williams, "A scalable neuristor built with Mott memristors," *Nature Materials* 12, 114–117 (2013)

551 N. Shukla, A. V. Thathachary, A. Agrawal, H. Paik, A. Aziz, D. G. Schlom, S. K.Gupta, R. Engel-Herbert and S. Datta, "A steep-slope transistor based on abrupt electronic phase transition," *Nature Communications* 6, 7812 (2015)

552 Tingting Wei, Teruo Kanki Kouhei Fujiwara, Masashi Chikanari and Hidekazu Tanaka, "Electric field-induced transport modulation in VO2 FETs with high-k oxide/ organic parylene-C hybrid gate dielectric," *Appl. Phys. Lett*. 108 (2016) 053503.

553 Tingting Wei, Teruo Kanki, Masashi Chikanari, T. Uemura, Tsuyoshi Sekitani, and Hidekazu Tanaka, "Enhancement of electronic transport modulation in single crystalline VO2 nanowire-based solid-state field-effect transistor," *Scientific Reports*, in press.

554 Y. Zhou, X. Chen, Z. Yang, C. Mouli and S. Ramanathan, "Voltage-Triggered Ultrafast Phase Transition in Vanadium Dioxide Switches," *IEEE Electron Device Letters*, 34, 220 (2013)

555 J. Leroy, A. Crunteanu, A. Bessaudou, F. Cosset, C. Champeaux, and J. C. Orlianges, "High-speed metal-insulator transition in vanadium dioxide films induced by an electrical pulsed voltage over nano-gap electrodes," *Applied Physics Letters*, 100, 213507 (2012).

556 Y. Zhang and S. Ramanathan, "Analysis of "on" and "off" times for thermally driven VO2 metal-insulator transition nanoscale switching devices," *Solid State Electronics*, August 2011, Vol. 62, pp. 161–164 (2011)

557 S. D. Ha, G. H. Aydogdu and S. Ramanathan, "Metal-insulator transition and electrically driven memristive characteristics of SmNiO3 thin films," *Appl. Phys. Lett*., 98, 012105 (2011)

558 P. Lacorre, J. B. Torrance, J. Pannetier, A. I. Nazzal, P. W. Wang and T. C. Huang, J., "Synthesis, crystal structure, and properties of metallic PrNiO3: Comparison with metallic NdNiO3 and semiconducting SmNiO3," *Sol. St. Chem*. 91, 225 (1991)

559 P.-H. Xiang, S. Asanuma, H. Yamada, H. Sato, I. H. Inoue, H. Akoh, A. Sawa, M. Kawasaki, and Y. Iwasa, "Electrolyte-gated $SmCoO_3$ thin-film transistors exhibiting thickness-dependent large switching ratio at room temperature," *Adv. Mat.*, 25, 2158–2161, 2013.

560 W. L. Lim, E. J. Moon, J. W. Freeland, D. J. Meyers, M. Kareev, J. Chakhalian, and S. Urazhdin, "Field-effect diode based on electron-induced Mott transition in NdNiO3," *Appl. Phys. Lett*. 101, 143111 (2012)

561 S. Asanuma, P.-H. Xiang, H. Yamada, H. Sato, I. H. Inoue, H. Akoh, A. Sawa, K. Ueno, H. Shimotani, H. Yuan, M. Kawasaki, and Y. Iwasa, "Tuning of the metal-insulator transition in electrolyte-gated NdNiO3 thin films," *Appl. Phys. Lett*. 97, 142110 (2010)

562 R. Scherwitzl, P. Zubko, I. G. Lezama, S. Ono, A. F. Morpurgo, G. Catalan and J.-M. Triscone, "Electric-Field Control of the Metal-Insulator Transition in Ultrathin NdNiO3 Films," *Adv. Mater*. 22, 5517 (2010).

563 S. D. Ha, U. Vetter, J. Shi, and S. Ramanathan, "Electrostatic gating of metallic and insulating phases in SmNiO3 ultrathin films," *Appl. Phys. Lett.* 102, 183102 (2013)

564 S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett*. 106, 122102 (2015)

565 S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett*. 106, 122102 (2015)

566 J. Shi, Y. Zhou and S. Ramanathan, “Colossal resistance switching and band gap modulation in a perovskite nickelate by electron doping,” *Nature Communications*, 5, 4860 (2014)

567 A. Khitun and K. Wang, “Nano scale computational architectures with Spin Wave Bus,” *Superlattices & Microstructures* 38, 184–200 (2005).

568 A. Khitun, M. Bao and K. L. Wang, “Spin Wave Magnetic NanoFabric: A New Approach to Spin-Based Logic Circuitry,” *IEEE Transactions on Magnetics* 44 (9), 2141–2152 (2008).

569 M. P. Kostylev, A. A. Serga, T. Schneider, B. Leven, and B. Hillebrands, “Spin-wave logical gates,” *Applied Physics Letters* 87, 153501 (2005).

570 K.-S. Lee and S.-K. Kim, “Conceptual design of spin wave logic gates based on a Mach–Zehnder-type spin wave interferometer for universal logic functions,” *Journal of Applied Physics* 104, 053909 (2008).

571 R. Verba, V. Tiberkevich, I. Krivorotov, and A. Slavin, “Parametric Excitation of Spin Waves by Voltage-Controlled Magnetic Anisotropy,” *Physical Review Applied* 1, 044006 (2014).

572 Y.-J. Chen, H. K. Lee, R. Verba, J. A. Katine, I. Barsukov, V. Tiberkevich, J. Q. Xiao, A. N. Slavin, and I. N. Krivorotov, “Parametric Resonance of Magnetization Excited by Electric Field,” *Nano Letters* 17, 572 (2017).

573 S. Dutta, et al. “Non-volatile Clocked Spin Wave Interconnect for Beyond-CMOS Nanomagnet Pipelines,” *Scientific Reports* 5, 09861 (2015).

574 A. Khitun, M. Bao, J.-Y. Lee, K. L. Wang, D. W. Lee, S. X. Wang and I. V. Roshchin, “Inductively Coupled Circuits with Spin Wave Bus for Information Processing,” *Journal of Nanoelectronics and Optoelectronics* 3 (1), 24–34 (2008).

575 A. Khitun and K. L. Wang, “Non-volatile magnonic logic circuits engineering,” *Journal of Applied Physics* 110 (3) 034306 (2011).

576 S. Klingler, et al. “Design of a spin-wave majority gate employing mode selection,” *Applied Physics Letters* 105, 152410 (2014).

577 A. Khitun, “Magnonic holographic devices for special type data processing,” *Journal of Applied Physics* 113 (16) 164503 (2013).

578 A. Khitun, “Multi-frequency magnonic logic circuits for parallel data processing,” *Journal of Applied Physics* 111 (5) 054307 (2012).

579 Y. Wu, M. Bao, A. Khitun, J.-Y. Kim, A. Hong and K. L. Wang, “A Three-Terminal Spin-Wave Device for Logic Applications,” *Journal of Nanoelectronics and Optoelectronics* 4 (3), 394–397 (2009).

580 S. Cherepov, P. K. Amiri, J. G. Alzate, K. Wong, M. Lewis, P. Upadhyaya, J. Nath, M. Bao, A. Bur, T. Wu, G. P. Carman, A. Khitun and K. L. Wang, “Electric-field-induced spin wave generation using multiferroic magnetoelectric cells,” *Applied Physics Letters* 104 (8) 082403 (2014).

581 F. Gertz, A. Kozhevnikov, Y. Filimonov and A. Khitun, Magnetics, “Magnonic Holographic Memory,” *IEEE Transactions on Magnetics* 51 (4), 4002905 (2015).

582 A. Kozhevnikov, F. Gertz, G. Dudko, Y. Filimonov and A. Khitun, “Pattern recognition with magnonic holographic memory device,” *Applied Physics Letters* 106 (14), 142409 (2015).

583 A. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod, "Majority logic gate for Magnetic Quantum-dot Cellular Automata," (in English), *Science*, vol. 311, no. 5758, pp. 205–208, Jan 13 2006.

584 M. Niemier et al., "Boolean and Non-Boolean Architectures for Out-of-Plane Nanomagnet Logic," *Procedings of the International Workshop on Cellular Nanoscale Networks and their Applications*, pp. 1–6, August 29–31 2012.

585 Á. Papp et al., "Threshold Gate Based Circuits from Nanomagnet Logic," *IEEE T. on Nanotechnology*, 2013.

586 S. Breitkreutz et al., "1-Bit Full Adder in Perpendicular Nanomagnet Logic using a Novel 5-Input Majority Gate," *accepted at the Joint European Magnetic Symposium (JEMS), Rhodos, Greece,* August 25–30 2013.

587 M. Becherer et al., "On-chip Extraordinary Hall-effect sensors for characterization of nanomagnetic logic devices," *Solid State Electronics*, vol. 54, no. 9, pp. 1027–1032, 2010.

588 J. M. Shaw et al., "Reversal mechanisms in perpendicularly magnetized nanostructures," *Physical Review B*, vol. 78, no. 2, p. 024414, 2008.

589 D. Bhowmik, L. You, and S. Salahuddin, "Spin Hall effect clocking of nanomagnetic logic without a magnetic field," *Nat Nano*, Letter vol. 9, no. 1, pp. 59–63, 01//print 2014.

590 G. Csaba, P. Lugli, and W. Porod, "Power dissipation in nanomagnetic logic devices," *IEEE Conference on Nanotechnology*, 2004, pp. 346–348.

591 A. Dingler, M. T. Niemier, X. S. Hu, and E. Lent, "Performance and Energy Impact on Locally Controlled NML Circuits," *ACM Journal on Emerging Technologies in Computing*, vol. 7, no. 1, pp. 1–24, 2011.

592 D. E. Nikonov, G. I. Bourianoff, and P. A. Gargini, "Suitability for Digital Logic and Scaling of Atomistic Magnetic QCA," *Device Research Conference*, pp. 163–164, 23–25 June 2008 2008.

593 E. Varga, A. Orlov, M. T. Niemier, X. S. Hu, G. H. Bernstein, and W. Porod, "Experimental Demonstration of Fanout for Nanomagnetic Logic," *IEEE Transactions on Nanotechnology*, vol. 9, no. 6, pp. 668–670, 2010.

594 E. Varga, M. T. Niemier, G. Csaba, G. H. Bernstein, and W. Porod, "Experimental Realization of a Nanomagnet Full Adder Using Slanted-Edge Input Magnets," *INTERMAG / MMM*, Chicago, IL, 2013.

595 M. A. J. Siddiq, M. T. Niemier, G. H. Bernstein, W. Porod, and X. S. Hu, "A Field Coupled Electrical Input for Nanomagnet Logic," *IEEE Transactions on Nanotechnology*, 2013.

596 M. A. Siddiq, M. Niemier, G. Csaba, X. S. Hu, W. Porod, and G. H. Bernstein, " Demonstration of Field Coupled Input Scheme on Line of Nanomagnets," *accepted in IEEE Transactions on Magnetics*, 2013.

597 A. Lyle et al., "Integration of spintronic interface for nanomagnetic arrays," *AIP Advances*, vol. 1, no. 4, pp. 042177–11, 2011.

598 A. Lyle et al., "Spin transfer torque programming dipole coupled nanomagnet arrays," *Applied Physics Letters*, vol. 100, no. 1, pp. 012402–3, 2012.

599 A. Lyle, A. Klemm, J. Harms, Y. Zhang, H. Zhao, and J.-P. Wang, "Probing dipole coupled nanomagnets using magnetoresistance read," *J. App. Phys.*, vol. 98, no. 9, p. 092502, 2011.

600 S. Liu, X. S. Hu, M. T. Niemier, J. J. Nahas, G. H. Bernstein, and W. Porod, "A Design Space Exploration of the Magnetic-Electrical Interfaces for Nanomagnet Logic," *IEEE Transactions on Nanotechnology*, vol. 12, no. 2, pp. 203–214, 2013.

601 S. Bhanja, D. Karunaratne, R. Panchumarthy, S. Rajaram, and S. Sarkar, "Non-Boolean computing with nanomagnets for computer vision applications," *Nature nanotechnology*, vol. 11, no. 2, p. 177, 2016.

602 M. T. Alam, M. J. Siddiq, G. H. Bernstein, M. Niemier, W. Porod, and X. S. Hu, "On-Chip Clocking for Nanomagnet Logic Devices," *IEEE Transactions on Nanotechnology*, vol. 9, no. 3, pp. 348–351, May 2010.

603 M. T. Alam et al., "On-chip Clocking of Nanomagnet Logic Lines and Gates," *IEEE Transactions on Nanotechnology*, vol. 11, no. 2, pp. 273–286, 2012.

604 P. Li et al., "Power Reduction in Nanomagnet Logic Clocking through High Permeability Dielectrics," *Device Research Conference*, State College, Pennsylvania, 2012, pp. 129–130.

605 P. Li et al., "Paths to Clock Power Reduction via High Permeability Dielectrics for Nanomagnet Logic Circuits," *Joint MMM/Intermag Conference*, Chicago, IL, 2013.

606 Y. H. Chu et al., "Electric-field control of local ferromagnetism using a magnetoelectric multiferroic," (in English), *Nat. Mat.*, Article vol. 7, no. 6, pp. 478–482, Jun 2008.

607 F. M. Salehi, K. Roy, J. Atulasimha, and S. Bandyopadhyay, "Magnetization dynamics, Bennett clocking and associated energy dissipation in multiferroic logic," *Nanotechnology*, vol. 22, no. 15, p. 155201, 2011.

608 N. D'Souza, M. Salehi Fashami, S. Bandyopadhyay, and J. Atulasimha, "Experimental clocking of nanomagnets with strain for ultralow power Boolean logic," *Nano letters*, vol. 16, no. 2, pp. 1069–1075, 2016.

609 D. Bhowmik, Y. Long, and S. Salahuddin, "Possible route to low current, high speed, dynamic switching in a perpendicular anisotropy CoFeB-MgO junction using Spin Hall Effect of Ta," *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 29.7.1–29.7.4.

610 S. Breitkreutz, J. Kiermaier, S. V. Karthik, G. Csaba, D. Schmitt-Landsiedel, and M. Becherer, "Controlled reversal of Co/Pt Dots for nanomagnetic logic applications," *Journal of Applied Physics*, vol. 111, no. 7, pp. 07A715–3, 04/01/ 2012.

611 I. Eichwald et al., "Nanomagnet Logic: error-free directed signal transmission by an inverter chain," *IEEE Transactions on Magnetics*, vol. 48, no. 11, pp. 4332–4335, 2012.

612 S. Breitkreutz et al., "Majority Gate for Nanomagnetic Logic with Perpendicular Magentic Anisotropy," *IEEE Transactions on Magnetics*, vol. 48, no. 11, pp. 4336–4339, 2012.

613 I. Eichwald et al., "Towards a Signal Crossing in double-layer Nanomagnetic Logic," *IEEE Transactions on Magnetics*, vol. 49, 2013.

614 S. Breitkreutz et al., "Experimental Demonstration of a 1-bit Full Adder in Perpendicular Nanomagnetic Logic," *IEEE Transactions on Magnetics*, vol. 49, 2013.

615 E. Irina, B. Stephan, Z. Grazvydas, C. György, P. Wolfgang, and B. Markus, "Majority logic gate for 3D magnetic computing," *Nanotechnology*, vol. 25, no. 33, p. 335202, 2014.

616 R. Perricone, Z. Yining, K. M. Sanders, X. S. Hu, and M. Niemier, "Towards systematic design of 3D pNML layouts," *Design, Automation & Test in Europe Conference & Exhibition*, 2015, pp. 1539–1542.

617 J. Kiermaier, S. Breitkreutz, G. Csaba, D. Schmitt-Landsiedel, and M. Becherer, "Electrical input structures for nanomagnetic logic devices," *Journal of Applied Physics*, vol. 111, no. 7, pp. 07E341–3, 04/01/ 2012.

618 X. Ju, M. Niemier, M. Becherer, W. Porod, P. Lugli, and G. Csaba, "Systolic Pattern Matching Hardware with Out-of-Plane Nanomagnet Logic Devices," *IEEE Transactions on Nanotechnology*, vol. 12, no. 3, pp. 399–407, 2013.

619 W.-G. Wang, M. Li, S. Hageman, and C. L. Chien, "Electric-field-assisted switching in magnetic tunnel junctions," *Nat Mater*, 10.1038/nmat3171 vol. 11, no. 1, pp. 64–68, 01//print 2012.

620 M. Becherer, J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Csaba, and D. Schmitt-Landsiedel, "Nanomagnetic Logic clocked in the MHz regime," *to appear in the Proceedings of the 43rd European Solid-State Device Research Conference (ESSDERC)*, Bucharest, Romania, 2013.

621 M. Becherer et al., "A monolithic 3D integrated nanomagnetic co-processing unit," *Solid-State Electronics*, vol. 115, pp. 74–80, 2016.

622 D. B. Carlton, N. C. Emley, E. Tuchfeld, and J. Bokor, "Simulation Studies of Nanomagnet-Based Logic Architecture," *Nano Letters*, vol. 8, no. 12, pp. 4173–8, 2008.

623 M. T. Niemier et al., "Nanomagnet Logic: Progress Toward System-Level Integration," *J. Phys. Con. Mat.*, vol. 23, p. 493202, 2011.

624 S. Kurtz, E. Varga, M. Niemier, W. Porod, G. H. Bernstein, and X. S. Hu, "Two Input, Non-Majority Magnetic Logic Gates: Experimental Demonstration and Future Prospects," *Journal of Physics: Condensed Matter*, vol. 23, no. 5, p. 053202, 2011.

625 H. Nomura, N. Yoshioka, S. Miura, and R. Nakatani, "Controlling operation timing and data flow direction between nanomagnet logic elements with spatially uniform clock fields," *Applied Physics Express*, vol. 10, no. 12, p. 123004, 2017.

626 H. Dey, G. Csaba, G. Bernstein, and W. Porod, "Exchange coupling between laterally adjacent nanomagnets," *Nanotechnology*, vol. 27, no. 39, p. 395202, 2016.

627 J. Cui, S. M. Keller, C.-Y. Liang, G. P. Carman, and C. S. Lynch, "Nanoscale magnetic ratchets based on shape anisotropy," *Nanotechnology*, vol. 28, no. 8, p. 08LT01, 2017.

628 Q. Weikang and M. D. Riedel, "The synthesis of robust polynomial arithmetic with stochastic logic," *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, 2008, pp. 648–653.

629 R. Venkatesan, S. Venkataramani, F. Xuanyao, K. Roy, and A. Raghunathan, "Spintastic: Spin-based stochastic logic for energy-efficient computing," *Design, Automation & Test in Europe Conference & Exhibition 2015*, 2015, pp. 1575–1578.

630 R. Perricone, Y. Liu, A. Dingler, X. S. Hu, and M. Niemier, "Design of stochastic computing circuits using nanomagnetic logic," *IEEE Transactions on Nanotechnology*, vol. 15, no. 2, pp. 179–187, 2016.

631 D. Nikonov and I. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," *International Electron Devices Meeting (IEDM)*, San Francisco, CA, 2012, pp. 25.4.1–25.4.4.

632 G. Csaba and W. Porod, "Behavior of Nanomagnet Logic in the presence of thermal noise," *Computational Electronics (IWCE), 2010 14th International Workshop on*, 2010, pp. 1–4.

633 E. Varga, G. Csaba, G. H. Bernstein, and W. Porod, "Domain-Wall Assisted Switching of Single-Domain Nanomagnets," *Magnetics, IEEE Transactions on*, vol. 48, no. 11, pp. 3563–3566, 2012.

634 Andreakou, S.V. Poltavtsev, J.R. Leonard, E.V. Calman, M. Remeika, Y.Y. Kuznetsova, L.V. Butov, J. Wilkes, M. Hanson, A.C. Gossard, "Optically controlled excitonic transistor," *Appl. Phys. Lett.* 104, 091101 (2014).

635 C.J. Dorow, Y.Y. Kuznetsova, J.R. Leonard, M.K. Chu, L.V. Butov, J. Wilkes, M. Hanson, A.C. Gossard, "Indirect excitons in a potential energy landscape created by a perforated electrode," *Appl. Phys. Lett.* 108, 073502 (2016).

636 A.G. Winbow, J.R. Leonard, M. Remeika, Y.Y. Kuznetsova, A.A. High, A.T. Hammack, L.V. Butov, J. Wilkes, A.A. Guenther, A.L. Ivanov, M. Hanson, A.C. Gossard, "Electrostatic Conveyer for Excitons," *Phys. Rev. Lett.* 106, 196806 (2011).

637 A.A. High, J.R. Leonard, A.T. Hammack, M.M. Fogler, L.V. Butov, A.V. Kavokin, K.L. Campman, A.C. Gossard, "Spontaneous coherence in a cold exciton gas," *Nature* 483, 584 (2012).

638 J.R. Leonard, A.A. High, A.T. Hammack, M.M. Fogler, L.V. Butov, K.L. Campman, A.C. Gossard, "Pancharatnam-Berry phase in condensate of indirect excitons," *SRC* publication P092121.

639 M.M. Fogler, L.V. Butov, K.S. Novoselov, "Indirect excitons in van der Waals heterostructures," *Nature Commun.* 5, 4555 (2014).

640 E.V. Calman, M.M. Fogler, L.V. Butov, S. Hu, A. Mishchenko, A.K. Geim, "Indirect excitons in van der Waals heterostructures at room temperature," *SRC* publication P092424.

641 J.R. Leonard, A.A. High, A.T. Hammack, M.M. Fogler, L.V. Butov, K.L. Campman, A.C. Gossard, Pancharatnam-Berry phase in condensate of indirect excitons, Nature Commun. 9, 2158 (2018).

642 C.J. Dorow, J.R. Leonard, M.M. Fogler, L.V. Butov, K.W. West, L.N. Pfeiffer, Split-gate device for indirect excitons, Appl. Phys. Lett. 112, 183501 (2018).

643 C.J. Dorow, M.W. Hasling, D.J. Choksy, J.R. Leonard, L.V. Butov, K.W. West, L.N. Pfeiffer, High-mobility indirect excitons in wide single quantum well, SRC publication P095019.

644 M. Feng, N. Holonyak, Jr., and W. Hafez, "Light-emitting transistor: Light emission from InGaP/GaAs heterojunction bipolar transistors," *Appl. Phys. Lett.* 84, pp. 151–153 (2004).

645 M. Feng, N. Holonyak, Jr., G. Walter, and R. Chan, "Room temperature continuous wave operation of a heterojunction bipolar transistor laser," *Appl. Phys. Lett.* 87, pp. 131103, (2005).

646 H.W. Then, M. Feng, and N. Holonyak, Jr., "The Transistor Laser: Theory and Experiment," *Proc. IEEE* 101, pp. 2271, (2013).

647 J.M. Dallesasse, P.L. Lam, and G. Walter, "Electronic-Photonic Integration Using the Light-Emitting Transistor," *Latin American Optics and Photonics Conference (LAOP) 2014*, paper LTu3A.2, (2014).

648 M. Feng, J. Qiu, and N. Holonyak, Jr., "Tunneling Modulation of Transistor Lasers: Theory and Experiment," *IEEE J. Quant. Electron*,vol. 54, issue 2, April 2018.

649 R. Bambery, C. Wang, J.M. Dallesasse, M. Feng, and N. Holonyak, Jr., "Effect of the energy barrier in the base of the transistor laser on the recombination lifetime," *Appl. Phys. Lett.* 8, pp 081117 (2014).

650 M. Feng, H.W. Then, and N. Holonyak, Jr., "Transistor Laser Optical NOR Gate for High-Speed Optical Logic Processors," http://www.dtic.mil/dtic/tr/fulltext/u2/1041445.pdf, AD1041445, 20 Mar (2017).

651 R. Bambery, C.Y. Wang, F. Tan, M. Feng, and N. Holonyak, Jr., "Single Quantum-Well Transistor Lasers Operating Error-Free at 22 Gb/s," *IEEE Photon. Tech. Lett.* 27, pp 600, (2015).

652 F. Tan, R. Bambery, M. Feng, and N. Holonyak, Jr., “Relative intensity noise of a quantum well transistor laser,” *Appl. Phys. Lett*. 101, 151118 (2012).

653 M. Feng, M. Liu, C. Wang and N. Holonyak, "Vertical cavity transistor laser for on-chip OICs," *2015 IEEE Summer Topicals Meeting Series (SUM)*, Nassau, 2015, pp. 146–147, doi: 10.1109/PHOSST.2015.7248238.

654 D. E. Nikonov and I. A. Young, "Overview of Beyond-CMOS Devices and a Uniform Methodology for Their Benchmarking," *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2498–2533, 2013

655 D. E. Nikonov and I. A. Young, "Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits," *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on*, vol. 1, pp. 3–11, 2015

656 C. Pan and A. Naeemi, "An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 3, pp. 101–110, 2017

657 A. Iyengar, K. Ramclam, and S. Ghosh, "DWM-PUF: A low-overhead, memory-based security primitive," *Hardware-Oriented Security and Trust (HOST), 2014 IEEE International Symposium on*, 2014, pp. 154–159: IEEE

658 G. S. Rose, N. McDonald, Y. Lok-Kwong, and B. Wysocki, "A write-time based memristive PUF for hardware security applications," *Computer-Aided Design (ICCAD), 2013 IEEE/ACM International Conference on*, 2013, pp. 830–833

659 S. T. C. Konigsmark, L. K. Hwang, C. Deming, and M. D. F. Wong, "CNPUF: A Carbon Nanotube-based Physically Unclonable Function for secure low-energy hardware design," *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*, 2014, pp. 73–78

660 J. Rajendran et al., "Nano meets security: Exploring nanoelectronic devices for security applications," *Proceedings of the IEEE*, vol. 103, no. 5, pp. 829–849, 2015

661 G. E. Suh, C. W. O'Donnell, I. Sachdev, and S. Devadas, "Design and implementation of the AEGIS single-chip secure processor using physical random functions," *ACM SIGARCH Computer Architecture News*, 2005, vol. 33, no. 2, pp. 25–36: IEEE Computer Society.

662 J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "Physical unclonable functions and public-key crypto for FPGA IP protection," *Field Programmable Logic and Applications, 2007. FPL 2007. International Conference on*, 2007, pp. 189–195: IEEE.

663 T. Marukame, T. Tanamoto, and Y. Mitani, "Extracting physically unclonable function from spin transfer switching characteristics in magnetic tunnel junctions," *IEEE Transactions on Magnetics*, vol. 50, no. 11, pp. 1–4, 2014.

664 J. Das, K. Scott, S. Rajaram, D. Burgett, and S. Bhanja, "MRAM PUF: A novel geometry based magnetic PUF with integrated CMOS," *IEEE Transactions on Nanotechnology*, vol. 14, no. 3, pp. 436–443, 2015.

665 L. Zhang, Z. H. Kong, C.-H. Chang, A. Cabrini, and G. Torelli, "Exploiting process variations and programming sensitivity of phase change memory for reconfigurable physical unclonable functions," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 921–932, 2014.

666 Y. Pang, H. Wu, B. Gao, N. Deng, D. Wu, R. Liu, S. Yu, A. Chen, and H. Qian, “Optimization of RRAM-Based Physical Unclonable Function With a Novel Differential Read-Out Method,” *IEEE Electron. Dev. Lett*. 38(2), 168-171 (2017).

667 Y. Pang, H. Wu, B. Gao, D. Wu, A. Chen, and H. Qian, “A Novel PUF Against Machine Learning Attack: Implementation on a 16 Mb RRAM Chip,” *IEDM* 290-293 (2017)

668 U. Rührmair, C. Jaeger, C. Hilgers, M. Algasinger, G. Csaba, and M. Stutzmann, "Security applications of diodes with unique current-voltage characteristics," *International Conference on Financial Cryptography and Data Security*, 2010, pp. 328–335: Springer.

669 D. P. C. Dimitrakopoulos, and J. Smith, "Authentication using graphene based devices as physical unclonable functions," 2004. Available: http://www.google.com/patents/US20140159040

670 S. C. Konigsmark, L. K. Hwang, D. Chen, and M. D. Wong, "Cnpuf: A carbon nanotube-based physically unclonable function for secure low-energy hardware design," *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*, 2014, pp. 73–78: IEEE

671 G. S. Rose, N. McDonald, L.-K. Yan, and B. Wysocki, "A write-time based memristive PUF for hardware security applications," *Proceedings of the International Conference on Computer-Aided Design*, 2013, pp. 830–833: IEEE Press

672 P. Koeberl, Ü. Kocabaş, and A.-R. Sadeghi, "Memristor PUFs: a new generation of memory-based physically unclonable functions," *Proceedings of the Conference on Design, Automation and Test in Europe*, 2013, pp. 428–431: EDA Consortium

673 Y. H. Tseng, C.-E. Huang, C.-H. Kuo, Y.-D. Chih, and C. J. Lin, "High density and ultra small cell size of contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits," *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1–4: IEEE

674 C.-Y. Huang, W. C. Shen, Y.-H. Tseng, Y.-C. King, and C.-J. Lin, "A contact-resistive random-access-memory-based true random number generator," *IEEE Electron Device Letters*, vol. 33, no. 8, pp. 1108–1110, 2012

675 A. Colli, S. Pisana, A. Fasoli, J. Robertson, and A. C. Ferrari, "Electronic transport in ambipolar silicon nanowires," *physica status solidi (b)*, vol. 244, no. 11, pp. 4161–4164, 2007

676 R. Martel et al., "Ambipolar Electrical Transport in Semiconducting Single-Wall Carbon Nanotubes," *Physical Review Letters*, vol. 87, no. 25, p. 256805, 12/03/ 2001

677 A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nat Mater*, 10.1038/nmat1849 vol. 6, no. 3, pp. 183–191, 03//print 2007

678 L. Yu-Ming, J. Appenzeller, J. Knoch, and P. Avouris, "High-performance carbon nanotube field-effect transistor with tunable polarities," *Nanotechnology, IEEE Transactions on*, vol. 4, no. 5, pp. 481–489, 2005

679 N. Harada, K. Yagi, S. Sato, and N. Yokoyama, "A polarity-controllable graphene inverter," *Applied Physics Letters*, vol. 96, no. 1, p. 012102, 2010

680 A. Heinzig, S. Slesazeck, F. Kreupl, T. Mikolajick, and W. M. Weber, "Reconfigurable Silicon Nanowire Transistors," *Nano Letters*, vol. 12, no. 1, pp. 119–124, 2012/01/11 2012

681 S. Das and J. Appenzeller, "WSe2 field effect transistors with enhanced ambipolar characteristics," *Applied Physics Letters*, vol. 103, no. 10, p. 103501, 2013

682 M. De Marchi et al., "Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire FETs," *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 8.4.1–8.4.4

683 T. Vasen, "Investigation of III-V Tunneling Field-Effect Transistors," A Dissertation submitted to the University of Notre Dame, 2014.

684 Y. Bi *et al.*, "Emerging Technology Based Design of Primitives for Hardware Security," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 13, no. 1, Article 3, 2016

685 B. Yu, P. E. Gaillardon, X. S. Hu, M. Niemier, Y. Jiann-Shiun, and J. Yier, "Leveraging Emerging Technology for Hardware Security - Case Study on Silicon Nanowire FETs and Graphene SymFETs," *Test Symposium (ATS), 2014 IEEE 23rd Asian*, 2014, pp. 342–347

686 A. Chen, X. S. Hu, Y. Jin, M. Niemier, and X. Yin, "Using Emerging Technologies for Hardware Security Beyond PUFs," *Design, Automation, and Test in Europe (DATE)*, Dresden, Germany, 2016

687 J. Rajendran *et al.*, "Fault Analysis-Based Logic Encryption," *Computers, IEEE Transactions on*, vol. 64, no. 2, pp. 410–424, 2015

688 Y. Bi *et al.*, "Emerging Technology Based Design of Primitives for Hardware Security," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 13, no. 1, Article 3, 2016

689 B. Yu, P. E. Gaillardon, X. S. Hu, M. Niemier, Y. Jiann-Shiun, and J. Yier, "Leveraging Emerging Technology for Hardware Security - Case Study on Silicon Nanowire FETs and Graphene SymFETs," *Test Symposium (ATS), 2014 IEEE 23rd Asian*, 2014, pp. 342–347

690 Y. Bi, K. Shamsi, J.-S. Yuan, F.-X. Standaert, and Y. Jin, "Leverage emerging technologies for dpa-resilient block cipher design," *Proceedings of the 2016 Conference on Design, Automation & Test in Europe*, 2016, pp. 1538–1543: EDA Consortium

691 Y. Bi, K. Shamsi, J.-S. Yuan, Y. Jin, M. Niemier, and X. S. Hu, "Tunnel fet current mode logic for dpa-resilient circuit designs," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 3, pp. 340–352, 2017

692 A. C. Seabaugh and Z. Qin, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095–2110, 2010

693 H. Lu and A. Seabaugh, "Tunnel field-effect transistors: State-of-the-art," *IEEE Journal of the Electron Devices Society*, vol. 2, no. 4, pp. 44–49, 2014

694 L. Britnell et al., "Resonant tunnelling and negative differential conductance in graphene transistors," *Nat Commun*, Article vol. 4, p. 1794, 04/30/online 2013

695 MishchenkoA et al., "Twist-controlled resonant tunnelling in graphene/boron nitride/graphene heterostructures," *Nat Nano*, Letter vol. 9, no. 10, pp. 808–813, 10//print 2014

696 M. O. Li, D. Esseni, J. J. Nahas, D. Jena, and H. G. Xing, "Two-Dimensional Heterojunction Interlayer Tunneling Field Effect Transistors (Thin-TFETs)," *Electron Devices Society, IEEE Journal of the*, vol. 3, no. 3, pp. 200–207, 2015

697 H. Mamiya, A. Miyaji, and H. Morimoto, "Efficient Countermeasures against RPA, DPA, and SPA," *Cryptographic Hardware and Embedded Systems – CHES 2004*, vol. 3156, M. Joye and J.-J. Quisquater, Eds. (Lecture Notes in Computer Science: Springer Berlin Heidelberg, 2004, pp. 343–356

698 B. Sedighi, J. J. Nahas, M. Niemier, and X. S. Hu, "Boolean circuit design using emerging tunneling devices," *Computer Design (ICCD), 2014 32nd IEEE International Conference on*, 2014, pp. 355–360

699 P. Kocher, “Design and validation strategies for obtaining assurance in countermeasures to power analysis and related attacks,” NIST Physical Security Workshop, September 26-29 (2005).

700 P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," *Annual International Cryptology Conference*, 1999, pp. 388–397: Springer

701 M.-L. Akkar and C. Giraud, "An implementation of DES and AES, secure against some attacks," *International Workshop on Cryptographic Hardware and Embedded Systems*, 2001, pp. 309–318: Springer

702 K. Tiri, M. Akmal, and I. Verbauwhede, "A dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis on smart cards," *Solid-State Circuits Conference, 2002. ESSCIRC 2002. Proceedings of the 28th European*, 2002, pp. 403–406: IEEE

703 Y. Bi, K. Shamsi, J.-S. Yuan, Y. Jin, M. Niemier, and X. S. Hu, "Tunnel fet current mode logic for dpa-resilient circuit designs," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 3, pp. 340–352, 2017

704 C. De Canniere, O. Dunkelman, and M. Knežević, "KATAN and KTANTAN—a family of small and efficient hardware-oriented block ciphers," *Cryptographic Hardware and Embedded Systems-CHES 2009*: Springer, 2009, pp. 272–288

705 A. Cevrero, F. Regazzoni, M. Schwander, S. Badel, P. Ienne, and Y. Leblebici, "Power-gated mos current mode logic (pg-mcml): A power aware dpa-resistant standard cell library," *Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE*, 2011, pp. 1014–1019: IEEE

706 M. J. Cotter, Y. Fang, S. P. Levitan, D. M. Chiarulli and V. Narayanan, "Computational Architectures Based on Coupled Oscillators," *2014 IEEE Computer Society Annual Symposium on VLSI*, Tampa, FL, 2014, pp. 130–135. doi: 10.1109/ISVLSI.2014.87

707 M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Ed., Cambridge University Press, 2011.

708 NIST, *Quantum Algorithm Zoo*, website, online at: https://math.nist.gov/quantum/zoo/. Accessed: Nov. 5, 2018.

709 C. G. Almudever, L. Lao, X. Fu, N. Khammassi, I. Ashraf, D. Iorga, S. Varsamopoulos *et al.*, "The engineering challenges in quantum computing." In 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 836-845. IEEE, 2017.

710 S. Joshi, C. Kim, S. Ha and G. Cauwenberghs, "From algorithms to devices: Enabling machine learning through ultra-low-power VLSI mixed-signal array processing," *2017 IEEE Custom Integrated Circuits Conference (CICC)*, Austin, TX, pp. 1–9, 2017.

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

712 R. Genov and G. Cauwenberghs, "Charge-mode parallel architecture for vector-matrix multiplication," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, no. 10, pp. 930–936, 2001.

713 V. Bohossian, P. Hasler, and J. Bruck, "Programmable neural logic," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging*, vol. 21, no. 4, pp. 346–351, Nov. 1998.

714 R. Robucci, J. Gray, Leung Kin Chiu, J. Romberg, and P. Hasler, "Compressive Sensing on a CMOS Separable-Transform Image Sensor," *IEEE proceedings*, vol. 98, no. 1, 2010, pp. 1089–1101.

715 S. Koziol, D. Lenz, S. Hilsenbeck, S. Chopra, P. Hasler, and A. Howard, "Using floating-gate based programmable analog arrays for real-time control of a game-playing robot,"*Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 3566–3571.

716 P. Hasler and J. Dugger, "An analog floating-gate node for supervised learning," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 5, pp. 834–845, May 2005.

717 S. Suh, A. Basu, C. Schlottmann, P. Hasler, and J. Barry, "Low-power discrete Fourier transform for OFDM: A programmable analog approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 2, pp. 290–298, 2011.

718 S. Ramakrishnan and J. Hasler, "Vector-matrix multiply and winner-take-all as an analog classifier," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 353–361, 2014.

719 S. Agarwal *et al.*, "Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and its Application to Sparse Coding," *Frontiers in Neuroscience*, vol. 9, p. 484, 2016, Art. no. 484.

720 M. J. Marinella *et al.*, "Multiscale Co-Design Analysis of Energy, Latency, Area, and Accuracy of a ReRAM Analog Neural Training Accelerator," *arXiv preprint arXiv:1707.09952*, 2017.

721 G. W. Burr *et al.*, "Access Devices for 3D Crosspoint Memory," *Journal of Vacuum Science and Technology B*, vol. 32, no. 4, p. 040802, 2014.

722 S. Agarwal *et al.*, "Resistive memory device requirements for a neural algorithm accelerator," *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 929–938.

723 W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, and H. Qian, "Improving Analog Switching in HfOx-Based Resistive Memory With a Thermal Enhanced Layer," *IEEE Electron. Dev. Lett.* 38(8), 1019-1022 (2017).

724 P. Yao, H. Wu, B. Gao, S.B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.S.P. Wong, and H. Qian, "Face classification using electronic synapses," *Nature Comm.* 8, 1-8 (2017).

725 M. Hu *et al.*, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," *2016 53nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2016, pp. 1–6.

726 B. Chakrabarti *et al.*, "A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit," *Scientific Reports*, Article vol. 7, p. 42429, 02/14/online 2017.

727 G. W. Burr, et al. "Recent progress in phase-change memory technology," *IEEE J. Emerg. Sel. Top. Circ. Sys.* 6(2), 146-162 (2016).

728 S. Kim et al., "A phase change memory cell with metallic surfactant layer as a resistance drift stabilizer," *Electron Devices Meeting (IEDM), 2013 IEEE International*, 30.7.

729 W. W. Koelmans, et al., "Projected phase-change memory devices," *Nature communications*, (6) 8181 (2015).

730 E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Advanced Materials*, vol. 29, no. 4, p. 1604310, 2017.

731 Y. van de Burgt *et al.*, "A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing," *Nat Mater*, Letter vol. 16, no. 4, pp. 414–418, 2017.

732 J. Hasler, C. Diorio, B. A. Minch, and C. A. Mead, "Single transistor learning synapses," Gerald Tesauro, David S. Touretzky, and Todd K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, MIT Press, Cambridge, MA, 1994, pp. 817–824.

733 A. Bandyopadhyay, G. Serrano, and P. Hasler, "Adaptive algorithm using hot-electron injection for programming analog computational memory elements within 0.2 percent of accuracy over 3.5 decades," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 9, pp. 2107–2114, Sept. 2006.

[734] S. Kim, J. Hasler, and S. George, "Integrated Floating-Gate Programming Environment for System-Level ICs," *IEEE Transactions on VLSI*, vol. 24, no. 6, 2016. pp. 2244–2252.

[735] G.G. Marotta, A. Macerola, A. D'Alessandro, A. Torsi, C. Cerafogli, C. Lattaro, C. Musilli, D. Rivers, E. Sirizotti, and F. Paolini, *et al.*, "A 3bit/cell 32Gb NAND flash memory at 34nm with 6MB/s program throughput and with dynamic 2b/cell blocks configuration mode for a program throughput increase up to 13MB/s," *ISSCC*, 2010.

[736] N. Shibata, K. Kanda, T. Hisada, K. Isobe1, et al., "A 19nm 112.8mm2 64Gb Multi-Level Flash Memory with 400Mb/s/pin 1.8V Toggle Mode Interface," *ISSCC*, 2012, pp. 422-423.

[737] Y. Li, S. Lee, K. Oowada, H Nguyen, et. al, "128Gb 3b/Cell NAND Flash Memory in 19nm Technology with 18MB/s Write Rate and 400Mb/s Toggle Mode," *ISSCC*, 2012, pp. 436–437.

[738] D. Lee, I. J. Chang, S.Y. Yoon, J. Jang, et al., "A 64Gb 533Mb/s DDR Interface MLC NAND Flash in Sub-20 nm Technology," *ISSCC*, 2012, pp. 430–431.

[739] S. Kim, T. Gokmen, H. M. Lee, and W. E. Haensch, "Analog CMOS-based resistive processing unit for deep neural network training," In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems*, 422–425, 2017.

[740] Y. Li, S. Kim, X. Sun, P. Solomon, T. Gokmen, H. Tsai, S. Koswatta, Z. Ren, R. Mo, C. C. Yeh, W. Haensch and E. Leobandung, "Capacitor-based Cross-point Array for Analog Neural Network with Record Symmetry and Linearity," IEEE VLSI Systems Symposium, 2018.

[741] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, et al. "Equivalent-Accuracy Accelerated Neural-Network Training Using Analogue Memory," *Nature*, vol. 558, no. 7708, pp. 60–67, 2018.

[742] R. Genov, G. Cauwenberghs, G. Mulliken, and F. Adil, "A 5.9mW 6.5GMACS CID/DRAM array processor," *Proceedings of the 28th European Solid-State Circuits Conference*, pp. 715–718, Sept 2002.

[743] R. Karakiewicz, R. Genov, and G. Cauwenberghs, "480-GMACS/mW Resonant Adiabatic Mixed-Signal Processor Array for Charge-Based Pattern Recognition," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 11, pp. 2573–2584, Nov 2007.

[744] R. Karakiewicz, R. Genov, and G. Cauwenberghs, "1.1 TMACS/mW Fine-Grained Stochastic Resonant Charge-Recycling Array Processor," *IEEE Sensors Journal*, vol. 12, no. 4, pp. 785–792, April 2012.

[745] S. Joshi, C. Kim, S. Ha, M. Y. Chi, and G. Cauwenberghs, "A 2pJ/MAC 14-b 8x8 Linear Transform Mixed-Signal Spatial Filter in 65 nm CMOS with 84 dB Interference Suppression," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, February 2017.

[746] E. H. Lee and S. S. Wong, "A 2.5 GHz 7.7 TOPS/W switched capacitor matrix multiplier with co-designed local memory in 40nm," *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, pp. 418–419, 2016.

[747] J. Zhang, Z. Wang, and N. Verma, "A Matrix-Multiplying ADC Implementing a Machine-Learning Classifier Directly with Data Conversion," *IEEE Int. Solid-State Circ. Conf. (ISSCC)*, 2015.

[748] E.H Lee and S.S. Wong, "A 2.5 GHz 7.7TOPS/W Switched-Capacitor Matrix Multiplier with Co-designed Local Memory in 40nm," *IEEE Int. Solid-State Circ. Conf. (ISSCC)*, 2016.

[749] F. N. Buhler, A. E. Mendrela, Y. Lim, J. A. Fredenburg, and M. P. Flynn, "A 16-channel Noise-Shaping Machine Learning Analog-Digital Interface," *IEEE VLSI Systems Symposium*, 2016.

[750] W. Guo and N. Sun, "A 12b-ENOB 61 µW noise-shaping SAR ADC with a passive integrator," *42nd IEEE European Solid-State Circuits Conference (ESSCIRC 2016)*, 405-408, 2016.

[751] D. Kadetotad et al., "Parallel Architecture With Resistive Crosspoint Array for Dictionary Learning Acceleration," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 5, no. 2, pp. 194–204, 2015.

[752] G. W. Burr et al., "Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element," *Electron Devices, IEEE Transactions on*, vol. 62, no. 11, pp. 3498–3507, 2015.

[753] P. Y. Chen, L. Gao, and S. Yu, "Design of Resistive Synaptic Array for Implementing On-Chip Sparse Learning," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 4, pp. 257–264, 2016.

[754] R. B. Jacobs-Gedrim et al., "Impact of Linearity and Write Noise of Analog Resistive Memory Devices in a Neural Algorithm Accelerator," presented at the *IEEE International Conference on Rebooting Computing (ICRC)* Washington, DC, November 2017.

[755] S. Agarwal et al., "Achieving ideal accuracies in analog neuromorphic computing using periodic carry," *VLSI Technology, 2017 Symposium on*, 2017, pp. T174–T175: IEEE.

[756] I. Boybat et al., "Improved Deep Neural Network Hardware-Accelerators Based on Non-Volatile-Memory: The Local Gains Technique," *2017 IEEE International Conference on Rebooting Computing (ICRC)*, 2017, pp. 1–8.

[757] S. Agarwal et al. (2017). *CrossSim*. Available: http://cross-sim.sandia.gov.

[758] T. Gokmen and Y. Vlasov, "Acceleration of deep neural network training with resistive cross-point devices," *arXiv preprint arXiv:1603.07341*, 2016.

[759] C. Schlottmann and P. Hasler, "A highly dense, low power, programmable analog vector-matrix multiplier: The FPAA implementation," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 3, pp. 403–411, 2011.

[760] C. Schlottmann, S. Shapero, S. Nease, and P. Hasler, "A digitally enhanced dynamically reconfigurable analog platform for low-power signal processing," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 9, pp. 2174–2184, 2012.

761 S. George, S. Kim, S. Shah, J. Hasler, M. Collins, F. Adil, R. Wunderlich, S. Nease, and S. Ramakrishnan, "A Programmable and Configurable Mixed-Mode FPAA SOC," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 24, no. 6, 2016. pp. 2253–2261.

762 C. M. Twigg, J. D. Gray, and P. Hasler, "Programmable floating gate FPAA switches are not dead weight," *IEEE International Symposium on Circuits and Systems*, May 2007, pp. 169–172.

763 S. Shapero and P. Hasler, "Mismatch characterization and calibration for accurate and automated analog design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 548–556, 2013.

764 S. Kim, S. Shah, and J. Hasler, "Calibration of Floating-Gate SoC FPAA System," IEEE Transactions on VLSI, 2017.

765 S. Shah, H. Toreyin, J. Hasler, and A. Natarajan, "Models and Techniques For Temperature Robust Systems On A Reconfigurable Platform," *Journal of Low Power Electronics Applications*, vol. 7, no. 21, August 2017. pp. 1–14.

766 J. Hasler, C. Scholttmann, and S. Koziol, "FPAA chips and tools as the center of an design-based analog systems education," *IEEE Microelectronic Systems Education (MSE)*, 2011, pp. 47–51.

767 M. Collins, J. Hasler, and S. George, "Analog systems education: An integrated toolset and FPAA SoC boards," *Microelectronics Systems Education (MSE)*, 2015, pp. 32–35.

768 J. Hasler and H. Wang, "A Fine-Grain FPAA fabric for RF+Baseband," *GOMAC*, March 2015.

769 J. Hasler, S. Kim, and F. Adil, "Scaling Floating-Gate Devices Predicting Behavior for Programmable and Configurable Circuits and Systems," *Journal of Low Power Electronics Applications*, vol. 6, no. 13, 2016, pp. 1–19.

770 I. Richter *et al.*, "Memristive Accelerator for Extreme Scale Linear Solvers," *40th Annual GOMACTech Conference*, St. Louis, MO, 2015.

771 S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, "Mixed-precision architecture based on computational memory for training deep neural networks." *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018.

772 M. Imani, D. Peroni, A. Rahimi, and T. Rosing, "Resistive CAM Acceleration for Tunable Approximate Computing," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, p. 1, 2016.

773 L. Yavits, S. Kvatinsky, A. Morad, and R. Ginosar, "Resistive Associative Processor," *IEEE Computer Architecture Letters*, vol. 14, no. 2, pp. 148–151, 2015.

774 C. Yakopcic, V. Bontupalli, R. Hasan, D. Mountain, and T. M. Taha, "Self-biasing memristor crossbar used for string matching and ternary content-addressable memory implementation," *Electronics Letters*, vol. 53, no. 7, pp. 463–465, 2017.

775 J. C. Magee, "Dendritic integration of excitatory synaptic input," *Nature Reviews Neuroscience*, vol. 1, pp. 181–190, 12//print 2000.

776 H. C. Lai and L. Y. Jan, "The distribution and targeting of neuronal voltage-gated ion channels," *Nature Reviews Neuroscience*, vol. 7, pp. 548–62, Jul 2006.

777 B. P. Bean, "The action potential in mammalian central neurons," *Nature Reviews Neuroscience*, vol. 8, pp. 451–65, Jun 2007.

778 H. Lim, V. Kornijcuk, J. Y. Seok, S. K. Kim, I. Kim, C. S. Hwang, *et al.*, "Reliability of neuronal information conveyed by unreliable neuristor-based leaky integrate-and-fire neurons: a model study," *Scientific Reports*, vol. 5, p. 9776, May 13, 2015.

779 T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, "Stochastic phase-change neurons," *Nature Nanotechnology*, vol. 11, pp. 693–9, Aug 2016.

780 H. Lim, H. W. Ahn, V. Kornijcuk, G. Kim, J. Y. Seok, I. Kim, *et al.*, "Relaxation oscillator-realized artificial electronic neurons, their responses, and noise," *Nanoscale*, vol. 8, pp. 9629-40, May 14, 2016.

781 *S*engupta and K. Roy, "Spin-Transfer Torque Magnetic neuron for low power neuromorphic computing," 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, 2015, pp. 1-7

782 S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems," *Nano Letters*, vol. 10, pp. 1297-1301, 2010.

783 D. Niu, Y. Xiao, and Y. Xie, "Low power memristor-based ReRAM design with error correcting code," in *Design Automation Conference (ASP-DAC), 2012 17th Asia and South Pacific*, 2012, pp. 79–84.

784 L. Yang, L. Chua, and K. Krieg, "VLSI implementation of cellular neural networks," *Circuits and Systems, IEEE International Symposium on*, 1990, pp. 2425–2427.

785 E. Y. Chou, B. J. Sheu, and R. C. Chang, "VLSI design of optimization and image processing cellular neural networks," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 44, pp. 12–20, 1997.

786 A. R. Trivedi and S. Mukhopadhyay, "Potential of ultralow-power cellular neural image processing with Si/Ge tunnel FET," *Nanotechnology, IEEE Transactions on*, vol. 13, pp. 627–629, 2014.

787 I. Palit, X. S. Hu, J. Nahas, and M. Niemier, "TFET-based cellular neural network architectures," *Proceedings of the 2013 International Symposium on Low Power Electronics and Design*, 2013, pp. 236-241.

788 D. C. Ralph and M. D. Stiles, "Spin transfer torques," *Journal of Magnetism and Magnetic Materials*, vol. 320, pp. 1190-1216, 4// 2008.

789 D. V. Berkov and J. Miltat, "Spin-torque driven magnetization dynamics: Micromagnetic modeling," *Journal of Magnetism and Magnetic Materials*, vol. 320, pp. 1238–1259, 2008.

790 C. Pan and A. Naeemi, "A proposal for energy-efficient cellular neural network based on spintronic devices," *Nanotechnology, IEEE Transactions on*, vol. 15, pp. 1-8, 2016.

791 C. Pan and A. Naeemi, "Non-Boolean computing benchmarking for beyond-CMOS devices based on cellular neural network," *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on*, vol. 2, pp. 36–43, 2016.

792 M. Namba and Z. Zhang, "Cellular neural network for associative memory and its application to Braille image recognition," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 2006, pp. 2409–2414.

793 D. Liu and A. N. Michel, "Cellular neural networks for associative memories," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 40, pp. 119–121, 1993.

794 P. Szolgay, I. Szatmari, and K. Laszlo, "A fast fixed point learning method to implement associative memory on CNNs," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 44, pp. 362–366, 1997.

795 S. Miller and P. McWhorter, "Physics of the ferroelectric nonvolatile memory field effect transistor," *Journal of Applied Physics*, vol. 72, pp. 5999–6010, 1992.

796 S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano letters*, vol. 8, pp. 405–410, 2008.

797 A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proceedings of the IEEE*, vol. 98, pp. 2095–2110, 2010.

798 S. Das, A. Prakash, R. Salazar, and J. Appenzeller, "Toward low-power electronics: tunneling phenomena in transition metal dichalcogenides," *ACS nano*, vol. 8, pp. 1681–1689, 2014.

799 M. O. Li, D. Esseni, G. Snider, D. Jena, and H. G. Xing, "Single particle transport in two-dimensional heterojunction interlayer tunneling field effect transistor," *Journal of Applied Physics*, vol. 115, p. 074508, 2014.

800 L. Wenjun, S. Sharmin, H. Ilatikhameneh, R. Rahman, L. Yeqing, W. Jingshan, Y. Xiaodong, A. Seabaugh, G. Klimeck, D. Jena, and P. Fay, "Polarization-Engineered III-Nitride Heterojunction Tunnel Field-Effect Transistors," *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on*, vol. 1, pp. 28–34, 2015.

801 D. Newns, B. Elmegreen, X. H. Liu, and G. Martyna, "A low-voltage high-speed electronic switch based on piezoelectric transduction," *Journal of Applied Physics*, vol. 111, p. 084509, 2012.

802 C. Pan and A. Naeemi, "Device- and system-level performance modeling for graphene P-N junction logic," *Quality Electronic Design (ISQED), 2012 13th International Symposium on*, 2012, pp. 262–269.

803 J. Son, S. Rajan, S. Stemmer, and S. J. Allen, "A heterojunction modulation-doped Mott transistor," *Journal of Applied Physics*, vol. 110, p. 084503, 2011.

804 L. Liu, Y. Lu, and J. Guo, "On Monolayer Field-Effect Transistors at the Scaling Limit," *Electron Devices, IEEE Transactions on*, vol. 60, pp. 4133–4139, 2013.

805 B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotechnology*, vol. 5, pp. 266–270, 2010.

806 S. Datta, S. Salahuddin, and B. Behin-Aein, "Non-volatile spin switch for Boolean and non-Boolean logic," *Applied Physics Letters*, vol. 101, p. 252411, 2012.

807 D. Morris, D. Bromberg, J.-G. J. Zhu, and L. Pileggi, "mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices," *Proceedings of the 49th Annual Design Automation Conference*, 2012, pp. 486–491.

808 S. Sayed, V. Q. Diep, K. Y. Camsari, and S. Datta, "Spin Funneling for Enhanced Spin Injection into Ferromagnets," *Scientific Reports*, vol. 6, 2016.

809 Y. Li, X. de Milly, F. Abreu Araujo, O. Klein, V. Cros, J. Grollier, and G. de Loubens, "Probing Phase Coupling Between Two Spin-Torque Nano-Oscillators with an External Source," *Physical Review Letters*, vol. 118, no. 24, p. 247202, Jun. 2017.

810 R. Lebrun, S. Tsunegi, P. Bortolotti, H. Kubota, A. S. Jenkins, M. Romera, K. Yakushiji, A. Fukushima, J. Grollier, S. Yuasa, and V. Cros, "Mutual synchronization of spin torque nano-oscillators through a long-range and tunable electrical coupling scheme," *Nature Communications*, vol. 8, p. ncomms15825, Jun. 2017.

811 M. R. Pufall, W. H. Rippard, G. Csaba, D. E. Nikonov, G. I. Bourianoff, and W. Porod, "Physical Implementation of Coherently Coupled Oscillator Networks," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 1, pp. 76–84, Dec. 2015.

812 N. Shukla, A. Parihar, E. Freeman, H. Paik, G. Stone, V. Narayanan, H. Wen, Z. Cai, V. Gopalan, R. Engel-Herbert, D. G. Schlom, A. Raychowdhury, and S. Datta, "Synchronized charge oscillations in correlated electron systems," *Scientific Reports*, vol. 4, p. 4964, May 2014.

813 P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully-programmable 100-spin coherent Ising machine with all-to-all connections," *Science*, p. aah5178, Oct. 2016.

814 C. L. Tang, W. R. Bosenberg, T. Ukachi, R. J. Lane, and L. K. Cheng, "Optical parametric oscillators," *Proceedings of the IEEE*, vol. 80, no. 3, pp. 365–374, 1992.

[815] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. L. Heras, R. Babbush, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. OMalley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, E. Solano, H. Neven, and J. M. Martinis, "Digitized adiabatic quantum computing with a superconducting circuit," *Nature*, vol. 534, no. 7606, pp. 222–226, Jun. 2016.

[816] S. Kumar, J. P. Strachan, and R. S. Williams, "Chaotic dynamics in nanoscale NbO2 Mott memristors for analogue computing," *Nature*, vol. 548, p. 318, 2017.

[817] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," *Nature Materials*, vol. 12, pp. 114-7, Feb 2013.

[818] P.-Y. Chen, J.-s. Seo, Y. Cao, and S. Yu, "Compact oscillation neuron exploiting metal-insulator-transition for neuromorphic computing," presented at the *Proceedings of the 35th International Conference on Computer-Aided Design*, Austin, Texas, 2016.

[819] K. Yogendra, D. Fan, and K. Roy, "Coupled Spin Torque Nano Oscillators for Low Power Neural Computation," *IEEE Transactions on Magnetics*, vol. 51, no. 10, pp. 1–9, Oct. 2015.

[820] A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim, and K. Roy, "Magnetic Tunnel Junction Mimics Stochastic Cortical Spiking Neurons," *Scientific Reports*, vol. 6, p. srep30039, Jul. 2016.

[821] S. Kaka, M. R. Pufall, W. H. Rippard, T. J. Silva, S. E. Russek, and J. A. Katine, "Mutual phase-locking of microwave spin torque nano-oscillators," *Nature*, vol. 437, no. 7057, pp. 389–392, Sep. 2005.

[822] L. Maleki, "Sources: The optoelectronic oscillator," *Nature Photonics*, vol. 5, no. 12, pp. 728–730, 2011.

[823] W. A. Vitale, C. F. Moldovan, A. Paone, A. Schüler, and A. M. Ionescu, "Fabrication of CMOS-compatible abrupt electronic switches based on vanadium dioxide," *Microelectronic Engineering*, vol. 145, pp. 117–119, Sep. 2015.

[824] H.-T. Kim, B.-G. Chae, D.-H. Youn, S.-L. Maeng, G. Kim, K.-Y. Kang, and Y.-S. Lim, "Mechanism and observation of Mott transition in VO_2-based two- and three-terminal devices," *New Journal of Physics*, vol. 6, no. 1, p. 52, 2004.

[825] C. Li, E. Wasige, L. Wang, J. Wang, and B. Romeira, "28 GHz MMIC resonant tunnelling diode oscillator of around 1 mW output power," *Electronics Letters*, vol. 49, no. 13, pp. 816–818, Jun. 2013.

[826] M. Feiginov, C. Sydlo, O. Cojocari, and P. Meissner, "Resonant-tunnelling-diode oscillators operating at frequencies above 1.1 THz," *Applied Physics Letters*, vol. 99, no. 23, p. 233506, 2011.

[827] S. Suzuki, M. Asada, A. Teranishi, H. Sugiyama, and H. Yokoyama, "Fundamental oscillation of resonant tunneling diodes above 1 THz at room temperature," *Applied Physics Letters*, vol. 97, no. 24, p. 242102, 2010.

[828] H. Eisele, "State of the art and future of electronic sources at terahertz frequencies," *Electronics Letters*, vol. 46, no. 26, p. S8, 2010.

[829] T. Silva and W. Rippard, "Developments in nano-oscillators based upon spin-transfer point-contact devices," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1260–1271, Apr. 2008.

[830] F. Dörfler and F. Bullo, "Exploring synchronization in complex oscillator networks," *Decision and control (cdc), 2012 IEEE 51st annual conference on*, 2012, pp. 7157–7170.

[831] A. T. Winfree, "Biological rhythms and the behavior of populations of coupled oscillators," *Journal of theoretical biology*, vol. 16, no. 1, pp. 15–42, 1967.

[832] Y. Kuramoto, "Self-entrainment of a population of coupled non-linear oscillators," *International symposium on mathematical problems in theoretical physics*, 1975, pp. 420–422.

[833] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*. Courier Dover Publications, 2003.

[834] D. E. Nikonov, G. Csaba, W. Porod, T. Shibata, D. Voils, D. Hammerstrom, I. A. Young, and G. I. Bourianoff, "Coupled-Oscillator Associative Memory Array Operation for Pattern Recognition," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 1, pp. 85–93, Dec. 2015.

[835] F. C. Hoppensteadt and E. M. Izhikevich, "Oscillatory neurocomputers with dynamic connectivity," *Physical Review Letters*, vol. 82, no. 14, p. 2983, 1999.

[836] E. Mallada and A. Tang, "Synchronization of weakly coupled oscillators: Coupling, delay and topology," *Journal of Physics A: Mathematical and Theoretical*, vol. 46, no. 50, p. 505101, 2013.

[837] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," *Reviews of Modern Physics*, vol. 77, no. 1, p. 137, 2005.

[838] B. Van der Pol, "The Nonlinear Theory of Electric Oscillations," *Proceedings of the Institute of Radio Engineers*, vol. 22, no. 9, pp. 1051–1086, 1934.

[839] N. Kopell and D. Somers, "Anti-phase solutions in relaxation oscillators coupled through excitatory interactions," *Journal of Mathematical Biology*, vol. 33, no. 3, pp. 261–280, Dec. 1995.

[840] D. Storti and R. Rand, "Dynamics of two strongly coupled Van der Pol oscillators," *International Journal of Non-Linear Mechanics*, vol. 17, no. 3, pp. 143–152, 1982.

[841] A. Kouda and S. Mori, "Mode analysis of a system of mutually coupled Van der Pol oscillators with coupling delay," *International Journal of Non-Linear Mechanics*, vol. 17, no. 4, pp. 267–276, 1982.

[842] T. Chakraborty and R. H. Rand, "The transition from phase locking to drift in a system of two weakly coupled Van der Pol oscillators," *International Journal of Non-Linear Mechanics*, vol. 23, no. 5, pp. 369–376, 1988.

[843] G. Ermentrout, "Stable Periodic Solutions to Discrete and Continuum Arrays of Weakly Coupled Nonlinear Oscillators," *SIAM Journal on Applied Mathematics*, vol. 52, no. 6, pp. 1665–1687, Dec. 1992.

[844] E. Izhikevich, "Weakly pulse-coupled oscillators, FM interactions, synchronization, and oscillatory associative memory," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 508–526, May 1999.

[845] R. Mirollo and S. Strogatz, "Synchronization of Pulse-Coupled Biological Oscillators," *SIAM Journal on Applied Mathematics*, vol. 50, no. 6, pp. 1645–1662, Dec. 1990.

[846] R. Adler, "A Study of Locking Phenomena in Oscillators," *Proceedings of the IRE*, vol. 34, no. 6, pp. 351–357, 1946.

[847] B. Razavi, "A study of injection locking and pulling in oscillators," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1415–1424, 2004.

[848] C.-L. Hu, "Self-sustained oscillation in an Rh-C or Rh-L circuit containing a hysteresis resistor Rh," *IEEE Transactions on Circuits and Systems*, vol. 33, no. 6, pp. 636–641, 1986.

[849] A. Parihar, N. Shukla, M. Jerry, S. Datta, and A. Raychowdhury, "Vertex coloring of graphs via phase dynamics of coupled oscillatory networks," *Scientific Reports*, vol. 7, no. 1, p. 911, Apr. 2017.

[850] T. Saito, "On a coupled relaxation oscillator," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 9, pp. 1147–1155, 1988.

[851] N. Shukla, A. Parihar, M. Cotter, M. Barth, X. Li, N. Chandramoorthy, H. Paik, D. G. Schlom, V. Narayanan, A. Raychowdhury, and S. Datta, "Pairwise coupled hybrid vanadium dioxide-MOSFET (HVFET) oscillators for non-boolean associative computing," *2014 IEEE International Electron Devices Meeting*, 2014, pp. 28.7.1–28.7.4.

[852] S. Levitan, Y. Fang, D. Dash, T. Shibata, D. Nikonov, and G. Bourianoff, "Non-Boolean associative architectures based on nano-oscillators," *2012 13th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, 2012, pp. 1–6.

[853] G. Csaba, M. Pufall, D. Nikonov, G. Bourianoff, A. Horvath, T. Roska, and W. Porod, "Spin torque oscillator models for applications in associative memories," *2012 13th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, 2012, pp. 1–2.

[854] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, January 1949

[855] M.P. Frank and E. P. DeBenedictis, "A novel operational paradigm for thermodynamically reversible logic: Adiabatic transformation of chaotic nonlinear dynamical circuits," *2016 IEEE International Conference on Rebooting Computing (ICRC)*, San Diego, CA, 2016, pp. 1–8. doi: 10.1109/ICRC.2016.7738679

[856] M.P. Frank and E.P. DeBenedictis, "Chaotic Logic," presentation, *ICRC 2016*, San Diego, CA, Oct. 2016, https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Frank_ICRC2016_ChaoticLogic_presUUR+notes.pdf.

[857] M.P. Frank, DYNAMIC, open-source software, Sandia National Laboratories, Center for Computing Research, 2017, https://cfwebprod.sandia.gov/cfdocs/CompResearch/templates/insert/softwre.cfm?sw=56.

[858] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982.

[859] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, Jul. 1985.

[860] L. O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1257–1272, Oct. 1988.

[861] A. Lucas, "Ising formulations of many NP problems," *Interdisciplinary Physics*, vol. 2, p. 5, 2014.

[862] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, "Coherent Ising machine based on degenerate optical parametric oscillators," *Physical Review A*, vol. 88, no. 6, p. 063853, Dec. 2013.

[863] M. Ercsey-Ravasz and Z. Toroczkai, "Optimization hardness as transient chaos in an analog approach to constraint satisfaction," *Nature Physics*, vol. 7, no. 12, pp. 966–970, Dec. 2011.

[864] H. Mostafa, L. K. Müller, and G. Indiveri, "An event-based architecture for solving constraint satisfaction problems," *Nature Communications*, vol. 6, p. 8941, Dec. 2015.

[865] F. L. Traversa, C. Ramella, F. Bonani, and M. Di Ventra, "Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states," *Science Advances*, vol. 1, no. 6, pp. e1500031–e1500031, Jul. 2015.

[866] V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 418–423, Jan. 2007.

[867] G. Y. Vichniac, "Simulating physics with cellular automata," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 96–116, Jan. 1984.

[868] T. Toffoli, "Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 117–127, Jan. 1984.

[869] T. Toffoli, "CAM: A high-performance cellular-automaton machine," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 195–204, Jan. 1984.

[870] G. Appa, L. S. Pitsoulis, and H. P. Williams, Eds., *Handbook on modelling for discrete optimization*. New York: Springer, 2006.

[871] S. Aaronson, *Quantum Computing since Democritus*, Cambridge University Press, 2013.

[872] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development*, vol. 5, no. 3, pp. 183-191, July 1961. doi: 10.1147/rd.53.0183

[873] E. P. DeBenedictis, "Computational Complexity and New Computing Approaches," *Computer*, vol. 49, no. 12, pp. 76–79, Dec. 2016.

[874] M. Mitzenmacher and E. Upfal, Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis, 2nd ed., Cambridge University Press, 2017.

[875] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature reviews. Neuroscience*, vol. 9, p. 292, 2008.

[876] M. D. McDonnell and L. M. Ward, "The benefits of noise in neural systems: bridging theory and experiment," *Nature reviews. Neuroscience*, vol. 12, p. 415, 2011.

[877] W. Maass, "Noise as a resource for computation and learning in networks of spiking neurons," *Proceedings of the IEEE*, vol. 102, pp. 860–880, 2014.

[878] M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. N. Salama, "Memristors empower spiking neurons with stochasticity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 5, pp. 242-253, 2015.

[879] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Applied Physics Letters* 93, 082508 (2008).

[880] W. Butler, X.-G. Zhang, T. Schulthess, and J. MacLaren, "Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches," *Physical Review B* 63, 054416 (2001).

[881] R. K. Henderson, E. A. G. Webster, and R. Walker, "A Gate Modulated Avalanche Bipolar Transistor in 130nm CMOS Technology," *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, pp. 226-229, IEEE (2012)

[882] E. A.G. Webster, J. A. Richardson, L. A. Grant, R. K. Henderson, "A single electron bipolar avalanche transistor implemented in 90 nm CMOS" *Solid-State Electronics*, *76*, pp.116-118 (2012)

[883] M. L. Gallo, T. Tuma, F. Zipoli, A. Sebastian, and E. Eleftheriou, "Inherent stochasticity in phase-change memory devices," *2016 46th European Solid-State Device Research Conference (ESSDERC)*, 2016, pp. 373–376.

[884] H. Jiang, D. Belkin, S. E. Savel'ev, S. Lin, Z. Wang, Y. Li, *et al.*, "A novel true random number generator based on a stochastic diffusive memristor," *Nature Communications*, vol. 8, 2017.

[885] Z. Wang, S. Joshi, S. E. Saveliev, H. Jiang, R. Midya, P. Lin, *et al.*, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature Materials*, vol. 16, pp. 101–108, 09/26/online 2016.

[886] M. Park, J. C. Rodgers, and D. P. Lathrop, "True random number generation using CMOS Boolean chaotic oscillator," *Microelectronics Journal* 46, 1364 (2015).

[887] K. Y. Camsari, S. Salahuddin, and S. Datta, "Implementing p-bits With Embedded MTJ," *IEEE Electron Device Letters* 38, 1767 (2017), ISSN 0741-3106.

[888] B. Behin-Aein, V. Diep, and S. Datta, "A building block for hardware belief networks," *Scientic Reports* 29893, 6 (2016).

[889] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-Bits for Invertible Logic," *Phys. Rev.* X **7**, 031014 (2017), doi:10.1103/PhysRevX.7.031014.

[890] R. Faria, K. Y. Camsari, and S. Datta, "Low-barrier nanomagnets as p-bits for spin logic," *IEEE Magnetics Letters* vol. **8**, 4105305 (2017).

[891] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," *Scientic Reports* **7**, 44370, (2017).

[892] A. Z. Pervaiz, L. A. Ghantasala, K. Y. Camsari, and S. Datta, "Hardware emulation of stochastic p-bits for invertible logic," *Scientic Reports* **7**, 10994 (2017), doi:10.1038/s41598-017-11011-8.

[893] A. Z. Pervaiz, B. M. Sutton, L. Ghantasala, and K. Y. Camsari, "Weighted p-bits for FPGA implementation of probabilistic circuits," arXiv preprint arXiv:1712.04166 (2017).

[894] R. Zand K. Y. Camsari, I. Ahmed, S. D. Pyle, C. H. Kim, S. Datta, and R. F. DeMara, "R-DBN: A Resistive Deep Belief Network Architecture Leveraging the Intrinsic Behavior of Probabilistic Devices," arXiv preprint arXiv:1710.00249 (2017).

[895] M.P. Frank, "Introduction to reversible computing: motivation, progress, and challenges." In *Proceedings of the 2nd conference on Computing frontiers (CF '05)*. ACM, New York, NY, USA, 385-390 (2005).

[896] M.P. Frank, "Foundations of Generalized Reversible Computing," in *Reversible Computation (RC 2017). Lecture Notes in Computer Science*, Phillips I., Rahaman H. (eds), vol 10301. Springer, Cham (2017).

[897] M. P. Frank, "Physical Foundations of Landauer's Principle," in *Reversible Computation (RC 2018). Lecture Notes in Computer Science*, Kari J., Ulidowski I. (eds), vol. 11106. Springer, Cham, 2018.

[898] N. G. Anderson, "Conditional erasure and the Landauer limit," in Energy Limits in Computation: A Review of Landauer's Principle, Theory, and Experiments (pp. 65–100), Lent, Orlov, Porod, and Snider, eds., Springer Nature, 2019.

[899] M.P. Frank, "Generalizations of the Reversible Computing Paradigm," Presentation, *SFI Workshop on Thermodynamics and Computation: Towards a New Synthesis*, Santa Fe, New Mexico, August 2017. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/SFI-talk-final2_ho2up.pdf

900 C. H. Bennett, “Logical Reversibility of Computation,” *IBM Journal of Research and Development*, vol. 17, no. 6, pp. 525–532, Nov. 1973. doi: 10.1147/rd.176.0525

901 M.P. Frank and M.J. Ammer, “Relativized separation of reversible and irreversible space-time complexity classes,” *arxiv preprint* arXiv:1708.08480 [cs.ET], 2017.

902 S. Pidaparthi and C. Lent, “Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata,” *Journal of Low Power Electronics and Applications* vol. 8, no. 3 (2018).

903 M. P. Frank, “On the interpretation of energy as the rate of quantum computation,” *Quantum Information Processing*, vol. 4, no. 4 (2005).

904 M. P. Frank, and T. F. Knight Jr. "Ultimate theoretical models of nanocomputers." *Nanotechnology* 9, no. 3 (1998).

905 K. Likharev, “Dynamics of some single flux quantum devices: I. Parametric quantron,” *IEEE Trans. Magn.* 13, 242 (1977).

906 M. Hosoya et al., “Quantum flux parametron: a single quantum flux device for Josephson supercomputer,” *IEEE Transactions on Applied Superconductivity*, vol. 1, no. 2, pp. 77-89, June 1991. doi: 10.1109/77.84613

907 K.E. Drexler, “Molecular machinery and manufacturing with applications to computation,” Ph.D. thesis, Massachusetts Institute of Technology, 1991. http://hdl.handle.net/1721.1/27999

908 S.G. Younis and T.F. Knight, Jr., “Practical implementation of charge recovering asymptotically zero power CMOS,” *Proceedings of the 1993 Symposium on Research on Integrated Systems*, pp. 234-250. MIT Press, 1993.

909 C. S. Lent, , P. D. Tougaw, W. Porod, and G. H. Bernstein, “Quantum cellular automata,” *Nanotechnology* vol. 4, no. 1 (1993): 49.

910 V. Anantharam, M. He, K. Natarajan, H. Xie, and M.P. Frank, “Driving Fully-Adiabatic Logic Circuits Using Custom High-Q MEMS Resonators,” ESA/VLSI, pp. 5-11. 2004.

911 Takeuchi, Naoki, Dan Ozawa, Yuki Yamanashi, and Nobuyuki Yoshikawa, “An adiabatic quantum flux parametron as an ultra-low-power logic device.” *Superconductor Science and Technology* 26, no. 3 (2013): 035010.

912 R.C. Merkle, R.A. Frietas, Jr., T. Hogg, T.E. Moore, M.S. Moses, J. Ryley, “Mechanical Computing Systems Using Only Links and Rotary Joints,” *arxiv preprint* arXiv:1801.03534 [cs.ET], 2018.

913 T. Hogg, M. S. Moses, and D. G. Allis, “Evaluating the friction of rotary joints in molecular machines,” *Molecular Systems Design & Engineering*, vol. 2, no. 3, May 2017. DOI: 10.1039/C7ME00021A.

914 Y. S. Ang, S. A. Yang, C. Zhang, Z. Ma, and L. K. Ang, “Valleytronics in merging Dirac cones: All-electric-controlled valley filter, valve, and universal reversible logic gate,” *Physical Review B* vol. 96, no. 24 (2017): 245410.

915 A. Adamatzky, ed., *Collision-Based Computing*, Springer, 2002.

916 K. Yamada, T. Asai, I.N. Motoike, and Y. Amemiya, “On digital VLSI circuits exploiting collision-based fusion gates.” *From Utopian to Genuine Unconventional Computers* (2008): 1.

917 T. Asai, K. Yamada, and Y. Amemiya, “Single-flux-quantum logic circuits exploiting collision-based fusion gates,” Physica C, vol. 468, 2008.

918 K. Yamada, T. Asai, and Y. Amemiya, “Combinational logic computing for single-flux quantum circuits with asynchronous collision-based fusion gates,” in 23rd Int’l. Tech. Conf. on Circuits/Systems, Computers, and Communications (ITC-CSCC 2008) (pp. 445–448), 2008.

919 Q. P. Herr, J. E. Baumgardner, and A. Y. Herr, “Method and apparatus for ballistic single flux quantum logic,” U.S. Patent No. 7,782,077, issued Aug. 2010.

920 K. D. Osborn, “Reversible computation with flux solitons,” *United States Patent 9812836 B1*, Nov. 2017.

921 M. P. Frank, “Asynchronous Ballistic Reversible Computing.” in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, pp. 172-179, 2017.

922 W. Wustmann and K. Osborn, “Reversible Fluxon Logic: Topological particles enable gates beyond the standard adiabatic limit,” abstract, *Bulletin of the American Physical Society* (2018). To be presented at the APS March meeting, Los Angeles, March 5–9, 2018. *Arxiv preprint* arXiv:1711.04339 [cond-mat.supr-con].

923 W. Wustmann and K. D. Osborn, “Reversible Fluxon Logic: Topological particles allow gates beyond the standard adiabatic limit,” *arXiv:1711.04339 [cond-mat.supr-con]*, Nov. 2017.

924 K. D. Osborn and W. Wustmann, “Ballistic Reversible Gates Matched to Bit Storage: Plans for an Efficient CNOT Gate Using Fluxons,” *Reversible Computation*, 2018, pp. 189–204. doi:10.1007/978-3-319-99498-7_13

925 M. P. Frank, R. M. Lewis, N. A. Missert, M. A. Wolak, and M. D. Henry, “Asynchronous Ballistic Reversible Fluxon Logic,” *IEEE Trans. on Appl. Supercond.*, 2018 (currently in peer review).

926 M.P. Frank, “Feasible demonstration of ultra-low-power adiabatic CMOS for cubesat applications using LC ladder resonators,” presentation, *Tenth Workshop on Fault-Tolerant Spaceborne Computing Employing New Technologies*, Albuquerque, NM, May 2017. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Space-computing-final12-2up.pdf

927 N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, “Simulation of sub-k_BT bit-energy operation of adiabatic quantum-flux-parametron logic with low bit-error-rate,” *Appl. Phys. Lett.*, vol.103, 62602 (2013). doi:10.1063/1.4817974

928 N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Reversible logic gate using adiabatic superconducting devices," *Sci. Rep.*, vol.4, 6354 (2014). doi:10.1038/srep06354

929 N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Reversibility and energy dissipation in adiabatic superconductor logic," *Sci. Rep.*, vol. 7, no. 1, p. 75, Mar. 2017. doi:10.1038/s41598-017-00089-9

930 N. Takeuchi and N. Yoshikawa, "Minimum energy dissipation required for a logically irreversible operation," *Phys. Rev. E*, vol. 97, no. 1, p. 012124, Jan. 2018. doi:10.1103/PhysRevE.97.012124

931 N. Takeuchi, Y. Yuki, and Y. Nobuyuki, "Recent Progress on Reversible Quantum-Flux-Parametron for Superconductor Reversible Computing," *IEICE Transactions on Electronics* vol. 101, no. 5, May 2018.

932 Y. Taiki, T. Naoki, Y. Yamanashi, and N. Yoshikawa, "Design and demonstration of reversible full adders using adiabatic quantum flux parameton logic," *IEEE Trans. Appl. Supercond*., 2018 (I assume they submitted a paper that is presently under peer review)

933 J. Ren and V. K. Semenov, "Progress with physically and logically reversible superconducting digital circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 780–786, Jun. 2011. doi:10.1109/TASC.2011.2104352

934 M. Lucci *et al.*, "Low-power digital gates in ERSFQ and nSQUID technology," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 3, pp. 1–5, Apr. 2016. doi:10.1109/TASC.2016.2535146

935 M.P. Frank, "Approaching the physical limits of computing," *35th International Symposium on Multiple-Valued Logic (ISMVL'05)*, 2005, pp. 168-185.

936 M.P. Frank, "Nanocomputer systems engineering," *Proceedings of the 2003 Nanotechnology Conference and Trade Show*, San Francisco, CA, Feb. 2003.

937 V. V. Zhirnov, R. K. Cavin, J. A. Hutchby, G. I. Bourianoff, "Limits to Binary Logic Scaling – A Gedankin Model," *Proc. IEEE*, November 2003.

938 Keyes, R.W, "The evolution of digital electronics towards VLSI," *IEEE Transactions on Electron Devices*, Volume 26, Issue 4, Apr 1979 Page(s):271 – 279.

939 Doug Matzke, "Will Physical Scalability Sabotage Performance Gains?" *IEEE Computer*, Volume 30, Issue 9, September, 1997, Page: 37 – 39.

940 J. Welser and K. Bernstein, "Challenges for Post-CMOS Devices & Architectures," *IEEE Device Research Conference Technical Digest*, Santa Barbara, CA, Jun 2011, pp. 183-186.

941 K. Bernstein, R.K. Cavin, W. Porod, A. Seabaugh, and J. Welser, "Device and Architecture Outlook for Beyond CMOS Switches," *Proceedings of the IEEE Special Issue - Nanoelectronics Research: Beyond CMOS Information Processing*, Volume 98, Issue 12, Dec 2010, pp. 2169-2184.

942 D.E. Nikonov and I.A. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," *IEDM Tech. Dig*., pp. 573–576, Dec. 2012.

943 D.E. Nikonov and I.A. Young, "Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits," *IEEE J. Expl. Sol-State Comp. Dev. & Circ.*, vol. 1, pp. 3–11 (2015).

944 T. N. Theis and P. M. Solomon, "In Quest of the 'Next Switch': Prospects for Greatly Reduced Power Dissipation in a Successor to the Silicon Field-Effect Transistor," *Proceedings of the IEEE Special Issue - Nanoelectronics Research: Beyond CMOS Information Processing*, Volume 98, Issue 12, Dec 2010, pp. 2005-2014.

945 C. Pan and A. Naeemi, "An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits," *IEEE J. Exploratory Solid-State Computational Devices and Circuits*, vol. 3, pp 101-110, December 2017.

946 I. Sutherland et al., *Logical Effort: Design Fast CMOS Circuits*, 1st ed. San Mateo, CA: Morgan Kaufmann, Feb. 1999, ISBN: 10:1558605576

947 An extremely valuable collection of different approaches to post-CMOS technology can be found in *Proceedings of the IEEE Special Issue - Nanoelectronics Research: Beyond CMOS Information Processing*, ed. G. Bourianoff, M. Brillouët, R. K. Cavin, III, T. Hiramoto, J. A. Hutchby, A. M. Ionescu, and K. Uchida, Volume 98, Issue 12, Dec 2010.