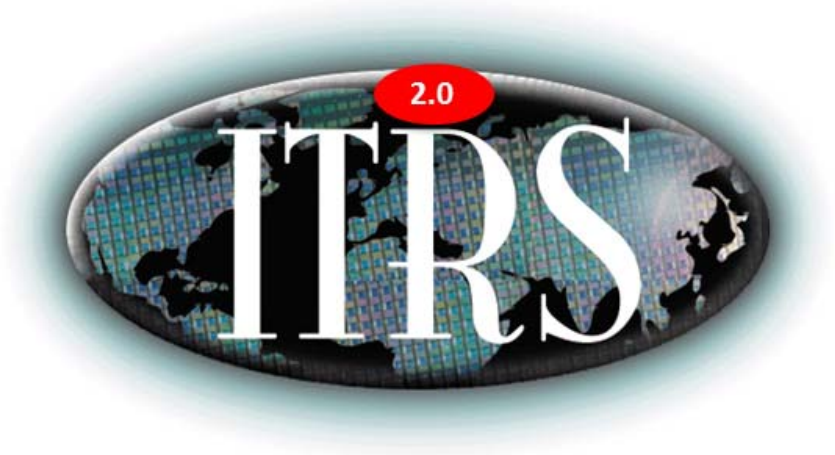


INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2.0

2015 EDITION

BEYOND CMOS

THE ITRS IS DEVISED AND INTENDED FOR TECHNOLOGY ASSESSMENT ONLY AND IS WITHOUT REGARD TO ANY COMMERCIAL CONSIDERATIONS PERTAINING TO INDIVIDUAL PRODUCTS OR EQUIPMENT.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EMERGING RESEARCH DEVICES

1. INTRODUCTION

1.1 SCOPE OF ERD AND BEYOND-CMOS FOCUS TEAM

Continued dimensional and functional scaling[i] of CMOS is driving information processing[ii] technology into a broadening spectrum of new applications. Many of these applications are enabled by performance gains and/or increased complexity realized by scaling. Because dimensional scaling of CMOS eventually will approach fundamental limits, several new alternative information processing devices and microarchitectures for existing or new functions are being explored to extend the historical integrated circuit scaling cadence and sustain performance gain beyond CMOS scaling. This is driving interest in new devices for information processing and memory, new technologies for heterogeneous integration of multiple functions (a.k.a. "More than Moore"), and new paradigms for system architecture. This chapter, therefore, provides an ITRS perspective on emerging research device technologies and serves as a bridge between CMOS and the realm of nanoelectronics beyond the end of CMOS dimensional and equivalent functional scaling. (Material challenges related to emerging research devices are addressed in a complementary chapter entitled *Emerging Research Materials*.)

An overarching goal of this chapter is to survey, assess and catalog viable new information processing devices and system architectures for their long-range potential, technological maturity, and to identify the scientific/technological challenges gating their acceptance by the semiconductor industry as having acceptable risk for further development. Another goal is to pursue long-term alternative solutions to technologies addressed in More-than-Moore (MtM) ITRS entries.

This is accomplished by addressing two technology-defining domains: 1) extending the functionality of the CMOS platform via heterogeneous integration of new technologies, and 2) stimulating invention of new information processing paradigms. The relationship between these domains is schematically illustrated in Figure ERD1. The expansion of the CMOS platform by conventional dimensional and functional scaling is often called "More Moore". The CMOS platform can be further extended by the "More-than-Moore" approach, which was first introduced into ERD chapter in 2011. On the other hand, new information processing devices and architectures are often called "Beyond CMOS" technologies and have been the main subjects of this chapter. The heterogeneous integration of "Beyond CMOS", as well as "More-than-Moore", into "More Moore" will extend the CMOS platform functionality to form ultimate "Extended CMOS".

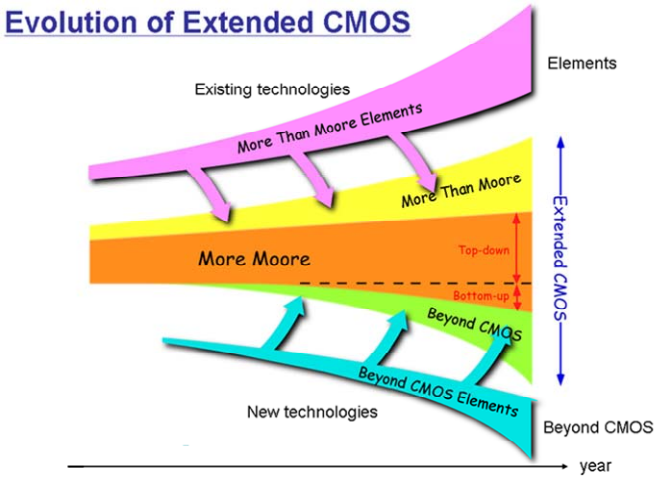


Figure ERD1 Relationship of More Moore, More-than-Moore, and Beyond CMOS (Courtesy of Japan ERD).

[i] *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.*

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

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, each directed to developing a new information processing technology. These communities include universities, research 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) More-than-Moore device technologies, 4) emerging research architectures, and 5) assessment of emerging memory and logic devices. Some detail is provided for each entry regarding operation principles, advantages, technical challenges, maturity, and current and projected performance. Also included is a device 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 memory device section is expanded to include a new technology entry: novel STTRAM devices. “Molecular memory” has been removed as a technology entry and included in the discussion of ultimately scaled memory devices. “Macromolecular memory” is included in the category of ReRAM, defined by polymer-based materials. The logic device section is organized based on state variables and novelty of materials/structures. “Piezotronic transistor” is added as a new entry. “Atomic switch” is removed as a logic device entry but included in the discussion of CBRAM. “BisFET” and “Excitonic FET” are removed in this version due to limited progress. The “More-than-Moore” section introduces new discussions on devices for hardware security. Devices with learning capabilities continue to be included. Emerging devices for RF applications are included in the discussion of low-power sensor nodes for growing applications of Internet of Things (IoT). Finally, the critical assessment section continues to cover a survey-based benchmark and quantitative benchmarks reported in literature to provide a balanced assessment of these emerging devices. The chapter ends with a brief section of a set of fundamental principles that will likely govern successful extension of information processing technology substantially beyond that attainable solely with ultimately scaled CMOS. Table ERD1 summaries the in and out transition of emerging research devices in this chapter.

Table ERD1 Transition Table for Emerging Research Devices.

1.2 DIFFICULT CHALLENGES

1.2.1 Introduction

The semiconductor industry is facing three classes of 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 information processing substantially beyond that 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 third class is to invent and reduce to practice long-term alternative solutions to technologies that address existing MtM ITRS topical entries. These difficult challenges, all addressing the long-term period of 2020 – 2030, are presented in Table ERD1.

Table ERD2 Emerging Research Devices Difficult Challenges

Difficult Challenges – 2020– 2030	*Summary of Issues and opportunities*
Scale high-speed, dense, embeddable, volatile/nonvolatile memory technologies to replace SRAM and FLASH in appropriate applications.	SRAM and FLASH scaling in 2D will reach definite limits within the next several years (see PIDS). These limits 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 nonvolatile memories. The desired material/device properties must be maintained through and after high temperature and corrosive chemical processing. Reliability issues should be identified & addressed early in the technology development.
Extend CMOS scaling	Develop 2nd generation new materials to replace silicon (or InGaAs, Ge) as an 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 for technology scaled to 2020 and beyond. 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 & addressed early in this development.
Extend ultimately scaled CMOS as a platform technology into new domains of application.	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.
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 eventually 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 & addressed early in the technology development.
Invent and reduce to practice long-term alternative solutions to technologies that address existing MtM ITRS topical entries.	The industry is now faced with the increasing importance of a new trend, "More than Moore" (MtM), where added value to devices is provided 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.

1.2.2 Device Technologies

Difficult challenges gating development of emerging research devices are divided into three parts: those related to memory technologies, those related to information processing or logic devices, and those related to 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 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.

A related challenge is to sustain scaling of CMOS logic technology beyond 2020. 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, “More than Moore” (MtM), where added value to devices is provided by incorporating functionalities that do not necessarily scale according to "Moore's Law“. A MtM section was first introduced into the ERD chapter in 2011 with the initial coverage on wireless technologies. The 2013 version added devices with learning capabilities. This version further expands the coverage to emerging devices for hardware security. Traditionally, the ITRS has taken a “technology push” approach for roadmapping “More Moore”, assuming the validity of Moore’s law. In the absence of such a law, different methodologies are needed to identify and guide roadmap efforts in the MtM-domain.

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.

1.2.3 Materials Technologies

The most difficult challenge for Emerging Research Materials 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 Emerging Research Materials.

1.3 NANO-INFORMATION PROCESSING TAXONOMY

Information processing systems to accomplish a specific function, in general, require several different interactive layers of technology. The objective of this section is to carefully delineate a taxonomy of these layers to further distinguish the scope of this chapter from that of the Emerging Research Materials chapter and the Design chapter.

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. As shown in Figure ERD2 below, 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 schematic 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. 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 the domain of the ERM 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 ERD chapter. The *architecture* plane 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. Subclass 1) is within the scope of the ERD chapter, and subclasses 2) and 3) above are within the scope of the Design chapter.

The elements shown in the red-lined yellow boxes represent the current CMOS platform technology that is based on electronic charge as a binary computational state variable. This state variable serves as the foundation for the von Neumann computational system architecture. The other entries grouped in these five categories summarize individual approaches that, combined in some yet to be determined highly innovative fashion, may provide a new highly scalable information processing paradigm.

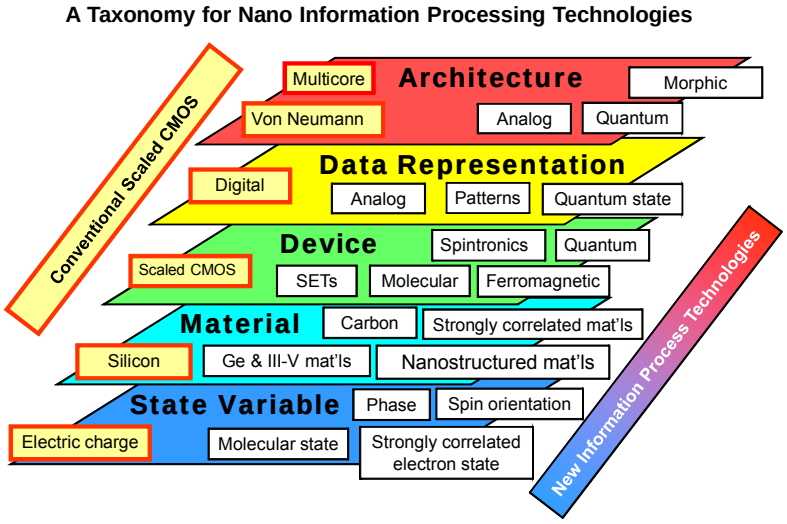


Figure ERD2 A Taxonomy for emerging research information processing devices (The technology entries are representative but not comprehensive.)

2. EMERGING MEMORY DEVICES

The emerging research memory technologies tabulated in this section are a representative sample of published research efforts (circa 2014 – 2015) describing attractive alternative approaches to established memory technologies.[iii] 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).

Figure ERD3 is a taxonomy of the prototypical and emerging memory technologies discussed in Table ERD3 and ERD4. 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 similar to the familiar silicon memory chip.

Because each of these new approaches attempts to mimic and improve on the capabilities of a present day memory technology, key performance parameters are provided in Table ERD3 for existing baseline and prototypical memory technologies. These parameters provide relevant benchmarks against which the current and projected performance of each new research memory technology may be compared.

The emerging research memory technology entries in the current version of the roadmap differ in several respects from the 2013 edition. These changes are described in the Transition Table for emerging research memory devices (Table ERD4). Major changes are as follows. First, due to significant research activity in the area of *Redox* memory (ReRAM) section has been expanded into a separate table (ERD5b) with subcategories describing four distinct device categories. Second, *Nanoelectricalmechanical* Memory has been removed from Table ERD5a and an entry for *Carbon* memory added. The motivations for these changes are given in Table ERD4.

This memory portion of this section is organized around a set of five technology entries shown in the column headers of Table ERD5a and four in ERD5b. 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, three 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 last row in Tables ERD5a and ERD5b contains the number of papers on the particular device technology published in the last two years. It is meant to be a gauge of the amount of research activity currently taking place in the research community and was a primary metric in determining which of the candidate devices were included in this table. 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 as well as 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 A_{cell}=(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 ERD6 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 ERD7 displays

[iii] *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.*

benchmark parameters required for a 2-terminal select device and Table ERD8 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 nonvolatile memory technology that can be manufactured at a very low cost per bit. Table ERD9 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 4.1.4. 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 ERD10 illustrates the potential for storage-class memory applications of a number of *prototypical* memory technologies (Table ERD3) and emerging research memory candidates (Table ERD5). The table shows qualitative assessments across a variety of device characteristics, based on the target system parameters from Table ERD9. These tables are discussed in more detail in Section 4.1.4.

Table ERD3 Current Baseline and Prototypical Memory Technologies.
Table ERD4a Emerging Research Memory Devices—Demonstrated and Projected Parameters.
Table ERD4b Emerging Research Memory Devices—Redox RAM Demonstrated and Projected Parameters.

2.1 MEMORY TAXONOMY

Figure ERD3 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. Nonvolatile 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. Nonvolatile memory technologies are further categorized by their maturity. Flash memory is considered the baseline nonvolatile memory because it is highly mature, well optimized, and has a significant commercial presence. Flash memory is the benchmark against which prototypical and emerging nonvolatile 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. These prototypical technologies are covered in Table ERD3, and in the PIDS Chapter. The focus of this section is Emerging Memory Technologies. These are the least mature memory technologies in Fig. ERD3, but 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.

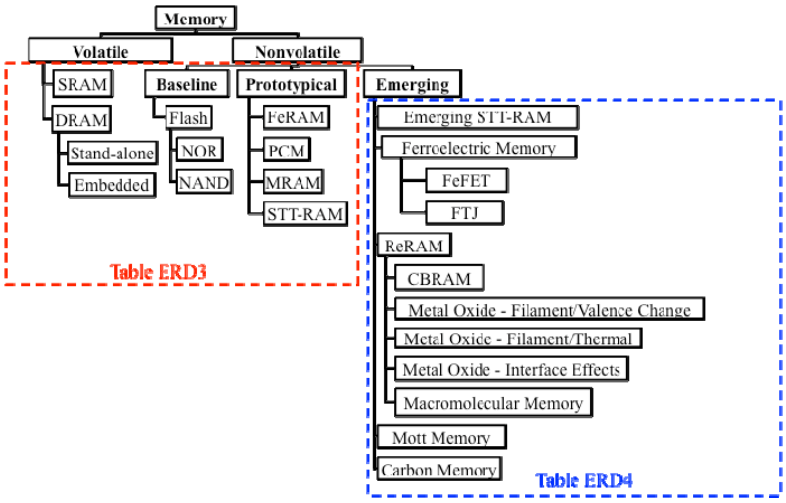


Figure ERD3 Taxonomy of emerging memory devices

2.2 EMERGING MEMORY DEVICES

2.2.1 Redox Memory

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 [2,3,4]. 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. Starting in 2013, ERD has categorized these using the type of switching (bipolar versus unipolar), metal oxide (involving oxygen vacancy motion) versus conducting bridge (involving metal ion motion), and filamentary versus non-filamentary. Most of the literature fits into four categories: electrochemical metallization bridge (EMB), metal oxide-bipolar filamentary (MO-BF), metal oxide-unipolar filamentary (MO-UF), and metal oxide-bipolar nonfilamentary (MO-BN).[5]

The material class for redox memory is comprised of oxides, higher chalcogenides (including glasses), semiconductors, as well as organic compounds including polymers. 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,[6] retention,[7] and endurance[8] are encouraging.

2.2.1.1 Electrochemical Metallization Bridge ReRAM

Electrochemical metallization bridge (EMB) ReRAM (also referred in literature as Conductive Bridge RAM (CBRAM), Programmable Metallization Cell (PMC), or as electrochemical metallization cells) utilizes electrochemical control of nano-scale quantities of metal in thin dielectric films or solid electrolytes to perform the resistive switching operation.[9] The basic EMB 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.[10] 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.

EMB ReRAM is a strong emerging memory candidate primarily due to scalability (~10nm),[11] ultra-low energy operations due to fast read, write and erase times, and low voltage requirements.[12] Maturity of the EMB technology development can be assessed by the fact that many companies are either shipping products based on EMB or in advance stages of commercialization. A recent publications show that CBRAM technology application in various markets including SSDs,[13] embedded NVM,[14] and serial interface nonvolatile memory replacement.[15] In 2012, an EMB based serial NVM replacement product became commercially available.[16] In 2014, a 16 GB ReRAM array based on a CuTe EMB cell was demonstrated.[17,18] Such efforts are critical to identify core technology challenges[19] and also on fundamental materials and mechanisms.[20] Novel applications such as reconfigurable switch[21] and synaptic elements in Neuromorphic systems[22] based on EMB ReRAM are also gaining prominence and are expected to expand the application base for this technology.

As with other filamentary ReRAM technologies, EMB ReRAM is challenged by bit level variability,[19] the random nature of reliability failure such as retention or endurance and random telegraph noise potentially contributing to read disturbs.[23] 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 EMB ReRAM 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 [15,21]. Active participation from semiconductor equipment vendors and material suppliers would assist overcoming manufacturing hurdles rapidly.

2.2.1.2 Metal Oxide-Bipolar Filamentary ReRAM

Metal oxide-bipolar filamentary (MO-BF) ReRAM operates on valence change mechanism of the dielectrics, or redox reaction of the dielectrics between electrodes.[24,25] It typically consists of one or a few layers of insulating materials[26] (i.e. 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 by two different electrodes. The electrode with relatively active metal interacts with dielectric material to generate vacancies of the anions (oxygen or nitrogen etc.) under fabrication or electroforming processes.[27] 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)[28]. Among the mentioned material systems, TaO_x and HfO_x are still the leading candidates of this category due to their overall superior performance and CMOS.

Most MO-BF switches require a one-time electroforming process to establish a filament linking both electrodes. The subsequent switching, operated at significantly smaller voltages, is believed to rupture and recover a portion of the filament in the vicinity of the inert metal electrodes.[29] The electroforming is an electrochemical process of reduction of the oxide or nitride dielectrics.[27,30] A thinner dielectric layer combined with a anion scavenging layer is typically adopted to mitigate the electroforming process and obtain electroforming-free devices.[27]

Since the demonstration of a single crossline HfO_x devices with a 10 nm dimension in 2011[31], scaling to a smaller size was achieved by employing a sidewall electrode in a 1×3 nm^2 cross-sectional HfO_x-based MO-BF device with reasonable performance in terms of both endurance and retention.[32] In terms of repeatability, up to 10^{12} cycles has been demonstrated with $Zr:SiO_x$ sandwiched by graphene oxide layers recently.[33] This is comparable to the previous endurance record with a Ta_2O_{5-x}/TaO_{2-x} structure developed in 2011.[34] The fastest MO-BF was demonstrated on a TaO_x device with a programing duration about 100 ps in 2011.[35] Remarkable progress has been made on improvement of the MO-BF 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 5×10^5 years with an activation energy of 1.6 eV.[36]

Unconventional electrodes (e.g., graphene, carbon nanotube) have been utilized for RRAM, e.g., graphene electrode on HfO_x-based RRAM with low write/erase energy of ~230 fJ/bit[37], graphene/CNT edge electrodes on bi-layer Ta_2O_{5-x}/TaO_y RRAM in a 3D array to explore the limit of RRAM cell size[38]. Notice that carbon nanotube biased AlO_x MO-BF device has exhibited even lower write/erase energy per transition in the ranging from 0.1 to10 fJ per bit.[39]

Large-scale integration of MO-BF switching based on 1T1R schemes has been reported. In 2013, a 32Gb RRAM test chip has been demonstrated using 24 nm CMOS process.[40] An encapsulated cell structure with an $Ir/Ta_2O_5/TaO_x/TaN$ stack was implemented in a 2-Mbit chip at 40nm node in 2014. A 16Mb array of HfO_x/TaO_x bi-layer RRAM with built-in verification function and a capacitive surge current reduction circuit was also reported in 2015[41]. In addition, passive integration of 1S1R scheme has been reported by Crossbar on a 4-Mbit chip but the material stack of the MO-BF switch was not revealed.[42]

Despite the significant advancement of MO-BF ReRAM, there are still a number of technical challenges for commercialization with those MO-BF ReRAM. The current for switching normally reported is still a bit too high for MO-BF devices. In addition, the variability of switching parameters (i.e. switching voltages, resistance distributions) is associated with the stochastic nature of the filament formation and rapture, which has a negative impact on applications like multilevel cell memory. But on the other hand, such variability of switching processes provides a way to realize stochastic learning rule. Moreover, passive integration of MO-BF switches requires either external selecting devices or built-in IV nonlinearity. An ideal external selecting device with good scalability, fast sweep, suitable operation current, and good ON/OFF ratio still remains to be demonstrated.

2.2.1.3 Metal Oxide-Unipolar Filamentary ReRAM

Metal Oxide - Unipolar filamentary (MO-UF) ReRAM is another resistive switching device, also referred to in the literature as thermochemical memory (TCM)2 due to its primary switching mechanism. The device structure is 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 was unipolar in nature (see reference [43] 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 has 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[44]). Only in symmetric structures (e.g. Pt/HfO_2/Pt), a-polar behavior can be obtained, where SET and RESET are occurring irrespective of voltage polarity[45].

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

The unipolar character of the switching indicates that drift (of charged defects) in an electric field does not play a role (as it does in bipolar switching resistive memory), but that thermal effects probably dominate.[46,47] On the other hand, polarity effects indicate anodic oxidation (e.g. at Ni or Pt electrodes) is responsible for RESET.[44] 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.[48,49,50,51] An interesting work reporting on the Scaling Effect on Unipolar and Bipolar Resistive Switching of Metal Oxides was recently published[52].

A unipolar switching device is seen as advantageous for making scaled memory arrays, as it enables the use of simple selector devices as a diode that can be stacked vertically with the memory device in a dense crossbar array[43]. Also, 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 plus side, unipolar switching mode typically shows a higher on/off resistance ratio. On the minus 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 is reported in references [53], [54], [55], [56], [57], and [58]. Reference 53 shows an endurance of over 10^6 cycles with a resistive window of over 5 orders of magnitude (and a reset current ~1mA). References 54, 55, and 56 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.[56] 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. Reference 57 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. Reference 58 reports on a novel approach using thermal assisted switching to lower the switching current.

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

Major challenges to be resolved are the high switching current that seems inherently to the unipolar operation mode. As shown in references 54, 55,and 56, reset currents less than 100 µA are achieved but need further reduction to less than 10 µA. Recently, a possible solution incorporating thermally assisted switching has been presented.[58]

2.2.1.4 Metal Oxide-Bipolar Non-Filamentary ReRAM

The *Bipolar Non-Filamentary* ReRAM 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:SrTiO3.[59] In contrast to *Unipolar* and *Bipolar Filamentary* ReRAM 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* ReRAM is *uniform*. Depending on 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 with field 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 ReRAM 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 512x4096 bit. No external circuitry is needed for current control during set operation. A continuous transition between on and off state 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 has 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 R&D 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.[60,61,62] Within the Bipolar Non-Filamentary RRAM device family the Tunnel RRAM is probably the furthest developed technology. Single device functionality is demonstrated down to 30nm. 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/RRAM functionality are verified for 200nm devices on 200mm 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 RRAM 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.[63] 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 have to be chosen to maximize the ratio between set/reset and retention times. 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.

Memory cells with a conductive perovskites 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 stroichiometry in the cell, oxygen loss in the presence of reducing atmospheres like H2, 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.

2.2.1.5 Macromolecular (Polymer) ReRAM

Macromolecular, also referred to as polymer ReRAM is a subcategory of redox RAM which focuses structures incorporating a layer of polymer. Via the two terminals, the resistance of the memory is modified and also read-out, using pulsed voltage. The polymers contain mainly carbon atoms and are largely amorphous. Non-volatile memory effects with

a non-destructive read have been reported for a surprisingly large variety of polymeric materials and blends of polymers with nanoparticles. 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 specificed. 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. For this reason, macromolecular ReRAM is not tracked in the ERD 4 Tables.

Depending on the structure of the polymer, a variety of mechanism 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. A number of studies have indicated the importance of certain interfaces and interlayers of the polymers with other conducting and semiconducting materials. In some cases, the mechanism and its relation between polymer structure seems obvious, but in many cases the mechanism of operation is largely unknown. In summary the macromolecular memory category encompasses a large variety of mechanisms.

Stability of the memory states at high temperatures (85^{o}C, 2×10^{4} s) has been demonstrated.[64,65] Programming at very low power (70 nW) has been realized.[65] Assuming a 15 ns switching time determined for the same system, one might acheive 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)[66]. Downscaling of polymer resistive memory cell to the 100 nm length scale has been reported.[67] At this length scale, integration of memory cells into an 8×8 array could be shown. Polymer memory cells on flexible substrate have been shown.[68] For amorphous carbon, downscaling to nanometer sized cell has been published (1×10^{3} nm^{2} [69]). 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.[70] 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 no comprehensive picture for the operation has 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.[71,72]

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 benchmarking the operation of the macromolecular materials. These research steps would be crucial for establishing and securing the collaboration of 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[73] 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.

2.2.2 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 the PIDS section and Table ERD3). 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. Current status and key challenges of these emerging ferroelectric memories will be assessed within this section.

2.2.2.1 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 remanent polarization P_r [74]. 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[75,76], but rather due to their low E_c (SBT: 10-100 kV/cm, PZT: ~50 kV/cm, summarized in[77]) 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[78]. This CMOS-compatible material innovation enabled the demonstration of a FeFET technology scaled to the 28 nm node utilizing a conventional HKMG technology already used in high volume production[79]. 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[80]. 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[81]. 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 and the possibility to use ultra-thin interfacial layers as well as by the depolarization resilience of the high E_c[77,82]. 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, 10-year data retention and endurance in the range of 10^{12} switching cycles have been demonstrated for FE-HfO_2-[78,79,83,84] as well as for perovskite-based FeFETs[81,85,86]. 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[87]. 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[83]. 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)[74], a reduced trap density at the interfaces[87], an optimized capacitive voltage divider by area scaling in the MFMISFET approach[88] or the realization of a MFSFET device by

implementing recent breakthroughs in the epitaxial growth of FE-HfO_2[89]. Despite promising results obtained for perovskite-based FeFET devices implemented into 64Kb NAND-Arrays at a feature size of 5 µm[86], 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. Most recently first data showing a clear low and high V_T separation in a sub-kB FE-HfO_2-based FeFET array with L_G: 34 nm nm has been demonstrated[90]. 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.

2.2.2.2 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 ([91] 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.[92]. A similarly high TER was demonstrated by Wen and co-workers[93] 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[94]. 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[95]. 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 5.2.2.1. The highly energy efficient electric field switching, common to all ferroelectric memories, enables fast (10 ns[96]) and low voltage (1.4 V[95]) 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[97]). 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 like for 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[98]. Simply based on the lateral dimensions of the FTJ element (assuming unlimited scalability of the selector), scaling below 50 nm^2 [97] and based on PFM data[99] most likely beyond 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/cm² current density at 300 nm feature size has been achieved by Bruno et al.[100] utilizing low resistivity nickelate electrodes. Based on these results maintaining 10 µA read current for feature sizes <100 nm appears possible.

FTJ based memories are currently at a very early development stage. 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[101]. 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 ([102] 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) as well as 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-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. However, polycrystalline ferroelectrics, unless very thin, tend to have other leakage current contributions overwhelming the desired tunneling current as well as an increased variability of FE-properties at the nanoscale. In this context it is worth noting that the CMOS-compatibility and advanced thickness scalability of FE-HfO_2 [84] as well as recent breakthroughs in its epitaxial growth[89], might yield great potential for the manufacturability of competitive FTJs.

2.2.3 Carbon Based Memory

Recently, various allotropes of carbon, such as amorphous carbon (α-C) and diamond-like (DLC) carbon, carbon nanotubes and graphenic carbon were explored for memory applications[103]. The device structure consists of an electrode/carbon/electrode capacitor structure, whose resistance can be changed by appropriate current pulses. One of the proposed switching mechanism is creation and destruction of conductive sp2-bonds in an otherwise insulating carbon matrix consisting of sp3-bonds[104]. Another switching mechanism is creation and closing of nanogaps. This can occur in low mass density carbon, like carbon nanotube ribbons or foams and in open systems like graphene or graphitic break-junctions which are operated in air or even vacuum[105]. In both cases the structures are not embedded by a protecting insulating layer but are held in air or in a vacuum, like a probe chamber, SEM or TEM. Even an unknown atmosphere is possible for integrated memory cells produced by the encapsulation process. By applying a high enough current density, a nanogap is created by evaporating carbon into the vacuum or hollow structure of the low density carbon material. Joule heating originating from the field emission current in the nanogap will cause atom diffusion and rearrangement of carbon, eventually supported by trapping of hydrocarbons from ubiquitous gas background contamination[105]. Recent Raman studies support the creation of carbon chains in the gap region[106,107].

Not considered here as carbon-based memories are experiments, which are diffusing oxygen or metal ions through insulating phases of carbon[108] or graphene oxide[109] to form a redox-based resistive memory[105].

Recently reported unipolar switching in 22 nm thick DLC relies on single cells with 360 nm diameter[104]. The SET pulse was 50 ns for 1.2V and pulses smaller than 25 ns at 0.4 V were needed for RESET. ON resistance varied between 2 kΩ - 10 kΩ. Switching currents, retention and endurance were not reported. Scaling of the cell diameter to 120 nm and SET voltage scaling to 0.8 V by reducing the DLC thickness to 16 nm has been demonstrated[110]. Unipolar switching has also been reported in single cells with 18 nm thick a-C:H films with cell diameters of 49 nm[111]. The FORMING voltage was 4V, SET pulses were 30 ns 4 V, and RESET was achieved after 5 ns at 4 V. ON resistance were around 10 kΩ , OFF to ON resistance ratio was 1000, 15 cycles and 16 hours retention have been demonstrated. For nanogap-based cells, the most advanced demonstration features a 4 Mbit chip established with carbon nanotube ribbons (NRAM)[112,113], where 108 write cycles in bipolar mode have been demonstrated on 256 bytes[113] with 114nm bottom electrode, but operating voltages and currents were not disclosed. On individual cells, with external transistors as current limiter, an endurance of 1012 write cycles with 20 ns pulses of 3.5 V for RESET and −2.4 V for SET have been demonstrated[114,115]. Data retention has been demonstrated with a 168 hours 250°C bake[116], again on different samples.

Memory cells integrated with on-chip transistors for current compliance need to be employed to study DLC a-C, and a-C:H based memory cells further and demonstrate the scalability to low current operation. It has already be shown that the reset pulse needs to be extremely short (<20 ns) not to induce strengthening of the filament[104,111]. The voltage drop on the bitline will collapse during a successful RESET in arrays, leaving the cell in the OFF-state with a much higher bias voltage, which could induce a SET operation again. The complete history of FORMING, SET, RESET voltages and currents needs to be revealed in the case of NRAM cells. Nothing is known about the virgin cell state after fabrication and whether the first operation is a SET or a RESET and how much current it will take. The origin of the bipolar operation of the NRAM cell needs to be investigated, especially in the context that water and acid-based spin-on deposition of CNTs can induce to the formation of WOx on the tungsten plug.

2.2.4 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[117]) 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[118]. The theory also predicted that the resistive switching due to the interfacial Mott transition has a nonvolatile-memory functionality, because the Mott transition is a first-order phase transition due to its nature[117]. 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.

The Mott transition induced by an electric field or carrier injection has been experimentally demonstrated in a correlated electron material of $Pr_{1-x}Ca_xMnO_3$.[119] After this demonstration, two-terminal devices such as switches and memories have been intensively studied with using such correlated electron oxides as $Pr_{1-x}Ca_xMnO_3$,[120,121] VO_2,[122,123,124] $SmNiO_3$,[125] NiO,[126,127] Ca_2RuO_4,[128] and NbO_2,[129,130] and using Mott-insulator chalcogenides of AM_4X_8 (A=Ga, Ge; M=V, Nb, Ta; X=S, Se).[131,132,133,134] In addition to these inorganic materials, recently, reversible and nonvolatile 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).[135]

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.[129,130] The programing and read voltages reported so far are < 2 V and < 0.2 V, respectively.[127] The nonvolatile resistive switching of AM_4X_8 single crystals was induced by the electric field of less than 10 kV/cm.[131,132,133,134] 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 nonvolatile 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.[131,132,133,134,135] In addition, the NbO_2-Mott memristors and VO_2-based devices are volatile switch.[122,123,124,129,130] 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 transition element.[117] However, because of disorders, defects, and spatial variation of chemical composition, 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[118] and the formation of conductive filament generated by local Mott transition.[131,133,134] However, a thorough understanding of the mechanism has not been achieved yet. Therefore, the elucidation of detailed mechanism is also a major research challenge.

2.2.5 Novel STT-RAM

Recently, some novel mechanisms have been discovered to switch spin-transfer-torque RAM (STT-RAM). Among them, voltage-incuded magnetization switching and spin Hall effects are the most promising candidates.

Voltage pulse induction of magnetization switching has emerged as a candidate for a high efficiency nonvolatile memory.[136] The voltage driven nature of the technology allows us to use high resistance magnetic tunneling junctions and fits to the existing CMOS design. It has the potential to offer ultra-low power, high speed and long endurance memory cells beyond conventional STT-MRAM technology, which is based on a spin-polarized current injection.

The voltage writing technique is based on a magnetic anisotropy change because of application of an electric field at the interface between magnetic metal and an insulator. The change in magnetic anisotropy switches magnetization from perpendicular to in-plane for a quasi-static process. With high-speed pulse application, however, the magnetization undertakes a precessional switching within around 1 nsec.[137] Since the precessional switching is a toggle switching, a read procedure is required before a writing pulse. Alternative method to use voltage effect is a combination with spin-transfer effect. For this case a precise control of device parameters[138] or a special pulse shape[139] is required. Voltage torque MRAM (VT-MRAM) is a very attractive candidate as for a storage class memory, DRAM and/or cash memory. For magnetic cells less than 20 nm in diameter, however, development of a material with a higher voltage effect is essential.

Voltage switching of magnetization requires the effect that cancels the magnetic anisotropy. Since smaller magnetic cell requires larger magnetic anisotropy to keep its retention, it also requires larger voltage effect. Here, both magnetic

anisotropy and voltage effects are proportional to the spin-orbit interaction. Therefore, larger voltage effect is expected for the smaller cells, in which materials with a larger spin-orbit interaction is implemented.

Endurance and temperature range of the VT-MRAM can be as good as STT-MRAM. Retention of the cells used in published papers are not enough. Write energy and current are already smaller than those in STT-MRAM but voltage higher. An integrated VT-MRAM array has not yet been demonstrated.[137]

Huge voltage effects associate with redox reaction, charge trapping, and/or atom migration has been reported.[140] In such cases, the effect can be slow (1 μsec or longer), hysteretic and the endurance can be limited. In addition, smaller pure electronic effects are an alternative mechanism. The largest effect, which is claimed as a purely electronic mechanism, has been observed in Cr/FeB/MgO system.[140] The size of the effect only allows to control 30 nm or larger cells. To switch 1x nm cells, an effect of 3x larger will be required.

VT-MRAM requires a major breakthrough to enable application implementation. Three advances are needed to make the VT-MRAM viable. The first is to create a much larger electronic voltage effect on the magnetic anisotropy. Efforts are needed which explore new combinations of ferromagnetic metal and insulating barrier material. The effect needs to be imporved by a factor 3 or more. The second necessary breakthough is enhancement of the switching speed and endurance of the redox type voltage effect. Finally, a STT-VT hybrid device array should be demonstrated.[138] However, the first pursuit should be to explore methods of having a larger and pure electronics voltage effect. Then, it is also important to make clear error rate of the voltage writing. After that we may start to think about array test designing peripheral circuits.

The spin Hall effect (SHE) refers to the conversion of a charge current to a transverse pure-spin current in non-magnetic materials[141,142]. Its origin stems from the asymmetric deflection of charge carriers via spin orbit coupling. The spin currents generated using the SHE can transfer spin angular momentum to an adjacent magnetic layer through the STT effect allowing for the controllable manipulation of magnetization using proximate electrical currents[143,144,145,146,147]. The effect is distinct from the conventional STT, whereby the spin current is created by passing a charge current through a ferromagnetic layer to spin-polarize the charge currents[148,149].

To date, the materials showing large SHE include Pt, β-Ta, β-W, and oxygen doped tungsten with spin Hall angles, θ_{SH}, of 0.05-0.1[145,150,151], -0.12[143], -0.33[152], and -0.49[153] respectively. However, the largest θ_{SH} to date have been reported in topological insulators[154,155]. Typically high θ_{SH} are observed in highly resistive materials, and a challenge is to look for low-resistivity materials with large θ_{SH} in order to minimize the device power consumption. Furthermore, materials with large θ_{SH} tend to increase the damping of the magnetic material through the spin pumping effect[156]; minimizing this effect would be beneficial for applications as the switching current threshold for spin torque switching is directly proportional to the Gilbert damping constant. It is also important that the interface is transparent to the spin currents to achieve maximum efficiency in spin angular momentum transfer[150,151]. A closely related effect to the SHE is the Rashba effect[157,158,159] that originates from the interfacial electric fields between dissimilar materials. In the reference frame of the moving electron, these interfacial electric fields appear as an effective magnetic field and causes the spins of the electrons to be re-oriented. This effect thus gives rise to a charge to spin current conversion, similar to the SHE.

The STT originating from SHE and Rashba effects are collectively known as spin orbit torques[160]. They have been used as the writing mechanism in magnetic tunnel junction (MTJ) devices[143]. One particular advantage of this writing mechanism is that it does not require large voltages through the MTJ tunnel barrier, which can, over extended time periods, lead to dielectric breakdown of the tunnel barrier. However, it does require a three terminal device, and is limited mostly to in-plane magnetized MTJ devices, although some recent promising studies have shown that through engineering the device structure to produce a field like spin orbit torque or applying an in-plane magnetic field, deterministic switching of perpendicular magnetized devices can be achieved[158]. The SHE- STT can also be used for the current induced manipulation of magnetic domain walls in perpendicular magnetized materials as in racetrack memory devices[146,147].

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[161], 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[162,163]. 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)[164]. Volatile switches can also be used as selector devices. Figure ERD4 shows a taxonomy of memory selector devices. In addition to external selector device, 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.

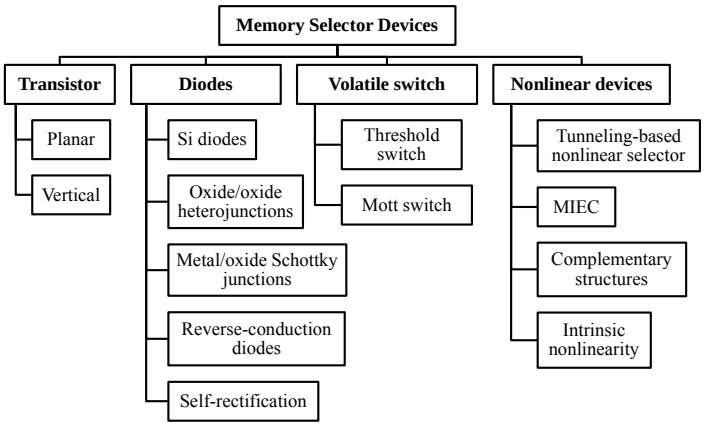


Figure ERD4. Taxonomy of memory select devices

2.3.1 Vertical transistors

Several examples of experimental demonstrations of vertical transistors used as selectors in memory arrays are presented in Table ERD5. While a vertical transistor selector allows for the highest planar array density ($4F^2$), it is challenging to integrate 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 ($4F^2$)[165]; although true $4F^2$ arrays can, in principle, be implemented with 3-terminal selector devices[166].

Table ERD5 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 ~ 1μA (Table ERD6). 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[167]. 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 ERD6 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[168], BARITT diodes[169], reverse breakdown Schottky diode[170].

2.3.2.1 Si diode selector devices

Both single-crystal Si[171] and poly-Si[172,173,174] 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[172]. 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 ON/OFF ratio of 10^8. It is believed that Si diode can be scaled beyond 20nm or 10nm. Poly-Si diode selector devices have been integrated in PCM crossbar arrays, 3D vertical chain-cell type PCM[173], and a 1Gb PCM test chip[174]. 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[175,176,177,178] or Schottky junction[181,182,183,184] diodes may be fabricated at lower temperature 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)[175]. 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[176,177] and with Al_2O_3 antifuse in a one-time-programmable (OTP) memory[178]. Si substrates can be used as a part of heterojunction diodes as demonstrated in n-ZnO/p-Si[179] and n-Ge-nanowire/p-Si diodes[180]. In a TiO_x-based diodes with Pt electrodes, temperature-dependent current-voltage (I-V) characteristics confirms a Schottky barrier of ~0.55eV at the TiO_x/Pt interface[181]. 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[182]. Another demonstration of Pt/TiO_2/Ti Schottky diodes improved ON current density to ~ 3×10^5 A/cm^2 at 2V on a 4μm^2 area[183]. 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 2V[184]. In addition to oxide Schottky diodes, Si Schottky diodes are also utilized as selector devices, e.g., Al/p-Si[185]. 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[186]. However, the feasibility of the Mott-transition switch 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 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 $SmNiO_3$ [187].

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[188]. Niobium oxide is found to possess both memory switching and threshold switching properties at different compositions, based on which a hybrid memory (W/bi-layer-NbO_x/Pt) was demonstrated in a 1kb array[189]. NbO_x-based selectors have also been integrated with TiO_x/TaO_x based RRAM in crossbar arrays at 5xnm node[190]. In Si-As-Te ternary alloy, the composition (controlled by the sputtering power during deposition) determines the emergence of threshold switching[191]. 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[192, 193]. 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} (0.2V), large on/off ratio (>10^7), fast speed (<10ns), long endurance (>10^9 cycles) and good thermal stability (180°C)[194]. A so-called "FAST" (Field Assisted Superliner Threshold) selector was recently reported, with abrupt switching (<5mV/dec), high on/off ratio (10^7), and long endurance (10^8 cycles)[195]. Unlike other threshold switch selectors, this device doesn't 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[196]. 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[197]. A so-called "varistor" selector device is based on a sandwiched $TaO_x/TiO_2/TaO_x$ structure[198]. 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 (<100nA)[199]. It is integrated with CBRAM in a 1kb 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[200].

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 create vacancies near BE. The hole and vacancy concentrations depend exponentially on the applied voltage. Symmetrical diode-like I-V characteristics is achieved with two inert electrodes. Large fraction of mobile Cu^+ enables high current density (> tens of MA/cm^2)[201]. Endurance above 10^8 cycles has been demonstrated on MIEC devices in small arrays[202]. The MIEC selector devices were also integrated with PCM in a 512kb testing array using 180nm CMOS process[203]. The scalability of MIEC select devices was tested to below 30nm in diameter and below 12nm in thickness[204].

2.3.4.3 Complementary resistive switches

Complementary resistive switch (CRS) provides a self-selecting memory by connecting two bipolar RRAM devices anti-serially[205]. It may be considered as "constructed nonlinearity". Both states "0" and "1" have high resistance in CSR, 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 CSR remains in HRS. Notice that the reading operation is destructive, although non-

destructive readout method was also proposed[206]. CRS has been demonstrated in different resistive switching devices, e.g., Cu/SiO_2/Pt bipolar resistive switches[207], amorphous carbon-based RRAM[208], TaO_x–based RRAM[209], multi-layer TiO_x device[210], HfO_x RRAM[211], ZrO_x/HfO_x bi-layer RRAM[212], Cu/TaO_2 atomic switch[213], Nb_2O_{5-x}/NbO_y RRAM[214], *etc*.

Table ERD7a summarizes experimentally demonstrated parameters of some two-terminal select devices, including diodes, volatile switches, and nonlinear devices. Table ERD7b 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 ERD6. For scaled two-terminal select devices, two fundamental challenges are *contact resistance*[172] and *lateral depletion effects*[215,216]. 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 ERD7a Experimentally demonstrated two-terminal memory select devices.
Table ERD7b Experimentally demonstrated self-selecting memory devices (self-rectifying).

2.4 STORAGE CLASS MEMORY DEVICES

2.4.1 Traditional storage: HDD and Flash solid-state drives

Conventionally, magnetic hard-disk drives are used for nonvolatile 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[217].

Nonvolatile 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 (~15V). 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[218].

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[219]. 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.

2.4.2 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.[220,221] Such a device requires a nonvolatile 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 nonvolatile 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 the more than two orders of magnitude between the latencies of off-chip DRAM (~80ns) and NAND Flash (20µs).

2.4.2.1 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[217]).

2.4.2.2 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 3-20 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[222]. 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.

2.4.3 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[223], and/or (2) Multiple level cell (MLC) techniques.

Table ERD8 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^{O}C should be available.

In order to make overall system power usage (as shown in Table ERD8) 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 ERD8 are the target specifications for M-type SCM devices. Given the faster latency target (which enables coherent access through a memory controller), 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 is sufficient to enable device lifetimes on the order of 3-10 years[224], 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 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 ERD8 Target device and system specifications for SCM.
Table ERD9 Potential of the current prototypical and emerging research memory candidates for SCM applications.

2.4.4 First SCM products reach the market

In July 2015, Intel and Micron jointly announced a new nonvolatile memory technology, called "3D-Xpoint." 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[225,226]. (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 nonvolatile memory device and a nonlinear access device [Micron3DXpoint]. The particular nonvolatile memory device was not specified, other than that it depends on bulk changes of resistance[227], 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[227].

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[220,221]. Furthermore, in a later presentation, a second, "Performance-focused" form of 3D-Xpoint memory was described as being under active development[228]. Compared to the initial "Cost-focused" form of 3D-Xpoint

memory, this variant is said to offer even faster latencies and higher endurance (Figure ERD5). 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 ERD5, 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 1-2 weeks would permit successful recovery of server data, even after a power outage due to a natural disaster or other major event.

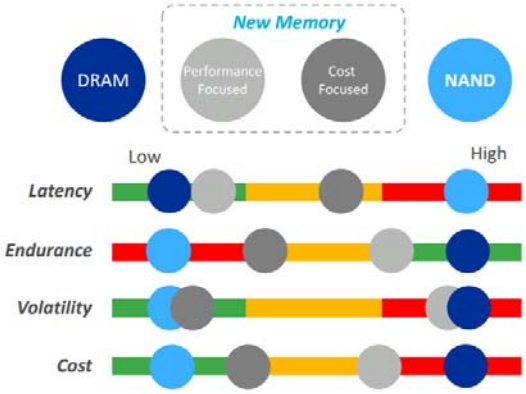


Figure ERD5 Comparison of performance of different memory technologies

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 potentially to be 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.

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 ERD6.

The resulting taxonomy of emerging logic devices is conveyed in the three tables associated with this section. Table ERD10a is titled "Extending MOSFETs to the End of the Roadmap" and tabulates characteristics of the devices in the lower-left quadrant of Figure ERD5. These devices are reviewed below in Section 3.2. Table ERD10b, 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 ERD10c details the "Alternative Information Processing Devices" that are listed in the upper-right quadrant of Figure ERD5 and that are reviewed below in Section 3.4.

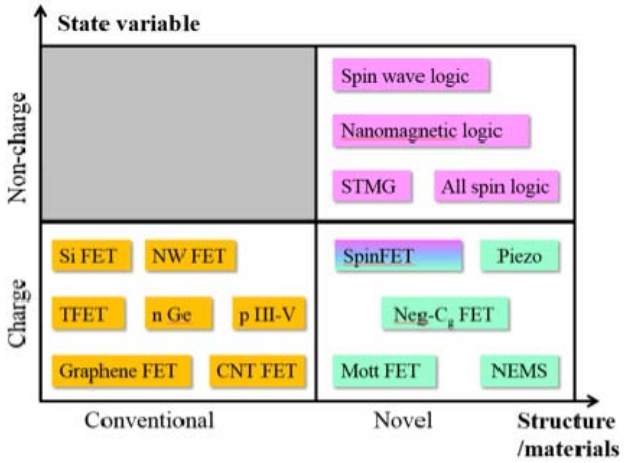


Figure ERD6 Taxonomy of options for emerging logic devices. The devices examined in this chapter are differentiated according to (1) whether the structure and/or materials are conventional or novel, and (2) whether the information carrier is electron charge or some non-charge entity. Since a conventional FET structure and material imply a charge-based device, this classification results in a three-part taxonomy.

Table ERD 10a MOSFETS: Extending MOSFETs to the End of the Roadmap.
Table ERD 10b Charge-Based Beyond CMOS: Non-Conventional FETs and Other Charge-Based Information Carrier Devices
Table ERD 10c: 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 & enhancing device performance in CNT FETs. These include (1) realizing end-bonded contacts having an effective contact length of 0 nm with reasonable performance,[229] (2) detailing the impact of contact scalability in CNT FETs,[230] (3) maintaining performance as the channel length is scaled down to 9 nm without observing short channel effects,[231] (4) fabricating complementary gate-all-around FETs,[232] (5) fabricating an FET with an intrinsic fT of 153 GHz,[233] (6) fabricating CMOS inverters and pass-transistor logic operating at 0.4 V with a non-doped CNT,[234] (7) fabricating a carbon nanotube computer composed of 178 FETs,[235] (8) progress towards reducing the variability in CNT FETs,[236] (9) understanding origins of hysteresis,[237] and (10) fabricating CNT FETs with ON-current of 0.5 mA/μm.[238]

In addition to improvements at the device level, continuous progress has been achieved toward overcoming the dominant material challenges,239 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 also remains 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 with some of the most substantial (and already demonstrated) potential in high-performance, low-voltage, sub-10 nm scaled transistor applications.

3.2.2 Graphene FETs

Graphene materials offer the potential of extremely high carrier mobilities that can exceed those of carbon nanotubes, and also offer the promise of patterning graphene nanoribbons using conventional top down processes. Work on graphene field effect transistors (FETs) is proceeding at a rapid pace, but there are still many issues to address[240]. Beginning with the first description of the electric field effect in graphene in 2004[241], graphene FETs using bottom gating[241], top-gating[242, 243, 244] dual-gating[245, 246] , and side-gating[247] have been demonstrated using exfoliated[248, 249] epitaxial[242, 243, 250] and CVD-grown graphene[251, 252].

Research on graphene FETs started with use of exfoliated graphene to form a transistor channel. Recently, there have been many studies using epitaxial graphene on SiC substrates and CVD-grown graphene. Exfoliated graphene still offers the highest mobility[253, 254, 255] but is not manufacturable. Back-gated graphene FETs with SiO_2 dielectric were typically shown to have field-effect mobilities up to around 10,000 cm^2/Vs[241] (note that although back-gating is not desirable for FET-based circuitry, this mobility value does act as a useful comparison to top-gate mobility values). It has been predicted that the room-temperature mobility of graphene on SiO_2 is limited to ~40,000 cm^2/Vs due to scattering by surface phonons at the SiO_2 substrate[256] . In fact, the highest field effect mobilities were 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, respectively[253, 254, 255] have been measured. Recently, hexagonal boron nitride (hBN), an inert and flat material, was used as a substrate for a graphene channel[257,258], and it was shown that the field effect mobility of such devices can exceed 100,000 cm^2/Vs at room temperature. Epitaxial graphene on SiC has exhibited carrier mobilities as high as 15,000 cm^2/Vs and 250,000 cm^2/Vs at room and liquid helium temperatures, respectively[259, 260]. On the other hand CVD graphene has shown carrier mobilities as high as 25,000 cm^2/Vs at room temperature[261].

As for top-gated graphene-channel transistors, field-effect mobilities are in general lower than those shown above, because the deposition of gate dielectric can degrade the electrical properties of graphene[262]. In order to avoid such degradation, a buffer layer is often used between graphene and high-k materials[263, 264]. It was also shown that SiO_2 layer formed by vacuum evaporation did not degrade graphene channel so much, and field-effect mobilities as high as 5,400 cm^2/Vs were achieved for top-gated transistors[265]. In another case, naturally-oxidized thin aluminum layer was used as a seeding layer of Al_2O_3 by atomic layer deposition and field effect mobilities as high as 8600 cm^2/Vs were achieved[266]. The highest field effect mobility of top-gated graphene transistors was obtained using Al_2O_3 nanowire as gate dielectric, which was as high as 23,600 cm^2/Vs[267]. In this case, exfoliated graphene was used as a channel. Hexagonal boron nitride has also been employed as a top-gate dielectric layer[268], however, the mobility of such devices is not yet as high as the value just shown above. Direct growth of hBN on graphene channels may be necessary for further improvement.

An important limitation of graphene for digital applications is its zero bandgap energy which in turn will result in a very small I_{on}/I_{off} ratio. Hhowever, several methods to open a bandgap have been proposed. One is to build devices with graphene nanoribbons(GNRs)[269, 270, 271, 272, 273]. Carrier transport through a GNR was first demonstrated using nanoribbons fabricated by a top-down approach[269]. An energy gap of ~200 meV was obtained for a GNR with a width of 15 nm. Recently, devices with multipleGNRs with a sub-10 nm half-pitch were fabricated using patterning with directed self-assembly of block copolymers[274]. The on-off ratios of these devices were, however, still ~10 at room temperature. GNRs were also made by several other methods[270, 271, 272, 273]. Specifically, it was recently demonstrated that GNRs with a precisely-controlled uniform width were synthesized by a bottom-up approach using 10,10'-dibromo-9,9'-bianthryl precursor monomers[275]. The band gap of this nanoribbon was estimated to be ~3.7 eV[276, 277]. Formation of wider GNRs with a smaller band gap was also demonstrated using a similar approach recently[278], as well as heterojunctions of GNRs with different widths (bandgaps)[279,280].These bottom-up GNRs are typically formed on a clean Au(111) surface in ultra-high vacuum, and it is still difficult to control their direction and position, which can be an obstacle for device fabrication. Devices using GNRs with a 2-nm width made by sonication of exfoliated graphite in a chemical solution showed an on-

off ratio of 10^7, with a field effect mobility of ~200 cm^2/Vs[270]. The relatively low mobility was considered to be caused by scattering at the edges of GNRs. In fact, recent theoretical studies showed that obtaining smooth edges is essential to obtain good electrical properties[281]. In addition, recent experimental studies suggest that transport in GNRs is greatly affected by defects at the edges and charged impurities[282, 283, 284]. Recently, however, it was shown that GNRs formed on SiC can have a ballistic length greater than 10 μm at room temperature, corresponding to mobility higher than 10^5 cm^2/Vs[285]. Although the width of GNRs in this case is relatively wide, typically 40 nm, the results clearly show a high potential of GNRs. More efforts are, however, required to realize GNR transistors for CMOS.

Another approach to open a band gap in graphene is applying an electric field perpendicular to AB-stacked bilayer graphene[286, 287, 288, 289, 290]. In fact, a transport gap of 130 meV was obtained at an electrical displacement of 2.2 V nm^{-1}, providing an on-off ratio of ~100 at room temperature[290]. The on-off ratio is not yet as low as theoretical predictions. In fact, it was recently pointed out that a small twist of AB-stacking bilayer can cause the bandgap to disappear[291], which is a serious problem. Alternatively, a band gap can also be created by molecular doping on one side or both sides of bilayer graphene[292, 293, 294, 295, 296]. It is predicted that a band gap exceeding 300 meV is formed by sandwiching bilayer graphene between n-type-dopant molecules and p-type-dopant molecules[295]. A transport gap can also be obtained by making holes in graphene, namely by making graphene nanomesh[297, 298, 299]. The transport gap can be controlled by the neck (distance between holes) width. An on-off ratio up to ~100 was obtained at room temperature using this method. Recently, it was demonstrated that a transport gap is created in graphene by introducing low-density defects with helium ion irradiation[300, 301]. An on-off ratio of ~100 was obtained at room temperature. The approaches introduced here have not yet provided an on-off ratio large enough for logic applications; therefore, more efforts are required for this approach to be used in CMOS applications.

An important application space for graphene may be RF with discrete elements and high linearity requirements. There have been many studies aiming at such high-frequency applications[249, 252, 302]. A unity current gain cut-off frequency of 427 GHz has been obtained by a self-aligned- processe using exfoliated graphene with a channel length of 67 nm[303]. Cut-off frequencies as high as 350 GHz and 300 GHz have also been obtained for devices using epitaxial and CVD graphene, respectively (channel length: 40 nm)[304]. Achieving a high maximum frequency of oscillation is the next important step for realizing RF applications.

3.2.3 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[305]. They may be composed of a wide variety of materials, including silicon, germanium, various III-V compound semiconductors (GaN, AlN, InN, GaP, InP, GaAs, InAs), II-VI materials (CdSe, ZnSe, CdS, ZnS), as well as semiconducting oxides (In_2O_3, ZnO, TiO_2), etc.[306] Importantly, at low diameters, these nanowires exhibit quantum confinement behavior, i.e, 1-D ballistic conduction[307], and match well with the gate-all-around structure that may permit 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 methods. The first method is litho&etch, by which semiconducting channels are formed through lithography and etch, followed by a printing or stamping process[iii]. The second is catalyzed chemical vapor deposition[308,309]. In particular, the vapor-liquid-solid (VLS) growth mechanism has been used to demonstrate a variety of nanowires, including core-shell and core-multishell heterostructures[310,311]. Heterogeneous composite nanowire structures have been configured in both core-shell and longitudinally segmented configurations using group IV and compound materials. The longitudinally segmented configurations are grown epitaxially so that the material interfaces are perpendicular to the axis of the nanowire. This allows significant lattice mismatches without significant defects. Vertical transistors have been fabricated in this manner using Si[312], InAs[313,314] and ZnO[315] , with quite good characteristics. Core-shell gate-all-around configurations[316] display excellent gate control and few short channel effects.

Circuit and system functionality of nanowire devices have been demonstrated, including vertical InAs MOSFETs with 103 GHz switching speed[317], a down-conversion mixer based on vertical InAs transistors that showed cut-off frequency of 2 GHz[318], and extended, programmable arrays ("tiles") of nonvolatile nanowire-based flash memory that are used to build circuits such as full-adder, full-substratcor, multiplexer, demultiplexer, clocked D-latch and finite state machines[319,320]. The measured operating speed of these various nanowire test circuits was limited by off-chip interconnect capacitance and thus did not achieve the THz operation that is predicted to be the intrinsic capability of nanowire devices[321,322]. Nearly ideal subthreshold swing of 60 mV/dec has been obtained in gate-all-around nanowire

transistors for both n and p type devices[323]. Intrinsic switching energy as low as 2×10^{-17} J has also been demonstrated[7324], although it is still ~ 3 orders of magnitude higher than predicted theoretically.

Despite the promising results that are mostly obtained from academic research labs, nanowire transistors still face significant challenges before commercialization is feasible. For both nanoimprint and grown nanowires, 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. Due to the high surface-volume ratio, the performance of nanowire transistors likely will be strongly influenced by surface roughness and surface defects, so proper surface treatment and passivation techniques need to be developed. Additionally, although individual devices with gate length < 40 nm has been demonstrated[7], it is unclear whether high density circuits with pitch size comparable to state-of-the-art CMOS can be fabricated. Instead of channel-replacement, nanowires may instead be better suited for alternative architectures such as crossbar memory and logic to fully take advantage of the 1D structure these devices offer. However, effective interface with peripheral circuitry still needs to be developed and series resistance can significantly limit the array size. Another option is to integrate vertical nanowire devices directly on top of CMOS circuits, but this approach still remains to be demonstrated on a large scale.

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[325]. 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[326], 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[326, 327, 328, 329]. Another method of improving hole mobility is the application of uniaxial strain[328, 330], 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[331]. 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[332].

The highest hole mobilities when using compressive strain were found in GaSb/AlAsSb heterostructures; they exhibit a mobility[329] 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[327]. In InSb, the highest mobility[331] 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[329]. 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[331] 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. First method is transfer of nano-ribbon[333], nanowire[14] or film[15]. In 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-k 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[334]. 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 case of nanowire, inverter circuit was fabricated from single nanowire with InAs and GaSb[14]. As another possibility of InAs/GaSb nanowire, it can apply tunnel FET[16]. Rapid melt growth (RMG) method [17] can also supply GaSb thin film on Si, and transistor operation by RMG was reported[18].

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[331]; 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[335]. 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[336]. 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[335]; however, shorter gate lengths were not studied in this report. The shortest reported gate length in a p-

MOSFET is 750 nm[336] 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[337]. 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, [338] is considered as 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, [339] although the electron mobility 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[340, 341] for the counterpart of Ge-pMOSFETs in 7-nm technology node and beyound. Strain technologies to enhance the electron mobility using source-drain stressors[342] and stress liners[343] 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[344] and on poly-Ge layers on SiO_2/Si substrates.[345] 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.[346] 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.[347,348] 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}.[349,350] 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.[340,351] The sub-threshold swing and off-state leakage current of Ge-nMOSFETs are getting better but still slightly worse than those of Ge-pMOSFETs[352], suggesting that the interface state density near the conduction-band edge should be more improved. The drain current for scaled Ge nMOSFETs have been greatly improved during the latest 2 years and reached to ~250 mA/mm for L_g=40 nm device at V_{dd}=1V.[340] 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 $1x10^{19}$ cm^{-3} by means of conventional ion-implantation and RTA techniques. Theoretical calculation predicted the specific contact resistivity required in the ITRS (< $1x10^{-8}$ 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.[351] 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[353]. 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.[354]

The most significant issue for the implementation of Ge-nMOSFETs in the 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 are key technologies. 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 60mV/decade.[355,356,357] In its simplest form, a TFET is a gated reverse-biased p-i-n junction. 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[358,359,360,361]. The thickness of the tunneling barrier is controlled by changing the electric field in the tunneling

junction. Alternatively, it is also possible use energy filtering or density of states switching. If the conduction and valence band don't 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.[357] If phonons or short channel lengths are accounted for, simulated subthreshold swings on the order of 20-30mV/decade are typical.[362] To get high on-state currents a small tunneling barrier, low effective mass and direct gap is needed. Group IV materials with indirect and larger band gaps such as Si[363,364] and Ge[364,365,366] have smaller ON-currents. Group III-V materials like InGaAs,[367] InAs,[368,369] and InSb[370] have higher ON-currents due to their narrower and direct band gaps. The use of staggered and broken-gap heterojunctions, such as AlGaSb/InAs[371] and InAs/GaSb[362], boosts the ON-current further by further lowering the tunneling barrier.

Following the review by Lu and Seabaugh,[355] there are at least 14 reports of subthreshold swings below 60 mV/decade: Most of the results are in group IV materials such as Si,[372,373,374,375,376,377] strained SiGe,[378] Si/Ge,[379] and strained Ge.[380] An $In_{0.53}Ga_{0.47}As/In_{0.7}Ga_{0.3}As$ heterojunction TFET that has a subthreshold swing of 58 mV/decade at 1 nA/μm drain current has been demonstrated.[381] Nanowire III-V TFETs have shown even steeper swings. A InP/GaAs heterojunction[382] has shown 30 mV/decade at 1 pA/μm. The steepest result ever reported is in a Si/InAs heterojunction[383] of approximately 20 mV/decade. However, there are only a few data points defining this result and most of the reports of sub-60-mV/decade swing. The most substantial sub-60-mV/decade measurements are given for Si[372,375,377] and for InGaAs.[381] The highest currents at which a subthreshold swing lower than 60mV/decade has ever been observed is at around 1-10 nA/μm. Even for low power applications, at least 1-10 μA/μm is needed.[384]

The experimental results are far worse than what simulations have been predicting. The best experimental results have typically relied on modulating the tunneling barrier thickness. Unfortunately, it seems that this mechanism typically only gives a step subthreshold swing at low current densities.[356] Consequently, density of states (DOS) switching may be preferred. An ideal DOS switch would switch abruptly from zero-conductance to the desired on conductance thus displaying zero subthreshold swing voltage.[385] Unfortunately, the band edges are not perfectly sharp and so there is a finite density of states extending into the band gap. Optical measurements of intrinsic GaAs imply a band edge steepness of 17meV/decade.[386] However, the electrically measured joint DOS in diodes has generally indicated a steepness >90mV/decade.[387] 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. This can also be seen from the optical absorption in doped GaAs. When GaAs is heavily doped to $10^{20}/cm^3$, the absorption falls of at a rate worse than 60meV/decade.[388] This means that if a tunnel switch is heavily doped, it will be unable to employ the DOS energy filtering mechanism to achieve a subthreshold swing voltage smaller than 60 mV/decade!

In transistors, the situation is even worse due to gate interface traps. The interface trap density in silicon is typically around 10^{10} /cm^2 while the density of states in a quantum well is only 10^{12} to 10^{13} /cm^2. These traps exist in the entire energy range of the band gap. This effectively means there is no band edge for energy filtering. Experimentally this results in a thermally activated tunneling process.[389]

To overcome these challenges we will need better material perfection than ever before. Every defect or dangling bond can create a trap that will ruin the band edge. 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 materials like MoS_2 are potentially a path towards atomic perfection as there will be no dangling bonds at the semiconductor oxide interface.

After reducing the non-ideal band edge DOS, the ideal DOS above the band edge can be optimized by taking advantage of quantum confinement. Quantum confinement in the direction of tunneling, or reduced dimensionality on each side of a pn junction can increase the on-state conductance by 10 to 100X.[356,390]

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" [391], and can be further divided into two categories: the spin-FETs proposed by Datta and Das [392] and spin-MOSFETs proposed by Sugahara and Tanaka. [393] 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. [391, 394] 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 nonvolatile 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. Nonvolatile 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. [394, 395, 396, 397, 398, 399, 400]

The read and write operations of spin-FETs [394, 401] and spin-MOSFETs [394, 402, 403, 404, 405] have not yet been fully experimentally verified. However, there has been important progress in the underlying technologies. Spin injection, detection and manipulation are essential [13]for realizing fully functional spin transistors. One key development has been half-metallic ferromagnetic materials for highly efficient spin injector/detectors. Theories [406, 407, 408, 409] 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 [410] and spin-polarized tunneling. [411, 412, 413, 414, 415, 416, 417, 418] The electrical creation and detection of spin accumulation in n-type and p-type Si were demonstrated using FM/tunnel contacts[411, 412, 414, 415, 416, 27, 28] and FM/Schottky-tunnel-barrier contacts[419] up to room temperature (RT). Electrical spin injection into Si channels has also been demonstrated up to 500 K [413] by using a spin relaxation measurement method (Hanle measurement). A relatively long spin lifetime has been observed in heavily doped Si at RT. [412, 415, 416, 28] Spin transport in SC has been studied intensively in various kinds of tunnel and channel materials. Observation of spin signals in heavily doped Ge[420, 421, 422] and GaAs[423, 424] at RT has also been reported. Another study[425, 426] reported low-resistance spin injection and detection into Si using a graphene and a 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. [427, 428] Recently, basic read operation was demonstrated [15] 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 [15] 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[429, 430, 431, 432, 433] are one of the most promising half-metallic ferromagnetic materials, because they are relatively easy to preparation. 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 [434, 435, 436, 437]. 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. [394, 399, 402, 438, 439] 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-MOSFET 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. [440] 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 [394] in order for the channel to sufficiently induce the Rashba spin-orbit interaction. However, materials with a strong spin-orbit interaction decrease spin life time. The use of a narrow wire channel structure[441, 442, 443] and the so-called persistent spin helix condition. [444, 445, 446, 447, 448] has been proposed in order to increase spin life time. 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 [12, 13] and the pseudo-spin-MOSFET [3948, 49] use spin-transfer-torque (*STT*) switching, similar to the *STT*-MRAM that was commercialized at the end of 2012. [449]

To date, long channel devices with center-to-center spacing between the FM source/drain electrodes larger than l_{FM-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-FMl} < 100 nm are needed.

3.3.2 Negative Gate Capacitance FET

Based on the energy landscapes of ferroelectric capacitors, it has been suggested[450] that by replacing the standard insulator of a MOSFET gate stack with a ferroelectric insulator of appropriate thickness, it should be possible to implement a step-up voltage transformer that will amplify the gate voltage. Such a device is called a negative gate capacitance FET. 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[451] is that it is a relatively straightforward replacement of conventional FET. Thus, high Ion 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 in 2008[452] and subsequently in 2010[453] in a more controlled structure. These experiments thus established the proof of concept of <60mV/decade operation using the principle of negative capacitance.

In addition to the above experiments on polymer based ferroelectrics, negative differential capacitance was demonstrated in a crystalline capacitor stack.[454] 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. Note that capacitance of two serially connected capacitors would normally be smaller than either constituent. It can be larger only if one of those constituent capacitors has a negative differential capacitance. Thus the enhanced capacitance 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.[455] More specifically, it was showed that when a ferroelectric capacitor is pulsed with an input voltage it shows an 'inductance-like' discharging in addition to a capacitive charging.

In the last few years a significant advance has been achieved in ferroelectric materials where it has been shown that it is 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.[456] Using this doped Hf based ALD ferroelectric, a number of recent experiments have demonstrated the negative capacitance effect.[457,458,459] For example, by using $HfZrO_2$ as a gate dielectric <60 mV/decade was demonstrated[459] in FINFETs with Lg=30 nm for both NFET and PFET structures. Future research will have to focus on the understanding of ALD ferroelectric material so that optimized structures can be designed for the ultra-scaled nodes.

3.3.3 NEMS Switch

Micro/Nano-Electro-Mechanical (M/NEM) switches are devices which employ electrostatic force to actuate a movable structure to make a conductive path between two electrodes. Mechanical switches feature two fundamental properties which are unattainable in MOSFETs: *zero off-state leakage* and *zero subthreshold swing*[460]. The first property provides for zero standby power dissipation, while the second property suggests the potential to operate at very low voltage 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[461]. Additional advantages of mechanical devices include robust operation across a wide temperature range[462], immunity to ionizing radiation and compatibility with inexpensive substrates such as glass or even plastic. Since M/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[463], it is conceivable for NEM switches to be implemented using multiple interconnect layers in an advanced back-end-of-line (BEOL) process[464]. Potential applications for mechanical switches in a hybrid NEMS-CMOS technology include CMOS power gating[465], configuration of CMOS FPGAs[466], and non-volatile storage of information in SRAM and CAM cells[464].

M/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.[467] This device is a vertically actuated switch featuring a 30 nm-thick, 300 nm-wide, 1.4 mm-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[468]. 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 M/NEM switches for improved device density, switching delay and switching energy.[469] 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,[460] 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.[461] 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 also can be connected to a logic signal. By incorporating multiple pairs of source and drain electrodes into each gated structure,[470] and/or by partitioning the gate electrode into multiple input electrodes,[471] 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.[472,473]

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 10× reduction in energy per operation as compared with MOSFETs, and can reach clock speeds in the GHz regime.[474] 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.[475]

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 leads 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 kW.[461]) MEM 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).[476] Endurance exceeding 10^{16} on/off cycles is projected for 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.[477,478]

In conclusion, M/NEM switches are intriguing candidates for ultra-low-power applications because they have negligible off-state leakage current and abrupt switching behavior which, in principle, enables 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[479,480]. 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.[481,482] 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[480]. 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)[483]. The phase transition time constant in VO_2 materials is in sub-picosecond range determined by optical pump-probe methods[484]. 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[485]. VO_2 Mott channels have been experimentally studied with thin film devices and the field effect has been demonstrated in preliminary device structures[486, 487, 488]. 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.[489, 490] The conductance modulation happens at a slow speed however, due to the large charging time constants.[491] 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.[491, 492, 493] 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[494] and circuits[495] 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[496] and steep-slope transistors[497].

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 quantitatively 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.[481] 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 possibility of sub-ns switching speeds in ultra-thin device elements in the vicinity of room temperature.[498,499] Such devices could also be of interest to Mott memory[500]. 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[501,502] or cobaltates[503] 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.[504, 505, 506, 507, 508] $SmNiO_3$ 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[509]. 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[510]. 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.

3.3.5 Piezotronic Logic Transduction Device

A promising approach to achieve low voltage transistors is transduction where an electrical signal is transformed into an intermediate form internal to the logic gates and then backs to an electrical signal. This transduction allows for novel switching mechanisms that do not have the subthreshold swing limitations of conventional MOSFETs. Piezotronic transduction logic devices convert an electrical signal to a mechanical stress using a piezoelectric. This is then converted back to an electrical signal using either a piezoresistor or another piezoelectric. The Piezoelectric Transistor (PET) operates by applying a voltage to a piezoelectric pillar which expands and compresses a piezoresistive element.[511,512,513,514,515] The piezoresistor undergoes a metal-insulator transition, lowering the resistance of the channel and turning the device on. Theoretically, the PET is projected to switch with voltages as low as 0.1V, with a 10^4 on/off ratio. By limiting the height to 20 nm mechanical speeds around 5 picoseconds should be achievable. As all materials are in continuous contact it is expected that this structure should not have the same stiction and contact induced wear out mechanisms of a nanomechanical relay.

An alternative piezotronic device uses a piezoelectric transformer in the gate stack.[516] A voltage is applied to one piezoelectric, causing it to expand and squeeze a second piezoelectric. This induces a voltage across the second piezoelectric that can be higher than the applied voltage. Multiple piezoelectric layers can be used to increase the stress in the lowest layer and therefore the voltage amplification. A 12.5X voltage amplification has been theoretically predicted. This voltage amplification reduces the energy to charge wires, but does not reduce the switching energy of the transistor as the transistor itself still operates at the same voltage. AC piezoelectric voltage transformers are commercially available, but a nanoscale DC transformer has not been demonstrated yet.

Proofs of concept Piezoelectric transistors have been fabricated.[512,515] In the initial demonstration, the piezoelectric (PZT) and the piezoresistor (SmSe) were fabricated separately on different substrates and brought together in the test setup. The piezotronic switching mechanism was demonstrated, but 20 volts were required to achieve a 7X on/off ratio.[512] The device could switch at 100 kHz and lasted for $2x10^9$ cycles. A monolithically integrated device was also demonstrated with worse performance (V_g=20V only gave I_{on}/I_{off} = 1.5).[515] The performance was limited by large dimensions and degradation of the piezoelectric polarization. In order to achieve low voltage switching, piezoelectrics such as PMN-PT with larger piezoelectric coefficients need to be integrated in nanoscale devices. There are also significant materials challenges in integrating both piezoelectric and piezoresistive materials in a single process. For instance, PZT, a metal oxide is readily depleted of oxygen which is electrically leaky and SmSe oxides readily and is unstable in air. Consequently encapsulation and compartmentalization becomes a significant design consideration. Furthermore, the devices need to be fully scaled down to 10's of nanometers to achieve low switching voltages. High levels of strain in the scaled devise may also pose reliability challenges.

3.4 BEYOND-CMOS DEVICES: ALTERNATIVE INFORMATION PROCESSING

3.4.1 Spin Wave Device

Spin Wave Device (SWD) is a type of magnetic logic devices exploiting collective spin oscillation (spin waves) for information transmission and processing[517,518]. The basic elements of the SWD include: (i) magneto-electric cells (e.g. multiferroic elements) aimed to convert voltage pulses into the spin waves and vice versa; (ii) magnetic waveguides - spin wave buses for spin wave signal propagation between the magneto-electric cells, (iii) magnetic junctions to couple two or several waveguides, and (iv) phase shifters to control the phase of the propagating spin waves. 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,[517] SWD concept has evolved in different ways encompassing volatile[519] and non-volatile,[520] Boolean[520] and non-Boolean,[521] single-frequency and multi-frequency circuits.[522] The primary expected advantage of SWD over Si CMOS are the following: (i) 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; (ii) power consumption minimization by exploiting built-in non-volatile magnetic memory, and (iii) 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.[523] It is based on $Ni_{81}Fe_{19}$ structure, operates within 1-3 GHz frequency range, and demonstrates ~10 signal-to-noise ratio at Room Temperature demonstrated[523]. The internal delay of SWD is defined by the spin wave group velocity (e.g. 3.1×10^6 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. 0.6MV/m for PMN-PT/Ni).[524] The later translates in ultra-low power consumption (e.g. 1aJ per multiferroic switching).

Recently, it was proposed a new type of SWD - Magnonic Holographic Memory (MHM).[521] The principle of operation of MHM is similar to optical holographic memory, while spin waves are utilized instead of optical beams. The first 2-bit MHM prototype based on yttrium iron garnet structure has been demonstrated.[525] MHM also possesses unique capabilities for pattern recognition by exploiting the correlation between the phases of the input waves and the output interference pattern. Pattern recognition using MHM has been recently demonstrated.[526] The potential advantage of spin wave utilization includes the possibility of in-chip integration with the conventional electronic devices via multiferroic elements. Also, magnonic holograms may have enormous capacity (about 1Tb/cm^2) due to their nanometer scale wavelength. According to estimates, the functional throughput of magnonic holographic devices may exceed 10^{18} bits/s/cm^2.[521]

There are several important milestones to be achieved for further SWD development: (i) nanomagnet switching by spin wave (e.g. by the combined effect of the voltage-assisted anisotropy change and a magnetic field produced by the incoming spin wave); (ii) integration of several magneto-electric cells on a single spin wave bus. In order to have an advantage over Si CMOS in functional throughput, the operational wavelength of SWDs should be scaled down below 100nm.[520] 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[527] and non-Boolean[528,529,530] 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.[531,532] This reversal behavior is markedly different from in-plane permalloy nanomagnets, which remain nearly single-domain during switching and rotate coherently. That said, Bhowmik *et al.* report using the spin Hall effect (SHE) to clock devices with PMA.[533] 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.[534] 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 is allowed to propagate through a part of said subgroup.[535] 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[533]), it is also possible to globally clock a large ensemble of pNML devices (e.g., with an oscillating, out-of-plane magnetic field[530]). 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 is capable of retaining 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.[527] 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.[536]

(iNML): Fanout structures,[537] a 1-bit full adder,[538] etc. have been experimentally demonstrated and successfully re-evaluated with new inputs. Both field-coupled[539,540] and spin transfer torque (STT)[541,542,543] electrical inputs have been realized. For electrical output, multiple NML-magnetic tunnel junction hybrids[544] have been proposed and simulated.

Field based, CMOS compatible line clock structures have been used to simultaneously switch the states of multiple magnetic islands,[545] as well as to re-evaluate NML lines and gates with new inputs.[546] Additionally, recent experiments have considered materials-based solutions to further reduce field/energy requirements for line clock structures. Notably, results from Li *et al.*[547,548] suggest that component energy metrics could be further reduced by as much as 16X.

Voltage controlled clocking – e.g., multiferroics[549] and magnetostriction[550] – have also been proposed as potential clocking mechanisms for iNML. With respect to magnetostriction, Salehi *et al.*[550] suggest that stress-based clocking

would lead to clock energy dissipation of just ~200 kT per device. The SHE could be employed as a path to reduced energy clocking (i.e., 10-100X).[551]

(pNML): Experimental results suggest that appropriately irradiated pNML structures[552] (to define dataflow directionality) can be controlled by a uniform, homogeneous, oscillating, global clock field.[553] Majority gates,[554] AF-lines,[553] multi-plane signal crossings,[555] and full adders[530,556] have all been experimentally demonstrated with this approach. Notably, Irina *et al.*[557] 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 in Perricone *et al.*[558]). When considering I/O and clocking, field coupled electrical inputs have been demonstrated.[559] Also, large arrays of pNML devices could be controlled with on-chip inductor structures[560] 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.[561]

Indeed, as noted above, small lines of pNML devices have been clocked via the SHE.[533] If 100 ps pulses are achievable, Bhowmik *et al.*[533] report 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.[562] Simulation-based studies[562] suggest that a NAND/NOR operation would require just 2.8 aJ (assuming 200 nm x 200 nm devices and 50 MHz frequency and 10 functional layers of devices.)

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.[563] At the device-level, magnetocrystalline biaxial anisotropy could further promote hard axis stability of an iNML ensemble.[563] Alternatively, adiabatic switching[564] and/or field gradients could potentially mitigate the effects of premature switching. Simulations suggest that NML circuits could tolerate some field misalignment.[565] 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.

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 processes information.[566] 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.[566] 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.*[567] and Perricone *et al.*[568] report examples of initial circuit design and benchmarking efforts in this area.

Energy efficient clocking: Voltage controlled clocking should continue to be pursued – not only for benefits with respect to energy (e.g,. as articulated by Nikonov and Young[569]), but also with respect to the fine-grained control it provides for an NML ensemble (which is useful in reducing error rates[570] as well as architecturally[560]). Domain-wall based switching could be another viable alternative for clocking.[571]

3.4.3 Spin Torque Majority Logic Gate

Spin torque majority gate (STMG) relies on spin torque switching mechanism for its operation[572]. STMG consists of a nanoscale ferromagnet, which serves as a common free layer for four fixed nanoscale ferromagnets separated from the free layer by a tunnel barrier[573]. Three of the fixed nanomagnets serve as input terminals of STMG and one fixed nanomagnet serves as the output terminal. Positive and negative voltage pulses applied to the input terminals cause spin torque switching of the free layer according to the majority of the input voltage pulse polarities. The output voltage is determined by the state of the free layer magnetization via the tunneling magneto-resistance effect.

The primary advantage of STMG over CMOS is non-volatility. The state of the free layer retains its latest state at zero standby power. Another advantage is relatively low footprint of the majority gate.

The best reported results to date are based on micromagnetic simulations for free layer with out-of-plane magnetization due to perpendicular magnetic anisotropy[574]. A simple cross geometry with three input and one output cross arms was considered. Switching current of 0.05 mA was predicted for this system. The footprint for this gate was 0.01 um^2. The switching speed is limited by the speed of domain wall propagation, which limits the switching time to approximately 0.1 ns. There are some notable recent publications or results since June 2013 [575,576,577].

Materials with higher spin torque efficiency should be developed in order to further reduce power consumption during the gate operation. Materials with higher domain wall speed should be studied for increased speed of operation of the gate – materials with strong interfacial Dzyaloshinskii-Moriya interaction are of particular interest. Another scheme of STMG is based on phase locking of spin torque oscillators – micromagnetic simulations for this type of STMG are needed.

One of the highest research priorities is first experimental realization of the proposed STMG – experimental studies of the device may reveal challenges not foreseen in simulations. Identification of materials with higher spin torque efficiency is important for realization of a competitive STMG.

3.4.4 All Spin Logic

The recently proposed concept of all spin logic (ASL)[578] uses magnets to represent non-volatile binary data while the communication between magnets is achieved using spin currents in spin coherent channels with the energy coming from the power supply. The ASL concept is based on key scientific advancements of the last decade.[579,580,581,582,583,584,585,586] These advancements have blurred the distinction between spintronics and magnetics, creating the possibility of a device capable of providing a low power alternative to charge-based information processing. In particular, the two key recent advances are (1) the demonstration of spin injection into metals[579,586,587] and semiconductors [588] from magnetic contacts and (2) the switching of a second magnet by the injected spins.[585,586] These demonstrations suggest an all-spin approach to information processing. Magnets inject spins and spins turn magnets (digital bits) forming a closed "ecosystem" which takes advantage of both analog (spin currents) and digital (bistable magnets) properties without the need to convert to charge. It has been shown that ASL can potentially reduce the switching energy-delay product[589] by a significant amount, but there are major challenges to be overcome. One is the room temperature demonstration of switching in multi-magnet networks interacting via spin currents. The other is the introduction of high anisotropy magnetic materials[590] into relevant experiments which can improve energy-delay. Issues such as current density and proper choice of channel materials also have to be carefully considered. The analog nature of ASL communication can be efficiently coupled with median function[591] to develop an architecture called Functionality Enhanced ASL (FEASL) to realize low-power, lower delay and lower area circuits. FEASL is especially suited for adder and multiplier circuits which are an integral part of arithmetic logic units (ALU). Moreover, it should be mentioned that ASL could also provide a natural implementation for biomimetic systems with architectures that are radically different from the standard von – Neumann architecture.

4. EMERGING DEVICES FOR MORE-THAN-MOORE

This section covers emerging devices that are neither targeted for Boolean logic nor memory. The target application may be either improved information processing (Emerging Research Architectures) or extensions to information processing to include peripheral or non-core functions (More-than-Moore). Most of ITRS covers devices expected to be used for Boolean logic in a target application comprising a microprocessor system. The ERD chapter also covers Emerging Research Architectures and More-than-Moore, the latter being essentially an architecture comprising a microprocessor core surrounded by analog devices for I/O. However, the microprocessor may also be augmented by other non-core functions, like power harvesting. This chapter will describe how new devices could raise the performance of microelectronic systems either through more functional diversity or greater performance of the core.

Non-memory uses of the Redox Random Access Memory (ReRAM) or memristors comprise the main new device class in this edition. ITRS sees a possibility that fabs will add capability to produce these devices. Initially, the devices will be as a replacement for Flash in memory sticks and solid state disk drives. ITRS is responding by enhanced coverage of ReRAM in the PIDS and ERD memory sections. However, there is nothing about the devices that confines them only for memory. In fact, the devices can become the basis of reconfigurable logic and neuromorphic architectures and thereby raise information processing performance substantially. ReRAM is not a good acronym because the devices are not used

as memory, so we will use the term memristor in these applications. Requirements for memristors in these applications will be described below.

This section will continue to cover emerging devices for RF applications in the context of low-power sensor nodes.

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 counterfeiting of semiconductor components, side-channel attacks, invasive/semi-invasive reverse engineering, and 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[592,593], 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, most emerging technologies being studied in the context of hardware security applications are related to designing physically unclonable functions (PUFs). Post-CMOS devices[594,595,596] have been suggested as a pathway to a PUF design. While intriguing, these approaches (i) only cover a small part of the hardware security landscape, and (ii) PUF designs often 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[592] and that few if any devices were proposed with hardware security as a "killer application," this document 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.

4.1.2 Background - Devices

While a detailed discussion is beyond the scope of this document, below we introduce the I-V characteristics and the devices that form the basis of this review.

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[597], carbon nanotubes[598] and graphene[599]. By controlling ambipolarity, device polarity can be adjusted/tuned post-deployment. Transistors with a configurable polarity – e.g., carbon nanotubes[600], graphene[601], silicon nanowires (SiNWs)[602], and transition metal dichalcogenides (TMDs)[603] – 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[604].

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[605]. 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[605] one can envision fabricating devices that could be better suited for logic as well as security-related applications. Given that screening length in TMD devices scales with their body thickness, one can achieve substantial tunneling currents.

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[606]. Such behavior has been observed experimentally[607,608]. More specifically, V_{TG} and V_{BG} change the carrier type/density of the drain and source graphene layers by electrostatic field, which can modulate I_{DS}. ITFETs or ThinTFETs may exhibit similar I-V characteristics[609].

4.1.3 Hardware security based on tunable polarity

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 gates: 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.

In[606,610], SiNW FET based polymorphic gates to prevent IP piracy was introduced. 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, through different configurations on the polymorphic control inputs, we can easily change the circuit functionality 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[611]. 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[611].

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 is 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[612]. 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[606,610], 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. 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[611]. 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 gates based logic obfuscation, camouflaging layout embraces higher security level but with larger area overhead.

4.1.4 Hardware security leveraging atypical switching behaviors

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.

Side-channel analysis, such as fault injection, power and timing analysis, 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[613]. 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[606,610]. 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[610]. In the event of a fault injection, the decreased supply voltage will power down the circuit rather than injecting a single-bit fault, 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 '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. (See[606] for more detail.) (SymFET-based Boolean logic is also possible[614].)

4.2 DEVICES WITH LEARNING CAPABILITIES

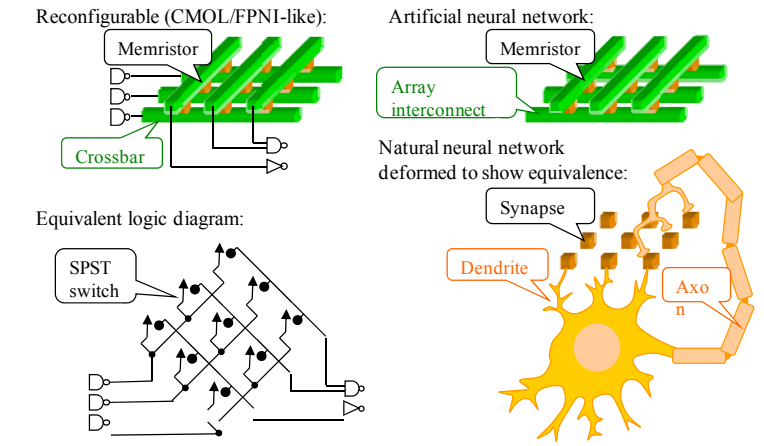
4.2.1 Devices That Learn Logic Configurations

The ITRS traditionally views computer performance as the product of device performance multiplied by a factor related to computer architecture – but it is possible to raise computer performance in a different way. An information processing function can be realized by the placement of gates and wires on a custom ASIC or by coding the function in software and running it on a microprocessor. The ASIC implementation will be several orders of magnitude faster and more power efficient because moving a bit from one gate or functional block to another via a wire is much more efficient than having a microprocessor move the bit in software, incurring as overhead the movement of hundreds of bits of instruction, opcode, cache access, etc. to interpret the machine's instruction set for the software. However, customers demand the ability to change a system's function in a split second. While the function of a microprocessor can be changed just by loading different software, changing the function of an ASIC requires redesign and refabrication of a chip that may take a year and cost a million dollars. The direction under discussion would implement "programmable wire." The vision is a chip that could be rewired and thus realize the flexibility of a microprocessor and the efficiency of an ASIC at the same time.

The Field Programmable Gate Array (FPGA) achieved the flexibility objective of the last paragraph long ago[615], but memristor-class devices show the possibility meeting the efficiency objective as well[616]. As illustrated in figure ERD7[617, 618], there are several approaches in the literature where memristors are used as essentially Single Pole Single Throw (SPST) switches that connect some form of general wiring array into specific wiring interconnections. By altering the pattern of switches, a series of gates in a base CMOS layer can be reconnected into an arbitrary configuration. This is essentially the same architectural concept as an FPGA, but the traditional FPGA implements a non-volatile switch with a sizeable CMOS circuit.

The efficiency of memristors in these types of reconfigurable architectures will depend on figures of merit in Table ERD11. Note that Table ERD11 covers a device called both a ReRAM and memristor in the ITRS. If the device is used for memory, it is called a ReRAM and covered in other sections of the ITRS. If the same device is used in reconfigurable logic (and also for neurmorphic applications) it is called a memristor, where it will be subject to the more stringent requirements in Table ERD11. ITRS considers memristors an "emerging" research device for this application because the device technology is less mature when measured against the more stringent requirements of Table ERD11.

Table ERD 11: Figure-of-merit of three reconfigurable architectures



Reconfigurable-memristor functional mapping. Memristors fill the role of a SPST switch in a crossbar wiring network. The logic circuit can be "rewired" by appropriate switch/memristor settings

Artificial-natural neural network mapping. Memristors fill the role of synapses with row/column wires filling the role of dendrites and axons.

Figure ERD7 Two variants of learning devices for configuration

The key issues for reconfigurable logic are described below:

- The low, on resistance requirement is due to the memristor switches being in series with the signal wires. This means the low, on resistance of the memristor Ron must be on the order of the on resistance of a drive transistor RTon otherwise the switches will lengthen transition times and reduce the throughput of the system.
- The high, off resistance of a memristor Roff must be high enough to avoid generating too much leakage current and thus raising static power dissipation. For a CMOS gate, the ratio of dynamic to static power dissipation will be approximately equal to RToff/RTon ~ 10,000, where RT is the resistance of the transistors. Say the circuits in figure Table ERD11 have f ~ 20 (for fanout) memristors on each of the array wires. In this case, f-1 of the memristors with resistance Roff will have the system power supply voltage across their leads. Thus, f × Roff must be on the order of RToff otherwise the memristors will significantly exacerbate leakage current and power.

4.2.2 Devices That Learn by Examples

Brain-inspired, neural systems are capable of performing tasks close to the level of human intellect and there is strong indication that these systems can be made more efficient using physical devices that support learning. Computers that learn by examples have been successfully applied to activities that have previously been beyond the reach of computers, such as the Watson computer playing Jeopardy[619]. To the extent a new type of computer can address formerly human-only tasks, the proper cost comparison becomes computer cost versus the cost of a human being paid a salary, working in an office, and so forth. The Jeopardy-playing computer demonstration shows that computer methods are known for performing at least some formerly human-only tasks, but the Watson computer was actually a 2500-core microprocessor cluster. If we assume the artificial intelligence community has correctly identified the artificial neural network as the

necessary computational primitive for this new type of computing, it would be important to find a way to build an artificial neural network that is more efficient than a microprocessor cluster.

The motivating example of an artificial neural network is shown in figure ERD7, where a new nanodevice fills the role of a synapse. Many mechanisms and devices have been proposed to implement the synapse, including software emulation, memristors, and other devices. As stated at the beginning of this section, the memristor is covered by ITRS (in the form of ReRAM for memory) and seems destined to be much more efficient than software emulation.

The basis of learning by example is illustrated in the example in figure ERD7. The example shows a device in the architecture configuration of a neural synapse. Information flowing across the device when it is performing its function causes a change in the non-volatile state of the device.

Table ERD11 includes rows for three variants of neural networks:

- There are projects in the literature that simulate artificial neural networks using software models for synapse behavior. Some models are biologically inspired[620] whereas others are device inspired[621]. The authors have tried to capture the common behavioral attributes of the device-inspired models. The memristor models are not derived from real memristors but may form a goal for device designers.
- The authors have tried to capture what would appear to be a reasonable target for artificial synapses.
- We include the final row on hybrid neuromorphic systems to emphasize that many projects in the literature have to accommodated the unavailability of acceptable devices by using methods such as multiple memristors per synapse, offline learning and other methods that produce working systems but at reduced efficiency. See reference[622] for an example.

4.3 EMERGING DEVICES FOR LOW-POWER SENSOR NODE

The More than Moore idea is to enhance the logic family of CMOS from strictly Boolean logic to additionally include a family of analog RF devices. In this instance, the term "family" implies both manufacturing and use of the manufactured device. For manufacturing, the objective would be to enhance fab lines so the new RF devices could either be fabricated with CMOS using the as few additional materials or process steps as possible. An alternative would be separate fabrication of RF components and subsequent integration into the same package. For usage, it implies that the RF components would operate with compatible voltages and signals. In this attempt to address the wide field of "More-than-Moore" emerging RF devices, this section is focused on a few devices and functional blocks.

4.3.1 Graphene RF Transistors

Owing to the potentially very high carrier velocity of graphene, RF transistors using this material possibly can reach very high unity current gain cutoff frequency, f_T: graphene RF transistors have been reported with higher f_T and lower f_{Max} than Si transistor at the same gate length. A cutoff frequency of 300 GHz is reported using Co_2Si-nanowire gate and an exfoliated graphene.[623] A f_T of 240 GHz was reported using wafer scale epitaxial graphene[624] and a f_T of 200 GHz was reported using a CVD-grown graphene layer[625]

To obtain higher f_T, the device configuration needs to be optimized. The source and drain of the graphene transistor are usually defined by depositing a metal film, which can induce parasitic capacitances: in top-gate configuration, where source, drain and gate are on the same side of graphene channel, gate-source or gate-drain capacitances can be large, which results in a reduced f_T. However, in back-gate configuration, where the gate is on the opposite side of the graphene layer than the source and drain, the gate-source or gate-drain capacitances could be lower even for an overlapped situation (L_{gs}<0). CVD graphene can be fabricated easily on the back-gate configuration; epitaxial graphene could not. Methods to fabricate a local embedded back gate on a SiC wafer are, however, suggested, although high growth temperature could make it difficult[626]. An inverted graphene integration process, where transistor gate/source/drain and passive circuit components were pre-fabricated before the transfer of CVD graphene, has been demonstrated on 200mm CMOS compatible process[627]. Graphene trasnsistors achieved f_T and f_{max} of 17GHz and 15.2GHz, respectively.

Since the cutoff frequency is inversely proportional to the channel length, the limit of f_T in graphene transistor has not been fully assessed by the reported wafer scale devices. Using a nanowire gate instead of a patterned metal, f_T was estimated down to a 45 nm channel length based on transition time. High Fermi velocity of carriers in graphene, resulting

in high drift velocity (~ 4 x 10^7 cm/s) in a channel, makes 1THz f_T achievable for a sub-70 nm channel length device[628]. This is greater than sub-30 nm channel length of HEMT and Si RF transistors[629].

The unity power gain frequency or maximum frequency of oscillation, f_{Max}, which is around 10-50 GHz even for devices with f_T of 200 GHz, could be increased by improving the device structure and reducing the parasitics. This is a field which is presently less investigated than the intrinsic properties of graphene.

4.3.2 Spin Torque Oscillators

Spin transfer torque in metallic spin-valves and magnetic tunnel junctions using nano-sized magnetic multilayer structures can drive uniform precession of the free layer magnetization under external magnetic field conditions[630,631]. When combined with the giant magnetoresistance (GMR) or tunneling magnetoresistance (TMR) effect, this precession produces voltage responses that make those magnetic multilayers high frequency spin torque oscillators. The oscillation frequency in a spin torque oscillator can be tuned by adjusting the electric current or the external magnetic field. Due to its high compactness, extremely wide tunability, and compatibility with standard CMOS process, spin torque oscillator has the potential to be an agile RF oscillator[632,633]

Currently, oscillation frequencies ranging from several hundred MHz to tens of GHz have been demonstrated depending on the magnetic structures, magnetic field and input current levels[634] . The output power in spin torque oscillators based on metallic spin valve structures is about several hundred pW and it was improved up to about tens of nW in MTJ based spin torque oscillators[635,636]. Despite these experimental advancements in spin torque oscillators, there are several challenges to be overcome for the practical application of spin torque oscillator. These challenges include 1) auto-oscillation structures, 2) increase of output power, and 3) high spectral purity (low phase noise).

Auto-oscillation structures are required to eliminate the need for an external magnetic field that is used in most recent experimental demonstrations. Spin torque oscillator structures with a perpendicular polarizer and a planar free layer[637], vortex magnetization state in free layer[638], and wavy angular dependence of spin torque[639] have been suggested.

For the spin torque oscillator to be useful, the output power of RF oscillation should be improved above a few microwatts. Although achieving higher magnetoresistance (MR) ratios through high spin polarization of magnetic layers or large precession angle of free layers could be a first approach for higher output power, phase locking of many weakly coupled oscillators is further needed for a sufficient increase of the output power. Many theoretical predictions and experimental demonstrations of synchronization of electrically connected spin torque oscillators have been reported so far[640,641,642].

Among the remaining challenges, obtaining spectral purity that is compatible to the levels of existing current oscillator devices may be the biggest obstacle for the spin torque oscillators to be applied in the telecommunication field. Issues in the spin torque oscillation linewidth are reportedly coming from the lack of temporal coherence[643] and from the non-linearity of the oscillation frequency[644,645]. The adoption of PLL circuits and synchronization of several spin torque oscillators can be one of the solutions for the higher spectral purity.

4.3.3 NEMS Resonators

There is an increasing interest to miniaturize and integrate the off-chip RF components, especially the quartz crystal used in the reference oscillator whose quality factor Q (>10^4-10^5) and temperature stability (better than 1 ppm/°C) are difficult to achieve with integrated devices. The quality factors of integrated LC-tank circuits are limited by the poor quality factor values of integrated inductors and capacitors (from 10's to 100's). As a consequence, the most promising solutions to miniaturize reference oscillators with uncompromised quality factors[646] are related to the classes of vibrating devices.

Among the most promising of these vibrating structures are the *capacitively transduced micro- and nano-electro-mechanical (M/NEM) resonators*. In recent years tremendous progress has been achieved in the frequency x quality factor product, the main figure of merit for MEM/NEM resonators. The general trend of increasing their resonance frequency to values exceeding the GHz domain pushed such resonators towards very small, very stiff and low mass NEM systems. However, their ability to conserve a high-Q at low dimensions is questionable when the main energy dissipation mechanisms are the gas friction, the clamping and surface losses[647]. Another important issue is how the (in)stabilities of these resonators scale with dimensions as the effects of fluctuations in numbers of photons, phonons, electrons and adsorbed molecules can significantly affect the noise characteristics[648]

4.3.4 NEM resonators based on silicon nanowires, carbon nanotubes and graphene

At micrometer scale, recent successful demonstrations of very high frequency resonators are the extensional wine-glass resonators with frequencies ranging from 400 MHz to 1.5 GHz (and Q > 3700)[649] and the dielectrically actuated and piezoresistively sensed 4.41 GHz silicon bar resonator exploiting internal dielectric actuation[650]. Piezoresistive sensing[651] of capacitively actuated resonator exceeding 4GHz with Q>8000 and using the 9th harmonic longitudinal model was implemented.

Very high frequency (VHF) NEM resonators were described using platinum nanowires, resonating at frequencies higher than 100MHz and with a quality factor of 8500 at 4K[652]. The same group later reported[653,654] VHF NEM resonators based upon single-crystal Si nanowires. Carbon nanotubes attracted major interest for building NEM resonators, due to their high stiffness (Young's Elastic Modulus, E, near 1Tpa), low density, defect-free structure and ultra-small cross section. Resonator responses have been[655] reported varying from 3 to 200MHz with voltage-tunable characteristics, in CNTs with 1-4nm in diameter, suspended over a trench. The NEM resonator with resonance frequencies up to ≈ 4 GHz were reported with a similar CNT device loaded in an abacus style with inertial metal clamps, yielding very short effective beam lengths[656,657]. One issue of such small vibrating SiNWs and CNTs is the early onset, at very low applied power, of non-linearities characterized by frequency bistability arising from the effect of tension built-up in the wire at large vibration amplitudes.

Recently, graphene material attracted further attention for its extremely high strength, stiffness, and thermal conductivity along the basal plane. In Ref.[658] exfoliated graphene sheets are suspended to form two-dimensional NEM resonators with resonance frequencies from 1MHz to 170MHz.

4.3.5 NEM resonators based on resonant gate or vibrating body transistors

The capacitively transduced signals of MEM resonators are very small and the impedance matching can be limiting. The movable gate and body FET transistor structures can operate as M/NEM resonators, with the main difference that the output is the drain current of the transistor, offering the possibility of building active resonators.

Resonant gate transistors were reported with out-of-plane AlSi resonant gate MOSFETs[659,660]and with in-plane resonant silicon gate transistors[661]. Aggressively scaled versions of the in-plane resonant gate transistors have been reported[662] based on Silicon-On-Nothing technology to achieve sub-100nm gaps and 400nm-thick single crystal resonators with a front-end process. The lateral MOS transistor could suffer from poor carrier mobility due to the roughness of the vertically etched sidewalls and have shown very little gain, but can be integrated with advanced CMOS[663] to minimize the influence of parasitic capacitances.

An alternative resonant transistor, called Vibrating-Body FET (VB-FET) has been proposed[664,665]; the movable body modulates both the inversion or accumulation charge in the lateral channels and the piezoreistance of the structure (carrier mobility and mass). Silicon nanowires show an unusually large piezoreistance effect compared with bulk Si. An outstanding gain of more than +30dB for the output signal was obtained in micrometer scale double-gate VB-FET when the output is taken from the transistor drain. Moreover, the device motional resistance is reduced from 16kΩ to below 100Ω, which enables excellent conditions for 50Ω matching in RF applications.

Another active resonator has been proposed[666] by using the mechanical strain rather than an electric field to modulate the conductivity of the silicon. This resonator can provide a gain higher than unity at 15MHz, similarly to the VB-FET.

4.3.6 RF Mixers

An RF mixer is an important building block of the RF front end and many emerging solutions seek attention[667]

Resonant tunneling diodes were explored for decades. Owing to their negative differential resistance and fast response, they had some potential in the RF domain, and subharmonic mixers were demonstrated[668,669]. The potential advantages of such an approach are a wide range of operating temperature, frequencies ranging up to 10 THz, and a reduced noise figure due to RTD shot noise suppression. While this field wasn't very active in the recent years it could regain interest due to the increased demand for THZ applications and the coming integration of III-V materials on Si.

For the same reason single electron transistors were considered with resonant frequencies in the range 1 – 10 GHz. A SET-based mixer with fully tunable band selection from 0 to 300 MHz was demonstrated but at cryogenic temperature[670,671].

Recently the ambipolar I-V characteristic of the graphene transistor which mimics the response of full-wave rectifier has been used to demonstrate a frequency-doubler circuit[672,673]. A double balanced RF mixer integrating four GFET has been demonstrated on 200mm CMOS wafers, with significantly improved linearity over single-GFET mixers[674].

Finally the nonlinear I-V characteristic of a carbon nanotube can be used to demodulate AM signal. However the demonstrations were limited to <100 kHz due to the external bias circuitry and to <2 GHz due to intrinsic parasitics (chip bond pads, etc.)[675,676].

5. EMERGING RESEARCH ARCHITECTURES

5.1 Introduction

This chapter covers two areas: Storage Class Memories and Emerging Computing Architectures. Storage class memories refers to large solid state memories intended to replace some or all of the traditional DRAM-disk hierarchy. Emerging Computing Architectures refers to non-traditional computer architectures intended to replace or supplement traditional architectures. Both sections include a review of the state of the art and a summary of open challenges.

5.2 STORAGE CLASS MEMORIES

5.2.1. 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, and 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 thus are rapidly exhausting 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 nonvolatile memory at speeds close to DRAM.

In order to implement SCM, both the emerging memory technologies discussed in Section 2 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.

5.2.2. 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.

Table ERD12 presents a summary of memory needs by application space. The table is organized as a cross-matrix of application drivers indexed against desired memory properties. It is not indexed by year but is meant to be read in the context of computing in the 2015 – 2025 timeframe.

Historically, roughly one-third of the power in a large computer system is consumed in the memory sub-system[677]. 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, Google[678]reports that servers are typically operating at over 50% of their peak power consumption even at very low utilization rates. 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 computational load. This provides additional incentive for the nonvolatile 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. 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[679]. 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. All types of IO software – from the filesystem, through the operating system and up to applications – will need to be redesigned in order to best leverage SCMs. The number of software interactions must be reduced and disk-centric features will need to be removed. Inefficiencies buried deep within conventional software can account for anywhere from 70% to 94% of the total IO latency[680]. 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. 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[681], where accumulated errors are periodically eliminated so that one or two bit error correction can remain sufficient.

Table ERD12. Anticipated Important Properties of Emerging Memories as driven by application need.

5.2.3. 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 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 a 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[682].

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)[683]. 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[684]. With proper caching, a carefully-designed M-SCM system could potentially even match DRAM performance despite its lower device latency[685]. 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 10^7, 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[686]. 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[687].

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,”[688] was predicted to offer 10-60x improvements in energy efficiency[689].

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"[690], 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 down-side that new types bugs might appear as OSs and programs effectively run indefinitely and can't "re-crate 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 a long time. Like others, they point out the problems that arise as NVMs

Table ERD13 presents the likely desirable properties of M-class SCM.

Table ERD13. Likely desirable properties of M (Memory) type and S (Storage) type Storage Class Memories

5.2.4. 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 what is 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. The standard SATA (Serial Advanced Technology Attachment) interface, which is a commonly used interface for SSD, was originally designed for HDD, and is not optimized for flash SSD[691].

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[692]and PCRAM/flash[693]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[694].

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 wear-leveling) 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[695]. Such asymmetry can affect the ordering and atomicity of writes, and needs to be taken into consideration in system or algorithm design[696]. 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[697], partial writes[698], or specialized algorithms/structures that trade writes for reads[20699 700]. (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[701 702 703] – or newly-proposed hardware primitives, such as atomic 8-byte writes and epoch barriers[704 705 20].

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 (<2KB) writes and for reads of all sizes [706]. 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[707]. 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[708], Mnemosyne[709] ,NVMalloc[710],and recovery and durable structures[711]– have been proposed to expose SCM as a persistent heap and thus ease its adoption. Schemes for filesystem support have been developed to transparently utilizes as byte-addressable persistent storage, including Intel's PMFS[712],BPFS, FRASH[713], ConquestFS[714], and SCMFS[715].

Table ERD13 presents the likely desirable properties of S-class SCM.

5.3 EMERGING COMPUTING ARCHITECTURES

5.3.1 INTRODUCTION

Emerging devices are creating opportunities for new computing machines, including machines with non-conventional architectures. At the same time, designers are developing emerging architectural concepts, such as neuro-inspired architectures that incorporate conventional computing devices. This sub-chapter includes a survey of a number of these technologies, together with a summary of open challenges.

Figure ERD8 presents a chart that summarizes and categorizes computing architectures, both conventional and unconventional. The chart is organized along multiple dimensions. Horizontally, the chart is first organized with "Program-Centric" architectures on the left and "Data-Centric" architectures on the right. "Program centric" refers to processors wherein the detailed design, including every step of the program flow, is dictated by the designer. Program-centric processors are further split into conventional stored program Von Neuman machines and non-Von Neuman processors that are not driven from a stored software program. "Data-centric" refers to processors wherein some of the details of the final "design" are driven by the data that is passed through the system during a training step, either by the system manufacturer at the factory, by a systems integrator before shipping to customers, or even by customers in the field. For example, the weights in a neural network are learned for a recognition task by showing the network many data-examples. The final horizontal dimension is the split of each column into CMOS and non-CMOS device enabled categories.

Vertically, the table is divided by a dashed red line between "deterministic" and "non-deterministic" computing. "Deterministic" computing refers to fixed-precision (whether floating or fixed point) and low bit-error-rate compute models. "Non-deterministic" computing refers to computing that does not rely on these paradigms, and the compute solution has some sort of probabilistic nature to it. A vertical distinction that is not specifically marked are that the top band in the deterministic row largely use logic devices while the lower entries in the band largely use emerging memory devices as crossbars. The grey boxes imply a neural inspired architecture.

In the next few sections, emerging program-centric and data-centric architectures are each surveyed in turn, followed by a discussion of crossbar trends. Table ERD14 summarizes some categories of non-CMOS emerging architectures.

Table ERD14. Categories of non-CMOS emerging architectures

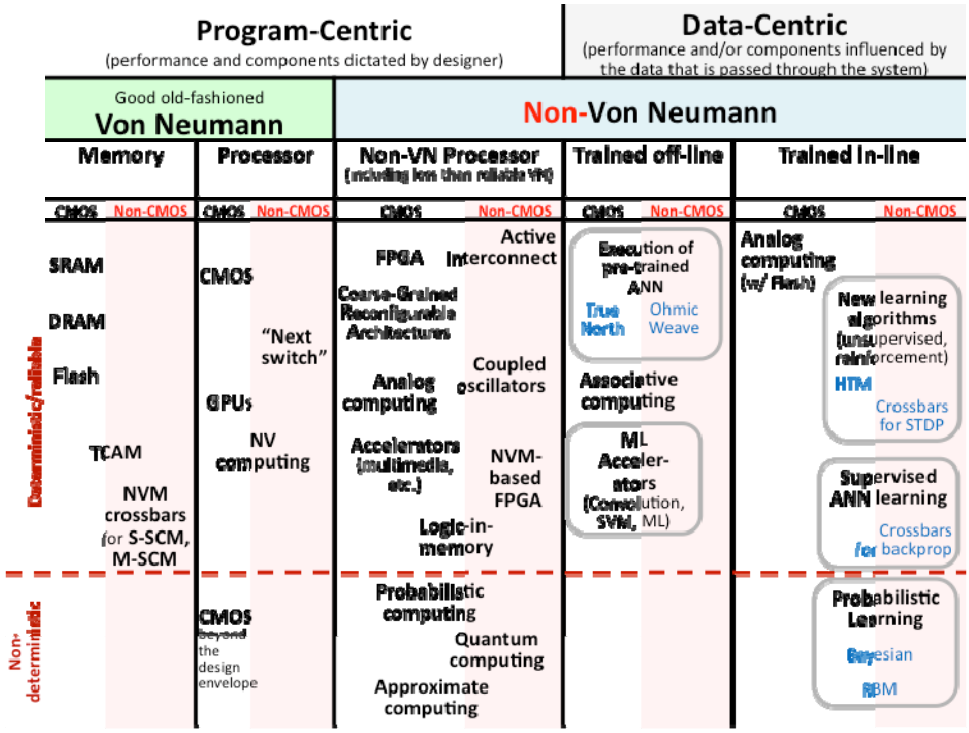


Figure ERD8. Categories of computing architectures.

5.3.2 Emerging Program-centric Architectures

Program-centric architecture are largely designed and programmed in the conventional way – all design is human directed, and processors are programmed using a programming language. As a result, the performance and components of the system are known before any data is presented to the system. At a circuit level, most SCM candidates (see above sections) are configured as one transistor, one "resistive memory device" (1T1R) cells. Using a high on-off ratio transistor as an access device simplifies the overall design but at the expense of area. A long-standing goal has been to create a memory array that consists of memory devices only (1R), i.e. a passive crossbar, or that only needs a two terminal diode (or switch) as an access device (1AD1R or 1S1R). The latter needs unipolar device operation. Outstanding challenges towards this goal has been (1) controlling the "sneak current path" so that the sum of the currents in the "off" devices in an array do not become comparable to the "on" current which we are trying to detect, and (2) the need for the access device to deliver sufficient write current with little power wasted on unselected devices. Together with high device variability, these issues limit the achievable size of a buildable array. Hamdioui *et al.*[716] summarizes these issues and presents some potential circuit level solutions. Narayanan et.al.[717] explores the limits for arrays with a specific type of access device. They explore arrays in the 1 Mb range and show that array size is mainly limited by write power.

Ternary Content Addressable Memories (TCAM) enable lookup via data value, rather than address, and are very useful in many applications. However, each cell requires 14 transistors or more, limiting the density and power consumption. There have been numerous simulation and physical demonstrations of smaller TCAM cells enabled by MTJs or RRAMs, and these typically require 6 CMOS transistors or less (and 2 resistive devices). For example, Li et.al. Demonstrate a 2T-2R 1Mb TCAM based on CMOS and PCM[718].

On the processor side, the concept of Non-Volatile (NV) computing relies on using non-volatile memories within the state registers inside the CPU. This enables the processor to be quickly placed into, and taken out of a full sleep mode, thus saving power overall despite the extra power required for the NV write. For example, Liu et.al.[719] demonstrate using NV flip-flops and NV SRAM to achieve a 170 ns restore time.

"Next switch" refers to using an emerging logic device to replace a CMOS switch in what would otherwise largely be a conventional design. Examples include all-spin logic, and employing TFETs for low voltage circuitry. Since MTJs have a high write current, compared with a CMOS FET, their use in program-centric computers is probably limited to enabling fast all-sleep modes. In contrast, all-spin devices have potential for low power computation but the level of device maturity is very low, and benchmarking results to date indicate poor potential for all spintronic and ferroelectric devices to have a better energy-delay product than CMOS[720] .

One "next switch" for which there is a lot of current interest is to use Josephson Junctions to build superconducting circuits operating at very low temperatures. Devices and circuits have been demonstrated. Even though the devices need to be cooled to 4 K, there is potential for a better energy-delay product than CMOS due to the expectation of very high speed operation[721]. One approach is to build relatively "conventional" logic using Josephson junctions modulating single-flux-quantum (SFQ) voltage pulses. A very different approach is to build a flux qubit that operates below 0.1 K and supports the superposition of two quantum states. An algorithm referred to as "quantum annealing" can be used to solve discrete optimization problems during which the qubits settle to one state.

Several new concepts are under active exploration for program-centric, non-Von Neuman processors, i.e. digital circuits generally designed for a fixed one or small set of applications. "Active interconnect" refers to using switchable low-resistance devices to enable switchable interconnect. For example the "complementary atom switch" has potential to replace the switchboxes in FPGAs for a substantial gain in power/performance[722].

This has also been considerable interest in replacing the Look Up Tables (LUTs) in FPGAs with LUTs enabled by NV memory crossbars. For example, Kumar et.al.[723] present a circuit for a RRAM based crossbar that reduces the sneak path current and half write problems, and predict an improvement in read energy delay product in the range of 10^3 to 10^4 J.

Logic-in-memory refers to the idea of attaching logic directly to the columns of an SRAM or DRAM, and has been proposed or employed multiple times to speed-up memory bound operations. Of course, replacing the SRAM or DRAM with a fast low power NV RAM, such as STT or RRAM gives potential for reducing power, delay and area.

When multiple oscillators are loosely coupled, they tend towards the lowest energy synchronized state. This can be used for associative computing tasks, e.g. image recognition. Several projects are underway today in which such systems are being built using CMOS and emerging-device-enabled oscillators, e.g. coupled spin networks. For example, Fan et.al.[724] describe one approach using spin torque oscillators.

Non-deterministic computing refers to computing paradigms in which different behaviors can be obtained on different runs. Examples include approximate computing, statistical computing, and quantum computing.

Approximate computing refers to the idea that you can size the precision (number of bits) of intermediate stages of the computation and achieve a sufficiently precise result. This reduced precision can be used to save energy. This concept can be quite useful in areas such as DSP, image processing, recognition and data mining. Just don't use it for finance! Current challenges are in hardware and software support to enable programmable approximation without user-driven optimization of the required precision. Approximate computing can also be used to manage lack of precision due to lack of reliability. See[725] for an example of recent work in this area. That work addresses still open issues as to how to represent probability reasoning to the designer and how to build an end-to-end compiler chain.

Statistical computing is a broad field covering the intersection of computing and statistics. The area of interest in the context of this chapter at the hardware layer is computing in the presence of errors, for example using an estimator to correct an output based on error statistics[726]. This in turn enables the use of emerging devices with poor yield or low bit-error-rates while enabling high density and low-power operation. A key challenge today is satisfying the need to have nanofunction models that encapsulate the statistical behavior needed to enable statistical information processing. A further need is to formulate the abstractions and supporting tools that permit software to be written that is agnostic to the specific statistics of the underlying hardware. A potentially broader agenda is how to tie probabilistic behavior to probabilistic programming, which are being used in machine learning techniques.

One quantum computing paradigm – quantum annealing – was discussed above. A more general technique is quantum superposition – for example if one has a pair of qubits, the quantum superposition can be in any one of four states. By manipulating this state through a set of quantum logic gates (e.g., by performing a quantum algorithm), a system of $\boldsymbol{n}$ qubits will decompose into one of $\boldsymbol{2^n}$ states, however with only a certain known probability that the answer will be correct. Vizotta[727] presents a review of the state of the art. Open challenges include maintaining coherence at scale and I/O between the outside world and qubits that must be carefully isolated from their surrounding environment. Numerous algorithms have been mapped onto quantum computers and prototype programming languages developed.

5.3.3 Emerging Data Centric Architectures

These are architectures that derive at least part of their "design" or "program" from the data they are processing. In effect, either the final performance or the makeup of the configuration that delivers a particular performance is not specified until data has been passed through the system. Such systems can be further divided into two types, systems that are trained off-line, and systems trained on-line. In the former, the weights and settings are trained on a separate computer on which a (usually) labeled training data set is run. In the latter, both training and recognition are done on the same hardware, though sometimes the training might consist of updating weights and settings initially trained off-line.

Artificial Neural Networks (ANN) have been under active investigation as a computation paradigm since the 1940s. Such systems consist of an interconnected network of multiple layers. Each node produces an output

$f(x) = K \Sigma_i w_i g_i(x)$

where $g_i(x)$ are the inputs, w_i are the weights, and K is an activation function, which can be linear, binary or non-linear. A continuously differentiable activation function permits gradient-based optimization methods to be used in training. The use of a nonlinear activation function allows multi-level systems to scale computing capability significantly as the number of layers in the network increases. In a spiking neural network, the output only changes ("fires") when the evaluated function reaches a specific value.

A recent CMOS implementation of a spiking neural network is the IBM True North chip[728]. True North is programmed via the "Corelet Programming Environment" which includes a compositional language and development library.

There is considerable interest in building ANNs with programmable resistive memories, including RRAM, PCM and spin-based. Building ANNs with crossbar arrays is of particular interest, using the programmable memory elements to store synaptic weights. With such, there is significant potential for power reduction and improvements in area efficiency. However, one challenge is that the NVM nonlinearity and asymmetry "prevents the back-propagation algorithm from locating optimum values during programming."[729].

Conventional logic can also be implemented using NVM crossbar arrays, by training the network with the complete truth table. This has been demonstrated (in simulation) on Ohmic Weave[730].

Micron has built a CMOS "automata processor" designed to accelerate regular expression evaluation, for use with complex unstructured data streams[731].

Recently, Deep Convolutional Networks (such as AlexNet) have arisen as algorithms that can achieve a high recognition success rate, once intensively trained off-line. Other successful ANN algorithms being pursued by the Machine Learning community include Long Short Term Memory. For example, deep learning is a feedforward ANN employing multiple layers of hidden neurons, and nonlinear activation functions. CMOS accelerators have been built for a number of these algorithms, e.g. Chen[732]. A key challenge is achieving high throughput in the presence of finite memory bandwidth.

The right-most column of Figure ERD8 refers to implementations of algorithms that can do in-situ training.

Some of these are variants of algorithms that can be trained off-line, for example using back-propagation to modify previously trained network weights. Currently, there is significant research into how to incorporate online modifiability into a number of machine learning algorithms. These normally operate in a supervised learning mode, i.e. the training data is clearly labelled.

Of particular interest are applying algorithms that take their inspiration about how the brain learns, and can learn even with unlabeled data (unsupervised learning). One approach is to implement spike timing dependent plasticity (STDP) on a spiking neural network. In this biologically inspired approach, a neuron's weight is increased if the input spike tends to occur just before the output spike (causal), and decreased if not (acausal). Of particular research interest is how to implement STDP in an NVM crossbar array.

Recently other algorithms have emerged that are inspired by higher level models implementing assumptions of how the brain works at the cognition level. Examples include Hierarchical Temporal Memory (HTM)[733] and Cogent Confabulation.[734] In Cogent Confabulation, the neurons are excited by a sum of Bayesian probabilities. Hebbian learning is employed. HTM is not inspired directly by ANN, but instead by neuroscientific observations about the neocortex. It is a sparse algorithm in which spatial and temporal sequences matter. In HTM, learning phases run between recognition or inference phases.

The general idea of analog computing predates digital computing. In traditional analog computing, analog multipliers etc. are built to perform calculations. Of recent interest has been the idea of using analog memories within an analog computing paradigm. This permits a programmable analog system. One approach is to use floating gates as analog memories in a field programmable analog array, and apply them to classification problems, e.g.[735] .

An area that has been little explored in the context of new devices is probabilistic learning, .e.g. in-situ training and classification in a Bayesian network. Such systems require representation of probability distributions within the system.

5.4 CROSSBAR PROGRESS

One relatively common theme is the construction of crossbar arrays, typically arrays that use a two-terminal access device instead of a transistor. Applications include FPGA LUTS, neuromorphic arrays, Storage-Class Memory, and TCAMs. Table ERD15 summarizes a selection of reported measured results for implemented crossbars.

Banno et.al.[736] describes a chip with an 8x16 crossbar switch. Each switch is built using two RRAM-based varistors, each with their own control line. Their motivating application was programmable logic devices. Prezioso et.al. [737] describe a 10x8 crosspoint built using bilayer metal-oxide memristors. Their motivating application are classifiers using deep learning neural networks. Luo et.al.[738] describe a threshold selector MIEC switch built with HfO2 RRAM device. They demonstrate the device in a 32x32 array. Yu et.al.[739] describe a 64x64 1T1R pseudo-crossbar with RRAM devices and integrate and fire neuron read CMOS circuits. Burr et.al.[740] describe a 3-layer perceptron network, with 164,885 synapses built using PCM. They demonstrate the capability of the network using the MNIST dataset. Jo et.al. [741] describe a 4 Mb RRAM array addressed using a new select device.

Table ERD15. A selection of reported measured results for implemented crossbars

6. EMERGING MEMORY AND LOGIC DEVICES ASSESSMENT

6.1 INTRODUCTION

The purpose of this section is to assess emerging research technology entries considered in this chapter benchmarked against current memory or CMOS technologies. Two methods are used to perform this assessment. In the "Quantitative Logic Device Benchmarking", each emerging logic 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 15nm CMOS applications. The second method, referred to as "Survey-Based Benchmarking", 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.

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, limited number of survey results sometimes raises questions of the accuracy of this survey. Therefore, a different voting-based survey was used in this version to replace the previous approach.

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 [742]concludes that the fundamental limit of scaling an electronic charge-based switch is only a factor of 3× smaller than the physical gate length of a silicon MOSFET in 2024. Furthermore, the density of these switches is limited by 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 replacement 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[743], 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" is an important means of connecting device performance and area to communication performance, relating time to space. The metric measures how other devices may be contacted within a characteristic delay of the

switch, and is dependent on not only switch delay but also switch area, as well as communication speed[744]. Successful architectures also need effective fan-out.

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[745] are particularly valuable.

6.2 QUANTITATIVE LOGIC DEVICE BENCHMARKING

The first method for benchmarking emerging information processing devices is based on their quantitative evaluation in conventional circuits. The Nanoelectronics Research Initiative (http://nri.src.org) 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. Several of the more promising NRI devices have been described in detail in the Logic and Emerging Information Processing Device Section, and the intermediate results on the benchmarking efforts were outlined in a 2010 IEEE Proceedings article[746]. The benchmark results were updated again based on refined device data in 2011[747]. Some results of this benchmark study were included in 2013 ERD chapter. Another independent benchmark was conducted by Intel researchers pursuing uniform methodology for benchmark on the same set of NRI device using similar reference circuits and parameters[748,749]. Earlier results of this study were included in 2013 ERD chapter and more results results are discussed in this version.

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 benchmark 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

Preliminary analyses sponsored by SRC/NRI and the Intel study surveyed the potential logic opportunities afforded by 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 15nm generation as captured by the ITRS. 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.

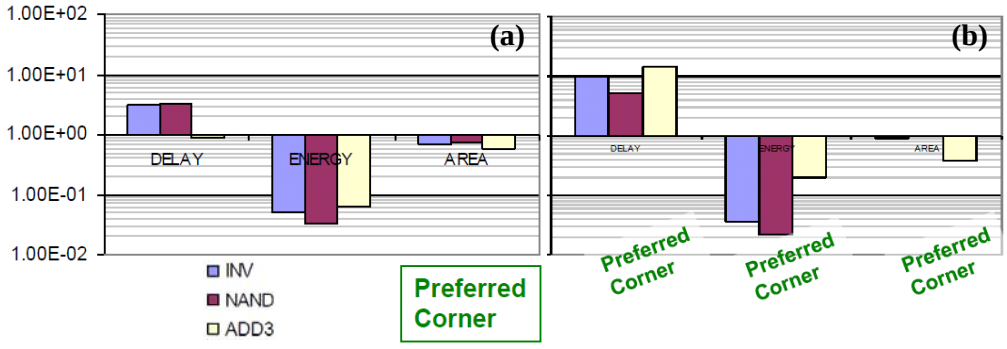


Figure ERD9 Median delay, energy, and area of proposed devices in NRI benchmark (normalized to ITRS 15-nm CMOS), based on principal investigators' data. (a) 2011 benchmark results[747]; (b) 2010 benchmark results[746].

At a high level, the data from these studies corroborates qualitative insights from earlier works, suggesting that many new logic switch structures are superior to CMOS in energy and area, but inferior to CMOS in delay, as shown in the plot of median data for the device (Figure ERD9). This is perhaps not surprising; the primary goal for nanoelectronics and NRI is to find a lower power device[750] since power density is a primary concern for future CMOS scaling. The power-speed tradeoffs commonly observed in CMOS is also extended into the emerging devices. The difference between panel (a) and panel (b) shows that the benchmarking results were evolving from 2010 to 2011, due to more refined device data input from NRI researchers in 2011. 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 ERD10 lists the devices benchmarked in the Intel study[749]. Simplified 32-bit arithmetic logic unit (ALU) was built from these devices to evaluate their performance. The energy-delay plot in Figure ERD11(a) reveals disadvantages of spin-based devices in speed, while no obvious energy advantage is demonstrated. Figure ERD11(b) shows power-constrained throughput of these ALUs. It should be pointed out that the evaluations from the NRI benchmark and the Intel study do not always agree with each other on the same types of devices, due to different device assumptions on device parameters, interpretations of physical mechanisms, and logic gate designs.

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 ERD10 List of devices considered in Intel benchmark with their computational variables and classification[749].

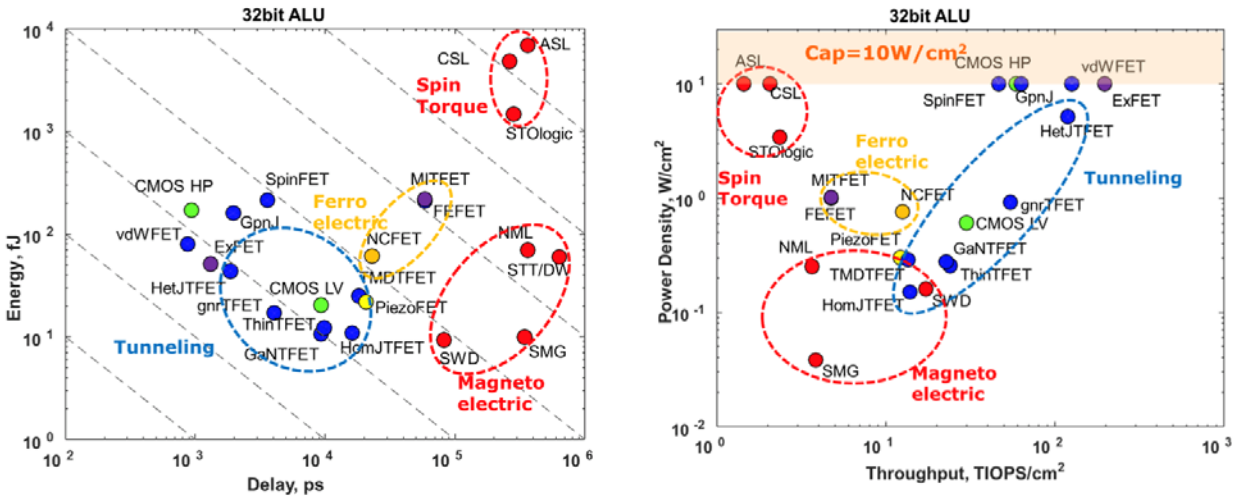


Figure ERD11 (a) Energy *vs*. delay plot of 32bit ALU built from benchmarked devices; (b) power *vs*. throughput of 32bit ALU built from these devices, reflecting power-constrained (< 10 W/cm^2) throughput[749].

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 100mW 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" [751] 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.

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 [752,753]. 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 and limitations 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 decade. 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. This is particularly attractive given the move to multi-core chips: while most are homogeneous today, 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 to 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

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 nice memory devices and fourteen logic devices were evaluated. Each device was evaluated by a proponent who focuses on the advantages of the device and a friendly critic who addresses challenges of the device. Each device was extensively discussed by participating experts. A survey was also conducted in the workshops for the experts to vote on the "most promising" devices and devices "needing more resources". Figure ERD12 shows the relative number of vote received by emerging memory devices in these two categories, ranked from high to low in the "most promising" category (read color bars). Figure ERD 13 is the relevant volte of emerging logic devices.

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 concensus among experts based on factual R&D status of these devices. The "need more resources" category reflects perceived value of these devices in the view of the experts who have diversified background and interest. The results also reflect 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 resource. The strong interest in emerging FeFET memory is closely linked to the discovery of ferroelectricity in doped HfO_x. Carbon-based memory is also recommended as a top choice for R&D resource, owing to general interest in carbon materials and devices based on various forms of carbon materials.

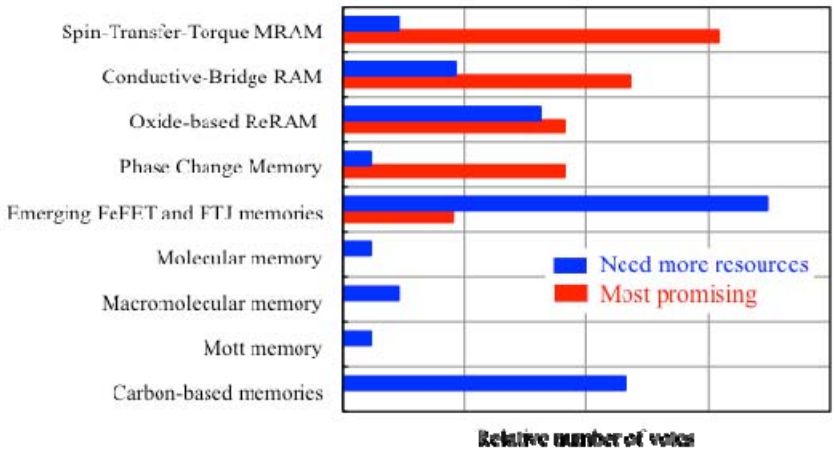


Figure ERD12 Survey of emerging memory devices in 2014 ERD Emerging Memory Workshop (Albuquerque, NM).

Similar trend and difference in the "most promising" and "need of resources" categories are also observed in emerging logic devices in Figure ERD13. "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.

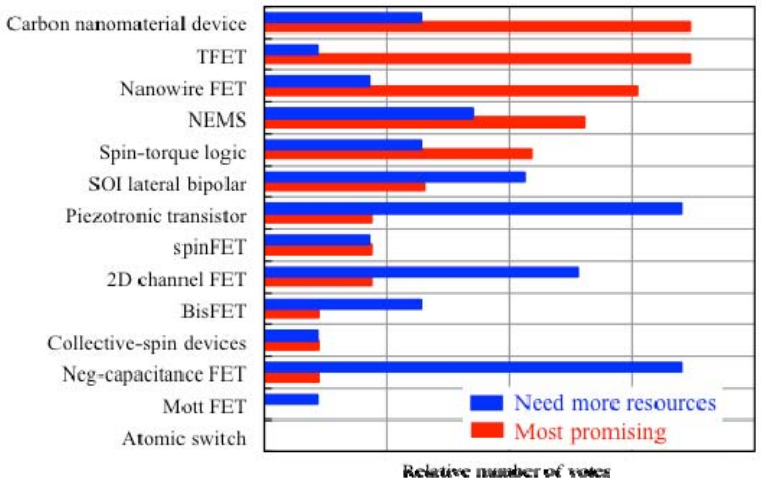


Figure ERD13 Survey of emerging logic devices in 2014 ERD Emerging Logic Workshop (Albuquerque, NM).

6.3.2 ERD Survey Criteria, Methodology, and results

The second method for benchmarking emerging memory and information processing devices is based on a survey of the Emerging Research Devices Working Group. Some emerging nanoscale devices discussed in this chapter are charge-based structures proposed to extend CMOS to the end of the current roadmap. Other emerging devices offer new computational state variables and will likely require new fabrication technologies. A set of relevance or evaluation criteria, defined below, are used to parameterize the extent to which proposed "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. Detailed description of each criterion can be found in 2013 ERD chapter.

Each CMOS extension and beyond-CMOS emerging research nanoscale memory and logic device technology is evaluated against each Relevance Criterion 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. These numbers are more precisely defined in that charts below. 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. Therefore, this version used the experts' vote in 2014 workshops as the qualitative assessment. Figures ERD14 and ERD15 summarize the critical review conducted in 2013 for emerging memory devices and emerging logic devices, respectively. Devices in the same category are compared in the same spider chart, while 2013 ERD chapter provided spider charts comparing evolving assessment of each device from 2009 to 2013. Notice that the technology entries in these figures are based on the 2013 ERD chapter, while some of them have been removed in this version (*e.g.*, molecular memory, atomic switch, *etc*.) and several new additions in this version are not included (*e.g.*, novel STTRAM, piezotronic transistors, *etc*.)

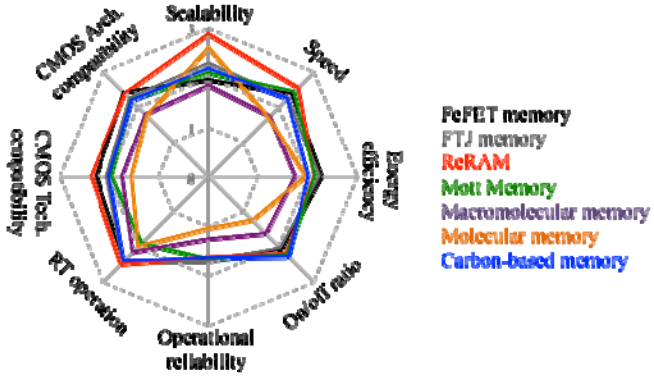
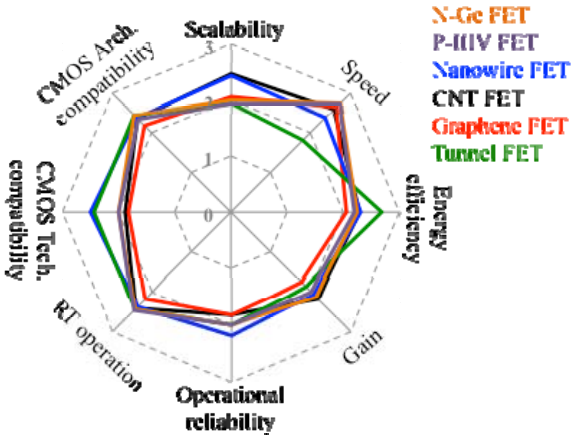
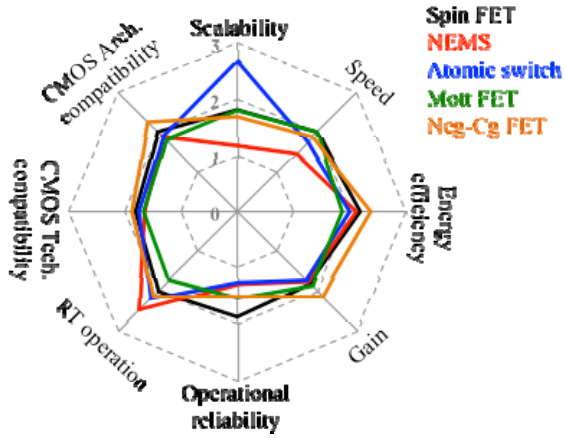


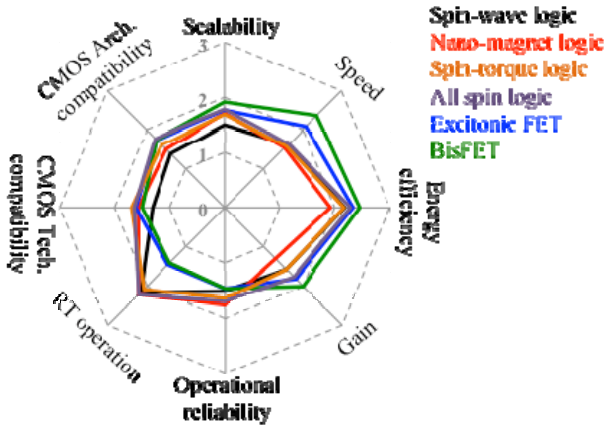
Figure ERD14 Comparison of emerging memory devices based on 2013 critical review.



(a) CMOS extension devices



(b) Charge-based byond-CMOS devices



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

Figure ERD15 Comparison of emerging logic devices based on 2013 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 ERD15 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 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. PROCESSING

7.1 INTRODUCTION

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 Emerging Research Devices Working Group proposes the following comprehensive set of guiding principles. We believe these "Guiding Principles" provide a useful structure for directing research on any "Beyond CMOS" information processing technology to dramatically enhance scaling of functional density and performance while simultaneously reducing the energy dissipated per functional operation. Further this new technology would need to be realizable using a highly manufacturable fabrication process.

7.2 GRAND CHALLENGES

7.2.1 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.

7.2.2 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.

7.2.3 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.

7.2.4 NANOSCALE THERMAL MANAGEMENT

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

7.2.5 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.

7.2.6 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] 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

[3] H. Akinaga and H. Shima, "Resistive random access memory (ReRAM) based on metal oxides", *Proc. IEEE* vol. 98, pp. 2237-2251, 2010.

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

[5] 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.

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

[7] Z. Wei*, et al.*, "Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism," in *IEDM*, 2008, pp. 1-4.

[8] 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.

[9] 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.

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

[11] 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

[12] 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.

[13] 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

[14] 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

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

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

[17] 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.

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

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

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

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

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] F. Miao, J. P. Strachan, J. J. Yang, M.-X. Zhang, I. Goldfarb, A. C. Torrezan*, et al.*, "Anatomy of a Nanoscale Conduction Channel Reveals the Mechanism of a High-Performance Memristor," *Advanced Materials,* vol. 23, pp. 5633-5640, 2011.

[30] 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.

[31] B. Govoreanu, G. S. Kar, Y. Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini*, et al.*, "10×10nm² Hf/HfO<inf>x</inf> crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2011, pp. 31.6.1-31.6.4.

[32] 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," in *VLSI Technology (VLSI-Technology): Digest of Technical Papers, 2014 Symposium on*, 2014, pp. 1-2.

[33] K.-C. Chang, T.-C. Chang, T.-M. Tsai, R. Zhang, Y.-C. Hung, Y.-E. Syu*, et al.*, "Physical and chemical mechanisms in oxide-based resistance random access memory," *Nanoscale Research Letters,* vol. 10, pp. 1-27, 2015.

[34] Y. B. Kim, S. R. Lee, D. Lee, C. B. Lee, M. Chang, J. H. Hur*, et al.*, "Bi-layered RRAM with unlimited endurance and extremely uniform switching," in *VLSI Technology (VLSIT), 2011 Symposium on*, 2011, pp. 52-53.

[35] 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.

[36] 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.

[37] 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.

[38] Y. Bai, H. Wu, K. Wang, R. Wu, L. Song, T. Li, J. Wang, Z. Yu, and H. Qian, "Stacked 3D RRAM array with graphene/CNT as edge electrodes," Scientific Reports **5**, 13785 (2015).

[39] C.-L. Tsai, F. Xiong, E. Pop, and M. Shim, "Resistive Random Access Memory Enabled by Carbon Nanotube Crossbar Electrodes," *ACS Nano,* vol. 7, pp. 5360-5366, June 2013.

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

[41] X. Ma, D. Wu, H. Wu, N. Deng, and H. Qian, "A 16Mb RRAM test chip based on analog power system with multiple write schemes," NVMTS 2015.

[42] 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," in *Electron Devices Meeting (IEDM), 2014 IEEE International*, 2014, pp. 6.7.1-6.7.4.

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

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

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

[46] 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," in *IEDM Tech. Dig.*, 2008, pp. 1-4.

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

[48] 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.

[49] 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.

[50] 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.

[51] YY. Chen et al. Appl. Phys. Lett. 100, 113513 (2012)

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

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

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

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

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

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

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

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

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

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

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

[63] 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.

[64] 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.

[65] 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.

[66] Low operation voltage macromolecular composite memory assisted by graphene nanoflakes 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. Mater. Chem. C 1,552-559, 2013.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] Q. Chen, B. F. Bory, A. Kiazadeh, P. R. F. Rocha, H. L. Gomes, F. Verbakel, D. M. De Leeuw, S. C. J. Meskers, Appl. Phys. Lett. 99, 083305, 2011.

[73] 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.

[74] S.L. Miller and P.J. McWhorter, Physics of the ferroelectric nonvolatile memory field effect transistor, J. Appl. Phys. 72, 5999 (1992).

[75] T. Oikawa, H. Morioka, A. Nagai, H. Funakubo, and K. Saito, Thickness scaling of polycrystalline Pb(Zr,Ti)O3 films downto 35nm prepared by metalorganic chemical vapor depositionhaving good ferroelectric properties, Appl. Phys. Lett. 85, 1754 (2004).

[76] 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).

[77] 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).

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

[79] 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.

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

[81] 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).

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

[83] 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).

[84] 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, in IEEE International Electron Devices Meeting (IEDM), Washington (2013.), pp. 10.8.1 - 10.8.4.

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

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

[87] 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, in IEEE International Reliability Physics Symposium (IRPS) (2014.), pp. 2E.5.1 - 2E.5.5.

[88] 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).

[89] 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).

[90] S. Mueller, S. Slesazeck, T. Mikolajick, J. Muller, P. Polakowski, and S. Flachowsky, Next-generation ferroelectric memories based on FE-HfO2, in Applications of Ferroelectric, International Symposium on Integrated Functionalities and Piezoelectric Force Microscopy Workshop (ISAF/ISIF/PFM), 2015 Joint IEEE International Symposium on the (2015.), pp. 233–236.

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

[92] 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).

[93] 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).

[94] 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).

[95] 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).

[96] 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).

[97] 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).

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

[99] 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).

[100] 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, submitted for publication (2015).

[101] 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).

[102] 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).

[103] F. Kreupl, R. Bruchhaus, P. Majewski, J. B. Philipp, R. Symanczyk, T. Happ, C. Arndt, M. Vogt, R. Zimmermann, A. Buerke, A. P. Graham, M. Kund, "Carbon-Based Resistive Memory", IEDM Tech. Dig., pp. 1-4 (2008)

[104] J. Xu, D. Xie, C. Zhang, X. Zhang, P. Peng, D. Fu, H. Qian, T. Ren, L. Liu, "Pulse widths dependence of programming and erasing behaviors for diamond like carbon based resistive switching memories", Applied Physics Letters, 105(17), 172101, 2014.

[105] F. Kreupl, "Carbon Memory Assessment", arxiv.org/abs/1408.4600, 2014.

[106] A. D. Liao, P. T. Araujo, R. Xu, M. S. Dresselhaus, "Carbon nanotube network-silicon oxide non-volatile switches", Nature communications, 5:5673, 2014.

[107] P. Araujo, A. Liao, J. Rodriguez-Nieva, E. Barros, H. Jung, J. Hao, M.S. Dresselhaus, "Carbon Nanotube Network Anti-fuses", In APS Meeting Abstracts, Vol. 1, p. 37003, March 2014.

[108] F. Zhuge, J. Li, H. Chen, J. Wang, L. Zhu, B. Bian, F. Bing, Q. Wang, L. Li, R. Pan, L. Liang, H. Zhang, H. Cao, H. Zhang, Z. Li, J. Gao, K. Li, "Single-crystalline metal filament-based resistive switching in a nitrogen-doped carbon film containing conical nanopores", Applied Physics Letters, 106(8), 083104, 2015.

[109] S. Porro, E. Accornero, C. F. Pirri, C. Ricciardi, "Memristive devices based on graphene oxide", Carbon, 85, 383-396, 2015

[110] J. Xu, D. Xie, T. Feng, C. Zhang, X. Zhang, P. Peng, D. Fu, H. Qian, T. Ren, L., "Scaling-down characteristics of nanoscale diamond-like carbon based resistive switching memories", Carbon 75, 255–261, (2014).

[111] L. Dellmann, A. Sebastian, P. Jonnalagadda, C. A. Santini, W. W. Koelmans, C. Rossel, E. Eleftheriou, "Nonvolatile resistive memory devices based on hydrogenated amorphous carbon", 43rd European Solid-State Device Research Conference (ESSDERC), 2013.

[112] Glen Rosendale, Sohrab Kianian, Monte Manning, Darlene Hamilton, X.M. Henry Huang, Karl Robinson, Young Weon Kim, Thomas Rueckes, "A 4 Megabit Carbon Nanotube-based Nonvolatile Memory (NRAM)", Proceedings of the ESSCIRC, 2010.

[113] Sheyang Ning, Tomoko Ogura Iwasaki, Eisuke Yanagizawa, Shogo Hachiya, Glen Rosendale, Monte Manning, Darlene Viviani, Thomas Rueckes, Ken Takeuchi, "Investigation of Carbon Nanotube Memory Cell Array Program Characteristics" SSDM, 30. September 2015

[114] Sheyang Ning, Tomoko Ogura Iwasaki, Kazuya Shimomura, Koh Johguchi, Glen Rosendale, Monte Manning, Darlene Viviani, Thomas Rueckes, Ken Takeuchi, 23% Faster Program and 40% Energy Reduction of Carbon Nanotube Non-volatile Memory with Over 1011 Endurance, VLSI Technology Symposium 2014

[115] Sheyang Ning, Tomoko Ogura Iwasaki, Eisuke Yanagizawa, Shogo Hachiya, Glen Rosendale, Monte Manning, Darlene Viviani, Thomas Rueckes, Ken Takeuchi, "Investigation and Improvement of Verify-Program in Carbon Nanotube-Based Nonvolatile Memory", Trans. Elec. Dev., Vol 62, 9, 2015.

[116] R.F. Smith, T.Rueckes, S. Konsek, J.W. Ward, D.K. Brock, and B.M. Segal, "Carbon Nanotube Based Memory Development and Testing", EEEAC paper #1310, (2006).

[117] N. F. Mott, Metal-Insulator Transitions, 2nd ed. (Taylor & Francis, London, 1990)

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

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

[120] 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

[121] 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

[122] 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

[123] 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

[124] 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

[125] 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

[126] 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

[127] 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

[128] F. Nakamura, M. Sakaki, Y. Yamanaka, S. Tamaru, T. Suzuki, and Y. Maeno, "Electric-field-induced metal maintained by current of the Mott insulator Ca2RuO4", Scientific Reports 3 (2013) 2536

[129] 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

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

[131] 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

[132] 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

[133] 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

[134] 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

[135] 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)

[136] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Yoshishige Suzuki, "Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses," *Nature Materials*, 11, 39-43, 2012.

[137] S. Kanai, M. Yamanouchi, S. Ikeda, Y. Nakatani, F. Matsukura, and H. Ohno, "Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction," Applied Physics Lett., 101, 122403 (2012).

[138] W.-G. Wang, M. Li, S. Hageman, and C. L. Chien, "Electric-field-assisted switching in magnetic tunnel junctions", Nature Materials 11, 64-68 (2012).

[139] S. Kanai, Y. Nakatani, M. Yamanouchi, S. Ikeda, H. Sato, F. Matsukura, and H. Ohno, "Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field-effect," Applied Physics Letters 104, 212406 (2014).

[140] T. Nozaki, "Voltage-control of magnetic anisotropy: what's the next target?," presented at Intl Colloquium on Magnetic Films and Surfaces (ICMFS) 2015 Cracow.

[141] Hoffmann, A. Spin Hall Effects in Metals, IEEE Transactions on Magnetics, **49**, 5172-5193, (2013).

[142] Sinova, J., Valenzuela, S. O., Wunderlich, J., Back, C. H. & Jungwirth, T. Spin Hall effects. Reviews of Modern Physics 87, 1213-1260 (2015).

[143] Liu, L. et al. Spin-Torque Switching with the Giant Spin Hall Effect of Tantalum. Science **336**, 555-558, (2012).

[144] Liu, L., Pai, C.-F., Ralph, D. C. & Buhrman, R. A. Magnetic Oscillations Driven by the Spin Hall Effect in 3-Terminal Magnetic Tunnel Junction Devices. Phys. Rev. Lett. **109**, 186602 (2012).

[145] Liu, L., Moriyama, T., Ralph, D. C. & Buhrman, R. A. Spin-Torque Ferromagnetic Resonance Induced by the Spin Hall Effect. Phys. Rev. Lett. **106**, 036601 (2011).

[146] Emori, S., Bauer, U., Ahn, S.-M., Martinez, E. & Beach, G. S. D. Current-driven dynamics of chiral ferromagnetic domain walls. Nat. Mater. **12**, 611-616, (2013).

[147] Ryu, K.-S., Thomas, L., Yang, S.-H. & Parkin, S. Chiral spin torque at magnetic domain walls. Nat Nano **8**, 527-533, (2013).

[148] Slonczewski, J. C. Current-driven excitation of magnetic multilayers Journal of Magnetism and Magnetic Materials **159**, L1-L7 (1996).

[149] Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. Phys. Rev. B **54**, 9353–9358 (1996).

[150] Zhang, W., Han, W., Jiang, X., Yang, S.-H. & S. P. Parkin, S. Role of transparency of platinum-ferromagnet interfaces in determining the intrinsic magnitude of the spin Hall effect. Nat. Phys. **11**, 496-502, (2015).

[151] Pai, C.-F., Ou, Y., Vilela-Leão, L. H., Ralph, D. C. & Buhrman, R. A. Dependence of the efficiency of spin Hall torque on the transparency of Pt/ferromagnetic layer interfaces. Phys. Rev. B. **92**, 064426 (2015).

[152] Pai, C.-F. et al. Spin transfer torque devices utilizing the giant spin Hall effect of tungsten. Appl. Phys. Lett. **101**, 122404, (2012).

[153] Demasius, K.-U. et al. Enhanced spin-orbit torques by oxygen incorporation in tungsten films. Nat Commun 7, (2016).

[154] Mellnik, A. R. et al. Spin-transfer torque generated by a topological insulator. Nature **511**, 449-451, (2014).

[155] Fan, Y. et al. Magnetization switching through giant spin–orbit torque in a magnetically doped topological insulator heterostructure. Nat. Mater. **13**, 699-704, (2014).

[156] Tserkovnyak, Y., Brataas, A. & Bauer, G. E. W. Enhanced Gilbert Damping in Thin Ferromagnetic Films. Phys. Rev. Lett. **88**, 117601 (2002).

[157] Mihai Miron, I. et al. Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. Nat. Mater. **9**, 230-234, (2010).

[158] Miron, I. M. et al. Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. Nature **476**, 189-193, (2011).

[159] Miron, I. M. et al. Fast current-induced domain-wall motion controlled by the Rashba effect. Nat. Mater. **10**, 419-423, (2011).

[160] Gambardella, P. & Miron, I. M. Current-induced spin–orbit torques. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **369**, 3175-3197 (2011).

[161] 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

[162] 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).

[163] 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

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

[165] 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

[166] 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

[167] 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

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

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

[170] 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

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

[172] 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.

[173] 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.

[174] 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.

[175] 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.

[176] 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.

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

[178] 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.

[179] 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.

[180] 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.

[181] 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.

[182] 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.

[183] 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.

[184] 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.

[185] 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.

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

[187] 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

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

[189] S. Kim, et al, "Ultrathin (<10nm) 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.

[190] 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).

[191] 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.

[192] 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.

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

[194] 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).

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

[196] 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.

[197] 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.

[198] 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.

[199] 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).

[200] 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.

[201] 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

[202] 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.

[203] 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.

[204] K. Virwani, et al, "Sub-30nm 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.

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

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

[207] 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.

[208] 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.

[209] 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.

[210] 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.

[211] 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.

[212] 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.

[213] N. Banno, et al, "Nonvolatile 32x32 crossbar atom switch block integrated on a 65-nm CMOS platform," Symposium VLSI Tech., pp. 39-40, Jun. 2012.

[214] 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.

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

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

[217] 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.

[218] Y. Deng and J. Zhou, "Architectures and optimization methods of flash memory based storage systems", J. Syst. Arch. 57 (2011) 214-227.

[219] 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, p.24-33

[220] 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).

[221] 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).

[222] 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://cmrr.ucsd.edu/education/workshops/documents/Franceschini_Michael.pdf

[223] 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).

[224] 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).

[225] www.micron.com/about/innovations/3d-xpoint-technology

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

[227] www.eetimes.com/author.asp?section_id=36&doc_id=1327313

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

[229] Q. Cao, S.-J. Han, J. Tersoff, A. D. Franklin, Y. Zhu, Z. Zhang, G. S. Tulevski, J. Tang, and W. Haensch, Science 350, 68 (2015).

[230] A. D. Franklin, D. B. Farmer, and W. Haensch, ACS Nano 8, 7333 (2014).

[231] A. D. Franklin, M. Luisier, S. J. Han, G. Tulevski, C. M. Breslin, L. Gignac, M. S. Lundstrom, and W. Haensch, Nano Lett. 12, 758 (2012).

[232] 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, Nano Lett. 13, 2490 (2013).

[233] 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, Appl. Phys. Lett. 101, 053123 (2012).

[234] L. Ding, Z. Y. Zhang, S. B. Liang, T. Pei, S. Wang, Y. Li, W. W. Zhou, J. Liu, and L. M. Peng, Nat. Commun. 3 (2012).

[235] M. M. Shulaker, G. Hills, N. Patil, H. Wei, H.-Yu Chen, H.-S. Philip Wong, and S. Mitra, Nature 501, 526 (2013).

[236] Q. Cao, J. Tersoff, S. J. Han, and A. V. Penumatcha, Phys. Rev. Appl. 4, 024022 (2015).

[237] R. S. Park, M. M. Shulaker, G. Hills, L. S. Liyanage, S. Lee, A. Tang, S. Mitra, and H. -S. P. Wong, ACS Nano 10, 4599 (2016).

[238] G. J. Brady, Y. Joo, M. -Y. Wu, M. J. Shea, P. Gopalan, and M. S. Arnold, ACS Nano 8, 11614 (2014).

[239] 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, ACS Nano 8, 8730 (2014).

[240] S. Sato, Jpn. J. Appl. Phys. 54, 040102 (2015)

[241] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov: "Electric Field Effect in Atomically Thin Carbon Films", Science, 306, 666 (2004).

[242] Gong Gu, Shu Nie, R. M. Feenstra, R. P. Devaty, W. J. Choyke, Winston K. Chan and Michael G. Kane, "Field effect in epitaxial graphene on a silicon carbide substrate," Applied Physics Letters, Vol. 90, 253507, (2007).

[243] Kedzierski, J., Pei-Lan Hsu, Healey, P., Wyatt, P.W., Keast, C.L., Sprinkle, M., Berger, C., de Heer, W.A., "Epitaxial Graphene Transistors on SiC Substrates," IEEE Transactions on Electron Devices, Vol. 55 (8), pp. 2078 - 2085 (Aug. 2008).

[244] Lemme, M.C., Echtermeyer, T.J., Baus, M., Kurz, H., "A Graphene Field-Effect Device," IEEE Electron Device Letters, Vol. 28 (4), pp. 282 - 284 (April 2007).

[245] Seyoung Kim,, Junghyo Nah, Insun Jo, Davood Shahrjerdi, Luigi Colombo, Zhen Yao, Emanuel Tutuc, and Sanjay K. Banerjee, "Realization of a high mobility dual-gated graphene field-effect transistor with Al2O3 dielectric," Applied Physics Letters, Vol. 94, 062107, (2009).

[246] Naoki Harada, Katsunori Yagi, Shintaro Sato, and Naoki Yokoyama, "A polarity controllable graphene invertor, " Applied Physics Letters, Vol. 96, 012102, (2010)

[247] L. A. Ponomarenko, F. Schedin, M. I. Katsnelson, R. Yang, E. W. Hill, K. S. Novoselov, and A. K. Geim, "Chaotic Dirac billiard in graphene quantum dots," Science, Vol. 320, 356 (2008)

[248] Inanc Meric, Melinda Y. Han, Andrea F. Young, Barbaros Ozyilmaz, Philip Kim, Kenneth L. Shepard, "Current saturation in zero-bandgap, top-gated graphene field-effect transistors," Nature Nanotechnology, Vol. 3, 654 - 659 (2008).

[249] Lei Liao, Yung-Chen Lin, Mingqiang Bao, Rui Cheng, Jingwei Bai, Yuan Liu, Yongquan Qu, Kang L. Wang, Yu Huang. And Xiangfeng Duan, "High-speed graphene transistors with a self-aligned nanowire gate." Nature, Vol. 467, 305 (2010)

[250] Y. Q. Wu, Y.-M. Lin, K. A. Jenkins, J. A. Ot1, C. Dimitrakopoulos, D.B. Farmer, F. Xia, A. Grill, D.A. Antoniadis, and Ph. Avouris, "RF performance of short channel graphene field-effect transistor, " IEEE IEDM Technical Digest 2010, p.226 (2010).

[251] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. K. Banerjee, L. Colombo, and R. S. Ruoff, *Science*, vol. 324, 1312 (2009).

[252] Yanqing Wu, Yu-ming Lin, Ageeth A. Bol, Keith A. Jenkins, Fengnian Xia, Damon B. Farmer, Yu Zhu, and Phaedon Avouris, Nature 472, 74 (2011)

[253] Xu Du, Ivan Skachko, Anthony Barker, and Eva Y. Andrei, Nature Nanotechnol. 3, 491 (2008)

[254] K. I. Bolotin, K. J. Sikes, J. Hone, H. L. Stormer, and P. Kim, Phys. Rev. Lett. 101, 096802 (2008)

[255] Eduardo V. Castro, H. Ochoa, M. I. Katsnelson, R.V. Gorbachev, D. C. Elias, K. S. Novoselov, A. K. Geim, and F. Guinea, Phys Rev. Lett. 105, 266601 (2010)

[256] Jian-Hao Chen, Chaun Jang, Shudong Xiao, Masa Ishigami, Michael S. Fuhrer, Nature Nanotechnol. 3, 206 (2008)

[257] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, J. Hone, Nature Nanotechnol. 5, 722 (2010)

[258] Alexander S. Mayorov, Roman V. Gorbachev, Sergey V. Morozov, Liam Britnell, Rashid Jalil, Leonid A. Ponomarenko, Peter Blake, Kostya S. Novoselov, Kenji Watanabe,| Takashi Taniguchi, and A. K. Geim, Nano Lett. dx.doi.org/10.1021/nl200758b

[259] Xiaosong Wu, Yike Hu, Ming Ruan, Nerasoa K Madiomanana, John Hankinson, Mike Sprinkle, Claire Berger, and Walt A. de Heer, Appl. Phys. Lett. 95, 223108 (2009)

[260] M. Orlita, C. Faugeras, P. Plochocka, P. Neugebauer, G. Martinez, D. K. Maude, A.-L. Barra, M. Sprinkle, C. Berger, W. A. de Heer, and M. Potemski, Phys. Rev. Lett. 101, 267601 (2008)

[261] A. W. Tsen, L. Brown, M. P. Levendorf, F. Ghahari, P. Y. Huang, R. W. Havener, C. S. Ruiz-Vargas, D. A. Muller, P. Kim, and J. Park, Science 336, 1143 (2012).

[262] M. C. Lemme, T. J. Echtermeyer, M. Baus, and H. Kurz: "A Graphene Field Effect Device", IEEE Electron Device Lett. 28, 282 (2007).

[263] J. R. Williams, L. DiCarlo, C. M. Marcus: "Quantum Hall Effect in a Gate-Controlled p-n Junction of Graphene", Sicence 317, 638 (2007).

[264] D. B. Farmer, H.-Y. Chiu, Y.-M. Lin, K. A. Jenkins, F. Xia, and Ph. Avouris: "Utilization of a Buffered Dielectric to Achieve High Field-Effect Carrier Mobility in Graphene Transistors", Nano Lett. 9, 4474 (2009).

[265] Y. Q. Wu, P. D. Ye, M. A. Capano, Y. Xuan, Y. Sui, M. Qi, J. A. Cooper, T. Shen, D. Pandey, G. Prakash, and R. Reifenberger: "Top-gated Graphene Field-effect-transistors Formed by Decomposition of SiC", Appl. Phys. Lett., 92, 092102 (2008).

[266] S. Kim, J. Nah, I. Jo, D. Shahrjerdi, L. Colombo, Z. Yao, E. Tutuc, S. K. Banerjee: "Realization of a High Mobility Dual-Gated Graphene Field-Effect Transistor with Al2O3 Dielectric", Appl. Phys. Lett. 94, 062107 (2009).

[267] Lei Liao, Jingwei Bai, Yongquan Qu, Yung-chen Lin, Yujing Li, Yu Huangb, and Xiangfeng Duan, PANS, 107, 6711 (2010)

[268] K. Kim et al., ACS Nano 6, 8583 (2012)

[269] M. Y. Han, B. Ozyilmaz, Y. B. Zhang, and P. Kim: "Energy Band-Gap Engineering of Graphene Nanoribbons" Phys. Rev. Lett. 98, 206805 (2007)

[270] X. Li, X. Wang, Li Zhang, S. Lee, and H. Dai: "Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors", Science 319, 1229 (2008).

[271] L. Jiao, L. Zhang, X. Wang, G. Diankov, and H. Dai: "Narrow Graphene Nanoribbons from Carbon Nanotubes", Nature 458, 877 (2009).

[272] D. V. Kosynkin, A. L. Higginbotham, A. Sinitskii, J. R. Lomeda, A. Dimiev, B. K. Price, and J. M. Tour: "Longitudinal Unzipping of Carbon Nanotubes to Form Graphene Nanoribbons", Nature 458, 872 (2009).

[273] J. Campos-Delgado, J. M. Romo-Herrera, X. Jia, D. A. Cullen, H. Muramatsu, Y. A. Kim, T. Hayashi, Z. Ren, D. J. Smith, Y. Okuno, T. Ohba, H. Kanoh, K. Kaneko, M. Endo, H. Terrones, M. S. Dresselhaus, and M. Terrones: "Bulk Production of a New Form of sp2 Carbon: "Crystalline Graphene Nanoribbons", Nano Lett. 8, 2773 (2008).

[274] Xiaogan Liang, and Sungjin Wi, ACS Nano 6, 9700 (2012).

[275] Jinming Cai, Pascal Ruffieux, Rached Jaafar, Marco Bieri, Thomas Braun, Stephan Blankenburg, Matthias Muoth, Ari P. Seitsonen, Moussa Saleh, Xinliang Feng, Klaus Mu¨llen & Roman Fasel, Nature 466, 470 (2010).

[276] ACS Nano, 6, 6930 (2012)

[277] L. Yang, C.-H.Park, Y.-W.Son, M. L. Cohen, S. G. Louie, Phys. Rev. Lett. 99, 186801 (2007)

[278] ACS Nano, DOI: 10.1021/nn401948e (2013)

[279] J. Cai et al., Nature Nanotech 9, 896 (2014)

[280] Y. Chen et al., Nature Nanotech 10, 156 (2015).

[281] M.Bresciani, A.Paussa, P.Palestri, D.Esseni, L.Selmi, IEEE IEDM Technical Digest 2010, p.724 (2010).

[282] Melinda Y. Han, Juliana C. Brant, and Philip Kim, Phys. Rev Lett. 104, 056801 (2010)

[283] Patrick Gallagher, Kathryn Todd, and David Goldhaber-Gordon, Phys. Rev. B **81**, 115409 (2010)

[284] Xinglan Liu, Jeroen B. Oostinga, Alberto F. Morpurgo, and Lieven M. K. Vandersypen, Phys. Rev. B 80, 121407(R) (2009)

[285] J. Baringhaus et al., Nature 506, 349 (2014)

[286] McCann and Fal'ko PRL 96, 086805 (2006)

[287] McCann PRB 74, 161403R (2006)

[288] Oostinga et al., Nature Mat 7, 151 (2008)

[289] Zhang et al., Nature 459, 821 (2009)

[290] Xia et al., Nano Lett. 10, 715 (2010)

[291] K. Kim et al., Nature Mater. 12, 887 (2013)

[292] Woo Jong Yu, Lei Liao, Sang Hoon Chae, Young Hee Lee, and Xiangfeng Duan, Nano Lett. 11, 4759 (2011)

[293] J. Park et al., Adv. Mater., 24, 407 (2012).

[294] Tian et al., J. Phys. Chem. B, 114, 11377 (2010)

[295] Dinh Loc Duong, Seung Mi Lee, Sang Hul Chae, Quang Huy Ta, Si Young Lee, Gang Hee Han, Jung Jun Bae, and Young Hee Lee, PHYSICAL REVIEW B 85, 205413 (2012)

[296] A. J. Samuels and J. D. Carey, ACS Nano, 7, 2790 (2013).

[297] J. Bai et al., Nature Nanotech 5, 190 (2010)

[298] X. Liang et al., Nano Lett. 10, 2454 (2010)

[299] H. Jippo et al., X. Liang, Nano Lett. 10, 2454 (2010)

[300] S. Nakaharai et al., Proc. IEEE IEDM 2012, 72 (2012).

[301] S. Nakaharai et al., ACS Nano, DOI: 10.1021/nn401992q (2013).

[302] Y.Q. Wu, Y.-M. Lin, K.A. Jenkins, J.A. Ott, C. Dimitrakopoulos, D.B. Farmer, F. Xia, A. Grill, D.A. Antoniadis, and Ph. Avouris, IEEE IEDM Technical Digest 2010, p.226 (2010).

[303] Cheng et al., PNAS 109,11588 (2012)

[304] Wu et al., Nano Lett. 12, 3062 (2012)

[305] 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.

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

[i306] Beckman, R., Johnston-Halperin, E., Luo, Y., Green, J. E., & Hearh, J. R. (2005, Oct. 21). Bridging dimensions: demultiplexing ultrahigh-density nanowire circuits. *Science, 310*(5747), 465-468.

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

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

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

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

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

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

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

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

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

[317] 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).

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

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

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

[321] Lu, W., Xie, P., & Lieber, C. M. (2008, Nov.). Nanowire transistor performance limits and applications. *IEEE Trans. Elect. Dev., 55*(11), 2859-2876.

[322] Hashemi, P., Teherani, J. T. & Hoyt, J. L. Investigation of hole mobility in gate-all-around Si nanowire p-MOSFETs with high-κ/metal-gate: Effects of hydrogen thermal annealing and nanowire shape. in *Electron Devices Meet. (IEDM), 2010 IEEE Int.* 34–35 (IEEE, 2010).

[323] Singh, N. *et al.* Ultra-Narrow Silicon Nanowire Gate-All-Around CMOS Devices: Impact of Diameter, Channel-Orientation and Low Temperature on Device Performance. *Electron Devices Meet. 2006. IEDM '06. Int.* 1–4 (2006). doi:10.1109/IEDM.2006.346840

[324] Hu, Y., Xiang, J., Liang, G., Yan, H. & Lieber, C. M. Sub-100 Nanometer Channel Length Ge/Si Nanowire Transistors with Potential for 2 THz Switching Speed. *Nano Lett.* **8,** 925–930 (2008).

[325] J. del Alamo, "Nanometre-Scale Electronics with III-V Compound Semiconductors," Nature, vol. 479, 317, 2011.

[326] A. Nainani, D. Kim, T. Krishnamohan, and K. Saraswat, "Hole Mobility and Its Enhancement with Strain for Technologically Relevant III-V Semiconductors," SISPAD, 2009.

[327] B. Bennett, M. Ancona, J. Boos, and B. Shanabrook, "Mobility Enhancement in Strained p-InGaSb Quantum Wells," Appl. Phys. Lett., vol. 91, 042104, 2007.

[328] 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.,

[329] B. Bennett, T. Chick, M. Ancona, and J. Boos, "Enhanced Hole Mobility and Density in GaSb Quantum Wells," Solid State Electron., 79, 274, 2013.

[330] 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.

[331] 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," in IEDM Tech. Dig., 727, 2008.

[332] 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.

[333] K. Takei, *et al*. "Nanoscale InGaSb Heterostructure Membranes on Si Substrates for High Hole Mobility Transistors," Nano Lett., 12, 2060, 2012.

[334] 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.

[335] A. Nainani, *et al*, "Development of High-k Dielectric for Antimonides and a Sub 350ºC III-V pMOSFET Outperforming Germanium," IEDM Tech. Dig., 138, 2010.

[336] 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

[337] M. Ancona, B. Bennet, and J. Boos, "Scaling projections for Sb-based p-channel FETs," Solid State Electron., vol. 54, 1349, 2010.

14. 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.

15. 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.

16. 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

17. 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.

18. 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", in Symp. VLSI Tech. Dig., 2013, T54, 2013.

[338] 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.

[339] 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", *in Extended Abstracts of SSDM*, 2006, p.1056.

[340] 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", *in IEDM Tech. Dig.* , p.235, 2014.

[341] 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", in IEDM Tech. Dig. , p.418, 2014.

[342] 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", *in Extended Abstracts of SSDM*, 2011, p.835.

[343] 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", *in IEDM Tech. Dig.,* 2012, p.131.

[344] Heng Wu, Nathan Conrad, Wei Luo, and Peide D. Ye, "First Experimental Demonstration of Ge CMOS Circuits", *in IEDM Tech. Dig.* p.227, 2014.

[345] 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.

[346] 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", in IEDM Tech. Dig., 2011, p.753.

[347] S. Gupta, et al, "Towards High Mobility GeSn Channel nMOSFETs: Improved Surface Passivation Using Novel Ozone Oxidation Method", *in IEDM Tech. Dig.,* 2012, p.375.

[348] S. Gupta, et al, "GeSn Channel nMOSFETs: Material Potential and Technological Outlook", *in Symposium on VLSI Technology Digest of Technical Papers,* 2012, p.95.

[349] 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", *in Symposium on VLSI Technology Digest of Technical Papers*, 2013, p.T28.

[350] 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, *in Symposium on VLSI Technology Digest of Technical Papers*, 2013, T26.

[351] 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", *in IEDM Tech. Dig.,* 2012, p.509.

[352] 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, *in Symposium on VLSI Technology Digest of Technical Papers,* 2012, p.83.

[353] 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.

[354] 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).

[355] 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.

[356] S. Agarwal and E. Yablonovitch, "Designing a Low Voltage, High Current Tunneling Transistor," in *CMOS and Beyond Logic Switches for Terascale Integrated Circuits*, T.-J. King-Liu and K. Kuhn, Eds., ed Cambridge University Press, 2015.

[357] A. C. Seabaugh and Z. Qin, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE,* vol. 98, pp. 2095-2110, 2010.

[358] 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.

[359] 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.

[360] 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.

[361] 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.

[362] U. E. Avci and I. A. Young, "Heterojunction TFET Scaling and resonant-TFET for steep subthreshold slope at sub-9nm gate-length," in *Electron Devices Meeting (IEDM), 2013 IEEE International*, 2013, pp. 4.3.1-4.3.4.

[363] L. De Michielis, L. Lattanzio, K. E. Moselund, H. Riel, and A. M. Ionescu, "Tunneling and Occupancy Probabilities: How Do They Affect Tunnel-FET Behavior?," *Electron Device Letters, IEEE,* vol. 34, pp. 726-728, 2013.

[364] M. Luisier and G. Klimeck, "Simulation of nanowire tunneling transistors: From the Wentzel–Kramers–Brillouin approximation to full-band phonon-assisted tunneling," *Journal of Applied Physics,* vol. 107, p. 084507, 2010.

[365] D. Verreck, A. S. Verhulst, K. Kuo-Hsing, W. G. Vandenberghe, K. de Meyer, and G. Groeseneken, "Quantum Mechanical Performance Predictions of p-n-i-n Versus Pocketed Line Tunnel Field-Effect Transistors," *Electron Devices, IEEE Transactions on,* vol. 60, pp. 2128-2134, 2013.

[366] R. Kotlyar, U. E. Avci, S. Cea, R. Rios, T. D. Linton, K. J. Kuhn*, et al.*, "Bandgap engineering of group IV materials for complementary n and p tunneling field effect transistors," *Applied Physics Letters,* vol. 102, p. 113106, 2013.

[367] U. E. Avci, S. Hasan, D. E. Nikonov, R. Rios, K. Kuhn, and I. A. Young, "Understanding the feasibility of scaled III-V TFET for logic by bridging atomistic simulations and experimental results," in *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 183-184.

[368] S. Agarwal, G. Klimeck, and M. Luisier, "Leakage-Reduction Design Concepts for Low-Power Vertical Tunneling Field-Effect Transistors," *Electron Device Letters, IEEE,* vol. 31, pp. 621-623, 2010.

[369] F. Conzatti, M. G. Pala, D. Esseni, E. Bano, and L. Selmi, "Strain-Induced Performance Improvements in InAs Nanowire Tunnel FETs," *Electron Devices, IEEE Transactions on,* vol. 59, pp. 2085-2092, 2012.

[370] S. S. Sylvia, M. A. Khayer, K. Alam, and R. K. Lake, "Doping, Tunnel Barriers, and Cold Carriers in InAs and InSb Nanowire Tunnel Transistors," *Electron Devices, IEEE Transactions on,* vol. 59, pp. 2996-3001, 2012.

[371] L. Yeqing, Z. Guangle, L. Rui, L. Qingmin, Z. Qin, T. Vasen*, et al.*, "Performance of AlGaSb/InAs TFETs With Gate Electric Field and Tunneling Direction Aligned," *Electron Device Letters, IEEE,* vol. 33, pp. 655-657, 2012.

[372] K. Jeon, W.-Y. Loh, P. Patel, C. Y. Kang, J. Oh, A. Bowonder*, et al.*, "Si tunnel transistors with a novel silicided source and 46mV/dec swing," in *VLSI Technology (VLSIT), 2010 Symposium on*, 2010, pp. 121-122.

[373] 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.

[374] 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.

[375] 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.

[376] 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," in *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 8.5.1-8.5.4.

[377] 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.

[378] 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," in VLSI Technology (VLSIT), 2012 Symposium on, 2012, pp. 49-50.

[379] K. Sung Hwan, K. Hei, H. Chenming, and L. Tsu-Jae King, "Germanium-source tunnel field effect transistors with record high Ion/Ioff," in VLSI Technology, 2009 Symposium on, 2009, pp. 178-179.

[380] 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," in Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 2008, pp. 1-3.

[381] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, et al., "Fabrication, characterization, and physics of III-V heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing," in Electron Devices Meeting (IEDM), 2011 IEEE International, 2011, pp. 33.6.1-33.6.4.

[382] 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.

[383] K. Tomioka, M. Yoshimura, and T. Fukui, "Steep-slope tunnel field-effect transistors using III-V nanowire/Si heterojunction," in *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 47-48.

[384] 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.

[385] 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.

[386] S. Johnson and T. Tiedje, "Temperature dependence of the Urbach edge in GaAs," *Journal of applied physics,* vol. 78, pp. 5609-5613, 1995.

[387] 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.

[388] J. I. Pankove, "Absorption Edge of Impure Gallium Arsenide," *Physical Review,* vol. 140, pp. A2059-A2065, 12/13/ 1965.

[389] S. Agarwal and E. Yablonovitch, "The low voltage TFET demands higher perfection than previously required in electronics," in *Device Research Conference (DRC), 2015 73rd Annual*, 2015, pp. 247-248.

[390] S. Agarwal and E. Yablonovitch, "Pronounced Effect of pn-Junction Dimensionality on Tunnel Switch Threshold Shape," *arXiv preprint arXiv:1109.0096,* 2011.

[391] ITRS 2013 edition.

[392] S. Datta and B. Das, "Electronic analog of the electro-optic modulator", *Appl. Phys. Lett.* vol. 56, pp.665-667, 1990.

[393] 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.

[394] S. Sugahara and J. Nitta, "Spin-Transistor Electronics: An Overview and Outlook", *Proc. IEEE*, vol. 98, pp. 2124-2154, 2010.

[395] 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.

[396] 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.

[397] 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.

[398] 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.

[399] 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.

[400] 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.

[401] 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.

[402] 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.

[403] 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.

[404] 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.

[405] 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.

[406] 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.

[407] 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.

[408] 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.

[409] 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.

[410] I. Appelbaum, B. Huang, and D. J. Monsma, "Electronic measurement and control of spin transport in silicon", *Nature*, vol. 447, pp. 295-298, 2007.

[411] 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.

[412] 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.

[413] 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.

[414] 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.

[415] 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.

[416] 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.

[417] 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.

[418] 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.

[419] 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.

[420] 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.

[421] 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.

[422] 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.

[423] 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.

[424] 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.

[425] 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.

[426] 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.

[427] 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.

[428] 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.

[429] 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.

[430] 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.

[431] 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.

[432] 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.

[433] 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.

[434] 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.

[435] 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.

[436] 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.

[437] 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.

[438] 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.

[439] 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.

[440] J. Nitta, T. Akazaki, H. Takayanagi, T. Enoki, "Gate Control of Spin-Orbit Interaction in an Inverted $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ Heterostructure", Phys. Rev. Lett. vol. 78, pp. 1335-1338 , 1997; 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.

[441] 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.

[442] S. Kettemann, "Dimensional control of antilocalozation and spin relaxation in quantum wires", *Phys. Rev. Lett.* vol. 98, pp. 176808/1–4, 2007.

[443] 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.

[444] J. Schliemann, J. C. Egues, D. Loss, "Nonballistic spin-field-effect transistor", *Phys. Rev. Lett.* vol. 90, pp. 1468011–1468014, 2003.

[445] 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.

[446] 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.

[447] 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.

[448] 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.

[449] 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.

[450] 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.

[451] 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.

[452] 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.

[453] A. Rusu, G. A. Salvatore, D. Jiménez, A. M. Ionescu, "Metal-Ferroelectric-Metal- Oxide-Semiconductor Field Effect Transistor with Sub-60mV/decade Subthreshold Swing and Internal Voltage Amplification", IEDM 2010 , San Francisco, USA, 06-08 December 2010.

[454] 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.

[455] 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.

[456] Muller, J., Boscke, T. S., Muller, S., Yurchuk, E., Polakowski, P., Paul, J., ... & Weinreich, W. (2013, December). Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories. In*2013 IEEE International Electron Devices Meeting*.

[457] Cheng, C. H., & Chin, A. (2014). Low-Voltage Steep Turn-On pMOSFET Using Ferroelectric High-Gate Dielectric. *Electron Device Letters, IEEE*,*35*(2), 274-276.

[458] M-H Lee et al, "Prospects for Ferroelectric HfZrOx FETs with Experimentally CET=0.98nm, SSfor=42mV/dec, SSrev=28mV/dec, Switch-OFF <0.2V, and Hysteresis-Free Strategies," Proceedings of IEDM, 2015.

[459] Kai-Shin Li et al, "Sub-60mV-Swing Negative-Capacitance FinFET without Hysteresis," Proceedings of IEDM, 2015.

[460] K. Akarvardar *et al*., "Design considerations for complementary nanoelectromechanical logic gates," IEEE International Electron Devices Meeting Technical Digest, pp. 299–302, 2007.

[461] F. Chen *et al*., "Integrated circuit design with NEM relays," IEEE/ACM International Conference on Computer-Aided Design, pp. 750–757, 2008.

[462] 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.

[463] 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.

[464] 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.

[465] H. Fariborzi *et al*., "Analysis and demonstration of MEM-relay power gating," Custom Integrated Circuits Conference, 2010.

[466] 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.

[467] J. O. Lee *et al*., "A sub-1-volt nanoelectromechanical switching device," *Nature Nanotechnology*, Vol. 8, pp. 36-40, 2013.

[468] R. Nathanael *et al*., "4-terminal relay technology for complementary logic," IEEE International Electron Devices Meeting Technical Digest, pp. 223-226, 2009.

[469] T.-J. K. Liu *et al*., "Prospects for MEM_relay logic switch technology," IEEE International Electron Devices Meeting Technical Digest, pp. 424-427, 2010.

[470] 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.

[471] 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.

[472] 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.

[473] H. Fariborzi *et al*., "Design and demonstration of micro-electro-mechanical relay multipliers," IEEE Asian Solid-State Circuits Conference, 2011.

[474] 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.

[475] L. Hutin *et al*., "Electromechanical diode cell scaling for high-density nonvolatile memory," *IEEE Transactions on Electron Devices*, Vol. 61, pp. 1382-1387, 2014.

[476] Y. Chen *et al*., "Reliability of MEM relays for zero leakage logic," SPIE Conference Proceedings Vol. 8614, 2013.

[477] 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.

[478] 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.

[479] C. Zhou, D. M. Newns, J. A. Misewich and P. C. Pattnaik, Appl Phys Lett **70** (5), 598-600 (1997)

[480] D. M. Newns, J. A. Misewich, C. C. Tsuei, A. Gupta, B. A. Scott and A. Schrott, Appl Phys Lett **73** (6), 780-782 (1998)

[481] 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.

[482] 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.

[483] Z. Yang, C. Ko and S. Ramanathan, Annual Review of Materials Research **41**, 8.1 (2011).

[484] A. Cavalleri, Cs. Toth, C. W. Siders, J. A. Squier, F. Raksi, P. Forget and J. C. Keiffer, Phy. Rev. Lett. 87, 237401 (2001)

[485] S. Hormoz and S. Ramanathan, Solid State Electron **54** (6), 654-659 (2010)

[486] H. T. Kim, B. G. Chae, D. H. Youn, S. L. Maeng, G. Kim, K. Y. Kang and Y. S. Lim, New J Phys **6**, 52 (2004)

[487] G. Stefanovich, A. Pergament and D. Stefanovich, J. Phys: Cond Mat, 12, 8837 (2000)

[488] D. Ruzmetov, G. Gopalakrishnan, C. Ko, V. Narayanamurti and S. Ramanathan, J Appl Phys **107** (11), 114516 (2010)

[489] Z. Yang, Y. Zhou, and S. Ramanathan, J. Appl. Phys. 111, 014506 (2012)

[490] M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa and Y. Tokura, Nature 487, 459–462 (2012)

[491] Y. Zhou and S. Ramanathan, J. Appl. Phys. 111, 084508 (2012)

[492] H. Ji , Jiang Wei , and D. Natelson, Nano Lett., 12, 2988–2992 (2012)

[493] J Jeong, N Aetukuri, T Graf, TD Schladt, MG Samant and S. S. Parkin, Science, 339(6126), 1402-1405.

[494] J. Shi, S. D. Ha, Y. Zhou, F. Schoofs and S. Ramanathan, Nature Communications, 4, 2676 (2013)

[495] S. D. Ha, J. Shi, Y. Meroz, L. Mahadevan and S. Ramanathan, Physical Review Applied, 2, 064003 (2014);

[496] M. D. Pickett, G. Medeiros-Ribeiro and R. Stanley Williams, Nature Materials 12, 114–117 (2013)

[497] N. Shukla, A. V. Thathachary, A. Agrawal, H. Paik, A. Aziz, D. G. Schlom, S. K.Gupta, R. Engel-Herbert and S. Datta, Nature Communications 6, 7812 (2015)

[498] Y. Zhou, X. Chen, Z. Yang, C. Mouli and S. Ramanathan, IEEE Electron Device Letters, 34, 220 (2013)

[499] J. Leroy, A. Crunteanu, A. Bessaudou, F. Cosset, C. Champeaux, and J. C. Orlianges, (2012). Applied Physics Letters, 100, 213507 (2012).

[500] Y. Zhang and S. Ramanathan, Solid State Electronics, August 2011, pp. 161-164 (2011)

[501] S. D. Ha, G. H. Aydogdu and S. Ramanathan, Appl. Phys. Lett., 98, 012105 (2011)

[502] P. Lacorre, J. B. Torrance, J. Pannetier, A. I. Nazzal, P. W. Wang and T. C. Huang, J. Sol. St. Chem. 91, 225 (1991)

[503] 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.

[504] W. L. Lim, E. J. Moon, J. W. Freeland, D. J. Meyers, M. Kareev, J. Chakhalian, and S. Urazhdin, Appl. Phys. Lett. 101, 143111 (2012)

[505] 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, Appl. Phys. Lett. 97, 142110 (2010)

[506] R. Scherwitzl, P. Zubko, I. G. Lezama, S. Ono, A. F. Morpurgo, G. Catalan and J.-M. Triscone, Adv. Mater. 22, 5517 (2010).

[507] S. D. Ha, U. Vetter, J. Shi, and S. Ramanathan, Appl. Phys. Lett. 102, 183102 (2013)

[508] S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, Appl. Phys. Lett. 106, 122102 (2015)

[509] S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, Appl. Phys. Lett. 106, 122102 (2015)

[510] J. Shi, Y. Zhou and S. Ramanathan, Nature Communications, 5, 4860 (2014)

[511] D. Newns, B. Elmegreen, X. Hu Liu, and G. Martyna, "A low-voltage high-speed electronic switch based on piezoelectric transduction," *Journal of Applied Physics,* vol. 111, p. 084509, 2012.

[512] P. M. Solomon, B. A. Bryce, M. A. Kuroda, R. Keech, S. Shetty, T. M. Shaw*, et al.*, "Pathway to the Piezoelectronic Transduction Logic Device," *Nano Letters,* vol. 15, pp. 2391-2395, 2015/04/08 2015.

[513] D. M. Newns, B. G. Elmegreen, X.-H. Liu, and G. J. Martyna, "The piezoelectronic transistor: A nanoactuator-based post-CMOS digital switch with high speed and low power," *MRS Bulletin,* vol. 37, pp. 1071-1076, 2012.

[514] I.-B. Magdău, X.-H. Liu, M. A. Kuroda, T. M. Shaw, J. Crain, P. M. Solomon*, et al.*, "The piezoelectronic stress transduction switch for very large-scale integration, low voltage sensor computation, and radio frequency applications," *Applied Physics Letters,* vol. 107, p. 073505, 2015.

[515] B. C. Josephine, M. Hiroyuki, C. Matthew, M. S. Paul, L. Xiao-Hu, M. S. Thomas*, et al.*, "First realization of the piezoelectronic stress-based transduction device," *Nanotechnology,* vol. 26, p. 375201, 2015.

[516] S. Agarwal and E. Yablonovitch, "A Nanoscale Piezoelectric Transformer for Low-Voltage Transistors," *Nano Letters,* vol. 14, pp. 6263-6268, 2014/11/12 2014.

[517] A. Khitun and K. Wang, Superlattices & Microstructures **38**, 184-200 (2005).

[518] A. Khitun, M. Bao and K. L. Wang, IEEE Transactions on Magnetics **44** (9), 2141-2152 (2008).

[519] A. Khitun, M. Bao, J.-Y. Lee, K. L. Wang, D. W. Lee, S. X. Wang and I. V. Roshchin, Journal of Nanoelectronics and Optoelectronics **3** (1), 24-34 (2008).

[520] A. Khitun and K. L. Wang, JOURNAL OF APPLIED PHYSICS **110** (3) (2011).

[521] A. Khitun, JOURNAL OF APPLIED PHYSICS **113** (16) (2013).

[522] A. Khitun, JOURNAL OF APPLIED PHYSICS **111** (5) (2012).

[523] Y. Wu, M. Bao, A. Khitun, J.-Y. Kim, A. Hong and K. L. Wang, Journal of Nanoelectronics and Optoelectronics **4** (3), 394-397 (2009).

[524] 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, Applied Physics Letters **104** (8) (2014).

[525] F. Gertz, A. Kozhevnikov, Y. Filimonov and A. Khitun, Magnetics, IEEE Transactions on **51** (4), 4002905-4400910 (2015).

[526] A. Kozhevnikov, F. Gertz, G. Dudko, Y. Filimonov and A. Khitun, Applied Physics Letters **106** (14), 142409-142405 (2015).

[527] A. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod, "Majority logic gate for Magnetic Quantum-dot Cellular Automata," *Science,* vol. 311, pp. 205-208, Jan 13 2006.

[528] M. Niemier, X. Ju, M. Becherer, G. Csaba, X. S. Hu, D. Schmitt-Landsiedel*, 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.

[529] Á. Papp, M. T. Niemier, Á. Csurgay, M. Becherer, S. Breitkreutz, J. Kiermaier*, et al.*, "Threshold Gate Based Circuits from Nanomagnet Logic," *submitted to IEEE T. on Nanotechnology,* 2013.

[530] S. Breitkreutz, I. Eichwald, J. Kiermaier, A. Papp, G. Csaba, M. Niemier*, 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.

[531] M. Becherer, J. Kiermaier, S. Breitkreutz, G. Csaba, X. Ju, J. Rezgani*, et al.*, "On-chip Extraordinary Hall-effect sensors for characterization of nanomagnetic logic devices," *Solid State Electronics,* vol. 54, pp. 1027-1032, 2010.

[532] J. M. Shaw, S. E. Russek, T. Thomson, M. J. Donahue, B. D. Terris, O. Hellwig*, et al.*, "Reversal mechanisms in perpendicularly magnetized nanostructures," *Physical Review B,* vol. 78, p. 024414, 2008.

[533] D. Bhowmik, L. You, and S. Salahuddin, "Spin Hall effect clocking of nanomagnetic logic without a magnetic field," *Nat Nano,* vol. 9, pp. 59-63, 01//print 2014.

[534] G. Csaba, P. Lugli, and W. Porod, "Power dissipation in nanomagnetic logic devices," in *IEEE Conference on Nanotechnology*, 2004, pp. 346-348.

[535] 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, pp. 1-24, 2011.

[536] 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.

[537] 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, pp. 668-670, 2010.

[538] 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," in *INTERMAG / MMM*, Chicago, IL, 2013.

[539] M. A. J. Siddiq, M. T. Niemier, G. H. Bernstein, W. Porod, and X. S. Hu, "A Field Coupled Electrical Input for Nanomagnet Logic," *accepted for publication in IEEE Transactions on Nanotechnology,* 2013.

[540] 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.

[541] A. Lyle, J. Harms, T. Klein, A. Lentsch, A. Klemm, D. Martens*, et al.*, "Integration of spintronic interface for nanomagnetic arrays," *AIP Advances,* vol. 1, pp. 042177-11, 2011.

[542] A. Lyle, J. Harms, T. Klein, A. Lentsch, D. Martens, A. Klemm*, et al.*, "Spin transfer torque programming dipole coupled nanomagnet arrays," *Applied Physics Letters,* vol. 100, pp. 012402-3, 2012.

[543] 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, p. 092502, 2011.

[544] 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, pp. 203-214, 2013.

[545] 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, pp. 348-351, May 2010.

[546] M. T. Alam, S. Kurtz, M. J. Siddiq, M. T. Niemier, G. H. Bernstein, X. S. Hu*, et al.*, "On-chip Clocking of Nanomagnet Logic Lines and Gates," *IEEE Transactions on Nanotechnology,* vol. 11, pp. 273-286, 2012.

[547] P. Li, G. Csaba, V. K. Sankar, X. S. Hu, M. Niemier, W. Porod*, et al.*, "Power Reduction in Nanomagnet Logic Clocking through High Permeability Dielectrics," in *Device Research Conference*, State College, Pennsylvania, 2012, pp. 129-130.

[548] P. Li, G. Csaba, V. K. Sankar, X. S. Hu, M. Niemier, W. Porod*, et al.*, "Paths to Clock Power Reduction via High Permeability Dielectrics for Nanomagnet Logic Circuits," in *Joint MMM/Intermag Conference*, Chicago, IL, 2013.

[549] Y. H. Chu, L. W. Martin, M. B. Holcomb, M. Gajek, S. J. Han, Q. He*, et al.*, "Electric-field control of local ferromagnetism using a magnetoelectric multiferroic," *Nat. Mat.,* vol. 7, pp. 478-482, Jun 2008.

[550] F. M. Salehi, K. Roy, J. Atulasimha, and S. Bandyopadhyay, "Magnetization dynamics, Bennett clocking and associated energy dissipation in multiferroic logic," *Nanotechnology,* vol. 22, p. 155201, 2011.

[551] 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," in *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 29.7.1-29.7.4.

[552] 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, pp. 07A715-3, 04/01/ 2012.

[553] I. Eichwald, A. Bartel, J. Kiermaier, S. Breitkeutz, G. Csaba, D. Schmitt-Landsiedel*, et al.*, "Nanomagnet Logic: error-free directed signal transmission by an inverter chain," *IEEE Transactions on Magnetics,* vol. 48, pp. 4332-4335, 2012.

[554] S. Breitkreutz, J. Kiermaier, I. Eichwald, X. Ju, G. Csaba, D. Schmitt-Landsiedel*, et al.*, "Majority Gate for Nanomagnetic Logic with Perpendicular Magentic Anisotropy," *IEEE Transactions on Magnetics,* vol. 48, pp. 4336-4339, 2012.

[555] I. Eichwald, J. Wu, J. Kiermaier, S. Breitkreutz, G. Csaba, D. Schmitt-Landsiedel*, et al.*, "Towards a Signal Crossing in double-layer Nanomagnetic Logic," *to appear in IEEE Transactions on Magnetics,* vol. 49, 2013.

[556] S. Breitkreutz, J. Kiermaier, I. Eichwald, C. Hildbrand, G. Csaba, D. Schmitt-Landsiedel*, et al.*, "Experimental Demonstration of a 1-bit Full Adder in Perpendicular Nanomagnetic Logic," *to appear in IEEE Transactions on Magnetics,* vol. 49, 2013.

[557] E. Irina, B. Stephan, Z. Grazvydas, C. György, P. Wolfgang, and B. Markus, "Majority logic gate for 3D magnetic computing," *Nanotechnology,* vol. 25, p. 335202, 2014.

[558] R. Perricone, Z. Yining, K. M. Sanders, X. S. Hu, and M. Niemier, "Towards systematic design of 3D pNML layouts," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*, 2015, pp. 1539-1542.

[559] 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, pp. 07E341-3, 04/01/ 2012.

[560] 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, pp. 399-407, 2013

[561] W.-G. Wang, M. Li, S. Hageman, and C. L. Chien, "Electric-field-assisted switching in magnetic tunnel junctions," *Nat Mater,* vol. 11, pp. 64-68, 01//print 2012.

[562] M. Becherer, J. Kiermaier, S. Breitkreutz, I. Eichwald, G. Csaba, and D. Schmitt-Landsiedel, "Nanomagnetic Logic clocked in the MHz regime," in *The Proceedings of the 43rd European Solid-State Device Research Conference (ESSDERC)*, Bucharest, Romania, 2013.

[563] D. B. Carlton, N. C. Emley, E. Tuchfeld, and J. Bokor, "Simulation Studies of Nanomagnet-Based Logic Architecture," *Nano Letters,* vol. 8, pp. 4173-8, 2008.

[564] M. T. Niemier, G. H. Bernstein, G. Csaba, A. Dingler, X. S. Hu, S. Kurtz*, et al.*, "Nanomagnet Logic: Progress Toward System-Level Integration," *J. Phys. Con. Mat.,* vol. 23, p. 493202, 2011.

[565] 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, p. 053202, 2011.

[566] Q. Weikang and M. D. Riedel, "The synthesis of robust polynomial arithmetic with stochastic logic," in *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, 2008, pp. 648-653.

[567] R. Venkatesan, S. Venkataramani, F. Xuanyao, K. Roy, and A. Raghunathan, "Spintastic: Spin-based stochastic logic for energy-efficient computing," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*, 2015, pp. 1575-1578.

[568] R. Perricone, Yang Liu, Aaron Dingler, X. Sharon Hu, and M. Niemier, "Design of Stochastic Computing Circuits using Nanomagnetic Logic," *accepted in IEEE Transactions on Nanotechnology,* 2015.

[569] D. Nikonov and I. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," in *International Electron Devices Meeting (IEDM)*, San Francisco, CA, 2012, pp. 25.4.1-25.4.4.

[570] G. Csaba and W. Porod, "Behavior of Nanomagnet Logic in the presence of thermal noise," in *Computational Electronics (IWCE), 2010 14th International Workshop on*, 2010, pp. 1-4.

[571] E. Varga, G. Csaba, G. H. Bernstein, and W. Porod, "Domain-Wall Assisted Switching of Single-Domain Nanomagnets," *Magnetics, IEEE Transactions on,* vol. 48, pp. 3563-3566, 2012.

[572] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers", J. Magn. Magn. Mater. vol. 159, 1996, p. L1

[573] D. E. Nikonov, G. I. Bourianoff, and T. Ghani, "Proposal of a spin torque majority gate logic", IEEE Electr. Device Lett. vol. 32, 2011, p. 1128

[574] D. E. Nikonov, S. Manipatruni, and I. A. Young, "Switching efficiency improvement in spin torque majority gates", J. Appl. Phys. 115, 17C736 (2014)

[575] Dmitri E. Nikonov, Sasikanth Manipatruni, and Ian A. Young, "Switching efficiency improvement in spin torque majority gates", Journal of Applied Physics 115, 17C736 (2014)

[576] Dmitri E Nikonov, Sasikanth Manipatruni and Ian A Young, "Cascade-able spin torque logic gates with input–output isolation", Physica Scripta, Volume 90, Number 7, 074047 (2015)

[577] T. Windbacher, A. Makarov, V. Sverdlov, and S. Selberherr, "Novel Buffered Magnetic Logic Gate Grid", ECS Transactions, 66 (4) 295-303 (2015)

[578]B. Behin-Aein, D. Datta, S. Salahuddin and S. Datta, *Nature Nanotech*. vol. 5, p.266, (2010).

[579] F.J. Jedeema, A.T. Filip and B.J. Van Wees, *Nature*, vol. 410, p.345, (2001).

[580]F. J. Jedeema, M. S. Nijboer, A. T. Filip and B. J. van Wees, *Phys. Rev. B.,* vol. 67, p. 085319 (2003).

[581]B.T. Jonker et al., *Nature Phys.*, vol. 5, p. 817 (2006).

[582]I. Appelbaum, B. Huang and D.J. Monsma, *Nature*, vol. 447, p. 295 (2007).

[583] N. Tombros, et al., *Nature*, 448, p. 571 (2007).

[584] Lou, X. et al., *Nature Physics*, vol. 3, p.197 (2007).

[585] T. Yang, K. Kimura and Y. Otani, *Nature Phys.*, vol. 4, p. 851, (2008).

[586] J.Z. Sun et al., *APL,* vol. 95, p.083506, (2009).

[587] F. J. Jedeema, M. S. Nijboer, A. T. Filip and B. J. van Wees, *Phys. Rev. B.,* vol. 67, p. 085319 (2003).

[588] B.T. Jonker et al., *Nature Phys.*, vol. 5, p. 817 (2006).

[589]B. Behin-Aein, A. Sarkar, S. Srinivasan and S. Datta, *APL,* vol. 98, p. 123510, (2011).

[590]D. Weller et al., IEEE Tran. Magn. 36, 10, (2000)

[591] S. B. Akers, Proc. 3rd Ann. Symp. on Switching Circuit Theory and Logical Design, pp. 150, (1962).

[592] D. E. Nikonov and I. A. Young, "Overview of Beyond-CMOS Devices and a Uniform Methodology for Their Benchmarking," Proceedings of the IEEE, vol. PP, pp. 1-1, 2013.

[593] A. E. Standard. (2001). Advanced Encryption Standard (AES) FIPS Pub 197 (2001, Nov.). Available: http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf

[594] A. Iyengar, K. Ramclam, and S. Ghosh, "DWM-PUF: A low-overhead, memory-based security primitive," in Hardware-Oriented Security and Trust (HOST), 2014 IEEE International Symposium on, 2014, pp. 154-159.

[595] G. S. Rose, N. McDonald, Y. Lok-Kwong, and B. Wysocki, "A write-time based memristive PUF for hardware security applications," in Computer-Aided Design (ICCAD), 2013 IEEE/ACM International Conference on, 2013, pp. 830-833.

[596] 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," in Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific, 2014, pp. 73-78.

[597] A. Colli, S. Pisana, A. Fasoli, J. Robertson, and A. C. Ferrari, "Electronic transport in ambipolar silicon nanowires," physica status solidi (b), vol. 244, pp. 4161-4164, 2007.

[598] R. Martel, V. Derycke, C. Lavoie, J. Appenzeller, K. K. Chan, J. Tersoff, et al., "Ambipolar Electrical Transport in Semiconducting Single-Wall Carbon Nanotubes," Physical Review Letters, vol. 87, p. 256805, 12/03/ 2001.

[599] A. K. Geim and K. S. Novoselov, "The rise of graphene," Nat Mater, vol. 6, pp. 183-191, 03//print 2007.

[600] 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, pp. 481-489, 2005.

[601] N. Harada, K. Yagi, S. Sato, and N. Yokoyama, "A polarity-controllable graphene inverter," Applied Physics Letters, vol. 96, p. 012102, 2010.

[602] A. Heinzig, S. Slesazeck, F. Kreupl, T. Mikolajick, and W. M. Weber, "Reconfigurable Silicon Nanowire Transistors," Nano Letters, vol. 12, pp. 119-124, 2012/01/11 2012.

[603] S. Das and J. Appenzeller, "WSe2 field effect transistors with enhanced ambipolar characteristics," Applied Physics Letters, vol. 103, p. 103501, 2013.

[604] M. De Marchi, D. Sacchetto, S. Frache, J. Zhang, P. Gaillardon, Y. Leblebici, et al., "Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire FETs," in Electron Devices Meeting (IEDM), 2012 IEEE International, 2012, pp. 8.4.1-8.4.4.

[605] T. Vasen, "Investigation of III-V Tunneling Field-Effect Transistors," A Dissertation submitted to the University of Notre Dame2014.

[606] Y. Bi, K. Shamsi, J. S. Yuan, P.-E. Gaillardon, G. D. Micheli, X. Yin, et al., "Emerging Technology Based Design of Primitives for Hardware Security," {ACM Journal on Emerging Technologies in Computing Systems (JETC), vol. to appear, 2015.

[607] L. Britnell, R. V. Gorbachev, A. K. Geim, L. A. Ponomarenko, A. Mishchenko, M. T. Greenaway, et al., "Resonant tunnelling and negative differential conductance in graphene transistors," Nat Commun, vol. 4, p. 1794, 04/30/online 2013.

[608] MishchenkoA, J. S. Tu, CaoY, R. V. Gorbachev, J. R. Wallbank, M. T. Greenaway, et al., "Twist-controlled resonant tunnelling in graphene/boron nitride/graphene heterostructures," Nat Nano, vol. 9, pp. 808-813, 10//print 2014.

[609] 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, pp. 200-207, 2015.

[610] 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," in Test Symposium (ATS), 2014 IEEE 23rd Asian, 2014, pp. 342-347.

[611] A. Chen, X. S. Hu, Y. Jin, M. Niemier, and X. Yin, "Using Emerging Technologies for Hardware Security Beyond PUFs," presented at the to appear in Design, Automation, and Test in Europe (DATE), Dresden, Germany, 2016.

[612] J. Rajendran, Z. Huan, Z. Chi, G. S. Rose, P. Youngok, O. Sinanoglu, et al., "Fault Analysis-Based Logic Encryption," Computers, IEEE Transactions on, vol. 64, pp. 410-424, 2015.

[613] H. Mamiya, A. Miyaji, and H. Morimoto, "Efficient Countermeasures against RPA, DPA, and SPA," in Cryptographic Hardware and Embedded Systems - CHES 2004. vol. 3156, M. Joye and J.-J. Quisquater, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 343-356.

[614] B. Sedighi, J. J. Nahas, M. Niemier, and X. S. Hu, "Boolean circuit design using emerging tunneling devices," in Computer Design (ICCD), 2014 32nd IEEE International Conference on, 2014, pp. 355-360.

[615] Kuon, I. & Rose, J., 2006. Measuring the gap between FPGAs and ASICs. In *Proceedings of the 2006 ACM/SIGDA 14th international symposium on Field Programmable Gate Arrays.* Monterey, CA, USA: ACM, pp. 21-30.

[616] Gojman, B., et al., 3-D nanowire-based programmable login. In *Proc. Nanonet Conf.* pp. 1-5.

[617] Jo, S. H. et al., 2010. Nanoscale memristor device as synapse in neurmorphic systems. *Nano letters*, 10(4), pp. 1297-1301.

[618] Snider, G. S. & Williams, R. S., 2007. Nano/CMOS architectures using a field-programmable nanowired interconnect. *NANOTECHNOLOGY, 18(3), p. 35204.*

[619] Ferrucci, D., et al., 2010. Building Watson: An overview of the DeepQA project., *AI magazine*, 31(3), pp. 59-79.

[620] Krichmar, J. L. et al., 2010. Neuromorphic modeling abstractions and simulatoin of large-scale cortical networks. In *Proceedings of the International Conference on Computer-Aided Design.* Pp. 334-338.

[621] Yakopcic, C. et al., 2012. Memristor SPICE Modeling. In *Advances in Neuromorphic Memristor Science and Applications.* Springer, pp. 211-244.

[622] Likharev, K. K., 2011. CrossNets: Neuromorphic hybrid CMOS/nanoelectronic networks. *Science of Advanced Materials,* 3(3), pp. 322-331.

[623] L. Liao, Y.-C. Lin, M.Bao, R. Cheng, J.Bai, Y. Liu, Y.Qu, K. L. Wang, Y. Huang, X.Duan, "High-speed graphene transistors with a self-aligned nanowire gate," Nature 467, 305 (2010)

[624] Ph. Avouris, Y.-M. Lin, F. Xia, D.B. Farmer, T. Mueller, C. Dimitrakopoulos, K. Jenkins, A. Grill, "Graphene-Based Fast Electronics and Optoelectronics," IEDM 2010, pp. 552 – 555

[625] J. Lee, H.-J. Chung, J. Lee, H. Shin, J. Heo, H. Yang, S.-H. Lee, J. Shin, S. Seo, U. Chung, I. Yoo, K. Kim, "RF Performance of Pre-Patterned Locally-Embedded-Back-Gate Graphene Device," IEDM 2010, pp. 568 – 571

[626] D.Waldmann, J.Jobst, F. Speck, T.Seyller, M. Krieger, H. B. Weber, "Bottom-gated epitaxial graphene," Nature Mat. 10, 357 – 360 (2011)

[627] H. Lv, H. Wu, J. Liu, C. Huang, J. Li, J. Yu, J. Niu, Q. Xu, Z. Yu, and H. Qian, "Inverted process for graphene integrated circuits fabrication," Nanoscale, vol. 6, pp. 5826-5830 (2014).

[628] L. Liao, J.Bai, R. Cheng, Y.-C. Lin, S. Jiang, Y.Qu, Y. Huang, X.Duan, "Sub-100 nm Channel Length Graphene Transistors," Nano Lett. 10, 3952 (2010)

[629] F.Schwierz, "Graphene Transistors," Nature Nanotechnology 5, 487 (2010)

[630] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, "Microwave oscillation of nanomagnet driven by a spin-polarized current", Nature 425, 380 (2003)

[631] W. H. Rippard, M. R. Pufall, S. Kaka, S. E. Russek, and T. J. Silva, "Direct-Current Induced Dynamics in Co90Fe10/Ni80Fe20 Point Contacts", Phys. Rev. Lett. 92, 027701(2004)

[632] J.A. Katine, and Eric E. Fullerton, "Device implications of spin-transfer torques ", J. of magnetism and magnetic materials 320, 1217(2008)

[633] P. Villard, U. Ebels, D. Houssameddine, J. Katine, D. Mauri, B. Delaet, P. Vincent, M.-C. Cyrille, B. Viala, J.-P. Michel, J. Prouvée, and F. Badets, "A GHz Spintronic-Based RF Oscillator", IEEE J. Solid-State Circ. 45, 214(2010)

[634] Y. Guan, *et al*, "Field and bias dependence of high-frequency magnetic noise in MgO-based magnetic tunnel junctions", J. Appl. Phys. 105, 07D127(2009)

[635] W. H. Rippard, M. R. Pufall, and S. E. Russek, "Comparison of frequency, linewidth, and output power in measurements of spin-transfer nanocontact oscillators", Phys. Rev. B 74, 224409(2006)

[636] A. M. Deac, A. Fukushima, H. Kubota, H. Maehara, Y. Suzuki, S. Yuasa, Y. Nagamine, K. Tsunekawa, D. Djayaprawira, and N. Watanabe, "Bias-driven high-power microwave emission from MgO-based tunnel magnetoresistance devices", Nature Phys. 4, 803(2008)

[637] D. Houssameddine, *et al*, "Spin-torque oscillator using a perpendicular polarizer and a planar free layer", Nature Mat. 6, 447(2007)

[638] V. S. Pribiag, *et al*, "Magnetic vortex oscillator driven by d.c. spin-polarized current", Nature Phys. 3, 498(2007)

[639] O. Boulle, V. Cros, J. Grollier, L. G. Pereira, C. Deranlot, F. Petroff, G. Faini, J. Barna, and A. Fert, "Shaped angular dependence of the spin-transfer torque and microwave generation without magnetic field", Nature Phys. 3, 492(2007)

[640] 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 437, 389(2005)

[641] F. B. Mancoff, N. D. Rizzo, B. N. Engel, and S. Tehrani, "Phase-locking in double-point-contact spin-transfer devices", Nature 437, 393(2005)

[642] A. Ruotolo, *et al*, "Phase-locking of magnetic vortices mediated by antivortices", Nature Nano. 4, 528(2009)

[643] D. Houssameddine, U. Ebels, B. Dieny, K. Garello, J.-P. Michel, B. Delaet, B. Viala, M.-C. Cyrille, J. A. Katine, and D. Mauri, "Temporal Coherence of MgO Based Magnetic Tunnel Junction Spin Torque Oscillators", Phys. Rev. Lett. 102, 257202 (2009)

[644] A. Slavin, and V. Tiberkevich, "Nonlinear Auto-Oscillator Theory of Microwave Generation by Spin-Polarized Current", IEEE Trans. Mag. 45, 1875(2009)

[645] J.-V. Kim, V. Tiberkevich, and A. N. Slavin, "Generation Linewidth of an Auto-Oscillator with a Nonlinear Frequency Shift: Spin-Torque Nano-Oscillator", Phys. Rev. Lett. 100, 017207(2008)

[646] C.T.C. Nguyen, "MEMS technology for timing and frequency control," IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control 54 (2), 251 (2007)

[647] K. Jensen, H.B. Peng, and A. Zetl, "Limits of Nanomechanical Resonators," Proc. of International Conference on Nanoscience and Nanotechnology ICONN 2006, pp. 68-71

[648] J.R. Vig and Y. Kim, "Noise in Microelectromechanical System Resonators," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control 46 (6), 1558 (1999)

[649] Y. Xie, S.-S. Li, Y.-W. Lin, Z, Ren, and C.T.-C. Nguyen, "UHF Microlectromechanical Extensional Wine-Glass Mode Ring Resonators," IEDM 2003, pp. 953-956

[650] D. Weinstein and S. Bhave, "Piezoresistive Sensing of a Dielectrically Actuated Silicon Bar Resonator," Proc. of Solid-State Sensors, Actuators and Microsystems Workshop 2008, pp. 368-371

[651] D. Weinstein and S. Bhave, "Internal dielectric transduction of a 4.5GHz silicon bar resonator," IEDM 2007, pp. 415-41

[652] A. Hussain, J. Hone, H.W. Postma, X.M.H. Huang, T. Drake, M. Narbic, A Scherer, and M.L. Roukes, "Nanowire–based very-high-frequency electromechanical resonator," Appl. Phys. Lett. 83 (6), 1240 (2003)

[653] X.L. Feng, R. He, P. Yang, and M.L. Roukes, "Very High Frequnecy Silicon Nanowire Electromechanical Resonators," Nano Lett. 7 (7), 1953 (2007)

[654] R. He, X.L. Feng, M.L. Roukes, and P. Yang, "Self-transducing Silicon Nanowire Electromechancial Systems at Room Temperature," Nano Lett. 8 (6), 1756 (2008)

[655] V. Sazonova, Y. Yaish, H. Ustunel, D. Roundy, T.A. Arlas, and P.L. McEuen, "A tunable carbon nanotube electromechanical oscillator," Nature 431, 284 (2004)

[656] H.B. Peng, C.W. Chang, S. Aloni, T.D. Yuzvinsky, and A. Zettl, "Ultrahigh Frequency Nanotube Resonators," Phys. Rev. Lett. 97, 087203 (2006)

[657] H.B. Peng, C.W. Chang, S. Aloni, T.D. Yuzvinsky, and A. Zettl, "Microwave Electromechanical Resonator Consisting of Clamped Carbon Nanotubes in an Abacus Arrangement," Phys. Rev. B76, 035405 (2007)

[658]] J. Scott Bunch, A.M. van der Zande, S.S. Verbridge, I.W. Frank, D.M. Tanenbaum, J.M. Parpia, H.G. Craighead, and P.L. McEuen, "Electromechanical Resonators from Graphene Sheets," Science 315, 490 (2007)

[659] N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, and A.M. Ionescu, "Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor," IEDM 2005, pp. 479 – 481

[660] N. Abele; K. Segueni; K. Boucart; F. Casset; B. Legrand; L. Buchaillot; P. Ancey, and A.M. Ionescu, "Ultra-Low Voltage MEMS Resonator Based on RSG-MOSFET," IEEE MEMS 2006, pp. 882 – 885

[661]] D. Grogg, D. Tsamados, N.D. Badila, and A.M. Ionescu, "Integration of MOSFET Transistors in MEMS Resonators for Improved Output Detection," Transducers 2007, pp. 1709 – 1712

[662] C. Durand, F. Casset, P. Renaux, N. Abele, B. Legrand, D. Renaud, E. Ollier, P. Ancey, A.M. Ionescu, and L. Buchaillot, "In-Plane Silicon-On-Nothing Nanometer-Scale Resonant Suspended Gate MOSFET for In-IC Integration Perspectives," IEEE Elect. Dev. Lett. 29 (5), 494 (2008

[663] E. Colinet, C. Durand, L. Duraffourg, P. Audebert, G. Dumas, F. Casset, E. Ollier, P. Ancey, J.-F. Carpentier, L. Buchaillot, and A.M. Ionescu, "Ultra-Sensitive Capacitive Detection Based on SGMOSFET Compatible With Front-End CMOS Process," IEEE J. Solid-State Circ. 44 (1), 247 (2009)

[664] D. Grogg, H.C. Tekin, N.D. Badila-Ciressan, M. Mazza, D. Tsamados, and A.M. Ionescu, "Laterally vibrating-body double gate MOSFET with improved signal detection," DRC 2008, pp. 155-156

[665] D. Grogg, M. Mazza, D. Tsamados, and A.M. Ionescu, "Multi-gate vibrating-body field effect transistor (VB-FETs)," IEDM 2008, pp. 663 – 666

[666] J.T.M.van Beek, K.L. Phan, G.J.A.M. Verheijden, G.E.J. Koops, C. van der Avoort, J. van Wingerden, D.E. Badaroglu, J.J.M. Bontemps, and R. Puers, "A piezo-resistive resonant MEMS amplifier," IEDM 2008, pp. 667-670

[667] L. Pierantonian and F. Coccetti, "Radio-frequency nanoelectronics: A new paradigm in electronic systems design," 2010 Asia-Pacific Microwave Conference Proceedings (APMC), pp. 1007-1014

[668] N. V. Alkeev, S. V. Averin, A. A. Dorofeev, E 1. Golant, and A. B. Pashkovskii, "New TeraHertz Mixer Based on Resonant-Tunneling Diode," Proc. Int'l Symp. on Phys. and Engr. of Microwaves, Millimeter and Submillimeter Waves (MSMW), Jun. 2007, pp. 192-194 (2007)

[669] I. Magrini and A. C. G. Manes, "A Low Local Oscillator Power K-Band Mixer Based on Tunneling Diodes," Microwave and Opt. Tech. Lett., 51(4), 1140 (2009)

[670] R. Knobel, C. S. Yung, and A. N. Cleland, "Single-electron transistor as a radio-frequency mixer," Appl. Phys. Lett., 81 (3), 532 (2002)

[671] K.-H. Oh, N. Shimizu, N. Kukutsu, Y. Kado, S. Kohjiro, K. Kikuchi, T. Yamada and A. Wakatsuki, "Heterodyne THz-wave receiver with a superconducting tunneling mixer driven by a high sweeping-speed photonics-based THz-wave local oscillator," IEICE Electronics Express, 6(10), 601 (2009)

[672] T. Palacios, A. Hsu, and H. Wang, "Applications of Graphene Devices in RF Communications," IEEE Comm. Mag., 48 (6), 122 (2010)

[673] H. Wang, A. Hsu, J. Wu, K. Jing, and T. Palacios, "Graphene-Based Ambipolar RF Mixers," IEEE Elec. Dev. Lett. 31 (9), 906 (2010)

[674] H. Lyu, H. Wu, J. Liu, Q. Lu, J. Zhang, X. Wu, J. Li, T. Ma, J. Niu, W. Ren, H. Cheng, Z. Yu, and H. Qian, "Double-balanced graphene integrated mixer with outstanding linearity," Nano Lett., vo. 15, pp. 6677-6682 (2015).

[675] C. Rutherglen and P. Burke, "Carbon Nanotube Radio," Nano Letters, 7 (11), 3296 (2007

[676] N. Rouhi, D. Jain, and P. J. Burke, "Nanoscale Devices for Large-Scale Applications," IEEE Microwave Mag. 11 (7), 72 (2010)

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

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

[679]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).

[680]S. Swanson, "System architecture implications for M/S-class SCMs," http://www.itrs.net/ITWG/ERD_files.html, ITRS SCM workshop, July 2012.

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

[682]K. H. Kim, "Memory Interfaces for M-Class SCMs," http://www.itrs.net/ITWG/ERD_files.html, ITRS SCM workshop, July 2012.

[683]. 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).

[684]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).

[685]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," in Proc. ACM/IEEE Int. Conf. High Perform. Comput.,Netw., Storage Anal., 2010, pp. 1–11.

[686]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).

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

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

[689]J. Chang, "Data-centric computing and Nanostores," http://www.itrs.net/ITWG/ERD_files.html, ITRS SCM workshop, July 2012.

[690]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.

[691]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

[692]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

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

[694]E. L. Miller, "Object-based interfaces for efficient and portable access to S-class SCMs," http://www.itrs.net/ITWG/ERD_files.html, ITRS SCM workshop, July 2012.

[695]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).

[696]S. Chen, P. B. Gibbons, and S. Nath, "Rethinking database algorithms for phase change memory," in Proc. CIDR, 2011,

pp. 21–31.

[697]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," in Proc. IEEE Int. Symp. Circuits Syst., 2007, pp. 3014–3017.

[698]B. C. Lee, E. Ipek, O. Mutlu, and D. Burger, "Architecting phase change memory as a scalable dram alternative," in Proc. 36th

Annu. Int. Symp. Comput. Archit., 2009, pp. 2–3.

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

[700700]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.

[701]T. Wang and R. Johnson, "Scalable logging through emerging non-volatile memory," Proc. VLDB Endowment, vol. 7,

pp. 865–876, 2014.

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

[703]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.

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

[705]J. Condit, E. B. Nightingale, C. Frost, E. Ipek, B. Lee, D. Burger, and D. Coetzee, "Better I/O through byte-addressable, persistent

memory," in Proc. ACM SIGOPS 22nd Symp. Operating Syst. Principles, 2009, pp. 133–146.

[706]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).

[707]. 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).

[708]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).

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

[710]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.

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

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

[713]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.

[714]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.

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

[716]Hamdioui, S.; Aziza, H.; Sirakoulis, G.C., "Memristor based memories: Technology, design and test," in Design & Technology of Integrated Systems In Nanoscale Era (DTIS), 2014 9th IEEE International Conference On , vol., no., pp.1-7, 6-8 May 2014

[717] Narayanan, P.; Burr, G.W.; Shenoy, R.S.; Stephens, S.; Virwani, K.; Padilla, A.; Kurdi, B.N.; Gopalakrishnan, K., "Exploring the Design Space for Crossbar Arrays Built With Mixed-Ionic-Electronic-Conduction (MIEC) Access Devices," in Electron Devices Society, IEEE Journal of the , vol.3, no.5, pp.423-434, Sept. 2015

[718]Jing Li; Montoye, R.K.; Ishii, M.; Chang, L., "1 Mb 0.41 µm² 2T-2R Cell Nonvolatile TCAM With Two-Bit Encoding and Clocked Self-Referenced Sensing," in Solid-State Circuits, IEEE Journal of , vol.49, no.4, pp.896-907, April 2014

[719]Y. Liu, et.al. "A 65nm ReRAM-enabled nonvolatile processor with 6x reduction in restore time and 4x higher clock frequency using adaptive data retention and self-write-termination nonvolatile logic," in ISSCC 2016, pp. 84-85.

[720]D. Nikonov and I. Young, "Benchmarking of beyond-CMOS devices and circuits," ITRS-ERD Architecture meeting, February 2015.

[721]D.S. Holmes, A.M. Kadin M.W. Johnson, "Superconducting computing in large-scale hybrid-systems," IEEE Computer, Dec. 2015, pp. 34-42.

[722]Miyamura, M.; Sakamoto, T.; Tada, M.; Banno, N.; Okamoto, K.; Iguchi, N.; Hada, H., "Low-power programmable-logic cell arrays using nonvolatile complementary atom switch," in Quality Electronic Design (ISQED), 2014 15th International Symposium on , vol., no., pp.330-334, 3-5 March 2014

[723]Kumar, T.N.; Almurib, H.A.F.; Lombardi, F., "A novel design of a memristor-based look-up table (LUT) for FPGA," in Circuits and Systems (APCCAS), 2014 IEEE Asia Pacific Conference on , vol., no., pp.703-706, 17-20 Nov. 2014

[724]D. Fan, S. Maji, K. Yogendra, M. Sharad, and K. Roy, "injection locked psing hall-induced coupled-oscillators for energy efficient associative computing," in IEEE Trans. Nano. Vol. 14, NO. 6, Nov. 2015, pp. 1083-1093.

[725]A. Sampson, "Hardware and software for approximate computing," PhD dissertation, University of Washington, 2015.

[726]N. Shanbhag, A. Singer, and P. Wong, "Statistical information processing – Computing on Nanoscale fabrics," ITRS-ERD Architecture meeting, February 2015.

[727]KaizerVizzotto, J., "Quantum Computing: State-of-Art and Challenges," in Theoretical Computer Science (WEIT), 2013 2nd Workshop-School on , vol., no., pp.9-13, 15-17 Oct. 2013

[728]A. S. Cassidy *et al*., "Real-Time Scalable Cortical Computing at 46 Giga-Synaptic OPS/Watt with ~100× Speedup in Time-to-Solution and ~100,000× Reduction in Energy-to-Solution," *High Performance Computing, Networking, Storage and Analysis, SC14: International Conference for*, New Orleans, LA, 2014, pp. 27-38.

[729]Shelby, R.M.; Burr, G.W.; Boybat, I.; di Nolfo, C., "Non-volatile memory as hardware synapse in neuromorphic computing: A first look at reliability issues," in Reliability Physics Symposium (IRPS), 2015 IEEE International , vol., no., pp.6A.1.1-6A.1.6, 19-23 April 2015

[730]D.J. Mountain, M. McLean, D. Palmer, J.D. Pucnal, C.D. Krieger, "Ohmic Weave: Memristor-based threshold gate networks," in Computer, Dec. 2015, pp. 65-71

[731]www.micronautomota.com

[732]Y-H Chen, T. Krishna, J. Emer and V. Zse, "Eyeriss: An energy efficient reconfigurable accelerator for deep convolutional neural networks," in ISSCC 2016, pp. 262-3.

[733]www.groksolutions.com

[734]R. Hecht-Nielsen, "Confabulation Theory: The mechanism of thought," 2007 (Springer)

[735]S. George; S. Kim; S. Shah; J. Hasler; M. Collins; F. Adil; R. Wunderlich; S. Nease; S. Ramakrishnan, "A Programmable and Configurable Mixed-Mode FPAA SoC," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems , vol.PP, no.99, pp.1-9

[736]N. Banno *et al*., "A novel two-varistors (a-Si/SiN/a-Si) selected complementary atom switch (2V-1CAS) for nonvolatile crossbar switch with multiple fan-outs," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2015, pp. 2.5.1-2.5.4.

[737]M. Prezioso *et al*., "Modeling and implementation of firing-rate neuromorphic-network classifiers with bilayer Pt/Al2O3/TiO2/TiOx/Pt Memristors," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2015, pp. 17.4.1-17.4.4.

[738]Luo *et al*., "Cu BEOL compatible selector with high selectivity (>107), extremely low off-current (¿¿¿pA) and high endurance (>1010)," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2015, pp. 10.4.1-10.4.4.

[739]S. Yu, P. Y. Chen, Y. Cao, L. Xia, Y. Wang and H. Wu, "Scaling-up resistive synaptic arrays for neuro-inspired architecture: Challenges and prospect," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, USA, 2015, pp. 17.3.1-17.3.4.

[740]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 Meeting (IEDM), 2014 IEEE International*, San Francisco, CA, 2014, pp. 29.5.1-29.5.4.

[741]Sung Hyun Jo, 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*, San Francisco, CA, 2014, pp. 6.7.1-6.7.4.

[742] V. V. Zhirnov, R. K. Cavin, J. A. Hutchby, G. I. Bourianoff, "Limits to Binary Logic Scaling – A Gedankin Model", Proc. IEEE, November 2003.

[743] 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.

[744] Doug Matzke, "Will Physical Scalability Sabotage Performance Gains"? IEEE Computer, Volume 30, Issue 9, September, 1997, Page: 37 – 39.

[745] 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.

[746] 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.

[747] K. Bernstein, "NRI Architecture Benchmarking Study Phase 1.5 Metrics Readout – Overview," NRI Annual Review, 2011.

[748] D.E. Nikonov and I.A. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," IEDM Tech. Dig., pp. 573-576, Dec. 2012.

[749] 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).

[750] 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.

[751] I. Sutherland et al., Logical Effort: Design Fast CMOS Circuits, 1st ed. San Mateo, CA: Morgan Kaufmann, Feb. 1999, ISBN: 10:1558605576

[752] George Bourianoff, et.al., "Boolean Logic and Alternative Information-Processing Devices," Computer, May 2008, pp. 38-46

[753] 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.