IEEE
Advancing Technology for Humanity

INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS™

2022 EDITION

BEYOND CMOS AND EMERGING MATERIALS INTEGRATION

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Table of Contents

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

List of Figures

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

List of Tables

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

ACKNOWLEDGMENTS

Sapan Agarwal
Brad Aimone
Hiro Akinaga
Otitoaleke Akinola
Mustafa Badaroglu
Gennadi Bersuker
Christian Binek
Geoffrey Burr
Leonid Butov
Kerem Camsari
Gert Cauwenberghs
An Chen
Winston Chern
Supriyo Datta
John Dallesasse
Shamik Das
Erik DeBenedictis
Peter Dowben
Tetsuo Endoh
Ben Feinberg
Thomas Ferreira de Lima
Akira Fujiwara
Elliot Fuller
Michael Frank
Paul Franzon
Michael Fuhrer
Satoshi Fujimura
Mike Garner
Chakku Goplan
Bogdan Govoreanu
Cat Graves
Kohei Hamaya
Masami Hane
Jennifer Hasler
Yoshihiro Hayashi
Toshiro Hiramoto
D. Scott Holmes
Sharon Hu
Francesca Iacopi
Gaku Ichihara
Danielle Ilmeni
Jean Anne Incorvia
Engin Ipek
Koji Izunome
Satoshi Kamiyama
Kiyoshi Kawabata
Asif Khan
Atsuhiro Kinoshita
Hajime Kobayashi
Takehito Kozasa
Suhas Kumar
Ilya Krivorotov
Xiuling Li
Xiang (Shaun) Li
Shy-Jay Lin
Tsu-Jae King Liu
Matthew Marinella
Bicky A. Marque
Koichi Matsumoto
Rivu Midya
Yoshiyuki Miyamoto
Johannes Muller
Azad Naeemi
Mitchell A. Nahmias
Hajime Nakabayashi
Emre Neftci
Kae Nemoto
Kei Noda
Mike Niemier
Dmitri Nikonov
Yutaka Ohno
Chenyun Pan
Shashi Paul
Ferdinand Peper
Paul R. Prucnal
Titash Rakshit
Shriram Ramanathan
Mingyi Rao
Arijit Raychowdhury
Graham Rowlands
Sayeef Salahuddin
Shintaro Sato
Michael Schneider
Bhavin J. Shastri
Xia Sheng
Takahiro Shinada
Urmita Sikder
Greg Snider
Satofumi Souma
John-Paul Strachan
Dimitri Strukov
Munehiro Tada
Tarek Taha
Alexander N. Tait
Shinichi Takagi
Norikatsu Takaura
Tsutomu Teduka
Yasuhide Tomioka
Wilman Tsai
Tohru Tsuruoka
Akiko Ueda
Ken Uchida
Zhongrui Wang
R. Stanley Williams
Justin Wong
Dirk Wouters
Patrick Xiao
J. Joshua Yang
Noboyuki Yoshikawa
Shusuke Yoshitake
Victor Zhirnov

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

BEYOND CMOS

1. INTRODUCTION

1.1. SCOPE OF BEYOND-CMOS FOCUS TEAM

Dimensional and functional scaling[1] of CMOS is driving information processing[2] technology into a broadening spectrum of new applications. Scaling has enabled many of these applications through increased performance and complexity. As dimensional scaling of CMOS will eventually approach fundamental limits, new information processing devices and microarchitectures for both existing and new functions are being explored. This is driving interest in new devices for information processing and memory, new technologies for heterogeneous integration of multiple functions, and new paradigms for system architecture. This chapter, therefore, provides an IRDS perspective on emerging research device technologies and serves as a bridge between conventional CMOS and the realm of nanoelectronics beyond the end of CMOS scaling.

An overarching goal of this chapter is to survey, assess, and catalog viable emerging devices and novel architectures for their long-range potential and technological maturity and to identify the scientific/technological challenges gating their acceptance by the semiconductor industry as having acceptable risk for further development. This chapter also surveys beyond-CMOS devices for more than Moore (MtM) applications.

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

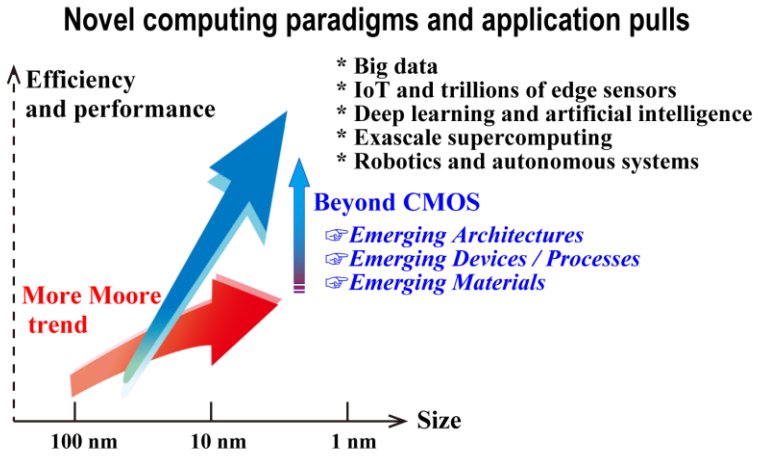


Figure BC1.1 Relationship of More Moore, Beyond CMOS, and Novel Computing Paradigms and Applications (Courtesy of Japan beyond-CMOS Group)

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

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

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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) emerging memory devices, 2) emerging logic and alternative information processing devices, 3) emerging device-architecture interaction, 4) beyond-CMOS devices for More-than-Moore applications, and 5) emerging materials integration (EMI). The former IRDS Emerging Research Materials (ERM) chapter is rolled into this chapter as a section named "emerging materials integration (EMI)". Some detail is provided for each entry regarding operation principles, advantages, technical challenges, maturity, and current and projected performance. The chapter also discusses applications and architectural focus combining emerging research devices offering specialized, unique functions as heterogeneous core processors integrated with a CMOS platform technology. This represents the nearer term focus of the chapter, with the longer-term focus remaining on discovery of an alternate information processing technology beyond digital CMOS.

1.2. DIFFICULT CHALLENGES

1.2.1. INTRODUCTION

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

1.2.2. DEVICE TECHNOLOGIES

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

One challenge is the need of a new memory technology that combines the best features of current memories in a fabrication technology compatible with CMOS process flow and that can be scaled beyond the present limits of SRAM and FLASH. This would provide a memory device fabrication technology required for both stand-alone and embedded memory applications. The ability of a chip 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 cache memory, thereby increasing the floor space that SRAM occupies on a chip. However, this trend eventually leads to a decrease of the net information throughput. Volatility of semiconductor memory requires external long-term storage media that tend to be slow to 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.

Another challenge is to sustain scaling of CMOS logic technology. One approach to continuing performance gains as CMOS scaling matures 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. Introduction of non-silicon materials into the channel and source/drain regions of an otherwise silicon MOSFET is fraught with difficult challenges. These challenges include 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, fabrication of high-κ gate dielectrics on the new channel materials, and elimination of Fermi level pinning in the channel/gate dielectric interface. Additional challenges are to sustain the required reduction in leakage currents and power dissipation in these ultimately scaled CMOS and to introduce new materials into MOSFET while simultaneously minimizing variations in critical dimensions and statistical fluctuations in the doping concentrations.

The industry is now addressing the increasing importance of a new trend of functional diversification, where added value to devices is provided by incorporating functionalities that do not necessarily scale according to "Moore's Law". In this chapter, an "Beyond-CMOS devices for More-than-Moore Applications" section covers unconventional applications of beyond-CMOS technologies. The section currently covers emerging devices for hardware security.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Table BC1.1 Beyond CMOS Difficult Challenges

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

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

1.2.3. MATERIALS TECHNOLOGIES

The most difficult challenge for Beyond CMOS is to deliver materials with controlled properties that will enable operation of emerging research devices in high density at the nanometer scale. To improve control of material properties for high-density

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

devices, research on materials synthesis must be integrated with work on new and improved metrology and modeling. These important objectives are addressed in the "Emerging Materials Integration" section.

1.3. Nano-information Processing Taxonomy

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

2. Emerging Memory Devices

The emerging research memory technologies tabulated in this section are a representative sample of published research efforts (circa 2020 – 2021) describing alternative approaches to established memory technologies.[3] The scope of this section also includes a new section on the Properties of Memory Devices for In Memory Computing, as well as updated subsections addressing the Select Device required for a crossbar memory application and updated section on Storage Class Memory.

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

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

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

The purpose of many memory systems is to store massive amounts of data, and therefore memory capacity (or memory density) is one of the most important system parameters. In a typical memory system, the memory cells are connected to form a two-dimensional array, and it is essential to consider the performance of memory cells in the context of this array architecture. A

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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. Planar transistors (e.g., FET or BJT) are typically used as select devices. In a two-dimensional layout using in-plane select FETs the cell layout area is Acell=(6-8)F^2. In order to reach the highest possible 2-D memory density of 4F^2, a vertical select transistor can be used. Table BC2.3 shows several examples of vertical transistor approaches. Another approach to obtaining a select device with a small footprint is a two-terminal nonlinear device, e.g., a diode. Table BC2.4 displays benchmark parameters required for a 2-terminal select device, and Table BC2.5 summarizes the operating parameters for several candidate 2-terminal select devices.

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

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

2.1. MEMORY TAXONOMY

Figure BC2.1 provides a simple visual method of categorizing memory technologies. At the highest level, memory technologies are separated by the ability to retain data without power. Non-volatile memory offers essential use advantages, and the degree to which non-volatility exists is measured in terms of the length of time that data can be expected to be retained. Volatile memories also have a characteristic retention time that can vary from milliseconds to (for practical purposes) the length of time that power remains on. Non-volatile memory technologies are further categorized by their maturity. Flash memory is considered the baseline non-volatile memory, because it is highly mature, well optimized, and has a significant commercial presence. Flash memory is the benchmark against which prototypical and emerging non-volatile memory technologies are measured. Prototypical memory technologies are at a point of maturity where they are commercially available (generally for niche applications), and have a large scientific, technological, and systems knowledge base available in the literature. The focus of this section is Emerging Memory Technologies. These are the least mature memory technologies in Figure BC2.1, but they have been shown to offer significant potential benefits if various scientific and technological hurdles can be overcome. This section provides an overview of these emerging technologies, their potential benefits, and the key research challenges that will allow them to become viable commercial technologies.

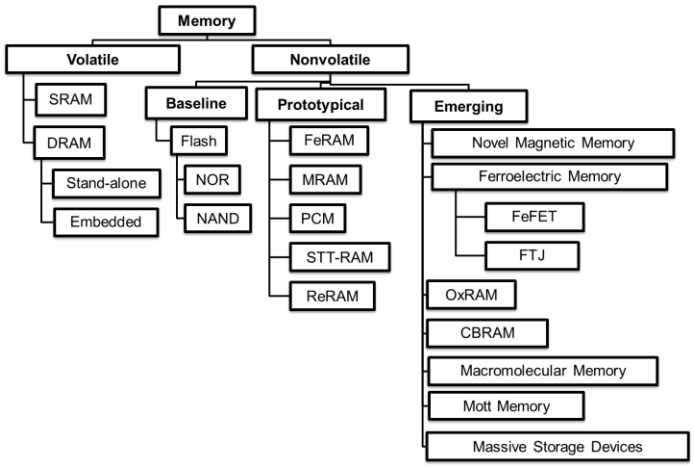


Figure BC2.1 Taxonomy of Emerging Memory Devices

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

2.2. EMERGING MEMORY DEVICES

2.2.1. NOVEL MAGNETIC MEMORIES

This section is divided into three categories of devices based on different mechanisms, i.e., spin-transfer torque, spin-orbital torque, and voltage-controlled magnetic anisotropy.

2.2.1.1. SPIN-TRANSFER TORQUE

Spin-transfer torque (STT) MRAM has recently entered the commercial production stage for both embedded and standalone Flash-like applications. Various foundries announce the readiness of embedded MRAM (TSMC, GlobalFoundries, Samsung) for production, and standalone MRAM products with various density (Mb-Gb) are also available on the market for IoT and data center applications (Avalanche Tech, Everspin). For embedded applications, STT-MRAM possesses non-volatility, high-endurance, scalability, low power, and fewer masks than the embedded Flash.[2] It also provides great area savings and lower leakage compared with SRAM. A STT-MRAM cell consists of a magnetic tunnel junction (MTJ) with two ferromagnetic layers, i.e., a free layer (FL) and a reference layer (RL) sandwiching a tunnel barrier. When a current enters the RL, the electrons' spin is polarized according to that of the RL. After tunneling through to the FL, the spin momentum of these spin-polarized electrons is transferred to the FL, thus writing the FL orientation to that of the RL. The opposite operation can be achieved by flipping the current polarity. Compared with its in-plane counterpart, perpendicular STT-MRAM has higher density, lower switching current, and is easier to control in large scale manufacturing.[3]

Since Flash-like STT-MRAM is entering mass production at all major foundries, this topic is covered in the More Moore Chapter of IRDS. For cache-level applications, research has shown 3 ns, 63 μA switching of 16-nm perpendicular STT-MRAM with endurance above 10^{12} cycles and WER below 10^{-6}.[4] Reliable 10 ns, 0.12 pJ switching of 50 μA with sub-ppm error rates is also recently demonstrated in 30 nm perpendicular STT-MTJs satisfying the last level cache (LLC) application requirements.[5] A system-level benchmarking study also indicates the use of 34 nm perpendicular STT-MRAM to replace SRAM as a LLC for high performance computing (HPC) at 5 nm node provides significant read and write energy gains at about 43% of the total macro area, achieving a nominal access latency <2.5 ns and <7.1 ns for read and write respectively.[6] (see Table BC2.1)

For SRAM-like STT-MRAM development, a major challenge is to reduce the write power consumption at sub-10 ns write speed. The large write power at sub-ns speed using STT requires a large access transistor, reducing density as well as reduced endurance due to tunnel barrier damage from higher writing current. Multiple factors contribute to this high switching current at sub-10 ns regime including limited spin torque efficiency, incubation delay, and intrinsic magnetization precession frequency on the order of GHz. A great amount of work has been devoted to increasing the switching speed, such as decreasing the FL saturation magnetization, lowering the FL damping factor,[7] and increasing the STT effect via double-RL design[8]. However, because at most 100% of the current can be polarized, there is an upper bound of the write efficiency using the STT effect. In contrast, the use of writing mechanisms of fundamentally different physics, such as spin-orbit torque (SOT) and voltage-controlled magnetic anisotropy (VCMA) effect, may help propel the next-generation of MRAM technologies for SRAM-like cache applications. Secondly, as the critical dimension of STT-MRAM scales down to less than 20 nm, the requirement of thermal stability calls for higher interfacial perpendicular magnetic anisotropy (PMA). Double-MgO barrier MTJs have shown 2x improvement in MTJ with diameters above 10 nm,[9] whereas the use of shape-anisotropy induced PMA from elongated ferromagnetic pillars may further the scaling of STT-MRAM below 10 nm diameter.[10] Lastly, STT-MRAM with higher density to replace DRAM is still an open area for research. Stacking of STT-MRAM dies with logic or memory dies using through silicon vias,[11,12] and high density 3D integration of STT-MRAM using selector-MTJ crossbar architecture[13] are two high potential approaches.

2.2.1.2. SPIN-ORBIT TORQUE

Spin-orbit torque (SOT)-driven magnetization switching recently emerges as an alternative write mechanism beyond STT for SRAM-like cache-level applications. Though at rather early stage of research, sub-ns SOT writing has been demonstrated at current density of 20-40 MA/cm^2,[14,15] compared with 3-10 ns switching of STT-MRAM at a current density of 7 MA/cm^2.[5,16](see Table BC2.1). A SOT-MRAM cell consists of a magnetic tunnel junction (MTJ) with its free layer (FL) sitting on top of a strip of material with large spin-orbit coupling (SOC), such as heavy metal[17,18]. When current flows through this long strip of SOT material, spin-polarized current emerges and diffuses into the adjacent FL. Like the STT case, the spin-polarized current exerts a damping-like spin torque on the ferromagnetic layer, thus switching the FL orientation. As the write path is separated from the read path, a much larger read voltage can increase read speed. The major advantage of SOT over STT is that unlike STT where the filtered spin-polarized current is smaller than the charge current, the SOT efficiency (spin-polarized current over charge current) can be larger than one in the SOT case.[19] From a physics perspective, multiple mechanisms have been found to contribute to this large damping-like SOT, including spin Hall effect (SHE)[17], Rashba-Edelstein effect[18], and spin-momentum locking from topological protected electronic states[19]. Most experimental work has discovered a damping-like SOT in the in-plane transverse

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

direction with respect to the current flow direction. However, there also exists a field-like SOT in many of the experimental works, which acts on the free layer like a static magnetic field with a fixed orientation.

There are three types of SOT-MRAM configurations, i.e., in-plane MTJ with easy axis oriented along the current direction (type X), in-plane MTJ with easy axis oriented orthogonal to the current direction (type Y), and perpendicular MTJ (type Z). Note that only type Y can achieve field-free switching, while both type X and Z require the breaking of symmetry for deterministic switching. Experimentally, 0.5-ns switching with a current density of 40 MA/cm^2 has been demonstrated in a 100 x 400 nm^2 type X MTJ, which will lead to 51-μA, 0.3-V, and 8-fJ write performance when scaled down to a channel width of 50 nm.[14] Another work shows 0.5-ns switching with a current density of 18 MA/cm^2 in 30 x 190 cm^2 type Y MTJ with a write error rate (WER) of 10^{-6}.[15] For perpendicular MTJ (type Z), research has shown 0.5-ns and 220-fJ write operation with a current density of 180 MA/cm^2 of a 60 nm perpendicular MTJ with endurance up to 10^{11} and WER down to 10^{-5}. Field-free switching is realized in this work by using an elongated biasing ferromagnet deposited on top of the SOT-MTJ.[20]

Lowering the write energy while maintaining sub-ns switching speed is the main challenge to push SOT-MRAM into cache-level applications. A multitude of novel SOT materials beyond traditional heavy metal are being intensively investigated for higher SOT efficiency. Several new research directions include heavy metal alloys[21], topological insulator and semimetal[22,23], antiferromagnets[24,25], and complex oxides[26,27]. The need of high SOT efficiency is especially critical for type Z as it inherently shows a larger switching current than type X and Y[28]. Second, SOT-MRAM suffers from a large cell size due to the three-terminal configuration needed to perform separate write and read functions.[29] A two-terminal perpendicular SOT-MRAM has been demonstrated by increasing the density of the current flowing in-plane in the SOT underlayer while suppressing that of the current flowing perpendicular in the MTJ[30]. Meanwhile, the scheme of multiple SOT-MTJs sharing one single SOT write line can partially alleviate the density disadvantage of SOT-MRAM.[31] Third, a tradeoff exists between writing speed and field-free switching. For type Y, the FL aligning to the transverse direction does not experience any SOT; thus, an initial perturbation of the FL away from the easy axis is required for fast switching. A large field-like SOT or Oersted field due to SOT current can provide this perturbation.[32] For type X and Z, a maximal SOT exists as the FL is aligned perpendicular to the spin current direction, thus enabling high-speed switching. There have been several approaches to break the symmetry for realizing field-free type X and Z switching, such as lateral structural and shape-induced asymmetry[28,33,34], use of interlayer exchange coupling[35], exchange bias[24], and dipolar external magnetic field[36]. Meanwhile, new studies show field-free switching of type Z MTJ using a damping-like SOT with perpendicular polarization arising from crystalline materials with broken in-plane symmetry[37] and interfaces with SOC[38,39]. Last, the scalability of all three SOT types remains an open question. Type X and Y scaling are challenging due to variations in MTJ shape, while type X and Z scaling face obstacles in implementing field-free switching at scaled nodes.

2.2.1.3. Voltage-Controlled Magnetic Anisotropy

Contrary to current-driven writing mechanisms such as STT and SOT, voltage-assisted writing using electron-mediated voltage-controlled magnetic anisotropy (VCMA) effect[40] enables lower write energy (~fJ) and smaller cell size due to reduced current, and therefore joule heating and select transistor size.(see Table BC2.1) A VCMA-MTJ is almost the same as an STT-MTJ, except that the tunnel barrier MgO thickness is increased to suppress the tunneling current and enhance the capacitive characteristics of the tunnel barrier.[41] When a voltage is applied across the VCMA-MTJ, charge accumulation or depletion takes place at the FL/barrier interface, leading to a change of electron occupancies among different Fe 3*d* orbitals. Because the interfacial perpendicular magnetic anisotropy (PMA) originates from the Fe 3*d* and O 2*p* orbitals hybridization, this change of electron occupancy results in the modulation of PMA and thus the energy barrier between the two FL stable states.[40,42,43] The VCMA effect is, therefore, a useful handle to reduce the energy barrier during the write operation, while the energy barrier is restored for retention purposes after writing by simply removing the VCMA bias.

There are two main types of VCMA-assisted magnetization switching schemes. First, removal of the entire energy barrier by the VCMA effect facilitates a precessional motion of the FL along an in-plane bias field direction (built-in or applied). By precise timing of the VCMA pulse width, the FL can switch from one state to the other in half the precession period.[44] Research has shown switching energy of 6 fJ/bit, switching speed of 0.5 ns, write voltage of 1.96 V, current density of 0.3 MA/cm^2 with a WER of 10^{-5} using perpendicular MTJs with a VCMA coefficient of 30 fJ/V-m.[45] Another recent work further shows 0.15 ns precessional switching of 120 nm perpendicular VCMA-MTJ at a write voltage of 3.06 V, a current density of 0.3 MA/cm^2 with a WER of $<10^{-6}$.[46] Second, the VCMA effect can be utilized to reduce the write energy in in-plane and perpendicular SOT-MTJs further.[31,47] Research has demonstrated VCMA-assisted (VCMA bias of 1 V) SOT writing of 30 x 80 nm^2 to 50 x 120 nm^2 in-plane MTJ using 2-ns pulse with a current density of 12 MA/cm^2 with a high endurance of 10^{13} write cycles.[48] Another work shows 5-ns 62 μA SOT current writing (VCMA bias of 1.2 V) of 30 x 80 nm^2 in-plane MTJ with WER $<10^{-8}$ and endurance over 10^{12} cycles, the VCMA coefficient in this device is about 100 fJ/V-m.[49]

The major roadblock of VCMA in either precessional switching or assisting SOT switching is the rather small VCMA coefficient of around 100 fJ/V-m, as defined by the interfacial PMA change under given electric field applied at the MgO barrier.[50] Though

Copyright © 2022 IEEE. All rights reserved.

~fJ-level write performance has already been demonstrated, further scaling of MTJs requires higher VCMA coefficient (>300 fJ/V-m) for advanced nodes cache or storage applications.[51] New materials research using Cr and Ir-based crystalline MTJs have shown a high VCMA coefficient of up to 1000 fJ/V-m.[52,53] Meanwhile, detailed chemical and structural characterizations of VCMA-MTJs recently reveal that metal-oxides at the FL/MgO interface lead to large VCMA effect.[54] Another challenge facing VCMA is the longer read time, because the thicker MTJ tunnel barrier leads to a much larger MTJ resistance. One way to resolve this is using a large read voltage (V_{DD}) which has reverse polarity compared with the write voltage to increase read speed and reduce read disturbance.[55] In terms of the precessional switching scheme, another significant challenge is the non-deterministic nature of the writing process, which results in large WER and narrow write pulse window. The use of pulse shape engineering and reverse biasing can partially help[56,57], whereas combining VCMA with deterministic writing mechanisms such as type Y SOT[47] and STT[58] may solve this challenge.

2.2.2. OXIDE-BASED RESISTIVE MEMORY (OXRAM)

The redox-based nanoionic memory operation is based on a *change in resistance* of a MIM structure caused by ion (cation or anion) migration combined with redox processes involving the electrode material or the insulator material, or both.[59,60,61] Three classes of electrically induced phenomena have been identified that involve chemical effects, i.e., effects that relate to redox processes in the MIM cell. In these three ReRAM classes, there is a competition between thermal and electrochemical driving forces involved in the switching mechanism. Two major types of ReRAM exist: i) those based on metal oxide (OxRAM), which involve oxygen ions/vacancies motion, and ii) conducting bridge-based RAM (CBRAM), which involves metal cation motion. This section covers the three categories of OxRAM, and conducting bridge-based RAM (CBRAM) is covered in the following section. Beyond CMOS has sub-categorized oxide ReRAM (OxRAM) based on the electrical switching type (bipolar versus unipolar) and whether a conductive filament is formed in the device. Most of the literature fits into the three categories: bipolar filamentary, unipolar filamentary, and nonfilamentary.[62]

In most cases, the conduction is of a filamentary nature, and hence a one-time formation process is required before the bipolar switching can be started. If this process can be controlled, memories based on this 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 below few nanoseconds, down to sub-nanoseconds range.[63,64] Many of the finer details of the ReRAM switching mechanisms are still under investigation. 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,[65] retention,[66] and endurance[67] are encouraging.

2.2.2.1. BIPOLAR-FILAMENTARY OXRAM

Bipolar filamentary OxRAM is the most common form of oxide-based ReRAM. At any given defect density, the number of current paths through the dielectric, in the virgin or fresh state, is proportional to the device area, and consequently the total current is area dependent. In addition, the current magnitude tends to fluctuate from device to device due to randomness of the initial distribution of vacancies/ions. However, cell area dependency is eliminated when the current is dominated by a single conductive path, called conductive filament (CF). The CF provides an ultimate scaling advantage since it is only limited to the active filament size, which potentially may be as small as a few nm.

A one-time forming process is required for most types of OxRAM devices to create a conduction filament across the dielectric layer linking the electrodes. A stable preferential conduction path is known to form through oxide films subjected to electrical stress: under the applied voltage, a current abruptly increases at some point in time indicating the occurrence of a dielectric breakdown (BD) resulting in the formation of a CF. During the forming process, electrons injected from the cathode electrode may lead to their trapping at defect sites in the dielectric material inducing chemical bonds breakage and the generation of anion vacancies (Oxygen or Nitrogen).[68,69]

Post-forming switching events between high and low conductive states, which are operated at significantly smaller voltages, are believed to modify the filament conductivity by rupturing/recovering a section of the filament (primarily in the vicinity of the metal electrode) or changing the filament cross-section. The specific mechanisms in filament-type switching depend on the materials (dielectric and metal electrodes) employed in the fabrication of the memory cell and may include more than one type of a conduction mode. The operation of these devices involves redox reactions of the dielectrics sandwiched between two electrodes.[70,71,72] The dielectrics are mostly comprised of one or a few layers of insulating materials[73] (e.g., oxide AlO_x, HfO_x, TaO_x, TiO_x, WO_x, ZrO_x, oxynitrides AlO_xN_y, or nitrides including AlN_x and CuN_x). TaOx and HfOx are the leading candidates among the aforementioned dielectrics, due to their superior performance (e.g., endurance) and CMOS compatibility.

Since the demonstration of a single crosspoint HfOx device with a 10 nm dimension in 2011[74], scaling to a smaller size has been achieved by employing a sidewall electrode in a 1×3 nm^2 cross-sectional HfO_x-based OxRAM device with reasonable performance in terms of both endurance and retention.[75] Up to 10^{12} cycles has been demonstrated with $Zr:SiO_x$ sandwiched by graphene oxide layers.[76] Some of the filament-based metal-oxide RRAMs implemented with metal electrodes and a variety of

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

fab-friendly transition-metal-oxides (i.e., HfO_2, ZrO_2, TiO_2, etc.) and nitride devices demonstrated sub-nanosecond,[77,78] switching with high (up to 10^{12} cycles) endurance[79] and retention of more than 10 years. Extrapolated retention at 85°C by stressing TaO_x in the temperature range from 300°C to 360°C is estimated to be years with an activation energy of 1.6 eV. [80] Reliable switching operations have been demonstrated at 340°C with devices based on 2D layered heterostructures (e.g., graphene/$MoS_{2-x}O_x$/graphene).[81]

Unconventional electrodes such as graphene have been paired with HfO_x dielectrics to yield a low power consumption, a write/erase energy of 230 fJ per bit for a single programming transition.[82] Pt/BMO((Bi, Mn)O_x)/Pt structured OxRAM device was used to demonstrate an even lower write/erase energy per transition, of the order of 3.8 pJ/bit for read and 20 pJ/bit for write operation.[83]

Large scale integration of OxRAM switching based on 1T1R schemes has been carried out by Toshiba, Panasonic and IMEC. In 2013, Toshiba announced the 32 Gb RRAM chip integrated with 24 nm CMOS.[84] In 2014, Panasonic and IMEC demonstrated the encapsulated cell structure with an Ir/Ta_2O_5/TaO_x/TaN stack on a 2-Mbit chip at the 40 nm node. In addition, passive integration of 1S1R scheme has been reported by Crossbar on a 4-Mbit chip, but the material stack of the OxRAM switch has not been revealed.[85] Ultra-fast (down to 100 ps), compliance-free, low power (< pJ) switching was demonstrated with 1R devices using TiN/HfO2/TiN stack.[86]

A number of technical challenges hampering the commercialization of OxRAM still remain despite the significant advancements made in the field. One of the main challenges is the fact that the switching currents for devices based on the currently most mature materials (e.g., HfO_x and TaO_x) are still too high (above tens of μA) for large arrays. Apart from that, the filament formation and rupture processes are stochastic in nature, which leads to variation in switching parameters like the voltage and resistance distribution of the switching. This is especially detrimental to certain applications such as multilevel cell memory.

2.2.2.2. Bipolar Non-Filamentary OxRAM

The *Bipolar Non-Filamentary* OxRAM is a non-volatile bipolar resistive switching device composed of one or more oxide layers. One layer is a conductive metal oxide (CMO), which is usually a perovskite such as PrCaMnO3 or Nb:$SrTiO_3$.[87] In contrast to *Unipolar* and *Bipolar Filamentary* OxRAM devices – typically based on binary oxides such as TiO_x, NiO_x, HfO_x, TaO_x or combinations thereof—the resistance change effect of the *Bipolar Non-Filamentary* OxRAM is *uniform*. Depending on the materials choice and structure the current is conducted across the entire electrode area, or at least across the majority of this area. A forming step to create a conductive filament is *not* needed. Non-volatile memory functionality is achieved by the field-driven redistribution of oxygen vacancies close to the contact resulting in a change of the electronic transport properties of the interface (e.g., by modifying the Schottky barrier height). Oxygen can be exchanged between layers due to the exponential increase in ion mobility at high fields. Low current densities, uniform conduction, and bipolar switching imply that substantial self-heating is not involved. Typical R_{OFF} to R_{ON} ratios are on the order of 10.

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

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

Currently the technology is in the research and development stage. Depending on material system and structure cycling endurance over 10,000 cycles and up to a billion cycles as well as data retention from days to months at 70°C has been achieved on single devices.[88,89,90] Within the Bipolar Non-Filamentary OxRAM device family the Tunnel OxRAM is probably the furthest developed technology. Single device functionality is demonstrated down to 30 nm. Set, reset, and read currents scale with area and tunnel oxide thickness facilitating sub μA switching currents with read currents in the order of a few nA to a few 100 nA. BEOL integration schemes and CMOS/OxRAM functionality are verified for 200 nm devices on 200 mm CMOS wafers. True cross-point array (1R) functionality utilizing the self-selecting non-linear device IV characteristics and transistor-less array operation is demonstrated on fully decoded 4kb true cross-point arrays (1R) build on top of CMOS base wafers. SLC and MLC operations are demonstrated within 4kb arrays.

Copyright © 2022 IEEE. All rights reserved.

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

The most important issue is the improvement of retention and the "voltage-time dilemma." This dilemma hypothesizes physical reasons as to why it is difficult in a particular device and material system to simultaneously obtain a long retention, with short low read voltages, and fast switching at moderate write voltages.[91] Even though the exact mechanism is still under investigation there is a common agreement that oxygen vacancies are moved by the external electric field resulting in different resistance states of the memory cell. Vacancy drift at room temperature is possible due to a field dependent mobility, which increases exponentially with field at fields of 1 MV/cm and larger. However, current models based on a field-dependent mobility underestimate the experimentally observed ratio between set/reset times and data retention indicating that the mechanism is only partly understood. More theoretical work is needed to understand the kinetics of programming and retention mechanisms. Once understood, materials need to be chosen to maximize the ratio between set/reset and retention times. The goal is to set/reset devices at low temperatures and meet retention requirement of 10 years at 70°C, 85°C, and 125°C, depending on the application. A multi-layer ReRAM structure (HfO_2/A_2O_3) was shown to improve retention by suppressing tail bit failure due to decreased oxygen ion diffusivity.[92]

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

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

2.2.2.3. Unipolar Filamentary OxRAM

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

Unipolar OxRAM is another resistive switching device, also referred to in the literature as thermochemical memory (TCM)[59] due to its primary switching mechanism. The device structure consists of a top electrode metal/insulator/bottom electrode metal (MIM) structure. Typical insulator materials are metal-oxides such as NiO_x, HfO_x, etc., and common metal electrodes include TiN, Pt, Ni, and W. In general, the device can be asymmetric (i.e., top electrode material differs from bottom electrode material), but unlike other types of ReRAM, asymmetry is not required.

The first reported resistive switching in these MIM structures after 2000 was unipolar in nature (see reference[93] for the first integrated device work that put metal oxide ReRAM in the spotlight). Unipolar is defined as switching where the same polarity of voltage needs to be applied for changing the resistance from high to low (SET) or from low to high (RESET). Note that in the general case, polarity is still important (e.g., repeatable SET/RESET switching only occurs for one polarity of voltage with respect to one of the electrodes[94]). Only in symmetric structures (e.g., $Pt/HfO_2/Pt$), nonpolar behavior can be obtained, where SET and RESET are occurring irrespective of voltage polarity.[95]

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

The unipolar character of the switching indicates that drift (of charged defects) in an electric field plays a less important role (than it does in bipolar switching resistive memory), but that thermal effects probably dominate.[96,97] On the other hand, polarity effects indicate anodic oxidation (e.g., at Ni or Pt electrodes) is responsible for RESET.[94] 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.[98,99,100,101] An interesting work reporting on the Scaling Effect on Unipolar and Bipolar Resistive Switching of Metal Oxides was published.[102]

Copyright © 2022 IEEE. All rights reserved.

A unipolar switching device is seen as advantageous for making scaled memory arrays, as it only requires a selector device as simple as a diode that can be stacked vertically with the memory device in a dense crossbar array.[93] In addition, the use of a single program voltage polarity greatly simplifies the circuitry.

On the other hand, as has been exemplified in mixed mode (unipolar/bipolar) operation of memory cells, there are important trade-offs between the unipolar and bipolar switching modes. On the positive side, unipolar switching mode typically shows a higher ON/OFF resistance ratio. On the negative side, unipolar switching is typically obtained at higher switching power (higher currents) than the bipolar mode, and also endurance is much more limited. As a result, major research and development work on resistive memories has shifted towards bipolar switching mechanisms. Yet, some interesting recent development work has been reported.[103,104,105,106,107,108] One paper[103] shows an endurance of over 10^6 cycles with a resistive window of over 5 orders of magnitude (and a reset current ~1mA). Others[104,105,106] 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 28 nm CMOS process was reported.[106] The key attributes were a small cell size (0.03 μm^2), switching voltage of less than 3V, RESET current of less than 60 µA, endurance > 10^6 cycles, and short SET and RESET times of 500 ns and 100 µs, respectively. One paper[107] shows 4Mb array data using this same Contact-RRAM technology, fabricated using a 65 nm CMOS process. To accommodate the low logic VDD process, an on-chip charge pump was applied. Set and reset voltages are less than 2V. Another paper[108] reports on a novel approach using thermal assisted switching to lower the switching current.

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

Major challenges to be resolved are the high switching current that seems inherent to the unipolar operation mode. Reset currents less than 100 µA are achieved but need further reduction to less than 10 µA.[104,105,106] Recently, a possible solution incorporating thermally assisted switching has been presented.[108]

2.2.3. CONDUCTING BRIDGE MEMORY

Conductive Bridge RAM (CBRAM), also referred to as Programmable Metallization Cell (PMC), and electrochemical metallization cells, solid-electrolyte switch and atom(ic) switch, is a device that utilizes electrochemical control of nano-scale quantities of metal in thin dielectric films or solid electrolytes to perform the resistive switching operation.[109, 110] The basic CBRAM cell is a metal–ion conductor–metal (MIM) system consisting of an electrode made of an electrochemically active material such as Ag, Cu or Ni, an electrochemically inert electrode such as W, Ta, Ru, or Pt, and a thin film of solid electrolyte sandwiched between both electrodes.[111] 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.[112]

CBRAM is a strong emerging memory candidate primarily due to scalability (~10 nm),[113] ultra-low energy operations due to fast read, write and erase times, and low voltage requirements.[114] Maturity of the CBRAM technology development can be assessed by the fact that many companies are either shipping products based on CBRAM or are in advanced stages of commercialization. Recent publications show CBRAM technology application in various markets including SSDs,[115] embedded NVM,[116] and serial interface non-volatile memory replacement.[117] In 2014, a 16 GB CBRAM array based on a CuTe CBRAM cell was demonstrated.[118,119] Such efforts are critical to identify core technology challenges[120] and fundamental materials and mechanisms.[121] An improved thermal stability of the conducting bridge in CBRAM gives a new opportunity for automotive applications.[122] Novel applications such as reconfigurable switch[123] and synaptic elements in Neuromorphic systems[124] based on CBRAM are also gaining prominence and are expected to expand the application base for this technology. An atom-switch-based field-programmable gate array (FPGA) is also released using Cu conduction bridge in a new polymer solid electrolyte.[125] Low-power and rad-hard operation of the FPGA using atom switch is demonstrated in obit.[126]

As with other filamentary ReRAM technologies, CBRAM is challenged by bit level variability,[120] the random nature of reliability failure such as retention or endurance, and random telegraph noise potentially contributing to read disturbs.[127] 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.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Engineering hurdles include the availability and integration of new materials used in CBRAM at advanced process nodes especially when there could be issues with compatibility of thermal budgets and process tooling. The availability of integrated array level information suggests that some of these challenges are being resolved in the recent years.[117,123] Active participation from semiconductor equipment vendors and material suppliers would assist in overcoming manufacturing hurdles rapidly.

2.2.4. MACROMOLECULAR (POLYMER) MEMORY

Macromolecular memory is a category of memory that focuses on structures incorporating a layer of polymer - the polymer perhaps containing nano-particles, small molecules and nanoparticles - that is sandwiched between two metal electrodes. This structure allows two different stable electrical states controlled through an external electrical voltage. These two stable electrical states, which are often called ON and OFF states (or 0 and 1), exhibit resistive, ferroelectric or capacitive natures according to the physical properties of the sandwich. The first fully-organic memory devices, based on nano-composite (a blend of poly-vinyl-phenol (PVP) and Bucky-ball (C60)) was presented in Materials Research Society in 2005.[128] Around the same time, memory devices using gold nano-particles and 8-hydroxyquinoline, dispersed in a polystyrene matrix, were also demonstrated.[129] Since then, the interest to use an admixture of nano-particles, small molecules and polymers in the manufacture of electronic memory devices, is on the rise. Non-volatile memory effects with a non-destructive read have been reported for a surprisingly large variety of polymeric/organic materials and blends of polymers with nanoparticles and molecules. Unlike the other four categories, this category is based on the material used in the switching layer(s) of the cell, but the mechanism is not specified. Both bipolar and unipolar (all pulses of the same polarity) switching have been demonstrated. Macromolecular ReRAM may have a mechanism placing it in one of the four main ReRAM categories listed above. However, the other mechanisms behind the electrical bistability, such as capacitive and ferroelectric, have also been reported.

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

Stability of the memory states at high temperatures (85°C, 2×10^4 s) has been demonstrated.[130,131] Programming at very low power (70 nW) has been realized.[131] Assuming a 15 ns switching time for the same system, one might achieve a write energy of 6×10^{-15} J/bit. Furthermore, low programming voltages have been realized: +1.4 and -1.3 V for the two states with good retention time (>10^4 s).[132] Downscaling of polymer resistive memory cell to the 100 nm length scale has been reported.[133] At this length scale, integration of memory cells into an 8×8 array could be shown. Polymer memory cells on a flexible substrate have been shown.[134] For amorphous carbon, downscaling to nanometer sized cells has been published (1×10^3 nm^2).[135] 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.[136] With regards to the mechanism of operation, extensive work on the class of polyimide polymers has shown clear correlations between electronic structure of the polymer and memory effects, although a comprehensive picture for the operation has not yet emerged. A number of studies have indicated an active role of the interface between macromolecular material and (native) oxide layers in the operation of the memory involving charge trapping.[137,138] The recent and past studies show resistive,[139] capacitive (charge storage, based on electric dipole formation)[140,141] and ferroelectric behavior[142] of such devices. Thus, there is a need to open up a further discussion on the right pathway to realize such memory.

In macromolecular memory, a large variety of operation mechanisms can be operative. A key research question concerns distinguishing different mechanisms and evaluating the potential and possibilities of each mechanism. A second subsequent step would be to identify model systems for each mechanism. Having such a model system then provides a possibility to benchmark the operation of the macromolecular materials. These research steps would be crucial for establishing and securing the collaboration of the chemical industry; for design, synthesis and development of the next generation macromolecular materials for memory applications, clear guidelines on the required structural and electronic properties of the macromolecular material are needed. For instance, memory effect originating for metallization and formation of metallic filaments requires macromolecular materials that support transport of ions and have appropriate internal free volume for ion conduction. Here the field could benefit from interaction with the field of polymer batteries. Ferroelectric polymers have been shown to give rise to resistive memory[143] 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.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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

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

2.2.5. FERROELECTRIC MEMORY

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

2.2.5.1. FERROELECTRIC FET

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

In order to assess the material and device requirements for a reliable and scalable FeFET technology the following two intrinsic relations in a ferroelectric gate stack need to be considered. First it is important to note that the extent of the aforementioned V_T-shift (memory window) in FeFET devices is primarily determined by the V_C of the implemented ferroelectric rather than by its remnant polarization P_r.[144] 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,[145,146] but rather due to their low E_c (SBT: 10-100 kV/cm, PZT: ~50 kV/cm, summarized in[147]) 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.[148] This CMOS-compatible material innovation enabled the demonstration of a FeFET technology scaled to the 28 nm node utilizing a conventional HKMG technology and is already used in high volume production.[149] 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[150]. 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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

capacitance as low as possible. In the perovskite-based FeFET this is achieved by utilizing high-k buffer layers and is additionally fostered by the unavoidably large physical thickness of the perovskite ferroelectrics.[4, 151]. In the case of the aggressively scaled FE-HfO_2-based FeFET, the small thickness of the ferroelectric is compensated by the comparably low permittivity of HfO_2, the possibility to use ultra-thin interfacial layers, and by the depolarization resilience of the high E_c.[147,152] This leads to the situation that despite the markedly different stack dimensions and materials used, the electrically obtained characteristics are quite similar. Fast switching speed (≤100 ns), switching voltages in the range of 4-6 V, and 10-year data retention and endurance in the range of 10^{12} switching cycles have been demonstrated for FE-HfO_2-[148,149,153,154] as well as for perovskite-based FeFETs.[144,155,156] 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.[157] 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.[153] 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)[144], a reduced trap density at the interfaces,[157] an optimized capacitive voltage divider by area scaling in the MFMISFET approach[158] or the realization of a MFSFET device by implementing recent breakthroughs in the epitaxial growth of FE-HfO_2.[159] Despite promising results obtained for perovskite-based FeFET devices implemented into 64Kb NAND-Arrays at a feature size of 5 µm[156], little is known about the variability and array characteristics of FeFET devices scaled to technology nodes approaching the grain or domain size of the implemented ferroelectrics. Initial investigations on phase and grain distribution in doped HfO_2 based ferroelectric thin films and the effects of such granularity on device level characteristics of scaled FeFETs (such as on the statistical nature of switching) have recently been reported in Refs. [160,161,162]. Recently, 64 kb and 32 Mb FeFET arrays were demonstrated in the 28 nm[163] and the 22 nm FD-SOI CMOS platform,[164] respectively—in each case, a clear low and high V_T separation at the array level was demonstrated. Nevertheless, in order to fully judge the variability of ferroelectric phase stability at the nanoscale and to guide material optimization and fundamental understanding of the phenomenon, larger array statistics in the kB to Mb range and high-resolution PFM data will be required. Besides, recent demonstration of non-volatile memory operation based on antiferroelectricity—a phenomenon closely related to ferroelectricity—in work-function engineered ZrO_2 thin film capacitors may allude to new way of addressing and potentially solving some of these challenges in FeFETs.[165]

2.2.5.2. FERROELECTRIC TUNNEL JUNCTION

The ferroelectric tunnel junction, a ferroelectric ultra-thin film commonly sandwiched by asymmetric electrodes and/or interfaces, exhibits ferroelectric polarization induced resistive switching by a non-volatile modulation of barrier height. With the tunneling current depending exponentially on the barrier height, the ferroelectric dipole orientation either codes for a high or a low resistance state in the FTJ, which can be read out non-destructively. The resulting tunneling electroresistance (TER) effect of FTJs, the ratio between HRS and LRS, is usually in the range of 10 to 100 ([166] 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.[167] A similarly high TER was demonstrated by Wen and co-workers[168] 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.[169] 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.[170] In the context of retention, it should be noted that despite improved TER, the newly proposed MFS-FTJ structure will give rise to a depolarization field, which will most likely degrade memory retention in a similar manner as described for the FeFET in Section 2.2.5.1. The highly energy efficient electric field switching, common to all ferroelectric memories, enables fast

[4]

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

(10 ns[171]) and low voltage (1.4 V[170]) 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[172]). Due to the availability of non-destructive read-out and the further reduced ferroelectric thickness in FTJ devices as compared to conventional FRAM, improved voltage scaling and total energy consumption may be expected from this technology.

Similar to most other two-terminal resistor-based memories with insufficient self-rectification, the elimination of sneak currents in large crossbar arrays is most efficiently suppressed utilizing 1T-1R or 1D-1R cell architecture. In terms of scaling, this two-element memory cell, as well as the scalability of the selector device itself, has to be considered.[173] Simply based on the lateral dimensions of the FTJ element (assuming unlimited scalability of the selector), scaling below 50 nm^2,[172]based on PFM data,[174] most likely appears possible. However, with further scaling a simultaneous enhancement of the LRS current density is required to maintain readability in massively parallel memory architectures. A recent breakthrough of $1.4x10^5$ A/cm^2 current density at 300 nm feature size has been achieved by Bruno et al.[175] utilizing low resistivity nickelate electrodes. Based on these results maintaining 10 µA read current for feature sizes <100 nm appears possible. New FTJ concepts are also emerging; for example, engineered domain walls within the ferroelectric layer in an FTJ structure can lead to exotic quantum phenomenon such as resonant electron tunneling and quantum oscillations in the electrical conductance albeit at low temperatures.[176]

FTJ based memories are currently at a very early development stage, and most of the research activity is focused on perovskite-based ferroelectrics. Further investigations reaching beyond single device characterization will be needed to fully judge the scalability of FTJ as well as its MLC capability suggested in Ref. [177]. 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 ([178] and references therein). For future development, the ferroelectric film in an ideal FTJ should be as thin as possible to allow scalability (while maintaining sufficient read current) and much less defective than that in ReRAM (e.g., with fewer oxygen vacancies), so that the mechanism of ferroelectric switching can dominate electrical behavior with little influence from mechanisms related to conducting filaments. The manufacturability of the rather complex electrode-perovskite ferroelectric-system of the FTJ concept will largely rely on the availability of high throughput and CMOS-compatible epitaxial growth techniques for large substrates or alternatively on the unrestricted feasibility demonstration of a polycrystalline FTJ. In this context it is worth noting that the CMOS-compatibility and advanced thickness scalability of ferroelectrics based on HfO_2 and its doped variant[154] as well as recent breakthroughs in its epitaxial growth[159] might yield great potential for the manufacturability of competitive FTJs. Experimental demonstrations of FTJs based on doped variant of HfO_2 were recently reported in Refs. [179,180,181].

2.2.6. MASSIVE STORAGE DEVICES

Device scaling has become a matter of strategic importance for modern and future information storage technologies, which motivates an exploration of unconventional materials with competitive performance attributes. By 2040 the conservative estimate the worldwide amount of stored data is 10^24 bits, and the high estimate is ~10^29 bits[182] (these estimates are based on research by Hilbert and Lopez[183]). In nature, much of the data about the structure and operation of a living cell is stored in the molecule of deoxyribonucleic acid (DNA) and using nucleic acids molecules, such as DNA, for memory storage has been proposed. DNA has an information storage density that is several orders of magnitude higher than any other known storage technology: 1 kg of DNA stores 10^24 bits, for which $>10^9$ kg of silicon Flash memory would be needed.[182] Thus, a few tens of kilograms of DNA could meet all of the world's storage needs for centuries to come.

A number of recent studies have shown that DNA can support scalable, random-access and error-free information storage.[184,185,186] A state-of-the-art operating system developed at the University of Washington with an industry partner is a DNA-based archival storage framework that supports random access from a DNA key-value store.[187] The DNA-stored files are compatible with mainstream digital format, and large-scale DNA storage up to 200 MB has been demonstrated.[188] There are still many unknowns regarding both DNA operations in cell and with regard to the potential of DNA technology for massive storage applications. DNA volumetric memory density far exceeds (10^3–10^7×) projected ultimate electronic memory densities. Also, in the living cell, the memory read/write operations occur at high speed (<100 µs/bit) and require very low energy of ~10^{-17} J/bit or 10^{-11} W/GB.[189] DNA can store information stably at room temperature for hundreds of years with zero power requirements, thus making it an excellent candidate for large-scale archival storage.[189] Also, DNA is an extremely abundant and totally recyclable material. Recently, a method for efficient encoding of information—including a full computer operating system—into DNA was presented, which approaches the theoretical maximum for information stored per nucleotide.[190] One of the goals for research efforts is to demonstrate miniaturized, on-chip integrated DNA storage. New methods for DNA synthesis and sequencing are key components for these developments.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Two major categories of technical challenges remain:

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

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

2.2.7. MOTT MEMORY

Mott memory is a metal/insulator/metal capacitor structure consisting of a correlated electron insulator (or Mott insulator). Correlated electron insulators often show the electronic phase transition accompanied by a drastic change in their resistivity under external stimuli such as temperature, magnetic field, electric field, and light. Mott memory exploits this electronic phase transition (called Mott metal-to-insulator transition or Mott transition[191]) 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.[192] The theory also predicted that the resistive switching due to the interfacial Mott transition has a non-volatile-memory functionality, because the Mott transition is a first-order phase transition due to its nature.[191] In addition, Mott memory based on the Mott transition involving a large number of carriers (more than 10^{22} cm^{-3}) has in principle an advantage in device scaling, because there are a sufficient number of carries for the Mott transition even in a nanoscale device. In an ideal Mott transition, the electrons localized due to the strong electron-electron correlation come to be itinerant, via the stimuli, such as application of an electric field, and so forth. It needs no dopants, and the mechanism withstands the miniaturization of the (silicon) devices.

The Mott transition induced by an electric field or carrier injection has been experimentally demonstrated in a correlated electron material of $Pr_{1-x}Ca_xMnO_3$.[193] After this demonstration, two-terminal devices such as switches and memories have been intensively studied using such correlated electron oxides as $Pr_{1-x}Ca_xMnO_3$,[194,195] VO_2,[196,197,198] $SmNiO_3$,[199] NiO,[200,201] Ca_2RuO_4,[202] and NbO_2,[203,204] and using Mott-insulator chalcogenides of AM_4X_8 (A=Ga, Ge; M=V, Nb, Ta; X=S, Se).[205,206,207,208] In addition to these inorganic materials, reversible and non-volatile resistive switching based on the electronic phase change between charge-crystalline state and quenched charge glass has recently been demonstrated in the organic correlated materials of θ-(BEDT-TTF)$_2X$ (where X denotes an anion).[209]

$SmNiO_3$ exhibits a colossal (8 orders in magnitude) resistance jump by hydrogenation. The $SmNiO_3$ channel with the solid state proton gate has demonstrated the electric base gated large ON/OFF switching.[199] The trigger for switching is based on the proton intercalation by electric field, and the DFT calculation explains the large gap opening by additional electron doping via protonation and is the origin for colossal resistance jump phenomena.[210] These results indicate that the device using the metal – insulator (Mott) transition driven by the strong electron-electron correlation is powerful as well as appropriate for the switching devices.

Scalability has been demonstrated down 110 × 110 nm^2 in Mott memristors consisting of NbO_2 that shows the temperature-driven Mott transition from a low-temperature insulator phase to a high-temperature metal phase. The switching speed, energies, and endurance of the NbO_2-Mott memristors have been evaluated to be less than 2.3 ns, of the order of 100 fJ, and $>10^9$, respectively.[203,204] The programing and read voltages reported so far are <2 V and <0.2 V, respectively.[201] The non-volatile resistive switching of AM_4X_8 single crystals was induced by the electric field of less than 10 kV/cm.[205,206,207,208] This suggests that if the device consisting of a 10-nm-thick AM_4X_8 film is fabricated, the switching voltage will be less than 0.01 V.

Although non-volatile switching has been reported in the devices based on AM_4X_8 and θ-(BEDT-TTF)$_2X$, their retention characteristics are not elucidated in detail.[205,206,207,208,209] In addition, the NbO_2-Mott memristors and VO_2-based devices are volatile switch.[196,197,198,203,204] The retention is thus a major concern of Mott memory. In principle, the Mott transition can be driven even by a small amount of carrier doping to the integer-filling or half-filling valence states of the transition element.[191] However, because of disorders, defects, and spatial variation of chemical composition, a rather large amount of carriers of more than 10^{22} cm^{-3} are required to drive the Mott transition in actual correlated electron materials, resulting in a relatively large switching voltage required in the Mott memory. Therefore, one of the key challenges is the control of crystallinity and chemical-composition in the thin films of correlated electron materials, including the integration of the correlated electron materials onto Si platform. There are some theoretical mechanisms proposed for Mott memories such as the interfacial Mott transition[192] and

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

the formation of conductive filament generated by local Mott transition.[205,207,208] However, a thorough understanding of the mechanism has not been achieved yet. Therefore, the elucidation of detailed mechanism is also a major research challenge.

2.3. CANDIDATES DEVICES FOR ANALOG IN-MEMORY COMPUTING

Analog In-Memory Computing is an emerging research architecture that promises to enable unprecedented performance per watt capabilities for certain workloads, such as deep neural networks, by performing matrix operations directly on weights stored in memory. This architecture, which is discussed in detail in Section 4, creates different requirements for memory devices than when used for traditional information storage applications. In the following we briefly describe the memory device candidates and their key properties as they relate to analog in memory computing.

2.3.1. RERAM

ReRAM memory is very dense and can be integrated in the back end of the line (BEOL) since the process temperature of ReRAM is usually lower than 400°C, thus avoiding the use of transistor area for the memory. Several ReRAM crossbars have been demonstrated for inference. [211,212,213] A conductance of CBRAM can be tuned as an effect analogous to the long-term potentiation of biological synapses, which enables analog memory and neural computing systems.[214] A key challenge is maintaining good analog properties and high resistance at the same time. Nanoscale oxide-based cross-bar memristors with analog properties at high resistance were demonstrated owing to a natural thermal-confinement-effect when reducing the cross-point area.[215,216] ReRAM for training remains a challenge due to the non-ideal electrical characteristics of synaptic devices.

2.3.2. PHASE CHANGE MEMORY

Phase-change memory (PCM) offers a wide range of analog memory states due to the large contrast between the amorphous and crystalline phases.[217] For memory applications, PCM devices can be switched between a high-resistance RESET state, formed by melting and quenching an amorphous plug that blocks a narrow constriction within the device; and a low-resistance SET state created by a crystallizing voltage pulse, which frequently ramps down in amplitude over a long duration to produce an extremely low-resistance state.

In contrast, for VMM applications where "device history" is a desirable feature, PCM devices programmed into the RESET state can be slowly brought to a much lower-resistance SET state using many repeated partial-crystallization pulses.[218] Careful choice of pulse condition can stretch this procedure out to many hundreds of pulses.[219] Some recent work has shown some evidence for gradual increases in resistance with multiple successive pulses,[220] although the operating regime must be carefully prepared, and the underlying mechanisms are not fully understood.

Many early VMM results using PCM focused on in-situ training.[725] Challenges for training include the one-sided nature of PCM programming, the nonlinear evolution of conductance with partial crystallization pulses, and the stochastic nature of PCM programming.

More recent VMM results using PCM have turned to inference of previously trained weights. Key challenges for inference include accurate programming of synaptic state despite the inherent stochasticity observed in PCM device programming,[221,222] reducing resistance drift due to long-term relaxation of the amorphous phase, and ensuring long retention at high operating temperatures. PCM unit-cell designs that sacrifice some of the inherently large resistance contrast in order to suppress resistance drift have been proposed and demonstrated, in both memory and VMM contexts.[223,224,225] It turns out that intra-device (e.g., shot-to-shot) variability in the rate at which devices drift over time (the "drift coefficient") is more problematic than the actual drift itself.[226] This is because any highly-predictable signal loss can be compensated by signal amplification, at least until background noise becomes strongly amplified.

2.3.3. ELECTROCHEMICAL MEMORY (ECRAM):

Electrochemical random access memory (ECRAM) has recently emerged as a promising candidate for analog memory.[227, 730] ECRAM is a family of three-terminal devices where charge sent to the gate electrode causes an electrochemical redox reaction in the bulk of the channel. The reaction modulates the source drain conductance. ECRAM channel and gate electrodes are made of redox-active materials (organic or inorganic) that conduct both electrons and ions (i.e., mixed conductors) and an ionically-conducting, electron-blocking electrolyte. To maintain global charge neutrality in the device, a counter-ion, typically lithium ions or protons, moves between the gate and channel and compensates for the changing oxidation state of the gate and channel. To retain the analog state and prevent the ECRAM from discharging, an access device such as a diffusive memristor is required on the gate.

In contrast to traditional semiconductor devices where dopants are static after manufacturing, ECRAM represents a class of devices with dynamic dopant control. For analog memory, a major advantage of ECRAM is the large, charge-neutral volumetric capacity that can be exploited to support gradual tuning of the transistor source-drain conductance. Such properties have promise for synaptic memory for artificial intelligence applications.[729] The storage capacity can be several orders of magnitude larger than

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

traditional semiconductor devices where information is stored at oxide interfaces. For example, flash-based memory store roughly 50 aC for a 14/16 nm node. By comparison, lithium containing metal oxides report volumetric capacities at ~5,000 C/cm^3 and polymer-based electrodes reported at ~50 C/cm^3 which could provide as much as 50 pF/μm^2 for a 10nm thick channel.

Additionally, ECRAMs may offer promise for low voltage digital logic.[228] Dynamic doping can lead to sharp metal insulator transitions, e.g., due to correlated electron effects in redox-active oxides,[229] and may result in abrupt low voltage switching.

2.3.4. MAGNETIC NEURAL NETWORK DEVICES

Several types of magnetic circuits can be used to implement neural networks including spin-diffusion-based devices,[230] charge-coupled spin logic (CSL),[231] and domain wall logic (mLogic).[232]. The use of these different devices has been benchmarked.[233]

Magnet switching dynamics that follow the Landau-Lifshitz-Gilbert (LLG) equation with a spin-transfer-torque term are quite similar to the cell dynamics in a Cellular Neural Network (CeNN). CeNNs have been designed based on spin diffusion using all-spin logic (ASL) with PMA magnets as the basic building block. A CeNN cell can also be implemented by using MTJs as synapses and using spin Hall effect or domain wall propagation-based devices as the neuron. In all these cases, the read-out circuit consists of read and reference MTJs and an inverter that amplifies the voltage division between the two MTJs.

In contrast to Boolean circuits, spintronic devices are more attractive compared to charge-based devices. This is because a single magnet can mimic the functionality of a neuron, and these spintronic devices operate at a low supply voltage. The domain wall device provides the best performance, in terms of Energy-Delay Product (EDP), thanks to its low critical current requirement. The spin diffusion based CeNN with IMA magnets consumes more energy due to the large critical current required to switch the magnet.

For optimal circuit-level performance using spintronic devices, several properties are desired including: MTJs with a large TMR and a moderate resistance-area product, large spin injection coefficient β, large perpendicular anisotropy Ku for PMA magnets, large spin Hall angle θ for SHE materials, and small critical depinning current for domain wall magnets.

2.3.5. FLOATING GATE

Floating gate synapses (so-called “synaptic transistors”) were first developed in 1994.[234] They are modified EEPROM devices which can be fabricated in a standard CMOS process and programmed to within 0.2% accuracy.[235] A number of sophisticated systems have been developed based on the arrays of synaptic transistors.[236,237,238]. The main advantage of analog and mixed-signal VMMs based on floating gate memories are very high input and output impedances, which help reducing overhead of peripheral circuitry. The main drawback of synaptic transistor approach is the relatively large cell area, i.e., > 1000F^2, where F is the minimum feature size[239], leading to higher interconnect capacitances and hence larger energy losses and time delays in analog computing circuits.

Recently, it was shown that much better area may be obtained re-designing, by simple re-wiring, the arrays of the ubiquitous NOR flash memories with their highly optimized cells.[240,241] One representative example is Embedded SuperFlash (ESF) memory from Silicon Storage Technology, Inc.[242] The areas of the modified arrays of the ESF1 and ESF3 NOR flash memories, with the latter technology scalable to F = 28 nm, are close to 120F^2 and may be further reduced to ~40F^2. (Note that such areas are much smaller compared to the contemporary 1T1R ReRAM.) Modified 180 nm ESF1 technology was successfully utilized to demonstrate a medium-scale (28×28-binary-input, 10-output, 2-layer, 101,780-synapse) network for pattern classification. The measured delay and energy dissipation compared very favorably with digital approaches, while the results for chip-to-chip statistics, long-term drift, and temperature sensitivity of the network were also very encouraging. Simulations have shown that similarly superior energy efficiency may also be reached in mixed-signal neuromorphic circuits based on industrial-grade SONOS floating gate memories.[243,244]

Even higher density floating gate neuromorphic circuits can be achieved by utilizing NAND flash memories. 2D NAND memory devices designed for digital memory application are already capable of storing 4 bits (16 levels) in a single transistor of 100 nm × 100 nm area in 32nm process.[245] Commercial NAND manufacturers have shown devices at 15 nm[246] and 19 nm.[247,248] EEPROM devices are found at every CMOS IC node, including 7 nm and 11 nm nodes. At these nodes, we still expect very small capacitors to retain 100s of quantization levels (7-10 bits) limited by electron resolution. In practice, larger capacitors are used, resulting in sufficiently high potential resolution. One expects EEPROM linear scaling down to 10 nm process to result in a 30 nm × 30 nm or smaller array pitch area. Perhaps, the most exciting opportunity is presented by the modern 3D NAND circuits. 3D NAND memories already feature 96 layers of floating gate cells.[249] The number of layers is projected to further increase to 512 to enable 10 Tb/in^2 density,[250] which will be essential for storing large-scale neuromorphic models. The very high density of 3D NAND circuits is achieved at the cost of certain restrictions at the circuit level, such as cells connected sequentially in strings and shared

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

gate (word) voltages for the cells in the same level. The sequential structure of NAND flash memory can be efficiently exploited by using time multiplexed computations at the architecture level, in which one cell from a string being utilized at a particular time step in a distributed VMM circuit.[251] Time-domain encoding of inputs was proposed to implement VMM circuits based on the existing 3D-NAND flash memory blocks with common word plane structure, not requiring any modification.[252]

2.3.6. CAPACITOR-ON-GATE

A recent proposal called for a small capacitor that can be programmed with standard CMOS devices to be tied to the gate of a read transistor.[253] In contrast to DRAM, where the charge on the capacitor is transferred through a select transistor onto a bit-line for readout, here the voltage on the capacitor modulates the conductance of a read transistor by direct attachment to its gate terminal. Although the charge leaks away with a time-constant of milliseconds, the training process can succeed if the time-per-example is at least 100,000× smaller than the decay constant (e.g., 20 ms decay constant and 200 ns per training example).[254] One method to obtain good update linearity, so that the amount of charge added and subtracted are balanced, is to use multiple large transistors to supply the current in each unit cell. Additional transistors are then added in order to ensure that, as per the weight-update algorithm, charge is added or subtracted only when both the upstream neuron (say, along the same row) and the downstream neuron (along the same column) agree that weight update should occur in a particular synapse shared by those two neurons. Recently, Ambrogio et al. introduced a combined PCM+capacitor-on-gate unit-cell, in which the PCM provided the non-volatile storage in a "higher significance" conductance, and the capacitor-on-gate devices provided high update linearity in a "lower significance" conductance, with periodic weight transfer from the lower to higher significance devices. The number of transistors associated with the capacitor could be reduced to three: The read transistor, an NFET for charge subtraction and a PFET for charge addition. This was made possible by giving the downstream neuron control over the NFET and PFET gates, and having the upstream neuron control the source contacts of these same two charge addition/subtraction transistors. Furthermore, variability between charge-addition and charge-subtraction due to process nonuniformity was compensated upon weight transfer from capacitor-on-gate cell to the PCM devices using a "polarity inversion" technique.[734] Later, it was shown that this same technique could suppress fixed device variabilities in other kinds of lower-significance conductances, including PCM devices.[255] This combined PCM+capacitor-on-gate unit-cell was shown to allow GPU-equivalent training accuracies, despite the known imperfections of PCM devices and typical fab-level CMOS variability in the capacitor-on-gate devices.[734]

2.3.7. CHARGE-BASED ANALOG ARRAYS FOR VMM

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

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

High-density mixed-signal adiabatic processors[256,257,258] using CIDs have been developed using these principles. To optimize for resonant adiabatic energy recovery a stochastic encoding and decoding scheme can be used to ensure a constant capacitive load of the CID array. This has resulted in better than 1.1 TMACS/mW efficiency excluding on-chip digitization.

Alternatively, several approaches have combined a capacitive charge-based VMM with analog-to-digital conversion to maintain high overall system efficiencies. Many analog multiply-and-accumulate operations can be performed for each digitization. High precision implementations of capacitive charge based VMM have achieved low-pJ/MAC energy efficiencies,[259] while low-precision versions have achieved efficiencies at the level of 100 fJ/MAC.[260] Comparison of key metrics with the state-of-the-art in analog capacitive VMM ICs[261,262,263,264,] is given in Table BC4.2 above.

Table BC2.2 Metrics for Analog Capacitive Vector-Matrix Multiply (VMM) ICs

2.4. MEMORY SELECTOR DEVICE

The *capacity* (or *density*) is one of the most important parameters for memory systems. In a typical memory array, a memory cell 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,

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

the storage node could in principle be scaled down below 10 nm,[265] 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 may 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.[266,267] 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).[268] Volatile switches with large on/off ratio can also be used as selector devices. Figure BC2.2 shows a taxonomy of memory selector devices.[269] In addition to external selector devices, some storage elements may have inherent self-selecting properties (e.g., intrinsic nonlinearity or self-rectification), which may enable functional crossbar arrays without external selectors.

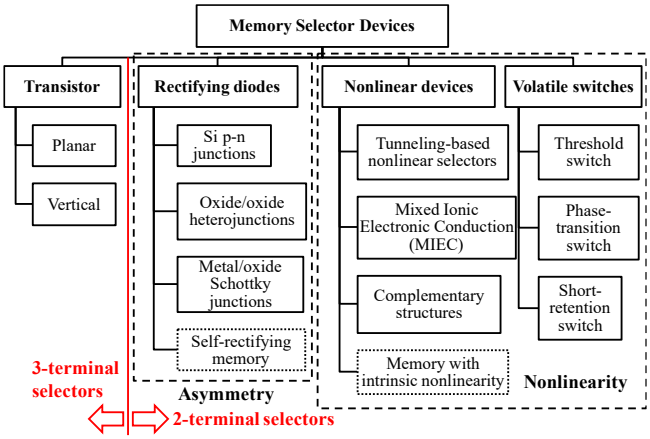


Figure BC2.2. Taxonomy of Memory Select Devices

Since the announcement of 3D cross-point (XP) memory by Intel and Micron in 2015, threshold switch selectors have become one of the most actively researched selector devices. With large on/off ratio and small turn-on voltage, threshold switch selectors can be effective in large crossbar arrays, especially for memory elements with small on/off ratio (e.g., STT-MRAM). As switching devices, threshold switch selectors also face challenges in cycling endurance. Another important challenge is the resistance and voltage balance between the memory element and the threshold switch selector, related to the voltage redistribution when the memory element and the selector, one or both, start switching. Historically, diode selectors were among the earliest two-terminal selectors in development; however, rectifying diodes are only suitable for unipolar memory elements. Nonlinear selectors usually rely on large resistance contrast in exponential transport characteristics to separate on and off states. They tend to have better endurance, but at the same time have non-negligible or even significant on-resistance that contribute to higher array operating voltage and power consumption.

Previous editions of the "Beyond-CMOS" chapter have described the selector devices in Figure BC2.2.[270] No major update is included in the 2022 edition.

2.5. 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.[271,272] Such a device requires a non-volatile memory (NVM) technology that could be manufactured at a very low cost per bit.

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

However, while these NVM candidate technologies are still relatively unproven compared to Flash, there is a strong opportunity for one or more of them to find success in applications that do not involve simply "replacing" NAND Flash. Storage Class Memory can be thought of as the realization that many of these emerging alternative non-volatile memory technologies can potentially offer significantly *more* than Flash, in terms of higher endurance, significantly faster performance, and direct-byte

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

access capabilities. In principle, SCM could engender two entirely new and distinct levels within the memory and storage hierarchy. These levels would be differentiated from each other by access time, with both levels located within more than two orders of magnitude between the latencies of off-chip DRAM (~80 ns) and NAND Flash (20 µs).

In July 2015, Intel and Micron jointly announced a new non-volatile 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.[273,274] (Note that it is most likely that the latency referred to here is write latency rather than read latency, since NAND write latency is much slower than its read latency.) 3D-Xpoint technology, said to have been implemented at the 128Gbit chip level, is based on a two-layer stacked crossbar array, with each intersection point containing a non-volatile memory device and a nonlinear access device.[275,276] The projected array specifications and the target applications of 3D-Xpoint memory are, for all intents and purposes, indistinguishable from those described for SCM. Thus we can consider 3D-Xpoint as the first commercial implementation of the SCM concept.

Previous editions of the "Beyond-CMOS" chapter have detailed description of SCM types, specifications, device candidates, and architecture considerations. No major update is included in the 2022 edition.

3. Emerging Logic and Alternative Information Processing Devices

3.1. Overview

Recent developments in the global semiconductor industry, particularly as documented in the More Moore chapter of this Roadmap, suggest that there is strong technical life in the near future maturation of devices such as lateral and vertical gate-all-around (GAA) transistors for extending Moore's Law. In this chapter, historically, the options for devices beyond conventional silicon transistors have been tracked, under the assumption that of the wide variety of options under research, some gradual winnowing and maturation of these devices to commercialization would manifest. Along such lines, this chapter categorized these options in three classes: (1) devices for CMOS extension, (2) charge-based devices beyond CMOS, and (3) non-charge-based or alternative information processing devices.

As the end-stage roadmap for conventional silicon-based CMOS becomes clearer, the focus of this section likewise becomes more refined and just two classes are considered: (1) CMOS extension devices consisting of a small set of candidates that might be viable for industrial incorporation at the end of the CMOS roadmap; and (2) Beyond CMOS devices that are primarily research-grade in their maturity but depart in significant and novel ways from CMOS in their behavior or performance.

The CMOS extension candidates include carbon-nanotube devices, 2D material channel FETs, and tunneling transistors. The following section provides a detailed review of present status and prospects for each. The section after provides a short synopsis of recent results for the wider variety of novel research candidates.

Table BC3.1 Beyond CMOS Devices for Logic and Computing

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 over twenty 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.

Copyright © 2022 IEEE. All rights reserved.

In the past several years, significant advances have been made in understanding and enhancing device performance in CNT FETs. These include (1) realizing end-bonded contacts having an effective contact length of 0 nm with reasonable performance,[277] (2) detailing the impact of contact scalability in CNT FETs,[278] (3) maintaining performance as the channel length is scaled down to 9 nm without observing short channel effects,[279] (4) fabricating complementary gate-all-around FETs,[280] (5) fabricating an FET with an intrinsic fT of 153 GHz,[281] (6) fabricating CMOS inverters and pass-transistor logic operating at 0.4 V with a non-doped CNT,[282] (7) fabricating a carbon nanotube computer composed of 178 FETs,[283] (8) progress towards reducing the variability in CNT FETs,[284] (9) understanding origins of hysteresis,[285] and (10) fabricating CNT FETs with ON-current of 0.5 mA/µm.[286]

In addition to improvements at the device level, continuous progress has been achieved toward overcoming the dominant material challenges,[287] including the need to achieve purified and sorted semiconducting CNTs with a relatively uniform diameter distribution and then position the CNTs into aligned, closely packed arrays with consistent pitch. With a target purity of 99.9999% semiconducting CNTs and placement density of >125 CNTs/µm (<8 nm pitch), much work still remains. However, it is important to note that progress continues to be steady and without fundamental obstacles barring these goals from being realized. There remains a need for further research toward improving other device-level aspects, including further reduction of contact effects at small contact lengths, demonstrated reduction in variability, improved control of gate dielectric interfaces and properties, and the experimental study of devices and circuits fabricated using the most scaled and relevant device structures and materials. In short, much work remains for CNT FETs, but they have some of the most substantial (and already demonstrated) potential in high-performance, low-voltage, sub-10 nm scaled transistor applications.

3.2.2. 2D Material Channel FETs

Two-dimensional (2D) semiconductors such as certain transition metal dichalcogenides (TMDs) are promising candidates as future channel materials in field-effect transistors (FETs) for very large-scale integration (VLSI). Since their structure does not require the formation of a three-dimensional crystal, ultra-thin body thicknesses (t_{body}) of sub-1nm are achievable in monolayer TMDs without dangling bonds in principle. At these dimensions, semiconductors as MoS_2, WS_2, $MoSe_2$, and WSe_2 exhibit bandgaps of 1.6 to 2eV[288] as desirable for achieving low currents in the device off-state. Mobilities are consistently extracted in the range of 20 to 60cm^2/Vs at room temperature from devices fabricated on traditional oxide substrates,[289,290] while higher mobilities are predicted assuming limited scattering induced by phonon and defects.[291] Note that some very high reported mobilities can be a result of using "incorrect" extraction methods as discussed in reference.[292] Considering the body thickness, these are outstanding transport properties that are unachievable in any three-dimensional (3D) crystal thinned to these dimensions. At the same time the ultra-thin body allows for most aggressive channel length scaling, since electrostatic gate control benefits not only from EOT scaling but also the thinning of t_{body}. A key parameter in this context is the so-called geometric screening length λ that describes any band bending inside the transistor channel along the source/drain direction and is thus a measure of how short of a gate length L_g can be achieved without introducing short-channel effects. The smaller EOT and t_{body}, the smaller λ and the more aggressively L_g can be scaled. Moreover, it is the same characteristic length scale λ that defines the band bending at the metal-to-TMD interface, i.e., Schottky barriers (SBs) in the source/drain contact region, in this way impacting the contact resistance in device structures without a sophisticated doping profile. Last, it is again this small λ that allows for the introduction of novel device concepts such a 2D based tunneling field-effect transistors (TFETs) as discussed below.

TMDs, in general, are characterized by the chemical formula MX_2, where M is a transition metal element and X is a chalcogen. They can be metallic, half-metallic, semiconducting, or superconducting depending on their compositions. The four above mentioned TMDs—MoS_2, WS_2, $MoSe_2$, and WSe_2—are the most popular semiconductors explored for transistor applications. A field-effect transistor (FET) from a monolayer (ML) molybdenum disulfide, MoS_2, was the first TMD FET demonstrated in 2011.[293] The key advantage over graphene is the presence of a bandgap, without the need to create the same through formation of nanoribbons that are notoriously hard to fabricate reproducibly, show typically substantial mobility degradation and allow only for the opening of moderate bandgaps.[294,295,296] Moreover, similar to carbon nanotubes, slightly different ribbon cuts can result in metallic or semiconducting behavior making this approach questionable for technologically relevant applications.

Electrical properties of semiconducting TMDs depend on the number of layers due to quantum confinement effects and changes in symmetry. For example, ML MoS_2 has a direct band gap of 1.9 eV, while bulk MoS_2 exhibits an indirect bandgap of 1.2 eV.[297] While the first ML TMD FET demonstrated by Radisavljevic et al.[293] from exfoliated MoS_2 showed a large on/off current ratio (~10^8) and good subthreshold swing (74 mV/decade) for a channel length of L_g=1.5µm and a gate dielectric film thickness of 30 nm HfO_2, a lot studies have focused on exploring transport properties of few-layer TMDs. Promising performance has been reported in more aggressively scaled few-layer TMD devices[298,299] and highlights the true scaling opportunities that TMDs offer. In fact, Desai et al. fabricated a bi-layer MoS_2-FET using a CNT with 1nm diameter as a gate.[298] For a ZrO_2 gate dielectric thickness of 5.8nm, an on/off current ratio ~10^6 and a subthreshold swing (SS) of 65 mV/decade were obtained, respectively. Note that due to fringing fields, a significant portion ~40 nm in length of the MoS_2 channel is impacted by the nanotube gate in the device off-state, aiding in achieving the aforementioned low SS value. Encouraging results in terms of scaling behavior were also reported by Pang et al. using few-layer WS_2 as the channel material in double-gated TMD FETs with channel lengths down

Copyright © 2022 IEEE. All rights reserved.

to 40 nm.[299] In their case the double gate consisted of an Al_2O_3/HfO_2 (2.5 nm/5.5 nm) as top gate dielectric and HfO_2 (2.8 nm) as bottom gate dielectric, allowing to achieve record high on-currents in excess of 600 μA/μm at drain bias of V_{ds}=1V and overdrive of V_{gs}-V_{th}=2V. While much work has been done on multi-layer TMD transistors due to higher achievable mobilities,[300] it is ultimately essential to focus on ML TMD FETs to benefit from the above discussed λ-scaling to the maximum extent. Some emerging work on ML devices will be discussed below.

While both of the above materials, i.e., MoS_2 and WS_2 exhibit n-type characteristics under typical gate field conditions due to a much preferred line-up of the source/drain metal Fermi levels with the conduction band edge, indeed WSe_2 allows for hole injection at sufficiently negative gate voltages due to a near mid-gap line-up of the source/drain metal Fermi levels.[301] This effect is not related to doping, but to the specifics of electron and/or hole injection from the contacts. The actual doping level determines the threshold voltage of the devices, as in any conventional transistor, but the absence of the so-called hole-branch in MoS_2 and WS_2 Schottky barrier (SB) FETs is a result of the suppressed hole injection from the source/drain contacts. This emphasizes the need to understand and control source/drain contacts to the TMD channel, which is key to lowering contact resistances in TMD FETs to unravel their intrinsic performance specs. Due to limited access to reliable doping approaches for TMDs, particular attention has been paid to engineering metal source/drain contacts to improve the current injection. Bismuth has recently been reported as enabling ~mA/μm on-currents at drain voltages of around 1V.[302] More work is needed to understand how to unlock the intrinsic channel properties, employing improved contact engineering schemes.

Another important finding in the context of TMD FET performance is related to their sensitivity to the exact environmental conditions. This is a natural result of the fact that TMD FETs in essence only consist of interfaces. It has been reported that scattering of charge carriers in the TMD channel with remote impurities via Coulomb interaction can result in a thickness dependent mobility,[290,303] i.e., a decrease of mobility (μ) for thinner t_{body}. It has also been pointed out that while on one hand improved screening is expected to suppress impurity scattering in gate stacks that include high-κ dielectrics,[304] on the other hand the same high-κ material has a detrimental impact on the TMD mobility through the interaction between soft phonon modes of the dielectric and charge carriers in the TMD channel.[291] Controlling the environmental impact and the interaction between charges in the TMD with the surrounding is thus of utmost importance. Note that the above statement about μ being a function of t_{body} should not be misinterpreted as if TMD FETs with a thicker body are preferred. The ultimate benefit of TMD FETs is related to their scalability, as discussed above, making monolayer devices undeniably is the favorable choice. Therefore, the goal should be to explore suitable gate stacks, including gate dielectrics, for *monolayer* TMD FETs that result in the most desirable transport in terms of reduced scattering inside the channel.

Most work on TMD FETs has been focusing on demonstrating either high on-state performance or good control of the off-state by employing devices with a long channel and/or thick gate dielectric. The argument is that extracting important intrinsic properties, such as mobility, allows projecting the performance of scaled devices. This argument, however, is only valid in a very limited sense. Contact effects cannot be accurately extracted in very long channel devices, while being ultimately limiting the device performance of ultra-scaled devices. Moreover, metal source/drain contacts that result in the formation of SBs at the metal/TMD interface do not give rise to a constant series resistance, since most back-gated device geometries impact the SB thickness that exponentially controls the current injection when scanning the gate voltage, while modulating the channel simultaneously. This in turn can prevent the correct extraction of intrinsic properties such as the carrier concentration in the channel and its mobility unless scaling of devices is utilized to clearly distinguish between contact and channel effects. Most importantly, short-channel effects cannot be studied unless both channel lengths and dielectric film thicknesses are properly scaled. These are just a few reasons why there is the urgent need for more work on scaled devices, since making predictions based on devices with long channels and/or thick gate dielectrics is likely to be misleading. It is the full device and circuits from the same that needs to be studied, since TMD FETs are *not* behaving as conventional CMOS devices.

This leads to another important topic, namely the need for both, n-type and p-type transistors of similar performance from TMDs —or ideally one type of TMD. Since substitutional doping through growth processes to achieve degenerate doping is not yet available, alternative approaches such as "charge transfer doping" have been employed to achieve e.g., n-doping in MoS_2 using benzyl viologen (BV) molecules[305] and p-doping in WSe_2 using MoO_3,[306] WO_x,[307] NO_2 [308] or NO.[309] While some of the above examples demonstrated a significantly enhanced on-current level due to improved contact resistances and shifted threshold voltages, the degenerate doping levels achieved resulted in the absence of a significant gate voltage response of the channel, i.e., the devices could not be turned off. It should be noted that unipolar transistor operation could be enabled if doping is only used in the contact regions rather than the entire TMD channel.[305,308] Last, it should be noted that it is very desirable to extend the scaled device approach, as described in the last paragraph, towards doped scaled TMD FETs with a particular focus on threshold voltage control and additional attention on dopants' air stability.[309]

The above section on doping of TMDs and two-dimensional (2D) materials obviously aims at developing CMOS type circuitry from 2D materials. As early as 2011, experimental circuit demonstrations[310] accompanied the work on improving individual

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

device characteristics. Interestingly, due to the relative ease of fabricating MoS_2 transistors at a sufficiently large scale, most circuits utilize exclusively n-type MoS_2 FETs[311,312] and only very limited work has truly focused on CMOS integration based on n-type MoS_2 and p-type black phosphorus (BP)[313] or n-type MoS_2 FETs and p-type WSe_2 FETs.[314] Using the same 2D material to create both, n-type and p-type devices and assemble the same into a circuit has only been accomplished so far using WSe_2 [315] and BP.[316,317]

Beyond the work that focuses on TMD-based FETs, 2D materials have also been employed to demonstrate the feasibility of novel device concepts, in particular focusing on steep slope devices. Using MoS_2 as the active channel material, germanium as the source electrode and a solid polymer electrolyte as the gate, tunneling field-effect transistor (TFET) operation with a subthreshold swing as small as ~31mV/dec over four decades of drain current at room temperature was demonstrated.[318] In another approach, the so-called Dirac source (DS) FET demonstrated a subthreshold swing of ~40mV/dec employing a 2D graphene injector and a carbon nanotube.[319] The operation principle is similar to a TFET but uses graphene without a bandgap to create a low-pass filter by making use of graphene's low density of states (DOS) at the Dirac point. Another interesting device concept employing 2D materials with multiple gates demonstrated the opportunity to reconfigure the same device into an n-MOSFET, p-MOSFET, n-TFET and p-TFET,[320,321] highlighting the unique device level opportunities that 2D materials offer in comparison to conventional, silicon-based CMOS.

Growth technology of TMDs is also in progress. In fact, the synthesis of TMDC film dates back to the 1980's, when the growth was performed with van der Waals epitaxy.[322] More recently, Lee *et al*. demonstrated CVD growth of MoS_2 using MoO_3 and S powder as precursors.[323,324] The growth temperature was 650˚C. Single-crystal monolayer MoS_2 flakes were successfully obtained. This method was further improved and single-crystal MoS_2 flakes as large as 120 μm in lateral size were obtained.[325] There have been a quite a few reports using similar methods for synthesizing TMDCs, including $MoSe_2$,[326] WS_2,[323,327] and WSe_2.[328] However, uniform growth over a large area is not easy using this powder-based CVD technique. In 2015, *Kang et al*. succeeded in large area growth of MoS_2 by MOCVD using $Mo(CO)_6$ and $(C_2H_5)_2S$ as precursors.[329] The electron mobility of MoS_2 in this case was 30 cm^2/Vs at room temperature. In general, mobility of CVD MoS_2 is lower than that of exfoliated MoS_2, which is still an important issue to address. Furthermore, MoS_2 deposition by using sputtering is also in progress.[330,331,332] The sputtering method is scalable, but it is still difficult to obtain film with a quality as high as that by MOCVD.

3.2.3. TUNNEL FETS

Tunneling Field Effect Transistors (TFETs) have the potential to achieve a low operating voltage by overcoming the thermally limited subthreshold swing voltage of 60 mV/decade by utilizing tunneling as a switching mechanism.[333,334,335] In its simplest form, a TFET is a gated, reverse-biased p-i-n diode with a gate controlled intrinsic channel. There are two mechanisms that can be used to achieve a low voltage turn on. The gate voltage can be used to modulate the thickness of the tunneling barrier at the source channel junction and thus modulate the tunneling probability.[336,337,338,339] The thickness of the tunneling barrier is controlled by changing the electric field in the tunneling junction. Alternatively, it is possible to use energy filtering or density of states switching: if the conduction and valence band do not overlap at the tunneling junction, no current can flow; once they do overlap, current can flow. Simulations have predicted arbitrarily steep subthreshold swings when relying on density of states switching as the current is abruptly cutoff when the conduction band and valence band no longer overlap.[335] If phonons or short channel lengths are accounted for, simulated subthreshold swings on the order of 20–30 mV/decade are typical.[340] It is possible to use the tunneling switching mechanisms in series with the standard MOSFET thermal switching mechanism to get an overall subthreshold swing that is steeper than 60 mV/decade when no individual mechanism is steeper than 60 mV/decade.[341] The best experimental results to date have relied on a combination of thermal switching and density of states switching.[342].So far, the experimental results are far worse than simulated device characteristics. The review by Lu and Seabaugh[333] shows a comprehensive benchmarking of published experimental results prior to 2014. The benchmarking shows two problems with TFETs as a MOSFET replacement: 1) Devices are unable to achieve SS <60 mV/decade over a large range or at useful current levels and 2) The on-state current is too low for reasonable performance.

The review shows 14 reports of subthreshold swings below 60 mV/decade, and a few additional results have been published since. Most of the results are for group-IV materials such as Si,[338,343,344,345,346,347] strained SiGe,[348] Si/Ge,[349] and strained Ge.[350] Nanowire III-V TFETs have shown even steeper swings. An InP/GaAs heterojunction[351] has shown 30 mV/decade at 1 pA/μm. The steepest result ever reported is in a Si/InAs heterojunction[352] of approximately 20 mV/decade at 0.1 pA/μm. However, there are only a few data points defining this result. Even for low power applications, at least 1–10 μA/μm is needed.[353] Recently, a promising InAs/GaAsSb/GaSb nanowire heterostructure TFET with a 48 mV/decade SS at 67 nA/μm and an I_{60} (current at 60 mV/decade) of 0.31 μA/μm was demonstrated.[342]

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

in on-state current. Unfortunately, this is a fundamental tradeoff when modulating the thickness of the tunneling barrier:[334] barrier thickness modulation only gives a step subthreshold swing at low current densities.

An ideal density of states (DOS) switch would switch abruptly from zero-conductance to the desired on-conductance, thus displaying zero subthreshold swing voltage.[354] Practically, the band edges are not perfectly sharp, so there is a finite density of states extending into the band gap. Optical measurements of intrinsic GaAs imply a band edge steepness of 17 meV/decade.[355] However, the electrically measured joint DOS in diodes has generally indicated a steepness >90 mV/decade.[356] This broadening is likely due to the spatial inhomogeneity and on heavy doping that appears in real devices. Effectively, there are many distinct channel thresholds in a macroscopic device, leading to threshold broadening. Fortunately, it has been experimentally demonstrated that a band edge worse than 60 mV/decade can be combined with a thermal switching mechanism to give an overall subthreshold swing better than 60 mV/decade.[341]

TFETs are reverse-biased diodes hence are subject to generation in the depletion region. These generation events include but are not limited to bulk and interface trap-assisted Shockley-Read-Hall (SRH)[357,358,359] and spontaneous and Auger generation.[360] Calculations based upon these mechanisms show that these significantly degrade the subthreshold swing and increase the leakage currents but do not prevent TFETs from achieving sub-60 mV/decade subthreshold swing. Material defects and gate interface traps make these effects worse and result in worse band edges.

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

3.3. Beyond-CMOS Devices

3.3.1. Spin FET and Spin MOSFET Transistors

Spin-transistors are classified as "non-conventional charge-based extended CMOS devices,"[361] and can be further divided into two categories: the spin-FETs proposed by Datta and Das[362] and spin-MOSFETs proposed by Sugahara and Tanaka.[363] 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.[361,364] 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 via the Rashba spin-orbit interaction. Both types of devices behave as a transistor and function as a magnetoresistive device. The important features of spin transistors are that they allow a variable current to be controlled by the magnetization configuration of the ferromagnetic electrodes (spin-MOSFETs) or the spin direction of the carriers (spin-FETs), and they offer the capability for non-volatile information storage using the magnetization configurations. These features are very useful for energy-efficient, low-power circuit architectures that cannot be achieved by ordinary CMOS circuits. Non-volatile logic and reconfigurable logic circuits have been proposed using the spin-MOSFET and the pseudo-spin-MOSFETs, which are suitable for power-gating systems with low static energy.[364,365,366,367,368,369,370]

The full operation suggested for spin-FETs[364,371] and spin-MOSFETs[364,372,373,374,375] have not yet been experimentally verified. For realizing fully functional spin transistors, some important progress in the underlying technologies such as electrical spin injection, spin detection, and spin manipulation in semiconductors (SCs) should be required.[376,377,378] Lots of theories[379,380,381,382] have predicted that the insertion of a tunnel barrier between the ferromagnet (FM) and SC is a promising method for highly efficient electrical spin injection and detection. In particular, large spin signals induced by the efficient spin injection and detection were observed in Si-based lateral spin-valve devices with FM/MgO tunnel-barrier contacts even at room temperature.[383,384,385,386] Also, by using a back-gated device structure with FM/MgO tunnel-barrier contacts,[373,374,375] a basic read operation of Si-spin-MOSFETs was demonstrated at room temperature. These are important developments for Si-based spin-MOSFETs. However, if an insulator tunnel barrier such as MgO was utilized, the large parasitic resistance can cause the obstacle for the development of source and drain structures in the spin transistors. Another key development for highly efficient spin injection/detection in SCs is half-metallic FM contacts. Thus far, electrical spin injection, transport, and detection in SCs without using insulator tunnel barriers have been demonstrated in lateral spin-valve devices with Co_2-Heusler alloy/SC Schottky-tunnel-barrier contacts.[387,388,389,390,391]To reduce the value of RA (interface resistance area product) at the source and drain structures in spin-MOSFETs, the delta-doping of dopant impurities near the Co_2-Heusler alloy/Ge Schottky-tunnel-barrier[392] has been demonstrated, leading to the room-temperature spin transport[393] including local magnetoresistance effect in Ge-based lateral devices with RA values of less than 0.2 $k\Omega\mu m^2$.[394] Recently, the quality of the Co_2-Heusler alloy/Ge heterointerface was significantly improved by inserting Fe atomic layers between Co_2-Heusler alloy and Ge.[395,396] This leads to the two-terminal magnetoresistance (MR) ratios of more than 0.1 % at room temperature even in the Ge-based lateral spin-valve devices.[396] For

Copyright © 2022 IEEE. All rights reserved.

enhancing the MR ratio in Ge-based spin-MOSFET structures, it is essential to further optimize the formation of $L2_1$-ordered Co_2-Heusler alloy/Ge Schottky-tunnel contacts and to reduce the spin relaxation in the Ge channel.[397]

Alternative approaches for realizing spin-MOSFETs have been proposed.[364,369,372,398,399] Pseudo-spin-MOSFETs are circuits that reproduce the functions of spin-MOSFETs using an ordinary MOSFET and a magnetic tunnel junction (*MTJ*) that is connected to the MOSFET in a negative feedback configuration. Although pseudo-spin-MOSFETs offer the same functionality as spin transistors, such as the ability to drive variable current, pseudo-spin-MOSFETs have larger resistance than spin-FETs or spin-MOSFETs.

For spin manipulation in SCs, channel materials with a strong spin-orbit interaction, such as InGaAs, InAs and InSb, are required[362] in order to sufficiently induce the Rashba spin-orbit interaction by an electric gate voltage. Using InAs[400] and InGaAs[401] 2DEG heterostructures with and without FM spin injector and detector, respectively, the spin manipulation was demonstrated by the electric field, meaning the operation of a spin-FET. However, the operation temperature was limited to the low temperature less than 40 K. The experimental proof of electrical spin injection, detection and manipulation in SCs with the strong spin-orbit interaction above room temperature is needed to create spin-FETs. Very recently, it was found that the Rashba spin-orbit interaction is induced by strong electric fields applied to the Si-MOS interfaces at room temperature.[402] For realizing spin-FETs using Si, the changes in MR ratios or magneto-current ratios in a Si-MOS structure should be explored at room temperature.

3.3.2. NEGATIVE GATE CAPACITANCE FET

Salahuddin and Datta originally proposed[403] that, based upon the energy landscape of ferroelectric capacitors, it should be possible to implement a step-up voltage transformer that will amplify the gate voltage of a MOSFET. This would be accomplished by replacing the standard insulator in the gate stack with a ferroelectric insulator of appropriate thickness. The resulting device is called a negative gate capacitance FET or NCFET. The gate operation in this device would lead to subthreshold swing (STS) lower than 60 mV/decade and might enable low voltage/low power operation. The main advantage of such a device[404] is that it is a relatively straightforward replacement of conventional FETs. Thus, high I_{on} levels similar to advanced CMOS would be achievable with lower voltages. An early experimental attempt to demonstrate a low-STS NCFET, based on a P(VDF-TrFE)/SiO_2 organic ferroelectric gate stack, was reported[405] in 2008, and subsequently[406] in a more controlled structure in 2010. These experiments established the proof of concept of sub-60 mV/decade operation using the principle of negative capacitance.

In addition to these experiments using polymer-based ferroelectrics, negative differential capacitance was demonstrated in a crystalline capacitor stack.[407] Essentially, it was demonstrated that in a bi-layer of dielectric Strontium Titanate ($SrTiO_3$: STO) and Lead Zirconate Titanate ($Pb_xZr_{1-x}TiO_3$: PZT), the total capacitance is larger than what it would be for just the STO of the same thickness as used in the bi-layer. This necessarily demonstrates the stabilization of PZT at a state of negative differential capacitance. More recently, in a single PZT capacitor, a direct measurement of negative capacitance was demonstrated.[408] That work determined that when a ferroelectric capacitor is pulsed with an input voltage, it shows an 'inductance-like' discharging in addition to a capacitive charging.

As a recent significant result, it is now possible to grow ferroelectric materials using the atomic layer deposition process (ALD) by doping the frequently used gate dielectric HfO_2 by constituents such as Zr, Al or Si.[409] Using this doped Hf based ALD ferroelectric, a number of experiments have demonstrated the negative capacitance effect.[410,411,412] For example, by using $HfZrO_2$ as a gate dielectric, sub-60 mV/decade STS was demonstrated[412] in finFETs with L_g=30 nm for both nFET and pFET structures. In the last two years, multiple papers have reported this effect for various material systems and channel lengths. Significant among them is the demonstration by GlobalFoundries of NCFETs in their 14 nm finFET technology, with improved subthreshold swing, lower OFF current and lower active power and ring oscillators running at GHz speed.[413]

3.3.3. NEMS SWITCH

Nano-Electro-Mechanical (NEM) switches use electrostatic force to mechanically actuate a movable structure to make or break physical contact between current-conducting *source* and *drain* electrodes. When the electrodes are separated physically by an air gap, no current flows across the gap, resulting in **zero OFF-state current**. The NEM switch undergoes an abrupt change in current conduction ability between non-contacting and contacting states, with nearly **zero subthreshold swing**.[414] While zero OFF-state current provides for zero standby power dissipation, zero subthreshold swing enables very low operating voltages for low dynamic power dissipation as well. Moreover, a NEM 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 NEM relay can be configured to turn on either with increasing control (*gate*) voltage or with decreasing *gate* voltage and can serve as either a pull-down device (passing a low voltage – ground – from the *source* to the *drain*, i.e., discharging the voltage at the *drain*) or a pull-up device (passing a high voltage – V_{DD} – from the *source* to the *drain*, i.e., charging the voltage at the *drain*).[415] Additional advantages of mechanical devices include robust operation across a wide temperature range, down to cryogenic temperatures,[416] and immunity to ionizing

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

radiation. NEM switches can be monolithically integrated with CMOS circuitry by a modular post-CMOS fabrication process with relatively low thermal budget.[417] Since state-of-the-art CMOS technology incorporates air-gapped interconnects,[418] NEM switches can be implemented using multiple interconnect layers formed in the back-end-of-line (BEOL) process.[419] Potential applications for hybrid NEMS-CMOS technology include CMOS power gating,[420] configuration of FPGAs,[421] non-volatile back-up storage of information in SRAM and CAM cells,[419] and energy-efficient, fast and reconfigurable look-up tables.[422]

Conventional planar processing techniques (i.e., thin-film deposition, lithography and etch processes) can be employed to fabricate the conducting electrodes of a mechanical switch. The air-gaps between electrodes are formed with a final "release" etch step in which a sacrificial material such as silicon dioxide, photoresist, polyimide or silicon is selectively removed. The switching delay and operating voltage of a NEM switch can be reduced by scaling down the size of these air-gaps. The smallest air-gap demonstrated to date for a functional NEM structure fabricated using a top-down approach is 4 nm, resulting in an actuation voltage of approximately 0.4 V.[423] As expected, it exhibited very low OFF-state current and sub-threshold swing; however, it is only a 2-terminal device, not suitable for logic switch application. SiC nanowires have been used as the movable structure for NEM switches that are suitable for high-temperature operation. Functioning 2-terminal SiC switches with air-gaps as small as 10 nm and switching voltage as low as 1V have been demonstrated.[424] 3-terminal devices and corresponding logic gates also were demonstrated.[425] Piezoelectric materials have been incorporated in NEM devices to enable sub-1V switching[426].

The operating speed of a NEM switch is much slower than that of a transistor because it is dominated by mechanical delay related to the physical motion of the movable structure rather than the electrical (charging/discharging) delay; therefore an optimized relay-based IC design should arrange for all mechanical movements to happen simultaneously, so that the delay per operation is essentially one mechanical delay.[427] For a pass-gate circuit topology, multiple switches are connected together in series to drive the output signal. This means that the *source* voltage can vary between the reference voltage (ground) and the supply voltage (V_{DD}). For proper pass-gate circuit operation, the state of the switch cannot be dependent on the *source* voltage; therefore a fourth electrode is necessary to provide a constant reference voltage, such that the voltage applied between the control (*gate*) electrode and the reference (*body*) electrode determines the state of the switch, i.e., whether a current path is established between the *source* and *drain* electrodes.[427] The contact and actuation air-gaps can be reduced by biasing the *body* electrode to reduce the magnitude of the *gate* voltage required to switch the device ON/OFF. Moreover, a constant-field scaling methodology can be applied to miniaturize NEM switches for reduced footprint, switching delay and switching energy.[428] With the aid of body biasing, 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.[429,430] Piezoelectric NEM devices have demonstrated 10 mV switching operation using body bias, providing for very low energy dissipation per switching cycle (23 aJ), and an extremely small subthreshold slope (0.013 mV/decade)[426]. A variety of relay-based computational and memory building blocks have been experimentally demonstrated to date.[431,432]

The prospective system-level benefits of mechanical logic have been analyzed by performing simulation-based assessments of VLSI circuit blocks implemented with NEM switches. These indicate that relays can provide for more than 10x reduction in energy per operation as compared with MOSFETs, and can reach clock speeds in the GHz regime.[433] Thus, a major potential advantage of NEM switch technology is **improved energy efficiency**. Moreover, the contact adhesive force and structural stiffness can be engineered to achieve bi-stable operation, which makes mechanical switches attractive for embedded non-volatile memory applications.[434] Recently hybrid CMOS-NEMS circuits have been demonstrated, with non-volatile NEM switches operating at the same V_{DD} as for the CMOS circuitry.[435] The BEOL metallic layers used to form interconnects in a conventional CMOS process can be leveraged to implement compact NV-NEM switches for dynamically reconfigurable circuit functionality[436]. Recently, monolithically integrated hybrid CMOS-NEM circuitry have been fabricated and circuit functionalities have been demonstrated utilizing BEOL NEM switches. Arrays of non-volatile NEM switches have been leveraged to demonstrate look-up table functionality[437] and data search operation.[438] Reprogrammable NEM switches, with air gaps as small as 32 nm, have been fabricated using a standard 16-nm CMOS manufacturing process.[438]

For all of the aforementioned applications, the NEM switches need to operate reliably and consistently for at least 10^9 cycles. Due to the extremely small mass of the movable electrode (less than 1 ng), neither gravitational acceleration nor mechanical vibration substantially affects their operation. Structural fatigue or creep can be easily avoided by designing the movable structure/anchor such that the maximum induced strain is well below the yield strength. However, permanent stiction can be a mode of device failure: for soft electrode materials such as gold and platinum, Joule heating at the contacting asperities can lead to atomic diffusion (welding). This issue can be mitigated by using a refractory electrode material such as tungsten to minimize contact wear and by reducing the peak voltage difference between the *source* and *drain* electrodes (V_{DD}) when they are in contact. NEM switches with tungsten electrodes 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), and endurance exceeding 10^{16} ON/OFF cycles is projected for voltages below 1 Volt and load capacitance below 1 pF. [439] The remaining reliability challenge for NEM logic switches is degradation (dramatic increase) of on-state resistance due to electrode surface oxidation or the formation of friction

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

polymers over the lifetime of device operation. Alternative contacting electrode materials such as Ru/RuO_2 and/or hermetic packaging are potential solutions to this issue.

Adhesive force between the contacting electrodes dictates the minimum required stiffness of the movable structure, which in turn determines the minimum *gate*-to-*body* voltage required to switch the NEM relay. The adhesion energy is determined by metallic bonding (at the contacting asperities) and by van der Waals force (in the non-contacting regions of the electrodes). Self-assembled molecular (SAM) coating of hydrophobic materials has been demonstrated to reduce the adhesive force and thereby enable switching operation at lower *gate* voltage.[440] Sub-50 mV operation of relay integrated circuits demonstrating OR, AND, and XOR gate functionality have been demonstrated with body-biased SAM-coated NEM switches.[441] Operation of NEM switches and integrated circuits at temperatures down to 4K were demonstrated recently for the first time.[442] Due to reduced adhesion energy and elimination of contact oxidation, relays can be operated reliably with voltages as low as 25 mV for more than 10^8 cycles at cryogenic temperatures, showing promise for monolithic integration of multiplexing control circuitry with qubits. Most recently, coupled MEM relays have been proposed to implement an Ising machine utilizing the bias-dependent oscillatory behavior of the relays.[443]In conclusion, the negligible OFF-state current and ultra-low-voltage operation capability of NEM switches make them a compelling option for ultra-low-power digital computing applications such as the Internet of Things, particularly where resilience to extreme temperatures and/or immunity to radiation are required. Furthermore, hybrid CMOS-NEM technology shows promise for enhanced chip functionality, e.g., with dynamically reconfigurable interconnects. Remaining challenges to realizing the promise of mechanical computing include materials and process optimization to achieve stably low contact resistance with minimal contact adhesive force. Further work is also needed to integrate NEM switches on the most advanced CMOS platforms and improve their reliability for future IoT device applications of hybrid CMOS-NEM circuits.

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.[444,445] 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.[446,447] 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.[445] 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).[448] The phase transition time constant in VO_2 materials is in sub-picosecond range determined by optical pump-probe methods.[449] 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.[450] 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.[451, 452] 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[453] and circuits[454] 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[455] and steep-slope transistors.[456] As a Mott device with purely electrostatic modulation in the solid-state base,[457] the VO_2-FET with a high-k oxide/organic hybrid dielectric gate has been proposed.[458,459] Their reversible as well as quick resistance switching upon an application of gate bias and the maximum resistance modulation at the Metal-Insulator transition temperature indicate the possibility of purely electrostatic field-induced metal-insulator (Mott) transition. The gate-tunable abrupt switching device based on a VO_2 microwire integrated monolithically with a two-dimensional tungsten diselenide semiconductor by van der Waals stacking has been reported.[460] Nanofabrication engineering has been demonstrated to enhance the performance of Mott FET[461]. The 3-dimensional Fe_3O_4 nanowires on the length scale of 10 nm exhibited the remarkable Verwey transition at about 112 K, which was found to be 6 times larger than that for the thin-film configuration.[462]

Experimental challenges with correlated electron oxide Mott FETs include fundamental understanding of gate oxide-functional oxide interfaces and local band structure changes in the presence of electric fields. Methods to extract quantitative properties (such as defect density) of the interfaces are an important topic that have not been explored much to date. The relatively large intrinsic carrier density in many of the Mott insulators requires the growth of ultra-thin channel materials and smooth gate oxide-functional oxide interfaces for optimized device performance. It is also important to understand the origin of low room-temperature carrier mobility in these materials.[446] Theoretical studies on the channel/dielectric interfacial electronic band structure are needed for the modeling of subthreshold behaviors of Mott FETs. Metal-insulator transition induced by resonant tunneling in double quantum well structures has been successfully demonstrated in strongly correlated oxides.[463] Understanding the electronic

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

transition mechanisms while de-coupling from structural Peierls (lattice) distortions is also of interest and important in the context of energy dissipation for switching.

While the electric field-induced transitions are typically explored with Mott FET, nanoscale thermal switches with Mott materials could also be of substantial interest. Recent simulation studies of "ON and "OFF" times for nanoscale two-terminal VO_2 switches indicate the possibility of sub-ns switching speeds in ultra-thin device elements in the vicinity of room temperature.[464,465] 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.[466, 467] Electronically driven transitions in perovskite-structured oxides such as rare-earth nickelates[468]-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.[469,470, 471,472] $SmNiO_3$, 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.[473] 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.[474] 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. *Topological Insulator Electronic Devices*

Topological insulators are recently discovered materials that possess a bandgap in their interior, however the topology of their electronic states necessitates that the existence of gapless, conducting modes on their boundaries – one dimensional (1D) edges in the case of two-dimensional (2D) topological insulators, and 2D surfaces in the case of three-dimensional (3D) topological insulators[475,476,477,478]. These conducting modes are protected from backscattering by symmetry, and in the case of 1D edge modes of 2D topological insulators, are expected to be ballistic conductors with conductance e2/h = 38.7 μS per edge, which has been confirmed experimentally in several materials[479,480,481].

Systematic searches of materials databases have found that topological materials are commonplace, representing a significant fraction of all known materials[482,483,484]. Two-dimensional topological insulators have been realized with very large bandgaps of 360 meV (Na3Bi[485]) and 800 meV (bismuthene[486]), significantly exceeding the thermal energy at room temperature (25 meV), which indicates that topological phenomena may be robust at room temperature in suitable materials.

Numerous proposals have been put forward to exploit topological materials in transistor design[487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499]. One topological transistor design envisions switching a 2D material from topological insulator ("on" with ballistic edge channels) to a conventional insulator ("off"). This switching may be accomplished via electric field through several mechanisms such as Rashba effect[488,498], staggered sublattice potential[487,488,491,492], inversion symmetry breaking[489], or Stark effect[485, 490, 493, 494, 497].

Electric field effect switching has been proposed in a number of materials including graphene[488], two-dimensional Xenes (e.g., germanene, stanene, etc.)[487,492,498], monolayer transition metal dichalcogenides in 1T' phase[494], SnTe and Pb1-xSnxSe(Te)[489], topological semimetals such as Cd2As3 and Na3Bi[485,493], and phosphorene[490]. Electric-field switching of the topological state has been demonstrated in at least one material, Na3Bi[485]. Other transistor proposals have focused on electric-field switching of tunnelling between topological edges[500], or using strong disorder to produce an off state in a bulk conducting topological insulator[495].

An important development in 2020-2021 has been the establishment of the basis for low-voltage switching in a topological transistor[498]. For staggered honeycomb lattice topological insulators in the Xene family, electric field acts in two ways; inducing a staggered sublattice potential and creating a Rashba spin-orbit coupling. Both effects act to open a bandgap in the topological insulator. A device termed a topological quantum field-effect transistor (TQFET) exploits this combined electric field effect on the bandgap, which is used as the barrier to electron flow in the "off" state of the transistor. Due to the combined effects of sublattice potential and Rashba spin-orbit coupling on the bandgap, the TQFET can overcome Boltmann's tyranny with a subthreshold swing lower than $kTln10/e \approx 60$ mV/decade at room temperature.

In additional important development in 2020-2021 was the realization that the TQFET concept can be combined with a negative capacitor to further reduce the subthreshold swing. Negative capacitance (realized, for example, in a ferroelectric layer) in series with a positive capacitance has previously been proposed to reduce the subthreshold swing in conventional MOSFETs[501]. However, in a well-designed MOSFET in the subthreshold region (in which the gate capacitance dominates over stray

Copyright © 2022 IEEE. All rights reserved.

capacitances to the channel), the gain due to negative capacitance is negligible[502]. This is because the negative capacitor in series with a positive capacitor acts to amplify the electric field in the positive capacitor at a given voltage. However, in an ideal conventional MOSFET in the subthreshold regime, the gate electric field is negligible as the channel electric potential follows that of the gate, hence the device cannot benefit from amplification of the electric field in the channel.

A TQFET switches by a qualitatively different mechanism, applying an electric field across the insulating channel, rather than changing the electrostatic potential of the channel as in a MOSFET. Thus, a TQFET can benefit maximally from electric field amplification in a negative capacitance structure. A negative capacitance topological quantum field-effect transistor (NC-TQFET) incorporates a stack consisting of gate electrode, ferroelectric, topological channel, ferroelectric, ground plane; with the ferroelectric layers having an effective negative capacitance[499]. The entire stack is designed to have a net positive capacitance to ensure absence of hysteresis and amplification of the electric field in the topological channel layer.

Benchmarking of a hypothetical NC-TQFET using a bismuthene topological channel and La-doped HfO2 ferroelectric layers demonstrated a 10X reduction in switching voltage and 8X reduction in switching energy compared to a GAAFET CMOS LV[503], with high on-current[499]. Further gains appear possible by using ferroelectrics with higher remnant polarization. While the NC-TQFET design is promising for low-voltage devices, many challenges remain to realize such a device. An appropriate topological channel material in the Xene family must be demonstrated experimentally and integrated with an appropriate ferroelectric layer of precisely controlled thickness.

In addition to switching via electric field, topology may also be controlled by magnetic fields[504,505], strain[506], temperature[507], or time-dependent electric fields (such as light)[508,509], hence topological transistors have the prospect of adding additional new functionality for “More than Moore” devices.

The deep connection between topological insulators and spin-orbit coupling also suggests further synergies between topological electronics and spin electronics. Indeed, 2D topological insulators exhibit a quantized spin Hall effect and can be used to produce completely polarised spin current[488], of potential use in spintronics devices. Magnetic 2D topological insulators (quantum anomalous Hall effect) may produce perfectly spin polarised current with spin direction determined by magnetization direction[478,510], opening new possibilities for spin transistors and non-volatile random-access memory[511]. Near-perfect ballistic conduction in the quantum anomalous Hall regime with current-induced magnetization switching at very low currents (1 nA) was recently demonstrated in graphene/boron nitride heterostructures[512] albeit at cryogenic temperatures. Analogous topological effects may be realized for other degrees of freedom such as the valley degree of freedom in materials with multiple conduction valleys, with analogous quantum valley Hall effects switchable by electric field[513,514,515].

3.3.6. 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.[516,517,518,519,520] 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;[521,522] (ii) magnetic waveguides – spin wave buses for spin wave signal propagation between the magneto-electric cells,[523] (iii) magnetic junctions to couple two or several waveguides,[524] (iv) spin wave amplifiers,[525] (v) phase shifters to control the phase of the propagating spin waves,[526] and (vi) spin wave phase error correctors.[527] 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,[516] SWD concept has evolved in different ways encompassing volatile[528] and non-volatile,[529] Boolean[529,530] and non-Boolean,[531] single-frequency and multi-frequency circuits.[532] The primary expected advantages 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 the intrinsically low energy of spin waves in ferromagnets (~10 μeV) and antiferromagnets (~1 meV) as well as built-in non-volatile magnetic memory and magnetic reconfigurability[533], 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.[534] It is based on ferromagnetic $Ni_{81}Fe_{19}$ structure, operates within 1-3 GHz frequency range, and exhibits signal-to-noise ratio of approximately 10 at room temperature.[534] The internal delay of SWD is defined by the spin wave group velocity (e.g., 3×10^6 cm/s in $Ni_{81}Fe_{19}$ waveguides). Power dissipation in SWD is mainly defined by the efficiency of the spin wave excitation. Recent experiments with synthetic multiferroics comprising piezoelectric (lead magnesium niobate-lead titanate PMN-PT) and magnetostrictive (Ni) materials have demonstrated spin wave generation by relatively low electric field (e.g., 0.6 MV/m for PMN-PT/Ni).[535] The later translates in ultra-low power consumption (e.g., 1aJ per multiferroic switching). Recent experimental demonstration[522] of parametric spin wave excitation by voltage-controlled magnetic anisotropy (VCMA) is another promising route towards energy-efficient generation and amplification of spin waves.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

A SWD-based magnonic holographic memory (MHM) has been proposed.[531] The principle of operation of MHM is similar to optical holographic memory, while spin waves are utilized instead of light waves. The first 2-bit MHM prototype based on yttrium iron garnet structure has been demonstrated.[536] MHM also possesses unique capabilities for pattern recognition by exploiting correlation between the phases of the input waves and the output interference pattern. Pattern recognition using MHM has been recently demonstrated.[537] The potential advantage of spin wave utilization includes the possibility of on-chip integration with the conventional electronic devices via multiferroic elements. In addition, magnonic holograms can show very high information density (about 1Tb/cm^2) due to the nanometer scale wavelength of spin waves. According to estimates, the functional throughput of magnonic holographic devices may exceed 10^{18} bits/s/cm.[2,531]

There are two important milestones crucial for further SWD development: (i) nanomagnet switching by spin waves;[538] (ii) integration of several magneto-electric cells on a single spin wave bus. In order to have an advantage over CMOS in functional throughput, the operational wavelength of SWDs should be scaled down below 100 nm.[529] The success of the SWD will also depend on the ability to restore/amplify spin waves (e.g., by multiferroic elements[539], by magneto-electric pumping[540] or by antidamping spin-transfer and spin-orbit torques[541,542]).

A very important recent direction of research is antiferromagnetic (AFM) SWD[543] that offers a thousand-time increase in operation speed due to the THz-scale frequencies of the AFM spin waves. The frequency-momentum dispersion of AFM spin waves is linear, which results in high group velocity of AFM spin waves and faster information transfer within SWD-based circuits. In addition, AFM spin waves usually exhibit degeneracy due to the presence of multiple (typically two) sublattices. This degeneracy provides an additional degree of freedom for encoding information carried by AFM spin waves. For example, AFM spin wave polarization can be used to encode information [544]. Alternatively, information can be encoded in the phase shift between the degenerate spin wave modes, which can be used in topologically protected spin wave gates[545]. Challenges in AFM SWD development include: (i) efficient interconversion between spin wave signals and electric signals[546,547] and (ii) efficient methods to control the AFM ground state, including the AFM domain control. Development of AFM-based magnetic tunnel junctions with high tunnel magneto-resistance is needed to address the former challenge. Magneto-electric control of AFM domains is a promising approach to the latter challenge.

An intriguing recent direction of research is SWDs based on two-dimensional van der Waals (2D vdW) materials[548]. Spin waves in both ferromagnetic[548] and antiferromagnetic[549] 2D vdW materials have been detected. The potential advantages of the 2D vdW materials are: (i) the ease of spin wave control by out-of-plane electric fields[550] and (ii) the ability to easily assemble designer magnetic heterostructures. Major challenges in this field include: (i) finding 2D materials with magnetic ordering temperatures much above the room temperature and (ii) identifying materials with low magnetic damping for long-range spin wave propagation.

3.3.7. EXCITONIC DEVICES

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

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

Experimental proof-of-principle for excitonic devices including IX transistors,[551] diodes,[552] and CCD[553] was demonstrated. IX condensate and coherent IX currents[554] and, recently, long range coherent IX spin currents[555] were observed at low temperatures. New heterostructures based on single-atomic-layers of transition-metal dichalcogenide (TMD) for room-temperature IX circuits were proposed.[556] Recent progress includes the realization of IXs at room temperature in TMD heterostructures,[557] discovery of IX coherent spin currents in GaAs heterostructures[558], development of first split-gate device for IXs, the basic mesoscopic device[559], development of advanced platform for high-mobility excitonic devices[560], development of TMD heterostructures with

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

small IX linewidth and finding charged IXs, indirect trions. At present, the studies are focused on the development of basic concepts of excitonic devices at low temperatures using GaAs heterostructures and development of room-temperature excitonic devices using TMD heterostructures.

3.3.8. TRANSISTOR LASER

The Light-Emitting Transistor (LET)[561] and Transistor Laser (TL)[562,563] utilize a fundamental characteristic of bipolar transistors - that electron-hole recombination in the base is an essential feature of the transistor and that the resulting photon signal in a direct-bandgap base is correlated to the electrical signal driving and being driven by the device. The TL can be thought of as a 5 terminal heterojunction bipolar transistor (HBT) with 3 conventional electrical terminals (emitter, base, and collector) and 2 optical terminals (input-generation and output-recombination).[564] Very-high-speed transmission and processing are enabled by the projected capability to achieve over 200 GHz bandwidths in the GaAs- or InP-based devices.[565] An advantage of the TL is that a single epitaxial layer structure can be used for devices that generate photons, detect photons, and perform electronic functions. The layer structure of the TL resembles a heterojunction bipolar transistor with features added to enhance base recombination and control base transit time.[563,566] When used to realize conventional logic architectures, for example NOR gates,[567] the key advantage is speed. With processing-intensive operations and using the energy-delay product as a metric for comparison, a 30–100 times improvement is expected over conventional CMOS, leading to both faster processing and improved energy efficiency. An even greater benefit might be achieved through the use of architectures that perform electronic-photonic processing in the analog domain.

The first demonstration of lasing in the TL occurred in 2004. Since that time, progress has been made on understanding the device physics and on using the TL for discrete optical interconnects. A key initial objective has been examining factors affecting device bandwidth. Edge-emitting TLs with large active regions (200 μm × 1 μm) have been modulated to 22 Gbps and have shown a measured bandwidth of 10.4 GHz.[568] Relative intensity noise (RIN) as low as -151 dB/Hz has also been demonstrated, showing an approximately 28 dB improvement over diode lasers.[569] To improve speed and enable integration, reducing the size of the active region is critical. For that reason, Vertical-Cavity Transistor Lasers (VCTL) have been examined and demonstrated.[570] Initial VCTLs had limited temperature range due to misalignment of the cavity reflectivity and gain peaks. More recently, work has been underway to show that electronic-photonic logic can be made using the transistor laser. The initial target of an integrated TL-based NOR gate has been demonstrated, but significant work is needed to improve performance.[571]

The ultimate performance in both power and speed will be achieved as the device is size is scaled, as projected performance to bandwidths in excess of 100 GHz has a sound rationale but has yet to be realized. The use of vertical-cavity structures to reduce the device footprint has been a first step in the scaling process but further work is needed on microcavity vertical-cavity transistor lasers (VCTLs). Scaling beyond micron-scale devices such as this is possible, but the key will be the design of optical structures such as photonic crystal cavities that will enable small numbers of photons to be captured and directed to act on other TL structures. As scaling advances, further examination of device physics to reduce the effective minority carrier radiative lifetime will be key, along with the examination of effects that might impact the modulation response. Further work is also needed on InP-based devices (1310 and 1550 nm emission) to facilitate the use of silicon waveguides for optical signal routing. Additional questions at the architecture level involve the best way to use the TL in computer systems. What is enabled by having very high speed optical links? What architectures make sense for electronic-photonic NOR gates? Are other approaches to computation enabled, such as analog methods? Initial work to address how the TL might impact computer architecture has been underway in the Li group at the University of Chicago, in collaboration with the University of Illinois at Urbana-Champaign.[572] Further work on how TLs might be used in machine learning applications is also underway by this group (unpublished). Other noteworthy results include demonstration of blue-emitting light-emitting transistors in the GaN/InGaN system.[573]

3.3.9. MAGNETOELECTRIC LOGIC

Several classes of magneto-electric devices, and their possible implementations as CMOS replacements, have now been investigated.[574] There is an extensive range of magneto-electric devices that have been explored as alternatives to CMOS including the magneto-electric magnetic tunnel junction (MEMTJ),[574,575,576,577,578,579,580,581,582,583,584,585,586,587] the composite–input magnetoelectric–based logic technology (CoMET),[588,589] perhaps with spin-orbit coupling,[590] the magneto-electric spin-orbit (MESO)[591,592,593,594] or inverse Rashba-Edelstein magneto-electric neuron (IRMEN)[595,596] devices and the voltage controlled spin switch.[597,598,599] Presently the most promising approach involves magneto-electric transistor (MEFET) schemes.[574,586,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614]

The CoMET style devices are limited by the switching speed of the ferroelectric and domain wall motion with the concept that the input voltage nucleates a domain wall, while the current would drive the domain wall to the output end of the device.[587] The nonvolatile MESO[591,592] and IRMEN[595,596] logic device concepts are laterally scaled spin valves that use a magneto-electric layer for electrical control of exchange bias. The delay time of the CoMET, MESO and IRMEN logic devices will be limited by the switching speed of the ferromagnetic layer (likely nanoseconds[615,616,617]).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

There have been major developments, involving magneto-electric transistor (MEFET) schemes, increasing the range of possible magneto-electric devices, that would serve as "beyond CMOS" 'plug-in' replacement logic.[574,587,602,603,607,612,614] There are also some benchmarking efforts,[574,587,602,603] where there has been an effort to compare the most competitive magneto-electric devices with CMOS. The result is that it is now increasingly clear that magneto-electric field effect transistors are more likely to be competitive or surpass CMOS[574] than the earliest magneto-electric device concepts were based on a magnetic tunnel junction structure.[574,575,576,577,578,579,580,581,582,583,584,585,586,587] The anti-ferromagnet spin-orbit read (AFSOR)[600,606] magneto-electric transistor (MEFET) device structure has interesting advantages:[574] the potential for high and sharp voltage "turn-on"; inherent non-volatility of magnetic state variables; absence of switching currents; large on/off ratios; and multistate logic and memory applications. The design will provide reliable room-temperature operation with large on/off ratios (>10^7) well beyond what can be achieved using magnetic tunnel junctions. Again, the core idea is the use of the boundary polarization of the magneto-electric to spin polarize or partly spin polarize a very thin semiconductor, ideally a 2D material, with very large spin orbit coupling.

Magneto-electric transistor schemes are based on polarization of the semiconductor channel, by the boundary polarization of the magneto-electric gate. The advantage to the magneto-electric field effect transistor is that such schemes avoid the complexity and detrimental switching energetics associated with magneto-electric exchange-coupled ferromagnetic devices. Spintronic devices based solely on the switching of a magneto-electric, will have a switching speed will be limited only by the switching dynamics of that magneto-electric material and above all are voltage controlled spintronic devices. Moreover, these magneto-electric devices promise to provide a unique field effect spin transistor (spin-FET)-based interface for input/output of other novel computational devices. This is spintronics without a ferromagnet, with faster write speeds (<20 ps/full adder), at a lower cost in energy (<20 aJ/full adder), greater temperature stability (operational to 400 K or more[618]), and scalability, requiring far fewer device elements (transistor equivalents) than CMOS. These do differ from the conventional field transistor in that the ME-FET must be both top and bottom gated, so the result is that these are 4, not 3 terminal transistors. Obviously, the semiconductor channel will only work if it is very thin, so the boundary polarization of chromia[618,619,620,621] effectively polarizes the semiconductor channel. The 2D semiconductors of the trichalcogenide class of quasi one-dimensional semiconductors, such as TiS_3, $HfSe_3$, as well as InP, have the potential to be scaled to transistor widths below 10 nm suggesting there is a plausible route forward. If the semiconductor channel retains large spin orbit coupling, then the spin current, mediated by the gate boundary polarization, may be enhanced and, to some extent, topologically protected. The latter implies that each spin current has a preferred direction.

The silicon CMOS majority gate requires 13 components. The ME-spinFET majority gate requires, including clocking, of 6 components. This represents an area improvement of over 50%, assuming similar size transistors.[574] This is equivalent to greater than one process node. If we split the magnetoelectric side of the gate so that the channel can independently be spin polarized up or down, this results in a component reduction from the previous best for the MEFET of six, down to four components, a further 50% reduction over the standard MEFET circuit, and a reduction to less than 30% in area compared to CMOS.[574]

There is a variant where inversion symmetry is not as strictly broken, that leads to a nonvolatile spintronics version of multiplexer logic (MUX).[574] The magneto-electric spin-FET multiplexer also exploits the modulation of the spin-orbit splitting of the electronic bands of the semiconductor channel through a "proximity" magnetic field derived from a voltage-controlled magneto-electric material. Here, by using semiconductor channels with large spin-orbit coupling, we expect to obtain a transverse spin Hall current, as well as a spin current overall. Depending on the magnitude of the effective magnetic field in the narrow channel, we anticipate two different operational regimes. Like the AFSOR magneto-electric spin FET, the magneto-electric spin-FET multiplexer uses spin-orbit coupling in the channel to modulate spin polarization and hence the conductance (by spin) of the device. There is a source-drain voltage and current difference, between the two FM source contacts, due to the spin-Hall effect when spin-orbit coupling is present. This output voltage can be modulated by the gate or gates, which influences the spin-orbit interaction in the channel especially when it is both top and bottom gated especially. The spin-Hall voltage in the device can be increased by using different ferromagnets in the source and drain. To increase the spin fidelity of current injection at the source end, one could add a suitable tunnel junction layer (basically a 1 nm oxide layer) between the magnetic source and the 2D semiconductor channel. This latter modification would result in diminished source-drain currents though. Again, there is a reduction in delay time and energy cost, because these devices are nonvolatile. There is no magnetization reversal of a ferromagnet involved and the implementation of this device concept would require only 5 components for a majority gate compared to the 13 components required of a silicon CMOS majority gate.[574] The related "steering" magneto-electric transistor (MEFET) device, which also exploits the spin orbit coupling, but unlike the two source, one drain MUXer just discussed, here there is one source and two drains.[614]

Magneto-electric coupling can be used to excite parametric resonance of magnetization by an electric field.[622] This has been considered for the development of spin wave devices based on voltage controlled magnetic anisotropy in in ferromagnetic nanowires.[623] This in turn, in effect, becomes a spin wave field effect transistor. The threshold voltage for parametric excitation

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

in this system is found to be well below 1 V, which is attractive for applications in energy-efficient spintronic and magnonic nanodevices such as spin wave logic.

The challenges in pushing forward these technologies extends not only to the fabrication and characterization of a new generation of nonvolatile magneto-electric devices, but also to ascertaining the optimal implementation of CMOS plug in replacement circuits. Questions that need to be resolved include demonstration that the magneto-electric devices can be scaled to less than 10 nm, and this includes finding a suitable 2D channel material that can be polarized by exchange coupling with the boundary polarization of the magneto-electric and yet does not suffer from large scale edge scattering. Experimental demonstration of limits to the switching speeds of any antiferromagnetic magnetic electric still remain absent but scaling studies[624] provide some promise the switching speed could be very fast ($\leq$ 10 ps). That said, the magneto-electric transistor has far fewer challenges to implementation than the magneto-electric magnetic tunnel junction, so, not surprisingly, there is a shifting of development effort toward those and related devices for both memory and logic. The actual demonstration of such devices appears almost "in hand" suggesting that experimental evidence of promise is not that far away. What is significant is that there are demonstrations of working devices based on the readout provided by anomalous Hall measurements, that change with the voltage controlled switching of the magneto-electric chromia (Cr_2O_3).[625,626,627] Furthermore, there are now compelling demonstrations of deterministic switching of the surface polarization of boron doped chromia by voltage alone (i.e., no applied magnetic field is required to break symmetry).[625]

3.3.10. DOMAIN WALL LOGIC

The domain wall-magnetic tunnel junction (DW-MTJ) or three-terminal magnetic tunnel junction (3T-MTJ) operates as an in-memory computing nonvolatile logic device through current-driven manipulation of a single domain wall in a magnetic patterned wire,[628,629,630] with readout performed using a magnetic tunnel junction. It can be considered as an extension of racetrack memory to a single domain wall racetrack for compute-in-memory applications, and it also has applications for computing in high radiation environments such as space. In the last years, there has also been extensive development on using DW-MTJ devices for neuromorphic computing and brain-inspired computing.

The domain wall track is composed of heavy metal/magnet/oxide thin films patterned into a wire shape. Traditionally the magnetic layer is a ferromagnet, but antiferromagnets and ferrimagnets can also be used. Standard materials examples are Ta/CoFeB/MgO. On top of the track is a patterned MTJ hard reference layer with additional related thin films to promote high switching field of the hard layer compared to the magnetic track.

The ends of the ferromagnetic track can be exchange-biased in opposing directions using antiferromagnets, and/or an additional electrode (called an Oersted field line) can be used to ensure a single domain wall is electrically created in the track, and/or an additional magnetic tunnel junction can be placed to nucleate the domain wall electrically, and/or lithographically-defined pinning notches can be fabricated to keep the domain wall in the track.

The simplest form of the device has three terminals: input (IN) and clock (CLK) contacting the ferromagnetic track, and output (OUT) contacting the top of the MTJ. During the write operation, a voltage applied between IN and CLK drives a current and moves the domain wall using either spin transfer torque (STT) and/or spin orbit torque (SOT). During the read operation, a voltage applied between CLK and OUT measures the resistance state of the MTJ relative to the domain wall, which will determine the amount of current to drive subsequent devices. The device can act as an analog universal NAND gate, in addition to other basic logic gates: if IN is connected to the OUT of two previous devices, and only if both are in a low resistance state (logic output 1) will there be sufficient current to depin and move the domain wall to the other side of the MTJ, changing the output of the device from a low resistance state (logic output 1) to high resistance state (logic output 0). The DW-MTJ can also exhibit fanout: a fanout of 2 has been partially demonstrated experimentally[630], and theoretically higher fanout is possible.

A related device is the mLogic four-terminal version, which has an additional non-magnetic, non-conducting spacer on top of the ferromagnetic track that couples the current-manipulated domain wall to a domain wall or nanomagnet in an electrically-separated layer, which then alters the output MTJ resistance.[631,632,633,634,635] The additional terminal comes from two terminals connected to the output MTJ. The four-terminal version provides complete input/output isolation.

The device operation has been modeled in micromagnetics including NAND, shift register, and full adder functions[629,642], and initial benchmarking was performed up to a full adder simulation.[636,637,638] A SPICE model has been developed to enable larger scale circuit simulations.[639] Single device operation and three-device circuit operation has been shown in experimental prototypes,[630,630,640] including inverter, buffer, and concatenation operation. The prototypes showed in experiment one of the first examples of concatenable magnetic logic devices for building larger circuits. The four-terminal mLogic spacer coupling has also been demonstrated.[633]

Larger scale system-level simulations have been performed to determine the benefits and drawbacks of DW-MTJ logic. One circuit-level energy-performance analysis showed that while very low voltage can be used to operate the essentially all-metallic

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

devices, it comes with increased extrinsic domain wall pinning effects and thermal noise.[641] They predicted the energy reduction from increasing the tunnel magnetoresistance (TMR) will saturate when TMR > 100%, but further increasing TMR can mitigate predicted thermal noise limits. Another work simulated a 32-bit adder communicating with registers with all DW-MTJ devices and shows SOT switching can make the technology competitive with a comparable CMOS sub-processor component.[642]

Seminal works came out in the 2021 year that made advances for experimental implementation of DW-MTJ computing. Prototypes were shown that benefit from perpendicular magnetic anisotropy (PMA) and SOT domain wall switching, therefore lowering the switching current density compared to previously demonstrated prototypes by over 10x [643], down to $1-5\times10^{11}\ A/m^2$. One challenge for DW-MTJs has been the precise etching needed to define both the MTJ and the domain wall track, without damaging the delicate thin film layers, which has previously lowered the device TMR after patterning to between 10-40%. Fabrication advancements in this work show that the TMR and resistance-area product of the unpatterned film stack can be maintained after patterning, and showed a TMR of 170%, around the highest that can be obtained to-date for PMA MTJs. The work also showed all-electrical operation of the DW-MTJ, including domain wall nucleation, except for an external DC bias field to overcome coupling of the domain wall track layer to the pinned layers. With these modernized prototypes that combine lower switching current density and high TMR, cycle-to-cycle studies of the domain wall operation were able to be performed, showing cycling variation of 7-10% and that the domain wall can be pushed back and forth in the device for continuous operation without the domain wall exiting the track. A two-device circuit was demonstrated showing an inverter operation.

A series of works also were released this year that apply to both DW-MTJs and domain wall-based majority logic[644,645]. These works also showed full integration of PMA, SOT, and MTJs into the device. Current through an MTJ is used to electrically initialize the domain wall. A hybrid free layer is designed and implemented to separate the delicate domain wall-containing track from the etching ions and therefore preserve the domain wall movement properties. Domain wall velocity is measured and show to be as high as 42 m/s.

These prototype results are a critical advancement since they provide heretofore unknown metrics for DW-MTJs, such as switching current density, switching speed, TMR, resistance-area product, cycle-to-cycle variation, device-to-device variation, scaling behavior, etc. These metrics can now be put into models to predict circuit operation and guide future development: for example, the 10% cycling variation with TMR = 165%, when put into a model of a full adder, predicts 90% accuracy, which could be increased to near-100% accuracy if the variation could be decreased to under 5% [643].

Here we will explain the numbers used in the comparison table for the DW logic device. The cell size of a DW-MTJ based NAND has been calculated to be $18F^2$ where F is the feature size [629], which is a 20x improvement compared to a CMOS-based NAND cell size of $360F^2$. This is because a single DW-MTJ can perform NAND. To accommodate the contacts, MTJ, room for the domain wall to switch, and spacing between devices, the minimum needed device pitch is $6F$, which is 90 nm for F = 15 nm, and the minimum device length is $4.5F$, which is ~70 nm for F = 15 nm. A feature size down to F = 50 nm has been demonstrated in a fully functional device[645]. Using a NAND feature size of $18F^2$, for F = 15 nm this is a projected density of $2.5\times10^{10}\ devices/cm^2$.

Recent results in new materials have shown experimentally that domain walls can obtain speeds up to 250 m/s in Pt-Bi/Co-Gd[646] and up to 5700 m/s in ferrimagnetic CoGd alloys[647]. Assuming the above scaled feature sizes, this would correspond to predicted switching speeds of 3.5-8 GHz, but these materials have not yet been integrated into a full device. The 42 m/s measured in a full device prototype would correspond to 0.6 GHz in a scaled device.

Circuit speed has been shown in simulation to be 33 MHz (30 ns) for a 1-bit full adder and 2 MHz (495 ns) for a 32-bit full adder; circuit energy has been shown in simulation to be 0.5 fJ for a 1-bit full adder and 31 fJ for a 32-bit full adder[642]. These numbers are considering a conventional CoFeB ferromagnetic track, and much higher speeds could be obtained using the new materials described above. Circuit speed and energy have not yet been demonstrated experimentally.

Scaled device switching energy is predicted to be approximately 5×10^{-18} Joules for F = 15 nm [629] . In experimental prototypes, the switching current density was measured ~ $1\times10^{11}\ A/m^2$ for F = 250-450 nm [643]. Assuming the switching current density stays size-independent as feature size scales down, and conservatively assuming a scaled F = 15 nm device would use switching voltage V = 100 mV and switching time t = 5 ns, this would correspond to approximately $E = IVt = 7.5\times10^{-15}$ Joules. Thus the number cited as demonstrated is based off of the measured switching current density in prototypes, but translated to a scaled energy with the described assumptions.

Subthreshold slope is not a good metric for DW logic, since thermionic emission is not the main source of switching error. For high enough switching current above the domain wall depinning energy, the switch can be considered as a collective switch of the magnetic object as it jumps out of its potential well and moves with momentum across the track. When the switching current density is close to or below the depinning energy, stochastic depinning will dominate: this stochasticity has been simulated to be useful for deep neural network applications[662].

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

The DW-MTJ is being studied by many in the community for non-Boolean applications in analog and neuromorphic circuits,[648,649,650,651,652] including spike-timing-dependent-plasticity synapses[653,654] and leaky, integrate, and fire neurons[655,656]. Since the DW-MTJ can act as both an artificial synapse and an artificial neuron, it provides a monolithic platform for neuromorphic computing. By exploiting the dynamical behavior of the domain wall, or an analogous device with skyrmion position being read by a tunnel junction, it can be used to mimic bio-inspired dynamical effects that are essential for learning and operation in the human brain[657,658]. Some example device-inherent behaviors have been shown to have system-level benefits, such as using magnetic stray field interactions between the DW-MTJs to implement lateral inhibition and winner take all[659], manipulating the resetting of the domain wall to mimic the leaking and edginess properties of neurons[660,661], and engineering the DW-MTJ shape to have necessary dynamics for training of deep neural networks[662] or for online learning[658].

While the cycle-to-cycle variation challenges are starting to be measured and addressed, Major challenges still exist in understanding and improving the device-to-device variation of the devices, which arises from variation in the domain wall location and pinning landscape. A discussion of influence of device variability on circuit performance is presented in Xiao et al.[642] Better understanding of experimental viability of the technology is needed, given TMR variability and constraints and scaling-induced variability and errors.[663] Thin film stack growth and engineering needs to be done to adapt the traditional MRAM-like stack for DW-MTJs, including minimizing stray magnetic fields from the pinning layers to remove the need for a DC bias field when the device is operated. Example ideal applications of the device technology are still needed, with some potential areas being radiation-hard environments, lower latency hardware accelerators, and low-area, low-energy needs of edge-computing internet of things devices. Experimental demonstrations of larger operating circuits and systems is still lacking. The experimental evidence of applications to bio-inspired neuromorphic computing is nascent with a few prototypes demonstrated, but many of the above predicted effects are only shown in simulation.

3.3.11. Spin Torque Majority Gate

A majority gate is a logic gate used to simplify circuit complexity to carry out the majority function where the output is a HIGH if and only if more than half of its inputs are HIGH, otherwise the output is a LOW.[664] A majority gate can have any number of inputs. However, the most common ones referred to as 3-input majority gates or MAJ3 are implemented with three inputs and one output. A majority gate can implement an AND by tying one of its inputs to a 0, or an OR by tying the input to a 1. A minority gate can be derived from a majority gate by appropriately using an inverter according to De Morgan's rule. In a full adder, the carry output is the majority function of the three inputs. Implementations of majority gates are currently done using complementary metal oxide semiconductor (CMOS), quantum cellular automata (QCA),[665,666,667,668] spin-wave majority gate (SWMG)[669,670,671] and domain-wall or spin-torque majority gate.[664,672,673,674]

Spin torque majority gate (STMG) is a 3-input majority gate that operates based on domain wall motion driven by either spin transfer torque (STT) or spin orbit torque (SOT) in magnetic tunnel junctions (MTJs) usually with perpendicular magnetic anisotropy (PMA). This has an advantage of energy efficiency, non-volatility, small area, low power, reconfigurability, and radiation hardness compared to other types of majority gate devices. In an STMG, three MTJs set the input states with STT and SOT while the fourth device reads the output state through tunneling magnetoresistance (TMR).[664,672,673,674]

There is extensive work on micromagnetic simulations of STMGs. The most common implementation consists of four discrete PMA ferromagnetic nanopillars of independent fixed layers placed on a common free cross-shaped PMA ferromagnetic layer.[672,673] It is observed that there is a minimum critical current density required to switch the device, which is inversely proportional to the applied pulse width. Attempts have been made to concatenate multiple STMGs and to implement a full adder circuit.[674] The STMG operates at a smaller drive current compared to its CMOS counterpart, but with slower switching.[674]

Majority gates with in-plane magnetic anisotropy have been demonstrated with Permalloy to have errors in the antiparallel configuration due to thermal noise introduced by the clocking field pulses.[675] Hence, implementation with PMA are more stable and energy efficient.[672]

Other types of implementation include a five-input majority gate with lateral spin valve,[676] and a fabricated 3-input majority gate in which the input and output nanomagnetic logic devices are field-coupled together rather than monolithically integrated.[677]

Significant advances have been made recently overcoming the experimental challenges of STMGs[678,679,680]. These advances apply to both domain wall logic and STMGs and are described more fully in the domain wall logic section. These works showed full integration of PMA, SOT, and MTJs into the device. Current through an MTJ or an Oersted field line is used to electrically initialize the domain wall. A hybrid free layer has been designed and implemented to separate the delicate domain wall-containing track from the etching ions and therefore preserve the domain wall movement properties. Domain wall velocity is measured and show to be as high as 42 m/s.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

There is currently a lack of an efficient spin torque inverter (STI). It is also still difficult to cascade multiple STMGs, though some proposals exist. Because the optimal functioning of the STMG occurs below some critical sizes, more work needs to be done experimentally in patterning devices with smaller dimensions.

4. EMERGING DEVICE-ARCHITECTURE INTERACTION

4.1. INTRODUCTION

Many new emerging Beyond-CMOS devices will require co-design between devices and higher levels of computer design (e.g., circuit, architecture and application). These emerging devices are not intended simply as "drop-in" replacements for standard CMOS devices, but will require new types of circuit designs, new functional module architectures, and even new software to best utilize the new devices' capabilities.

In particular, novel design issues spanning the device and architecture levels especially need to be considered when adopting new low-level computing paradigms. In such situations, devices may be organized in radically new ways to carry out computation in a very different manner from what we may consider the most "conventional" computing paradigm, which has relied on von Neumann architectures using standard combinational and sequential irreversible Boolean logic. Examples of unconventional or alternative computing architectures and systems that comprise new computing paradigms are listed in Figure BC4.1 and include the following:

- **Neuro-Inspired Systems** (§4.3.1)—These computing architectures take direct inspiration from the brain to develop more efficient systems. Local learning rules (§4.3.1.1) are a neuro-inspired method of training neural networks that keeps all communication local. Hyperdimensional computing (§4.3.1.3) is a cognitive computing model based on the high-dimensional properties of neural circuits in the brain. Spiking neural networks (§4.3.1.1.1) minimize communication energy by using sparse spike-based communications. Reservoir computing (§4.3.2.4) nonlinearly transforms time-dependent signals into a new, random high-dimensional basis and then a conventional machine learning algorithm is used to make predictions from the new basis set.
- **Dynamical Systems** (§4.3.2)—Analog dynamical systems can be used to solve a variety of problems. Optimization problems can be solved using simulated annealing (§4.3.3.1) or coupled-oscillator-based approaches (§4.3.2.1). Dynamical systems can be used to encode associative memories (§4.2.3.7) or to solve differential equations (§4.3.2.3). Chaotic logic can theoretically enable sub-kT computing (§4.3.2.5).
- **Probabilistic Systems** (§4.3.3)—Devices and circuits that produce truly nondeterministic or random outputs at the hardware level may be useful for accelerating probabilistic algorithms such as Monte Carlo or simulated annealing, for generating secure cryptographic keys, and for other applications.
- **Reversible Systems** (§4.3.4)—Computing systems that approach *logical and physical reversibility* offer the potential to greatly exceed the energy efficiency of all other approaches for general digital computation. While devices for reversible computing may perform fairly conventional functions (such as switching or oscillating), they should be optimized to use quasi-reversible physical processes such as near-adiabatic state transitions, near-ballistic signal propagation, highly elastic interactions, and highly underdamped oscillations. For maximal efficiency, circuits and architectures must approach reversibility at the logical as well as physical level.[681,682] Careful fine-tuning and optimization of analog circuit characteristics (e.g., resonator quality factors, elasticity of ballistic interactions) remains a difficult and crucially important engineering challenge that must be met to fully realize the promise of the reversible computing paradigm. In the meantime, potentially commercially viable near-term applications of reversible computing are beginning to emerge for specialized cryogenic applications.[683,684]
- **Digital in-Memory Systems** (§4.3.5)—Boolean logic can be performed in memory arrays and digital logic is being embedded closer and closer to memory to reduce data movement costs.
- **Quantum computing** (CEQIP chapter)—Quantum computing[685] offers the potential to carry out exponentially more efficient algorithms for a variety of specialized problem classes.[686] Devices for quantum computing are very different from conventional devices, and fine-tuning device characteristics to avoid decoherence while organizing them effectively into scalable architectures has so far proved to be a formidable engineering challenge.[687] The 2022 IRDS Cryogenic Electronics & Quantum Information Processing (CEQIP) roadmap chapter addresses quantum computing. For more details refer to that roadmap chapter.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

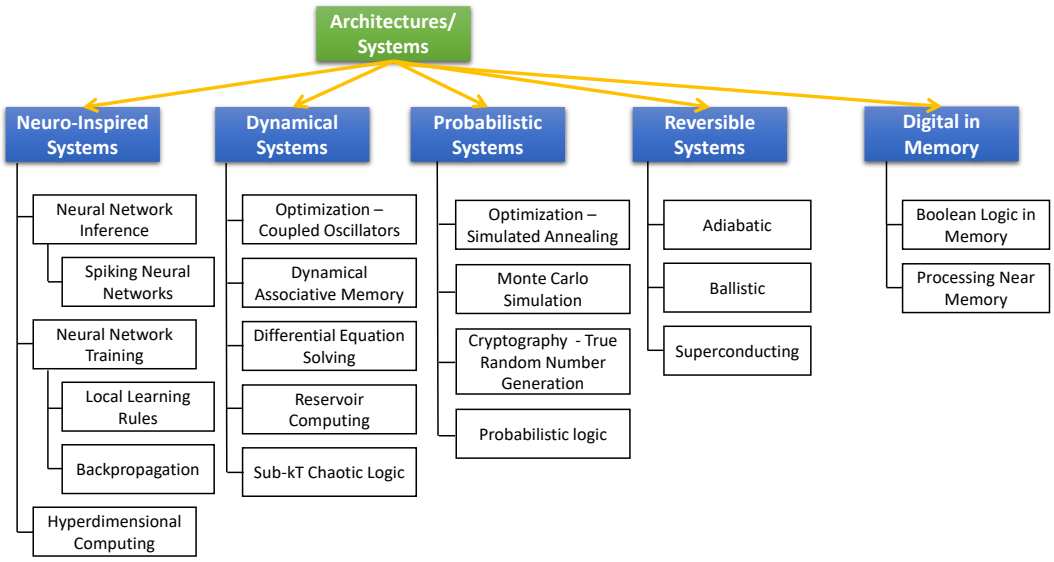


Figure BC4.1. Architectures/Systems for Novel Computing Paradigms Discussed in This Chapter Requiring Codesign with Emerging Devices

Computational Paradigms	Basic Dichotomies			
	Analog vs. Digital	Deterministic vs. Stochastic	Irreversible vs. Reversible	Classical vs. Quantum
CONVENTIONAL DIGITAL COMPUTING	D	D	I	C
Classical Analog & Neural Computing	A	*	I	C
Probabilistic Digital Computing	D	S	I	C
Reversible Digital Computing	D	D	R	C
Stochastic Reversible Digital	D	S	R	C
Reversible Analog Computing	A	*	R	C
Quantum Computing Systems (the various thermodynamically irreversible implementations now in development)	*	*	I	Q
Future, thermodynamically-reversible implementations of quantum computing?	*	*	R	Q

* = Variants of either type are included in the given row.

Figure BC4.2. Conventional vs. Alternative Computing Paradigms

Note on above figure: The conventional computing paradigm is explicitly designed to be digital, deterministic, irreversible, and classical. Typical classical analog computing schemes (the focus of this section), including the neural approaches, relax the digital requirement while leaving determinism optional, and they are typically physically irreversible. In classic probabilistic computing, we abandon the requirement for determinism, while preserving the irreversible, digital nature of the computation. And typical classical reversible computing techniques maintain the digital and usually deterministic nature of computation while attempting to minimize irreversibility. Quantum computing machines, as they are conceived and engineered today, are highly thermodynamically irreversible at the system level, and many useful quantum algorithms are nondeterministic. Although achieving digital stabilization of quantum information via fault-tolerant error correction is a major goal of the field, it remains very challenging. Quantum machines that are also thermodynamically reversible at the system level are rarely considered but may be conceivable.

These systems are composed of new computational kernels and enabling devices as shown in Figure BC4.3.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

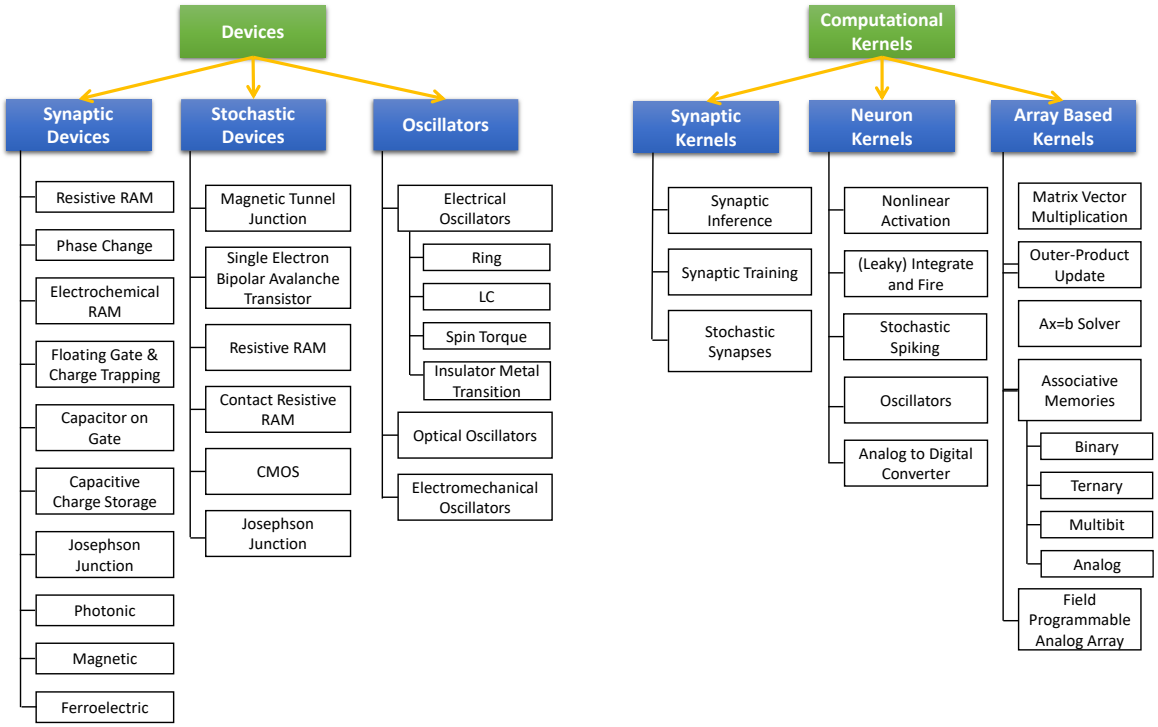


Figure BC4.3. Emerging Devices and Computational Kernels Requiring Codesign between the Device Layer and Higher Layers of the Technology Stack

Neuro-inspired, dynamical, and stochastic systems all typically leverage analog computing. Analog computing attempts to "let physics do the computation" by using physical processes directly (as opposed to, by going through the traditional digital abstraction barrier) to compute complex functions. Historically, this required inefficient analog circuitry for all elements, and expensive analog to digital conversion, resulting in limited applications, specifically those requiring analog signal processing. Recently, new analog devices have enabled a new generation of efficient analog architectures. This is especially true for hybrid analog and digital systems where efficient designs may exploit analog preprocessing and computation prior to digitization. Analog preprocessing can reduce the required A/D precision and therefore reduce the system energy consumption by orders of magnitude.[688] Additionally, new architectures can be used for ultra-low-power co-processors for conventional CMOS designs. A key challenge is that analog signals are typically low precision, with energy and latencies increasing exponentially with higher bit precision. Fortunately, many machine learning and other applications are being developed that can tolerate such lower precision computation.

Furthermore, all these architectures except for reversible computing are special purpose accelerators designed to accelerate a specific function. Reversible computing has the unique property that it can do general digital computing at lower energy than is possible with conventional methods.

The reader should note that the material in this section is not intended to comprise an exhaustive list of *all* possible new computing paradigms, new devices, new circuits, or new architectures. It is only intended to serve as a representative sample of several new general computing paradigms and specific technology concepts.

In this section we first describe some enabling computational kernels (§4.2), then some systems and architectures exhibiting novel computing paradigms (§4.3), and finally we discuss some of the enabling devices (§4.4).

4.2. Computational Kernels

Emerging computing systems are composed of enabling kernels that perform some novel computational function more efficiently than a corresponding digital implementation. At the device level these kernels can be divided into "synaptic kernels" and "neuron kernels." The synaptic and neuron kernels can then be composed into larger array-based kernels that perform larger scale computations (§4.2.3).

Copyright © 2022 IEEE. All rights reserved.

4.2.1. SYNAPTIC KERNELS

Synaptic kernels typically represent the memory in a novel system and must have a programmable internal state and use that state to modulate an input, most commonly by multiplying the internal state with the input. By merging memory with a simple computation, synaptic devices allow emerging systems to overcome the von-Neumann bottleneck by merging processing and memory. Synaptic devices are also typically passive devices that modulate an incoming signal.

4.2.1.1. SYNAPTIC INFERENCE

The most common kernel is synaptic inference where a stored state is used to modulate an input. The simplest form of this is to use an analog memory device as a programmable resistor and to do multiplication using Ohm's law $V = I \times R$, or equivalently $I = V \times G$. Many variations on this are possible. Nonlinear devices can be used to implement non-linear functions and multiple devices can be combined to provide pattern matching or associative memory functionality (§4.2.3.7). Enabling devices are surveyed in §4.4.1.

4.2.1.2. SYNAPTIC TRAINING

In addition to using a stored value to modulate an input, the stored value needs to be programmable. If the values can be programmed in parallel, the devices can also implement training functionality like outer product updates (§4.2.3.4), three factor local learning rules (§4.3.1.2.2) and spike timing dependent plasticity.

4.2.1.3. STOCHASTIC SYNAPSES

For probabilistic systems (§4.3.3), having a good source of randomness is critical. Often a stochastic synapse will have a programmable mean state that varies stochastically around that mean state. For many applications it is also critical to be able to tune the probability distribution or standard deviation of the variation around the mean. This functionality can be implemented in a single device or in a combination of devices (one to tune the mean state and one to tune the standard deviation). Possible stochastic devices are surveyed in §4.4.3.

4.2.2. NEURON KERNELS

Neuron kernels perform more complex computations on an input, may integrate an input over time, and typically do not need to have a programmable internal state. This allows for a complex expensive function to be computed using a single or few devices in an energy efficient manner, obviating the need for time- and energy-consuming ADC/DAC in the circuits. Neuron kernels are also often active devices that restore a signal and inject energy into a system. If a passive device is used in large systems, it is often paired with an active driver circuit to power downstream synaptic kernels. Just as in digital logic, energy periodically needs to be injected into a system or the signal will decay. In many applications neurons provide a natural place to do so. For systems that accelerate learning and update synaptic weights, neuron blocks should be capable of programming the synaptic blocks.

The benchmarks used to evaluate electronic neurons generally measure the energy per operation, the fabrication cost, or the chip area of the integrated functional block, and the fidelity to the desired neuron function (e.g., integrate and fire). High reliability and low variation of devices are two key factors for the viability of a neuron technology. Device failure will require a lot more circuitry for error detection and correction.[689] Large variation increases the difficulty for designing peripheral circuits and degrades the adaptability of the block.

4.2.2.1. NONLINEAR ACTIVATION (THRESHOLD, RELU, SIGMOID, TANH)

One of the most basic and common neuron functionalities is to provide a nonlinear activation function. In many neural networks, after a linear multiply accumulate is performed on a set of inputs, or in a hidden layer, a non-linear activation or transfer function is applied. The presence of this function prevents the network from mathematically collapsing into a single linear equation, which helps improve the computational capabilities as the number of layers increases.

This can be as simple as a binary threshold that outputs a 1 if the input crosses a predefined threshold or a 0 otherwise. A binary threshold can be implemented both with an analog comparator or a new device. In more conventional neural networks, functions like rectified linear units, sigmoid functions and Tanh functions are common. There are many device candidates that can approximate these functions. If a device transfer function does not exactly model the intended numerical function, the rest of the system can often be adapted to compensate (i.e., a neural network can be trained to use the transfer function available)

4.2.2.2. (LEAKY) INTEGRATE AND FIRE

For systems that operate using temporal dynamics (leaky) integrate and fire neurons are common. An input signal is integrated until a threshold is reached and then a binary spike is created. Mott memristors,[690] phase change based memristive switches,[691] and chalcogenide threshold switches[692] have all been reported to be capable of performing temporal voltage signal integration in which the effects of non-simultaneous unitary post-synaptic potentials add in time.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

4.2.2.3. STOCHASTIC SPIKING

Probabilistic systems can have randomness be injected at the synapse or neuron level. At the neuron level, the randomness can be represented as random threshold shifts or as stochastic firing of a neuron after a threshold is reached. Possible stochastic devices are surveyed in §4.4.3.

4.2.2.4. OSCILLATORS

Many dynamical systems (§4.3.2) are based on coupled oscillators. While oscillators are not a conventional "neuron" they provide many of the same functionalities for dynamical systems. Possible device implementations are surveyed in §4.4.2.

4.2.2.5. ANALOG-TO-DIGITAL CONVERTER

While conventional analog-to-digital converters are the least efficient approach to a neuron, we explicitly mention them as they are still used by many proposed architectures. Because the bulk of a system's computation is accelerated by the synaptic kernels, it is often possible to average out the cost of an analog-to-digital converter and still provide system level advantages. Using digital signals between analog arrays/synapses also allows for the use of flexible digital routing to enable more flexible computing accelerators.

4.2.3. ARRAY-BASED KERNELS

Analog crossbars or memory arrays can perform low-precision matrix operations in parallel, by processing analog data directly at each memory element. Thus, in 1990, Carver Mead projected that custom analog matrix vector multiplications would be thousands of times more energy efficient than custom digital computation.[693] Because a digital memory must individually access each memory cell and move the data to a separate computation unit, digital systems consume more energy and incur longer latencies. Computing on larger crossbars/matrices allows for any analog overhead to be averaged out over many matrix elements. Any two- or three-terminal device that features a modifiable internal physical state variable (which might be, for example, a variable resistance, a variable capacitance, a stored charge, or a stored magnetic field) that modifies the device's behavior can be used as a building block for analog operations. Several different types of array architectures are summarized in the following sections.

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

MVM and VMM are key computational kernels underlying many different algorithms. Although these terms are sometimes used interchangeably, we refer to MVM as the operation $\mathbf{A}x$, where $\mathbf{A}$ is a weight matrix and x is the input vector. VMM uses the transpose of the same matrix: $\mathbf{A}^{\mathrm{T}}x$. In an analog accelerator, MVM is used during the forward propagation step (inference) while VMM is needed during the backpropagation steps (training).

There are several approaches to accelerating these kernels. Any programmable resistor such as a two-terminal resistive memory or a three-terminal floating gate cell can be used.[694] Alternatively, a capacitive MVM can be designed by adding charge from capacitive memory elements.[695]

For many algorithms such as neural network inference (of an already-trained network), accelerating MVM accelerates the bulk of the computation,[696] allowing for large system level efficiency gains. An $N \times N$ crossbar accelerates $\mathrm{O}(N^2)$ operations, leaving only $O(N)$ inputs and outputs that need to be processed and communicated. This allows each unit of communication and processing cost (such as analog-to-digital conversion) to be amortized over N memory elements. This changes the tradeoff for some neural-network algorithms as larger crossbars are more efficient than smaller ones, provided that the algorithm can make use of them. There are ultimately several factors that limit how large a crossbar can be made, including: (1) accumulation of device errors on a column[711,697], (2) accumulation of voltage drops induced by array parasitic resistance,[707] and (3) ability to isolate individual devices during programming.[708]

Analog MVMs have been used for experimental demonstration of threshold logic,[698] compressed sensing initial filtering,[699] robotic navigation and control,[700] adaptive filtering,[701] Fourier transforms[702] and more. Additionally, Analog MVM techniques have been used for ultra-low power classification and neural networks.[703]

4.2.3.2. RESISTIVE MVM AND VMM

Resistive MVMs are based on using Ohm's law, $V = I \times R$, to perform multiplication, and Kirchhoff's current law to perform addition by summing currents. Programmable resistors are used to program the weights. Arranging the memory elements in an array allows for the entire operation to be performed in a single parallel step, giving a fundamental $O(N)$ energy and latency advantage over a standard digital memory that, at best, must access a memory array one row at a time.[704] An MVM and the transpose VMM can be performed on the same memory array, by changing whether the rows of an array are driven and the columns are read, or vice versa.[705,706]

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

The key metrics for a resistive MVM are 1) the energy per multiply-and-accumulate (MAC) operation, 2) latency per MVM, 3) crossbar and supporting circuitry area per matrix element, 4) crossbar dimensions, 5) input/output bit precision for digitally driven MVMs, and 6) the standard deviation of the noise or error per conductance when programmed as a percentage of the absolute conductance range. Relating to #6, the absolute error is more important than the error relative to a given conductance level. This is because in an MVM, it is the absolute conductance errors that are summed by Kirchoff's law along a column of the array, not the relative errors. For example, a 10% error in a 1 μS device contributes less error to the MVM than a 1% error in a 100 μS device.

If high-resistance memory elements ($R_{on} = 100$ MΩ) with good analog properties are developed, one ReRAM based crossbar design projects that each multiply and accumulate operation will require 12 fJ when using 8-bit A/Ds and only 0.4 fJ when using 2-bit A/Ds.[706] The latency for a 1024×1024 MVM will only be 384 ns or 11 ns for 8-bit or 2-bit A/Ds, respectively. This is over 100× better than an optimized SRAM-based accelerator, which would require 2,700 fJ and 4,000 ns for 8 bits. The area per weight for the 8-bit A/D ReRAM accelerator is 0.05 μm^2, 16× better than the 0.8 μm^2 needed for an SRAM accelerator. The energy and latency are dominated by the A/D circuitry and not by the crossbar itself, with the A/D converters and digital circuitry occupying 10× the area of the ReRAM array itself.

To allow for large arrays and minimize parasitic resistance drops, high resistance (~100 MΩ) memory elements are needed. The higher the resistance, the larger the array possible, and the more any A/D costs and system level communication costs are amortized out.[707] However, such high resistances would prolong and potentially complicate the process of programming each conductance value accurately to encode already-trained neural network weights or matrix element values.

A key design choice is the bit precision of the inputs and outputs to the crossbar. The fewer bits are needed by an algorithm, the more efficient the crossbar is. If analog or binary inputs/outputs can be used, the A/D costs can also be avoided. The inputs to the crossbar can be encoded in voltage, time, or digitally. Voltage encoding applies different voltages to represent different analog input values. This requires circuitry to create different input voltages, and it requires that the memory elements have a linear I-V relationship, greatly complicating the use of nonlinear access devices.[708] Encoding inputs in variable length pulses requires longer reads and an integrator to sum the resulting current. Digital encoding applies each bit of the input sequentially and then combines the result digitally.[695] For digital encoding, the usefulness of the lower-order bits in the input is limited by the noise/errors on the highest-order bit. To save on ADC costs, each bit position can also be combined in analog using successive integration and rescaling[709].

The precision with which each resistor needs to be programmed depends on the application. Neural networks trained for simple image classification tasks such as MNIST digit recognition can often tolerate as much as a 15% error or noise on their conductance values before losing accuracy. The requirement is much more stringent for the ImageNet dataset, which is much more complex (100× larger images, 100× more classes than MNIST) and is considered more representative of a real-world application. For example, the ResNet50 neural network, commonly used for benchmarking digital accelerators[710], tolerates only about 1% weight error and requires 8-bit activations and ADCs[711].

For many applications, it is useful not to treat the required resistor programming precision as a single number, but to instead consider the desired precision as a function of the programmed resistance (or conductance). This is because the actual weight values programmed into the resistive array may not be uniformly distributed, and thus certain conductance ranges may be much more heavily utilized than others. The precision of the more heavily used weight values generally have a larger effect on the end-to-end algorithm accuracy. A good example is neural network inference. It is well known that for many neural network models, the weights have a skewed distribution where values near zero are the most abundant.[712] Therefore, for this application, the resistances which encode low weight values should be made as precise as possible, while precision for the high weight values is less important because these tend to be outliers.[711] One class of devices that has high precision at low conductance is flash transistors operated in the subthreshold regime.[713]

To extend the precision of computation beyond the limits of reliable programming, a technique called bit slicing can be used.[714] With bit slicing, a matrix of wide operands is striped across multiple crossbars, enabling the crossbars to collectively perform computation on arbitrarily wide operands at the cost of additional digital circuitry to reduce partial results from multiple bit slices. When bit slicing is combined with digital input encoding, each bit of the input must be applied to each bit slice, analogous to multibit scalar multiplication. Leveraging bit slicing, accelerators for a wide range of applications have been proposed, including combinatorial optimization,[714] neural network inference,[715,716] graph analytics,[717] and scientific computing.[718]

4.2.3.3. CAPACITIVE MVM

Some analog accelerators propose crossbar-based parallel MVMs using volatile capacitive memories, where a weight value (binary or multi-bit) is stored as charge on a capacitor. Analog multiplication of the capacitive weight and a binary voltage input can be carried out using a multiplying digital-to-analog converter (MDAC) circuit; the resulting currents can be summed on a

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

wire as in the resistive MVM case.[719] Another implementation uses a binary input to control the redistribution of charge between two capacitors. This charge movement induces a voltage change on a metal lines, and the voltage changes induced by all the capacitive cells on a column are summed to form a dot product.[695]

Some other techniques combine capacitive and resistive MVM. For example, an analog weight can be stored as charge on a capacitor, which is also connected to the gate of a transistor. The amount of charge then controls the source-drain resistance of the transistor.[720,721] One challenge of these schemes is the fact that capacitor charge is not truly non-volatile and can decay within milliseconds to seconds.

4.2.3.4. Outer Product Update (OPU)

Analog resistive memory crossbars can also perform a parallel write or an outer product, rank 1 update where all the weights are incremented by the outer product of a vector applied to the rows and a different vector applied to the columns. This is a key kernel for many learning algorithms such as backpropagation[722] and sparse coding.[704,723] Row inputs are encoded in time and column inputs are encoded in either time[723,724,723] or voltage.[706]

When an MVM, VMM and OPU are combined on the same crossbar, extremely efficient learning accelerators can be designed,[706,725,726] with the potential to be 100–1000× more energy efficient and faster than an optimized digital ReRAM or SRAM based accelerator.[706]

The same figures of merit and design considerations for a VMM apply to the OPU. Additionally, the 1) write noise and 2) asymmetric write nonlinearity are important for determining how well a learning algorithm will perform. The 3) ability to withstand failures, and 4) endurance are also important for training systems. To have an efficient learning accelerator, parallel blind updates are needed where weights are updated without knowing the previous value and without verifying that the correct value is written. To obtain ideal accuracies, a low write noise is needed, less than 0.4% of the weight range. Even more important is having low asymmetries in the write process. The change in conductance for a positive pulse should be the same as that for a negative pulse for all starting states.[727] Often the conductance will saturate near a maximum where a positive pulse will not change the conductance, while a single negative pulse will cause a large decrease in conductance. This significantly lowers accuracy as it only takes a single negative write pulse to cancel many positive pulses.

Several devices have been examined for neural network training, including phase change memory,[725] resistive memory[706,728] and novel lithium-based devices.[729,730] Currently no devices meet all the ideal requirements for training (high resistance >10 MΩ, low write noise <0.4%, low write asymmetry[727]). Nevertheless, algorithmic approaches such as periodic carry,[731] Local Gains,[732] Tiki-Taka,[733] or the inclusion of semi-volatile capacitor-on-gate devices[734] can be used to help compensate and achieve ideal accuracies. Several co-design tools have been developed to model the impact of device level properties on algorithmic performance,[725,735,736] which have allowed for the algorithmic development needed. Additionally, lower resistance devices can be used to give smaller near-term gains in performance.

The need for high on-state resistance and good analog characteristics means that filamentary resistive memories may not work as well as non-filamentary devices. A resistance higher than a quantum of conductance, 13 kΩ, requires current to tunnel through a barrier. This presents a fundamental problem for a filamentary device: a single atom can halve that tunneling barrier, resulting in huge variability and poor analog characteristics. OPUs have recently been physically demonstrated using memory cells with >10 MΩ resistance: the cell consists of an ionic floating-gate memory element in series with a conductive-bridge RAM select device.[737]

One requirement of the parallel OPU is that the update has to be rank 1. This is satisfied by stochastic gradient descent updates but not batch or minibatch gradient descent which are also commonly used. One way to make a batch update computable with the parallel OPU is to approximate it as a series of M rank-1 updates where M is much smaller than the batch size. These approximations can be built using streaming principal component analysis, though there is a tradeoff between efficiency and the approximation quality, which determines the final accuracy.[738]

4.2.3.5. Large-Scale Field Programmable Analog Arrays (FPAAs)

A field programmable analog array has configurable analog components, digital components, configurable interconnects between those components and off-chip communications.[739,740,741] FPAAs allow users to build analog applications without having expertise in IC design. FPAA I/O lines can transmit or receive analog signals, digital signals and create direct connection lines typical of analog circuits. The routing between analog and digital blocks can occur between the blocks of devices, with converters between these blocks, or more finely connected heterogeneous analog and digital component populations. The components are often organized into regions called computational analog blocks (CABs). CAB components vary considerably between implementations but often include nFET and pFET transistors, transconductance amplifiers [TA or operational transconductance amplifier (OTA)], other amplifiers, passives (e.g., capacitors), as well as more complicated elements (e.g., multipliers). The most advanced FPAAs to date utilize Floating-Gate (FG) devices, dramatically improving the analog parameter storage and therefore

Copyright © 2022 IEEE. All rights reserved.

the resulting computational capability.[742] FPAAs include aspects of digital computation, such as FPGA blocks or shift registers or microprocessors, to complete the full end-to-end configurable system.[742]

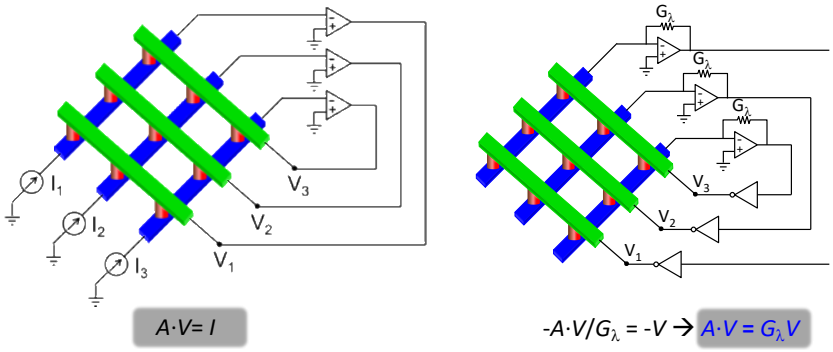
The fundamental breakthrough was recognizing that a switch matrix of single floating gate elements could be used for analog computation. The routing crossbar networks were, in fact, crossbar networks that could support VMM and other computations.[743] Routing was no longer dead weight, as perceived for FPGA architectures. The floating gate cells could also allow for mismatch calibration at the mismatch source.[744,745] The density for VMM in FPAA architectures is nearly the level of custom IC design. These analog computations can be made robust to temperature fluctuations.[746] These techniques have been utilized by a number of students in university courses.[747,748] FPAA based VMMs can be scaled to small geometry (e.g., 40 nm and smaller) and operated at RF frequencies.[749,750] FPAAs have been used for command-word recognition in less than 23 µW with standard digital interfaces.[741] The full classification results in less than 1 µJ per classification (or inference), which has 1000× improvement over similar digital neuromorphic solutions requiring roughly 1 mJ or higher for just an inference.[751]

4.2.3.6. RESISTIVE MEMORY CROSSBAR SOLVER

Resistive memory crossbars can be used to solve matrix problems, such as the linear system of equations $Ax = b$, where x is the unknown vector and b is the known vector, represented by output voltage and input current, respectively, in Figure BC4.4(a).[752,753] Each resistive memory element is a programmable resistor that represents an element in the coefficient matrix A. (Alternatively, any programmable analog element can also be used.) The equation $Ax = b$ can be mapped to Ohm's law, $\sum G_{ij}(V_j - V_i) = I_i$. The V_i are set to zero by the virtual ground of the op-amps. Currents, I_i, are applied to the crossbar and the resulting voltages, V_j, are measured. The op-amps provide feedback allowing the V_j to be determined.

Similarly, eigenvectors of a matrix A can be calculated according to the circuit in Figure BC4.4(b). Here, the maximum eigenvalue is mapped in the feedback conductance G_λ, while the voltage yields the corresponding eigenvector x. To solve problems with positive/negative coefficients in A, two crossbars can be used in the circuits of Figure BC5.2. In all cases, crossbar solvers yield their solution in one computational step, without any iteration, and the solution is generally obtained in less than 1 µs, depending on the poles of the analogue feedback circuit.[754] The same scheme can be extended to one-shot learning by linear/logistic regression.[755]

Figure BC4.4 A Resistive Memory Crossbar $Ax = b$ Solver is Illustrated



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

The biggest challenge in taking advantage of analog solvers for HPC is that analog operations only offer low precision, ~8 bits fixed point, while HPC applications often demand 32 or more bits of floating-point precision. This can be potentially overcome by hybrid analog/digital systems where the computationally intense parts of a calculation can be done in analog, while the required precision can be achieved by refining the solution in digital using a method with lower computational complexity.[756] This allows for some digital computation, while still getting a reduction in the overall computational complexity. The precision can potentially be improved by using iterative refinement or by using the crossbar to initialize a digital solver. The analog solution can also be used as a preconditioner within a Krylov method like CG or GMRES. Large matrices can be broken down into smaller blocks compatible with the accelerator and scaling can be used to compensate for finite on/off ranges. Work is still needed to show how noisy crossbar solvers can be used with ill-conditioned matrices. In general, iterative refinement will only converge if the noise is less than $1/\kappa$ where κ is the condition number of a matrix.

4.2.3.7. ASSOCIATIVE MEMORIES

Associative memories (AMs), which efficiently "associate" input queries with appropriate data words/locations in the memory, are powerful in-memory-computing cores. AMs search for data words/locations in the memory in a highly parallel fashion

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

according to (i) an input query and (ii) the desired association (matching) function.[757] AMs are widely used for network routers[758], and recently have been exploited to accelerate machine learning models such as memory augmented neural networks (MANNs)[759], clustering[758], and bioinformatics tasks[758].

CMOS AM or CAM designs suffer from several fundamental challenges:

(i) The density of CMOS AMs, especially ternary CAM (TCAMs)—where "don't care" states as well as "1s" and "0s" need to be stored/searched—is about half of the SRAM density.
(ii) The standby power of CMOS AMs is high.
(iii) CMOS AMs are typically limited to implementing Hamming (HM) distance-based matching.

Recent research on AMs based on beyond-CMOS devices suggests new directions for addressing the above challenges. For example, a ferroelectric FET (FeFET) based multi-bit CAM (MCAM)[760] implements nearest-neighbor (NN) search according to a sigmoid-like (SG) distance function and offers significant performance/energy advantages versus competing designs for memory augmented neural networks (MANNs). This, as well as other recent work suggest that harnessing the benefits of AM for data intensive applications requires cross-layer efforts spanning from devices and circuits to architectures and applications.

Different non-volatile memory (NVM) technologies (such as RRAM, FeFET memory and Flash) can be exploited to implement various AMs (e.g., exact and best match searches, multi-level and analog data representations, and different distance functions). Like with the much-studied resistive crossbar array architecture, the large performance benefit for using CAMs for in-memory computation comes from the vast reduction of data movement for target applications with large numbers of compare or lookup operations. We classify AMs according to (i) the data representation (binary, ternary, multi-bit and analog) in a cell, and (ii) the matching type (Table BC4.1). Representative AM designs based on different technologies are discussed below.

FeFET AM cell designs—were originally proposed to accomplish exact match (EX), best match (BE), and threshold match (TH) TCAM based on HM distance. Different match types employ different sensing circuits[759]. The same cell design can also function as a EX/BE/TH-MCAM when FeFETs are programmed to store multiple bits[760], which improves storage density. The FeFET BE/TH-MCAM can implement SG and squared Euclidean (SE) distance functions[761], which is useful for machine learning applications. This design can also function as an EX-ACAM where the threshold voltage (V_{th}) values in FeFETs define either upper or lower bounds, and an analog input matches stored cell data if it is within the bounds defined by the FeFETs[762]. ACAMs can encode more information per cell than MCAMs but may suffer more from noise and variation effects. Other FeFET AM cell designs have also been considered for reducing search latency, lowering search energy, simplifying peripherals, and/or improving scalability[763,764].

RRAM and PCM AM Designs—Resistive memory (RRAM), has been widely used for crossbar type IMC core designs. RRAM has also been used to realize TCAMs, MCAMs, and ACAMs. There are a variety of RRAM TCAM designs (2T2R, 5T2R, 3T1R, and 2.5T1R)[763] that support EX match and BE/TH match based on the HM distance. A 6T2R EX-ACAM[765] was proposed and can directly process analog inputs. This design has high static power consumption and does not support BE/TH search. Phase-change memory (PCM) has similar characteristics to RRAM. A PCM EX-TCAM design was proposed and evaluated in Ref [766]. This design can perform BE/TH searches based on HM distance with appropriate sensing circuits.

Flash AM Designs—Floating-gate MOSFET (flash) is a mature NVM technology that is suitable for AM design.[767] and[768] have proposed EX-TCAM/MCAM designs based on 3D NAND Flash. The same 2-transistor design the FeFET approach was implemented in Ref. [761] with flash technology where it can perform EX match and BE/TH match based on SG and SE distance functions. The main drawbacks of flash technology for AM design are high write voltages, low endurance, and scalability.

MTJ AM Designs—Magnetic tunnel junction (MTJ) memory (MRAM) offers fast write operations and high endurance. EX-TCAMs based on voltage-dividers (9T-2MTJ) and latches (15T-4MTJ) were proposed in Ref [769] and Ref [770], respectively. MTJs cannot offer high R_{OFF}/R_{ON} ratios and usually require several transistors to cope with this issue. Hence it may not be a good candidate for MCAM or ACAM.

While new technologies may enable more efficient CAMs, and enable new/existing distance functions to be computed directly within the memory, said solutions may also *fundamentally change algorithms* developed and deployed by the ML community. Data precision might change due to the multi- bit capacity of NVMs, different distance metrics might be used for an original ML algorithm (e.g., cosine instead of HM distance), etc. To ensure technology-enabled CAM solutions have value, evaluations must extend to the application level (e.g., to ensure iso-accuracy for ML applications). Recent work has proposed using associative memories as in-memory computation blocks for applications such as associative processing,[771,772,773] approximate computing,[774] spiking NNs,[775] string matching,[776] and regular expression matching finite state machines.[777] Given the broad range of application spaces – from machine learning models like MANNs which reduce training costs, to deep random forests (i.e., interpretable AI)[778], to bioinformatics[758], to security[779], etc., further study is warranted.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Table BC4.1 Classification of associative memory based on the representation and matching function. Match-line (ML_i) functions are defined based on input query (q_j) and memory content (C_{ij})

	Binary CAM (BCAM)	Ternary CAM (TCAM)	Multi-bit CAM (MCAM)	Analog CAM (ACAM)
Data type of q_j and C_{ij}	0 1	0 1 X	S0 S1 S2 S3	
Exact match (EX-)	ML_i = 1 if C_{ij} == q_j for all j / 0 otherwise			ML_i = 1 if $q_j \cup C_{ij}$ for all j / 0 otherwise
Best match (BE-)	ML_i = 1 if $\sum_j d(C_{ij}, q_j)$ is the smallest among all MLs, where $d(C_{ij}, q_j)$ is the distance between C_{ij} and q_j			
Threshold match (TH-)	ML_i = 1 if $\sum_j d(C_{ij}, q_j) \leq Thr$, where $d(C_{ij}, q_j)$ is the distance between C_{ij} and q_j and *Thr* is the threshold			

4.3. Emerging Architectures And Systems

4.3.1. Neural-Inspired Computing

There are several characteristics of how the brain computes that have been proposed for efficient computing technologies. Architecturally, there are two attractive approaches: processing in memory, which is analogous to the analog computation that occurs at synapses within the brain, and event-based communication, which is analogous to neuronal spiking. Both approaches potentially yield considerable energy savings, and the brain clearly benefits from both. As discussed in §4.2.3, analog crossbars can be used as building blocks for conventional neural networks to give significant improvements in energy, latency and area over digital accelerators. For accelerators specialized to a particular algorithm, various analog neurons can be used to process the crossbar outputs and avoid the need for and high cost of analog-to-digital conversion.

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

4.3.1.1. Neural Network Inference

In-memory analog computing has been extensively explored as energy-efficient implementations of deep learning algorithms, or deep neural networks (DNNs). To be useful for a particular task, a DNN must first be trained so that the values of its internal parameters or weights are optimized to yield high accuracy. After training, the DNN can be deployed for the intended application, where it may encounter unseen data that deviate significantly from the training data. During this phase, called inference, the weights are left fixed. Emerging accelerators for neural network inference are a more near-term goal compared to training accelerators. During inference, there is no need to apply frequent updates to the memory device conductances, or to propagate information backward through the network, both of which greatly simplify the architecture. Since MVMs are the dominant computational kernel used during inference,[696] large system-level energy savings are possible by efficiently computing the MVMs in the analog domain, as described in Section 4.2.3.1. Furthermore, executing MVMs inside memory arrays largely eliminates the very significant energy cost in conventional digital systems of moving weight matrix data between memory and the processor.[780,1087]

The basic architecture of an analog in-memory inference accelerator consists of resistive arrays that accelerate MVMs, combined with digital elements to perform a small amount of computation and handle the data routing. Since weight values are directly encoded in physical device resistance values, the computation is inherently weight stationary: weight values stay inside their

Copyright © 2022 IEEE. All rights reserved.

MVM arrays, but the digital activation values are moved between processing elements as the computation proceeds from layer to layer.[781] The MVM array receives digital inputs at a specified precision (e.g., 8 bits) and the analog MVM result is digitized by an ADC. These outputs are then further processed digitally (e.g., activation function, pooling) and digitally routed to the MVM array of the next neural network layer. For all but the simplest multi-layer perception (MLP) networks, digital memory arrays are needed to buffer these intermediate values between layers.

Inference accelerators have been demonstrated fully in hardware using flash memory[782] and ReRAM.[783] So far, these demonstrations have shown high accuracy on simple benchmark tasks such as MNIST. There are many more proposed inference accelerator architectures, whose performance has been evaluated in simulation on larger-scale machine learning problems such as ImageNet.[715,714,713,697,784,716,785,786,787] Physically demonstrating these larger-scale systems is challenging, due to the very large number of resistive devices needed, and the sensitivity of these more difficult tasks to analog errors. Nonetheless, these systems can be realistically modeled by combining measured device- or array-level properties with system-level simulation.[735,786] A common architectural choice in these proposed large-scale architectures is to group a set of analog and digital circuit elements into a modular unit called a *tile*.[714,715,784,713] Many tiles are connected together, e.g., using a concentrated mesh interconnection topology, to implement the many layers of a deep neural network. Since the tiles have the same structure, this type of system can scale to larger networks by adding more tiles.

The main challenge with analog in-memory inference accelerators is to ensure high accuracy along with the high energy efficiency. Analog computing differs fundamentally from digital in that (non-catastrophic) errors at the device and circuit level can directly affect the inference prediction. In a DNN, these errors can propagate and grow from one layer to the next. At the same time, nonlinear operations such as activation functions, ADC quantization, and the argmax operation (at the network output) can help suppress the propagation of these errors. Given these complex mechanics, it is difficult to know how much precision is needed at the device level, or even the MVM level. Therefore, analog systems should be designed with the *end-to-end* neural network accuracy as the figure of merit, rather than simply the MVM precision. A common, alternative design choice is the full-precision guarantee: the system is parameterized (via the array size, ADC resolution, number of bit slices, etc.) such that each analog MVM result, after being digitized, can match the precision of an *N*-bit digital processor.[715,714,697] This choice often leads to an MVM-level precision that is far more than sufficient for high end-to-end inference accuracy, and thus can cause energy and area inefficiency.[711]

The sources of error that affect inference accuracy in an analog system are the same ones described in Section 4.2.3.1 that degrade the accuracy of an MVM. These include: (1) errors in device resistance programming due to process variations or device/circuit noise in the write process, (2) noise in the device resistance during an MVM, (3) resolution and noise in the ADC, (4) parasitic voltage drops due to array interconnect resistance, (5) drift over time in the programmed device resistances, and (6) noise, process variation, and transients in the peripheral analog circuits, and (7) temperature dependence.

For neural network inference specifically, the effect of these errors (especially #1-4 above) can depend very strongly on the distribution of the data values in the neural network. A near-universal property of neural networks is the fact that the distribution of weight values is not uniform, but rather heavily concentrated near zero. This property is partly why pruning (eliminating low-valued weights) has been so successful in compressing neural networks, making them easier to process by digital inference accelerators.[712] For analog inference accelerators, one implication is that the low-valued weights need to be more precise than the much less common high-valued weights. As a result, the method that is used to map weight values to the device resistances can have a very large effect on the accuracy. If weights with low absolute value correspond to low conductance, higher accuracy to be achieved by devices that have *low error at low conductance*.[711] This also leads to *high On/Off ratio* as a desirable property for neural network inference. Since low-valued weights are abundant, an improvement in the On/Off ratio can greatly reduce the array currents, reduce the parasitic voltage drops,[707] and enable larger arrays with higher energy efficiency. This is an important example of how the device and system architecture (e.g., mapping method) are designed with the algorithm in mind to ensure robustness to errors, as well as high efficiency.

There are two other prominently used methods to reach high accuracy even with very limited precision in the memory device technology. One of these is bit slicing,[715] described in Section 4.2.3.2, where the *N* bits of weight precision are "sliced" across multiple devices, and the MVM results from these slices are aggregated digitally. Bit slicing can be helpful, or even necessary, when device precision is limited. However, there are two important caveats: (1) representing weight values with high fidelity does not imply MVMs are computed with high fidelity, because errors accumulate across many devices within a slice, and (2) bit slicing incurs significant energy and area overheads, due to requiring a separate MVM for each slice.[711]

Neural network retraining is also commonly used to compensate for device errors. The most common style of compensation is to inject noise during the forward and/or backward propagation phase of training, by an amount that emulates the amount of error or noise in the device conductance.[786,788,789,790,791] The drawbacks of these methods are that: (1) the retrained network that compensates for device errors may not have the same accuracy as the original network and (2) the device compensation method

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

may be difficult to integrate with state-of-the-art training workflows. In a similar vein, methods have been developed to optimize the mapping of weights to conductances to mitigate the effects of parasitic voltage drops.[785,789,792,793]

Conductance drift is a concern for inference accelerators and arises from physical changes in the memory devices that occur over time. This causes a drift in the weight values; if this drift is predictable, it can be compensated.[794] However, the physical processes underlying drift are often thermally driven, and result not only in weight drift but also an increase in the weight variance over time.[713] Eventually, the accuracy may be degraded to the point where the device resistances must be refreshed from a digital reference: the exact refresh frequency depends on the device and the application.

4.3.1.1.1. Spiking-based Neural Networks (SNNs)

Like the brain that couples both the processing in memory and spike-based communication for maximal space and energy efficiency, a good spiking based neural network needs to couple both. Ultimately, the spiking function by neurons has two features that must be captured by a proposed device or circuit: its non-linearity and its efficient long-distance communication. First, it must accomplish the analog-to-spiking conversion, which in its simplest form is a compact 1-bit analog-to-digital conversion, but ideally would also enable holding some additional state or history from the analog inputs. There have been many proposed devices for spiking neurons including neuristors,[795] spin-torque-based devices,[796] stochastic phase change neurons,[797] superconducting neurons,[798] and others. Second, future spiking systems must be able to communicate this information efficiently to downstream neurons.

Currently, CMOS systems rely on event-driven communication of packets that contain some source or destination address relevant to routing the spike to appropriate destinations. Thus, CMOS systems do benefit from the relative rarity of the communication (only transmit when an event occurs), and are already achieving considerable savings from that, but do not benefit significantly from the theoretical 1-bit precision of the spike as a multibit address is needed. In this respect, more efficient mechanisms for direct point to point communication, such as superconducting systems, 3D-nanowires, or perhaps optical interconnects are needed. The challenge in these systems is how to achieve the necessary level of fan-in/fan-out (i.e., number of synapses per neuron) between non-local regions and how to deliver that information in a suitable form for processing in the analog memory circuits used as synapses.

4.3.1.2. Neural Network Training

In-memory accelerators for neural network training is inherently more difficult than inference accelerators, due to the need to support frequent updates to the conductances of the memory devices. These conductance updates must not only be frequent, but also must be both blind and sufficiently accurate. Having to verify the conductance after each programming step is costly and would eliminate the energy benefits of an in-memory training accelerator.

For deep neural networks, the most ubiquitously used training algorithm is gradient-descent-based learning with backpropagation. In backpropagation, a loss function is computed at the output of the network and the weights in each layer are tuned to optimize the value of the global loss function. In spiking neural networks, local learning rules have also been used.

4.3.1.2.1. Backpropagation

In backpropagation, training data is first fed in the forward direction through a neural network and a loss function is computed at the output. The gradient of the loss function with respect to each layer (the layer-wise error) is then computed for each layer. These errors are computed from the last layer backward to the first, using successive VMMs between the layer-wise error and the weight matrix. Within a layer, the optimal weight update is an outer product between two vectors: the layer-wise error and the layer's activations. Compared to an analog inference accelerator, supporting the backpropagation of errors within a training accelerator requires additional digital logic and memory, since activations computed during the forward pass must be saved to compute the correct weight update.[694] For high efficiency, the same array used for MVM should be used to compute the VMM (transpose operation) by allowing the inputs to drive either the rows or the columns of the array.[706]

The in-memory OPU computational primitive is very useful for accelerating backpropagation, since the update needed for each weight in the matrix can be computed in analog by the array itself: see Section 4.2.3.4. This reduces the digital processing overhead and is faster and more efficient than programming row by row.[704] However, one limitation of the method is that OPU updates assume a batch size of 1 during training. Larger batch sizes are often desired, which can reduce wear on the devices, enable greater pipelining, and improve training convergence.[694] Thus, some works have proposed computing the weight updates within a batch in digital hardware, then transferring these updates to the memory array at the end of a batch. This mixed-signal technique can also be used to account for the low bit precision of a device update, by digitally storing the residual of the update and accumulating over this residual in the next batch.[799] Another method is to construct an optimal rank-1 approximation of a batch update that allows the efficiency of the in-memory OPU to be leveraged while also retaining the convergence and endurance benefits of batched training.[738]

Copyright © 2022 IEEE. All rights reserved.

The greatest challenge for analog in-memory training accelerators is ensuring accurate, blind conductance updates. This accuracy is affected by: (1) cycle-to-cycle write noise or limited programming resolution, (2) linearity in the update response, e.g., the conductance change in response to a pulse should be the same regardless of the starting conductance, (3) symmetry in the update response, e.g., the conductance change in response to a pulse should be the same regardless of the update polarity, and (4) device-to-device variations in the update response.[800] In general, it has been found that an asymmetric nonlinearity is much more harmful than a symmetric nonlinearity, because it tends to cause the weight to settle toward zero after a sequence of updates.[727,801] In addition to meeting these requirements, the resistive device should further have high endurance, low write latency, and low write energy. Low write currents are also desired to reduce parasitic voltage drops during write, which can further distort the desired weight update.[694]

At the device level, a linear update response requires a device to be smoothly tunable between its minimum and maximum conductance. Early ReRAM devices proposed for this application were highly nonlinear, responding with an abrupt conductance change to a single positive or negative update pulse.[706] This can partially be attributed to the filamentary nature of many ReRAM devices that can make their conductance respond highly nonlinearly to small changes in the arrangement of atoms or ions in the filament. One way to alleviate this is to rely not on the creation and annihilation of a filament for conductance change, but smoothly modulate the conductivity of the bulk channel material. This has been explored using ionic electronics, or electrochemical RAM (ECRAM), where mobile ions are driven into and out of a channel material to continuously modulate the conductivity. Exploiting this bulk phenomenon has led to devices with highly linear, symmetric, and deterministic update characteristics.[730,729,737,802,803] Another way to obtain linear and symmetric behavior is to embed the conductance states in the device geometry, rather than modulating electrical doping or material properties. Linearity has been showing in simulation using spintronic devices where a domain wall (separating two oppositely magnetized regions) can be pinned at one of many lithographically defined notches along a long track. The domain wall can be moved between notches via a spin-transfer-torque or spin-orbit-torque current, and its position can be read out as a conductance using a magnetic tunnel junction.[804,805,806]

There are also system architecture solutions to non-ideal device properties. The periodic carry method splits the bits of a weight across multiple memory devices, similar to bit slicing for MVM, in order to accommodate the limited precision or noise of a conductance update. Updates are accumulated on the least significant bits and these are periodically "carried" to the devices encoding the higher bits. By limiting the range of the least significant bit, periodic carry can also mitigate nonlinearity.[731] The Tiki-Taka method splits a weight matrix into two, where one matrix is always composed of devices that are in their linear regime, and their states are periodically transferred to another, less frequently updated matrix.[733] Another method is to split the bits of significance between two different types of devices: a nonlinear non-volatile memory device (e.g., PCM) for the higher bits and a linear volatile memory device (e.g., capacitor) for the lower bits. Updates are carried out on the volatile device and periodically carried to the non-volatile device.[734] For asymmetry, one proposed solution is to use the conductance difference between two devices to encode a weight, and always apply updates of the same polarity to one of the two devices. Periodically, when a device conductance saturates, the pair is re-programmed to represent the same weight with lower conductances.[725,807] All of these techniques rely on the fact that expensive operations such as carries are infrequent and their energy is amortized over many updates, most of which can be carried out using more efficient parallel OPU methods.

4.3.1.2.2. Local Learning Rules

Spiking neural networks can be viewed as a type of recurrent neural network, where activities are binary and recurrence is both due to explicit connections and internal dynamics (referred to as implicit connections)[808,809]. This analogy enables the transfer of learning algorithms based on gradient descent to local synaptic plasticity. The synaptic plasticity dynamics that result from these derivations are "three-factor rules." Three factor rules are popular among computational neuroscientists for reward-based learning[810]. Because gradient-based rules vastly outperform other traditional learning rules borrowed from neuroscience on industry relevant benchmarks, our focus here is on gradient-based learning.

As in deep learning, estimating the gradients with respect to parameters embedded in deeper layers requires solving a spatial and temporal credit assignment problem. Through approximations that rely on ignoring gradients caused by explicit connections, it is possible to solve the temporal credit assignment problem with a very small penalty compared to the exact gradients[808,811].

Solving the spatial credit assignment problem remains challenging, however. This is because all interlayer connections are explicit and ignoring explicit connections is equivalent to ignoring the effect of deeper layers. Solving the spatial credit assignment problem is under heavy investigation, ranging from approximating gradient back-propagation[812], predictive coding[813,814], contrastive learning to auxiliary loss functions[815]. Understanding how such approximations in the context of novel in-memory architectures is key to local learning using both conventional and spiking neural networks.

The machine learning community has been dominated so far by GPUs high bandwidth access to global memory, making gradient backpropagation like rules compatible. However, the demands on scale and power are increasing pushing to distribute computations across multiple nodes and processors with more limited access to global memory[816]. As a result, one can expect an

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

increase in output on the topics of local learning from the machine learning community. Furthermore, local learn rules that keep all data local naturally map to in-memory computing technologies with primarily local interconnects.

Regardless of the credit assignment techniques used, local learning is subject to constraints that do not verify the assumptions of deep learning and statistical learning theory: data is potentially independent, and identically distributed (i.i.d.), causing catastrophic forgetting; data consists of a unique stream ("batch size is one"), leading to long training times and the requirement of impractically small learning rates; and large amounts of data are required. These problems are compounded by the realities of limited precision and hardware non-idealities, especially in the case of emerging nanotechnologies. A result of these challenges is that hybrid learning schemes that combine offline pre-training and online learning need to be explored.

4.3.1.3. HYPERDIMENSIONAL COMPUTING

Hyperdimensional (HD) computing is a cognitive computing model based on the high-dimensional properties of neural circuits in the brain.[817] Information is encoded into high-dimensional distributed representations in the form of high-dimensional vectors or *hypervectors.*[817,818,819] A hypervector distributes information uniformly across all of its dimensions, resulting in a distributed or *holographic* representation.[820,821] This contrasts with conventional *positional representations* in which different digits or bits can convey vastly different amounts of information depending on their position. Consequently, incurring bit errors in particular positions can result in *catastrophic failure*. In contrast, the distributive nature of hypervectors combined with high dimensionality provides a robustness to bit errors: errors that occur in any dimension result in the same information loss, and this information loss is small due to the large number of dimensions. Thus, HD computers exhibit *graceful degradation* as hardware components fail, analogous to in the brain when neurons die[821,822].

HD computing relies on the algebraic properties of their key operations to incorporate the advantages of structured symbolic and vector distributed representations. It incorporates learning capability along with typical memory functions of storing/loading information.[817] It mimics important functionalities of the human memory model with vector operations, which are computationally tractable and mathematically rigorous in describing human cognition.[823] HD computing operates over a well-defined and hardware-friendly set of mathematics: *Binding* is well suited for associating two hypervectors and used for variable-value association. *Bundling* is a memorization function that keeps the information of input data into a bundled vector. *Permutation* is an operation to represent sequences by creating a near-orthogonal but reversible hypervector. *Reasoning* is done by measuring the similarity of hypervectors.

HD computing has shown several advantages as the next generation of cognitive machines. First, its training capability in one or few shots, where partial data are learned from one or few examples as opposed to many iterations.[822,824,825] Second, HD operations are highly parallel and do not require frequent weight updates, thus empowering online learning on *processing in-memory.*[824,826] Third, HD computing has natural robustness to noise and bit errors provided by the high-dimensional distributed representations. This enables greater tolerance for device variability and unreliable emerging technologies, such as non-volatile memory[827,828,829] or nanoscale devices with low signal-to-noise ratios.[826,830,831] In fact, HD computing enables further size and energy scaling by abandoning the deterministic requirement set by the variability and reliability of the composing devices. Finally, HD computing has been shown great potential to enable lightweight privacy and security.[832,833,834,835]

4.3.2. COMPUTING WITH DYNAMICAL SYSTEMS

In computing with dynamical systems, the built-in dynamical behavior of a physical system exhibiting continuous degrees of freedom is used to compute. The entire computational process can be analog, with only the results being digitized. The following subsections give a few examples of different types of dynamical systems-based approaches. Also exemplifying this category are the hardware-based reservoir computing or liquid-state machines, which were already discussed above in §4.2.2.4.

Although some of the below methods target NP-hard problems and provide approximate solutions, it's important to note that, to date, no general physical computing method (including analog and quantum approaches) has yet been clearly demonstrated to be capable of exactly solving NP-hard problems without requiring exponential physical resources (energy and/or time) to be invested in the physical process performing the computation. The prevailing belief among computational complexity theorists[836] is that solving NP-hard problems efficiently would require uncovering new physics (i.e., beyond standard quantum mechanics). (Note that although quantum computers can efficiently find prime factorizations, this is not known to be an NP-hard problem.)

4.3.2.1. COUPLED-OSCILLATOR BASED OPTIMIZATION

Coupled-oscillator machines are another class of analog accelerators for combinatorial optimization that share a similar architecture: they are composed of a network of decentralized nonlinear oscillators, and the programmable strength of the coupling between them encodes the specific problem to be solved. These networks have been proposed and demonstrated with electrical,[837,838,839,840,841] optical,[842,843,844] and electromechanical[845] oscillators. These systems are often called "Ising machines", because they map readily to the Ising graph optimization problem, with each oscillator representing one bistable spin. Any

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

combinatorial optimization problem can be converted into an equivalent Ising problem that is programmed onto and solved by the machine. The fixed points of the equations of motion of such networks correspond to the solutions of an NP-hard combinatorial optimization problem and can model other NP-hard problems as well.[841]

Networks of coupled oscillators have been shown to embed the energy (cost function) of the Ising problem in their physical dynamics and relax to configurations that minimize this energy. Networks of coupled parametric or second-harmonic injection locked nonlinear oscillators can be designed such that the network physical dynamics implement the method of Lagrange multipliers on the Ising problem.[846] They can thus rapidly sample the local minima of a problem,[838] and can potentially also be used to arrive at solutions close to the global minimum.[837,840] The sampling speed is determined fundamentally by the oscillator frequency and the quality factor of the oscillator network. The accuracy of the different architectures has been benchmarked using large instances of the Ising or Max-Cut problems generated by the operations research community. Solutions have been found that match those obtained by state-of-the-art digital algorithms.[839,842] Energy and delay benchmarking remain a future step.

Of the proposed coupled oscillator optimization schemes, the systems that use electrical (LC) or electronic (ring oscillator) oscillators are the most compatible with CMOS technology. Since the Ising problem is specified by a connectivity matrix, the oscillators can be densely connected using a resistive crossbar.[837,840] In these networks, each oscillator takes on a discrete binary nature by synchronizing to its associated second-harmonic pump oscillator through a circuit nonlinearity. In an N-spin Ising problem, the bistable voltage signals of the N fundamental oscillators play the role of the N Ising spins while the N second-harmonic pump voltage signals play the role of the corresponding Lagrange multipliers that enforce bistable constraints on the spins.[847] Novel nonlinear dynamical system approaches that go beyond Lagrange multiplier optimization and exceed state-of-the-art performance on certain benchmark problems have been proposed[848] and their physical implementation is an area of ongoing research. The nonlinearities in all these approaches can be easily implemented by semiconductor components such as diodes, MOS capacitors, or amplifiers. In fact, phase-based Boolean computing using such nonlinear oscillator circuits was first proposed decades ago.[849,850,851,852]

The physical limits of coupled oscillator systems are governed by the achievable level of weight precision and circuit delay. The resistive connections must be linear, programmable to high precision, and have minimal drift. The precision and retention of the resistive connections limit the accuracy to which a problem can be programmed onto the hardware, and the precision requirements for an adequate representation increase with problem size. For very large problems, device or process limitations will impose an upper bound on the quality of the solution; the target error bound depends on the application, but it must be superior to bounds that are guaranteed by digital approximation algorithms. Since a coupled oscillator network relies on synchronization between the oscillators, signal delays can also impact performance, especially in high-frequency circuits needed for rapid optimization. To this end, architectures have been proposed that separate the oscillators in time rather than space using optics,[843] but this comes at the loss of continuous-time communication among the oscillators, leading to slower convergence.

Related constraint satisfaction systems have been built.[853] An interesting approach based on memory co-processors was introduced as Memcomputing.[854] Useful insights can also be obtained by looking into dynamical systems like iterated maps,[855] and 0-1 continuous reformulations of discrete optimization problems.[856]

4.3.2.2. Dynamical Associative Memories

Hopfield networks are attractor networks proposed for associative memories[857] where the fixed points (or stable states) of the system correspond to memories, and the dynamics of the network is such that the system settles to the fixed point, which is closest to the initial state the system starts from. These networks can be implemented with coupled oscillators.[858,859,860]

The associative memory application is widely used in the tasks of voice and image recognition, which can be performed in a cellular neural network architecture.[861,862] In this work, five decimal digits, '1'–'5', are associated with the other five digits, '6'–'0'. Hebbian learning is used for storing patterns.[863] Patterns with noisy input pixels can still be recalled. The delay per cellular neural network operation is dependent on the input pattern, input noise, and thermal noise.

A key challenge for oscillator-based associative memories is storage capacity. A fully connected net with N units can only store around $0.15N$ memories while requiring N^2 weights, resulting in a poor memory density.[864]

Another application for Hopfield networks is discussed later, in §4.3.3.1.

4.3.2.2.1. Cellular Neural Networks

The cellular neural network[865] (CeNN) is a non-Boolean computing architecture that contains an array of computing cells that are connected to nearby cells. Since interconnects are major limitations in modern VLSI systems, CeNN systems take advantage of the local communication and encounter fewer constraints imposed by interconnects. The CeNN is a brain-inspired computing architecture that relies on neurons to integrate the incoming currents. The accumulated and activated output signal drives nearby neurons through weighted synapses. CeNNs can be used to create associative memories for voice and image recognition.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

CMOS based CeNNs can be implemented by analog circuits using operational amplifiers and operational transconductance amplifiers (OTAs) as neurons and synapses, respectively.[866,867] Some recent work has also investigated CeNN using beyond-CMOS charge-based devices, such as TFETs, to potentially improve energy efficiency[868,869] thanks to their steep subthreshold slope and low operating voltage.

Using novel devices such as all-spin logic (ASL),[870] charge-coupled spin logic (CSL),[871] and domain wall logic (mLogic)[872] whose dynamics match the dynamical state of cells in CeNN can be far more efficient than op-amp and OTA based CeNNs. The use of these different devices has been benchmarked.[873] It was shown that the digital CeNNs are quite power hungry and slow. This is because multiple cycles are required to read out the weights from the register and perform the summation in the adder, which is energy and time consuming. In general, analog CeNNs implemented by TFETs dissipate less energy thanks to their steep subthreshold slope and lower supply voltage. In contrast to Boolean circuits, spintronic devices are more competitive. This is because a single magnet can mimic the functionality of a neuron, and these spintronic devices operate at a low supply voltage. The domain wall device provides the best performance, in terms of energy-delay product, thanks to its low critical current requirement.

4.3.2.3. Differential Equation Modeling

The circuit dynamics of CMOS systems can serve as computational primitives for solving differential equations. The premise of this style of computing is to map continuous-valued and continuous-time solutions to ordinary differential equations onto the currents and voltages of analog electronic circuits, the dynamics of which are also described by ODEs. This basic idea of analog computing for solving ODEs originated in mid-20th Century analog computers and was extended to combine analog and digital representations to offer a wider tradeoff of dynamic range and precision.[874] The idea of analog computing has since been revisited in CMOS; the implementations in modern device technologies demonstrate reconfigurability, scalability, and energy efficiency.[875,876,877,878]

The fundamental operation of solving ODEs is integration, which in analog computers is carried out via capacitors that integrate current over time. To form arbitrarily complex systems of ODEs, the analog computer can include constant coefficient multipliers, variable-variable multipliers, and summers to construct polynomial functions, and non-polynomial functions can also be made via hybrid techniques that involve lookup tables.[879, 880] An important extension to ODEs in the form of stochastic differential equations can be tackled by introducing controllable noise sources.[881]

Using analog systems to solve ODEs confers several benefits. 1) Deeply embedded cyber-physical systems can take advantage of analog computation to eliminate the need to digitize analog sensory inputs and actuator outputs. 2) Analog integration avoids some difficulties in algorithmic numerical integration on digital architectures such as round-off error due to digital representations and truncation error due to numerical approximations.[882] 3) With appropriate design of the analog circuits, analog computing can handle numerical difficulties due to stiffness in the ODEs being solved.

Potential downsides include: 1) Difficulty in realizing more advanced ODE solving algorithms such as high-order quadrature. 2) Difficulty in offering both high dynamic range and high precision due to physical costs in area and power consumption. 3) Difficulty in implementing in device technologies that are not optimized for mixed-signal design.

Solving partial differential equations (PDEs) is even more important to modern scientific computation than solving ODEs. In PDEs, problem variables span space in addition to time. This class of problems introduces a wide range of options in terms of whether continuous space, time, and variable values are discretized. A guiding principle for research in this area is to survey the taxonomy of PDE types and demonstrate analog computer implementations of key benchmark solution methods and problems, such as the finite-difference time-domain approach to Maxwell's equations.[883,884] Given the sheer diversity in PDE problem types and numerical methods supported by digital high-performance computers, a fruitful approach has been to retain the space and time discretization steps in numerical methods for PDEs, and only map the core ODE solving kernels to analog co-processors.

An even broader use case for analog computing is to use the circuit dynamics to solve algebraic and satisfiability problems entirely outside of differential equations.[885] The key idea here is to view the differential equations as a continuous interpolation of problem variables as they get updated over the course of discrete algorithm steps. For example, steepest gradient descent algorithms for linear optimization can be rephrased as ODEs such that analog computers can tackle linear algebra.[886,887] Likewise, Newton methods can be rephrased such that analog computers can provide initial seeds for solving nonlinear systems of equations.[888]

4.3.2.4. Hardware-Based Reservoir Computing and Liquid State Machines

Software-based reservoir computing (RC) arose from the surprising insight that a single randomly connected recurrent neural network can be tailored to solve various tasks solely by means of adjusting its output weights.[889] Time dependent signals are nonlinearly transformed into a new, random high-dimensional basis and then a conventional machine learning algorithm is used to make predictions from the new basis set. Hardware-based RC uses physical systems whose temporal evolution—much as in dynamical systems computing (see §4.2.3 below)—form the basis of computation.[890] Such hardware need not implement tunable

Copyright © 2022 IEEE. All rights reserved.

weights that add considerable complexity to hardware accelerators of neural networks. These systems are well-suited to time-series data processing and forecasting tasks such as learning to emulate a chaotic attractor,[890] though they are generally limited to supervised learning. Naturally, a wide variety of systems have been pressed into service as reservoirs, including photonic,[891,892] mechanical,[893] electronic,[894,895,896] spintronic,[897] and more recently superconducting systems[898] among many others. An active area of research is extending RC into the quantum regime whose large Hilbert space dimensions may afford additional reservoir performance.[899,900] Spiking neural networks (see §4.2.2.3 above) can also be used for RC, though this is typically referred to as liquid state machine (LSM) computing.[901] The possibility also exists to create "deep" hardware RC architectures,[902] and to otherwise combine RCs with other novel computing paradigms in order to combine their strengths.

Other approaches such as recurrent neural networks (and structured variants such as long short-term memory) are also suitable for processing temporal data. However, these approaches are difficult to train and may struggle with complex nonlinear dynamics across many timescales

4.3.2.5. Sub-kT Chaotic Logic and Chaos Computing

Shannon's noisy channel coding theorem[903] shows that one can reliably communicate information on a channel subject to noise (at a sufficiently low bit-rate) even when the transmitted signal power is below the noise floor (i.e., at a signal-to-noise ratio of less than 1). Moreover, any computational process can be viewed as just a special case of a communication channel, namely, one that simply happens to transform the encoded data in transit—since the derivation of Shannon's theorem relies solely on counting distinguishable signals, and nothing about how the signals are being counted in Shannon's argument precludes the encoded data from being transformed as it passes through the channel. This observation suggests that performing reliable computation utilizing signal energies (that is, energies associated with the information-bearing variability in the dynamical degrees of freedom in the system) that are at average levels $\ll kT$ (i.e., well below the thermal noise floor) should theoretically also be possible—although the output bit rate (per unit signal bandwidth) will scale down with the average signal energy.

In 2016, Frank and DeBenedictis investigated a theoretical approach for implementing digital computation using chaotic dynamical systems,[904,905,906] which provided evidence that the above theoretical observation is correct. In that approach, the long-term average value of a chaotically evolving dynamical degree of freedom encodes a digital bit. The interactions between degrees of freedom are tailored such that the bit-values represented by different degrees of freedom correspond to the results that would be computed in an ordinary Boolean circuit. This method can also be considered to be related to analog energy-minimization-based approaches (§4.3.3.1). However, this method does not require cooling the system to low noise temperatures for annealing, as is frequently done in energy-minimization approaches. Instead, the dynamical network uses a variation on reversible computing principles (§4.3.4) to adiabatically cause the system to transition between different warm, chaotic "strange attractors" that represent different computational states; this transformation can take place reversibly, without energy loss. The dynamical energy of the signal variables is itself conserved within the (Hamiltonian) dynamical system, and so the total energy dissipated per result computed can approach zero in this model as the rate of transformation decreases.

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

An earlier, more extensively developed proposal that is similar to the chaotic logic concept is called *chaos computing*.[907]

4.3.3. Probabilistic Systems

Traditionally, conventional computational processes are designed to be deterministic, with computational results determined by the machine's initial state and inputs. Nevertheless, computations that are *intentionally* designed to behave randomly or stochastically, even at the level of individual bit-operations, are of interest and can have many useful applications in *naturally* probabilistic tasks. Applications of this approach include powerful statistical sampling algorithms used in Monte Carlo simulation (Markov Chain Monte Carlo, direct sampling, simulated annealing, parallel Temperin, etc.), machine learning and artificial intelligence algorithms (Bayesian networks, Boltzmann machines, deep Boltzmann machines), randomized algorithms[908,909] as studied in computational complexity theory, and cryptographically secure random number generation for generating secure private keys. Noise in biological neurons is beneficial for information processing in nonlinear systems and is essential for computation and learning in cortical microcircuits.[910,911,912,913]

Obtaining randomness in traditional CMOS is difficult and typically relies on costly pseudo-random number generators. This requires a large circuit block and significant computational effort to obtain high quality random numbers. Several new devices have been proposed to obtain true randomness as discussed in §4.4.3. These allow for a random bit to be generated with a single device, often leveraging the intrinsic noise out of natural phenomena. Chaotic devices can be used to turn poor quality randomness into high quality random numbers[914].

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

One example of probabilistic computation is achieved by p-bits. The main function of a *p*-bit is to provide *tunable* randomness of a digitized voltage at its output terminal controlled by an analog input terminal.[932,933] The main advantage of a device-based p-bit comes from its compact, low-power implementation of a complex functionality (tunable randomness) compared to digital implementations.[940] In the context of machine learning, this functionality is an approximate hardware representation of a *binary stochastic neuron*, allowing a natural mapping of powerful algorithms developed for such stochastic neural networks.

4.3.3.1. STOCHASTIC/CHAOTIC OPTIMIZATION—SIMULATED ANNEALING

Many of the optimization problems that are found in modern operations research—such as routing, scheduling, and other types of resource allocation—are intractably hard. Consider this example: finding an optimal route among three cities can be done using the digits on two hands, but with 15 cities, we are left with more than 40 billion routes to choose from. As the size of the problem grows, the resources needed to solve the problem increases exponentially. Finding even approximate solutions to large combinatorial optimization problems are prohibitively resource-intensive with the best supercomputers we have. Exact solutions to these problems are known in computational complexity theory to be NP-hard (non-deterministic polynomial time hard), meaning that it is too hard for any computer, analog or digital, to solve exactly, in general, and specifically at large scale. However, there are many ways to compute approximate solutions, meaning finding a good solution but not necessarily the best one. In recognition of this, there have been many analog hardware approaches that exploit the inherent computational ability and parallelism in physical processes to solve these hard optimization problems.

An example solution uses energy minimization using multiple runs on a Hopfield network (§4.3.2.2).[915] A Hopfield network is a popular neural network with its output being calculated via a simple decision system (e.g., thresholding of input), which is then weighted and fed back to its input. The feedback weights define an energy landscape based on values emerging from the output. If a Hopfield network is initialized to a particular value, the network will follow a trajectory that will take it to a minimum, or local minimum, of the energy landscape.

As an example, consider using a Hopfield network to solve the Traveling Salesman Problem. The Traveling Salesman Problem is to find the best route for a salesman that needs to visit a series of cities, each pair separated by some distance. The salesman seeks the shortest route that visits each city exactly once and then returns to the starting city. In an analog system the weights are set to encode the intercity distances. The system drives the outputs to a starting point for the salesman's route. The system will settle into a candidate salesman's route in an amount of time equal to a few time constants of the feedback loop.

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

Such a system leverages multiple new circuit blocks including both a crossbar to encode the n intercity distances (that can be built using memristors), an analog neuron to run the Hopfield network, and either chaos or a noise generator to get randomness. As the Traveling Salesman problem is NP-hard, no solution method can solve it exactly at scale. Nevertheless, the analog solution seems to be at least comparable in efficiency with some software algorithms. For instance, a memristor based Hopfield network has been built.[916]

Another important application for combinatorial optimization, specifically graph coloring, was described[917] and developed theoretically with support from experimental demonstrations using relaxation oscillators based on phase-change IMT materials. An architecture based on non-repeating phase relations[918] between fabricated CMOS oscillators tries to emulate stochastic local search (SLS) for constraint satisfaction problems.

Recently, solving hard combinatorial optimization problems with powerful sampling algorithms such as simulated annealing and parallel tempering has received significant attention from the circuit design and architecture communities. Many special CMOS solvers from companies (Hitachi[919], Fujitsu[920], Toshiba[921], and others) and academic laboratories[922,923,924] to solve hard optimization problems have been designed and implemented to solve select optimization problems. These special purpose computing systems, sometimes, called Ising Machines map the underlying mathematics of generalized Ising Models to their computing architecture to accelerate the sampling of Ising Models. Such generalized Ising Models have been shown to map to many combinatorial optimization problems[925] and have been compared to D-Wave, despite the technical difference between D-Wave's simulated quantum annealing machines.

4.3.3.2. MONTE CARLO SIMULATION

Probabilistic circuits have also been discussed in the context of traditional Monte Carlo simulation[926]. A broad application of Monte Carlo simulation is that of integration. At high dimensions, Monte Carlo integration is known to scale better than

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

deterministic numerical integration techniques[927] and probabilistic circuits accelerating Monte Carlo simulations can potentially result in energy and area savings with massively parallel probabilistic bits in hardware.

4.3.3.3. CRYPTOGRAPHY

Digital computers often rely on pseudo-random number generators that produce random numbers deterministically given an input seed. For cryptographic applications, certifiable true random number generation (TRNGs) is an important requirement and many stochastic noise sources can be used to generate TRNGs. Among others, optoelectronic (single photon avalanche diodes[928]), magnetic (magnetic tunnel junctions[929]) and metal-insulator phase transition[930] (stochasticity in VO_2 transitions) based true random number generators have been proposed and tested against statistical suites to test the quality of randomness.

4.3.3.4. PROBABILISTIC (P)- LOGIC

In a series of recent papers, Camsari, Datta and collaborators proposed a type of probabilistic computing model introducing the concept of *p*-bits and *p*-circuits.[931,932] The authors explored how *p*-bits can be compactly realized by leveraging existing magnetoresistive RAM (MRAM) technology[933] and showed different applications of p-circuits including image recognition (inference),[934] combinatorial optimization[935,936] Bayesian networks,[937] and an enhanced type of Boolean logic that allows invertible operation.[932] More recently, potential applications have been extended to include emulation of a class of quantum systems[938] and on-chip learning for stochastic neural networks.[939] Further, a prototype realization of an 8 p-bit circuit demonstrating a quantum-inspired integer factorization algorithm that uses MRAM-based p-bits has recently been realized.[940]

The tunability allows a network of *p*-bits (*p*-circuits) to be able to get correlated with one another when appropriately connected through a programmable feedback circuit. The p-bit concept is hardware agnostic and digital implementations of invertible logic have been realized,[941,942] The generic p-circuit consists of autonomously operating p-bits without any digital clocking circuitry, leading to a massively parallel architecture whose performance increases with the number of p-bits in the system.[943]

Recent breakthroughs in modern MRAM industry have led to production-ready integrated chips with up to 1 Gb cell densities, thus leveraging this technology could lead to application specific probabilistic coprocessors with broad applications for the active fields of Quantum Computing and Machine Learning.[944]

4.3.4. REVERSIBLE COMPUTING

Referring back to Figure BC4.2, we can see that, besides analog computing and probabilistic computing, a third dimension along which we may explore departures from the conventional computing paradigm is *reversible computing*.[945] In the present context, when we say that a computation is *reversible*, we mean that the lowest-level physical computational processes should be arranged to approach a condition of being both *logically reversible* and *thermodynamically reversible*. To say that a computational process is *logically reversible* means that known (or deterministically computed) information is not obliviously discarded from the digital state of the machine and ejected to a randomizing thermal environment. To say that the computation process approaches being *thermodynamically reversible* here means that the total increase in physical entropy incurred by the machine's operation per useful computational operation performed should be extremely small, with the vision that this quantity can approach zero asymptotically as the technology continues to be improved.

In 1961, Rolf Landauer of IBM argued[946] that there is a fundamental physical limit on the energy efficiency of conventional *irreversible* digital operations, meaning those that carry out a many-to-one transformation of the space of computational states that is used. Landauer's limit states that an amount $kT \ln 2$ of available energy (where k is Boltzmann's constant and T is the temperature of the heat bath) must be (irreversibly) dissipated to heat per bit's worth of (known or correlated) information that is lost from the computational state. Landauer's limit can be rigorously derived from fundamental physical considerations.[681,947,948]

An important caveat to be aware of is that Landauer's limit only applies to computational information that is *correlated* with other available information, as opposed to independent random information.[681,949,948] A computational bit that bears *no* correlations with other available bits is, in effect, *already* entropy, and thus it can be transferred back and forth between a stable, digital form in a computer and a rapidly-fluctuating physical form in a thermal environment with asymptotically zero net increase in total entropy, by, for example, adiabatically raising and lowering a potential energy barrier separating two degenerate states.[681]

However, most bits in a digital computer are *correlated* bits, having been computed deterministically from other available bits. Performing a many-to-one transformation such as destructively overwriting or erasing such a bit *obliviously* (i.e., without regards to its existing correlations) therefore typically increases total entropy by one bit's worth ($k \ln 2$) and thus implies at least $kT \ln 2$ energy consumption (loss of available energy). Fundamentally, then, the *only* way to avoid Landauer's limit, in a deterministic computational process, is to avoid many-to-one transformations of the computational state. Bennett[950] showed that indeed, this is always possible; that is, any desired irreversible computation can always be embedded into a functionally equivalent reversible one. Such an embedding generally appears to incur some algorithmic overheads,[951,952] in terms of (abstract) time or space complexity, but if reversible devices continue to become cheaper and more energy-efficient over time, then, in principle, these

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

resource overheads can be outweighed by the achievable energy savings, and total cost may be reduced compared to an irreversible design.[953] In the long run, reversible computing is the *only* physically possible path by which the amount of general digital computation that can be performed per unit energy (and cost!) might continue to be increased indefinitely, without any known fundamental limit.[681]

In existing *adiabatic* implementations of reversible computing in today's device technologies (see §§4.3.4.1–4.3.4.3 below), one typically finds that there is a linear tradeoff, at the device level, between the energy dissipation E_{diss} resulting from, and the time interval or *delay* t_{del} required to carry out, a given primitive digital operation in the adiabatic limit. We can express this tradeoff relation by stating that the *dissipation-delay product* (DdP) of the technology is a constant, over some range of achievable delay values. *E.g.*, within that range, we can write

$$E_{\mathrm{diss}} \cdot t_{\mathrm{del}} \cong c_{\mathrm{E}},$$

where c_{E} is the constant DdP, which we may also call the *energy coefficient* of the technology. Further, a very new analysis[954] suggests that asymptotically, this tradeoff relation may even be *fundamental*; that is, that it may apply to *all* physically possible implementation technologies for reversible computing. Nevertheless, other analyses[955,956] suggest that an exponential downscaling of energy dissipation with delay may sometimes be possible, within a limited regime, when quantum effects are leveraged, and the system is well-isolated from the thermal bath. However, even among cases where the linear relation still applies, there are no known *technology-independent* lower bounds on the *value* of the dissipation-delay constant. Although there are indeed firm quantum lower limits on the product of energy *invested* in performing an operation times the delay,[957] there are no known fundamental lower limits above zero on energy *dissipated* for any given delay value. Further, even when a fixed value of the constant is given, thermally limited parallel processors can still benefit from reversible computing in terms of their aggregate performance. For example, in cooling-limited stacked 3D logic scenarios, the per-area performance advantage of time-proportionally adiabatic technologies increases with the square root of 3D processor thickness.[958,953] And in loosely-coupled, arbitrarily-massively-parallelizable applications with fixed power budgets, the aggregate performance gain from adiabatic computing scales up with energy efficiency arbitrarily far, at least up to astronomical scales.[953,954]

However, as of today, experimentally realizing reversible computing's promise to vastly exceed the system-level energy efficiency of all conventional computers in practice remains a difficult engineering challenge. Although a variety of different adiabatic[959,960,961,962,963,964,965,966,967] and ballistic[968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985] schemes for the realization of reversible computation have been proposed, it has so far turned out to be challenging to actually achieve large energy efficiency gains at the system level in practice while accounting for all of the complexity overheads that are incurred from using a mostly-reversible design discipline, together with a variety of real-world parasitic energy dissipation mechanisms that exist and would need to be systematically eliminated or reduced. (See §4.3.4.4 for further discussion.) There is not (yet) any known "magic bullet" physical implementation strategy that automatically addresses all of the many possible energy-loss mechanisms that would typically exist in a complete computing system all at once. Logical reversibility (when suitably generalized[682]) is indeed a *necessary* condition for approaching physical reversibility in deterministic digital computations, but it is by no means a *sufficient* one.

However, while approaching the ideal of physically reversible computing is by no means an *easy* path forward, it is the *only* way that general digital computing can continue to move forward indefinitely, with no clear limits in the foreseeable future—in contrast with the conventional, non-reversible computing paradigm, which is necessarily limited by Landauer's principle.[681] Plausibly, even in CMOS, adiabatic circuits might be able to demonstrate useful energy efficiency gains for highly energy-limited applications (such as spacecraft) even in the relatively near term if sufficiently high-Q resonators can be developed.[986,987] Further, even some of the existing reversible superconducting logic styles (such as RQFP[988,989,990,991,992,993] and nSQUID[994,995,996] logic) already appear to be capable of achieving energy dissipation below the Landauer limit in principle, although the available analyses don't include dissipation in the clock-power supply. However, in cryogenic applications, if the dissipation in the power supply can take place in a higher-temperature exterior environment, this can translate to a significant and highly practically useful reduction in the amount of power that is dissipated *internally* within the low-temperature system.[683,684,987] Finally, superconducting technologies operate with extremely small signal energies, which, if transferred nondissipatively to the room-temperature environment, become relatively insignificant in absolute terms; this can reduce pressure on AC supply design even for general HPC applications. See §4.3.4.2 below for further discussion of superconducting reversible computing technologies.

Reversible computing can also be potentially usefully combined with probabilistic computing (§4.3.3); if random digital bits are obtained by taking in entropy from the thermal environment and capturing it in a stable form, this can actually reduce environment entropy temporarily—albeit without reducing total entropy, of course, since the entropy of the digital state is increased.[681] Once a randomized reversible computation utilizing such bits of "true" entropy has completed, those random bits can later be returned to the thermal environment with no net thermodynamic cost.[681] Thus, the requirement for such a nondeterministic computation to be thermodynamically reversible is somewhat looser than is the case for a deterministic computation; many-to-one

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

(irreversible) transformations can be permitted together with compensating one-to-many (nondeterministic) transformations in a computation,[997] so long as, overall over the course of the computation, previously-established correlations are not lost.

Reversible computing is normally conceived of as a strategy for making *digital* computation more energy efficient. More generally, can a broad variety of *analog* computing schemes be developed that are also thermodynamically reversible? Record energy efficiencies for charge-based analog vector-matrix multiplication have been demonstrated using adiabatic principles as discussed in §4.2.3.3.[998] Further, fundamental physics is reversible at the microscale, which suggests that a sufficiently carefully engineered analog computer might be made to approach macroscopic reversibility, and that its energy efficiency might thus be increased without limit as its technology is further refined. The degrees of freedom utilized for the analog physical computation would likely have to be very well-isolated from the system's thermal degrees of freedom, and the usual tendency for complex dynamical systems to devolve towards chaotic behavior would have to be suppressed in some way, or else made into a useful feature of the computational process, such as in reservoir computing (see §4.3.2.4 above). Also, the previously mentioned work on chaotic logic (§4.3.2.5) suggests a potential technique for harnessing the chaotic analog behavior of conservative dynamical systems usefully for general digital computational purposes, but many other, more sophisticated methods may be possible.

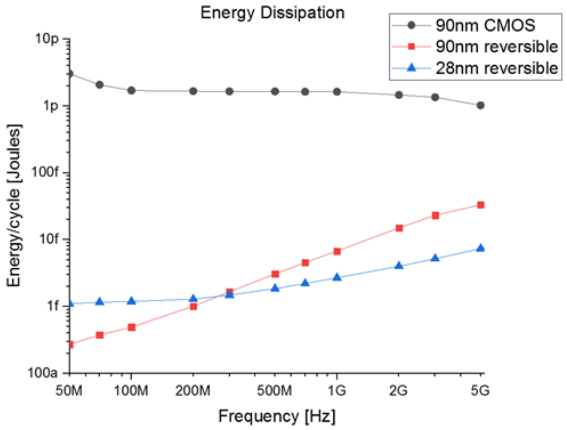


Figure BC4.5 Energy Dissipation per Stage vs. Frequency in an Adiabatic CMOS Shift Register

4.3.4.1. REVERSIBLE ADIABATIC CMOS

As mentioned above, currently the most well-developed implementation technologies for reversible computing are those that utilize classical quasi-adiabatic transformations to carry out digital state transitions. Among such technologies, the most well-developed class of them at present has been referred to variously as *adiabatic CMOS*, *adiabatic transistor circuits*,[684] or just *adiabatic circuits*. In traditional (irreversible, non-adiabatic) CMOS circuits, the full digital circuit-node signal energy of $\frac{1}{2}CV^2$ is dissipated to heat on every digital switching event. In contrast, the use of classical quasi-adiabatic transitions for switching reduces the associated local dissipation by a factor of $\sim t/2RC$, where t is the transition time for a linear voltage ramp and R is the resistance of the charging path. This reduction yields a linear tradeoff between speed and energy dissipation per operation over a certain range of frequencies, where the *dissipation-delay product* or *energy coefficient* scales as $c_{\mathrm{E}} \propto C^2V^2R$.

The earliest complete circuit families for sequential, pipelined reversible computing with adiabatic CMOS were developed in the 1990s in Tom Knight's group at MIT.[962,999,1000,1001,1002,1003] These methods did not gain widespread traction at the time, perhaps in part because, to save energy, adiabatic CMOS must operate relatively slowly compared to the inherent *RC* propagation delay of the gates. However, in the period since the end of Dennard scaling in ~2005, multi-core processor performance has become increasingly limited by power dissipation rather than by raw gate delays (witness the increasing amounts of "dark silicon" in modern processor designs), so, revisiting the adiabatic energy-delay tradeoff appears timely at present. Adiabatic switching holds promise as a design technique for the future, since it offers a means by which the dissipation-delay pareto frontier of any given CMOS technology might be expanded beyond the limits of what can be achieved using more conventional low-power design techniques, such as subthreshold operation.[987]

Subsequent to the original MIT adiabatic logic families cited above, a number of other adiabatic logic design styles were also explored, such as *two-level adiabatic logic* (2LAL),[964,1004] *positive feedback logic* (PFAL),[1005,1006] and *efficient charge recovery logic* (ECRL).[1007,1008] In addition, applications to the design of secure circuits have also been explored, since the unique electrical behavior of adiabatic CMOS circuits can help to prevent non-invasive side channel attacks.[1009,1010] And, general-purpose computing has been pursued in the design of adiabatic microprocessors.[1011,1012] The simulation results shown in Figure BC4.5[1012]

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

indicate that, when operated adiabatically, advanced CMOS nodes such as 28 nm technology can dissipate less energy per cycle than prior nodes at relatively high frequencies, at which dynamic power dissipation exceeds the leakage losses. In contrast, if the same circuit is operated irreversibly, it dissipates an energy orders of magnitude higher at all frequencies up into the GHz range.

Another very recent development is the introduction of fully adiabatic CMOS logic families that are also *fully static*,[1013,1014] meaning that all nodes are at all times connected to a supply reference, which eliminates voltage-level drift from leakage and capacitively-induced voltage sag, which could otherwise occur on floating notes and contribute to non-adiabatic losses. In principle, as leakage is reduced, these "perfectly adiabatic" logic styles are predicted to be capable of exceeding the energy efficiency of *any other* semiconductor-based form of digital logic.

4.3.4.2. REVERSIBLE ADIABATIC SUPERCONDUCTING LOGIC

After adiabatic CMOS, currently the second most well-developed type of hardware technology for reversible computing based on classical adiabatic transformations is the class of adiabatic superconducting logic families.[959–960,965,988–995] A significant motivation for the consideration of superconducting circuits for energy-efficient computation is the lossless nature of charge transport in Josephson junctions and superconducting wires, which act as switching elements and interconnects, respectively. Also, the naturally discrete phenomenon of *flux quantization* facilitates the restoration and stabilization of digital signals.

Two basic families of superconducting reversible digital elements, the parametric quantron (PQ)[959] and quantum flux parametron (QFP)[1015,960] were proposed in early studies. Both approaches were based conceptually on the abstract physical model of adiabatic digital operations introduced by Landauer,[946,950] in which the potential energy function of the digital element is transformed adiabatically between single-well and double-well configurations in the course of an operation. As with adiabatic CMOS, superconducting reversible logic gates utilizing this approach require AC driving waveforms, in this case to provide a time-dependent flux bias to each gate. More recently, a DC-powered superconducting reversible logic gate based on a negative-inductance SQUID (nSQUID) was proposed,[1016] and its energy dissipation was estimated to be a few kT per reversible bit-operation.[994]

Recently, a further reduction of the energy dissipation of the QFP approach was achieved by appropriately optimizing the circuit parameters for the adiabatic mode of operation[965] and eliminating the junction shunt resistance.[1017] The energy dissipation of this improved adiabatic QFP (AQFP) was investigated numerically, taking thermal noise into account,[1018] and found to be well below kT with a low error rate.[988] The energy dissipation of a single AQFP gate was estimated to be 10 zJ per gate at 5 GHz by measuring the scattering parameters of a superconducting resonator coupled to an AQFP gate.[1019] The energy dissipation per operation of an (irreversible) AQFP 8-b carry-lookahead adder was experimentally evaluated to be 24 kT per Josephson junction.[1020]

The first demonstration of logically and physically reversible operation of superconducting logic was performed using a newer reversible QFP (RQFP) design style.[989] The basic RQFP element is a logic gate having three binary inputs x_0, x_1, x_2 and three outputs y_0, y_1, y_2 that are related by

$$(y_0, y_1, y_2) = \left(\mathrm{MAJ}(\overline{x_0}, x_1, x_2), \mathrm{MAJ}(x_0, \overline{x_1}, x_2), \mathrm{MAJ}(x_0, x_1, \overline{x_2})\right)$$

where $\mathrm{MAJ}(i, j, k) = (i \wedge j) \vee (j \wedge k) \vee (k \wedge i)$. This logically reversible element is composed of three AQFP splitter gates and three AQFP majority gates. The bidirectionality and time reversal symmetry of the RQFP gate were investigated, revealing the cause of the energy dissipation in logically irreversible AQFP logic.[990] Using RQFP gates, the functionality of 1-bit[993] and 8-bit[1021] reversible full adders and an 8-word by 1-bit RQFP register file[1022] were demonstrated, and their energy dissipation was numerically calculated, while accounting for thermal noise. Figure BC4.6 shows the simulation results for the energy dissipation of reversible and irreversible full adders as a function of the frequency of the driving clock.[993] It was found that the energy dissipation of the reversible full adder is much lower than that of the irreversible full adder; it becomes lower than the kT thermal energy at 4.2 K at frequencies below 20 MHz.

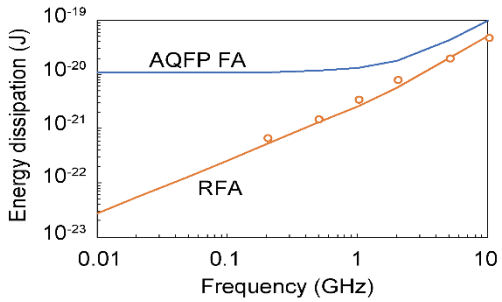


Figure BC4.6 Energy dissipation of RQFP and irreversible AQFP 1-b full adders

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Note: The above figure is from Ref. [993]. Simulation results are plotted as a function of the frequency of the excitation (driving) clock. Lines are calculation results at 0 K, markers show results accounting for thermal noise at $T = 4.2$ K.

4.3.4.3. OTHER REVERSIBLE TECHNOLOGY CONCEPTS

Beyond the adiabatic semiconductor/superconductor technologies discussed in §§4.3.4.1–4.3.4.2 above, over the years, a rather wide variety of disparate aspirational concepts for physical implementation technologies for adiabatic reversible computing have been described in the literature, although typically without accompanying physical demonstrations as of yet.

In the 1980s, the pioneering nanotechnology visionary K. Eric Drexler at MIT outlined a variety of concepts for nanoscale computing technologies, including adiabatic reversible versions, that were based on nanoscale *mechanical*, rather than electrical, interactions.[1023,1024,961,1025] More recently, a group led by Ralph Merkle at the Institute for Molecular Manufacturing (IMM) has been developing an even more advanced nanomechanical reversible logic concept based on (single-atomic-bond) rotary bearings, which were analyzed to dissipate as little as $\sim 4\times10^{-26}$ J per operation at room temperature at speeds of 100 MHz.[966,967] This is roughly 74,000× lower energy than the Landauer limit at 300 K, and roughly 10^6× smaller dissipation-delay product (DdP) than even *end-of-roadmap* CMOS. Although we do not yet have a nanomanufacturing technology that is capable of fabricating the atomically precise nanostructures envisioned in the IMM designs, this analysis nevertheless suggests how much farther the dissipation-delay frontier might someday be extended, beyond what is possible in today's semiconductor- and superconductor-based technologies.

Meanwhile, back in the electronic domain, since the 1990s, a number of adiabatic reversible computing concepts based on *single-electron devices* have been proposed.[963,1026,955,1027] Notable among these is the *quantum-dot cellular automata* (QCA or QDCA) technology concept pioneered at Notre Dame, which has been taken up to the level of complete simulated processor designs.[1028,1029] However, there is not, as of yet, a viable manufacturing process for fabricating scalable QCA-based processors.

To conclude our review of the adiabatic approaches, we mention in passing an interesting concept for *adiabatic capacitive logic*,[1030,1031,1032] which avoids the limitations on the efficiency of adiabatic CMOS due to leakage.

In addition to the various adiabatic approaches to reversible computing, there are also a number of *ballistic* reversible computing concepts.[968–985] These are based on a rather different picture of the basic physical mechanism of reversible computing than the adiabatic approach suggested by Landauer.[946] In the adiabatic approach, some *external* system (i.e., separate from the logic circuits) drives the adiabatic transformations of the computing system that carry out transitions between digital states. Whereas, in the *ballistic* picture, first conceived by Ed Fredkin,[968] the physically reversible dynamics of the system is instead *self-contained*; in other words, individual entities (such as particles or pulses) carrying information-bearing degrees of freedom evolve forwards reversibly under their own (generalized) inertia, as it were, with no direct external influence.

We should note that the distinction between the two classes of approaches is not a perfectly crisp one, since, even in the adiabatic approach, the driving system (such as a resonant oscillator) can be viewed as evolving ballistically, and even in the ballistic approach, the interactions (*e.g.*, elastic collisions) between individual ballistically-propagating information-bearing entities can be analyzed, on a sufficiently fine timescale, as adiabatic processes. So, to some extent, the distinction between the approaches is primarily just one of perspective and emphasis. However, generally speaking, the *adiabatic* approaches are characterized by a large-scale *separation* of the ballistic driving systems from the adiabatic logic transitions being driven, whereas in the *ballistic* approaches, the ballistic properties are *distributed* throughout the system, and are associated to the lowest-level information-bearing entities themselves. We can also imagine that other, future approaches could interpolate between these two extremes. *E.g.*, one could imagine systems comprising large numbers of small ballistic oscillators, each driving just a small region of local adiabatic logic, with the various subsystems communicating timing information and data to each other via elastic interactions transmitted via (short- or long-range) couplings between individual oscillators.

In terms of practical realizations of a (fully-distributed) ballistic approach to reversible computing, the approaches to this that have been developed most intensively to date are based on superconducting electronics.[970–985] This is a particularly convenient technology for ballistic computing, because, unlike in semiconductors, superconductors exhibit the phenomenon of naturally-discrete *single flux quanta* (SFQ), which can propagate near-ballistically along interconnects consisting of *passive transmission lines* (PTLs)[1033] or *long Josephson junctions* (LJJs).[1034] Currently active efforts to develop reversible computing technologies focused on SFQ-based approaches include the *synchronous ballistic* approach, which has been explored since around 2010 at U. Maryland,[974–978,983–984] and the *asynchronous ballistic* approach, which has been in development since 2016 at Sandia National Laboratories.[979–982,985,1035]

4.3.4.4. CHALLENGES FOR REVERSIBLE COMPUTING

Despite the great long-term promise of reversible computing, many fundamental engineering challenges associated with the development of a practical reversible computing technology remain to be solved at this time. These include the following:

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

- Even at the level of very basic physics, a more complete understanding is needed of the fundamental (technology-independent) physical limitations of important cost metrics for reversible computing, such as the dissipation-delay product (DdP), or, more generally, energy dissipation as a function of delay, $E_{\mathrm{diss}}(t_{\mathrm{del}})$. An important question is: Are there universal lower bounds on this quantity that we can derive based on parameters such as temperature, or the length scale of devices, or perhaps based on some kind of generalized viscosity characteristics, or on other fundamental physical or materials-dependent parameters?[1036,1037,948]

- New, more complete abstract (but still realistic) physical models of reversible computing should be crafted to illustrate how we might more closely saturate the above fundamental limits in real artifacts, pointing the way to new device and circuit concepts for reversible computing. Are there quantum-mechanical approaches or phenomena that could be usefully harnessed, such as shortcuts to adiabaticity (STA),[1038] topological invariants, dynamical variations of the quantum Zeno effect (QZE)[1039,1040,954] or others, to help reversible computing technologies to further suppress the rate of entropy increase while still operating as quickly as possible?

- Facilitated by fundamental advances such as the above, new device and circuit concepts for reversible computing need to be developed that significantly reduce $E_{\mathrm{diss}}(t_{\mathrm{del}})$ at useful operating speeds while still being inexpensively manufacturable. Novel physical mechanisms for computing need to be developed *with reversible operation in mind from the start.*

- Meanwhile, to advance the achievable energy efficiency of adiabatic CMOS for cryogenic applications, novel FET device structures that are optimized to minimize leakage at particular cryogenic temperatures of interest with minimal impact on device performance (expressed in terms of, say, DdP) need to be developed.[987]

- For adiabatic reversible computing technologies operating at room temperature, the logic signal energy (*e.g.*, $\frac{1}{2}CV^2$ in CMOS) remains a concern, since it still exists even in adiabatic circuits, and is merely transferred dynamically to the power-clock generator system, rather than being dissipated locally within the logic. Thus, to achieve significant overall energy savings at the system level, compared to the corresponding irreversible technology, this generator must be designed to efficiently recover a large fraction of this signal energy, *e.g.*, by comprising a resonant oscillator with a high *quality factor* (Q). Designing extremely high-Q resonators and clock distribution networks already demands advanced, high-precision engineering. Further, as RF designers know, achieving high Q implies narrow bandwidth. This in turn implies that the returned clock waveform must be extremely pristine—*e.g.*, any data-dependent back-action from the logic must be avoided. Thus, we must maintain a careful load balancing discipline, *e.g.,* via complementary signaling. And if bulk semiconductors are used, this adds another level of challenges relating to time-varying loads during transitions, since device capacitances are more voltage-dependent when depletion regions are not structurally constrained. Thus, fully depleted SOI, thin-film, or gate-all-around (GAA) nanosheet/nanowire FET geometries may be preferred.

- At higher levels, many advances in areas such as reversible architectures, EDA tool enhancements to support reversible design styles, reversible algorithms[952] and so forth still need to be developed. As useful reversible computing hardware technologies emerge and develop, systems engineering practice will also need to evolve to best leverage the opportunities and tradeoffs offered by reversible design.[953] However, all of these R&D areas remain in their infancy at this time.

4.3.5. Digital In-Memory Systems

4.3.5.1. Boolean Logic Operations Using Memory

Boolean operations are another class of operations that can be performed *in situ* using memory cells. Memories that can perform Boolean operations *in situ* typically share two key properties: (1) they can enable multiple cells and/or multiple rows/columns to be active simultaneously, (2) the analog interaction of the stored values in two or more of the memory's cells produces an output that is a Boolean function of the cells. While it is not a strict requirement, Boolean-capable memories typically arrange their memory cells into two-dimensional arrays, which enables them to perform *in situ* Boolean operations on input memory cells that sit in the same row and/or column of the array as each other, and to store the result in an output memory cell that also sits in the same row and/or column as the input memory cells. The specific family of Boolean operations that can be performed using memory are highly dependent on the underlying memory technology and the architecture, as discussed below.

Boolean operations can be performed using conventional bitline-based volatile memories (e.g., DRAM[1041,1042,1043,1044], SRAM[1045,1046,1047,1048,1049]) by taking advantage of, and making minor modifications to, the access circuitry. As a motivating example, Boolean AND and OR operations can be performed between the DRAM cells and the sense amplifier that share a bitline in a conventional 1T1C DRAM[1050]. By simultaneously activating three cells on the same bitline, *charge sharing* shifts the bitline's voltage slightly towards the *average* of the voltages stored in the three cells, causing the sense amplifier to drive a voltage that represents the *majority* (MAJ) function of the three cells. This 3-input MAJ function can be converted into 2-input AND or OR by presetting one of the three cells to 0 or 1, respectively. Additional support is required in the array to provide Boolean

Copyright © 2022 IEEE. All rights reserved.

completeness, such as the inclusion of dedicated rows of DRAM cells with special inverted bitlines to enable NOT[1042,1051]. An alternative approach uses 3T1C DRAM cells to perform in-DRAM NOR operations[1043]. Similar architectures can be used to perform Boolean operations using SRAM[1052,1048,1049], by taking advantage of the dual bitlines (one for the true value and one for the inverted value) connected to each SRAM cell and making minor modifications to the sensing circuitry to support multi-row reading.

Recent works propose to perform Boolean operations using crossbar-based non-volatile memories (e.g., ReRAM[1053,1054,1055], PCM[1056], STT-MRAM[1057,1058], SOT-MRAM[1059]). While the precise implementations differ based on the underlying memory technology, we discuss one such example, MAGIC[1055], which can perform NOR operations in a ReRAM crossbar. For two input cells and one output cell connected to the same row line in a crossbar, MAGIC performs 2-input NOR by applying a fixed voltage by applying a column selection voltage of V_{nor} to the column lines connected to the two input cells, while keeping the row selection voltage of the row line with the three cells floating. (Note that we can also perform NOR on two input cells connected to the same column, but the voltages can differ.) By floating the row selection voltage, the current on the row line settles to a value that is equivalent to the NOR function of the two cells. Setting a column selection voltage of *GND* to the column line connected to the output cell stores the NOR value in the output cell.

One downside of performing Boolean logic operations in memory is the long latency of performing common operations such as addition and multiplication, which often need to be performed in a bit-serial manner (or require additional arithmetic units in memory). To amortize these latencies, Boolean-capable memories typically perform bit-parallel operations on an entire row or column of data at a time[1045,1051,1060]. Recent works provide various architectural abstractions and/or frameworks that expose bit-parallel operations as vector operations[1045,1042,1060] or as single-instruction multiple-thread (SIMT; e.g., GPU) instructions[1046]. A key limiting factor of bit-parallel Boolean operations is the current carrying capacity of the crossbar wires[1060]. For example, in ReRAM, this limits practically achievable column-wide NOR to arrays where the column length is less than approximately 200 cells, requiring other techniques (e.g., pipelining[1060]) to fully amortize the overheads of bit serialization.

The Boolean-capable operations described above have been proposed for single-level cell memories (i.e., memories that store only one bit of data per cell). While this limitation reduces the information density compared to other in-memory functions that can operate on multi-level cells, it eases the ability to fabricate logic-capable memories, by reducing the need to be sensitive to non-linear cell functions that can be difficult to control with peripheral circuitry.

4.3.5.2. PROCESSING-NEAR-MEMORY CIRCUITS

An alternative to *in situ* processing *using* memory cells is *processing-near-memory* (PNM) or *near-memory processing*, where dedicated logic elements for compute are integrated near (but not inside) a memory array[1061]. Logic elements for processing-near-memory are made up of digital transistor-based circuits. While any arbitrary CMOS-like logic can be implemented, PNM logic elements tend to be area-constrained and/or power-constrained, and thus tend to consist of simpler logic than modern conventional CPUs. There are three main approaches to integrating PNM circuits with memory arrays.

The first approach, *near-bank acceleration*[1062,1063,1064,1065,1066] integrates logic elements close to each memory bank, with the logic sitting adjacent to the peripheral circuitry of a bank within the same die as the memory array. Near-bank acceleration avoids the need for data to go through (semi-)global I/O drivers during PNM compute and allows the number of logic elements to scale linearly with the number of memory banks. However, the complexity of logic that can be implemented near the bank is significantly limited, in order to avoid significant disruption of the peripheral circuit design.

The second approach, *logic layer acceleration*[1067,1068,1069], takes advantage of the design of modern 3D-stacked memory chips. The High-Bandwidth Memory (HBM) specification[1070] includes an optional logic layer known as a base logic die, which is integrated as the lowest die in the vertical 3D stack. (The Hybrid Memory Cube (HMC)[1071], which is not currently in active production, also includes a logic layer.) The logic layer can be used to implement global I/O circuits and logic elements, which have access to the through-silicon vias (TSVs) that span the chip. Logic elements implemented in the logic layer can (but do not need to) be distributed across the chip-level memory partitions (e.g., pseudo channels, vaults) that exist in HBM and HMC chips. One challenge of logic layer acceleration is that the logic layer may be fabricated using a DRAM-optimized CMOS manufacturing process technology and may be suboptimal for implementing logic transistors.

The third approach, *discrete die acceleration*, is similar to logic layer acceleration, but allows for entirely separate dies/chips containing the logic elements to be tightly integrated with memory dies/chips. One method for discrete die acceleration makes use of 2.5D/3D integration to connect discrete CMOS logic dies to the internal pins/TSVs of a memory[1072,1073]. Silicon interposers can facilitate 2.5D integration and can potentially allow for relatively large logic dies that make use of a logic-optimized CMOS manufacturing process technology. However, logic dies attached with silicon interposers must be physically placed side-by-side with the memory, which can create power delivery and/or thermal dissipation challenges for large dies. A second method for discrete die acceleration uses buffer chips that are mounted on certain memory modules to house the logic elements[1074]. While

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

implementing PNM circuits in buffer chips allows for easy and cost-effective integration with commodity memory chips, communication between the buffer chip and the memory chips can often still traverse pin-limited buses that consume significant energy and have tight bandwidth constraints.

To date, there are two commercial PNM-based products that are available, both of which employ near-bank acceleration. UPMEM produces PIM-enabled DIMMs (dual inline memory modules), which include PIM chips that contain a DDR4 DRAM array integrated with general-purpose in-order CPU cores[1075,1076]. Samsung produces HBM-PIM, where application-specific PNM logic is integrated with HBM arrays[1063,1064,1077]. Currently available HBM-PIM products support deep neural network (DNN) computation, but products with logic elements for other applications are expected to be released in the future.

4.4. Enabling Devices

4.4.1. Synaptic Devices

This section lists some specific device technologies that are useful in analog computing.

4.4.1.1. Memory Based Devices

In general, synaptic devices also need to store a synaptic state and are therefore often based on a memory element. Consequently, all but the single flux quantum and photonic based systems are described in the memory section, §2. The memory-based synaptic devices surveyed include:

- ReRAM
- Phase Change Memory
- Ion-Insertion Redox Transistor:
- Floating Gate
- Capacitor-on-Gate
- Charge-Based Analog Arrays for VMM
- Single Flux Quantum Neural Network Inference
- Magnetic Neural Network Devices

4.4.1.2. Single Flux Quantum Neural Network Inference

Josephson junctions assembled into single flux quantum (SFQ) circuits form a natural neuromorphic system. In this technology, SFQ pulses act as the action potential for a spiking neuron, and superconducting transmission lines act as axons. The synaptic or weighting function can be implemented with a flux-biased Josephson junction[1078] or with a clustered magnetic Josephson junction.[1079] These elements represent the basic functionality of an artificial spiking neural network (SNN).

Superconducting and magnetic device technologies are relatively mature among emerging devices. On the magnetic device side, large non-volatile memories based on magnetic tunnel junctions, using spin-polarized currents to write, have been commercialized.[1080] On the superconducting device side, high speed microprocessors and communication systems have been developed, based mostly on superconducting tunnel junctions.[1081]

When implementing the synaptic function with a magnetic nanocluster Josephson junction, the superconducting order parameter is modulated by the magnetic state of the nanoclusters in the barrier. The magnetic state of embedded nanoclusters can be changed by applying small current or field pulses, enabling both unsupervised and supervised learning. Maximum operating frequencies of these systems are above 100 GHz, while spiking and training energies have been demonstrated at roughly 10^{-20} J and 10^{-18} J, respectively.[1079] High speed and low-power operation are promising for direct hardware implementation of neural networks that could perform inference at higher speeds and potentially lower power than alternatives even when including the cooling overhead to operate at 4 K.

4.4.1.3. Photonic Vector-Matrix Multiply (VMM)

There are currently two major bottlenecks in the energy efficiency of artificial intelligence accelerators: data movement, and the performance of multiply-accumulate (MAC) operations, or matrix multiplications. Light is an established communication medium and has traditionally been used to address data movement on a larger scale. As photonic links are scaled smaller and some of their practical problems addressed, photonic devices have the potential to address both of these bottlenecks on-chip simultaneously. Such photonic systems have been proposed in various configurations to accelerate neural network operations (see [1082,1083,1084]). However, their main advantage comes from addressing MAC operations directly. Here, we will look at the

Copyright © 2022 IEEE. All rights reserved.

advantages of a simple matrix vector multiplication (MVM) unit made of integrated photonic components, in which inputs and outputs are encoded as light signals, and analog matrix multiplications are performed using a passive optical array.

Figure BC4.7 One Generalized Instantiation of a Photonic MVM unit, with Wavelength Multiplexed Inputs and Outputs and a Coupler-based Tunable Array. Reproduced from Ref. 1099.

One possible instantiation of a photonic MVMs is shown in the figure above. Power or phase can be used to encode information, while wavelength or phase selectivity can be used to program the network into a desired configuration. Wavelength division multiplexing (WDM) can further increasing the compute density of the approach. Classic examples include arrays of resonator weight banks[1083,1085,1086] or Mach Zehnder interferometers[1084]. The most important metrics are *energy efficiency (energy/MAC), throughput per unit area (MACs/s/mm^2), speed* (*MVM/s), and latency* (*s),* where both speed and latency are measured across an entire matrix-vector (MVM) operation. In CMOS, MVM operations are typically instantiated using systolic arrays[1087] or SIMD units,[1088] although there are some other architectures that use aspects of both.[1089] Digital systems are limited by the use of many transistors to represent simple operations and require machinery to coordinate the data movement involved in both weights and activations. The state-of-the-art values typically hover around 0.5–1 pJ/MAC, 0.5–1 TMACs/mm^2, 0.5–1 GMVM/s, and 1–2 us, respectively. In contrast, photonics MVM units could perform in range 2–10 fJ/MAC, 50 TMACs/mm^2, and ~3 ps (1 clock cycle) per MVM operation. This performance depends on solving a number of practical problems which are possible to address in the short term. These are discussed below.

The largest bottleneck in efficient photonic MVM operations is the use of heaters for coarse tuning. Typically, the thermo-optic coefficient (dn/dT) is the strongest effect in most materials of interest (i.e., silicon), leading to heavy use of heaters in almost any tunable passive photonic system. There are several ways these can be eradicated, via the use of post-fabrication trimming[1090,1091] or devices with an enhanced electro-optic coefficient (dn/dT, dalpha/dT) such that heaters are not as necessary.[1092,1093] The second largest problem is fabrication variation, which can result in parameter drifts for devices in an array. Resonators, for example, are highly sensitive to such variation, particularly across a wafer. This can also be remedied by enhancing the electro-optic coefficient of devices and some other tricks (see [1094,1095] for resonators). Third, the signal-to-noise ratio of the output must be optimized by reducing the intrinsic loss of photonic components together with the noise on the receiver. There are a variety of technologies that can address this—for example, lasers can be coupled on-chip with < 1 dB of loss,[1096] photonic devices in state-of-the-art silicon foundries can be designed with low scattering,[1097] while detectors such as avalanche photodiodes,[1098] can reduce the relative contribution of thermal noise to the signal at the receiver.

Photonic arrays ultimately have very similar limits to analog electronic crossbar arrays, as analyzed in Ref. [1099]: single-digit aJ/MAC efficiencies, and 100s of PMACs/s/mm^2 compute densities. However, photonic MVMs garner an advantage for larger MVM units, both in the size of the matrix and in the physical footprint of the core. Generally speaking, optimized photonic systems tend to perform worse than their electrical counterparts for smaller arrays (distances approximately < 100 um), but perform better for larger arrays (distances approximately > 100 um)[1099]. In that sense, photonic MVM arrays have a similar profile to photonic communication channels, with better performance over larger distances. However, photonic systems tend to have worse signal-to-noise ratios, as a result of several factors: (1) photonic channels are ultimately shot noise limited, which is more than an order of magnitude greater than the thermal noise limits on resistors,[1099] and (2) to achieve similar compute densities to electronics, photonic MVMs must run faster to compensate for their larger device sizes, and noise is speed dependent. That being said, there are some architectural tricks to reduce this issue—for example, optical unitary operations[1084] can conserve the variation of the input and output signals, in contrast to other approaches such as resistive crossbar arrays,[704] which by default, see a $\sqrt{N}$ decrease in effective signal variation from input to output for an $N \times N$ matrix operation.

Although photonic arrays exhibit some fundamental advantages over analog electronics (particularly for large matrix sizes or large physical sizes), a more important question is whether photonics arrays are practical. Thankfully, the transceiver industry has created a silicon photonic ecosystem fully compatible with high volume manufacturing (HVM). Compared to CMOS chips, photonics has costlier packaging, largely because light generation cannot be done easily in silicon—in fact, the cost of a production photonic chip is dominated by packaging. In addition, the tools required for the design and testing of large-scale photonic systems (>10k components) are still early in early development—analog photonic systems must grapple with the challenge of addressing yield, variability, precision, and tunability. Nonetheless, the total cost to produce a photonic chip package

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

at high volume is dipping below one hundred dollars, and it is expected that the trend will continue[1100]. The orders of magnitude advantages offered by photonics, and its potential for HVM scalability, makes it a viable inroad for the breakneck performance and innovation required by artificial intelligence algorithms in the years to come.

4.4.2. OSCILLATOR DEVICES

The principles that underlie the mechanism of combinatorial optimization in coupled oscillator systems can be considered universal, present in many different oscillator technologies. These general physical principles include parametric oscillation, injection locking, and sub-harmonic injection locking.[846,838] While the same properties have been exhibited in a number of highly differentiated technologies, these solutions can differ considerably in their scalability, coupling scheme, energy efficiency, susceptibility to process variations or parasitics, and foundry compatibility.

4.4.2.1. ELECTRICAL OSCILLATORS:

Electrical oscillators have the advantage that they can be coupled electrically, which is relatively straightforward to do in a CMOS process. For ring oscillators, LC oscillators, and other electrical oscillators, positive and negative coupling coefficients can be implemented with resistors. [837,838,840,1101] Arbitrary continuous-valued coupling coefficients can be realized using precisely programmed resistances. This allows an array of synaptic memory devices (such as ReRAM, flash, PCM) to be leveraged to connect together the oscillators, so as long as they are compatible with the oscillators' operating voltage and frequency. Such an array can provide all-to-all connectivity among a large number of oscillators. High programming precision in these resistors enables scaling to larger problem sizes while retaining high solution quality.[837]

4.4.2.1.1. RING OSCILLATORS:

Digital ring oscillators, based on CMOS inverter feedback loops, are perhaps the most readily implemented coupled oscillator system in today's foundry process; a large network of ring oscillators was recently demonstrated that solves combinatorial optimization problems.[1101]

4.4.2.1.2. LC OSCILLATORS:

Electrical oscillators can also be built from transistor-driven LC oscillator circuits, with CMOS cross-coupled pairs supplying the gain needed to sustain the oscillations.[837,838,840] Gain can also be supplied using the negative differential resistance of a tunnel diode.[1102] The oscillator amplitudes remain fixed during optimization but their phases are induced into bistability (separated by 180°) through sub-harmonic injection locking. Alternatively, phase bistability can be induced in an LC oscillator by modulating the capacitance with a pump at twice the resonance frequency.[1103] This can be done using the nonlinear capacitance-voltage characteristics of a semiconductor varactor. Pumping the varactor introduces phase-sensitive gain, and a phase-bistable parametric oscillation grows from noise.[837] In each case above, which of the two bistable phases is chosen is random for an isolated oscillator, depending on thermal noise and phase jitter. Coupling to other oscillators changes the probability distribution, but the process remains stochastic. This allows the system to find a different local minimum of the optimization problem when re-initialized. For scalability of on-chip solutions, LC oscillator systems can leverage the fact that at high frequencies (>10 GHz), integrated inductors have about the same area as the capacitor, so the oscillator can remain compact.[1104,1105]

4.4.2.1.3. SPIN TORQUE OSCILLATORS:

Among electrical oscillators, one prominent effort is the use of spin torque oscillators (STOs) for providing a computational platform for machine learning, spiking neural networks, optimization, and other applications.[1106,1107,1108,1109,1110,1111] STOs are based on the phenomenon that when a spin-polarized current passes through a ferromagnetic material, precession of the film's magnetization can be induced. This magnetic precession (which has ~1-10 GHz resonance) can be converted to an oscillating electrical signal by incorporating the ferromagnetic inside a magnetic tunnel junction, which would has an oscillating tunnel resistance when driven by a DC current. STOs can be coupled together resistively and sub-harmonic injection locking can be used to solve discrete optimization problems.[1112] However, the high current densities of STOs and the limited range of spin diffusion currents continue to pose serious challenges in creating coupled networks of such oscillators.

4.4.2.1.4. INSULATOR-METAL TRANSITION OSCILLATORS:

Another promising non-silicon technology for very compact oscillators is the IMT (insulator-metal transition) material-based oscillator technology.[1113,1114,1115,1116,1117,1118] IMT devices, using materials such as VO_2, have a resistance that can change abruptly with the applied electric field (voltage). Connecting the device to a load circuit can induce self-sustained oscillations between the two resistance states, and sub-harmonic injection locking can be used to induce phase-bistable oscillations.[1119,1120,1121,1122] IMT oscillators can operate at room temperature, and the IMT transition can occur on nanosecond time scales. As the oscillation mechanism is completely electrical, the coupling of oscillators can be done easily using electrical components. There have been other implementation efforts for electrical oscillators[1123,1124,1125,1126,1127] but the focus has been to build high frequency and low power individual oscillators, as opposed to the demonstration of coupled systems of oscillators, or the generation of interesting dynamics for computing.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

4.4.2.2. Optical Oscillators:

Optical oscillators have been studied[1128,1129] and used for computing,[842] but challenges include bulky components, difficult interfacing between the electrical and optical domains, and lack of programmability to enable an optical computing apparatus. Optical parametric oscillators are particularly useful for discrete optimization applications. Like their electrical counterpart, rely on a second-harmonic pump to induce a phase-bistable coherent oscillation from noise. This has been shown to solve the Ising problem with success,[842,843,844] but coupling the oscillators together is more practically difficult in the optical domain. The Coherent Ising Machine implemented these oscillators as pulses traveling along a long optical fiber, but the coupling had to be accomplished using electronic circuits that read out optical signals, digitally applied coupling weights, and re-injected pulses into the fiber at precise times to achieve coupling.[843]

4.4.2.3. Electromechanical Oscillators:

Parametric oscillation has also been demonstrated in electromechanical systems.[1130] For binary optimization, two spins can be encoded in the symmetric and anti-symmetric vibrational modes of a single mechanical resonator, which are independent. Modulating the spring constant of the resonator material at the appropriate frequency can implement specific coupling schemes between two spins. The modulation can be achieved using a voltage signal by exploiting the piezoelectric effect.[845]

4.4.3. Stochastic Device Technologies for Random Bit Generation

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

4.4.3.1. Magnetic Tunnel Junctions (MTJ)

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

Low-barrier MTJs can convert ambient thermal noise on nanomagnets into a fluctuating resistance, which is then used to build a device with tunable randomness when integrated with minimal CMOS periphery. The fluctuating resistance change due to thermal magnetic noise in MTJs can be measured by Tunneling Magneto resistance (TMR). State-of-the-art TMR values range from upwards of 100% to 600% demonstrated by the Tohoku Group,[1131] and commercial STT-MRAM devices exhibit >100% TMR. A large TMR would enable a robust functional unit for controllable randomness. The theoretical limit for TMR in MgO-based MTJs has been reported[1132] to be 1,000% and can presumably be larger. There is currently intense research activity in half-metallic ferromagnets to increase TMR.

Two key experimental breakthroughs on MTJs, one from IBM researchers[1133] and one from Tohoku University[1134], have been the demonstration of nanosecond (GHz) fluctuations of low-barrier ferromagnets which could find use in probabilistic circuits implemented with MRAM technology

4.4.3.2. Single-Electron Bipolar Avalanche Transistor (SEBAT)

The single-electron bipolar avalanche transistor (SEBAT) is a novel Geiger-mode avalanche bipolar transistor structure.[1135,1136] The device generates Poisson-distributed digital output pulses at rates between 1kHz and 20MHz. The pulse rate is linearly proportional to the emitted current. A MOS transistor is also formed within the base region of the device, allowing for voltage control of the pulse rate. The device is fully compatible with low-voltage CMOS circuits and standard digital process steps.

4.4.3.3. Memristors/Resistive RAM

The intrinsic variability of memristive switching, particularly the switching delay time of memristors, can be a good source of stochasticity.[1137,1138,1139] Such stochasticity originates from the ionic dynamics within the memristors.[1140]

4.4.3.4. Contact-Resistive Random Access Memory (CRRAM)

CRRAM can be used for random number generation.[1137, 1141] A CRRAM device may be based on a layer of silicon dioxide that is sandwiched between two electrodes; the bottom electrode could simply be the drain of a CMOS transistor. [1142] During operation, the current flowing in a filament channel will be (randomly) impacted by any electrons trapped in the insulating layer. If a high voltage is applied to a device, the current in the filament channel will be large and not impacted by trapped electrons. However, with the application of a lower voltage, the width of a filament will shrink, and the trapped electrons will (randomly) influence output current.

Copyright © 2022 IEEE. All rights reserved.

4.4.3.5. CMOS

There are different ways to obtain random number generators (RNG) in CMOS using different physical noise sources, one being "jitter" in ring oscillators.[1143] These TRNGs can be tuned into tunable random number generators as required but require significant amounts of area when compared to single device alternatives.

4.4.3.6. STOCHASTIC JOSEPHSON JUNCTION

Single flux quantum (SFQ) logic relies on voltage pulses generated by 2π phase slips of the superconducting order parameter across a Josephson junction. These voltage pulses have a time-integrated amplitude given by the flux quantum $\Phi_0 = 2\times10^{-15}$ Wb. A standard circuit model of a Josephson junction is the parallel connection of a supercurrent up to I_c the critical current, a normal state resistor, a capacitor, and a channel for the thermal noise term. The resulting dynamics are the same as a forced damped pendulum. The energy barrier is given by $(I_c\Phi_0)/(2\pi)$. For stochastic operation one can operate with junctions that meet the condition $\delta = (2\pi\, k_B T)/(I_c\Phi_0) \sim 10$, where δ is the stochasticity, k_B is the Boltzmann constant and T is the temperature in kelvin. The exact value of the stochasticity is an important circuit parameter as it determines the frequency of spiking events. In this regime, the energy in a single flux quantum spike is sub-attojoule while the frequency can be greater than 100 GHz.[1144]

Because Josephson junctions can be operated near the thermal stability limit, the amount of stochasticity can be varied by changing the temperature a few degrees. For values of the stochasticity less than 10, the dynamics are basically deterministic, whereas when the stochasticity is larger there is a significant stochastic component. The value of the stochasticity can be effectively tuned between the deterministic state and the stochastic state by changing the temperature of the circuit by a few degrees.[1145] Circuits based on these devices have many promising potential applications. For example, stochastic Josephson junctions have been shown to make effective pseudo-sigmoid generators,[1146] and stochastic Josephson junction spiking can perform the neural accumulate operation at speeds up to 70 GHz.[1147]

4.5. DEVICE-ARCHITECTURE INTERACTION: CONCLUSIONS/RECOMMENDATIONS

In this section, we have surveyed a variety of concepts and R&D directions for the development of novel Beyond CMOS computing technologies that represent an effort to think "outside the box," in the sense of looking beyond just developing simple drop-in replacements for traditional logic and memory cells. More broadly, new hardware designs spanning multiple levels of the technology stack from the devices up through circuits and architectures must be considered, and the interactions between the various levels explored. More specifically, we expand the scope of future computing technologies beyond traditional irreversible, deterministic digital logic to include a broad range of alternative, unconventional computational paradigms, such as analog, probabilistic, and (classical) reversible computing paradigms.

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

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

5. BEYOND CMOS DEVICES FOR MORE-THAN-MOORE APPLICATIONS

5.1. EMERGING DEVICES FOR SECURITY APPLICATIONS

5.1.1. INTRODUCTION

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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,[1148,1149,1150] emerging technologies also demonstrate unique features that could drastically simplify circuit structures for protection against hardware security threats. Security applications could not only benefit from the non-traditional I-V characteristics of some emerging devices, but also help shape research at the device level by raising security measures to the level of other design metrics.

At present, many emerging technologies being studied in the context of hardware security applications are related to designing physically unclonable functions (PUFs). Many post-CMOS devices[1151,1152,1153] have been suggested as a pathway to a PUF design. (More detailed reviews are also available.[1154]) With a PUF, challenge/response pairs are mapped (typically in a trusted environment). Responses are derived from natural/random variations and disorders in an integrated circuit that cannot be copied (or cloned) by an adversary. PUFs have been employed for tasks such as device authentication,[1154] to securely extract software,[1155] in trusted Field Programmable Gate Arrays (FPGAs),[1156] and for encrypted storage.[1155] Post-CMOS devices also find utility as random number generators (RNGs) that may be employed for secure communication channels (e.g., to generate session keys[1154]). That said, while intriguing, PUFs and RNGs may only cover a small part of the hardware security landscape. (Furthermore, one must be careful that PUF designs based on emerging technologies do not depend on device characteristics that a designer would like to eliminate when considering utility for logic or memory.)

Given the many emerging devices being studied[1148] and that few if any devices were proposed with hardware security as a "killer application," this document also reports initial efforts as to how the unique I-V characteristics of emerging transistors that are not found in traditional MOSFETs could benefit hardware security applications.

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

5.1.2. Physically Unclonable Functions (PUFs) and Emerging Technologies

A variety of different emerging logic and memory technologies have been considered in the context of PUFs. As has been reviewed,[1154] variations in the required write time in spin torque transfer random access memory (STT-RAM) was proposed to create a domain wall memory PUF.[1151] Other structures based on magnetic tunnel junctions have also been proposed.[1157,1158] Variations in write times have also been exploited to produce unique responses in phase change memory (PCM) arrays.[1159] The variability of ReRAM presents a natural opportunity for PUF implementation, and array demonstration has been reported.[1160, 1161] At the array-level, variations in diode resistivity have also been used to derive challenge/response pairs from crossbar structures.[1162] PUFs based on graphene[1163] and carbon nanotubes have also been proposed/considered.[1164]

As a more representative case study, prior work[1154] considers an array structure based on process variation in memristors[1165,1166] to create a PUF structure (referred to as *NanoPUF*[1154]). NanoPUF is based on 1) a crossbar with memristors. 2) A challenge is applied to the memristor array by using a row decoder to apply a voltage amplitude (V_{dd}) to a given row that can vary in duration; a column decoder connects a given column to a resistance R_{load}. All other rows and columns remain floating. 3) A response circuit (to collect outputs to different challenges) would consists of R_{load} and a current comparator that compares I_{out} from a given column to a reference current I_{ref}. A logic 1 might be recorded if $I_{out} > I_{ref}$, while a logic 0 might be recorded if $I_{out} < I_{ref}$. With respect to PUF functionality, when a write pulse is applied, natural process variations will cause some memristors to turn on (leading to a logic 1), and others to remain off (leading to a logic 0). While the time of the right pulse serves as one variable,[1165] the pulse's duration and amplitude may also be varied.[1166]

5.1.3. Random Number Generators (RNGs) and Emerging Technologies

The inherent randomness in emerging devices can also be used to generate random numbers.[1154] As a representative case study, prior work[1154] explores an approach based on contact-resistive random access memory (CRRAM).[1167] (Note that a CRRAM device may be based on a layer of silicon dioxide that is sandwiched between two electrodes; the bottom electrode could simply be the drain of a CMOS transistor—which in turn suggests that RNGs based on emerging technologies can be CMOS compatible.[1168])

During operation, the current flowing in a filament channel will be (randomly) impacted by any electrons trapped in the insulating layer. If a high voltage is applied to a device, the current in the filament channel will be large and not impacted by trapped electrons. However, with the application of a lower voltage, the width of a filament will shrink, and the trapped electrons *will* (randomly) influence output current.[1168] Indeed, RNGs based on emerging devices[1168] can successfully pass randomness tests such as those provided by the National Institute of Standards and Technology (NIST).

Copyright © 2022 IEEE. All rights reserved.

As random number are derived from current passing through filaments, memristors, PCM, and RRAM devices can also be leveraged to build similar RNGs.[1154]

5.1.4. Other Hardware Security Primitives based on Emerging Technologies

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

5.1.4.1. Emerging Technologies for IP Protection

Tunable Polarity—In many nanoscale FETs (45 nm 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,[1169] carbon nanotubes[1170] and graphene.[1171] By controlling ambipolarity, device polarity can be adjusted/tuned post-deployment. Transistors with a configurable polarity—e.g., carbon nanotubes,[1172] graphene,[1173] silicon nanowires (SiNWs),[1174] and transition metal dichalcogenides (TMDs)[1175] —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.[1176]

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.[1177] 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,[1177] one can envision fabricating devices that could be better suited for logic as well as security-related applications. Given that the screening length in TMD devices scales with their body thickness, one can achieve substantial tunneling currents.

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

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

SiNW FET based polymorphic gates to prevent IP piracy have been introduced.[1178,1179] If the control gate (CG) of a SiNW FET is connected to a normal input while the polarity gate (PG) is treated as the polymorphic control input, we can easily change the circuit functionality through different configurations on the polymorphic control inputs without a performance penalty. For example, a SiNW FET-based NAND gate can be converted to a NOR gate, whereas a CMOS-based NAND cannot be converted to a fully functioning NOR by switching power and ground.

TFET-based polymorphic logic circuits have also recently been developed.[1180] 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.[1180]

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

Copyright © 2022 IEEE. All rights reserved.

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.[1181] A CMOS camouflaging layout that can function either as an XOR, NAND or NOR gate requires at least 12 transistors. Emerging technologies help reduce the area overhead. Recent work has demonstrated that only four SiNW FETs with tunable polarity are required to build a camouflaging layout that can perform NAND, NOR, XOR or XNOR functionality.[1182,1183] 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.[1180] The probability that an attacker can retrieve the correct key becomes $1/2^{32}$ (2.33×10^{-10}). Obviously, polymorphic based logic obfuscation techniques are resistant to a conventional brute-force attack. With respect to camouflaging layouts, given that our proposed SiNW based camouflaging layout can perform four different functions, the probability that an attacker can retrieve the correct layout is 25%. Therefore, if *N* SiNW FET camouflaging layouts are incorporated in a design, the attacker has to compute up to 4^N times to resolve the correct layout design. Compared to polymorphic gate-based logic obfuscation, camouflaging layout embraces higher security level but with larger area overhead.

5.1.4.2. Emerging Technologies to Prevent Side-Channel Attacks

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

Steep slope transistors—TFETs have been exploited to design current mode logic (CML) style light-weight ciphers.[1184,1185] The high energy carriers in TFETs can be filtered by the gate-voltage-controlled tunneling such that a sub-60 mV/decade subthreshold swing is achievable at room temperature.[1186] With improved steep slope and high on-current at a low supply voltage, TFETs could enable supply voltage scaling to address challenges such as undesirable leakage currents, threshold voltage reduction, etc. Different types of TFETs have been developed and fabricated.[1186,1187]

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.[1178] Such behavior has been observed experimentally.[1188,1189] More specifically, V_{TG} and V_{BG} change the carrier type/density of the drain and source graphene layers by the electrostatic field, which can modulate I_{DS}. ITFETs or ThinTFETs may exhibit similar I-V characteristics.[1190]

Preventing fault injection—Side-channel analysis, such as fault injection, power, and timing, allows attackers to learn about internal circuit signals without destroying the fabricated chips. Countermeasures have been proposed to balance the delay and power consumption when performing encryption/decryption at either the algorithm or circuit levels.[1191] 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.[1178,1179] With only two SymFETs, the power supply protector can easily monitor the supply voltage to ensure that the supply voltage to the circuit-under-protection is within a predefined range. In the event of a fault injection, the decreased supply voltage will power down the circuit rather than injecting a single-bit fault,[1179] and can thus protect the circuit from fault injection attacks. If one uses V_{out} as the power supply to a circuit under protection (e.g., an adder), due to the bell-shaped I-V characteristic of the SymFET, an intentional lowering of V_{DD} cuts off the power supply. Thus, the sum and carry-out of the full adder output is '0', and no delay related faults are induced. A similar CMOS power supply protector would require op-amps for voltage comparison. As a result of the voltage/current monitors developed thus far, voltage/current-based fault injections can be largely prevented. By inserting the protectors in the critical components of a given circuit design, the power supply to these components can be monitored and protected.[1180] (SymFET-based Boolean logic is also possible.[1192])

Copyright © 2022 IEEE. All rights reserved.

Preventing differential power analysis (DPA)—As an advanced side-channel attack scheme, DPA employs analysis of statistic power consumption measurements from a crypto system to obtain secret keys. Since the introduction of DPA[1194], there has been many efforts to develop low-cost and efficient countermeasures. Countermeasures are generally classified into two categories: 1) algorithm-level solutions and 2) hardware-level solutions.

Algorithm-level solutions—aim to design cryptographic algorithms that can withstand a certain amount of information leakage,[1193] e.g., frequently changing the keys to prevent the attacker from collecting enough power traces[1194] or using masking bits during the internal stages to limit information leakage.[1195]

A more practical circuit-level method for preventing DPA attack leverages a *sense amplifier-based logic* (SABL) or *current mode logic* (CML) for cryptographic algorithm implementations.[1196] A CML gate includes a tail current source, a current steering core and a differential load. A CML gate will switch the constant current through the differential network of input transistors, utilizing the reduced voltage swing on the two load devices as the output. Although CML is not widely used in mainstream circuit design, its unique features, namely low latency and stable power consumption, can be leveraged to serve as a countermeasure against a DPA attack.

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

Recent work has implemented a standard cell library of TFET CML gates and conducted a detailed study of their performance, power and area with respect to CMOS equivalents.[1197] Standard cells were used to implement and evaluate TFET-based CML on a 32-bit KATAN cipher (a light-weight block cipher). All KATAN ciphers share the same key schedule with the key size of 80 bits as well as the 254-round iteration with the same non-linear function units.[1198]

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

Moreover, the correlation coefficient of a TFET static KATAN32 reaches its highest when the correct keys are applied. By comparison, the correlation coefficient of TFET CML KATAN32 is much more scattered, and all four hypothetical keys are equally distributed. Thus, the TFET CML KATAN32 implementation can successfully counteract CPA. Because the power consumption is mainly determined by AND/XOR logic gates of two nonlinear functions—and the effect of CPA is maximized—the correlation coefficients for KATAN32 are higher on average than other block ciphers, e.g., CPA on S-box[1199].

6. Emerging Materials Integration

6.1. Introduction and Scope

6.1.1. Current State of Technology

The semiconductor industry was historically driven by a strong correlation between technology scaling and performance of most integrated circuits. The PC market required more complex and faster microprocessors that largely drove the development and scaling of transistors and memory. These devices required new materials and processes such as strained silicon, high-k gate dielectrics and metal gate electrodes that are now widely, and will continue, to be used in IC manufacturing. In the past decade, a completely new ecosystem has emerged. New system integrators, from mobile to data centers to the Internet of Everything, have appeared with new and complex technology requirements. These system integrators will have impact that includes microprocessors, but extends towards new applications including medicine, energy and the environment.

6.1.2. Drivers and Technology Targets

As transistors and memory begin to run out of horizontal space and ICs continue to be limited by power, device technologies will enter a phase characterized by vertical integration and performance specifications driven towards reduction of power. New transistor, memory, interconnect, lithography materials and processes will be required to enable this new More Moore scaling paradigm. As conventional information processing and storage technology reaches its ultimate limits, entirely new non-CMOS logic and memory devices and even new, non-Von Neumann circuit architectures are potential Beyond CMOS solutions. Such

Copyright © 2022 IEEE. All rights reserved.

solutions ideally can be integrated onto the Si-based platform to take advantage of the established processing infrastructure, as well as being able to include Si devices such as memories onto the same chip. However, while these technologies will likely be integrated on a Si-based platform, the vast majority of these Beyond CMOS technologies are based on entirely new materials and physics. Finally, new system integrators require materials that enable potentially trans-disciplinary advances in monolithically integrated complex functionality, i.e., functional scaling. Significant challenges must be overcome for these emerging materials to provide viable solutions for future integrated circuit technologies. To deliver these capabilities, enhanced Metrology will be needed to accelerate material evaluation, improvement and capabilities. The ultimate goal is to provide timely guidance on emerging material and process performance, cost, reliability, and sustainability options that will drive breakthrough advances in future manufacturing technology.

6.1.3. SCOPE

The IRDS represents a strategic repositioning of the community's scope, needs, and set of emergent opportunities. In alignment with this new perspective, this edition of the emerging materials integration (EMI) sub-chapter represents a work in transition with a primary goal of aligning with the needs of related IRDS working groups. Much of the associated information in the detailed requirements and solutions tables comes from prior ERM chapters and input from current IRDS working groups, and will be updated in future editions. The chapter emphasizes strategic difficult challenges and/or enabling of novel, breakthrough and potentially disruptive opportunities for emerging material properties, synthetic methods, and metrology, organized in the following areas:

1. ***Scaled technology materials needs for More Moore:*** transistors, memory, interconnects, lithography, heterogeneous integration, assembly and packaging.
2. ***Novel materials for Beyond CMOS:*** emerging logic and information processing devices, emerging memory and storage devices, and novel computational paradigms and architectures.
3. ***Potentially disruptive material opportunities for functional scaling and convergent applications****:* Heterogenous components, outside system connectivity, and high impact application areas such as energy, environment, agriculture, health, medical, etc.

For all areas, the advancement requires an intergration of emerging materials as illustrated in Figure BC-EMI 1.

Figure EMI1 Emerging Material Integration Promotes the Advancement of Existing Technologies

6.2. CHALLENGES

6.2.1. NEAR-TERM CHALLENGES

Table EMI1 Near-term Difficult Challenges

Near-Term Challenges: 2022–2028	***Description***
Materials and processes that achieve performance and power scaling of lateral fin- and nanowire FETs (Si, SiGe, Ge, III-V).	Integrated high k dielectrics with EOT <0.5nm and low leakage. Integrated contact structures that have ultralow contact resistivity. Achieving high hole mobility in III-V materials in FET structures. Achieving high electron mobility in Ge with low contact resistivity in FET structures. Processes for achieving low dislocations and anti-phase boundary generating interface between Ge/III-V channel materials and Si. Dopant placement and activation i.e., deterministic doping with desired number at precise location for Vth control and S/D formation in Si as well as alternate materials.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Materials and processes that improve copper interconnect resistance and reliability	Mitigate impact of size effects in interconnect structures. Patterning, cleaning, and filling at nano dimensions. Cu wiring barrier materials must prevent Cu diffusion into the adjacent dielectric but also must form a suitable, high quality interface with Cu to limit vacancy diffusion and achieve acceptable electromigration lifetimes. Reduction of the k value of inter-metal dielectrics.
Materials and processes for continued scaling of DRAM/SRAM and embedded NVM	Low temperature materials for high performance vertical transistor memory select structures. High-k, low leakage DRAM dielectrics. Processes for stacking of 3D flash.
Materials and processes that extend lithography to sub-10 nm dimensions with reproducible properties	Novel resists to extend 193 nm lithography and support EUV lithography. Directed self-assembly (DSA) with materials such as block-copolymers to potentially extend lithography though pattern rectification and pattern density multiplication.
Materials for heterogeneous integration of multi-chip, multi-function packages.	Materials to modify polymer properties to enable increased product reliability. Novel electrical attaching materials to allow lower assembly temperatures and improved product reliability. Simultaneously achieve package polymer CTE, modulus, electrical and thermal properties, with moisture and ion diffusion barriers. Nanosolders compatible with <200C assembly, multiple reflows, high strength, and high electromigration resistance. Nanoinks that can be printed as die attach adhesives with required electrical, mechanical, thermal, and reliability properties.

6.2.2. LONG-TERM CHALLENGES

Table EMI2 Long-term Difficult Challenges

Long-term Challenges: 2029–2037	***Description***
Materials and processes that achieve 3D monolithic and vertical integration of high mobility and steep subthreshold transistors	Processes for sequential 3D vertical integration of transistors. Methods to lower the synthesis temperature of vertical semiconductor nanowires. Methods to dope and contact vertical semiconductor nanowire transistors. Lithography-free and low-temperature methods to achieve gate stack on vertical transistors.
Materials and processes that replace copper interconnects with improved reliability and electromagnetic performance at the nanoscale	Synthesis or assembly of CNTs in predefined locations and directions with controlled diameters, chirality and site-density. Carbon and collective excitations. Novel interlayer dielectrics: Metal Organic Framework (MOF) and Carbon Organic Framework (COF). Metals with less size effects such as silicides.
Materials and processes for charge-based and non-charge-based beyond CMOS logic that replaces or extends CMOS	Achieving a bandgap and full interfaces control in graphene in FET structures and alternative FETs (TFETs etc). Synthesis of CNTs with tight distribution of bandgap and mobility. Complex metal oxides with low defect density. High mobility transition metal dichalcogenides with low defect density and low resistance ohmic contacts. Spin materials: characterization of spin, magnetic and electrical properties and correlation to nanostructure. Topological materials: large bandgaps much greater that kT at room temperature, ability to modulate bandgap efficiently with electric field. BiSFET heterostructures: achieving exciton condensation at room temperature.
Materials and processes for emerging memory and select devices to replace DRAM/NVM.	Multiferroic with Curie temperature >400K and high remnant magnetization to >400K. Ferromagnetic semiconductor with Curie temperature >400K. Complex Oxides: Control of oxygen vacancy formation at metal interfaces and interactions of electrodes with oxygen and vacancies. Switching mechanism of atomic switch: Improvements in switching speed, cyclic endurance, uniformity of the switching bias voltage and resistances both for the on-state and the off-state.
Materials and processes that enable monolithically 3D integrated complex functionality including thermal and yield challenges	Integration on CMOS Platforms. Integration with flexible electronics. Biocompatible functional materials. Leveraging convergent materials expertise in adjacent sectors, including More than Moore functionalities (photonics, optics/metamaterials, outside connectivity, energy transfer/storage, power circuits).

6.3. TECHNOLOGY REQUIREMENTS AND POTENTIAL SOLUTIONS

6.3.1. SUMMARY

The IRDS seeks a framework for managing the convergence of scaled information processing and storage, i.e., More Moore (MM) and Beyond CMOS (BC), with the next emerging era of monolithically integrated systems that achieve enhanced overall functional density. The trend towards the convergence of monolithically integrated functional diversification with miniaturization manifests as increasing complexity in the road-mapping process. The IRDS reflects this growing complexity, with an increasing number of projected roadmap parameters and requirements associated with new functionalities. While EMI continues to support the evolutionary, and semiconductor centric needs of the traditional semiconductor community, emerging architectures could

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

benefit from new device functionality, which may require new materials and new physical mechanisms. New waves of emerging materials technologies may represent potentially disruptive opportunities.

Candidate EMI materials and processes exhibit unique and useful properties that may require atomic level structural, interface, defect and compositional control. In some cases, current synthetic or manufacturing technologies are not yet capable of producing such materials with the required level of control. The difficulties could be due to: 1) The inability of a research environment to produce materials with the required level of control that would express the desired properties; or 2) scaling up the synthetic and fabrication processes to satisfy commercial manufacturing requirements. In some cases, current materials growth processes effect unacceptable levels of defect formation, which drive the need for new and more robust fabrication methods. In other cases, synthetic methods exist for producing high quality materials, but these processes cannot be scaled to the higher growth rates, yields, or purity needed for insertion into viable commercial applications. While these materials may provide proof of concept and suggest a potential solution, new cost effective fabrication technologies may be required to warrant a candidate material's insertion into high volume manufacturing.

6.3.2. SCALED TECHNOLOGY MATERIALS FOR MORE MOORE

As described in the More Moore chapter, after 2027 there is no headroom for 2D geometry scaling and 3D VLSI integration of circuits and systems using sequential/stacked integration approaches will likely begin. Whether one is considering 2D geometry scaling or 3D integration, there are numerous materials challenges to achieving increasing device density and integrated performance. The following outlines key materials challenges for transistor scaling and integration, lithography, interconnects, heterogenous integration, assembly and packaging, and outside system connectivity.

6.3.2.1. MATERIALS FOR TRANSISTOR SCALING AND INTEGRATION

Continued increases in transistor device density require a variety of new materials and processes including new channels (Ge, III-V), improved doping techniques, gate stacks and contacting structures. Table EMI3 provides a set of materials and processes priorities for transistor scaling and integration.

Table EMI3 Materials for Transistor Scaling and Integration

6.3.2.2. MATERIALS FOR LITHOGRAPHY AND PATTERNING

The future of scaled technologies depends upon emerging patterning materials (resist or self-assembled) to enable extensible lithographic capabilities. New resist materials must concurrently exhibit higher resolution, higher sensitivity, reduced line edge roughness, and sufficient etch resistance to enable robust pattern transfer. 193nm and EUV extension materials are being developed, which can improve LWR, pattern shrink materials, and topcoats for EUV to ameliorate issues with out-of-band optical flare and outgassing. Evolutionary approaches for enhancing positive, negative, and chemically amplified families of resists will continue to be evaluated. Leading process approaches to pitch division include multiple patterning (MP) and spacer patterning (SP) as options for extending 193 nm immersion lithography. Alternate technologies are utilizing patterning materials to create guide patterns for directed self-assembly, which can include resists to form chemoepitaxy and graphoepitaxy guides, or directly patternable brushes and SAMs. Directed self-assembly (DSA) with block-copolymers or polymer pairs has made significant progress in characterizing sources of defect formation and in applications such as contact rectification, fin patterning, and pattern density multiplication. Table EMI4 provides a set of materials and processes priorities for lithography and patterning.

Table EMI4 Materials for Lithography and Patterning

6.3.2.3. INTERCONNECT MATERIALS

Key challenges for continued increased performance of future integrated circuit interconnects consist of maintaining reductions of RC time constants for delivery of signals and power with high reliability. For copper interconnects, the sidewall copper barrier thickness must continue to be reduced, which is a significant challenge. For post copper interconnect scaling, novel interconnects, such as carbon nanotubes, are being explored. Several elemental metals have been studied by simulation and experiments to identify potential candidates for post-Cu interconnect materials at highly scaled dimension.[1200] The interconnect material space becomes significantly larger when alloys and compounds are considered, and machine learning assisted data analysis has been employed to filter this vast material space. Also, lower dielectric constant (κ for both intra and inter level dielectric are needed; however, each of these emerging families of materials must overcome significant challenges for them to warrant adoption. Airgap, another approach to reducing the effective κ, places additional requirements on barrier layers or novel interconnects. Table EMI5 provides a set of materials and processes priorities for interconnects.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Table EMI5 Interconnect Materials

6.3.2.4. HETEROGENEOUS INTEGRATION, ASSEMBLY AND PACKAGING MATERIALS

The EMI and Heterogeneous Integration teams are in the process of prioritizing key heterogeneous integration and assembly & packaging EMI challenges, which include:

- New engineered materials: substrate, mold, underfill, wafer bond alloys, solder alloys
- Conductors: Nanomaterials (CNT, graphene, NWs), metals (Cu, Al, W, Ag, etc.), composites
- Dielectrics: Oxides, polymers, porous materials, composites
- Semiconductors: Elemental (Si, Ge), Compounds (SiC, III-V, II-VI, tertiary), polymers
- Critical factors: Cost, CTE differential, thermal conductivity, fracture toughness, modulus, processing temperature, interfacial adhesion, operating temperature, and breakdown field strength

Table EMI6 provides a set of heterogeneous integration and assembly and packaging priorities for EMI.

Table EMI6 Heterogeneous Integration, Assembly and Packaging Materials

6.3.2.5. MATERIALS CHALLENGES FOR OUTSIDE SYSTEM CONNECTIVITY

Table EMI7 provides a set of top Outside System Connectivity material priorities for EMI.

Table EMI7 Emerging Research Materials Needs for Outside System Connectivity

6.3.3. EMERGING MATERIALS FOR MEMORY, BEYOND CMOS LOGIC AND COMPUTING

Beyond 2030, MOSFET scaling will likely become ineffective and/or very costly. As described in this chapter, completely new, non-CMOS types of memory, logic devices and maybe even new circuit architectures are potential solutions. Such solutions ideally can be integrated onto the Si-based platform to take advantage of the established processing infrastructure, as well as being able to include Si devices such as memories onto the same chip. The following outlines key materials challenges for emerging materials for memory, beyond CMOS logic and alternative information processing.

6.3.3.1. EMERGING MATERIALS FOR MEMORY

Emerging memory devices includes capacitive memories (Fe FET), and resistive memories including ferroelectric devices, resistance change devices, devices based on Mott transitions and novel magnetic memories. Another key requirement for memory technology is the development of corresponding select devices that access only the selected memory cell of interest without perturbing non-selected cells. Table EMI8 provides a set of materials and associated challenges for emerging memory materials, and Table EMI9 provides materials and associated challenges for memory select.

Table EMI8 Emerging Materials for Memory

Table EMI9 Emerging Materials for Memory Select

6.3.3.2. EMERGING MATERIALS FOR ADVANCED AND BEYOND-CMOS LOGIC DEVICES

There are generally two classes of devices/materials for advanced and beyond-CMOS logic devices. The first are those that do not involve spin or magnetism such as ferroelectric FETs, nanoelectromechanical (NEM) switches, topological FETs, and transistors based on collective electron phenomena such as Mott effect or exciton condensation (BiSFET). The second are those based on spin and magnetism that each uses a variety of materials. Negative-capacitance topological quantum FET was proposed to achieve extremely low switching voltages and energies.[1201] Table EMI10 contains materials and associated challenges for the non-spin devices. Table EMI11 maps various spin device concepts to associated materials types and Table EMI12 describes the requirements of these materials.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

Table EMI10 Emerging Materials for Advanced and Beyond-CMOS Logic Devices

Table EMI11 Spin Devices versus Materials

Table EMI12 Spin Material Requirements and Properties

6.3.3.3. EMERGING MATERIALS FOR NOVEL COMPUTING

Emerging materials spur developments of novel computing, such as neuromorphic computing, reinforcement learning, topological quantum computing, and reversible computing, and probabilistic computing. Since the performance of the computing is considered to be dependent on the intrinsic properties of the emerging material, the material research will further boost the performance, such as the energy efficiency [1202,1203,1204,1205,1206,1207]. Two-dimensional materials are also expected to play an important role in memristors for neuromorphic computing.[1208, 1209]

6.3.4. METROLOGY NEEDS AND CHALLENGES FOR EMERGING RESEARCH MATERIALS

Metrology is needed to characterize composition, properties, and understand structure of emerging research materials, at nanometer dimensions and below. The difficult EMI metrology challenges would be those associated with the introduction of directed self-assembly (DSA), such as evaluating critical material properties, size and location of features, registration, and defects. Also needed are non-destructive methods for characterizing embedded materials and interfaces defects in nano-scale devices,[1210] as well as platforms that enable simultaneous measurement of complex nanoscopic properties. Table EMI13 summarizes the current set of continuing and prioritized metrology related EMI challenges and needs.

Table EMI13 Metrology Needs and Challenges for Emerging Research Materials

6.4. EMERGING/DISRUPTIVE CONCEPTS AND TECHNOLOGIES

As mentioned at the beginning of the chapter, new system integrators, from mobile to data centers to the Internet of Everything, have appeared with new and complex technology requirements. These application domains require a highly interdisciplinary set of expertise, e.g., electrical and mechanical engineering; as well as materials, biological, medical, energy, aerospace, transportation, communication, and sustainability sciences. The trend towards the convergence of monolithically integrated functional diversification with miniaturization manifests as increasing complexity in the road-mapping process. Collaborative transdisciplinary research is needed to identify materials and processes that catalyze breakthrough and convergent advances in these technologies. Initiatives that leverage the expertise of colleagues in adjacent spaces who know the local environment, e.g., biology, energy, etc., will help to drive novel approaches and more optimal materials, process, manufacturing, and performance solutions to emerging IoT challenges than can be achieved by semiconductor centric approaches. As examples, (6.4.1) transient concept and (6.4.2) development aided by machine learning (Table EMI14) are shown. Table EMI15 identifies several emerging application opportunities that will drive and enhance future EMI working group activities.

6.4.1. TRANSIENT ELECTRONICS

Transient electronics is an emerging field that requires materials, devices, and systems to be capable of disappearing with minimal or non-traceable remains in a controllable period of time. The spontaneous and transient function appears after the stable operation. This emerging electronics with the disintegrating capability will bring about intelligent applications in various scenes, such as bioelectronics, environmentally friendly electronics [1211,1212,1213,1214]. The conductance change with the controllable decay characteristics was demonstrated in molecular gap atomic switches [1215,1216].

6.4.2. MODELING AND SIMULATION FOR EMERGING MATERIALS DEVELOPMENTS

Simulation for designing material structures matching to targeted performance needs hierarchical understanding of material from nanometer scale to centimeter scale. First-principles approaches are developed to provide theoretical insight into emerging materials and in silico exploration of as-yet-unreported materials.[1217] Oppositely, despite its importance for the future design of device structures, the development of simulation for emerging devices is just beginning.[1218] This requires multiscale understanding in space and time for phenomena in materials. For instance, the individual atom-atom interaction derived from chemical bonding given by details of electronic structures can be scaled up to mechanical response of many atoms against macroscopic load, which gives understanding of elastic behavior of crystals[1219]. Meanwhile, a performance of fuel cell can be

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

derived from macro-model originated from the rate equation of chemical reactions of individual molecules[1220]. Multiscale view of carrier transport in organic devices were treated by considering molecular orbital levels to mesoscopic scale of electron-hopping [1221,1222,1223]. Beside the direct approach on multiscale physics, aid of machine learning can bridge the difference scales phenomena in composite material [1224]. These approaches will accelerate the simulation, while proper modeling with understanding of natural physics in difference scales of time and space are still necessary. Finite element analysis (FEA) models will have to be further developed so that they can adequately represent 2D and other nanomaterials.

Table EMI14 Modeling and Simulation

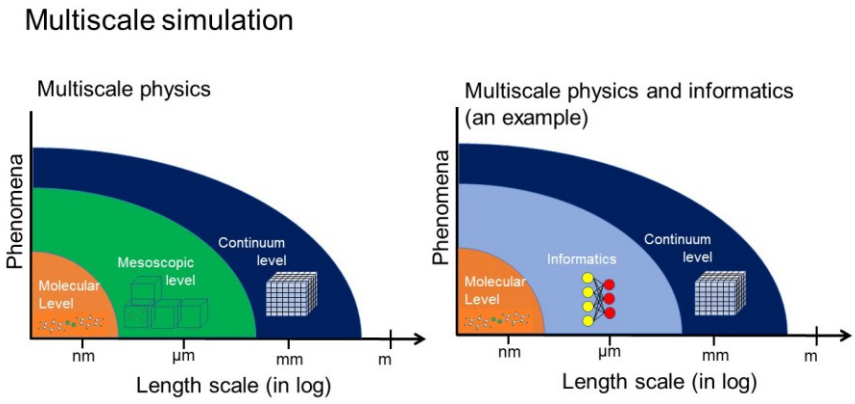


Figure EMI2 An Example of the Role of Machine Learning in the Multiscale Simulation

Note: The left shows conventional multiscale physics where physics to explain mesoscopic level phenomena is required to link molecular and continuum level phenomena. The right shows a potential role of machine learning that provides output of continuum level phenomena from inputs of molecular level phenomena.

Table EMI15 Summary of Potentially Disruptive Emerging Research Materials Application Opportunities

6.5. Conclusions and Recommendations

The IRDS represents a strategic repositioning of the community's scope, needs, and set of emergent opportunities. In alignment with this new perspective, this edition of the EMI chapter represents a work in transition that has aligned difficult challenges with the needs of related IRDS working groups. Much of the information in the detailed tables comes from prior ERM chapters and current IRDS working groups. Future editions of EMI will provide additional detailed descriptions and continue to adapt its scope to engage with a new set of EMIs, many of which will be identified by the IRDS working groups.

7. Assessment

7.1. Introduction

It is important to evaluate beyond-CMOS devices considered in this chapter against current CMOS technologies. Two methods of assessments have been reported: a "survey-based assessment" conducted previously by the ITRS ERD working group and a "quantitative device benchmarking" conducted by the Nanoelectronics Research Initiative (NRI).

Up to the 2013 ERD Chapter, a survey-based critical review was conducted based on eight criteria to compare emerging devices against CMOS. Spider chart has been used to visualize the perceived potential of these technology entries. However, the limited number of survey results sometimes raises questions of the accuracy of this survey. The most recent "survey-based assessment" was conducted in the 2014 ERD Emerging Memory and Logic Device Assessment Workshops (Albuquerque, NM). The survey collects voting on emerging technologies evaluated in the workshops in the categories of the "most promising" and the "most need of resources" to assess the potential of these technology entries perceived by ERD experts. A summary of previous survey-based assessments is included in section 7.3 "Archive".

Copyright © 2022 IEEE. All rights reserved.

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

Most of the proposed beyond-CMOS devices are very different from their CMOS counterparts, and often pass computational state variables (or tokens) other than charge. Alternative state variables include collective or single spins, excitons, plasmons, photons, magnetic domains, polarization, etc. 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 metrics used for CMOS and new metrics which take into consideration the idiosyncrasies of the new device behavior. Even more challenging is to extend this process to consider new circuits and architectures beyond the Boolean logic used by CMOS today, which may enable these devices to complete transactions more effectively.

7.1.1. Architectural Requirements for a Competitive Logic Device

The circuit designers and architects depend on the logic switch to exhibit specific desired characteristics in order to successfully realize a wide range of applications. These characteristics,[1225] which have since been supplemented in the literature, include:

- Inversion and flexibility (can form an infinite number of logic functions)
- Isolation (output does not affect input)
- Logic gain (output may drive more than one following gate and provides a high I_{on}/I_{off} ratio)
- Logical completeness (the device is capable of realizing any arbitrary logic function)
- Self-restoring / stable (signal quality restored in each gate)
- Low cost manufacturability (acceptable process tolerance)
- Reliability (aging, wear-out, radiation immunity)
- Performance (transaction throughput improvement)
- Span of control (measures number of devices that may be reached within a characteristic delay of the switch[1226])

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[1227] are particularly valuable.

7.2. NRI Beyond-CMOS Benchmarking

The Nanoelectronics Research Initiative (https://www.src.org/program/nri/) has quantitatively evaluated quite a few beyond-CMOS technologies.[1228, 1229,1230] Several NRI devices have been described in detail in the Logic and Emerging Information Processing Device Section. While beyond-CMOS device benchmarking is still very much a work in progress – and no concrete decisions have been made on which devices should be chosen or eliminated as candidates to significantly extend or augment CMOS – this section summarizes some of the data and insights gained from these studies. Further benchmarks may alter some of the conclusions here and the outlook on some of these devices, but the overall message on the challenges of beyond-CMOS devices remains valid.

7.2.1. Quantitative Results

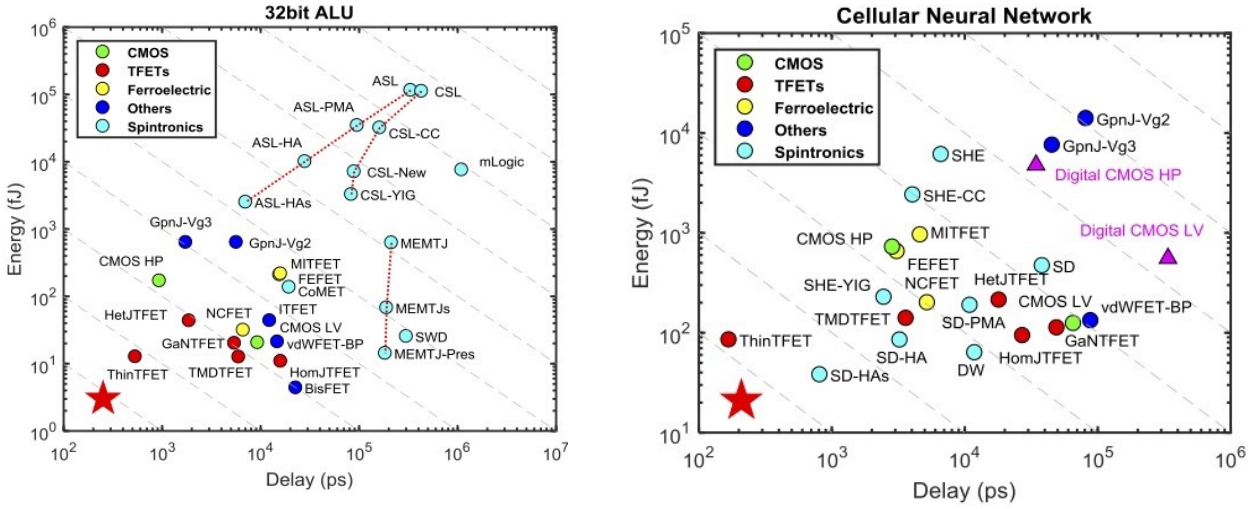
NRI benchmarking analyzes the potential of major emerging switches using a variety of information tokens and communication transport mechanisms. Specifically, the projected effectiveness of these devices used in a number of logic gate configurations was evaluated and normalized to CMOS at the 5nm generation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

At a high level, the data from these studies corroborates qualitative insights from earlier works, suggesting that many new logic switch structures may have some advantages over CMOS in terms of power or energy, but they are also inferior to CMOS in delay. This is perhaps not surprising; the primary goal for nanoelectronics and NRI is to find a lower power device[1231] since power density is a primary concern for future CMOS scaling. The power-speed tradeoffs commonly observed in CMOS also extend to beyond-CMOS devices. It is 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

Copyright © 2022 IEEE. All rights reserved.

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.

A simplified 32-bit arithmetic logic unit (ALU) was built from these devices to evaluate their performance, as shown in Figure BC7.1(a) [1232]. While tunneling devices show limited advantages over CMOS in terms of energy-delay product, most beyond-CMOS devices are inferior to CMOS in energy and/or delay. For example, spintronic devices tend to be slower than CMOS and also show no energy advantage.



(a) (b)

Figure BC7.1 (a) Energy versus Delay of a 32-bit ALU for a Variety of Charge- and Spin-based Devices; (b) Energy versus Delay per Memory Association Operation Using Cellular Neural Network (CNN) for a Variety of Charge- and Spin-based Devices[1232]

At the architecture level, the ability to speculate on how these devices will perform is still in its infancy. While the ultimate goal is to compare at a very high level—e.g., how many MIPS can be produced for 100 mW on 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"[1233] for these technologies, a figure of merit that ties fundamental technology to a resulting logic transaction. Several of the devices appear to offer advantage over CMOS in logical effort, particular for more complex functions, which increases the urgency of doing more joint device-architecture co-design for these emerging technologies.

The direction of device-architecture co-optimization has driven NRI benchmarking to explore non-Boolean applications of beyond-CMOS devices. Cellular neural network (CNN) has been utilized as a benchmarking model that has been implemented with various novel devices.[1232] The energy and delay of CNN based on beyond-CMOS devices are compared with CMOS-based CNN in Figure BC7.1(b). Tunneling devices have significant performance improvement because of their steep subthreshold slopes and large driving current at ultra-low supply voltage. Interestingly, spintronic devices are much closer to the preferred corner in CNN implementation in comparison with 32-bit ALU. This is because some characteristics of spintronic devices (e.g., spin diffusion, domain wall motion) may mimic the functionality of a neuron (e.g., integration) more naturally in a single device.

7.2.2. OBSERVATIONS

A number of common themes have emerged from these benchmarking studies and in the observations made during recent studies of beyond-CMOS devices.

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 achieve sufficient throughput with "slow" devices must be developed.
2. Most of the architectures that have been considered to date in the context of new devices utilize binary logic to implement von Neumann computing structures. In this area, CMOS implementations are difficult to supplant because they are very competitive across the spectrum of energy, delay and area – not surprising since these architectures have evolved over several decades to exploit the properties of CMOS most effectively. Novel electron-based devices—which can include

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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. 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.
5. Assessment of novel architectures using new switches must also include the transport mechanism for the information tokens. Fundamental relationships connecting information generation with information communication spatially and temporally will dictate CMOS' successor.

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

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 advantageous 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, performance, and power constraints) and the changing landscape of workloads and processing needs (e.g., big data, unstructured information, real-time computing, 3D rich graphics) are increasing the need for new computing solutions. Therefore, future beyond-CMOS research should focus on specific emerging functions and device-architecture co-optimization to achieve solutions that can break through the current power/performance limits.

7.3. ARCHIVE

The survey-based emerging device assessment has not been continued after the 2015 International Technology Roadmap for Semiconductors (ITRS) Emerging Research Devices (ERD) chapter. Previous survey-based assessments are summarized here for references.

7.3.1. EMERGING DEVICE ASSESSMENT IN 2014 ERD WORKSHOPS

In August 2014, ERD organized an "Emerging Memory Device Assessment Workshop" and an "Emerging Logic Device Assessment Workshop", where nine memory devices and fourteen logic devices were evaluated. A survey was conducted in the workshops for the experts to vote on the "most promising" devices and devices "needing more resources". Figure BC7.2 shows the relative number of votes received by emerging devices in these two categories, ranked from high to low in the "most promising" category (red color bars).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

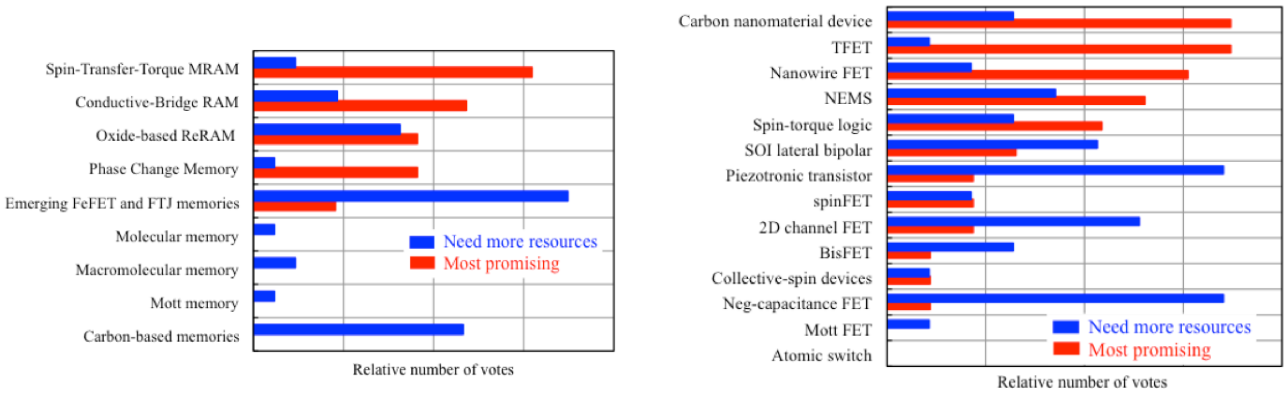


Figure BC7.2 (a) Survey of Emerging Memory Devices and (b) Survey of Emerging Logic Devices in 2014 ERD Emerging Logic Workshop (Albuquerque, NM)

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

7.3.2. 2013 ERD Survey Criteria, Methodology, and Results

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

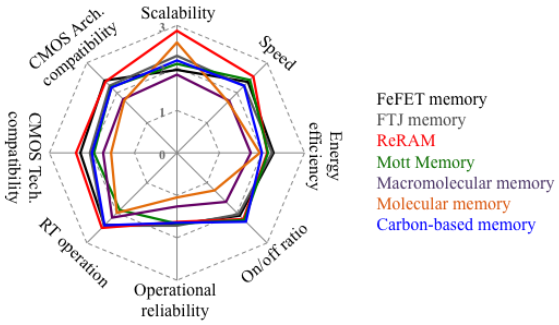


Figure BC7.3 Comparison of Emerging Memory Devices Based on 2013 Critical Review

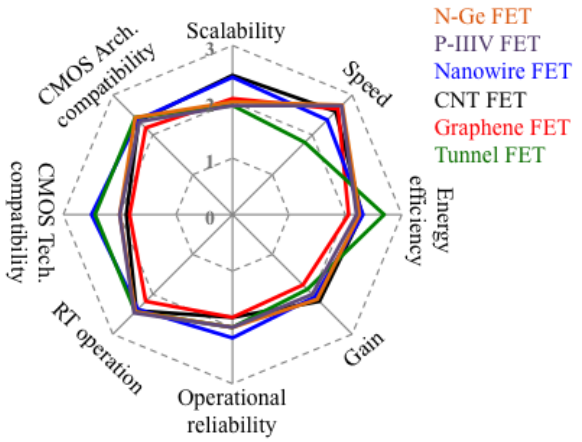
Each CMOS extension and beyond-CMOS emerging memory and logic device technology is evaluated against these criteria according to a single factor. For logic, this factor relates to the *projected potential performance* of a nanoscale device technology, assuming its successful development to maturity, *compared to that for silicon CMOS scaled to the end of the Roadmap.* For memory, this factor relates the *projected potential performance* of each nanoscale memory device technology, assuming its successful development to maturity, *compared to that for ultimately scaled 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-

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

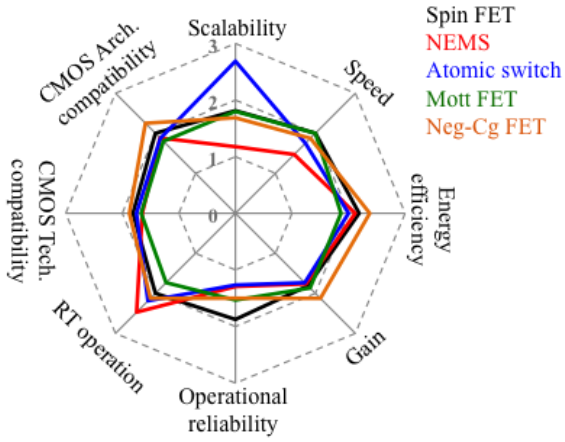
scaled CMOS, and "1" substantially inferior to CMOS or a comparable existing memory technology. This evaluation is determined by a survey of the ERD Working Group members composed of individuals representing a broad range of technical backgrounds and expertise. Details of the assessment results can be found in the 2013 ERD chapter.

Although this survey-based critical review has been conducted in ERD for several versions and has been widely cited in literatures, the decreasing number of votes of some less popular devices has raised concerns about the accuracy of some of the results. Figures BC7.3 and BC7.4 summarize the last critical review conducted in 2013 for emerging memory devices and emerging logic devices, respectively. Notice that the technology entries in these figures are based on the 2013 ERD chapter, while some of them have been removed in this chapter (e.g., molecular memory, atomic switch, etc.).

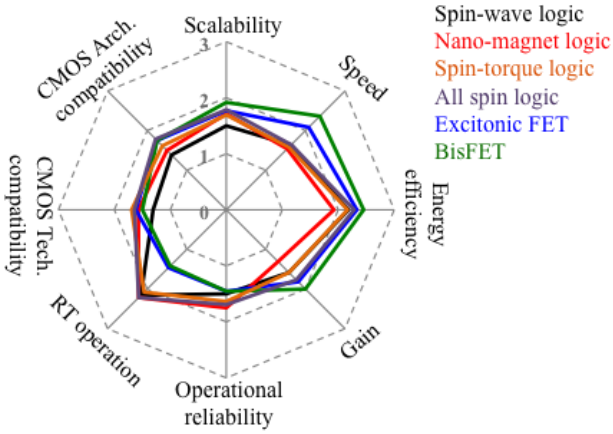
Since "3" represents the best result and "1" the worst in the spider chart, devices with larger circle area represent more promising devices. In Figure BC7.4 for emerging logic devices, the perceived potential of "beyond-CMOS devices" is generally poorer than "CMOS-extension devices". Within "beyond-CMOS devices", "non-charge-based devices" area perceived slightly less promising than "charge-based devices". The general trend is consistent with the quantitative NRI assessment in section 7.2. Multiple factors contribute to this result, including the strength of CMOS as a platform technology, the challenges of beyond-CMOS devices in materials and fabrication, the lack of memory and interconnect solutions for beyond-CMOS devices, *etc*.



(a) CMOS extension devices



(b) Charge-based beyond-CMOS devices



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

Figure BC7.4 Comparison of Emerging Logic Devices Based on 2013 ITRS ERD Critical Review: (a) CMOS Extension Devices; (b) Charge-based Beyond-CMOS Devices; (c) Non-charge-based Beyond-CMOS Devices

8. SUMMARY

The "Beyond CMOS" chapter systematically surveys emerging memory and logic devices (sections 2 and 3), novel technologies (section 4), and alternative architectures and computing paradigms (section 5), to explore potential solutions beyond the

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

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

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

- **Computational State Variable(s) other than Solely Electron Charge**

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

- **Non-thermal Equilibrium Systems**

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

- **Novel Energy Transfer Interactions**

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

- **Nanoscale Thermal Management**

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

- **Sub-lithographic Manufacturing Process**

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

- **Alternative Architectures**

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

9. 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] K. Lee and S. H. Kang, "Development of Embedded STT-MRAM for Mobile System-on-Chips," *Ieee Transactions on Magnetics,* vol. 47, pp. 131-136, Jan 2011.

[3] S. Mangin, D. Ravelosona, J. A. Katine, M. J. Carey, B. D. Terris, and E. E. Fullerton, "Current-induced magnetization reversal in nanopillars with perpendicular anisotropy," *Nature Materials,* vol. 5, pp. 210-215, 2006.

[4] D. Saida, S. Kashiwada, M. Yakabe, T. Daibou, M. Fukumoto, S. Miwa, Y. Suzuki, K. Abe, H. Noguchi, J. Ito, and S. Fujita, "1x-to 2x-nm perpendicular MTJ Switching at Sub-3-ns Pulses Below 100 mu A for High-Performance Embedded STT-MRAM for Sub-20-nm CMOS," *Ieee Transactions on Electron Devices,* vol. 64, pp. 427-431, Feb 2017.

[5] G. Jan, L. Thomas, S. Le, Y. Lee, H. Liu, J. Zhu, J. Iwata-Harms, S. Patel, R. Tong, V. Sundar, S. Serrano-Guisan, D. Shen, R. He, J. Haq, Z. J. Teng, V. Lam, Y. Yang, Y. Wang, T. Zhong, H. Fukuzawa, and P. Wang, "Demonstration of Ultra-Low Voltage and Ultra Low Power STT-MRAM designed for compatibility with 0x node embedded LLC applications," in *2018 IEEE Symposium on VLSI Technology*, 2018, pp. 65-66.

[6] S. Sakhare, M. Perumkunnil, T. H. Bao, S. Rao, W. Kim, D. Crotti, F. Yasin, S. Couet, J. Swerts, S. Kundu, D. Yakimets, R. Baert, H. Oh, A. Spessot, A. Mocuta, G. S. Kar, and A. Furnemont, "Enablement of STT-MRAM as last level cache for the high performance computing domain at the 5nm node," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 18.3.1-18.3.4.

[7] L. Thomas, G. Jan, S. Serrano-Guisan, H. Liu, J. Zhu, Y. Lee, S. Le, J. Iwata-Harms, R. Tong, S. Patel, V. Sundar, D. Shen, Y. Yang, R. He, J. Haq, Z. Teng, V. Lam, P. Liu, Y. Wang, T. Zhong, H. Fukuzawa, and P. Wang, "STT-MRAM devices with low damping and moment optimized for LLC applications at Ox nodes," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 27.3.1-27.3.4.

[8] G. Hu, J. H. Lee, J. J. Nowak, J. Z. Sun, J. Harms, A. Annunziata, S. Brown, W. Chen, Y. H. Kim, G. Lauer, L. Liu, N. Marchack, S. Murthy, E. J. O. Sullivan, J. H. Park, M. Reuter, R. P. Robertazzi, P. L. Trouilloud, Y. Zhu, and D. C. Worledge, "STT-MRAM with double magnetic tunnel junctions," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, pp. 26.3.1-26.3.4.

[9] H. Sato, M. Yamanouchi, S. Ikeda, S. Fukami, F. Matsukura, and H. Ohno, "Perpendicular-anisotropy CoFeB-MgO magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure," *Applied Physics Letters,* vol. 101, p. 022414, 2012.

[10] K. Watanabe, B. Jinnai, S. Fukami, H. Sato, and H. Ohno, "Shape anisotropy revisited in single-digit nanometer magnetic tunnel junctions," *Nature Communications,* vol. 9, p. 663, 2018/02/14 2018.

[11] M. M. S. Aly, M. Gao, G. Hills, C. S. Lee, G. Pitner, M. M. Shulaker, T. F. Wu, M. Asheghi, J. Bokor, F. Franchetti, K. E. Goodson, C. Kozyrakis, I. Markov, K. Olukotun, L. Pileggi, E. Pop, J. Rabaey, R. C, H. S. P. Wong, and S. Mitra, "Energy-Efficient Abundant-Data Computing: The N3XT 1,000x," *Computer,* vol. 48, pp. 24-33, 2015.

[12] S. Fujita, H. Noguchi, K. Ikegami, S. Takeda, K. Nomura, and K. Abe, "Novel memory hierarchy with e-STT-MRAM for near-future applications," in *2017 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, 2017, pp. 1-2.

[13] H. Yang, X. Hao, Z. Wang, R. Malmhall, H. Gan, K. Satoh, J. Zhang, D. H. Jung, X. Wang, Y. Zhou, B. K. Yen, and Y. Huai, "Threshold switching selector and 1S1R integration development for 3D cross-point STT-MRAM," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 38.1.1-38.1.4.

[14] S. Fukami, T. Anekawa, A. Ohkawara, Z. Chaoliang, and H. Ohno, "A sub-ns three-terminal spin-orbit torque induced switching device," in *2016 IEEE Symposium on VLSI Technology*, 2016, pp. 1-2.

[15] S. Shi, Y. Ou, S. V. Aradhya, D. C. Ralph, and R. A. Buhrman, "Fast Low-Current Spin-Orbit-Torque Switching of Magnetic Tunnel Junctions through Atomic Modifications of the Free-Layer Interfaces," *Physical Review Applied,* vol. 9, p. 011002, 01/30/ 2018.

[16] E. Kitagawa, S. Fujita, K. Nomura, H. Noguchi, K. Abe, K. Ikegami, T. Daibou, Y. Kato, C. Kamata, S. Kashiwada, N. Shimomura, J. Ito, and H. Yoda, "Impact of ultra low power and fast write operation of advanced perpendicular MTJ on power reduction for high-performance mobile CPU," *2012 Ieee International Electron Devices Meeting (Iedm),* 2012.

[17] L. Liu, C. F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin-torque switching with the giant spin Hall effect of tantalum," *Science,* vol. 336, pp. 555-8, May 4 2012.

[18] I. M. Miron, K. Garello, G. Gaudin, P. J. Zermatten, M. V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, and P. Gambardella, "Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection," *Nature,* vol. 476, pp. 189-93, Aug 11 2011.

[19] A. R. Mellnik, J. S. Lee, A. Richardella, J. L. Grab, P. J. Mintun, M. H. Fischer, A. Vaezi, A. Manchon, E. A. Kim, N. Samarth, and D. C. Ralph, "Spin-transfer torque generated by a topological insulator," *Nature,* vol. 511, pp. 449-51, Jul 24 2014.

[20] K. Garello, F. Yasin, H. Hody, S. Couet, L. Souriau, S. H. Sharifi, J. Swerts, R. Carpenter, S. Rao, W. Kim, J. Wu, K. K. V. Sethu, M. Pak, N. Jossart, D. Crotti, A. Furnémont, and G. S. Kar, "Manufacturable 300mm platform solution for Field-Free Switching SOT-MRAM," in *2019 Symposium on VLSI Circuits*, 2019, pp. T194-T195.

[21] L. J. Zhu, D. C. Ralph, and R. A. Buhrman, "Highly Efficient Spin-Current Generation by the Spin Hall Effect in Au1-xPtx," *Physical Review Applied,* vol. 10, Sep 6 2018.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[22] Y. Fan, P. Upadhyaya, X. Kou, M. Lang, S. Takei, Z. Wang, J. Tang, L. He, L. T. Chang, M. Montazeri, G. Yu, W. Jiang, T. Nie, R. N. Schwartz, Y. Tserkovnyak, and K. L. Wang, "Magnetization switching through giant spin-orbit torque in a magnetically doped topological insulator heterostructure," *Nat Mater,* vol. 13, pp. 699-704, Jul 2014.

[23] S. Shi, S. Liang, Z. Zhu, K. Cai, S. D. Pollard, Y. Wang, J. Wang, Q. Wang, P. He, J. Yu, G. Eda, G. Liang, and H. Yang, "All-electric magnetization switching and Dzyaloshinskii-Moriya interaction in WTe2/ferromagnet heterostructures," *Nat Nanotechnol,* vol. 14, pp. 945-949, Oct 2019.

[24] S. Fukami, C. Zhang, S. DuttaGupta, A. Kurenkov, and H. Ohno, "Magnetization switching by spin-orbit torque in an antiferromagnet-ferromagnet bilayer system," *Nat Mater,* vol. 15, pp. 535-41, May 2016.

[25] P. Wadley, B. Howells, J. Železný, C. Andrews, V. Hills, R. P. Campion, V. Novák, K. Olejník, F. Maccherozzi, S. S. Dhesi, S. Y. Martin, T. Wagner, J. Wunderlich, F. Freimuth, Y. Mokrousov, J. Kuneš, J. S. Chauhan, M. J. Grzybowski, A. W. Rushforth, K. W. Edmonds, B. L. Gallagher, and T. Jungwirth, "Electrical switching of an antiferromagnet," *Science,* vol. 351, pp. 587-590, 2016.

[26] Y. Wang, R. Ramaswamy, M. Motapothula, K. Narayanapillai, D. Zhu, J. Yu, T. Venkatesan, and H. Yang, "Room-Temperature Giant Charge-to-Spin Conversion at the SrTiO3–LaAlO3 Oxide Interface," *Nano Letters,* vol. 17, pp. 7659-7664, 2017/12/13 2017.

[27] L. Liu, Q. Qin, W. Lin, C. Li, Q. Xie, S. He, X. Shu, C. Zhou, Z. Lim, J. Yu, W. Lu, M. Li, X. Yan, S. J. Pennycook, and J. Chen, "Current-induced magnetization switching in all-oxide heterostructures," *Nature Nanotechnology,* 2019/09/09 2019.

[28] S. Fukami, T. Anekawa, C. Zhang, and H. Ohno, "A spin–orbit torque switching scheme with collinear magnetic easy axis and current configuration," *Nature Nanotechnology,* vol. 11, p. 621, 03/21/online 2016.

[29] F. Oboril, R. Bishnoi, M. Ebrahimi, and M. B. Tahoori, "Evaluation of Hybrid Memory Technologies Using SOT-MRAM for On-Chip Cache Hierarchy," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,* vol. 34, pp. 367-380, 2015.

[30] N. Sato, F. Xue, R. M. White, C. Bi, and S. X. Wang, "Two-terminal spin–orbit torque magnetoresistive random access memory," *Nature Electronics,* vol. 1, pp. 508-511, 2018/09/01 2018.

[31] H. Yoda, N. Shimomura, Y. Ohsawa, S. Shirotori, Y. Kato, T. Inokuchi, Y. Kamiguchi, B. Altansargai, Y. Saito, K. Koi, H. Sugiyama, S. Oikawa, M. Shimizu, M. Ishikawa, K. Ikegami, and A. Kurobe, "Voltage-control spintronics memory (VoCSM) having potentials of ultra-low energy-consumption and high-density," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 27.6.1-27.6.4.

[32] S. V. Aradhya, G. E. Rowlands, J. Oh, D. C. Ralph, and R. A. Buhrman, "Nanosecond-Timescale Low Energy Switching of In-Plane Magnetic Tunnel Junctions through Dynamic Oersted-Field-Assisted Spin Hall Effect," *Nano Letters,* vol. 16, pp. 5987-5992, Oct 2016.

[33] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L. T. Chang, Y. Jiang, M. Lang, J. Tang, Y. Wang, Y. Tserkovnyak, P. K. Amiri, and K. L. Wang, "Switching of perpendicular magnetization by spin-orbit torques in the absence of external magnetic fields," *Nat Nanotechnol,* vol. 9, pp. 548-54, Jul 2014.

[34] L. You, O. Lee, D. Bhowmik, D. Labanowski, J. Hong, J. Bokor, and S. Salahuddin, "Switching of perpendicularly polarized nanomagnets with spin orbit torque without an external magnetic field by engineering a tilted anisotropy," *Proceedings of the National Academy of Sciences,* vol. 112, pp. 10310-10315, 2015.

[35] Y.-C. Lau, D. Betto, K. Rode, J. M. D. Coey, and P. Stamenov, "Spin–orbit torque switching without an external field using interlayer exchange coupling," *Nature Nanotechnology,* vol. 11, p. 758, 05/30/online 2016.

[36] Z. Zhao, A. K. Smith, M. Jamali, and J.-p. Wang, "External-field-free spin Hall switching of perpendicular magnetic nanopillar with a dipole-coupled composite structure," *arXiv preprint arXiv:1603.09624,* 2016.

[37] D. MacNeill, G. M. Stiehl, M. H. D. Guimaraes, R. A. Buhrman, J. Park, and D. C. Ralph, "Control of spin–orbit torques through crystal symmetry in WTe2/ferromagnet bilayers," *Nature Physics,* vol. 13, p. 300, 11/07/online 2016.

[38] S. H. C. Baek, V. P. Amin, Y. W. Oh, G. Go, S. J. Lee, G. H. Lee, K. J. Kim, M. D. Stiles, B. G. Park, and K. J. Lee, "Spin currents and spin-orbit torques in ferromagnetic trilayers," *Nature Materials,* vol. 17, pp. 509-+, Jun 2018.

[39] A. M. Humphries, T. Wang, E. R. J. Edwards, S. R. Allen, J. M. Shaw, H. T. Nembach, J. Q. Xiao, T. J. Silva, and X. Fan, "Observation of spin-orbit effects with spin rotation symmetry," *Nature Communications,* vol. 8, p. 911, 2017/10/13 2017.

[40] T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami, Y. Ando, and Y. Suzuki, "Large voltage-induced magnetic anisotropy change in a few atomic layers of iron," *Nat Nanotechnol,* vol. 4, pp. 158-61, Mar 2009.

[41] W. G. Wang, M. Li, S. Hageman, and C. L. Chien, "Electric-field-assisted switching in magnetic tunnel junctions," *Nat Mater,* vol. 11, pp. 64-8, Jan 2012.

[42] C.-G. Duan, J. Velev, R. Sabirianov, Z. Zhu, J. Chu, S. Jaswal, and E. Tsymbal, "Surface Magnetoelectric Effect in Ferromagnetic Metal Films," *Physical Review Letters,* vol. 101, p. 137201, Sep 26 2008.

[43] K. Nakamura, R. Shimabukuro, Y. Fujiwara, T. Akiyama, T. Ito, and A. J. Freeman, "Giant Modification of the Magnetocrystalline Anisotropy in Transition-Metal Monolayers by an External Electric Field," *Physical Review Letters,* vol. 102, 2009.

[44] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, "Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses," *Nat Mater,* vol. 11, pp. 39-43, Jan 2012.

[45] C. Grezes, F. Ebrahimi, J. G. Alzate, X. Cai, J. A. Katine, J. Langer, B. Ocker, P. Khalili Amiri, and K. L. Wang, "Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product," *Applied Physics Letters,* vol. 108, p. 012403, 2016.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[46] T. Yamamoto, T. Nozaki, H. Imamura, Y. Shiota, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, Y. Suzuki, and S. Yuasa, "Improvement of write error rate in voltage-driven magnetization switching," *Journal of Physics D-Applied Physics,* vol. 52, Apr 17 2019.

[47] L. Liu, C.-F. Pai, D. Ralph, and R. Buhrman, "Gate voltage modulation of spin-Hall-torque-driven magnetic switching," *arXiv preprint arXiv:1209.0962,* 2012.

[48] T. Inokuchi, H. Yoda, K. Koi, N. Shimomura, Y. Ohsawa, Y. Kato, S. Shirotori, M. Shimizu, H. Sugiyama, S. Oikawa, B. Altansargai, and A. Kurobe, "Real-time observation of fast and highly reliable magnetization switching in Voltage-Control Spintronics Memory (VoCSM)," *Applied Physics Letters,* vol. 114, p. 192404, 2019.

[49] Y. Kato, H. Yoda, T. Inokuchi, M. Shimizu, Y. Ohsawa, K. Fujii, M. Yoshiki, S. Oikawa, S. Shirotori, K. Koi, N. Shimomura, B. Altansargai, H. Sugiyama, and A. Kurobe, "Low write current and strong durability in high-speed spintronics memory (spin-Hall MRAM and VoCSM) through development of a shunt-free design process and W spin-Hall electrode," *Journal of Magnetism and Magnetic Materials,* vol. 491, p. 165536, 2019/12/01/ 2019.

[50] X. Li, A. Lee, S. A. Razavi, H. Wu, and K. L. Wang, "Voltage-controlled magnetoelectric memory and logic devices," *MRS Bulletin,* vol. 43, pp. 970-977, 2018.

[51] C. Grezes, X. Li, K. Wong, F. Ebrahimi, P. Khalili Amiri, and K. L. Wang, "Voltage-controlled magnetic tunnel junctions with synthetic ferromagnet free layer sandwiched by asymmetric double MgO barriers," *Journal of Physics D: Applied Physics,* 2019/09/26 2019.

[52] T. Nozaki, A. Kozioł-Rachwał, W. Skowroński, V. Zayets, Y. Shiota, S. Tamaru, H. Kubota, A. Fukushima, S. Yuasa, and Y. Suzuki, "Large Voltage-Induced Changes in the Perpendicular Magnetic Anisotropy of an MgO-Based Tunnel Junction with an Ultrathin Fe Layer," *Physical Review Applied,* vol. 5, p. 044006, 04/15/ 2016.

[53] Y. Kato, H. Yoda, Y. Saito, S. Oikawa, K. Fujii, M. Yoshiki, K. Koi, H. Sugiyama, M. Ishikawa, T. Inokuchi, N. Shimomura, M. Shimizu, S. Shirotori, B. Altansargai, Y. Ohsawa, K. Ikegami, A. Tiwari, and A. Kurobe, "Giant voltage-controlled magnetic anisotropy effect in a crystallographically strained CoFe system," *Applied Physics Express,* vol. 11, p. 053007, 2018.

[54] X. Li, T. Sasaki, C. Grezes, D. Wu, K. L. Wong, C. Bi, P.-V. Ong, F. Ebrahimi, G. Yu, N. Kioussis, W. Wang, T. Ohkubo, P. Khalili Amiri, and K. L. Wang, "Predictive Materials Design of Magnetic Random-Access Memory Based on Nanoscale Atomic Structure and Element Distribution," *Nano Letters,* vol. 19, pp. 8621-8629, 2019/11/07 2019.

[55] C. Grezes, H. Lee, A. Lee, S. Wang, F. Ebrahimi, X. Li, K. Wong, J. A. Katine, B. Ocker, J. Langer, P. Gupta, P. Khalili Amiri, and K. L. Wang, "Write Error Rate and Read Disturbance in Electric-Field-Controlled Magnetic Random-Access Memory," *IEEE Magnetics Letters,* vol. 8, pp. 1-5, 2017.

[56] H. Lee, A. Lee, S. D. Wang, F. Ebrahimi, P. Gupta, P. K. Amiri, and K. L. Wang, "A Word Line Pulse Circuit Technique for Reliable Magnetoelectric Random Access Memory," Ieee Transactions on Very Large Scale Integration (Vlsi) Systems, vol. 25, pp. 2027-2034, Jul 2017.

[57] T. Ikeura, T. Nozaki, Y. Shiota, T. Yamamoto, H. Imamura, H. Kubota, A. Fukushima, Y. Suzuki, and S. Yuasa, "Reduction in the write error rate of voltage-induced dynamic magnetization switching using the reverse bias method," *Japanese Journal of Applied Physics,* vol. 57, p. 040311, 2018.

[58] 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,* vol. 104, p. 212406, 2014.

[59] 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

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

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

[62] 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.

[63] D. M. Nminibapiel *et al.*, "Characteristics of resistive memory read fluctuations in endurance cycling," *El. Dev. Letters*, vol. 38, p. 326, 2017.

[64] D. Veksler *et al.,* "Switching variability factors in compliance-free metal oxide RRAM," *2019 IEEE International Reliability Physics Symposium (IRPS)*, 1-5, 2019.

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

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

[67] 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.

[68] 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.

[69] G. Bersuker *et al.*, "Metal oxide RRAM switching mechanism based on conductive filament microscopic properties," *J. Appl. Phys.,* vol. 110, p. 125518, 2011.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[70] 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.

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

[72] G. Bersuker, D. C. Gilmer, D. Veksler, "Metal-oxide resistive random access memory (RRAM) technology: Material and operation details and ramifications" in Advances in Non-Volatile Memory and Storage Technology, pp. 35-102, 2019

[73] 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.

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

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

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

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

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

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

[80] 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.

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

[82] 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.

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

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

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

[86] G. Bersuker et al., "Toward reliable RRAM performance: macro- and micro-analysis of operation processes,"*J. Comput. Electronics*, vol. 16, pp. 1085–1094, 2017.

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

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

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

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

[91] 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.

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

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

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

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

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[97] U. Russo, *et al.*, "Conductive-filament switching analysis and self-accelerated thermal dissolution model for reset in NiO-based RRAM," *IEDM Tech. Dig.*, 2007, pp. 775–778.
[98] 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.
[99] 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.
[100] 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.
[101] YY. Chen *et al*. "Insights into Ni-filament formation in unipolar-switching Ni/HfO2/TiN resistive random access memory device," *Appl. Phys. Lett.* 100, 113513 (2012)
[102] T. Yanagida, *et al.*, "Scaling Effect on Unipolar and Bipolar Resistive Switching of Metal Oxides," *Sci. Rep.,* vol. 3, p. 1657, Apr. 2013.
[103] X. A. Tran, *et al.*, "High performance unipolar AlO_yHfO_xNi based RRAM compatible with Si diodes for 3D application," *VLSI Technology (VLSIT), 2011 Symposium on*, 2011, pp. 44–45.
[104] Y.-H. Tseng, *et al.*, "High density and ultra-small cell size of Contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits," *IEDM Tech. Dig.*, 2009, pp. 1–4.
[105] Y.-H. Tseng, *et al.*, "Electron trapping effect on the switching behavior of contact RRAM devices through random telegraph noise analysis," *IEDM Tech. Dig.*, 2010, pp. 28.5.1–28.5.4.
[106] W. C. Shen, *et al.*, "High-κ metal gate contact RRAM (CRRAM) in pure 28 nm CMOS logic process," *IEDM Tech. Dig.*, 2012, pp. 31.6.1–31.6.4.
[107] M.-F. Chang, *et al.*, "A 0.5V 4Mb logic-process compatible embedded resistive RAM (ReRAM) in 65 nm CMOS using low-voltage current-mode sensing scheme with 45 ns random read time," *ISSCC Tech. Dig.*, 2012, pp. 434–436.
[108] W.-C. Chien, *et al.*, "A novel high performance WOx ReRAM based on thermally-induced SET operation," *VLSI Technology*, 2013, pp. T100–T101.
[109] 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.
[110] K. Terabe, T. Hasegawa, T. Nakayama and M. Aono, "Quantized conductance atomic switch", Nature 433, 47-50 (2005).
[111] Valov, Ilia, et al. "Electrochemical metallization memories—fundamentals, applications, prospects." *Nanotechnology* 22.25 (2011): 254003.
[112] M. Tada, et al., "Polymer Solid-electrolyte Switch Embedded on CMOS for Nonvolatile Crossbar Switch", IEEE Transactions on Electron Devices, volume 58, Issue 12, pp. 4398-4406, 2011.
[113] 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
[114] 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.
[115] S. Sills, et al., "A Copper ReRAM Cell for Storage Class Memory Applications", 2014 Symposium on VLSI Technology Digest of Technical Papers, DOI: 10.1109/VLSIT.2014.6894368.
[116] T. Sakamoto et al., "Atom switch with Improved Cycle Endurance using Field Enhancement for Nonvolatile SoC", 2018 IEEE International Memory Workshop (IMW), doi: 10.1109/IMW.2018.8388781.
[117] Gopalan, C., et al. "Demonstration of conductive bridging random access memory (CBRAM) in logic CMOS process." *Solid-State Electronics* 58.1 (2011): 54-61.
[118] 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.
[119] J. Zahurak *et al*., "Process integration of a 27 nm, 16Gb Cu ReRAM," *2014 IEEE International Electron Devices Meeting*, San Francisco, CA, 2014, pp. 6.2.1–6.2.4.
[120] Prall, Kirk, et al. "An Update on Emerging Memory: Progress to 2Xnm." Memory Workshop (IMW), 2012 4th IEEE International. IEEE, 2012.
[121] Sankaran, Kiroubanand, et al. "Modeling of Copper Diffusion in Amorphous Aluminum Oxide in CBRAM Memory Stack." ECS Transactions 45.3 (2012): 317-330.
[122] J. R. Jameson et al., "Towards Automotive Grade Embedded RRAM" 2018 48th European Solid-State Device Research Conference (ESSDERC) doi: 10.1109/ESSDERC.2018.8486890.
[123] Miyamura, Makoto, et al. "Programmable cell array using rewritable solid-electrolyte switch integrated in 90 nm CMOS." *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International.* IEEE, 2011.
[124] 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.
[125] Xu Bai, Makoto Miyamura, Ryusuke Nebashi, Ayuka Morioka, Naoki Banno, Koichiro Okamoto, Hideaki Numata, Hiromitsu Hada, Tadahiko Sugibayashi, Toshitsugu Sakamoto, and Munehiro Tada, "An atom-switch-based field-programmable gate array with

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

optimized driving capability buffer," Japanese Journal of Applied Physics 58, SBBB04 (2019), pp. SBBB04-1-6, https://doi.org/10.7567/1347-4065/aaf899.

[126] https://www.kenkai.jaxa.jp/eng/research/innovative/innovative.html

[127] 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.

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

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

[130] 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.

[131] 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.

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

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] 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.

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

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

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

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

[142] 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.

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

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

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

[146] 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).

[147] 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).

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

[149] 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.

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

[151] 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).

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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[153] 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).
[154] J. Müller, T.S. Böscke, S. Müller, E. Yurchuk, P. Polakowski, J. Paul, D. Martin, T. Schenk, K. Khullar, A. Kersch, W. Weinreich, S. Riedel, K. Seidel, A. Kumar, T.M. Arruda, S.V. Kalinin, T. Schlösser, R. Boschke, R. van Bentum, U. Schröder, and T. Mikolajick, "Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories," *IEEE International Electron Devices Meeting (IEDM)*, Washington (2013.), pp. 10.8.1 - 10.8.4.
[155] W. Zhang, M. Takahashi, and S. Sakai, "Electrical properties of CaxSr1− xBi2Ta2O9 ferroelectric-gate field-effect transistors," *Semicond. Sci. Technol*. 28, 85003 (2013).
[156] X. Zhang, M. Takahashi, K. Takeuchi, and S. Sakai, "64 kbit ferroelectric-gate-transistor-integrated NAND flash memory with 7.5 V program and long data retention," *Jap. J. Apl. Phys.* 51, 04DD01 (2012).
[157] E. Yurchuk, S. Mueller, D. Martin, S. Slesazeck, U. Schroeder, T. Mikolajick, J. Müller, J. Paul, R. Hoffmann, J. Sundqvist, T. Schlosser, R. Boschke, R. van Bentum, and M. Trentzsch, "Origin of the endurance degradation in the novel HfO2-based 1T ferroelectric non-volatile memories," *IEEE International Reliability Physics Symposium (IRPS)* (2014.), pp. 2E.5.1 - 2E.5.5.
[158] 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).
[159] 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).
[160] E. D. Grimley, T. Schenk, T. Mikolajick, U. Schroeder, and J. M. LeBeau, "Atomic Structure of Domain and Interphase Boundaries in Ferroelectric HfO2. Advanced Materials Interfaces," 1701258 (2018). Doi: 10.1002/admi.201701258
[161] M. Pešić, F. P. G. Fengler, L/. Larcher, A. Padovani, T. Schenk, E. D. Grimley, X. Sang, J. M. LeBeau, S. Slesazeck, U. Schroeder, and T. Mikolajick, "Physical Mechanisms behind the Field-Cycling Behavior of HfO2-Based Ferroelectric Capacitors," *Advanced Functional Materials* 26, 4601 (2016).
[162] H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, and S. Slesazeck, "Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors," *ACS applied materials & interfaces* 9, 3792 (2017).
[163] M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Müller, S. Slesazeck, J. A. Ocker. "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," in *Electron Devices Meeting (IEDM), 2016 IEEE International,* 2016 Dec 3 (pp. 11-5). IEEE.
[164] S. Dünkel S, M. Trentzsch, R. Richter, P. Moll, C. Fuchs, O. Gehring, M. Majer, S. Wittek, B. Müller, T. Melde, H. Mulaosmanovic, "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond,"in *Electron Devices Meeting (IEDM), 2017 IEEE International*, 2017 Dec 2 (pp. 19-7). IEEE.
[165] M. Pesic, S. Knebel, M. Hoffmann, C. Richter, T. Mikolajick, U. Schroeder, "How to make DRAM non-volatile? Anti-ferroelectrics: A new paradigm for universal memories," *Electron Devices Meeting (IEDM), 2016 IEEE International* 2016 Dec 3 (pp. 11-6). IEEE.
[166] V. Garcia and M. Bibes, "Ferroelectric tunnel junctions for information storage and processing," *Nature Communications* 5 (2014).
[167] 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).
[168] 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).
[169] 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).
[170] 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).
[171] 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).
[172] 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).
[173] A. Chen, "Emerging memory selector devices," *13th Non-Volatile Memory Technology Symposium (NVMTS)*, Singapore (2013.), pp. 1–5.
[174] 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).
[175] F.Y. Bruno, S. Boyn, V. Garcia, S. Fusil, S. Girod, C. Carrétéro, M. Marinova, A. Gloter, S. Xavier, C. Deranlot, M. Bibes, and A. Barthélémy, "Million-fold Resistance Change in Ferroelectric Tunnel Junctions Based on Nickelate Electrodes," *Advanced Electronic Materials* 2 (2016). DOI: 10.1002/aelm.201500245
[176] G. Sanchez-Santolino, J. Tornos, D. Hernandez-Martin, J. I. Beltran, C. Munuera, M. Cabero, A. Perez-Muñoz, J. Ricote, F. Mompean, M. Garcia-Hernandez, Z. Sefrioui, "Resonant electron tunneling assisted by charged domain walls in multiferroic tunnel junctions," *Nature Nanotechnology* 12, 655 (2017).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[177] 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).
[178] 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).
[179] X. Tian, and A. Toriumi, "New opportunity of ferroelectric tunnel junction memory with ultrathin HfO_2-based oxides," *Electron Devices Technology and Manufacturing Conference (EDTM), 2017 IEEE* (pp. 36-64). IEEE.
[180] F. Ambriz-Vargas, G. Kolhatkar, M. Broyer, A. Hadj-Youssef, R. Nouar, A. Sarkissian, R. Thomas, C. Gomez-Yáñez, M. A. Gauthier, A. Ruediger, "A Complementary Metal Oxide Semiconductor Process-Compatible Ferroelectric Tunnel Junction," *ACS applied materials & interfaces* 9, 13262 (2017).
[181] S. Fujii, Y. Kamimuta, T. Ino, Y. Nakasaki, R. Takaishi, M. Saitoh, "First demonstration and performance improvement of ferroelectric HfO_2-based resistive switch with low operation current and intrinsic diode property," in VLSI *Technology, 2016 IEEE Symposium on*, 2016 Jun 14 (pp. 1-2). IEEE.
[182] Rebooting the IT Revolution: A Call for Action. SIA-SRC Report (2015): https://www.semiconductors.org/clientuploads/Resources/RITR%20WEB%20version%20FINAL.pdf
[183] M. Hilbert and P. Lopez, "The World's Technological Capacity to Store, Communicate, and Compute Information," *Science* 332 (2011) 60
[184] G. M. Church, Y. Gao, K. Yuan, S. Kosuri, "Next-generation digital information storage in DNA," *Science* 337 (2012) 1628
[185] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature* 494 (2013) 77
[186] S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, "CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria," *Nature* 547 (2017) 345
[187] J. Bornholt, L. Ceze, R. Lopez, G. Seelig, D. M. Carmean, K. Strauss, "A DNA-Based Archival Storage System," *OPERATING SYSTEMS REVIEW* 50 (2017) 637-649
[188] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, K. Strauss, "Toward a DNA-based archival storage system," *IEEE* Micro 37 (2017)
[189] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, "Nucleic acid memory," *Nature Materials* 15 (2016) 366
[190] Y. Erlich and D. Zielinski, "DNA Fountain enables a robust and efficient storage architecture," *Science* 355 (2017) 950
[191] N. F. Mott, *Metal-Insulator Transitions*, 2nd ed. (Taylor & Francis, London, 1990)
[192] T. Oka and N. Nagaosa, "Interfaces of Correlated Electron Systems: Proposed Mechanism for Colossal Electroresistance," *Phys. Rev. Lett.* 95 (2005) 266403
[193] A. Asamitsu, Y. Tomioka, H. Kuwahara, and Y. Tokura, "Current switching of resistive states in magnetoresistive manganites," *Nature* 388 (1997) 50
[194] 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
[195] 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
[196] 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
[197] 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
[198] 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
[199] 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
[200] K-H. Xue, C. A. Paz de Araujo, J. Celinska, C. McWilliams, "A non-filamentary model for unipolar switching transition metal oxide resistance random access memories,," *J. Appl. Phys*. 109 (2011) 091602
[201] 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
[202] F. Nakamura, M. Sakaki, Y. Yamanaka, S. Tamaru, T. Suzuki, and Y. Maeno, "Electric-field-induced metal maintained by current of the Mott insulator Ca2RuO,4," *Scientific Reports* 3 (2013) 2536
[203] 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
[204] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors,," *Nature Materials* 12 (2013) 114
[205] 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
[206] 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

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[207] 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

[208] 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

[209] 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)

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

[211] M. Hu *et al.*, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," *2016 53nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2016, pp. 1–6.

[212] B. Chakrabarti *et al.*, "A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit," *Scientific Reports,* Article vol. 7, p. 42429, 02/14/online 2017.

[213] P. Yao, H. Wu, B. Gao, S.B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.S.P. Wong, and H. Qian, "Face classification using electronic synapses," *Nature Comm*. 8, 1-8 (2017).

[214] T. Tsuruoka, T. Hasegawa, K. Terabe and M. Aono "Conductance Quantization and synaptic behavior in a Ta2O3-based atomic switch", Nanotechnology, 23(43), 435705 (2012).

[215] X. Sheng, et al. "Low-Conductance and Multilevel CMOS-Integrated Nanoscale Oxide Memristors." *Advanced Electronic Materials* (2019): 1800876

[216] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, and H. Qian, "Improving Analog Switching in HfOx-Based Resistive Memory With a Thermal Enhanced Layer," *IEEE Electron. Dev. Lett.* 38(8), 1019-1022 (2017).

[217] G. W. Burr, et al. "Recent progress in phase-change memory technology," *IEEE J. Emerg. Sel. Top. Circ. Sys.* 6(2), 146-162 (2016).

[218] G. W. Burr, R. M. Shelby, C. di Nolfo, et al., "Experimental Demonstration and Tolerancing of a Large–Scale Neural Network (165,000 Synapses), Using Phase–Change Memory as the Synaptic Weight Element," *IEDM Technical Digest,* 29.5, 2014.

[219] W. Kim, R.L. Bruce, T. Masuda, et al., "Confined PCM-based Analog Synaptic Devices offering Low Resistance-drift and 1000 Programmable States for Deep Learning," *VLSI Technology Symposium*, 2019.

[220] S. La Barbera, D. R. B. Ly, G. Navarro, et al., "Narrow Heater Bottom Electrode-Based Phase Change Memory as a Bidirectional Artificial Synapse," *Advanced Electronic Materials,* vol 4, no. 9, 1800223, 2018.

[221] C. Mackin, H. Tsai, S. Ambrogio, P. Narayanan, A. Chen, and G. W. Burr, "Weight Programming in DNN Analog Hardware Accelerators in the Presence of NVM Variability," *Advanced Electronic Materials* vol 5, no. 9, 1900026 (2019).

[222] H. Tsai, S. Ambrogio, C. Mackin, et al., "Inference of Long-Short Term Memory networks at software-equivalent accuracy using 2.5M analog Phase Change Memory devices," *VLSI Technology Symposium*, 2019.

[223] S. Kim et al., "A phase change memory cell with metallic surfactant layer as a resistance drift stabilizer," *Electron Devices Meeting (IEDM), 2013 IEEE International*, 30.7.

[224] W. W. Koelmans, et al., "Projected phase-change memory devices," *Nature communications*, (6) 8181 (2015).

[225] I. Giannopoulos, A. Sebastian, M. Le Gallo, et al., "8-bit precision in-memory multiplication with projected phase-change memory." *2018 IEEE International Electron Devices Meeting (IEDM)*, pp. 27-7. 2018.

[226] S. Ambrogio, M. Gallot, K. Spoon, et al., "Reducing the Impact of Phase-Change Memory Conductance Drift on the Inference of large-scale Hardware Neural Networks," *Electron Devices Meeting, 2019 IEEE International*.

[227] E. J. Fuller *et al.*, Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* **364**, 570-574 (2019).

[228] Y. Li *et al.*, Low-Voltage, CMOS-Free Synaptic Memory Based on Li X TiO2 Redox Transistors. *ACS Applied Materials & Interfaces* **11**, 38982-38992 (2019).

[229] C. A. Marianetti, G. Kotliar, G. Ceder, A first-order Mott transition in Li_xCoO_2. *Nature Materials* **3**, 627-631 (2004).

[230] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotechnology,* vol. 5, pp. 266–270, 2010.

[231] S. Datta, S. Salahuddin, and B. Behin-Aein, "Non-volatile spin switch for Boolean and non-Boolean logic," *Applied Physics Letters,* vol. 101, p. 252411, 2012.

[232] D. Morris, D. Bromberg, J.-G. J. Zhu, and L. Pileggi, "mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices," *Proceedings of the 49th Annual Design Automation Conference*, 2012, pp. 486–491.

[233] C. Pan and A. Naeemi, "Non-Boolean computing benchmarking for beyond-CMOS devices based on cellular neural network," *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on,* vol. 2, pp. 36–43, 2016.

[234] J. Hasler, C. Diorio, B. A. Minch, and C. A. Mead, "Single transistor learning synapses," Gerald Tesauro, David S. Touretzky, and Todd K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, MIT Press, Cambridge, MA, 1994, pp. 817–824.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[235] A. Bandyopadhyay, G. Serrano, and P. Hasler, "Adaptive algorithm using hot-electron injection for programming analog computational memory elements within 0.2 percent of accuracy over 3.5 decades," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 9, pp. 2107–2114, Sept. 2006.

[236] S. Kim, J. Hasler, and S. George, "Integrated Floating-Gate Programming Environment for System-Level ICs," *IEEE Transactions on VLSI*, vol. 24, no. 6, 2016. pp. 2244–2252.

[237] Chakrabartty S and Cauwenberghs G 2007 Sub-microwatt analog VLSI trainable pattern classifier *IEEE J. Solid-State Circuits* **42** 1169-1179

[238] Lu J, Young S, Arel I, and Holleman J 2015 A 1 TOps/W analog deep machine-learning engine with floating-gate storage in 0.13 µm CMOS *IEEE J. Solid-State Circuits* **50** 270-281

[239] Hasler J and Marr H 2013 Finding a roadmap to achieve large neuromorphic hardware systems *Frontiers Neurosci.* **7** 118

[240] Guo X, Merrikh Bayat F, Bavandpour M, Klachko M, Mahmoodi M R, Prezioso M, Likharev K K and Strukov D 2017 Fast, energy-efficient, robust, and reproducible mixed-signal neuromorphic classifier based on embedded NOR flash memory technology *IEEE Int. Electron Device Meeting* (San Francisco, CA) pp. 6.5.1-4

[241] X. Guo, F. Merrikh Bayat, M. Prezioso, Y. Chen, B. Nguyen, N. Do, and D.B. Strukov, "Temperature-insensitive analog vector-by-matrix multiplier based on 55 nm NOR flash memory cells", in: Proc. CICC'17, Austin, TX, Apr.-May 2017, pp. 1-4.

[242] Silicon Storage Technology, Inc. http://www.sst.com/products-and-services/superflash-technology-products

[243] Fick L, Blaauw D, Sylvester D, Skrzyniarz S, Parikh M, and Fick D 2017 Analog in-memory subthreshold deep neural network accelerator *IEEE Custom Integrated Circuits Conf.* (Austin TX) pp. 1-4

[244] Agarwal S, Garland D, Niroula J, Jacobs-Gedrim R B, Hsia A, Van Heukelom M S, Fuller E, Draper B and Marinella M 2019 Using floating gate memory to train ideal accuracy neural networks *IEEE J. Explor. Solid-State Computat. Devices Circuits* **5** 52-7

[245] G.G. Marotta, A. Macerola, A. D'Alessandro, A. Torsi, C. Cerafogli, C. Lattaro, C. Musilli, D. Rivers, E. Sirizotti, and F. Paolini, *et al.*, "A 3bit/cell 32Gb NAND flash memory at 34nm with 6MB/s program throughput and with dynamic 2b/cell blocks configuration mode for a program throughput increase up to 13MB/s," *ISSCC*, 2010.

[246] D. Lee, I. J. Chang, S.Y. Yoon, J. Jang, et al., "A 64Gb 533Mb/s DDR Interface MLC NAND Flash in Sub-20 nm Technology," *ISSCC*, 2012, pp. 430–431.

[247] N. Shibata, K. Kanda, T. Hisada, K. Isobe1, et al., "A 19nm 112.8mm2 64Gb Multi-Level Flash Memory with 400Mb/s/pin 1.8V Toggle Mode Interface," *ISSCC*, 2012, pp. 422-423.

[248] Y. Li, S. Lee, K. Oowada, H Nguyen, et. al, "128Gb 3b/Cell NAND Flash Memory in 19nm Technology with 18MB/s Write Rate and 400Mb/s Toggle Mode," *ISSCC*, 2012, pp. 436–437.

[249] C. Kim, D.H. Kim, W. Jeong, H.J. Kim, I.H. Park, H.W. Park, J. Lee, J. Park, Y.L. Ahn, J.Y. Lee, and S.B. Kim, "A 512Gb 3b/cell 64-stacked WL 3D V-NAND flash memory," in IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, pp. 202-203, 2017.

[250] C.M. Compagnoni, A. Goda, A.S. Spinelli, P. Feeley, A.L. Lacaita, and A. Visconti, "Reviewing the evolution of the NAND flash technology," Proc. IEEE, vol. 105, no. 9, pp. 1609-1633, 2017.

[251] Bavandpour M, Sahay S, Mahmoodi M R and Strukov D 2019 3D-aCortex: An ultra-compact energy-efficient neurocomputing platform based on commercial 3D-NAND flash memories *ArXive (submitted)*

[252] Bavandpour M, Sahay S, Mahmoodi M R and Strukov D 2019 3D-aCortex: An ultra-compact energy-efficient neurocomputing platform based on commercial 3D-NAND flash memories *ArXive (submitted)*

[253] S. Kim, T. Gokmen, H. M. Lee, and W. E. Haensch, "Analog CMOS-based resistive processing unit for deep neural network training," In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems*, 422–425, 2017.

[254] Y. Li, S. Kim, X. Sun, P. Solomon, T. Gokmen, H. Tsai, S. Koswatta, Z. Ren, R. Mo, C. C. Yeh, W. Haensch and E. Leobandung, "Capacitor-based Cross-point Array for Analog Neural Network with Record Symmetry and Linearity," IEEE VLSI Systems Symposium, 2018.

[255] G. Cristiano, M. Giordano, S. Ambrogio, L. P. Romero, C. Cheng, P. Narayanan, et al., "Perspective on Training Fully Connected Networks with Resistive Memories: Device Requirements for Multiple Conductances of Varying Significance," *J. Appl. Phys.*, vol. 124, no. 15, 151901, 2018.

[256] R. Genov, G. Cauwenberghs, G. Mulliken, and F. Adil, "A 5.9mW 6.5GMACS CID/DRAM array processor," *Proceedings of the 28th European Solid-State Circuits Conference*, pp. 715–718, Sept 2002.

[257] R. Karakiewicz, R. Genov, and G. Cauwenberghs, "480-GMACS/mW Resonant Adiabatic Mixed-Signal Processor Array for Charge-Based Pattern Recognition," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 11, pp. 2573–2584, Nov 2007.

[258] R. Karakiewicz, R. Genov, and G. Cauwenberghs, "1.1 TMACS/mW Fine-Grained Stochastic Resonant Charge-Recycling Array Processor," *IEEE Sensors Journal*, vol. 12, no. 4, pp. 785–792, April 2012.

[259] S. Joshi, C. Kim, S. Ha, M. Y. Chi, and G. Cauwenberghs, "A 2pJ/MAC 14-b 8x8 Linear Transform Mixed-Signal Spatial Filter in 65 nm CMOS with 84 dB Interference Suppression," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, February 2017.

[260] E. H. Lee and S. S. Wong, "A 2.5 GHz 7.7 TOPS/W switched capacitor matrix multiplier with co-designed local memory in 40nm," *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, pp. 418–419, 2016.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[261] J. Zhang, Z. Wang, and N. Verma, "A Matrix-Multiplying ADC Implementing a Machine-Learning Classifier Directly with Data Conversion," *IEEE Int. Solid-State Circ. Conf. (ISSCC)*, 2015.
[262] E.H Lee and S.S. Wong, "A 2.5 GHz 7.7TOPS/W Switched-Capacitor Matrix Multiplier with Co-designed Local Memory in 40nm," *IEEE Int. Solid-State Circ. Conf. (ISSCC)*, 2016.
[263] F. N. Buhler, A. E. Mendrela, Y. Lim, J. A. Fredenburg, and M. P. Flynn, "A 16-channel Noise-Shaping Machine Learning Analog-Digital Interface," *IEEE VLSI Systems Symposium*, 2016.
[264] W. Guo and N. Sun, "A 12b-ENOB 61 µW noise-shaping SAR ADC with a passive integrator," *42nd IEEE European Solid-State Circuits Conference (ESSCIRC 2016)*, 405-408, 2016.
[265] 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
[266] 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).
[267] 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
[268] A. Chen, "Nonlinearity and Asymmetry for Device Selection in Cross-bar Memory Arrays," *IEEE Trans. Electron Dev*., 62(9), 2857-2864, (2015)
[269] A. Chen, "Memory selector devices and crossbar array design: a modeling-based assessment," J Comput Electron (2017) 16:1186-1200 DOI 10.1007/s10825-017-1059-7
[270] The IRDS "Beyond-CMOS" Chapter 2021.
[271] 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).
[272] 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).
[273] www.micron.com/about/innovations/3d-xpoint-technology
[274] www.computerworld.com/article/2951869/computer-hardware/intel-and-micron-unveil-3d-xpoint-a-new-class-of-memory.html
[275] https://www.micron.com/products/advanced-solutions/3d-xpoint-technology
[276] https://www.intel.com/content/www/us/en/architecture-and-technology/intel-micron-3d-xpoint-webcast.html
[277] Q. Cao, S.-J. Han, J. Tersoff, A. D. Franklin, Y. Zhu, Z. Zhang, G. S. Tulevski, J. Tang, and W. Haensch, "End-bonded contacts for carbon nanotube transistors with low, size-independent resistance," *Science* 350, 68 (2015).
[278] A. D. Franklin, D. B. Farmer, and W. Haensch, "Defining and Overcoming the Contact Resistance Challenge in Scaled Carbon Nanotube Transistors," *ACS Nano* 8, 7333 (2014).
[279] A. D. Franklin, M. Luisier, S. J. Han, G. Tulevski, C. M. Breslin, L. Gignac, M. S. Lundstrom, and W. Haensch, "Sub-10 nm Carbon Nanotube Transistor," *Nano Lett.* 12, 758 (2012).
[280] A. D. Franklin, S. O. Koswatta, D. B. Farmer, J. T. Smith, L. Gignac, C. M. Breslin, S. J. Han, G. S. Tulevski, H. Miyazoe, W. Haensch, and J. Tersoff, "Carbon Nanotube Complementary Wrap-Gate Transistors," *Nano Lett*. 13, 2490 (2013).
[281] M. Steiner, M. Engel, Y. M. Lin, Y. Q. Wu, K. Jenkins, D. B. Farmer, J. J. Humes, N. L. Yoder, J. W. T. Seo, A. A. Green, M. C. Hersam, R. Krupke, and P. Avouris, "High-frequency performance of scaled carbon nanotube array field-effect transistors," *Appl. Phys. Lett*. 101, 053123 (2012).
[282] L. Ding, Z. Y. Zhang, S. B. Liang, T. Pei, S. Wang, Y. Li, W. W. Zhou, J. Liu, and L. M. Peng, "CMOS-based carbon nanotube pass-transistor logic integrated circuits," *Nat. Commun*. 677 (2012).
[283] M. M. Shulaker, G. Hills, N. Patil, H. Wei, H.-Yu Chen, H.-S. Philip Wong, and S. Mitra, "Carbon nanotube computer," *Nature* 501, 526 (2013).
[284] Q. Cao, J. Tersoff, S. J. Han, and A. V. Penumatcha, "Scaling of Device Variability and Subthreshold Swing in Ballistic Carbon Nanotube Transistors," *Phys. Rev.* Appl. 4, 024022 (2015).
[285] R. S. Park, M. M. Shulaker, G. Hills, L. S. Liyanage, S. Lee, A. Tang, S. Mitra, and H. -S. P. Wong, "Hysteresis in Carbon Nanotube Transistors: Measurement and Analysis of Trap Density, Energy Level, and Spatial Distribution," *ACS Nano* 10, 4599 (2016).
[286] G. J. Brady, Y. Joo, M. -Y. Wu, M. J. Shea, P. Gopalan, and M. S. Arnold, "Polyfluorene-Sorted, Carbon Nanotube Array Field-Effect Transistors with Increased Current Density and High On/Off Ratio," *ACS Nano* 8, 11614 (2014).
[287] G. S. Tulevski, A. D. Franklin, D. Frank, J. M. Lobez, Q. Cao, H. Park, A. Afzali, S. -J. Han, J. B. Hannon, and W. Haensch, "Toward High-Performance Digital Logic Technology with Carbon Nanotubes," *ACS Nano* 8, 8730 (2014).
[288] Kang, J., Tongay, S., Zhou, J., Li, J., Wu, J., "Band offsets and heterostructures of two-dimensional semiconductors," Applied Physics Letters 2013, 102 (1), 012111.
[289] Radisavljevic, B., Kis, A., "Mobility engineering and a metal–insulator transition in monolayer MoS_2," Nature Materials 2013, 12, 815.
[290] Li, S.-L., Wakabayashi, K., Xu, Y., Nakaharai, S., Komatsu, K., Li, W.-W., Lin, Y.-F., Aparecido-Ferreira, A., Tsukagoshi, K., "Thickness-Dependent Interfacial Coulomb Scattering in Atomically Thin Field-Effect Transistors," Nano Letters 2013, 13 (8), 3546-3552.
[291] Ma, N. and Jena, D., "Charge Scattering and Mobility in Atomically Thin Semiconductors," *Phys. Rev. X*, **4**, 011043 (2014)

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[292] C.-S. Pang, R. Zhou, X. Liu, P. Wu, T.Y.T. Hung, S. Guo, M.E. Zaghloul, S. Krylyuk, A.V. Davydov, J. Appenzeller, and Z. Chen, "Mobility Extraction in 2D Transition Metal Dichalcogenide Devices—Avoiding Contact Resistance Implicated Overestimation", Small, 2100940 (2021). DOI: https://doi.org/10.1002/smll.202100940

[293] Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., Kis, A., "Single-layer MoS2 transistors," Nature Nanotechnology 2011, 6, 147.

[294] Z. Chen, Y.-M. Lin, M. J. Rooks, Ph. Avouris, " Graphene nano-ribbon electronics," *Physica E,* Vol. **40**, p. 228 (2007)

[295] Li, X., Wang, X., Zhang, L., Lee, S. & Dai, H., "Chemically derived, ultrasmooth graphene nanoribbon semiconductors," Science 319, 1229–1232 (2008).

[296] Jiao, L., Zhang, L., Wang, X., Diankov, G. & Dai, H., "Narrow graphene nanoribbons from carbon nanotubes," Nature 458, 877–880 (2009).

[297] Kuc, A., Zibouche, N., Heine, T., "Influence of quantum confinement on the electronic structure of the transition metal sulfide TS_2," Physical Review B 2011, 83 (24), 245213.

[298] Desai, S. B., Madhvapathy, S. R., Sachid, A. B., Llinas, J. P., Wang, Q., Ahn, G. H., Pitner, G., Kim, M. J., Bokor, J., Hu, C., Wong, H.-S. P., Javey, A., "MoS2 transistors with 1-nanometer gate lengths," Science 2016, 354 (6308), 99-102.

[299] C.-S. Pang, P. Wu, J. Appenzeller, and Z. Chen, "Thickness-Dependent Study of High- Performance WS2-FETs With Ultra-scaled Channel Lengths," IEEE Transactions on Electron Devices 68, 2123-2129 (2021).

[300] Bao, W., Cai, X., Kim, D., Sridhara, K., Fuhrer, M. S., "High mobility ambipolar MoS2 field-effect transistors: Substrate and dielectric effects," Applied Physics Letters 2013, 102 (4), 042104.

[301] J. Appenzeller, F. Zhang, S. Das, and J. Knoch, "Transition Metal Dichalcogenide Schottky Barrier Transistors – A Device Analysis and Material Comparison," chapter 8, pp. 207-240 in 2D Materials for Nanoelectronics, (Editors: M. Houssa, T. Dimoulas and A. Molle), Taylor and Francis (2016).

[302] P.-C. Shen, C. Su, Y. Lin, A.-S. Chou, C.-C. Cheng, J.-H. Park, M.-H. Chiu, A.-Y. Lu, H.-L. Tang, M.M. Tavakoli, G. Pitner, X. Ji, Z. Cai, N. Mao, J. Wang, V. Tung, J. Li, J. Bokor, A. Zettl, C.-I. Wu, T. Palacios, L.-J. Li, and J. Kong, "Ultralow contact resistance between semimetal and monolayer semiconductors," Nature 593, 211- 217 (2021).

[303] Das, S., Appenzeller, J., "Where Does the Current Flow in Two-Dimensional Layered Systems?" Nano Letters 2013, 13 (7), 3396-3402.

[304] D. Jena and A. Konar, "Enhancement of Carrier Mobility in Semiconductor Nanostructures by Dielectric Engineering", Physical Review Letters **98**, 136805 (2007).

[305] D. Kiriya, M. Tosun, P. Zhao, J. S. Kang, and A. Javey, "Air-Stable Surface Charge Transfer Doping of MoS2 by Benzyl Viologen," *J. Am. Chem. Soc.*, vol. 136, no. 22, pp. 7853–7856, Jun. 2014, doi: 10.1021/JA5033327.

[306] L. Cai, C. J. McClellan, A. L. Koh, H. Li, E. Yalon, E. Pop, and X. Zheng, "Rapid Flame Synthesis of Atomically Thin MoO_3 down to Monolayer Thickness for Effective Hole Doping of WSe2," *Nano Lett.*, vol. 17, no. 6, pp. 3854–3861, Jun. 2017, doi: 10.1021/ACS.NANOLETT.7B01322.

[307] C.-S. Pang, T.Y.T. Hung, A. Khosravi, R. Addou, Q. Wang, M.J. Kim, R.M. Wallace, Z. Chen, "Atomically Controlled Tunable Doping in High Performance WSe2 Devices," *Advanced Electronic Materials*, **6**, p.1901304 (2020)

[308] H. Fang, S. Chuang, T. C. Chang, K. Takei, T. Takahashi, and A. Javey, "High-Performance Single Layered WSe2 p-FETs with Chemically Doped Contacts," *Nano Lett.*, vol. 12, no. 7, pp. 3788–3792, Jul. 2012, doi: 10.1021/NL301702R.

[309] Chiang, C.-C., Lan, H.-Y., Pang, C.-S., Appenzeller, J., Chen, Z., "Air-Stable P-doping in Record High-Performance Monolayer WSe2 Devices," IEEE EDL, DOI: 10.1109/LED.2021.3135312

[310] Radisavljevic, B., M. B. Whitwick, and A. Kis "Integrated circuits and logic operations based on single-layer MoS2", *ACS Nano*, 5: 9934-8, (2011).

[311] Wang, H., L. Yu, Y. H. Lee, Y. Shi, A. Hsu, M. L. Chin, L. J. Li, M. Dubey, J. Kong, and T. Palacios, "Integrated circuits based on bilayer MoS(2) transistors", *Nano Lett*, 12: 4674-80, (2012).

[312] Li, Na, Qinqin Wang, Cheng Shen, Zheng Wei, Hua Yu, Jing Zhao, Xiaobo Lu, Guole Wang, Congli He, Li Xie, Jianqi Zhu, Luojun Du, Rong Yang, Dongxia Shi, and Guangyu Zhang, "Large-scale flexible and transparent electronics based on monolayer molybdenum disulfide field-effect transistors", *Nature Electronics*, 3: 711-17, (2020).

[313] Liu, H., A. T. Neal, Z. Zhu, Z. Luo, X. Xu, D. Tomanek, and P. D. Ye, "Phosphorene: an unexplored 2D semiconductor with a high hole mobility," *ACS Nano*, 8: 4033-41, (2014).

[314] Sachid, A. B., M. Tosun, S. B. Desai, C. Y. Hsu, D. H. Lien, S. R. Madhvapathy, Y. Z. Chen, M. Hettick, J. S. Kang, Y. Zeng, J. H. He, E. Y. Chang, Y. L. Chueh, A. Javey, and C. Hu, 'Monolithic 3D CMOS Using Layered Semiconductors', *Adv Mater*, 28: 2547-54, (2016).

[315] C.-S. Pang, N. Thakuria, S. Gupta, Z. Chen, "First Demonstration of WSe2 Based CMOS-SRAM", *IEEE International Electron Devices Meeting (IEDM) Proceeding*, 2018

[316] P. Wu and J. Appenzeller, "Towards CMOS like devices from two-dimensional channel materials," Applied Physics Letters Materials 7, 100701-1–100701-7 (2019).

[317] P. Wu, D. Reis, X.S. Hu and J. Appenzeller, "Two-Dimensional Transistors with Reconfigurable Polarities for Secure Circuits," Nature Electronics **4**, 45-53 (2020).

[318] Sarkar, D., X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, P. M. Ajayan, and K. Banerjee, "A subthermionic tunnel field-effect transistor with an atomically thin channel," *Nature*, 526: 91-5, (2015).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[319] Qiu, C., F. Liu, L. Xu, B. Deng, M. Xiao, J. Si, L. Lin, Z. Zhang, J. Wang, H. Guo, H. Peng, and L. M. Peng, "'Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches," *Science*, 361: 387-92, (2018).

[320] P. Wu and J. Appenzeller, "Reconfigurable Black Phosphorus Vertical Tunneling Field-Effect Transistor with Record High ON-Currents," IEEE Electron Device Letters **40**, 981-984 (2019).

[321] C.-S. Pang, C.-Y. Chen, T. Ameen, S. Zhang, H. Ilatikhameneh, R. Rahman, G. Klimeck, and Z. Chen, "WSe_2 Homojunction Devices: Electrostatically Configurable as Diodes, MOSFETs, and Tunnel FETs for Reconfigurable Computing," *Small*, **15**, 2019, p.1902770

[322] Koma, A., Sunouchi, K., Miyajima, T., "Fabrication and characterization of heterostructures with subnanometer thickness," *Microelectronic Engineering* 1984**,** *2* (1), 129-136.

[323] Lee, Y.-H., Yu, L., Wang, H., Fang, W., Ling, X., Shi, Y., Lin, C.-T., Huang, J.-K., Chang, M.-T., Chang, C.-S., Dresselhaus, M., Palacios, T., Li, L.-J., Kong, J., "Synthesis and Transfer of Single-Layer Transition Metal Disulfides on Diverse Surfaces," *Nano Letters* 2013**,** *13* (4), 1852-1857.

[324] Lee, Y.-H., Zhang, X.-Q., Zhang, W., Chang, M.-T., Lin, C.-T., Chang, K.-D., Yu, Y.-C., Wang, J. T.-W., Chang, C.-S., Li, L.-J., Lin, T.-W., "Synthesis of Large-Area MoS2 Atomic Layers with Chemical Vapor Deposition," *Advanced Materials* 2012**,** *24* (17), 2320-2325.

[325] van der Zande, A. M., Huang, P. Y., Chenet, D. A., Berkelbach, T. C., You, Y., Lee, G.-H., Heinz, T. F., Reichman, D. R., Muller, D. A., Hone, J. C., "Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide," *Nature Materials* 2013**,** *12*, 554.

[326] Shaw, J. C., Zhou, H., Chen, Y., Weiss, N. O., Liu, Y., Huang, Y., Duan, X., "Chemical vapor deposition growth of monolayer MoSe2 nanosheets," *Nano Research* 2014**,** *7* (4), 511-517.

[327] Cong, C., Shang, J., Wu, X., Cao, B., Peimyoo, N., Qiu, C., Sun, L., Yu, T., "Synthesis and Optical Properties of Large-Area Single-Crystalline 2D Semiconductor WS2 Monolayer from Chemical Vapor Deposition," *Advanced Optical Materials* 2014**,** *2* (2), 131-136.

[328] Lin, Y.-C., Lu, N., Perea-Lopez, N., Li, J., Lin, Z., Peng, X., Lee, C. H., Sun, C., Calderin, L., Browning, P. N., Bresnehan, M. S., Kim, M. J., Mayer, T. S., Terrones, M., Robinson, J. A., "Direct Synthesis of van der Waals Solids," ACS Nano 2014, 8 (4), 3715-3723.

[329] Kang, K., Xie, S., Huang, L., Han, Y., Huang, P. Y., Mak, K. F., Kim, C.-J., Muller, D., Park, J., "High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity," Nature 2015, 520, 656.

[330] Zeng, L., Tao, L., Tang, C., Zhou, B., Long, H., Chai, Y., Lau, S. P., Tsang, Y. H., "High-responsivity UV-Vis Photodetector Based on Transferable WS2 Film Deposited by Magnetron Sputtering," *Scientific Reports* 2016**,** *6*, 20343.

[331] Takumi, O., Kohei, S., Seiya, I., Naomi, S., Shimpei, Y., Kentaro, M., Kuniyuki, K., Nobuyuki, S., Akira, N., Yoshinori, K., Kenji, N., Kazuo, T., Hiroshi, I., Atsushi, O., Hitoshi, W., "Multi-layered MoS 2 film formed by high-temperature sputtering for enhancement-mode nMOSFETs," *Japanese Journal of Applied Physics* 2015**,** *54* (4S), 04DN08.

[332] Hayashi, K., Kataoka, M., Jippo, H., Ohfuchi, M., Iwai, T., Sato, S., "Two-dimensional SnS_2 for detecting gases causing Sick Building Syndrome," *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017, 2017, pp 18.6.1-18.6.4.

[333] 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.

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

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

[336] 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.

[337] 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.

[338] 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.

[339] 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.

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

[341] E. Memisevic, E. Lind, M. Hellenbrand, J. Svensson and L. E. Wernersson, "Impact of Band-Tails on the Subthreshold Swing of III-V Tunnel Field-Effect Transistor," *IEEE Electron Device Letters*, vol. 38, no. 12, pp. 1661-1664, Dec. 2017.

[342] E. Memisevic, J. Svensson, M. Hellenbrand, E. Lind, and L. E. Wernersson, "Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with S = 48 mV/decade and Ion = 10 μA/μm for Ioff = 1 nA/μm at Vds = 0.3 V," *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 19.1.1–19.1.4.

[343] 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.

[344] 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.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[345] 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.
[346] H. Qianqian, H. Ru, Z. Zhan, Q. Yingxin, J. Wenzhe, W. Chunlei*, et al.*, "A novel Si tunnel FET with 36mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration," *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 8.5.1–8.5.4.
[347] 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.
[348] A. Villalon, C. Le Royer, M. Casse, D. Cooper, B. Previtali, C. Tabone, et al., "Strained tunnel FETs with record Ion: first demonstration of ETSOI TFETs with SiGe channel and RSD," *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 49–50.
[349] K. Sung Hwan, K. Hei, H. Chenming, and L. Tsu-Jae King, "Germanium-source tunnel field effect transistors with record high Ion/Ioff," *VLSI Technology, 2009 Symposium on, 2009*, pp. 178–179.
[350] T. Krishnamohan, K. Donghyun, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and <<60mV/dec subthreshold slope," *Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 2008*, pp. 1–3.
[351] 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.
[352] K. Tomioka, M. Yoshimura, and T. Fukui, "Steep-slope tunnel field-effect transistors using III-V nanowire/Si heterojunction," *VLSI Technology (VLSIT), 2012 Symposium on*, 2012, pp. 47–48.
[353] 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.
[354] 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.
[355] S. Johnson and T. Tiedje, "Temperature dependence of the Urbach edge in GaAs," *Journal of Applied Physics,* vol. 78, pp. 5609-5613, 1995.
[356] 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.
[357] U. E. Avci *et al.*, "Study of TFET non-ideality effects for determination of geometry and defect density requirements for sub-60mV/dec Ge TFET," *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, p. 34.5.1–34.5.4.
[358] R. N. Sajjad, W. Chern, J. L. Hoyt, and D. A. Antoniadis, "Trap Assisted Tunneling and Its Effect on Subthreshold Swing of Tunnel FETs," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4380–4387, Nov. 2016.
[359] M. Pala and D. Esseni, "(Invited) Modeling the Influence of Interface Traps on the Transfer Characteristics of InAs Tunnel-FETs and MOSFETs," *Meet. Abstr.*, vol. MA2014-01, no. 36, pp. 1372–1372, Apr. 2014.
[360] J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, "Auger generation as an intrinsic limit to tunneling field-effect transistor performance," *J. Appl. Phys*., vol. 120, no. 8, p. 084507, Aug. 2016.
[361] IRDS 2021 edition.
[362] S. Datta and B. Das, "Electronic analog of the electro-optic modulator," *Appl. Phys. Lett.* vol. 56, pp.665–667, 1990.
[363] 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.
[364] S. Sugahara and J. Nitta, "Spin-Transistor Electronics: An Overview and Outlook," *Proc. IEEE*, vol. 98, pp. 2124–2154, 2010.
[365] T. Tanamoto, H.Sugiyama, T.Inokuchi, T. Marukame, M. Ishikawa, K. Ikegami, and Y. Saito, "Scalability of spin field programmable gate array: A reconfigurable architecture based on spin metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.* vol. 109, pp. 07C312/1-3, 2011.
[366] 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.
[367] 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.
[368] 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.
[369] Y. Shuto, S. Yamamoto, H. Sukegawa, Z. C. Wen, R. Nakane, S. Mitani, M. Tanaka, K. Inomata, and 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.
[370] S. Yamamoto, Y. Shuto, and S. Sugahara, "Nonvolatile flip-flop based on pseudo-spin transistor architecture and its nonvolatile power-gating applications for low-power CMOS logic," *EPJ Appl. Phys.* vol. 63, pp. 14403, 2013.
[371] D. Osintev, V. Sverdlov, Z. Stanojevic, A. Makarov, and 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.
[372] Y. Saito, T. Marukame, T. Inokuchi, M. Ishikawa, H. Sugiyama, and T. Tanamoto, "Spin injection, transport and read/write operation in spin-based MOSFET," *Thin Solid Films* vol. 519, pp. 8266 – 8273, 2011.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[373] T. Sasaki, Y. Ando, M. Kameno, T. Tahara, H. Koike, T. Oikawa, T. Suzuki, and 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.

[374] T. Tahara, H. Koike, M. Kameno, T. Sasaki, Y. Ando, K. Tanaka, S. Miwa, Y. Suzuki, and M. Shiraishi, "Room-temperature operation of Si spin MOSFET with high on/off spin signal ratio," *Appl. Phys. Express* vol. 8, pp. 113004/1-3, 2015.

[375] S. Sato, M. Ichihara, M. Tanaka, and R. Nakane "Electron spin and momentum lifetimes in two-dimensional Si accumulation channels: Demonstration of Schottky-barrier spin metal-oxide-semiconductor field-effect transistors at room temperature", *Phys. Rev. B* 99, pp. 165301/1-9, 2019.

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

[377] X. Lou, C. Adelmann, S. A. Crooker, E. S. Garlid, J. Zhang, K. S. M. Reddy, S. D. Flexner, C. J. Palmstrøm, and P. A. Crowell, "Electrical detection of spin transport in lateral ferromagnet-semiconductor devices", *Nat. Phys.* vol. 3, pp. 197-202, 2007.

[378] O. M. J. van't Erve, A. T. Hanbicki, M. Holub, C. H. Li, C. Awo-Affouda, P. E. Thompson, and B. T. Jonker, "Electrical injection and detection of spin-polarized carriers in silicon in a lateral transport geometry", *Appl. Phys. Lett.* vol. 91, pp. 212109, 2007.

[379] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, and 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.

[380] 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.

[381] A. Fert, and H. Jaffrès, "Conditions for efficient spin injection from a ferromagnetic metal into a semiconductor," *Phys. Rev.* B vol. 64, pp. 184420/1-9, 2001.

[382] T. Tanamoto, H. Sugiyama, T. Inokuchi, M. Ishikawa, and Y. Saito, "Effects of interface resistance asymmetry on local and non-local magnetoresistance structures," *Jpn. J. Appl. Phys.* vol. 52, pp. 04CM03/1-4, 2013.

[383] M. Ishikawa, T. Oka, Y. Fujita, H. Sugiyama, Y. Saito, and K. Hamaya, "Spin relaxation through lateral spin transport in heavily doped n-type silicon", *Phys. Rev. B* vol. 95, pp. 115302/1-6, 2017.

[384] A. Spiesser, H. Saito, Y. Fujita, S. Yamada, K. Hamaya, S. Yuasa, and R. Jansen, "Giant Spin Accumulation in Silicon Nonlocal Spin-Transport Devices", *Phys. Rev. Appl.* vol. 8, pp. 064023/1-10, 2017.

[385] R. Jansen, A. Spiesser, H. Saito, Y. Fujita, S. Yamada, K. Hamaya, and S. Yuasa, "Nonlinear Electrical Spin Conversion in a Biased Ferromagnetic Tunnel Contact", *Phys. Rev. Appl.* vol.10, pp. 064050/1-13, 2018.

[386] A. Spiesser, H. Saito, S. Yuasa, and R. Jansen, "Tunnel spin polarization of Fe/MgO/Si contacts reaching 90% with increasing MgO thickness", *Phys. Rev. B* vol. 99, pp. 224427/1-11, 2019.

[387] T. Akiho, J. Shan, H. X. Liu, K. Matsuda, M. Yamamoto, and T. Uemura, "Electrical injection of spin-polarized electrons and electrical detection of dynamic nuclear polarization using a Heusler alloy spin source", *Phys. Rev. B* vol. 87, pp. 235205/1-7, 2013.

[388] T. Saito, N. Tezuka, M. Matsuura, and 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.

[389] K. Kasahara, Y. Fujita, S. Yamada, K. Sawano, M. Miyao, and 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.

[390] T. A. Peterson, S. J. Patel, C. C. Geppert, K. D. Christie, A. Rath, D. Pennachio, M. E. Flattè, P. M. Voyles, C. J. Palmstrø m, and P. A. Crowell, "Spin injection and detection up to room temperature in Heusler alloy/n-GaAs spin valves", *Phys. Rev. B* vol. 94, pp. 235309/1-12, 2016.

[391] Y. Fujita, M. Yamada, M. Tsukahara, T. Oka, S. Yamada, T. Kanashima, K. Sawano, and K. Hamaya, "Spin Transport and Relaxation up to 250 K in Heavily Doped n-Type Ge Detected Using $Co_2FeAl_{0.5}Si_{0.5}$ Electrodes", *Phys. Rev. Appl.* vol. 8, pp. 014007/1-9 2017.

[392] M. Yamada, Y. Fujita, S. Yamada, K. Sawano, and K. Hamaya, "Control of electrical properties in Heusler-alloy/Ge Schottky tunnel contacts by using phosphorous δ-doping with Si-layer insertion", *Mat. Sci. Semicon. Proc.* vol. 70, pp. 83-85, 2017.

[393] M. Yamada, M. Tsukahara, Y. Fujita, T. Naito, S. Yamada, K. Sawano, and K. Hamaya, "Room-temperature spin transport in *n*-Ge probed by four-terminal nonlocal measurements", *Appl. Phys. Express* vol. 10, pp. 093001/1-4, 2017.

[394] M. Tsukahara, M. Yamada, T. Naito, S. Yamada, K. Sawano, V. K. Lazarov, and K. Hamaya, "Room-temperature local magnetoresistance effect in n-Ge devices with low-resistive Schottky-tunnel contacts", *Appl. Phys. Express* vol. 12, pp. 033002/1-4, 2019.

[395] M. Yamada, F. Kuroda, M. Tsukahara, S. Yamada, T. Fukushima, K. Sawano, T. Oguchi, and K. Hamaya, "Spin injection through the energy-band symmetry matching with high spin polarization in atomically controlled ferromagnet/ferromagnet/semiconductor structures", *NPG Asia Mater.* vol. 12, pp. 47/1-9 (2020).

[396] K. Kudo, M. Yamada, S. Honda, Y. Wagatsuma, S. Yamada, K. Sawano, and K. Hamaya, "Room-temperature two-terminal magnetoresistance ratio reaching 0.1 % in semiconductor-based lateral devices with $L2_1$-ordered Co_2MnSi", *Appl. Phys. Lett.* vol. 118, pp. 162404/1-6 2021.

[397] T. Naito, M. Yamada, S. Yamada, K. Sawano, and K. Hamaya, "Suppression of Donor-Driven Spin Relaxation in Strained $Si_{0.1}Ge_{0.9}$", *Phys. Rev. Appl.* vol. 13, 054025/1-8 2020.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[398] 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.

[399] R. Nakane, Y. Shuto, H. Sukegawa, Z. C. Wen, S. Yamamoto, S. Mitani, M. Tanaka, K. Inomata, and S. Sugahara, "Fabrication of pseudo-spin-MOSFETs using a multi-project wafer CMOS chip," *Solid-State Elec.*, vol. 102, pp. 52–58, 2014.

[400] H. C. Koo, J. H. Kwon, J. Eom, J. Chang, S. H. Han, and M. Johnson, "Control of Spin Precession in a Spin-Injected Field Effect Transistor", *Science* vol. 325, pp. 1515–1518, 2009.

[401] P. Chuang, S.-C. Ho, L.W. Smith, F. Sfigakis, M. Pepper, C.-H. Chen, J.-C. Fan, J. P. Griffiths, I. Farrer, H. E. Beere, G. A. C. Jones, D. A. Ritchie, and T.-M. Chen, "All-electric all-semiconductor spin field-effect transistors", *Nat. Nanotechnol.* vol. 10, pp. 35-39, 2015.

[402] S. Lee, H. Koike, M. Goto, S. Miwa, Y. Suzuki, N. Yamashita, R. Ohshima, E. Shigematsu, Y. Ando, and M. Shiraishi, "Synthetic Rashba spin–orbit system using a silicon metal-oxide semiconductor", *Nat. Mater.* vol. 20, pp. 1228-1233, 2021.

[403] 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.

[404] 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.

[405] G. A. Salvatore, D. Bouvet, A. M. Ionescu, "Demonstration of Subthreshold 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.

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

[407] 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.

[408] 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.

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

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

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

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

[413] Krivokapic et al., "14nm Ferroelectric FinFET Technology with Steep Subthreshold Slope for Ultra Low Power Applications," *Proc. IEDM*, 2017.

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

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

[416] 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.

[417] U. Sikder et al., "Back-End-of-Line Nano-Electro-Mechanical Switches for Reconfigurable Interconnects," *IEEE Electron Device Letters*, vol. 41, no. 4, pp. 625-628, April 2020.

[418] 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.

[419] 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.

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

[421] 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.

[422] K. Kato, V. Stojanović and T.-J. K. Liu, "Embedded nano-electro-mechanical memory for energy-efficient reconfigurable logic," *IEEE Electron Device Letters*, vol. 37, no. 12, pp. 1563–1565, 2016.

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

[424] Feng, X. L., et al. "Low voltage nanoelectromechanical switches based on silicon carbide nanowires." Nano Letters, vol.10, no. 8, pp. 2891-2896, 2010.

[425] T. He et al., "Silicon carbide (SiC) nanoelectromechanical switches and logic gates with long cycles and robust performance in ambient air and at high temperature," 2013 IEEE International Electron Devices Meeting, Washington, DC, 2013, pp. 4.6.1-4.6.4.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[426] U. Zaghloul and G. Piazza, "Sub-1-volt Piezoelectric Nanoelectromechanical Relays With Millivolt Switching Capability," in IEEE Electron Device Letters, vol. 35, no. 6, pp. 669-671, June 2014.
[427] R. Nathanael et al., "4-terminal relay technology for complementary logic," *IEEE International Electron Devices Meeting Technical Digest*, pp. 223–226, 2009.
[428] T.-J. K. Liu et al., "Prospects for MEM_relay logic switch technology," *IEEE International Electron Devices Meeting Technical Digest*, pp. 424–427, 2010.
[429] 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.
[430] 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.
[431] 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.
[432] H. Fariborzi et al., "Design and demonstration of micro-electro-mechanical relay multipliers," *IEEE Asian Solid-State Circuits Conference*, 2011.
[433] 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.
[434] L. Hutin et al., "Electromechanical diode cell scaling for high-density nonvolatile memory," *IEEE Transactions on Electron Devices*, Vol. 61, pp. 1382–1387, 2014.
[435] H. S. Kwon et al., "Monolithic three-dimensional 65-nm CMOS-nanoelectromechanical reconfigurable logic for sub-1.2V operation," IEEE Electron Device Letters, vol. 38, no. 9, pp. 1317–1320, 2017.
[436] T.-J. K. Liu et al., "There's plenty of room at the top," in 2017 IEEE 30th International Conference on Micro Electro Mechanical Systems (MEMS). IEEE, 2017, pp. 1–4.
[437] U. Sikder, et al., "Vertical NV-NEM switches in CMOS Back-End-of-Line: first experimental demonstration and array programming scheme," 2020 IEEE International Electron Devices Meeting (IEDM), 2020, pp. 21.2.1-21.2.4.
[438] U. Sikder et al., "Toward Monolithically Integrated Hybrid CMOS-NEM Circuits," in *IEEE Transactions on Electron Devices*, vol. 68, no. 12, pp. 6430-6436, Dec. 2021.
[439] Y. Chen et al., "Reliability of MEM relays for zero leakage logic," *SPIE Conference Proceedings* Vol. 8614, 2013.
[440] B. Saha et al., "Reducing adhesion energy of micro-relay electrodes by ion beam synthesized oxide nanolayers," in APL Materials, vol. 5, no. 3, pp. 036103, 2017.
[441] Z. A. Ye et al., "Demonstration of 50-mV Digital Integrated Circuits with Microelectromechanical Relays," 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2018, pp. 4.1.1-4.1.4.
[442] X. Hu et al., "Ultra-Low-Voltage Operation of MEM Relays for Cryogenic Logic Applications," 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2019, pp. 34.2.1-34.2.4.
[443] X. Hu et al., "Study of DC-driven MEM Relay Oscillators for Implementation of Ising Machines," 2021 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2022, pp. 23.6.1-23.6.4.
[444] C. Zhou, D. M. Newns, J. A. Misewich and P. C. Pattnaik, "A field effect transistor based on the Mott transition in a molecular layer," *Appl Phys Lett* 70 (5), 598–600 (1997)
[445] D. M. Newns, J. A. Misewich, C. C. Tsuei, A. Gupta, B. A. Scott and A. Schrott, "Mott transition field effect transistor," *Appl Phys Lett* 73 (6), 780–782 (1998)
[446] 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.
[447] 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," *Jpn. J. Appl. Phys.*, 52(100001), 2013.
[448] Z. Yang, C. Ko and S. Ramanathan, "Oxide Electronics Utilizing Ultrafast Metal-Insulator Transitions," *Annual Review of Materials Research* 41, 8.1 (2011).
[449] A. Cavalleri, Cs. Toth, C. W. Siders, J. A. Squier, F. Raksi, P. Forget and J. C. Keiffer, "Femtosecond Structural Dynamics in VO2 during an Ultrafast Solid-Solid Phase Transition," *Phy. Rev. Lett*. 87, 237401 (2001)
[450] S. Hormoz and S. Ramanathan, "Limits on vanadium oxide Mott metal–insulator transition field-effect transistors," *Solid State Electron* 54 (6), 654–659 (2010)
[451] H. Ji , Jiang Wei , and D. Natelson, "Modulation of the Electrical Properties of VO2 Nanobeams Using an Ionic Liquid as a Gating Medium," *Nano Lett*., 12, 2988–2992 (2012)
[452] J Jeong, N Aetukuri, T Graf, TD Schladt, MG Samant and S. S. Parkin, "Suppression of Metal-Insulator Transition in VO2 by Electric Field–Induced Oxygen Vacancy Formation," *Science*, 339(6126), 1402–1405 (2013).
[453] J. Shi, S. D. Ha, Y. Zhou, F. Schoofs and S. Ramanathan, "A correlated nickelate synaptic transistor," *Nature Communications*, 4, 2676 (2013)
[454] S. D. Ha, J. Shi, Y. Meroz, L. Mahadevan and S. Ramanathan, "Neuromimetic Circuits with Synaptic Devices Based on Strongly Correlated Electron Systems," *Physical Review Applied*, 2, 064003 (2014),

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[455] M. D. Pickett, G. Medeiros-Ribeiro and R. Stanley Williams, "A scalable neuristor built with Mott memristors," *Nature Materials* 12, 114–117 (2013)
[456] N. Shukla, A. V. Thathachary, A. Agrawal, H. Paik, A. Aziz, D. G. Schlom, S. K.Gupta, R. Engel-Herbert and S. Datta, "A steep-slope transistor based on abrupt electronic phase transition," *Nature Communications* 6, 7812 (2015)
[457] Philipp Scheiderer, Matthias Schmitt, Judith Gabel, Michael Zapf, Martin Stübinger, Philipp Schütz, Lenart Dudy, Christoph Schlueter, Tien‐Lin Lee, Michael Sing, Ralph Claessen, "Tailoring Materials for Mottronics: Excess Oxygen Doping of a Prototypical Mott Insulator", Advanced Materials 30, 1706708 (2018), https://doi.org/10.1002/adma.201706708
[458] T. Wei, T. Kanki, K. Fujiwara, M. Chikanari and H. Tanaka, "Electric field-induced transport modulation in VO2 FETs with high-k oxide/ organic parylene-C hybrid gate dielectric," Appl. Phys. Lett. 108 (2016) 053503.
[459] T. Wei, T. Kanki, M. Chikanari, T. Uemura, T. Sekitani, and H. Tanaka, "Enhancement of electronic transport modulation in single crystalline VO2 nanowire-based solid-state field-effect transistor," Scientific Reports, 7, 17215 (2017), https://doi.org/10.1038/s41598-017-17468-x
[460] M. Yamamoto, R. Nouch, T. Kanki, A. N. Hattori, K. Watanabe, T. Taniguchi, K. Ueno and H. Tanaka, "Gate-Tunable Thermal Metal-Insulator Transition in VO2 Monolithically Integrated into a WSe2 Field-Effect Transistors", ACS Appl. Mater. and Inter. 11, 3224-3230 (2019), https://doi.org/10.1021/acsami.8b18745
[461] A. N. Hattori, A. I. Osaka, K, Hattori, Y. Naitoh, H. Shima, H. Akinaga, H. Tanaka, " Investigation of Statistical Metal-Insulator Transition Properties of Electronic Domains in Spatially Confined VO2 Nanostructure", Crystals 10, 631 (2020), https://doi.org/10.3390/cryst10080631
[462] R. Rakshit, A. N. Hattori, Y. Naitoh, H. Shima, H. Akinaga, H. Tanaka, "Three-Dimensional Nanoconfinement Supports Verwey Transition in Fe3O4 Nanowire at 10 nm Length Scale", Nano Lett. 19, 5003-5010 (2019), https://doi.org/10.1021/acs.nanolett.9b01222
[463] R. Yukawa, M. Kobayashi, T. Kanda, D. Shiga, K. Yoshimatsu, S. Ishibashi, M. Minohara, M. Kitamura, K. Horiba, A. F. Santander-Syro and H. Kumigashira, "Resonant tunneling driven metal-insulator transition in double quantum-well structures of strongly correlated oxide", Nat. Commun. 12, 7070 (2021). https://doi.org/10.1038/s41467-021-27327-z
[464] Y. Zhou, X. Chen, Z. Yang, C. Mouli and S. Ramanathan, "Voltage-Triggered Ultrafast Phase Transition in Vanadium Dioxide Switches," *IEEE Electron Device Letters*, 34, 220 (2013)
[465] J. Leroy, A. Crunteanu, A. Bessaudou, F. Cosset, C. Champeaux, and J. C. Orlianges, "High-speed metal-insulator transition in vanadium dioxide films induced by an electrical pulsed voltage over nano-gap electrodes," *Applied Physics Letters*, 100, 213507 (2012).
[466] T. Yajima, T. Nishimura, T. Tanaka, K. Uchida and A. Toriumi, "Modulation of VO2 Metal–Insulator Transition by Ferroelectric HfO2 Gate Insulator", Adv. Electron.Mater., 6, 1901356 (2020). https://doi.org/10.1002/aelm.201901356
[467] J.del Valle, N. M. Vargas, R. Rocco, P. Salev, Y. Kalcheim, P. N. Lapa, C. Adda, M.-H. Lee, P. Y. Wang, L. Fratino, M. J. Rozenberg and I. K. Schuller, "Spatiotemporal characterization of the field-induced insulator-to-metal transition", Science 373, 907 (2021). https://doi.org/10.1126/science.abd9088
[468] S. D. Ha, G. H. Aydogdu and S. Ramanathan, "Metal-insulator transition and electrically driven memristive characteristics of SmNiO3 thin films," *Appl. Phys. Lett*., 98, 012105 (2011)
[469] W. L. Lim, E. J. Moon, J. W. Freeland, D. J. Meyers, M. Kareev, J. Chakhalian, and S. Urazhdin, "Field-effect diode based on electron-induced Mott transition in NdNiO3," *Appl. Phys. Lett*. 101, 143111 (2012)
[470] R. Scherwitzl, P. Zubko, I. G. Lezama, S. Ono, A. F. Morpurgo, G. Catalan and J.-M. Triscone, "Electric-Field Control of the Metal-Insulator Transition in Ultrathin NdNiO3 Films," *Adv. Mater*. 22, 5517 (2010).
[471] S. D. Ha, U. Vetter, J. Shi, and S. Ramanathan, "Electrostatic gating of metallic and insulating phases in SmNiO3 ultrathin films," *Appl. Phys. Lett.* 102, 183102 (2013)
[472] S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett*. 106, 122102 (2015)
[473] S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabinyc, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett*. 106, 122102 (2015)
[474] J. Shi, Y. Zhou and S. Ramanathan, "Colossal resistance switching and band gap modulation in a perovskite nickelate by electron doping," *Nature Communications*, 5, 4860 (2014)
[475] Chiu, C.-K., et al., *Classification of topological quantum matter with symmetries.* Reviews of Modern Physics, 2016. **88**(3).
[476] Hasan, M.Z. and C.L. Kane, *Colloquium: Topological insulators.* Reviews of Modern Physics, 2010. **82**(4): p. 3045-3067.
[477] Ren, Y., Z. Qiao, and Q. Niu, *Topological phases in two-dimensional materials: a review.* Rep Prog Phys, 2016. **79**(6): p. 066501.
[478] Weng, H., et al., Quantum anomalous Hall effect and related topological electronic states. Advances in Physics, 2015. **64**(3): p. 227-282.
[479] Fei, Z., et al., *Edge conduction in monolayer WTe2.* Nature Physics, 2017. **13**(7): p. 677-682.
[480] König, M., et al., Quantum Spin Hall Insulator State in HgTe Quantum Wells. Science, 2007. **318**: p. 766
[481] Liu, C., et al., Helical Edge Transport in Millimeter-Scale Thin Films of Na3Bi. Nano Lett, 2020. **20**(9): p. 6306-6312.
[482] Tang, F., et al., Comprehensive search for topological materials using symmetry indicators. Nature, 2019. **566**(7745): p. 486-489.
[483] Vergniory, M.G., et al., *A complete catalogue of high-quality topological materials.* Nature, 2019. **566**(7745): p. 480-485.
[484] Zhang, T., et al., *Catalogue of topological electronic materials.* Nature, 2019. **566**(7745): p. 475-479.
[485] Collins, J.L., et al., Electric-field-tuned topological phase transition in ultrathin Na_3Bi. Nature, 2018. **564**(7736): p. 390-394.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[486] F. Reis, G.L., L. Dudy, M. Bauernfeind, S. Glass, W. Hanke, R. Thomale, J. Schäfer, R. Claessen, *Bismuthene on a SiC substrate: A candidate for a high-temperature quantum spin Hall material.* Science, 2017. **357**: p. 287-290.
[487] Ezawa, M., Quantized conductance and field-effect topological quantum transistor in silicene nanoribbons. Applied Physics Letters, 2013. **102**(17).
[488] Kane, C.L. and E.J. Mele, *Z2 topological order and the quantum spin Hall effect.* Phys Rev Lett, 2005. **95**(14): p. 146802.
[489] Liu, J., et al., Spin-filtered edge states with an electrically tunable gap in a two-dimensional topological crystalline insulator. Nat Mater, 2014. **13**(2): p. 178-83.
[490] Liu, Q., et al., Switching a normal insulator into a topological insulator via electric field with application to phosphorene. Nano Lett, 2015. **15**(2): p. 1222-8.
[491] Molle, A., et al., *Buckled two-dimensional Xene sheets.* Nat Mater, 2017. **16**(2): p. 163-169.
[492] Ni, Z., et al., *Tunable bandgap in silicene and germanene.* Nano Lett, 2012. **12**(1): p. 113-8.
[493] Pan, H., et al., Electric control of topological phase transitions in Dirac semimetal thin films. Sci Rep, 2015. **5**: p. 14639.
[494] Qian, X., et al., Quantum spin Hall effect in two-dimensional transition metal dichalcogenides. Science, 2014. **346**: p. 1344-1347.
[495] Vandenberghe, W.G. and M.V. Fischetti, Imperfect two-dimensional topological insulator field-effect transistors. Nat Commun, 2017. **8**: p. 14184.
[496] Wray, L.A., *Topological transistor.* Nature Physics, 2012. **8**(10): p. 705-706.
[497] Zhang, Z., et al., Magnetic quantum phase transition in Cr-doped Bi2(SexTe1-x)3 driven by the Stark effect. Nat Nanotechnol, 2017. **12**(10): p. 953-957.
[498] Nadeem, M., et al., Overcoming Boltzmann's Tyranny in a Transistor via the Topological Quantum Field Effect. Nano Lett, 2021. **21**(7): p. 3155-3161.
[499] Fuhrer, M.S., et al., Proposal for a Negative Capacitance Topological Quantum Field-Effect Transistor, in 67th Annual IEEE International Electron Devices Meeting (IEDM). 2021: San Francisco.
[500] Xu, Y., et al., *Quantized Field-Effect Tunneling between Topological Edge or Interface States.* Phys Rev Lett, 2019. **123**(20): p. 206801.
[501] Salahuddin, S. and S. Datta, Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices. Nano Letters, 2008. **8**: p. 405.
[502] Cao, W. and K. Banerjee, *Is negative capacitance FET a steep-slope logic switch?* Nat Commun, 2020. **11**(1): p. 196.
[503] Nikonov, D.E. and I.A. Young, *Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits.* IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 2015. **1**: p. 3-11.
[504] Kadykov, A.M., et al., *Terahertz detection of magnetic field-driven topological phase transition in HgTe-based transistors.* Applied Physics Letters, 2015. **107**(15): p. 152101.
[505] Ezawa, M., Topological Switch between Second-Order Topological Insulators and Topological Crystalline Insulators. Phys Rev Lett, 2018. **121**(11): p. 116801.
[506] Guan, S., et al., *Artificial gravity field, astrophysical analogues, and topological phase transitions in strained topological semimetals.* npj Quantum Materials, 2017. **2**(1): p. 23.
[507] Krishtopenko, S.S., et al., *Pressure- and temperature-driven phase transitions in HgTe quantum wells.* Physical Review B, 2016. **94**(24): p. 245402.
[508] Lindner, N.H., G. Refael, and V. Galitski, *Floquet topological insulator in semiconductor quantum wells.* Nature Physics, 2011. **7**(6): p. 490-495.
[509] Zheng, J., et al., *All-Optically Controlled Topological Transistor Based on Xenes.* Physical Review Applied, 2020. **14**(3): p. 034027.
[510] Chang, C.-Z., et al., Experimental Observation of the Quantum Anomalous Hall Effect in a Magnetic Topological Insulator. Science 2013. **340**: p. 167-170
[511] Bhattacharyya, S., et al., Recent Progress in Proximity Coupling of Magnetism to Topological Insulators. Adv Mater, 2021. **33**(33): p. e2007795.
[512] Serlin, M., et al., Intrinsic quantized anomalous Hall effect in a moiré heterostructure. Science, 2019. **367**(6480): p. 900-903.
[513] Ju, L., et al., Topological valley transport at bilayer graphene domain walls. Nature, 2015. **520**(7549): p. 650-5.
[514] Li, J., et al., Gate-controlled topological conducting channels in bilayer graphene. Nat Nanotechnol, 2016. **11**(12): p. 1060-1065.
[515] Li, J., et al., *A valley valve and electron beam splitter.* Science 2018. **362**: p. 1149-1152.
[516] A. Khitun and K. Wang, Superlattices & Microstructures **38**, 184-200 (2005).
[517] A. Khitun, M. Bao and K. L. Wang, IEEE Transactions on Magnetics **44** (9), 2141-2152 (2008).
[518] M. P. Kostylev, A. A. Serga, T. Schneider, B. Leven, and B. Hillebrands, Applied Physics Letters **87**, 153501 (2005).
[519] K.-S. Lee and S.-K. Kim, Journal of Applied Physics **104**, 053909 (2008).
[520] G. Csaba, A. Papp, and W. Porod, Physics Letters A **381**, 1471–1476 (2017).
[521] R. Verba, V. Tiberkevich, I. Krivorotov, and A. Slavin, Physical Review Applied **1**, 044006 (2014).
[522] Y.-J. Chen et al., Nano Letters **17** (1), 572–577 (2017).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[523] S. Dutta et al., Scientific Reports **5**, 09861 (2015).
[524] K. Vogt et al., Nature Communications **5**, 1–5 (2014).
[525] B. Divinskiy et al., Advanced Materials **30**, 1802837 (2018).
[526] J. Han, P. Zhang, J. T. Hou, A. A. Siddiqui and L. Liu, Science **366**, 1121–1125 (2019).
[527] R. Verba et al., Physical Review Applied **11**, 054040 (2019).
[528] A. Khitun et al., Journal of Nanoelectronics and Optoelectronics **3** (1), 24–34 (2008).
[529] A. Khitun and K. L. Wang, Journal of Applied Physics **110** (3), 034306 (2011).
[530] S. Klingler et al., Applied Physics Letters **105**, 152410 (2014).
[531] A. Khitun, Journal of Applied Physics **113** (16), 164503 (2013).
[532] A. Khitun, Journal of Applied Physics **111** (5), 054307 (2012).
[533] K. Wagner et al., Nature Nanotechnology **11**, 432–436 (2016).
[534] 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).
[535] S. Cherepov et al., Applied Physics Letters **104** (8), 082403 (2014).
[536] F. Gertz, A. Kozhevnikov, Y. Filimonov and A. Khitun, Magnetics, IEEE Transactions on Magnetics **51** (4), 1–5 (2014).
[537] A. Kozhevnikov, F. Gertz, G. Dudko, Y. Filimonov and A. Khitun, Applied Physics Letters **106** (14), 142409 (2015).
[538] Y. Wang et al., Science **366**, 1125–1128 (2019).
[539] A. Khitun, D. E. Nikonov, and K. L. Wang, Journal of Applied Physics **106** (12), 123909 (2009).
[540] R. Verba, V. Tiberkevich, I. Krivorotov, and A. Slavin, Physical Review Applied **1**, 044006 (2014).
[541] B. Divinskiy, V. E. Demidov, S. Urazhdin, R. Freeman, A. B. Rinkevich, and S. O. Demokritov, Advanced Materials **30**, 1802837 (2018).
[542] P. Graczyk and M. Krawczyk, Scientific Reports **11**, 15692 (2021).
[543] R. Lebrun et al., Nature **561**, 222–225 (2018).
[544] J. Han, P. Zhang, Z. Bi, Y. Fan, T. S. Safi, J. Xiang, J. Finley, L. Fu, R. Cheng, and L. Liu, Nature Nanotechnology **15**, 563–568 (2020).
[545] C. Jia, M. Chen, A. F. Schäffer, and J. Berakdar, Npj Computational Materials **7**, 101 (2021).
[546] J. Li, C. B. Wilson, R. Cheng, M. Lohmann, M. Kavand, W. Yuan, M. Aldosary, N. Agladze, P. Wei, M. S. Sherwin, and J. Shi, Nature **578**, 70–74 (2020).
[547] P. Vaidya, S. A. Morley, J. van Tol, Y. Liu, R. Cheng, A. Brataas, D. Lederman, and E. del Barco, Science **368**, 160–165 (2020).
[548] C. Gong, L. Li, Z. Li, H. Ji, A. Stern, Y. Xia, T. Cao, W. Bao, C. Wang, Y. Wang, Z. Q. Qiu, R. J. Cava, S. G. Louie, J. Xia, and X. Zhang, Nature **546**, 265–269 (2017).
[549] D. MacNeill, J. T. Hou, D. R. Klein, P. Zhang, P. Jarillo-Herrero, and L. Liu, Physical Review Letters **123**, 047204 (2019).
[550] G. Chen, S. Qi, J. Liu, D. Chen, J. Wang, S. Yan, Y. Zhang, S. Cao, M. Lu, S. Tian, K. Chen, P. Yu, Z. Liu, X. C. Xie, J. Xiao, R. Shindou, and J.-H. Chen, Nature Communications **12**, 6279 (2021).

[551] Andreakou, S.V. Poltavtsev, J.R. Leonard, E.V. Calman, M. Remeika, Y.Y. Kuznetsova, L.V. Butov, J. Wilkes, M. Hanson, A.C. Gossard, "Optically controlled excitonic transistor," *Appl. Phys. Lett*. 104, 091101 (2014).
[552] C.J. Dorow, Y.Y. Kuznetsova, J.R. Leonard, M.K. Chu, L.V. Butov, J. Wilkes, M. Hanson, A.C. Gossard, "Indirect excitons in a potential energy landscape created by a perforated electrode," *Appl. Phys. Lett*. 108, 073502 (2016).
[553] A.G. Winbow, J.R. Leonard, M. Remeika, Y.Y. Kuznetsova, A.A. High, A.T. Hammack, L.V. Butov, J. Wilkes, A.A. Guenther, A.L. Ivanov, M. Hanson, A.C. Gossard, "Electrostatic Conveyer for Excitons," *Phys. Rev. Lett*. 106, 196806 (2011).
[554] A.A. High, J.R. Leonard, A.T. Hammack, M.M. Fogler, L.V. Butov, A.V. Kavokin, K.L. Campman, A.C. Gossard, "Spontaneous coherence in a cold exciton gas," *Nature* 483, 584 (2012).
[555] J.R. Leonard, A.A. High, A.T. Hammack, M.M. Fogler, L.V. Butov, K.L. Campman, A.C. Gossard, "Pancharatnam-Berry phase in condensate of indirect excitons," *SRC* publication P092121.
[556] M.M. Fogler, L.V. Butov, K.S. Novoselov, "Indirect excitons in van der Waals heterostructures," *Nature Commun*. 5, 4555 (2014).
[557] E.V. Calman, M.M. Fogler, L.V. Butov, S. Hu, A. Mishchenko, A.K. Geim, "Indirect excitons in van der Waals heterostructures at room temperature," *SRC* publication P092424.
[558] J.R. Leonard, A.A. High, A.T. Hammack, M.M. Fogler, L.V. Butov, K.L. Campman, A.C. Gossard, Pancharatnam-Berry phase in condensate of indirect excitons, Nature Commun. 9, 2158 (2018).
[559] C.J. Dorow, J.R. Leonard, M.M. Fogler, L.V. Butov, K.W. West, L.N. Pfeiffer, Split-gate device for indirect excitons, Appl. Phys. Lett. 112, 183501 (2018).
[560] C.J. Dorow, M.W. Hasling, D.J. Choksy, J.R. Leonard, L.V. Butov, K.W. West, L.N. Pfeiffer, High-mobility indirect excitons in wide single quantum well, SRC publication P095019.
[561] M. Feng, N. Holonyak, Jr., and W. Hafez, "Light-emitting transistor: Light emission from InGaP/GaAs heterojunction bipolar transistors," *Appl. Phys. Lett*. 84, pp. 151–153 (2004).
[562] M. Feng, N. Holonyak, Jr., G. Walter, and R. Chan, "Room temperature continuous wave operation of a heterojunction bipolar transistor laser," *Appl. Phys. Lett*. 87, pp. 131103, (2005).
[563] H.W. Then, M. Feng, and N. Holonyak, Jr., "The Transistor Laser: Theory and Experiment," *Proc. IEEE* 101, pp. 2271, (2013).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[564] J.M. Dallesasse, P.L. Lam, and G. Walter, "Electronic-Photonic Integration Using the Light-Emitting Transistor," *Latin American Optics and Photonics Conference (LAOP) 2014*, paper LTu3A.2, (2014).

[565] M. Feng, J. Qiu, and N. Holonyak, Jr., "Tunneling Modulation of Transistor Lasers: Theory and Experiment," *IEEE J. Quant. Electron*,vol. 54, issue 2, April 2018.

[566] R. Bambery, C. Wang, J.M. Dallesasse, M. Feng, and N. Holonyak, Jr., "Effect of the energy barrier in the base of the transistor laser on the recombination lifetime," *Appl. Phys. Lett.* 8, pp 081117 (2014).

[567] M. Feng, H.W. Then, and N. Holonyak, Jr., "Transistor Laser Optical NOR Gate for High-Speed Optical Logic Processors," http://www.dtic.mil/dtic/tr/fulltext/u2/1041445.pdf, AD1041445, 20 Mar (2017).

[568] R. Bambery, C.Y. Wang, F. Tan, M. Feng, and N. Holonyak, Jr., "Single Quantum-Well Transistor Lasers Operating Error-Free at 22 Gb/s," *IEEE Photon. Tech. Lett*. 27, pp 600, (2015).

[569] F. Tan, R. Bambery, M. Feng, and N. Holonyak, Jr., "Relative intensity noise of a quantum well transistor laser," *Appl. Phys. Lett*. 101, 151118 (2012).

[570] M. Feng, M. Liu, C. Wang and N. Holonyak, "Vertical cavity transistor laser for on-chip OICs," *2015 IEEE Summer Topicals Meeting Series (SUM)*, Nassau, 2015, pp. 146–147, doi: 10.1109/PHOSST.2015.7248238.

[571] A. Winoto, J. Qiu, D. Wu, and M. Feng, "Optical NOR Gate Transistor Laser Integrated Circuit," in *Conference on Lasers and Electro-Optics*, OSA Technical Digest (Optical Society of America, 2019), paper JTh2A.62.

[572] NOCS '19: Proceedings of the 13th IEEE/ACM International Symposium on Networks-on-Chip, Oct. 2019, Article #10, pp. 1-8, https://doi.org/10.1145/3313232.3352368.

[573] H. Lan, I. Tseng, Y. Lin, S. Chang and C. Wu, "Characteristics of Blue GaN/InGaN Quantum-Well Light-Emitting Transistor," in *IEEE Electron Device Letters*, vol. 41, no. 1, pp. 91-94, Jan. 2020.

[574] N. Sharma, J. P. Bird, Ch. Binek, P. A. Dowben, D. Nikonov, and A. Marshall, "Evolving Magneto-electric Device Technologies", *Semiconductor Science and Technology* **35** (2020) 073001; doi: 10.1088/1361-6641/ab8438

[575] D. E. Nikonov, I. A. Young, "Overview of beyond-CMOS devices and a uniform methodology for their benchmarking", *Proc. IEEE* **101,** 2498-2533 (2013)

[576] . E. Nikonov, I. A. Young, "Benchmarking spintronic logic devices based on magnetoelectric oxides", *J. Mater. Res.* **29**, 2109-2115 (2014).

[577] C. Binek, B. Doudin, "Magnetoelectronics with magnetoelectrics", *J. Phys.: Condens. Matter* **17,** L39-44 (2005)

[578] X. Chen, A. Hochstrat, P. Borisov, W. Kleemann, "Magnetoelectric exchange bias systems in spintronics", *Appl. Phys. Lett.* **89,** 202508 (2006)

[579] M. Bibes, A. Barthélémy, "Multiferroics: Towards a magnetoelectric memory", *Nat. Mater.* **7,** 425-426 (2008)

[580] N. Sharma, A. Marshall, J. P. Bird, P. A. Dowben, "Magneto-Electric Magnetic Tunnel Junction as process adder for non-volatile memory applications", *Proceedings on IEEE Xplore; Circuits and Systems Conference (DCAS)*, 2015 IEEE Dallas (2015) 978-1-4673-7980-9/15; doi: 10.1109/DCAS.2015.7356588

[581] N. Sharma, A. Marshall, J. Bird, P. Dowben, "Multi-bit Adder Design using the Compact Model of the ME-MTJ device and its Derived Structures." The 11th International Conference on Advanced Semiconductor Devices and Microsystems (ASDAM) 2016, *IEEE Explore* (2017) 89-92, INSPEC Assession number 16580073; doi: 10.1109/ASDAM.2016.7805902

[582] N. Sharma, A. Marshall, J. Bird, and P. A. Dowben, "VerilogA based Compact model of a three-terminal ME-MTJ device", Proceedings of the 16th International Conference on Nanotechnology; Sendai, Japan, August 22-25, 2016, *IEEE Xplore* 16498661 (2016) 145-148; DOI: 10.1109/NANO.2016.7751550

[583] N. Sharma, A. Marshall, J. Bird, P. Dowben, "Magneto-Electric Magnetic Tunnel Junction Logic devices", 2015 4th Berkeley Symposium on Energy Efficient Electronic Systems (E3S), Oct. 2, 2015; *IEEE Explore* (2015), doi: 10.1109/E3S.2015.7336817

[584] N. Sharma, J. P. Bird, P. A. Dowben, and A. Marshall, "Compact-Device Model Development for the Energy-Delay Analysis of Magneto-Electric Magnetic Tunnel Junction Structures", *Semiconductor Science and Technology* **31**, 065022 (2016); doi:10.1088/0268-1242/31/6/065022

[585] R. Ahmed, R. H. Victora, "A fully electric field driven scalable magnetoelectric switching element", *Appl. Phys. Lett.* **112,** 182401 (2018)

[586] P. A. Dowben, C. Binek, and D. E. Nikonov, "Potential of nonvolatile magnetoelectric devices for spintronic applications", Chapter 11 in *Nanoscale Silicon Devices*; edited by Shuni Oda and David Ferry; Taylor and Francis (London) (2016) pp 255-278, ISBN 9781482228670

[587] C. Pan, A. Naeemi, "An expanded benchmarking of beyond-CMOS devices based on boolean and neuromorphic representative circuits", IEEE Journal on Exploratory Solid-State Computational Devices and Circuits 3, 101-110 (2018)

[588] M. G. Mankalale, Z. Liang, Z. Zhao, C. H. Kim, J.–P. Wang, and S. S. Sapatnekar, "CoMET: Composite-input magnetoelectric- based logic technology", *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* 3, 27 (2017)

[589] R. Perricone, Z. Liang, M. G. Mankalale, M. Niemier, S. S. Sapatnekar, J.-P. Wang and X. S. Hu, "An Energy Efficient Non-Volatile Flip-Flop based on CoMET Technology", Proceedings of the 2019 Design, Automation and Test in Europe Conference and Exhibition, *IEEE Explore* 8714916, 390-39 (2019); doi: 10.23919/DATE.2019.8714916

[590] M. Kazemi, "An electrically reconfigurable logic gate intrinsically enabled by spin-orbit materials", *Sci. Rep.* **7,** 15358 (2006)

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[591] D. C. Vaz, C.-C. Lin, J. Plombon., W. Y. Choi, I. Groen, I. Arango, V. T Pham, D. E. Nikonov, H. Li, P. Debashis, S. B. Clendenning, T. A. Gosavi, V. Garcia, S. Fusil, M. Bibes, Y.-L. Huang, B. Prasad, R. Ramesh, F. Casanova, and I. A. Young, "Functional Demonstration of a Fully Integrated Magneto-Electric Spin-Orbit Device", IEEE International Electron Devices Meeting (IEDM), IEEE Xplore (2022) IEDM21-689

[592] S. Manipatruni, D. E. Nikonov, C.-C. Lin, T. A. Gosavi, H. Liu, B. Prasad, Y.-L. Huang, E. Bonturim, R. Ramesh, I. A. Young, "Scalable energy-efficient magnetoelectric spin–orbit logic", *Nature* **565**, 35 (2019); doi: 10.1038/s41586-018-0770-2

[593] J. Zeng, P. Yi, B. Chen, C. Huang, X. Qi, S. Qiu, L. Fang, "MESO-ADC: The ADC design using MESO device", *Microelectronics Journal* **116**, 105235 (2021); doi: 10.1016/j.mejo.2021.105235

[594] Li, H.; Lin, C.C.; Nikonov, D.E.; Young, I.A. "Differential Electrically Insulated Magnetoelectric Spin-Orbit Logic Circuits", *IEEE J. Exploratory Solid-State Computational Devices and Circuits* **7**, 18-25 (2021); doi: 10.1109/JXCDC.2021.3105524

[595] A. W. Stephan, J. Hu, S. J. Koester, "Performance Estimate of Inverse Rashba-Edelstein Magnetoelectric Devices for Neuromorphic Computing", *IEEE J. Exploratory Solid-State Computational Devices and Circuits* **5,** 25 (2019); doi: 10.1109/JXCDC.2019.2903286

[596] A. W. Stephan, Q. Lou, M. Niemier, X. S. Hu, S. J. Koester, "Nonvolatile Spintronic Memory Cells for Neural Networks", arXiv:1905.12679v1 [cs.ET] 29 May 2019

[597] D. Datta, V. N. Naik, M. V. R. M. Kota, "Voltage Controlled spin switches for low power applications", U.S. Patent no. 9,646,666 B2 issued May 9, 2017

[598] M. Myronbov, C. Morrison, "Spin electronic Device", United Kingdon Patent Application GB 2560936-A, published October 3, 2018

[599] S. Wang, J. Zheng, Y. Yang, G. Cong, H. Cui, and Li Cai, "Proposal of a magneto-electric effect assisted all spin logic device", *J. Phys. D: Appl. Phys.* **52**, 125001 (2019)

[600] P. A. Dowben, C. Binek, K. Zhang, L. Wang, W.-N. Mei, J. P. Bird, U. Singisetti, X. Hong, K. L. Wang, D. Nikonov, "Towards a Strong Spin-Orbit Coupling Magnetoelectric Transistor", *IEEE J. Exploratory Solid-State Computational Devices and Circuits* **4,** 1-9 (2018); Online ISSN: 2329-9231; doi: 10.1109/JXCDC.2018.2809640

[601] N. Sharma, A. Marshall, J. Bird, P. A. Dowben, "Verilog-A based Compact Modeling of the ME-FET device", 2017 Fifth Berkeley Symposium on *Symposium & Steep Transistors Workshop (E3S) IEEE Xplore* (2018) INSPEC Accession Number: 17453928; doi: 10.1109/E3S.2017.8246186

[602] N. Sharma, C. Binek, A. Marshall, J. P. Bird, P. A. Dowben and D. Nikonov, "Compact Modeling and Design of Magneto-Electric Transistor Devices and Circuits," *2018 31st IEEE International System-on-Chip Conference (SOCC)*, Arlington, VA, USA, *IEEE Explore* (2018) 146-151; doi: 10.1109/SOCC.2018.8618494

[603] C. Pan, A. Naeemi, "Complementary Logic Implementation for Antiferromagnet Field-Effect Transistors", *IEEE J. Exploratory Solid-State Computational Devices and Circuits* **4**, 69-75 (2018)

[604] J. Kelber, C. Binek, P. A. Dowben, K. Belashchenko, "Magneto-electric voltage controlled spin transistors", U.S. Patent 9,379,232 B2; issued June 28, 2016

[605] S. C. Stuart, B.Gray, D. Nevola, L. Su, E. Sachet, M. Ulrich, D. B. Dougherty, "Magnetoelectric oxide films for spin manipulation in graphene", *Phys. Status Solidi RRL* **10**, 242–7 (2016)

[606] D. E. Nikonov, C. Binek, X. Hong, J. P. Bird, K. L. Wang, P. A. Dowben, "Magneto-Electric Logic Devices Using Semiconductor Channel with Large Spin-Orbit Coupling", U.S. Patent 10,361,292 B2, issued July 23, 2019

[607] N. Sharma Gaul, A. Marshall, P. A Dowben, D. E Nikonov, "Circuits Based on Magnetoelectric Transistor Devices", U.S. Patent application 16581691, publication date: March 26, 2020, filed September 24, 2019

[608] K. He, B. Barut, S. Yin, M. D. Randle, R. Dixit, N. Arabchigavkani, J. Nathawat, A. Mahmood, W. Echtenkamp, C. Binek, P. A. Dowben, and J. P. Bird, "Graphene on Chromia: A System for Beyond-Room-Temperature Spintronics", *Advanced Materials* (2022), in the press; doi: 10.1002/adma.202105023

[609] R. Choudhary, P. Kumar, P. Manchanda, D. J. Sellmyer, P. A. Dowben, A. Kashyap, and R. Skomski, "Interface-induced spin polarization in graphene on chromia", *IEEE Magnetics Letters* **7** (2016) 3101604; doi: 10.1109/LMAG.2016.2530628

[610] S. Cao, Z. Xiao, C.-P. Kwan, K. Zhang, J. P. Bird, L. Wang, W.-N. Mei, X. Hong, and P. A. Dowben, "Moving towards the Magnetoelectric Graphene Transistor", *Applied Physics Letters* **111**, 182402 (2017); doi: 10.1063/1.4999643

[611] P. A. Dowben, D. E. Nikonov, A. Marshall and Ch. Binek, "Magneto-electric antiferromagnetic spin-orbit logic devices", *Applied Physics Letters* **116**, 080502 (2020); doi: 10.1063/1.5141371

[612] A. Marshal and P. A. Dowben, "Magneto-electric Transistor Devices and Circuits with Steering Logic", 2020 IEEE 14th Dallas Circuits and Systems Conference (DCAS), *IEEE Explore* (2021) pp 1-4, doi: 10.1109/DCAS51144.2020.9330662

[613] S. Angizi, N. Khoshavi, A. Marshall, P. Dowben, and D. Fan, "MeF-RAM: A New Non-Volatile Cache Memory Based on Magneto-Electric FET", *ACM Transactions on Automation of Electronic Systems (TODAES)* **27** (2021) 18; doi: 10.1145/3484222

[614] A. Marshall and P. A. Dowben, "Magneto-electric Transistor Devices and Circuits with Steering Logic", in Advances in Semiconductor Technologies, edited by An Chen, Wiley (2022), in the press

[615] A. Parthasarathy, S. Rakheja, "Dynamics of magnetoelectric reversal of an antiferromagnetic domain", *Physics Review Applied* **11**, 034051 (2019)

[616] K. D. Belashchenko, O. Tchernyshyov, A. A. Kovalev, and O. A. Tretiakov, "Magnetoelectric domain wall dynamics and its implications for magnetoelectric memory", *Appl. Phys. Lett.* **108**, 132403 (2016); doi: 10.1063/1.4944996

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[617] E. Raymenants, D. Wan, S. Couet, Y. Canvel, A. Thiam, D. Tsvetanova, L. Souriau, I. Asselberghs, R. Carpenter, N. Jossart, M. Manfrini, A. Vaysset, O. Bultynck, S. Van Beek, M. Heyns, D. E. Nikonov, I. A.Young, S. Ghosh, L. Vila, K. Garello, S. Pizzini, V. D. Nguyen and I. P. Radu, "Magnetic domain walls: from physics to devices", IEEE International Electron Devices Meeting (IEDM), IEEE Xplore (2022) IEDM-685

[618] M. Street, W. Echtenkamp, T. Komesu, S. Cao, P. A. Dowben, and Ch. Binek, "Increasing the Néel Temperature of Magnetoelectric Chromia for Voltage-Controlled Spintronics", *Applied Physics Letters* **104**, 222402 (2014)

[619] Xi He, Yi Wang, Ning Wu, Anthony N. Caruso, Elio Vescovo, Kirill D. Belashchenko, Peter A. Dowben and Christian Binek, "Robust isothermal electric control of exchange bias at room temperature", *Nature Materials* **9** (2010) 579-585

[620] Ning Wu, Xi He, Aleksander Wysocki, Uday Lanke, Takashi Komesu, Kirill D. Belashchenko, Christian Binek and Peter A. Dowben, "Imaging and control of surface magnetization domains in a magnetoelectric antiferromagnet", *Physical Review Letters* **106**, 087202 (2011).

[621] Shi Cao, Xin Zhang, Ning Wu, A. T. N'Diaye, G. Chen, A. K. Schmid, Xumin Chen, W. Echtenkamp, A. Enders, Ch. Binek, P. A. Dowben, "Spin Polarization Asymmetry at the Surface of Chromia", *New Journal of Physics* **16**, 073021 (2014); doi:10.1088/1367-2630/16/7/073021

[622] Y.-J. Chen, H. K. Lee, R. Verba, J. A. Katine, I. Barsukov, V. Tiberkevich, J. Xiao, A. Slavin, I. Krivorotov, "Parametric resonance of magnetization excited by electric field", *Nano Letters* **17**, 572 (2017)

[623] R. Verba, M. Carpentieri., G. Finocchio, V. Tiberkevich, A.Slavin, "Excitation of spin waves in an in-plane-magnetized ferromagnetic nanowire using voltage-controlled magnetic anisotropy", *Physical Review Applied* 7, 064023 (2017)

[624] E. Parsonnet, Y.-L. Huang, T. Gosavi, A. Qualls, D. Nikonov, C.-C. Lin, I. Young, J. Bokor, L. W. Martin, and R. Ramesh, "Toward Intrinsic Ferroelectric Switching in Multiferroic $BiFeO_3$", *Phys. Rev. Lett.***125**, 067601 (2020)

[625] Ather Mahmood, Will Echtenkamp, Mike Street, Jun-Lei Wang, Shi Cao, Takashi Komesu, Peter A. Dowben, Pratyush Buragohain, Haidong Lu, Alexei Gruverman, Arun Parthasaraty, Shaloo Rakheja, Christian Binek, "Voltage controlled Néel vector rotation in zero magnetic field at CMOS-compatible temperatures", *Nature Communications* **12**, 1674 (2021); doi: 10.21203/rs.3.rs-38435/v1

[626] J.-L. Wang, W. Echtenkamp, A. Mahmood, C. Binek, "Voltage controlled magnetism in Cr_2O_3 based all-thin-film systems", *J. Magnetism and Magnetic Materials* **486**, 165262 (2019)

[627] T. Kosub, M. Kopte, R. Hühne, P. Appel, B. Shields, P. Maletinsky, R. Hübner, M. O. Liedke, J. Fassbender, O. G. Schmidt & D. Makarov, "Purely antiferromagnetic magnetoelectric random access memory", *Nature Comm.* **8**, 13985 (2017); DOI: 10.1038/ncomms13985

[628] J. A. Currivan, M. A. Baldo & C. A. Ross. (12) Patent Application Publication (10) Pub . No .: US 2013 / 0181158 A1. *US Pat.* **1,** 12–15 (2013).

[629] J. A. Currivan, Y. Jang, M. D. Mascaro, M. A. Baldo & C. A. Ross. Low Energy Magnetic Domain Wall Logic in Short, Narrow, Ferromagnetic Wires. *IEEE Magn. Lett.* **3,** 3000104 (2012)

[630] J. A. Currivan-Incorvia, S. Siddiqui, S. Dutta, E. R. Evarts, J. Zhang, D. Bono, C. A. Ross & M. A. Baldo. Logic circuit prototypes for three-terminal magnetic tunnel junctions with mobile domain walls. *Nat. Commun.* **7,** 10275 (2016).

[631] D. Morris, D. Bromberg, J.-G. J.-. G. J. Zhu & L. Pileggi. mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices. *Proc. 49th Annu. Des. Autom. Conf.* 486–491 (2012).

[632] V. Sokalski, D. M. Bromberg, D. Morris, M. T. Moneck, E. Yang, L. Pileggi & J.-G.-. G. Zhu. Naturally Oxidized FeCo as a Magnetic Coupling Layer for Electrically Isolated Read/Write Paths in mLogic. *IEEE Trans. Magn.* **49,** 4351–4354 (2013).

[633] D. M. Bromberg, M. T. Moneck, V. M. Sokalski, J. Zhu, L. Pileggi & J. G. Zhu. Experimental demonstration of four-terminal magnetic logic device with separate read- and write-paths. *Tech. Dig. - Int. Electron Devices Meet. IEDM* **2015-Febru,** 33.1.1-33.1.4 (2015).

[634] D. M. Bromberg, J. Zhu, L. Pileggi, V. Sokalski, M. Moneck & P. C. Richardson. (12) United States Patent.

[635] J. G. J. Zhu, D. M. Bromberg, M. Moneck, V. Sokalski & L. Pileggi. MLogic: All spin logic device and circuits. *2015 4th Berkeley Symp. Energy Effic. Electron. Syst. E3S 2015 - Proc.* 1–2 (2015). doi:10.1109/E3S.2015.7336781

[636] D. E. Nikonov & I. A. Young. Overview of beyond-CMOS devices and a uniform methodology for their benchmarking. *Proc. IEEE* **101,** 2498–2533 (2013).

[637] D. E. Nikonov & I. A. Young. Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **1,** 3–11 (2015).

[638] D. E. Nikonov & I. A. Young. Uniform methodology for benchmarking beyond-CMOS logic devices. *Electron Devices Meet. (IEDM), 2012 IEEE Int.* 24–25 (2012).

[639] X. Hu, A. Timm, W. H. H. Brigner, J. A. C. A. C. Incorvia & J. S. S. Friedman. SPICE-only model for STT domain-wall MTJ logic. *IEEE Electron Device Lett*

[640] J. A. Currivan-Incorvia, S. Siddiqui, S. Dutta, E. R. Evarts, C. A. Ross & M. A. Baldo. Spintronic logic circuit and device prototypes utilizing domain walls in ferromagnetic wires with tunnel junction readout. *2015 IEEE Int. Electron Devices Meet.* 32–36 (2015).

[641] Y. Kang, J. Bokor & V. Stojanović. Design Requirements for a Spintronic MTJ Logic Device for Pipelined Logic Applications. *IEEE Trans. Electron Devices* **63,** 1754–1761 (2016).

[642] T. P. Xiao, M. J. Marinella, C. H. Bennett, X. Hu, B. Feinberg, R. Jacobs-Gedrim, S. Agarwal, J. Brunhaver, J. S. Friedman & J. A. C. Incorvia. Energy and Performance Benchmarking of a Domain Wall-Magnetic Tunnel Junction Multibit Adder. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **PP,** 1 (2019).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[643] M. Alamdar, T. Leonard, C. Cui, B. P. Rimal, L. Xue, O. G. Akinola, T. P. Xiao, J. S. Friedman, C. H. Bennett, M. J. Marinella, & J. A. C. Incorvia. Domain wall-magnetic tunnel junction spin-orbit torque devices and circuits for in-memory computing. *Applied Physics Letters* **118**, 112401 (2021).

[644] E. Raymenants, O. Bultynck, D. Wan, T. Devolder, K. Garello, L. Souriau, A. Thiam, D. Tsvetanova, Y. Canvel, D. E. Nikonov, I. A. Young, M. Heyns, B. Soree, I. Asselberghs, I. Radu, S. Couet & V. D. Nguyen. Nanoscale domain wall devices with magnetic tunnel junction read and write. *Nat. Electron.* **4**, 392–398 (2021).

[645] E. Raymenants, D. Wan, S. Couet, L. Souriau, A. Thiam, Di. Tsvetanova, Y. Canvel, K. Garello, G. S. Kar, M. Heyns, I. Asselberghs, D. E. Nikonov, I. A. Young, S. Pizzini, I. Radu & V. Dai Nguyen. All-Electrical Control of Scaled Spin Logic Devices Based on Domain Wall Motion. *IEEE Trans. Electron Devices* **68**, 2116–2122 (2021).

[646] C. Garg, S. H. Yang, L. Thompson, T. Topuria, A. Capua, B. Hughes, T. Phung, P. C. Filippou & S. S. P. Parkin. Efficient chiral-domain-wall motion driven by spin-orbit torque in metastable platinum films. *Phys. Rev. Appl.* **14**, 1 (2020).

[647] K. Cai, Z. Zhu, J. M. Lee, R. Mishra, L. Ren, S. D. Pollard, P. He, G. Liang, K. L. Teo & H. Yang. Ultrafast and energy-efficient spin–orbit torque switching in compensated ferrimagnets. *Nat. Electron.* **3**, 37–42 (2020).

[648] A. Sengupta, Y. Shim & K. Roy. Proposal for an all-spin artificial neural network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets. *IEEE Trans. Biomed. Circuits Syst.* **10,** 1152–1160 (2016).

[649] S. Dutta, S. A. Siddiqui, F. Buttner, L. Liu, C. A. Ross & M. A. Baldo. A logic-in-memory design with 3-Terminal magnetic tunnel junction function evaluators for convolutional neural networks. *Proc. IEEE/ACM Int. Symp. Nanoscale Archit. NANOARCH 2017* 83–88 (2017). doi:10.1109/NANOARCH.2017.8053724

[650] D. Blaauw & Q. Dong. Binary-threshold, Rectified-linear and Recurrent Neural Networks Built with Spintronic Devices. *SPIN Neuron DAC (1).pdf* (2016).

[651] R. Zand, K. Y. Camsari, S. D. Pyle, I. Ahmed, C. H. Kim & R. F. DeMara. Low-energy deep belief networks using intrinsic sigmoidal spintronic-based probabilistic neurons. in *Proceedings of the 2018 on Great Lakes Symposium on VLSI - GLSVLSI '18* 15–20 (ACM Press, 2018). doi:10.1145/3194554.3194558

[652] G. Srinivasan, A. Sengupta & K. Roy. Magnetic Tunnel Junction Based Long-Term Short-Term Stochastic Synapse for a Spiking Neural Network with On-Chip STDP Learning. *Sci Rep* **6,** 29545 (2016).

[653] O. Akinola, X. Hu, C. H. Bennett, M. Marinella, J. S. Friedman & J. A. C. Incorvia. Three-terminal magnetic tunnel junction synapse circuits showing spike-timing-dependent plasticity. *J. Phys. D. Appl. Phys.* **52,** 49LT01 (2019).

[654] D. Bhowmik, U. Saxena, A. Dankar, A. Verma, D. Kaushik, S. Chatterjee & U. Singh. On-chip learning for domain wall synapse based Fully Connected Neural Network. *J. Magn. Magn. Mater.* **489,** 165434 (2019).

[655] N. Hassan, X. Hu, L. Jiang-Wei, W. H. Brigner, O. G. Akinola, F. Garcia-Sanchez, M. Pasquale, C. H. Bennett, J. A. C. Incorvia & J. S. Friedman. Magnetic domain wall neuron with lateral inhibition. *J. Appl. Phys.* **124,** (2018).

[656] W. H. Brigner, X. Hu, N. Hassan, C. H. Bennett, J. A. C. Incorvia, F. Garcia-Sanchez & J. S. Friedman. Graded-Anisotropy-Induced Magnetic Domain Wall Drift for an Artificial Spintronic Leaky Integrate-and-Fire Neuron. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **5,** 1–1 (2019).

[657] H. Vakili, W. Zhou, C. T. Ma, S. J. Poon, M. G. Morshed, M. N. Sakib, S. Ganguly, M. Stan, T. Q. Hartnett, P. Balachandran, J. W. Xu, Y. Quessab, A. D. Kent, K. Litzius, G. S. D. Beach & A. W. Ghosh. Skyrmionics-Computing and memory technologies based on topological excitations in magnets. *J. Appl. Phys.* 130, (2021).

[658] T. Leonard, S. Liu, M. Alamdar, C. Cui, O. G. Akinola, L. Xue, T. P. Xiao, J. S. Friedman, M. J. Marinella, C. H. Bennett & J. A. C. Incorvia. Shape-Dependent Multi-Weight Magnetic Artificial Synapses for Neuromorphic Computing. *Under review.* Arxiv Prepr. arXiv: 2111.11516 (2021).

[659] C. Cui, O. G. Akinola, N. Hassan, C. H. Bennett, M. J. Marinella, J. S. Friedman & J. A. C. Incorvia. Maximized lateral inhibition in paired magnetic domain wall racetracks for neuromorphic computing. *Nanotechnology* 31, (2020).

[660] W. H. Brigner, J. S. Friedman, N. Hassan, L. Jiang-Wei, X. Hu, D. Saha, C. H. Bennett, M. J. Marinella, J. A. C. Incorvia & F. Garcia-Sanchez. Shape-Based Magnetic Domain Wall Drift for an Artificial Spintronic Leaky Integrate-and-Fire Neuron. *IEEE Trans. Electron Devices* **66**, 4970–4975 (2019).

[661] S. Liu, C. H. Bennett, J. S. Friedman, M. J. Marinella, D. Paydarfar & J. A. Incorvia. Controllable reset behavior in domain wall-magnetic tunnel junction artificial neurons for task-adaptable computation. *IEEE Magn. Lett.* **11**, 1–5 (2021).

[662] S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett & J. Matthew. A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks. *Appl. Phys. Lett.* 202405, 1–6 (2021).

[663] S. Fukami, N. Ishiwata, N. Kasai, M. Yamanouchi, H. Sato, S. Ikeda & H. Ohno. Scalability prospect of three-terminal magnetic domain-wall motion device. *IEEE Trans. Magn.* **48,** 2152–2157 (2012).

[664] Amarù, L., Testa, E., Couceiro, M., Zografos, O., De Micheli, G. & Soeken, M. Majority logic synthesis. *IEEE/ACM Int. Conf. Comput. Des. Dig. Tech. Pap. ICCAD* 1–6 (2018). doi:10.1145/3240765.3267501

[665] Sivasubramani, S., Mattela, V., Pal, C. & Acharyya, A. Nanomagnetic logic design approach for area and speed efficient adder using ferromagnetically coupled fixed input majority gate. *Nanotechnology* **30,** (2019)

[666] Cowburn, R. P. & Welland, M. E. Room temperature magnetic quantum cellular automata. *Science (80-.).* **287,** 1466–1468 (2000).

[667] Spedalieri, F. M., Jacob, A. P., Nikonov, D. E. & Roychowdhury, V. P. Performance of magnetic quantum cellular automata and limitations due to thermal noise. *IEEE Trans. Nanotechnol.* **10,** 537–546 (2011).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[668] Imre, A., Csaba, G., Ji, L., Orlov, A., Bernstein, G. H. & Porod, W. Majority Logic Gate for Magnetic Quantum-Dot Cellular Automata. *Science (80-.).* **311,** 205–208 (2006).
[669] Fischer, T., Kewenig, M., Bozhko, D. A., Serga, A. A., Syvorotka, I. I., Ciubotaru, F., Adelmann, C., Hillebrands, B. & Chumak, A. V. Experimental prototype of a spin-wave majority gate. *Appl. Phys. Lett.* **110,** (2017).
[670] Klingler, S., Pirro, P., Brächer, T., Leven, B., Hillebrands, B. & Chumak, A. V. Design of a spin-wave majority gate employing mode selection. *Appl. Phys. Lett.* **105,** 1–5 (2014).
[671] Zografos, O., Dutta, S., Manfrini, M., Vaysset, A., Sorée, B., Naeemi, A., Raghavan, P., Lauwereins, R. & Radu, I. P. Non-volatile spin wave majority gate at the nanoscale. *AIP Adv.* **7,** (2017).
[672] Radu, I. P., Zografos, O., Vaysset, A., *et al.* Spintronic majority gates. *Tech. Dig. - Int. Electron Devices Meet. IEDM* **2016-February,** 32.5.1-32.5.4 (2015).
[673] Nikonov, D. E., Bourianoff, G. I. & Ghani, T. Proposal of a spin torque majority gate logic. *IEEE Electron Device Lett.* **32,** 1128–1130 (2011).
[674] Nikonov, D. E., Bourianoff, G. I. & Ghani, T. Nanomagnetic circuits with spin torque majority gates. *Proc. IEEE Conf. Nanotechnol.* 1384–1388 (2011). doi:10.1109/NANO.2011.6144490
[675] Csaba, G. & Porod, W. Behavior of nanomagnet logic in the presence of thermal noise. *2010 14th Int. Work. Comput. Electron. IWCE 2010* 311–314 (2010). doi:10.1109/IWCE.2010.5677954
[676] Sharad, M., Augustine, C., Panagopoulos, G. & Roy, K. Spin-based neuron model with domain-wall magnets as synapse. *IEEE Trans. Nanotechnol.* **11,** 843–853 (2012).
[677] Breitkreutz, S., Kiermaier, J., Eichwald, I., Ju, X., Csaba, G., Schmitt-Landsiedel, D. & Becherer, M. Majority gate for nanomagnetic logic with perpendicular magnetic anisotropy. *IEEE Trans. Magn.* **48,** 4336–4339 (2012).
[678] E. Raymenants, D. Wan, S. Couet, L. Souriau, A. Thiam, Di. Tsvetanova, Y. Canvel, K. Garello, G. S. Kar, M. Heyns, I. Asselberghs, D. E. Nikonov, I. A. Young, S. Pizzini, I. Radu & V. Dai Nguyen. All-Electrical Control of Scaled Spin Logic Devices Based on Domain Wall Motion. *IEEE Trans. Electron Devices* 68, 2116–2122 (2021).
[679] E. Raymenants, O. Bultynck, D. Wan, T. Devolder, K. Garello, L. Souriau, A. Thiam, D. Tsvetanova, Y. Canvel, D. E. Nikonov, I. A. Young, M. Heyns, B. Soree, I. Asselberghs, I. Radu, S. Couet & V. D. Nguyen. Nanoscale domain wall devices with magnetic tunnel junction read and write. *Nat. Electron.* 4, 392–398 (2021).
[680] M. Alamdar, T. Leonard, C. Cui, B. P. Rimal, L. Xue, O. G. Akinola, T. P. Xiao, J. S. Friedman, C. H. Bennett, M. J. Marinella, & J. A. C. Incorvia. Domain wall-magnetic tunnel junction spin-orbit torque devices and circuits for in-memory computing. *Applied Physics Letters* **118**, 112401 (2021).
[681] M. P. Frank, "Physical Foundations of Landauer's Principle," in *Reversible Computation (RC 2018)*. *Lecture Notes in Computer Science*, Kari J., Ulidowski I. (eds), vol. 11106. Springer, Cham, 2018. Extended postprint: arxiv:1901.10327 [cs.ET], 2019.
[682] M.P. Frank, "Foundations of Generalized Reversible Computing," in *Reversible Computation (RC 2017). Lecture Notes in Computer Science*, Phillips I., Rahaman H. (eds), vol 10301. Springer, Cham (2017). Extended postprint: arxiv:1806.10183 [cs.ET], 2018.
[683] M. P. Frank and E. P. DeBenedictis, "New Design Principles for Cold Electronics," *IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference*, San Jose, CA, Oct. 2019. Preprint: https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/SAND2019-8199_C_S3S_v12.pdf. Slides: https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/S3S-2019-talk.pdf.
[684] E. P. DeBenedictis, "New Design Principles for Cold, Scalable Electronics," technical report, Nov. 2019. http://www.zettaflops.org/CATC/DPfC_52ver4.pdf.
[685] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Ed., Cambridge University Press, 2011.
[686] NIST, *Quantum Algorithm Zoo*, website, online at: https://math.nist.gov/quantum/zoo/. Accessed: Nov. 5, 2018.
[687] C. G. Almudever, L. Lao, X. Fu, N. Khammassi, I. Ashraf, D. Iorga, S. Varsamopoulos *et al.*, "The engineering challenges in quantum computing." In 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 836-845. IEEE, 2017.
[688] S. Joshi, C. Kim, S. Ha and G. Cauwenberghs, "From algorithms to devices: Enabling machine learning through ultra-low-power VLSI mixed-signal array processing," *2017 IEEE Custom Integrated Circuits Conference (CICC)*, Austin, TX, pp. 1–9, 2017.
[689] D. Niu, Y. Xiao, and Y. Xie, "Low power memristor-based ReRAM design with error correcting code," in *2012 17th Asia and South Pacific Design Automation Conference (ASP-DAC),* pp. 79–84, 2012.
[690] H. Lim, V. Kornijcuk, J. Y. Seok, S. K. Kim, I. Kim, C. S. Hwang, *et al.*, "Reliability of neuronal information conveyed by unreliable neuristor-based leaky integrate-and-fire neurons: a model study," *Scientific Reports,* vol. 5, p. 9776, May 13, 2015.
[691] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, "Stochastic phase-change neurons," *Nature Nanotechnology,* vol. 11, pp. 693–9, Aug 2016.
[692] H. Lim, H. W. Ahn, V. Kornijcuk, G. Kim, J. Y. Seok, I. Kim, *et al.*, "Relaxation oscillator-realized artificial electronic neurons, their responses, and noise," *Nanoscale,* vol. 8, pp. 9629-40, May 14, 2016.
[693] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct 1990. doi: 10.1109/5.58356
[694] T. P. Xiao, C. H. Bennett, B. Feinberg, S. Agarwal, and M. J. Marinella, "Analog architectures for neural network acceleration based on non-volatile memory," *Applied Physics Reviews*, vol. 7, 031301, 2020.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[695] R. Genov and G. Cauwenberghs, "Charge-mode parallel architecture for vector-matrix multiplication," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing,* vol. 48, no. 10, pp. 930–936, 2001.
[696] A. Coates, B. Huval, T. Wang, D. J. Wu, A. Y. Ng, and B. Catanzaro, "Deep learning with COTS HPC systems," *Proceedings of the 30th International Conference on Machine Learning (ICML),* Atlanta, Georgia, USA, 2013.
[697] T.-H. Yang, H.-Y. Cheng, C.-L. Yang, I.-C. Tseng, H.-W. Hu, H.-S. Chang, and H.-P. Li, "Sparse ReRAM Engine: Joint Exploration of Activation and Weight Sparsity in Compressed Neural Networks," *ACM/IEEE International Symposium on Computer Architecture (ISCA)*, Phoenix, AZ, USA, 2019.
[698] V. Bohossian, P. Hasler, and J. Bruck, "Programmable neural logic," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging*, vol. 21, no. 4, pp. 346–351, Nov. 1998.
[699] R. Robucci, J. Gray, Leung Kin Chiu, J. Romberg, and P. Hasler, "Compressive Sensing on a CMOS Separable-Transform Image Sensor," *IEEE proceedings*, vol. 98, no. 1, 2010, pp. 1089–1101.
[700] S. Koziol, D. Lenz, S. Hilsenbeck, S. Chopra, P. Hasler, and A. Howard, "Using floating-gate based programmable analog arrays for real-time control of a game-playing robot." *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 3566–3571.
[701] P. Hasler and J. Dugger, "An analog floating-gate node for supervised learning," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 5, pp. 834–845, May 2005.
[702] S. Suh, A. Basu, C. Schlottmann, P. Hasler, and J. Barry, "Low-power discrete Fourier transform for OFDM: A programmable analog approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 2, pp. 290–298, 2011.
[703] S. Ramakrishnan and J. Hasler, "Vector-matrix multiply and winner-take-all as an analog classifier," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 353–361, 2014.
[704] S. Agarwal *et al.*, "Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and its Application to Sparse Coding," *Frontiers in Neuroscience,* vol. 9, p. 484, 2016, Art. no. 484.
[705] P. Narayanan *et al.*, "Toward on-chip acceleration of the backpropagation algorithm using nonvolatile memory," *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 11:1-11:7, 2017.
[706] M. J. Marinella *et al.*, "Multiscale Co-Design Analysis of Energy, Latency, Area, and Accuracy of a ReRAM Analog Neural Training Accelerator," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 1, pp. 86-101, 2018.
[707] T. P. Xiao et al., "Analysis and mitigation of parasitic resistance effects for analog in-memory neural network acceleration", *Semiconductor Science and Technology*, vol. 36, no. 11, pp. 114004, 2021.
[708] G. W. Burr *et al.*, "Access Devices for 3D Crosspoint Memory," *Journal of Vacuum Science and Technology B,* vol. 32, no. 4, p. 040802, 2014.
[709] M. Bavandpour, S. Sahay, M. R. Mahmoodi, and D. Strukov, "Efficient Mixed-Signal Neurocomputing via Successive Integration and Rescaling," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* 2019.
[710] V. Reddi et al., "MLPerf Inference Benchmark," ACM/IEEE International Symposium on Computer Architecture (ISCA), Valencia, 2020.
[711] T. P. Xiao et al., "On the Accuracy of Analog Neural Network Inference Accelerators," *arXiv preprint*, arXiv:2109.01262, 2021.
[712] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," *arXiv preprint*, arXiv:1510.00149, 2016.
[713] T. P. Xiao et al., "An Accurate, Error-Tolerant, and Energy-Efficient Neural Network Inference Engine Based on SONOS Analog Memory," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1-14, 2022.
[714] M. N. Bojnordi and E. Ipek, "Memristive Boltzmann machine: A hardware accelerator for combinatorial optimization and deep learning," *2016 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, Barcelona, 2016, pp. 1-13.
[715] A. Shafiee *et al.*, "ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars," *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Seoul, 2016, pp. 14-26.
[716] P. Chi *et al.*, "PRIME: A Novel Processing-in-Memory Architecture for Neural Network Computation in ReRAM-Based Main Memory," *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Seoul, 2016, pp. 27-39
[717] L. Song, Y. Zhuo, X. Qian, H. Li and Y. Chen, "GraphR: Accelerating Graph Processing Using ReRAM," *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, Vienna, 2018, pp. 531-543
[718] B. Feinberg, U. K. R. Vengalam, N. Whitehair, S. Wang and E. Ipek, "Enabling Scientific Computing on Memristive Accelerators," *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, Los Angeles, CA, 2018, pp. 367-382
[719] B. E. Boser, E. Sackinger, J. Bromley, Y. Le Cun, and L. D. Jackel, "An analog neural network processor with programmable topology," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 12, pp. 2017-2025, 1991.
[720] F. J. Kub, K. K. Moon, I. A. Mack, and F. M. Long, "Programmable analog vector-matrix multipliers," *IEEE Journal of Solid-State Circuits,* vol. 25, no. 1, pp. 207-214, 1990.
[721] S. Kim, T. Gokmen, H.-M. Lee, and W. F. Haensch, "Analog CMOS-based resistive processing unit for deep neural network training," *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017.
[722] D. Soudry, D. D. Castro, A. Gal, A. Kolodny, and S. Kvatinsky, "Memristor-Based Multilayer Neural Networks With Online Gradient Descent Training," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, 2015.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[723] D. Kadetotad *et al.*, "Parallel Architecture With Resistive Crosspoint Array for Dictionary Learning Acceleration," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on,* vol. 5, no. 2, pp. 194–204, 2015.
[724] I. Gao *et al*., "Fully parallel write/read in resistive synaptic array for accelerating on-chip learning," *Nanotechnology*, vol 26, 455204, 2015.
[725] G. W. Burr *et al.*, "Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element," *Electron Devices, IEEE Transactions on,* vol. 62, no. 11, pp. 3498–3507, 2015.
[726] P. Y. Chen, L. Gao, and S. Yu, "Design of Resistive Synaptic Array for Implementing On-Chip Sparse Learning," *IEEE Transactions on Multi-Scale Computing Systems,* vol. 2, no. 4, pp. 257–264, 2016.
[727] S. Agarwal et al., "Resistive memory device requirements for a neural algorithm accelerator," *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016.
[728] R. B. Jacobs-Gedrim et al., "Impact of Linearity and Write Noise of Analog Resistive Memory Devices in a Neural Algorithm Accelerator," presented at the *IEEE International Conference on Rebooting Computing (ICRC)* Washington, DC, November 2017.
[729] E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Advanced Materials,* vol. 29, no. 4, p. 1604310, 2017.
[730] Y. van de Burgt *et al.*, "A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing," *Nat Mater,* Letter vol. 16, no. 4, pp. 414–418, 2017.
[731] S. Agarwal et al., "Achieving ideal accuracies in analog neuromorphic computing using periodic carry," *VLSI Technology, 2017 Symposium on*, 2017, pp. T174–T175: IEEE.
[732] I. Boybat et al., "Improved Deep Neural Network Hardware-Accelerators Based on Non-Volatile-Memory: The Local Gains Technique," *2017 IEEE International Conference on Rebooting Computing (ICRC)*, 2017, pp. 1–8.
[733] Gokmen, Tayfun, and Wilfried Haensch. "Algorithm for Training Neural Networks on Resistive Device Arrays." *arXiv preprint arXiv:1909.07908* (2019).
[734] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, et al. "Equivalent-Accuracy Accelerated Neural-Network Training Using Analogue Memory," *Nature*, vol. 558, no. 7708, pp. 60–67, 2018.
[735] S. Agarwal et al. (2017). *CrossSim*. Available: http://cross-sim.sandia.gov.
[736] T. Gokmen and Y. Vlasov, "Acceleration of deep neural network training with resistive cross-point devices," *arXiv preprint arXiv:1603.07341,* 2016.
[737] E. J. Fuller *et al.*, "Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing," *Science*, vol. 364, no. 6640, pp. 570-574, 2019.
[738] B. D. Hoskins, et al., "Streaming Batch Eigenupdates for Hardware Neural Networks," *Frontiers in Neuroscience*, vol. 13, no. 793, 2019.
[739] C. Schlottmann and P. Hasler, "A highly dense, low power, programmable analog vector-matrix multiplier: The FPAA implementation," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 3, pp. 403–411, 2011.
[740] C. Schlottmann, S. Shapero, S. Nease, and P. Hasler, "A digitally enhanced dynamically reconfigurable analog platform for low-power signal processing," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 9, pp. 2174–2184, 2012.
[741] S. George, S. Kim, S. Shah, J. Hasler, M. Collins, F. Adil, R. Wunderlich, S. Nease, and S. Ramakrishnan, "A Programmable and Configurable Mixed-Mode FPAA SOC," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 24, no. 6, 2016. pp. 2253–2261.
[742] J. Hasler, "Large-Scale Field-Programmable Analog Arrays,'' IEEE Proceedings, Dec 2019
[743] C. M. Twigg, J. D. Gray, and P. Hasler, "Programmable floating gate FPAA switches are not dead weight," *IEEE International Symposium on Circuits and Systems*, May 2007, pp. 169–172.
[744] S. Shapero and P. Hasler, "Mismatch characterization and calibration for accurate and automated analog design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 548–556, 2013.
[745] S. Kim, S. Shah, and J. Hasler, "Calibration of Floating-Gate SoC FPAA System," IEEE Transactions on VLSI, 2017.
[746] S. Shah, H. Toreyin, J. Hasler, and A. Natarajan, "Models and Techniques For Temperature Robust Systems On A Reconfigurable Platform," *Journal of Low Power Electronics Applications*, vol. 7, no. 21, August 2017. pp. 1–14.
[747] J. Hasler, C. Scholttmann, and S. Koziol, "FPAA chips and tools as the center of an design-based analog systems education," *IEEE Microelectronic Systems Education (MSE),* 2011, pp. 47–51.
[748] M. Collins, J. Hasler, and S. George, "Analog systems education: An integrated toolset and FPAA SoC boards," *Microelectronics Systems Education (MSE),* 2015, pp. 32–35.
[749] J. Hasler and H. Wang, "A Fine-Grain FPAA fabric for RF+Baseband," *GOMAC*, March 2015.
[750] J. Hasler, S. Kim, and F. Adil, "Scaling Floating-Gate Devices Predicting Behavior for Programmable and Configurable Circuits and Systems," *Journal of Low Power Electronics Applications*, vol. 6, no. 13, 2016, pp. 1–19.
[751] P. Blouw, X. Choo, E. Hunsberger, and C. Eliasmith, "Benchmarking keyword spotting efficiency on neuromorphic hardware," Apr. 2019. arXiv:1812.01739. [Online]. Available: https://arxiv.org/abs/1812.01739
[752] I. Richter *et al.*, "Memristive Accelerator for Extreme Scale Linear Solvers," *40th Annual GOMACTech Conference*, St. Louis, MO, 2015.
[753] Sun, Z., Pedretti, G., Ambrosi, E., Bricalli, A., Wang, W., Ielmini, D. Solving matrix equations in one step with crosspoint resistive arrays. Proc. Natl. Acad. Sci. (PNAS) 116(10), 4123-4128 (2019).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[754] Sun, Z., Pedretti, G., Ielmini, D. Fast solution of linear systems with analog resistive switching memory (RRAM). 2019 International Conference on Rebooting Computing (ICRC) 2019.
[755] Sun, Z., Pedretti, G., Bricalli, A., Ielmini, D. One-step regression and classification with crosspoint resistive memory arrays. Sci. Adv. (2019). Accepted for publication
[756] S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, "Mixed-precision architecture based on computational memory for training deep neural networks." *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018.
[757] Xiaobo Sharon Hu, Michael T. Niemier, Arman Kazemi, Ann Franchesca B. Laguna, Kai Ni, Ramin Rajaei, Mohammad Mehdi Sharifi, Xunzhao Yin, "In-Memory Computing with Associative Memories: A Cross-Layer Perspective," *IEEE International Electron Devices Meeting (IEDM)*, 2021.
[758] H. Amrouch *et al*., "Cross-layer Design for Computing-in-Memory: From Devices, Circuits, to Architectures and Applications," *2021 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2021, pp. 132-139.
[759] Ni, Kai, et al. "Ferroelectric ternary content-addressable memory for one-shot learning." *Nature Electronics* 2.11 (2019): 521-529.
[760] Kazemi, Arman, et al. "In-memory nearest neighbor search with fefet multi-bit content-addressable memories." *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2021.
[761] Kazemi, Arman, et al. "A flash-based multi-bit content-addressable memory with euclidean squared distance." *2021 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. IEEE, 2021.
[762] Yin, Xunzhao, et al. "FeCAM: A universal compact digital and analog content addressable memory using ferroelectric." *IEEE Transactions on Electron Devices* 67.7 (2020): 2785-2792.
[763] Rajaei, Ramin, et al. "Compact single-phase-search multistate content-addressable memory design using one FeFET/cell." *IEEE Transactions on Electron Devices* 68.1 (2020): 109-117.
[764] Li, Chao, et al. "A scalable design of multi-bit ferroelectric content addressable memory for data-centric computing." *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2020.
[765] Li, C., Graves, C. E., Sheng, X., Miller, D., Foltin, M., Pedretti, G., & Strachan, J. P. (2020). Analog content-addressable memories with memristors. *Nature communications*, *11*(1), 1-8.
[766] Li, Jing, Robert Montoye, Masatoshi Ishii, Kevin Stawiasz, Takeshi Nishida, Kim Maloney, Gary Ditlow et al. "1Mb 0.41 µm 2 2T-2R cell nonvolatile TCAM with two-bit encoding and clocked self-referenced sensing." In *2013 Symposium on VLSI Technology*, pp. C104-C105. IEEE, 2013.
[767] Yang, H. Z., et al. "A Novel High-Density and Low-Power Ternary Content Addressable Memory Design Based on 3D NAND Flash." *2020 IEEE Silicon Nanoelectronics Workshop (SNW)*. IEEE, 2020.
[768] Wang, Fei, et al. "Implementation of data search in multi-level NAND flash memory by complementary storage scheme." *IEEE Electron Device Letters* 41.8 (2020): 1189-1192.
[769] Matsunaga, Shoun, et al. "Design of a nine-transistor/two-magnetic-tunnel-junction-cell-based low-energy nonvolatile ternary content-addressable memory." *Japanese Journal of Applied Physics* 51.2S (2012): 02BM06.
[770] C. Wang, D. Zhang, L. Zeng, E. Deng, J. Chen and W. Zhao, "A Novel MTJ-Based Non-Volatile Ternary Content-Addressable Memory for High-Speed, Low-Power, and High-Reliable Search Operation," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 4, pp. 1454-1464, April 2019, doi: 10.1109/TCSI.2018.2885343.
[771] Guo, Qing, et al. "A resistive TCAM accelerator for data-intensive computing." *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*. ACM, 2011.
[772] L. Yavits, S. Kvatinsky, A. Morad, and R. Ginosar, "Resistive Associative Processor," *IEEE Computer Architecture Letters,* vol. 14, no. 2, pp. 148–151, 2015.
[773] Shinde, Rajendra, et al. "Similarity search and locality sensitive hashing using ternary content addressable memories." *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010.
[774] M. Imani, D. Peroni, A. Rahimi, and T. Rosing, "Resistive CAM Acceleration for Tunable Approximate Computing," *IEEE Transactions on Emerging Topics in Computing,* vol. PP, no. 99, p. 1, 2016.
[775] Ly, Denys RB, et al. "In-depth Characterization of Resistive Memory-Based Ternary Content Addressable Memories." *2018 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2018
[776] C. Yakopcic, V. Bontupalli, R. Hasan, D. Mountain, and T. M. Taha, "Self-biasing memristor crossbar used for string matching and ternary content-addressable memory implementation," *Electronics Letters,* vol. 53, no. 7, pp. 463–465, 2017.
[777] Graves, Catherine E., et al. "Regular Expression Matching with Memristor TCAMs." *2018 IEEE International Conference on Rebooting Computing (ICRC)*. IEEE, 2018.
[778] Zhi-Hua Zhou and Ji Feng, Deep Forest: Towards An Alternative to Deep Neural Networks, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, p 3553—3559, 2017.
[779] IMCRYPTO: An In-Memory Computing Fabric for AES Encryption and Decryption, Dayane Reis and Haoran Geng and Michael Niemier and Xiaobo Sharon Hu, 2021, 2112.02231, arXiv.
[780] W. A. Wulf and S. A. McKee, "Hitting the memory wall: implications of the obvious", *SIGARCH Computer Architecture News,* vol 23, no. 1, 1995.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[781] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295-2329, 2017.
[782] X. Guo et al, "Fast, energy-efficient, robust, and reproducible mixed-signal neuromorphic classifier based on embedded NOR flash memory technology," *International Electron Devices Meeting (IEDM)*, pp. 6.5.1-6.5.4, 2017.
[783] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, "Fully hardware-implemented memristor convolutional neural network," *Nature*, vol. 577, no. 7792, pp. 641-646, 2020.
[784] A. Ankit, et al., "PUMA: A programmable ultra-efficient memristor-based accelerator for machine learning inference," *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS),* pp. 715-731, 2019.
[785] M. Hu, *et al.*, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," *Design Automation Conference (DAC)*, pp. 1-6, 2016.
[786] V. Joshi, *et al.*, "Accurate deep neural network inference using computational phase-change memory," *Nature Communications,* vol. 11, no. 2473, 2020.
[787] A. Nag, *et al.*, "Newton: Gravitating towards the physical limits of crossbar acceleration," *IEEE Micro,* vol. 38, no. 5, pp. 41-49, 2018.
[788] Z. He, J. Lin, R. Ewetz, J.-S. Yuan, and D. Fan, "Noise injection adaption: End-to-end ReRAM crossbar non-ideal effect adaption for neural network mapping," *Design Automation Conference (DAC)*, pp. 57:1-57:6, 2019.
[789] S. Jain and A. Raghunathan, "CxDNN: Hardware-software compensation methods for deep neural networks on resistive crossbar systems," *ACM Transactions on Embedded Computer Systems*, vol. 18, no. 6, 2019.
[790] M. Klachko, M. R. Mahmoodi, and D. Strukov, "Improving noise tolerance of mixed-signal neural networks," *International Joint Conference on Neural Networks (IJCNN),* pp. 1-8, 2019.
[791] Y. Long, X. She, and S. Mukhopadhyay, "Design of reliable DNN accelerator with unreliable ReRAM," *Design, Automation Test in European Conference Exhibition (DATE)*, pp. 1769-1774, 2019.
[792] Y. Jeong, M. A. Zidan, and W. D. Lu, "Parasitic effect analysis in memristor-array-based neuromorphic systems," *IEEE Transactions on Nanotechnology*, vol. 17, no. 1, pp. 184-193, 2018.
[793] F. Zhang and M. Hu, "Mitigate parasitic resistance in resistive crossbar-based convolutional neural networks," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 16, no. 3, 2020.
[794] S. Ambrogio, et al., "Reducing the Impact of Phase-Change Memory Conductance Drift on the Inference of Large-Scale Hardware Neural Networks," *IEEE International Electron Devices Meeting (IEDM)*, 2019.
[795] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," *Nature materials,* vol. 12, no. 2, pp. 114-117, 2013.
[796] Locatelli, N., Cros, V. & Grollier, J. Spin-torque building blocks. *Nature Mater* **13,** 11–20 (2014).
[797] Tuma, T., Pantazi, A., Le Gallo, M. *et al.* Stochastic phase-change neurons. *Nature Nanotech* **11,** 693–699 (2016).
[798] J. M. Shainline, S. M. Buckley, A. N. McCaughan, J. T. Chiles, A. J. Salim, M. Castellanos-Beltran, C. A. Donnelly, M. L. Schneider, R. P. Mirin, and S. W. Nam, "Superconducting optoelectronic loop neurons," *Journal of Applied Physics,* vol. 126, no. 4, p. 044902, 2019
[799] S. R. Nandakumar, *et al.*, "Mixed-precision deep learning based on computational memory," *Frontiers in Neuroscience*, vol. 14, no. 406, 2020.
[800] C. H. Bennett, D. Garland, R. B. Jacobs-Gedrim, S. Agarwal, and M. J. Marinella, "Wafer-Scale TaO_x device variability and implications for neuromorphic computing applications," *IEEE International Reliability Physics Symposium (IRPS)*, pp. 1-4, 2019.
[801] X. Sun and S. Yu, "Impact of Non-Ideal Characteristics of Resistive Synaptic Devices on Implementing Convolutional Neural Networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 3, pp. 570-579, 2019.
[802] Y. Li, et al., "Filament-Free Bulk Resistive Memory Enables Deterministic Analogue Switching," *Advanced Materials*, vol. 32, no. 45, 2003984, 2020.
[803] Y. Li, et al., "In situ parallel training of analog neural network using electrochemical random access memory," *Frontiers in Neuroscience*, vol. 15, 636127, 2021.
[804] S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett, and M. J. Marinella, "A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks," *Applied Physics Letters*, vol. 118, 202405, 2021.
[805] M. Alamdar, et al., "Domain wall-magnetic tunnel junction spin-orbit torque devices and circuits for in-memory computing," *Applied Physics Letters,* vol. 118, 112401, 2021.
[806] D. Kaushik, J. Sharda, and D. Bhowmik, "Synapse cell optimization and back-propagation algorithm implementation in a domain wall synapse based crossbar neural network for scalable on-chip learning," *Nanotechnology,* no. 31, 364004, 2020.
[807] M. Suri, et al., "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," *IEEE International Electron Devices Meeting (IEDM)*, pp. 4.4.1-4.4.4, 2011.
[808] Zenke, F., and E. O. Neftci. "Brain-Inspired Learning on Neuromorphic Substrates." Proceedings of the IEEE, 2021, 1–16. https://doi.org/10.1109/JPROC.2020.3045625.
[809] Richards, Blake A, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, et al. "A Deep Learning Framework for Neuroscience." Nature Neuroscience 22, no. 11 (2019): 1761–70.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[810] Gerstner, Wulfram, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. “Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of Neohebbian Three-Factor Learning Rules.” Frontiers in Neural Circuits 12 (2018): 53.
[811] Bellec, Guillaume, Franz Scherr, Anand Subramoney, Elias Hajek, Darjan Salaj, Robert Legenstein, and Wolfgang Maass. “A Solution to the Learning Dilemma for Recurrent Networks of Spiking Neurons.” Nature Communications 11, no. 1 (2020): 1–15.
[812] Lillicrap, Timothy P, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. “Backpropagation and the Brain.” Nature Reviews Neuroscience, 2020, 1–12.
[813] Illing, Bernd, Wulfram Gerstner, and Guillaume Bellec. “Towards Truly Local Gradients with CLAPP: Contrastive, Local And Predictive Plasticity.” ArXiv Preprint ArXiv:2010.08262, 2020.
[814] Whittington, James C. R., and Rafal Bogacz. “Theories of Error Back-Propagation in the Brain.” Trends in Cognitive Sciences 23, no. 3 (2019): 235–50. https://doi.org/10.1016/j.tics.2018.12.005.
[815] Jaderberg, Max, Volodymyr Mnih, Wojciech Marian Czarnecki, Tom Schaul, Joel Z Leibo, David Silver, and Koray Kavukcuoglu. “Reinforcement Learning with Unsupervised Auxiliary Tasks.” ArXiv Preprint ArXiv:1611.05397, 2016.
[816] Laskin, Michael, Luke Metz, Seth Nabarrao, Mark Saroufim, Badreddine Noune, Carlo Luschi, Jascha Sohl-Dickstein, and Pieter Abbeel. “Parallel Training of Deep Networks with Local Updates.” ArXiv Preprint ArXiv:2012.03837, 2020.
[817] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive computation*, vol. 1, no. 2, pp. 139-159, 2009.
[818] E. P. Frady, D. Kleyko, C. J. Kymn, B. A. Olshausen, and F. T. Sommer, "Computing on Functions Using Randomized Vector Representations," *arXiv preprint arXiv:2109.03429,* 2021.
[819] Justin C. Wong. “Negative capacitance and hyperdimensional computing for unconventional low-power computing.” PhD thesis, EECS Department, University of California, Berkeley, Dec 2018.
[820] G. E. Hinton, "Mapping part-whole hierarchies into connectionist networks," *Artificial Intelligence,* vol. 46, no. 1-2, pp. 47-75, 1990.
[821] T. Plate, "Holographic Reduced Representations: Convolution Algebra for Compositional Distributed Representations," in *IJCAI*, 1991: Citeseer, pp. 30-35.
[822] Z. Zou, Y. Kim, F. Imani, H. Alimohamadi, R. Cammarota, and M. Imani, "Scalable edge-based hyperdimensional learning system with brain-like neural adaptation," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2021, pp. 1-15.
[823] A. Mitrokhin, P. Sutor, C. Fermüller, and Y. Aloimonos, "Learning sensorimotor control with neuromorphic sensors: Toward hyperdimensional active perception," *Science Robotics,* vol. 4, no. 30, 2019.
[824] A. Hernandez-Cane, N. Matsumoto, E. Ping, and M. Imani, "Onlinehd: Robust, efficient, and single-pass online learning using hyperdimensional system," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021: IEEE, pp. 56-61.
[825] G. Karunaratne *et al.*, "Robust high-dimensional memory-augmented neural networks," *Nature communications,* vol. 12, no. 1, pp. 1-12, 2021.
[826] D. Kleyko *et al.*, "Vector Symbolic Architectures as a Computing Framework for Nanoscale Hardware," *arXiv preprint arXiv:2106.05268,* 2021.
[827] M. Imani, A. Rahimi, D. Kong, T. Rosing, and J. M. Rabaey, "Exploring hyperdimensional associative memory," in *2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2017: IEEE, pp. 445-456.
[828] G. Karunaratne, M. Le Gallo, G. Cherubini, L. Benini, A. Rahimi, and A. Sebastian, "In-memory hyperdimensional computing," *Nature Electronics,* vol. 3, no. 6, pp. 327-337, 2020.
[829] H. Li *et al.*, "Hyperdimensional computing with 3D VRRAM in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016: IEEE, pp. 16.1. 1-16.1. 4.
[830] P. Poduval, Z. Zou, H. Najafi, H. Homayoun, and M. Imani, "Stochd: Stochastic hyperdimensional system for efficient and robust learning from raw data," in *2021 58th ACM/IEEE Design Automation Conference (DAC)*, 2021: IEEE, pp. 1195-1200.
[831] A. Rahimi, P. Kanerva, and J. M. Rabaey, "A robust and energy-efficient classifier using brain-inspired hyperdimensional computing," in *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*, 2016, pp. 64-69.
[832] A. Hérnandez-Cano, R. Cammarota, and M. Imani, "Prid: Model inversion privacy attacks in hyperdimensional learning systems," in *2021 58th ACM/IEEE Design Automation Conference (DAC)*, 2021: IEEE, pp. 553-558.
[833] M. Imani *et al.*, "A framework for collaborative learning in secure high-dimensional space," in *2019 IEEE 12th International Conference on Cloud Computing (CLOUD)*, 2019: IEEE, pp. 435-446.
[834] R. Thapa, D. Ma, and X. Jiao, "HDXplore: Automated Blackbox Testing of Brain-Inspired Hyperdimensional Computing," *arXiv preprint arXiv:2105.12770,* 2021.
[835] F. Yang and S. Ren, "Adversarial Attacks on Brain-Inspired Hyperdimensional Computing-Based Classifiers," *arXiv preprint arXiv:2006.05594,* 2020.
[836] S. Aaronson, *Quantum Computing since Democritus*, Cambridge University Press, 2013.
[837] T. P. Xiao, “Optoelectronics for refrigeration and analog circuits for combinatorial optimization,” Ph.D. dissertation, Chapter 6, University of California, Berkeley, 2019.
[838] T. Wang and J. Roychowdhury, “Oscillator-based Ising machine,” arXiv e-prints, 2017. arXiv: 1709.08102

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[839] T. Wang, L. Wu, and J. Roychowdhury, "New computational results and hardware prototypes for oscillator-based Ising machines," *Proceedings of the 56th Annual Design Automation Conference*, 2019.
[840] J. Chou, S. Bramhavar, S. Ghosh, and W. Herzog, "Analog coupled oscillator based weighted Ising machine," arXiv e-prints, 2019. arXiv: 1906.06312
[841] A. Lucas, "Ising formulations of many NP problems," *Interdisciplinary Physics*, vol. 2, p. 5, 2014.
[842] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, "Coherent Ising machine based on degenerate optical parametric oscillators," *Phys. Rev. A*, vol 88, p. 063853, 2013.
[843] T. Inagaki et al, "A coherent Ising machine for 2000-node optimization problems," *Science*, vol. 354, no. 6312, pp. 603-606, 2016.
[844] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully-programmable 100-spin coherent Ising machine with all-to-all connections," *Science*, p. aah5178, Oct. 2016.
[845] I. Mahboob, H. Okamoto, and H. Yamaguchi, "An electromechanical Ising Hamiltonian," *Science Advances*, vol 2, no. 6, 2016.
[846] S. K. Vadlamani, T. P. Xiao, E. Yablonovitch, "Physics successfully implements Lagrange multiplier optimization," *Proceedings of the National Academy of Sciences*, vol. 117, no. 43, pp. 26639-26650, 2020.
[847] S. K. Vadlamani, "Sharp Switching in Tunnel Transistors and Physics-based Machines for Optimization," Ph. D. dissertation, Chapter 5, University of California, Berkeley, 2021.
[848] T. Leleu, Y. Yamamoto, P. L. McMachon, K. Aihara, "Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity," *Physical Review Letters*, vol. 122, no. 4, p. 040607, 2019.
[849] E. Goto, "New Parametron circuit element using nonlinear reactance," *KDD Kenkyu Shiryo*, 1954.
[850] J. von Neumann, "Non-linear Capacitance or Inductance Switching, Amplifyinig, and Memory Organs," U. S. Patent No. 2,815,488, Dec. 1957.
[851] E. Goto, "The parametron, a digital computing element which utilizes parametric oscillation," *Proceedings of the IRE*, vol. 47, no. 8, pp. 1304-1316, 1959.
[852] R. L. Wigington, "A New Concept in Computing," in *Proceedings of the IRE*, vol. 47, no. 4, pp. 516–523, April 1959, doi:10.1109/JRPROC.1959.287311.
[853] M. Ercsey-Ravasz and Z. Toroczkai, "Optimization hardness as transient chaos in an analog approach to constraint satisfaction," *Nature Physics*, vol. 7, no. 12, pp. 966–970, Dec. 2011.
[854] F. L. Traversa, C. Ramella, F. Bonani, and M. Di Ventra, "Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states," *Science Advances*, vol. 1, no. 6, pp. e1500031–e1500031, Jul. 2015.
[855] V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 418–423, Jan. 2007.
[856] G. Appa, L. S. Pitsoulis, and H. P. Williams, Eds., *Handbook on modelling for discrete optimization*. New York: Springer, 2006.
[857] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982.
[858] F. C. Hoppensteadt and E. M. Izhikevich, "Oscillatory neurocomputers with dynamic connectivity," *Physical Review Letters*, vol. 82, no. 14, p. 2983, 1999.
[859] S. Levitan, Y. Fang, D. Dash, T. Shibata, D. Nikonov, and G. Bourianoff, "Non-Boolean associative architectures based on nano-oscillators," *2012 13th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, 2012, pp. 1–6.
[860] G. Csaba, M. Pufall, D. Nikonov, G. Bourianoff, A. Horvath, T. Roska, and W. Porod, "Spin torque oscillator models for applications in associative memories," *2012 13th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, 2012, pp. 1–2.
[861] M. Namba and Z. Zhang, "Cellular neural network for associative memory and its application to Braille image recognition," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 2006, pp. 2409–2414.
[862] D. Liu and A. N. Michel, "Cellular neural networks for associative memories," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on,* vol. 40, pp. 119–121, 1993.
[863] P. Szolgay, I. Szatmari, and K. Laszlo, "A fast fixed point learning method to implement associative memory on CNNs," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 44, pp. 362–366, 1997.
[864] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks" 1985 *Phys. Rev. Lett.* 55:1530. 10.1103/PhysRevLett.55.1530
[865] L. O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1257–1272, Oct. 1988.
[866] L. Yang, L. Chua, and K. Krieg, "VLSI implementation of cellular neural networks," *Circuits and Systems, IEEE International Symposium on*, 1990, pp. 2425–2427.
[867] E. Y. Chou, B. J. Sheu, and R. C. Chang, "VLSI design of optimization and image processing cellular neural networks," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 44, pp. 12–20, 1997.
[868] A. R. Trivedi and S. Mukhopadhyay, "Potential of ultralow-power cellular neural image processing with Si/Ge tunnel FET," *Nanotechnology, IEEE Transactions on,* vol. 13, pp. 627–629, 2014.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[869] I. Palit, X. S. Hu, J. Nahas, and M. Niemier, “TFET-based cellular neural network architectures,” *Proceedings of the 2013 International Symposium on Low Power Electronics and Design*, 2013, pp. 236-241.
[870] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, “Proposal for an all-spin logic device with built-in memory,” *Nature Nanotechnology,* vol. 5, pp. 266–270, 2010.
[871] S. Datta, S. Salahuddin, and B. Behin-Aein, “Non-volatile spin switch for Boolean and non-Boolean logic,” *Applied Physics Letters,* vol. 101, p. 252411, 2012.
[872] D. Morris, D. Bromberg, J.-G. J. Zhu, and L. Pileggi, “mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices,” *Proceedings of the 49th Annual Design Automation Conference*, 2012, pp. 486–491.
[873] C. Pan and A. Naeemi, “Non-Boolean computing benchmarking for beyond-CMOS devices based on cellular neural network,” *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on,* vol. 2, pp. 36–43, 2016.
[874] Ulmann, Bernd. Analog Computing, München: Oldenbourg Wissenschaftsverlag, 2013. https://doi.org/10.1524/9783486755183
[875] G. E. R. Cowan, R. C. Melville and Y. P. Tsividis, "A VLSI analog computer/digital computer accelerator," in IEEE Journal of Solid-State Circuits, vol. 41, no. 1, pp. 42-53, Jan. 2006, doi: 10.1109/JSSC.2005.858618.
[876] A. Basu et al., "A Floating-Gate-Based Field-Programmable Analog Array," in IEEE Journal of Solid-State Circuits, vol. 45, no. 9, pp. 1781-1794, Sept. 2010, doi: 10.1109/JSSC.2010.2056832.
[877] N. Guo et al., "Energy-Efficient Hybrid Analog/Digital Approximate Computation in Continuous Time," in IEEE Journal of Solid-State Circuits, vol. 51, no. 7, pp. 1514-1524, July 2016, doi: 10.1109/JSSC.2016.2543729.
[878] S. George et al., "A Programmable and Configurable Mixed-Mode FPAA SoC," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 24, no. 6, pp. 2253-2261, June 2016, doi: 10.1109/TVLSI.2015.2504119.
[879] G. E. R. Cowan, R. C. Melville and Y. P. Tsividis, "A VLSI analog computer/math co-processor for a digital computer," ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005., 2005, pp. 82-586 Vol. 1, doi: 10.1109/ISSCC.2005.1493879.
[880] N. Guo et al., "Continuous-time hybrid computation with programmable nonlinearities," ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC), 2015, pp. 279-282, doi: 10.1109/ESSCIRC.2015.7313881.
[881] Y. Huang, N. Guo, S. Sethumadhavan, M. Seok and Y. Tsividis, "A Case Study in Analog Co-Processing for Solving Stochastic Differential Equations," 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), 2018, pp. 1-5, doi: 10.1109/ICDSP.2018.8631831.
[882] J. Hasler, “Starting Framework for Analog Numerical Analysis," *Journal of Low Power Electronics Applications*, vol. 7, no. 17, June 2017.
[883] N. Udayanga, S. I. Hariharan, S. Mandal, L. Belostotski, L. T. Bruton and A. Madanayake, "Continuous-Time Algorithms for Solving Maxwell’s Equations Using Analog Circuits," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 66, no. 10, pp. 3941-3954, Oct. 2019, doi: 10.1109/TCSI.2019.2915200.
[884] N. Udayanga et al., "A Radio Frequency Analog Computer for Computational Electromagnetics," in IEEE Journal of Solid-State Circuits, vol. 56, no. 2, pp. 440-454, Feb. 2021, doi: 10.1109/JSSC.2020.3018432.
[885] Huang, Yipeng. Hybrid Analog-Digital Co-Processing for Scientific Computation. Columbia University, 2018.
[886] Y. Huang, N. Guo, M. Seok, Y. Tsividis and S. Sethumadhavan, "Analog Computing in a Modern Context: A Linear Algebra Accelerator Case Study," in IEEE Micro, vol. 37, no. 3, pp. 30-38, 2017, doi: 10.1109/MM.2017.55.
[887] Zhong Sun, Giacomo Pedretti, Elia Ambrosi, Alessandro Bricalli, Wei Wang, Daniele Ielmini. Solving matrix equations in one step with cross-point resistive arrays. Proceedings of the National Academy of Sciences Mar 2019, 116 (10) 4123-4128; DOI: 10.1073/pnas.1815682116
[888] Y. Huang, N. Guo, M. Seok, Y. Tsividis, K. Mandli and S. Sethumadhavan, "Hybrid Analog-Digital Solution of Nonlinear Partial Differential Equations," 2017 50th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), 2017, pp. 665-678.
[889] H. Jaeger, “Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication,” *Science*, vol. 304, no. 5667, pp. 78–80, Apr. 2004, doi: 10.1126/science.1091277.
[890] G. Tanaka *et al.*, “Recent advances in physical reservoir computing: A review,” *Neural Networks*, vol. 115, pp. 100–123, Jul. 2019, doi: 10.1016/j.neunet.2019.03.005.
[891] G. V. der Sande, D. Brunner, and M. C. Soriano, “Advances in photonic reservoir computing,” *Nanophotonics*, vol. 6, no. 3, pp. 561–576, May 2017, doi: 10.1515/nanoph-2016-0132.
[892] Y. Paquot *et al.*, “Optoelectronic Reservoir Computing,” *Sci Rep*, vol. 2, no. 1, p. 287, Dec. 2012, doi: 10.1038/srep00287.
[893] G. Dion, S. Mejaouri, and J. Sylvestre, “Reservoir computing with a single delay-coupled non-linear mechanical oscillator,” *Journal of Applied Physics*, vol. 124, no. 15, p. 152132, Oct. 2018, doi: 10.1063/1.5038038.
[894] L. Appeltant *et al.*, “Information processing using a single dynamical node as complex system,” *Nat Commun*, vol. 2, no. 1, p. 468, Sep. 2011, doi: 10.1038/ncomms1476.
[895] N. D. Haynes, M. C. Soriano, D. P. Rosin, I. Fischer, and D. J. Gauthier, “Reservoir computing with a single time-delay autonomous Boolean node,” *Phys. Rev. E*, vol. 91, no. 2, p. 020801, Feb. 2015, doi: 10.1103/PhysRevE.91.020801.
[896] M. J. Marinella and S. Agarwal, “Efficient reservoir computing with memristors,” *Nature Electronics*, vol. 2, pp. 437–438, Oct. 2019.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[897] J. Torrejon *et al.*, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, no. 7664, pp. 428–431, Jul. 2017, doi: 10.1038/nature23011.
[898] G. Ribeill, M.-H. Nguyen, L. Govia, A. Wagner, W. Barbosa, D. Gauthier, T. Ohki and G. Rowlands, "Superconducting circuits as reservoir computers for high speed data processing," presented at Applied Superconductivity Conference, IEEE, Nov. 2020.
[899] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. H. Liew, "Quantum reservoir processing," *npj Quantum Inf*, vol. 5, no. 1, p. 35, Dec. 2019, doi: 10.1038/s41534-019-0149-8.
[900] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, H. K. Krovi, and T. A. Ohki, "Quantum reservoir computing with a single nonlinear oscillator," *arXiv:2004.14965 [cond-mat, physics:quant-ph]*, Apr. 2020, Accessed: Jul. 09, 2020. [Online]. Available: http://arxiv.org/abs/2004.14965.
[901] W. Maass, T. Natschläger, and H. Markram, "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, Nov. 2002, doi: 10.1162/089976602760407955.
[902] C. Gallicchio, A. Micheli, and L. Pedrelli, "Deep reservoir computing: A critical experimental analysis," *Neurocomputing*, vol. 268, pp. 87–99, Dec. 2017, doi: 10.1016/j.neucom.2016.12.089.
[903] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, January 1949.
[904] M.P. Frank and E. P. DeBenedictis, "A novel operational paradigm for thermodynamically reversible logic: Adiabatic transformation of chaotic nonlinear dynamical circuits," *2016 IEEE International Conference on Rebooting Computing (ICRC)*, San Diego, CA, 2016, pp. 1–8, 2016. doi:10.1109/ICRC.2016.7738679.
[905] M.P. Frank and E.P. DeBenedictis, "Chaotic Logic," presentation, *ICRC 2016*, San Diego, CA, Oct. 2016. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Frank_ICRC2016_ChaoticLogic_presUUR+notes.pdf.
[906] M.P. Frank, DYNAMIC, open-source software, Sandia National Laboratories, Center for Computing Research, 2017. https://cfwebprod.sandia.gov/cfdocs/CompResearch/templates/insert/softwre.cfm?sw=56.
[907] Wikipedia, "Chaos Computing," https://en.wikipedia.org/wiki/Chaos_computing, retrieved Dec. 6, 2019.
[908] M. Mitzenmacher and E. Upfal, Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis, 2nd ed., Cambridge University Press, 2017.
[909] Buluc, Aydin, et al. "Randomized algorithms for scientific computing (RASC)." *arXiv preprint* arXiv:2104.11079 (2021).
[910] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature reviews. Neuroscience,* vol. 9, p. 292, 2008.
[911] M. D. McDonnell and L. M. Ward, "The benefits of noise in neural systems: bridging theory and experiment," *Nature reviews. Neuroscience,* vol. 12, p. 415, 2011.
[912] W. Maass, "Noise as a resource for computation and learning in networks of spiking neurons," *Proceedings of the IEEE,* vol. 102, pp. 860–880, 2014.
[913] M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. N. Salama, "Memristors empower spiking neurons with stochasticity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 5, pp. 242-253, 2015.
[914] Suzuki, Hideyuki, et al. "Chaotic boltzmann machines." *Scientific reports* 3.1 (2013): 1-5.
[915] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, Jul. 1985.
[916] S. Kumar, J. P. Strachan and R. S. Williams, "Chaotic dynamics in nanoscale NbO2 Mott memristors for analogue computing," Nature, 548, 318 (2017)
[917] A. Parihar, N. Shukla, M. Jerry, S. Datta, and A. Raychowdhury, "Vertex coloring of graphs via phase dynamics of coupled oscillatory networks," *Scientific Reports*, vol. 7, no. 1, p. 911, Apr. 2017.
[918] H. Mostafa, L. K. Müller, and G. Indiveri, "An event-based architecture for solving constraint satisfaction problems," *Nature Communications*, vol. 6, p. 8941, Dec. 2015.
[919] Yamaoka, Masanao, et al. "A 20k-spin Ising chip to solve combinatorial optimization problems with CMOS annealing." *IEEE Journal of Solid-State Circuits* 51.1 (2015): 303-309.
[920] Aramon, Maliheh, et al. "Physics-inspired optimization for quadratic unconstrained problems using a digital annealer." *Frontiers in Physics* 7 (2019): 48.
[921] H. Goto, K. Tatsumura, and A. R. Dixon, "Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems," *Sci. Adv.*, vol. 5, no. 4, 2019 Apr 1, Art no. eaav2372
[922] Patel, Saavan, et al. "Ising model optimization problems on a FPGA accelerated restricted Boltzmann machine." *arXiv preprint arXiv:2008.04436* (2020).
[923] Yamamoto, Kasho, et al. "STATICA: A 512-Spin 0.25 M-Weight Annealing Processor With an All-Spin-Updates-at-Once Architecture for Combinatorial Optimization With Complete Spin–Spin Interactions." *IEEE Journal of Solid-State Circuits* 56.1 (2020): 165-178.
[924] Su, Yuqi, et al. "A 252 Spins Scalable CMOS Ising Chip Featuring Sparse and Reconfigurable Spin Interconnects for Combinatorial Optimization Problems." *2021 IEEE Custom Integrated Circuits Conference (CICC)*. IEEE, 2021.
[925] Lucas, Andrew. "Ising formulations of many NP problems." *Frontiers in physics* 2 (2014)
[926] Kaiser, Jan, and Supriyo Datta. "Probabilistic computing with p-bits." Applied Physics Letters 119.15 (2021): 150503.
[927] Joseph, Anosh. "Markov Chain Monte Carlo Methods in Quantum Field Theories: A Modern Primer." *arXiv preprint arXiv:1912.10997* (2019).

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[928] Burri, Samuel, et al. "SPADs for quantum random number generators and beyond." *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE, 2014.
[929] Vodenicarevic, Damir, et al. "Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing." *Physical Review Applied* 8.5 (2017): 054045.
[930] Valle, Javier del, et al. "Generation of Tunable Stochastic Sequences Using the Insulator–Metal Transition." *Nano Letters* (2022).
[931] Behin-Aein, Behtash, Vinh Diep, and Supriyo Datta. "A building block for hardware belief networks." *Scientific reports* 6 (2016): 29893.
[932] Camsari, Kerem Yunus, et al. "Stochastic p-bits for invertible logic." *Physical Review X* 7.3 (2017): 031014.
[933] Camsari, Kerem Yunus, Sayeef Salahuddin, and Supriyo Datta. "Implementing p-bits with embedded MTJ." IEEE Electron Device Letters 38.12 (2017): 1767-1770.
[934] Zand, Ramtin, et al. "Composable probabilistic inference networks using MRAM-based stochastic neurons." ACM Journal on Emerging Technologies in Computing Systems (JETC) 15.2 (2019): 17.
[935] Aadit, Navid Anjum, et al. "Massively Parallel Probabilistic Computing with Sparse Ising Machines." *arXiv preprint arXiv:2110.02481* (2021).
[936] Sutton, Brian, et al. "Intrinsic optimization using stochastic nanomagnets." *Scientific reports* 7 (2017): 44370.
[937] Faria, Rafatul, Kerem Y. Camsari, and Supriyo Datta. "Implementing Bayesian networks with embedded stochastic MRAM." AIP Advances 8.4 (2018): 045101
[938] Camsari, Kerem Y., Shuvro Chowdhury, and Supriyo Datta. "Scalable Emulation of Sign-Problem–Free Hamiltonians with Room-Temperature p-bits." Physical Review Applied 12.3 (2019): 034061.
[939] Kaiser, Jan, et al. "Probabilistic Circuits for Autonomous Learning: A simulation study." arXiv preprint arXiv:1910.06288 (2019)
[940] Borders, William A., et al. "Integer factorization using stochastic magnetic tunnel junctions." *Nature* 573.7774 (2019): 390-393
[941] Smithson, Sean C., et al. "Efficient CMOS invertible logic using stochastic computing." IEEE Transactions on Circuits and Systems I: Regular Papers 66.6 (2019): 2263-2274.
[942] Onizawa, Naoya, et al. "In-Hardware Training Chip Based on CMOS Invertible Logic for Machine Learning." IEEE Transactions on Circuits and Systems I: Regular Papers (2019).
[943] Sutton, Brian, et al. "Autonomous Probabilistic Coprocessing with Petaflips per Second." arXiv preprint arXiv:1907.09664 (2019).
[944]Camsari, Kerem Y., Brian M. Sutton, and Supriyo Datta. "p-bits for probabilistic spin logic." Applied Physics Reviews 6.1 (2019): 011305.
[945] M. P. Frank, "Introduction to reversible computing: motivation, progress, and challenges." In *Proceedings of the 2nd conference on Computing frontiers (CF '05)*. ACM, New York, NY, 385–390, 2005.
[946] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development*, vol. 5, no. 3, pp. 183-191, July 1961. doi:10.1147/rd.53.0183.
[947] N. G. Anderson, "Conditional erasure and the Landauer limit," in Energy Limits in Computation: A Review of Landauer's Principle, Theory, and Experiments (pp. 65–100), Lent, Orlov, Porod, and Snider, eds., Springer Nature, 2019.
[948] M. P. Frank and K. Shukla, "Quantum Foundations of Classical Reversible Computing," *Entropy* vol. 23, no. 6, 2021. doi:10.3390/e23060701.
[949] M. P. Frank, "Why reversible computing is the only way forward for general digital computing," invited talk, *IEEE International Nanodevices and Computing Conference*, Grenoble, France, Apr. 2019. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/INC19-talk-v5.pdf.
[950] C. H. Bennett, "Logical Reversibility of Computation," *IBM Journal of Research and Development*, vol. 17, no. 6, pp. 525–532, Nov. 1973. doi:10.1147/rd.176.0525.
[951] M. P. Frank and M.J. Ammer, "Relativized separation of reversible and irreversible space-time complexity classes," preprint, arxiv:1708.08480 [cs.ET], 2017.
[952] E. D. Demaine, J. Lynch, G. J. Mirano, and N. Tyagi, "Energy-efficient algorithms," in *Proc. 2016 ACM Conf. on Innovations in Theoretical Computer Science*, pp. 321–332, 2016.
[953] M. P. Frank, "Architectural, Algorithmic, and Systems Engineering Issues for Reversible Computing," presented at the CCC Workshop on Physics & Engineering Issues in Adiabatic/Reversible Classical Computing, Oct. 2020. Slides: https://cra.org/ccc/wp-content/uploads/sites/2/2020/10/CCC-Frank-Day3-v3.pdf, video: https://www.youtube.com/watch?v=-SzIZEEPPGY&feature=emb_logo.
[954] A. H. Earley, "On the performance and programming of reversible molecular computers," Ph.D. thesis, University of Cambridge, 2021. doi:10.17863/CAM.79111.
[955] S. Pidaparthi and C. Lent, "Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata," *Journal of Low Power Electronics and Applications* vol. 8, no. 3, 2018. doi:10.3390/jlpea8030030.
[956] S. S. Pidaparthi and C. S. Lent, "Energy dissipation during two-state switching for quantum-dot cellular automata," *Journal of Applied Physics* vol. 129, art. 024304, 2021. doi:10.1063/5.0033633.
[957] M. P. Frank, "On the interpretation of energy as the rate of quantum computation," *Quantum Information Processing*, vol. 4, no. 4, 2005. doi:10.1007/s11128-005-7851-5.
[958] M. P. Frank and T. F. Knight Jr., "Ultimate theoretical models of nanocomputers." *Nanotechnology* 9, no. 3, 1998. doi:10.1088/0957-4484/9/3/005.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[959] K. Likharev, "Dynamics of some single flux quantum devices: I. Parametric quantron," *IEEE Trans. Magn.* vol. 13, no. 242, 1977. doi:10.1109/TMAG.1977.1059351.
[960] M. Hosoya et al., "Quantum flux parametron: a single quantum flux device for Josephson supercomputer," *IEEE Transactions on Applied Superconductivity*, vol. 1, no. 2, pp. 77–89, June 1991. doi:10.1109/77.84613.
[961] K. E. Drexler, "Molecular machinery and manufacturing with applications to computation," Ph.D. thesis, Massachusetts Institute of Technology, 1991. http://hdl.handle.net/1721.1/27999.
[962] S. G. Younis and T. F. Knight, Jr., "Practical implementation of charge recovering asymptotically zero power CMOS," *Proceedings of the 1993 Symposium on Research on Integrated Systems*, pp. 234–250. MIT Press, 1993.
[963] C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, "Quantum cellular automata," *Nanotechnology* vol. 4, no. 1, p. 49, 1993. doi:10.1088/0957-4484/4/1/004.
[964] V. Anantharam, M. He, K. Natarajan, H. Xie, and M.P. Frank, "Driving Fully-Adiabatic Logic Circuits Using Custom High-Q MEMS Resonators," ESA/VLSI, pp. 5–11, 2004.
[965] N. Takeuchi, D. Ozawa, Y. Yamanashi, and N. Yoshikawa, "An adiabatic quantum flux parametron as an ultra-low-power logic device." *Superconductor Science and Technology* 26, no. 3, p. 035010, 2013.
[966] R. C. Merkle, R. A. Frietas, Jr., T. Hogg, T. E. Moore, M. S. Moses, J. Ryley, "Mechanical Computing Systems Using Only Links and Rotary Joints," preprint arXiv:1801.03534 [cs.ET], 2018.
[967] T. Hogg, M. S. Moses, and D. G. Allis, "Evaluating the friction of rotary joints in molecular machines," *Molecular Systems Design & Engineering*, vol. 2, no. 3, May 2017. doi:10.1039/C7ME00021A.
[968] E. Fredkin and T. Toffoli, "Conservative logic," *International Journal of Theoretical Physics*, vol. 21, nos. 3–4, pp. 219–253, Apr. 1982.
[969] A. Adamatzky, ed., *Collision-Based Computing*, Springer, 2002.
[970] K. Yamada, T. Asai, I.N. Motoike, and Y. Amemiya, "On digital VLSI circuits exploiting collision-based fusion gates." *From Utopian to Genuine Unconventional Computers*, vol. 1, 2008.
[971] T. Asai, K. Yamada, and Y. Amemiya, "Single-flux-quantum logic circuits exploiting collision-based fusion gates," Physica C, vol. 468, 2008.
[972] K. Yamada, T. Asai, and Y. Amemiya, "Combinational logic computing for single-flux quantum circuits with asynchronous collision-based fusion gates," in 23rd Int'l. Tech. Conf. on Circuits/Systems, Computers, and Communications (ITC-CSCC 2008), pp. 445–448, 2008.
[973] Q. P. Herr, J. E. Baumgardner, and A. Y. Herr, "Method and apparatus for ballistic single flux quantum logic," U.S. Patent No. 7,782,077, issued Aug. 2010.
[974] K. D. Osborn, "Reversible computation with flux solitons," *United States Patent 9812836 B1*, Nov. 2017.
[975] W. Wustmann and K. Osborn, "Reversible Fluxon Logic: Topological particles enable gates beyond the standard adiabatic limit," abstract, *Bulletin of the American Physical Society*, 2018. Presented at the APS March meeting, Los Angeles, March 5–9, 2018. Preprint: arXiv:1711.04339 [cond-mat.supr-con], 2017.
[976] W. Wustmann and K. D. Osborn, "Reversible Fluxon Logic: Topological particles allow gates beyond the standard adiabatic limit," preprint, arXiv:1711.04339 [cond-mat.supr-con], Nov. 2017.
[977] K. D. Osborn and W. Wustmann, "Ballistic Reversible Gates Matched to Bit Storage: Plans for an Efficient CNOT Gate Using Fluxons," *Reversible Computation*, 2018, pp. 189–204. doi:10.1007/978-3-319-99498-7_13. Preprint: arxiv:1806.08011 [cond-mat.mes-hall].
[978] K. D. Osborn and W. Wustmann, "Reversible Fluxon Logic for Future Computing," *IEEE International Superconductive Electronics Conference*, Riverside, CA, Jul.-Aug. 2019. Published in *IEEE Superconductivity News Forum (SNF)*, art. no. STP652, 2019, https://snf.ieeecsc.org/sites/ieeecsc.org/files/documents/snf/abstracts/STP652_Final_ISECExtendedAbstract.pdf.
[979] M. P. Frank, "Asynchronous Ballistic Reversible Computing." in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, pp. 172-179, 2017.
[980] M. P. Frank, R. M. Lewis, N. A. Missert, M. A. Wolak, and M. D. Henry, "Asynchronous Ballistic Reversible Fluxon Logic," *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1–7, 2019.
[981] M. P. Frank, R. M. Lewis, N. A. Missert, M. D. Henry, M. A. Wolak and E. P. DeBenedictis, "Semi-automated Design of Functional Elements for a New Approach to Digital Superconducting Electronics: Methodology and Preliminary Results," *IEEE International Superconductive Electronics Conference*, Riverside, CA, Jul.-Aug. 2019. Preprint: https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/ISEC2019_Frank-etal_final_v7.pdf.
[982] M. P. Frank, R. M. Lewis and K. Shukla, "Implementing the Asynchronous Reversible Computing Paradigm in Josephson Junction Circuits," *21st Biennial U.S. Workshop on Superconductor Electronics, Devices, Circuits, and Systems*, Skytop, PA, Oct. 2019. Slides: https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/JJ-workshop-v3.pdf.
[983] K. Osborn and W. Wustmann, "Logic gates with flux solitons," U.S. patent #1077829B1, Sep. 2020.
[984] K. D. Osborn and W. Wustmann, "Reversible Fluxon Logic with optimized CNOT gate components," arxiv:2010.15991v1 [quant-ph], Oct. 2020.
[985] M. P. Frank, "Reversible Computing as a Path Forward for Improving Dissipation-Delay Efficiency in Superconducting Computing," presented at the Applied Superconductivity Conference, Nov. 2020, https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/ASC20-talk-final+SAND.pdf.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[986] M. P. Frank, "Feasible demonstration of ultra-low-power adiabatic CMOS for cubesat applications using LC ladder resonators," presentation, *Tenth Workshop on Fault-Tolerant Spaceborne Computing Employing New Technologies*, Albuquerque, NM, May 2017. Slides: https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Space-computing-final12-2up.pdf.

[987] M. P. Frank, R. W. Brocato, T. M. Conte, A. H. Hsia, A. Jain, N. A. Missert, K. Shukla and B. D. Tierney, "Special Session: Exploring the Ultimate Limits of Adiabatic Circuits," presented at the 38th IEEE International Conference on Computer Design (ICCD 2020), Oct. 2020. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/ICCD-CRC_1v8.pdf, https://youtu.be/URp6MqApoXk.

[988] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Simulation of sub-k_BT bit-energy operation of adiabatic quantum-flux-parametron logic with low bit-error-rate," *Appl. Phys. Lett.*, vol. 103, p. 62602, 2013. doi:10.1063/1.4817974.

[989] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Reversible logic gate using adiabatic superconducting devices," *Sci. Rep.*, vol. 4, p. 6354, 2014. doi:10.1038/srep06354.

[990] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Reversibility and energy dissipation in adiabatic superconductor logic," *Sci. Rep.*, vol. 7, no. 1, p. 75, Mar. 2017. doi:10.1038/s41598-017-00089-9.

[991] N. Takeuchi and N. Yoshikawa, "Minimum energy dissipation required for a logically irreversible operation," *Phys. Rev. E*, vol. 97, no. 1, p. 012124, Jan. 2018. doi:10.1103/PhysRevE.97.012124.

[992] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Recent Progress on Reversible Quantum-Flux-Parametron for Superconductor Reversible Computing," *IEICE Transactions on Electronics* vol. 101, no. 5, May 2018.

[993] T. Yamae, N. Takeuchi, N. Yoshikawa, "A reversible full adder using adiabatic superconductor logic," *Supercond. Sci. Technol.*, vol. 32, art. 035005, 2019.

[994] J. Ren and V. K. Semenov, "Progress with physically and logically reversible superconducting digital circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 780–786, Jun. 2011. doi:10.1109/TASC.2011.2104352.

[995] J. Ren, "Physically and Logically Reversible Superconducting Circuit," Ph.D. dissertation, Stony Brook University, May 2011. https://dspace.sunyconnect.suny.edu/handle/1951/56102.

[996] M. Lucci *et al.*, "Low-power digital gates in ERSFQ and nSQUID technology," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 3, pp. 1–5, Apr. 2016. doi:10.1109/TASC.2016.2535146.

[997] M. P. Frank, "Approaching the physical limits of computing," *35th International Symposium on Multiple-Valued Logic (ISMVL'05)*, 2005, pp. 168-185.

[998] R. Karakiewicz, R. Genov, and G. Cauwenberghs, "1.1 TMACS/mW Fine-Grained Stochastic Resonant Charge-Recycling Array Processor," *IEEE Sensors Journal*, vol. 12, no. 4, pp. 785–792, April 2012.

[999] S. G. Younis and T. F. Knight, Jr., "Asymptotically zero energy split-level charge recovery logic," *International Workshop on Low Power Design*, pp. 177–182, Apr. 1994.

[1000] S. G. Younis, "Asymptotically zero energy computing using split-level charge recovery logic," Ph.D. dissertation, MIT Dept. of EECS, 1994. https://dspace.mit.edu/handle/1721.1/11620

[1001] C. J. Vieri, "Pendulum—A reversible computer architecture," M.S. thesis, MIT Dept. of EECS, 1995

[1002] M. P. Frank, "Reversibility for Efficient Computing," Ph.D. dissertation, MIT Dept. of EECS, 1999. https://dspace.mit.edu/handle/1721.1/9464.

[1003] C. J. Vieri, "Reversible computer engineering and architecture," Ph.D. dissertation, MIT Dept. of EECS, 1999. https://dspace.mit.edu/handle/1721.1/80144.

[1004] M. P. Frank, "Reversible Computing and Truly Adiabatic Circuits: The Next Great Challenge for Digital Engineering," *5th IEEE Dallas Circuits and Systems Workshop on Design, Applications, Integration and Software (DCAS-06)*, Oct. 2006, Dallas, TX. doi:10.1109/DCAS.2006.321027.

[1005] A. Vetuli, S. Di Pascoli, and L. M. Reyneri, "Positive feedback in adiabatic logic," *Electronics Letters* vo. 32, no. 20, 1996.

[1006] P. Teichman, "Adiabatic Logic: Future Trend and System Level Perspective," Springer Science & Business Media, vol. 24, New York: Springer, 2012.

[1007] Y. Moon and D.-K. Jeong, "Efficient charge recovery logic," Digest of Technical Papers, Symposium on VLSI Circuits, IEEE, 1995.

[1008] S. Maheshwari, V. A. Bartlett and I. Kale, "Adiabatic flip-flops and sequential circuit design using novel resettable adiabatic buffers," *2017 European Conference on Circuit Theory and Design (ECCTD)*, Catania, pp. 1–4, 2017.

[1009] M. Avital, H. Dagan, I. Levi, O. Keren and A. Fish, "DPA-Secured Quasi-Adiabatic Logic (SQAL) for Low-Power Passive RFID Tags Employing S-Boxes," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 1, pp. 149–156, Jan. 2015.

[1010] S. D. Kumar, H. Thapliyal and A. Mohammad, "EE-SPFAL: A Novel Energy-Efficient Secure Positive Feedback Adiabatic Logic for DPA Resistant RFID and Smart Card," IEEE Transactions on Emerging Topics in Computing, vol. 7, no. 2, pp. 281–293, Apr.-Jun. 2019.

[1011] I. K. Hanninen, C. O. Campos-Aguillon, R. Celis-Cordova, and G. L. Snider. "Design and Fabrication of a Microprocessor using Adiabatic CMOS and Bennett Clocking" *Lecture Notes in Computer Science*, *Reversible Computation: 7th International Conference*, Springer, Cham, 2015.

[1012] R. Celis-Cordova, A. O. Orlov, T. Lu, J.M. Kulick, and G. L. Snider. "The design of 16-bit Adiabatic Microprocessor". *2019 IEEE International Conference on Rebooting Computing (ICRC)*, San Mateo, CA, 2019.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1013] M. P. Frank, R. W. Brocato, B. D. Tierney, N. A. Missert, and A. H. Hsia, "Reversible Computing with Fast, Fully Static, Fully Adiabatic CMOS," arxiv:2009.00448v2 [cs.AR], Sep. 2020.
[1014] E. P. DeBenedictis, "Inversion for S2LAL," Zettaflops LLC tech. report ZF004, Sep. 2020, https://zettaflops.org/zf004/.
[1015] K. Loe, E. Goto, "Analysis of flux input and output Josephson pair device," *IEEE Trans. Magn.*, vol. 21, pp. 884–887, 1985.
[1016] V. K. Semenov, G. V. Danilov, D. V. Averin, "Negative-inductance SQUID as the basic element of reversible Josephson-junction circuits," *IEEE Trans. Appl. Supercond.*, vol. 13, pp. 938–943, 2003.
[1017] N. Takeuchi, Y. Yamanashi, N. Yoshikawa, "Energy efficiency of adiabatic superconductor logic," *Supercond. Sci. Technol.*, vol. 28, art. 015003, 2015.
[1018] N. Takeuchi, Y. Yamanashi, N. Yoshikawa, "Thermodynamic Study of Energy Dissipation in Adiabatic Superconductor Logic," *Phys. Rev. Applied*, vol. 4, art. 034007, 2015.
[1019] N. Takeuchi, Y. Yamanashi and N. Yoshikawa, "Measurement of 10 zJ energy dissipation of adiabatic quantum-flux-parametron logic using a superconducting resonator," *Appl. Phys. Lett.*, vol. 102, art. 052602, 2013.
[1020] N. Takeuchi, T. Yamae, C. L. Ayala, H. Suzuki, N. Yoshikawa, "An adiabatic superconductor 8-bit adder with 24 k_BT energy dissipation per junction," *Appl. Phys. Lett.*, vol. 114, art. 042602, 2019.
[1021] N. Takeuchi, T. Yamae, and N. Yoshikawa, "A reversible 8-bit adder using adiabatic quantum-flux-parametron logic," IEEE Applied Superconductivity Conference, 2020.
[1022] T. Yamae, N. Takeuchi, and N. Yoshikawa, "Design and demonstration of an 8 word by 1 bit reversible quantum flux parametron register file," IEEE Applied Superconductivity Conference, 2020.
[1023] K. E. Drexler, "Molecular engineering: An approach to the development of general capabilities for molecular manipulation." *Proceedings of the National Academy of Sciences*, vol. 78, no. 9, pp. 5275–5278, 1981.
[1024] K. E. Drexler, "Rod logic and thermal noise in the mechanical nanocomputer," in *Molecular Electronic Devices*, F. L. Carter, R. Siatkowski and H. Wohltjen, eds. Amsterdam: North-Holland, 1988.
[1025] K. E. Drexler, Nanosystems: molecular machinery, manufacturing, and computation, New York: Wiley, 1992.
[1026] K. K. Likharev and A. N. Korotkov, "Single-electron parametron: reversible computation in a discrete-state system," *Science*, vol. 273, no. 5276, pp.763–765, 1996.
[1027] E. Blair, "Electric-Field Inputs for Molecular Quantum-Dot Cellular Automata Circuits," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 453–460, 2019.
[1028] K. Walus, M. Mazur, G. Schulhof, and G.A. Jullien. "Simple 4-bit processor based on quantum-dot cellular automata (QCA)." *2005 IEEE International Conference on Application-Specific Systems, Architecture Processors (ASAP '05)*, pp. 288–293, IEEE, July 2005.
[1029] E. Fazzion, O.L. Fonseca, J. A. M. Nacif, O. P. V. Neto, A. O. Fernandes, and D. S. Silva, "A quantum-dot cellular automata processor design," *Proceedings of the 27th Symposium on Integrated Circuits and Systems Design*, p. 29, ACM, Sep. 2014.
[1030] G. Pillonnet, H. Fanet, and S. Houri, "Adiabatic Capacitive Logic: a paradigm for low-power logic," in IEEE International Symposium on Circuits and Systems (ISCAS). IEEE: Baltimore, MD, 2017.
[1031] R. Celis-Cordova, A. O. Orlov and G. L. Snider, "Adiabatic Capacitive Logic using Voltage-controlled Variable Capacitors," 2020 IEEE Silicon Nanoelectronics Workshop (SNW), pp. 115–116, 2020.
[1032] Y. Perrin, A. Galisultanov, L. Hutin, P. Basset, H. Fanet and G. Pillonnet, "Contact-Free MEMS Devices for Reliable and Low-Power Logic Operations," in *IEEE Transactions on Electron Devices* vol. 68, no. 6, pp. 2938-2943, June 2021.
[1033] S. V. Polonsky, V. K. Semenov, D. F. Schneider, "Transmission of single-flux-quantum pulses along superconducting microstrip lines," *IEEE Trans. Appl. Supercond.*, vol. 3, no. 1, pp. 2598–2600, Mar. 1993.
[1034] R. D. Parmentier, K. Lonngren, A. C. Scott, "Fluxons in long Josephson junctions" in *Solitons in Action*, Cambridge, MA, USA:Acad. Press, pp. 173–196, 1978.
[1035] M. P. Frank, E. P. DeBenedictis, "Ballistic Reversible Superconducting Memory Element," U.S. Patent Application, Pub. No. US 2022/0036943 A1. United States Patent and Trademark Office, Feb. 2022.
[1036] M. P. Frank, "Fundamental Physical Limits of Reversible Computing—An Introduction," presented at the CCC Workshop on Physics & Engineering Issues in Adiabatic/Reversible Classical Computing, Oct. 2020. Slides: https://cra.org/ccc/wp-content/uploads/sites/2/2020/10/CCC-Frank-Day1-v2.pdf, video: https://youtu.be/Oj_cRr828-c, https://youtu.be/PEqiOSLzFRE.
[1037] K. Shukla, "Foundations of the Lindbladian Approach to Adiabatic and Reversible Computing," presented at the CCC Workshop on Physics & Engineering Issues in Adiabatic/Reversible Classical Computing, Oct. 2020. Slides: https://cra.org/ccc/wp-content/uploads/sites/2/2020/12/Shukla-Foundations-of-the-Lindbladian-Approach-to-Adiabatic-and-Reversible-Computing-2.pdf, video: https://youtu.be/mzoL6m-2rrA.
[1038] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, "Shortcuts to adiabaticity: concepts, methods, and applications," *Reviews of Modern Physics*, vol. 91, no. 4, art. 045001, 2019.
[1039] W. M. Itano, D. J. Heinzen, J. J. Bollinger, and D. J. Wineland, "Quantum zeno effect," *Physical Review A*, vol. 41, no. 5, p. 2295, 1990.
[1040] P. Facchi, H. Nakazato, and S. Pascazio. "From the quantum Zeno to the inverse quantum Zeno effect," *Physical Review Letters*, vol. 86, no. 13, p. 2699, 2001.
[1041] F. Gao, G. Tziantzioulis, and D. Wentzlaff, "ComputeDRAM: In-Memory Compute Using Off-the-Shelf DRAMs," in *MICRO*, 2019.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1042] N. Hajinazar, G. F. Oliveira, S. Gregorio, J. D. Ferreira, N. M. Ghiasi, M. Patel, M. Alser, S. Ghose, J. Gomez-Luna, and O. Mutlu, "SIMDRAM: A Framework for Bit-Serial SIMD Processing Using DRAM," in *ASPLOS*, 2021.
[1043] S. Li, D. Niu, K. T. Malladi, H. Zheng, B. Brennan, and Y. Xie, "DRISA: A DRAM-Based Reconfigurable In-Situ Accelerator," in *MICRO*, 2017.
[1044] V. Seshadri, D. Lee, T. Mullins, H. Hassan, A. Boroumand, J. Kim, M. A. Kozuch, O. Mutlu, P. B. Gibbons, and T. C. Mowry, "Ambit: In-Memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology," in *MICRO*, 2017.
[1045] S. Aga, S. Jeloka, A. Subramaniyan, S. Narayanasamy, D. Blaauw, and R. Das, "Compute Caches," in *HPCA*, 2017.
[1046] D. Fujiki, S. Mahlke, and R. Das, "Duality Cache for Data Parallel Acceleration," in *ISCA*, 2019.
[1047] S. Jeloka, N. B. Akesh, D. Sylvester, and D. Blaauw, "A 28 nm Configurable Memory (TCAM/BCAM/SRAM) Using Push-Rule 6T Bit Cell Enabling Logic-in-Memory," in *JSSC*, 2016.
[1048] M. Kang, E. P. Kim, M.-S. Keel, and N. R. Shanbhag, "Energy-Efficient and High Throughput Sparse Distributed Memory Architecture," in *ISCAS*, 2015.
[1049] J. Wang, X. Wang, C. Eckert, A. Subramaniyan, R. Das, D. T. Blaauw, and D. Sylvester: "A Compute SRAM with Bit-Serial Integer/Floating-Point Operations for Programmable In-Memory Vector Acceleration," in *ISSCC*, 2019.
[1050] V. Seshadri, D. Lee, T. Mullins, H. Hassan, A. Boroumand, J. Kim, M. A. Kozuch, O. Mutlu, P. B. Gibbons, and T. C. Mowry, "Ambit: In-Memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology," in *MICRO*, 2017.
[1051] V. Seshadri, D. Lee, T. Mullins, H. Hassan, A. Boroumand, J. Kim, M. A. Kozuch, O. Mutlu, P. B. Gibbons, and T. C. Mowry, "Ambit: In-Memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology," in *MICRO*, 2017.
[1052] S. Jeloka, N. B. Akesh, D. Sylvester, and D. Blaauw, "A 28 nm Configurable Memory (TCAM/BCAM/SRAM) Using Push-Rule 6T Bit Cell Enabling Logic-in-Memory," in *JSSC*, 2016.
[1053] S. Gupta, M. Imani, and T. Rosing, "FELIX: Fast and Energy-Efficient Logic in Memory," in *ICCAD*, 2018.
[1054] S. Hamdioui, L. Xie, H. A. D. Nguyen, M. Taouil, K. Bertels, H. Corporaal, H. Jiao, F. Catthoor, D. Wouters, L. Eike, and J. van Lunteren, "Memristor-Based Computation-in-Memory Architecture for Data-Intensive Applications," in DATE, 2015.
[1055] S. Kvantinsky, D. Belousov, S. Liman, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "MAGIC: Memristor-Aided Logic," TCAS II, Sep. 2014.
[1056] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, "Pinatubo: A Processing-in-Memory Architecture for Bulk Bitwise Operations in Emerging Non-Volatile Memories," in DAC, 2016.
[1057] S. Angizi, Z. He, A. Awad, and D. Fan, "MRIMA: An MRAM-Based In-Memory Accelerator," TCAD, May 2020.
[1058] J. Louis, B. Hoffer, and S. Kvatinsky, "Performing Memristor-Aided Logic (MAGIC) Using STT-MRAM," in ICECS, 2019.
[1059] S. Angizi, Z. He, and D. Fan, "PIMA-Logic: A Novel Processing-in-Memory Architecture for Highly Flexible and Energy-Efficient Logic Computation," in *DAC*, 2018.
[1060] M. S. Q. Truong, E. Chen, D. Su, A. Glass, L. Shen, L. R. Carley, J. A. Bain, and S. Ghose, "RACER: Bit-Pipelined Processing Using Resistive Memory," in *MICRO*, 2021.
[1061] S. Ghose, A. Boroumand, J. S. Kim, J. Go´mez-Luna, and O. Mutlu, "Processing-in-Memory: A Workload-Driven Perspective," *IBM JRD*, Nov. 2019.
[1062] M. He, C. Song, I. Kim, C. Jeong, S. Kim, I. Park, M. Thottethodi, and T. N. Vijaykumar, "Newton: A DRAM-Maker's Accelerator-in-Memory (AiM) Architecture for Machine Learning," in *MICRO*, 2020.
[1063] Y. C. Kwon, et all, "A 20nm 6GB Function-In-Memory DRAM, Based on HBM2 with a 1.2TFLOPS Programmable Computing Unit Using Bank-Level Parallelism, for Machine Learning Applications," in *ISSCC*, 2021.
[1064] S. Lee, S.-h. Kang, J. Lee, H. Kim, E. Lee, S. Seo, H. Yoon, S. Lee, K. Lim, H. Shin, J. Kim, S. O, A. Iyer, D. Wang, K. Sohn, and N. S. Kim, "Hardware Architecture and Software Stack for PIM Based on Commercial DRAM Technology," in *ISCA*, 2021.
[1065] H. Shin, D. Kim, S. P. Eunhyeok Park, Y. Park, and S. Yoo, "McDRAM: Low Latency and Energy-Efficient Matrix Computations in DRAM," *TCAD*, Nov. 2018.
[1066] A. Yazdanbakhsh, C. Song, J. Sacks, P. Lotfi-Kamran, H. Esmaeilzadeh, and N. S. Kim, "In-DRAM Near-Data Approximate Acceleration for GPUs," in *PACT*, 2018.
[1067] Y. Eckert, N. Jayasena, and G. H. Loh, "Thermal Feasibility of Die-Stacked Processing in Memory," in WoNDP, 2014.
[1068] S. H. Pugsley, J. Jestes, H. Zhang, R. Balasubramonian, V. Srinivasan, A. Buyuktosunoglu, A. Davis, and F. Li, "NDC: Analyzing the Impact of 3D-Stacked Memory+Logic Devices on MapReduce Workloads," in ISPASS, 2014.
[1069] D. Zhang, N. Jayasena, A. Lyashevsky, J. L. Greathouse, L. Xu, and M. Ignatowski, "TOP-PIM: Throughput-Oriented Programmable Processing in Memory," in HPDC, 2014.
[1070] JEDEC Solid State Technology Assn., JESD235C: High Bandwidth Memory (HBM) DRAM, Jan. 2020.
[1071] Hybrid Memory Cube Consortium, Hybrid Memory Cube Specification Rev. 2.0, 2013.
[1072] Q. Zhu, B. Akin, H. E. Sumbul, J. C. Hoe, L. Pileggi, and F. Franchetti, "A 3D-Stacked Logic-in-Memory Accelerator for Application-Specific Data Intensive Computing," in *3DIC*, 2013.
[1073] [Y. Zhu, B. Wang, D. Li, and J. Zhao, "Integrated Thermal Analysis for Processing in Die-Stacking Memory," in *MEMSYS*, 2016.
[1074] S. Aga, S. Jeloka, A. Subramaniyan, S. Narayanasamy, D. Blaauw, and R. Das, "Compute Caches," *HPCA*, 2017.
[1075] F. Devaux, "The True Processing in Memory Accelerator," in *HotChips*, 2019.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1076] UPMEM SAS, *Technology*, https://www.upmem.com/technology/.
[1077] Samsung Electronics Co., Ltd., HBM Processing in Memory, https://www.samsung.com/semiconductor/ solutions/technology/hbm-processing-in-memory/.
[1078] P. Crotty, D. Schult and K. Segall, "Josephson junction simulation of neurons." *Phys. Rev. E*, vol. **82**, no. 8, doi:10.1103/PhysRevE.82.011914, 2010.
[1079] M. L. Schneider, *et al.*, "Ultralow power artificial synapses using nanotextured magnetic Josephson junctions," *Science Advances* vol. 4, e1701329, doi:10.1126/sciadv.1701329, 2018.
[1080] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive Random Access Memory," *Proceedings of the IEEE* vol. 104, pp. 1796–1830, doi:10.1109/jproc.2016.2590142, 2016.
[1081] S. K. Tolpygo, "Superconductor digital electronics: Scalability and energy efficiency issues (Review Article)." *Low Temperature Physics* vol. 42, 361–379, doi:10.1063/1.4948618, 2016.
[1082] Hamerly, Ryan, et al. "Large-scale optical neural networks based on photoelectric multiplication." *Physical Review X* 9.2 (2019): 021032.
[1083] Tait, Alexander N., et al. "Neuromorphic photonic networks using silicon photonic weight banks." *Scientific reports* 7.1 (2017): 7430
[1084] Shen, Yichen, et al. "Deep learning with coherent nanophotonic circuits." *Nature Photonics* 11.7 (2017): 441.
[1085] Yang, Lin, et al. "On-chip CMOS-compatible optical signal processor." *Optics express* 20.12 (2012): 13560-13565.
[1086] Tait, Alexander N., et al. "Microring weight banks." *IEEE Journal of Selected Topics in Quantum Electronics* 22.6 (2016): 312-325.
[1087] Jouppi, Norman P., et al. "In-datacenter performance analysis of a tensor processing unit." *2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 2017
[1088] Papamichael, Jeremy Fowers Kalin Ovtcharov Michael, et al. "A Configurable Cloud-Scale DNN Processor for Real-Time AI." (2018).
[1089] Nicol, Chris. "A coarse grain reconfigurable array (cgra) for statically scheduled data flow computing." *Wave Computing White Paper* (2017).
[1090] Schrauwen, Jonathan, Dries Van Thourhout, and Roel Baets. "Trimming of silicon ring resonator by electron beam induced compaction and strain." *Optics express* 16.6 (2008): 3738-3743.
[1091] Atabaki, Amir H., et al. "Accurate post-fabrication trimming of ultra-compact resonators on silicon." *Optics express* 21.12 (2013): 14139-14145.
[1092] Atabaki, Amir H., et al. "Accurate post-fabrication trimming of ultra-compact resonators on silicon." *Optics express* 21.12 (2013): 14139-14145.
[1093] Padmaraju, Kishore, and Keren Bergman. "Resolving the thermal challenges for silicon microring resonator devices." *Nanophotonics* 3.4-5 (2014): 269-281.
[1094] Krishnamoorthy, Ashok V., et al. "Exploiting CMOS manufacturing to reduce tuning requirements for resonant optical devices." *IEEE Photonics Journal* 3.3 (2011): 567-579.
[1095] Su, Zhan, et al. "Reduced wafer-scale frequency variation in adiabatic microring resonators." *OFC 2014*. IEEE, 2014.
[1096] Mekis, Attila, et al. "A grating-coupler-enabled CMOS photonics platform." *IEEE Journal of Selected Topics in Quantum Electronics* 17.3 (2010): 597-608.
[1097] Bogaerts, Wim, et al. "Silicon-on-insulator spectral filters fabricated with CMOS technology." *IEEE journal of selected topics in quantum electronics* 16.1 (2010): 33-44.
[1098] Assefa, Solomon, Fengnian Xia, and Yurii A. Vlasov. "Reinventing germanium avalanche photodetector for nanophotonic on-chip optical interconnects." *Nature* 464.7285 (2010): 80.
[1099] Nahmias, Mitchell A., et al. "Photonic multiply-accumulate operations for neural networks." *IEEE Journal of Selected Topics in Quantum Electronics* 26.1 (2019): 1-18.
[1100] Glick, Madeleine, Lionel C. Kimmerling, and Robert C. Pfahl. "A roadmap for integrated photonics." *Optics and Photonics News* 29.3 (2018): 36-41.
[1101] I. Ahmed, P.-W. Chiu, W. Moy, and C. H. Kim, "A Probabilistic Compute Fabric Based on Coupled Ring Oscillators for Solving Combinatorial Optimization Problems," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 9, 2021.
[1102] C. T. Van Degrift, "Tunnel diode oscillator for 0.001 ppm measurements at low temperatures," *Review of Scientific Instruments,* vol. 46, 599, 1975.
[1103] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements – Part I. general energy relations," *Proceedings of the Institute of Radio Engineers,* vol. 44, no. 7, pp. 904-913, 1956.
[1104] A. M. Niknejad, "Inductors and Transformers," *IEEE Solid-State Circuits Magazine*¸2014.
[1105] L. Iotti, S. Krishnamurthy, G. LaCaille, A. M. Niknejad, "A Low-Power 70-100 GHz Mixer-First RX Leveraging Frequency-Translational Feedback," *IEEE Journal of Solid-State Circuits,* vol. 55, no. 8, pp. 2043-2054, 2020.
[1106] K. Yogendra, D. Fan, and K. Roy, "Coupled Spin Torque Nano Oscillators for Low Power Neural Computation," *IEEE Transactions on Magnetics*, vol. 51, no. 10, pp. 1–9, Oct. 2015.
[1107] A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim, and K. Roy, "Magnetic Tunnel Junction Mimics Stochastic Cortical Spiking Neurons," *Scientific Reports*, vol. 6, p. srep30039, Jul. 2016.
[1108] S. Kaka, M. R. Pufall, W. H. Rippard, T. J. Silva, S. E. Russek, and J. A. Katine, "Mutual phase-locking of microwave spin torque nano-oscillators," *Nature*, vol. 437, no. 7057, pp. 389–392, Sep. 2005.

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1109] Y. Li, X. de Milly, F. Abreu Araujo, O. Klein, V. Cros, J. Grollier, and G. de Loubens, "Probing Phase Coupling Between Two Spin-Torque Nano-Oscillators with an External Source," *Physical Review Letters*, vol. 118, no. 24, p. 247202, Jun. 2017.

[1110] R. Lebrun, S. Tsunegi, P. Bortolotti, H. Kubota, A. S. Jenkins, M. Romera, K. Yakushiji, A. Fukushima, J. Grollier, S. Yuasa, and V. Cros, "Mutual synchronization of spin torque nano-oscillators through a long-range and tunable electrical coupling scheme," *Nature Communications*, vol. 8, p. ncomms15825, Jun. 2017.

[1111] M. R. Pufall, W. H. Rippard, G. Csaba, D. E. Nikonov, G. I. Bourianoff, and W. Porod, "Physical Implementation of Coherently Coupled Oscillator Networks," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 1, pp. 76–84, Dec. 2015.

[1112] B. C. McGoldrick, J. Z. Sun, and L. Liu, "Ising Machine Based on Electrically Coupled Spin Hall Nano-Oscillators," *Physical Review Applied*, vol. 17, no. 1, pp. 014006, 2022.

[1113] W. A. Vitale, C. F. Moldovan, A. Paone, A. Schüler, and A. M. Ionescu, "Fabrication of CMOS-compatible abrupt electronic switches based on vanadium dioxide," *Microelectronic Engineering*, vol. 145, pp. 117–119, Sep. 2015.

[1114] H.-T. Kim, B.-G. Chae, D.-H. Youn, S.-L. Maeng, G. Kim, K.-Y. Kang, and Y.-S. Lim, "Mechanism and observation of Mott transition in VO_2-based two- and three-terminal devices," *New Journal of Physics*, vol. 6, no. 1, p. 52, 2004.

[1115] N. Shukla, A. Parihar, E. Freeman, H. Paik, G. Stone, V. Narayanan, H. Wen, Z. Cai, V. Gopalan, R. Engel-Herbert, D. G. Schlom, A. Raychowdhury, and S. Datta, "Synchronized charge oscillations in correlated electron systems," *Scientific Reports*, vol. 4, p. 4964, May 2014.

[1116] S. Kumar, J. P. Strachan, and R. S. Williams, "Chaotic dynamics in nanoscale NbO2 Mott memristors for analogue computing," *Nature,* vol. 548, p. 318, 2017.

[1117] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," *Nature Materials,* vol. 12, pp. 114-7, Feb 2013.

[1118] P.-Y. Chen, J.-s. Seo, Y. Cao, and S. Yu, "Compact oscillation neuron exploiting metal-insulator-transition for neuromorphic computing," presented at the *Proceedings of the 35th International Conference on Computer-Aided Design*, Austin, Texas, 2016.

[1119] S. Dutta, et al., "An Ising Hamiltonian solver based on coupled stochastic phase-transition nano-oscillators," *Nature Electronics*, vol. 4, pp. 502-512, 2021.

[1120] A. Raychowdhury, et al., "Computing with Networks of Oscillatory Dynamical Systems," *Proceedings of the IEEE,* vol. 107, no. 1, pp. 73-89, 2019.

[1121] S. Dutta, et al., "Programmable coupled oscillators for synchronized locomotion," *Nature Communications*, vol. 10, 3299, 2019.

[1122] S. Dutta, A. Khanna, J. Gomez, K. Ni, Z. Toroczkai, and S. Datta, "Experimental Demonstration of Phase Transition Nano-Oscillator Based Ising Machine," *IEEE International Electron Devices Meeting (IEDM)*, 2019.

[1123] C. Li, E. Wasige, L. Wang, J. Wang, and B. Romeira, "28 GHz MMIC resonant tunnelling diode oscillator of around 1 mW output power," *Electronics Letters*, vol. 49, no. 13, pp. 816–818, Jun. 2013.

[1124] M. Feiginov, C. Sydlo, O. Cojocari, and P. Meissner, "Resonant-tunnelling-diode oscillators operating at frequencies above 1.1 THz," *Applied Physics Letters*, vol. 99, no. 23, p. 233506, 2011.

[1125] S. Suzuki, M. Asada, A. Teranishi, H. Sugiyama, and H. Yokoyama, "Fundamental oscillation of resonant tunneling diodes above 1 THz at room temperature," *Applied Physics Letters*, vol. 97, no. 24, p. 242102, 2010.

[1126] H. Eisele, "State of the art and future of electronic sources at terahertz frequencies," *Electronics Letters*, vol. 46, no. 26, p. S8, 2010.

[1127] T. Silva and W. Rippard, "Developments in nano-oscillators based upon spin-transfer point-contact devices," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1260–1271, Apr. 2008.

[1128] C. L. Tang, W. R. Bosenberg, T. Ukachi, R. J. Lane, and L. K. Cheng, "Optical parametric oscillators," *Proceedings of the IEEE*, vol. 80, no. 3, pp. 365–374, 1992.

[1129] L. Maleki, "Sources: The optoelectronic oscillator," *Nature Photonics*, vol. 5, no. 12, pp. 728–730, 2011.

[1130] I. Mahboob and H. Yamaguchi, "Piezoelectrically pumped parametric amplification and Q enhancement in an electromechanical oscillator," *Applied Physics Letters,* vol. 92, 173109, 2008.

[1131] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Applied Physics Letters* 93, 082508 (2008).

[1132] W. Butler, X.-G. Zhang, T. Schulthess, and J. MacLaren, "Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches," *Physical Review B* 63, 054416 (2001).

[1133] Safranski, Christopher, et al. "Demonstration of nanosecond operation in stochastic magnetic tunnel junctions." *Nano Letters* 21.5 (2021): 2040-2045.

[1134] Hayakawa, Keisuke, et al. "Nanosecond random telegraph noise in in-plane magnetic tunnel junctions." *Physical Review Letters* 126.11 (2021): 117202.

[1135] R. K. Henderson, E. A. G. Webster, and R. Walker, "A Gate Modulated Avalanche Bipolar Transistor in 130nm CMOS Technology," *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, pp. 226-229, IEEE (2012)

[1136] E. A.G. Webster, J. A. Richardson, L. A. Grant, R. K. Henderson, "A single electron bipolar avalanche transistor implemented in 90 nm CMOS" *Solid-State Electronics*, *76*, pp.116-118 (2012)

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1137] J. Rajendran et al., "Nano meets security: Exploring nanoelectronic devices for security applications," *Proceedings of the IEEE,* vol. 103, no. 5, pp. 829–849, 2015
[1138] M. L. Gallo, T. Tuma, F. Zipoli, A. Sebastian, and E. Eleftheriou, "Inherent stochasticity in phase-change memory devices," *2016 46th European Solid-State Device Research Conference (ESSDERC)*, 2016, pp. 373–376.
[1139] H. Jiang, D. Belkin, S. E. Savel'ev, S. Lin, Z. Wang, Y. Li*, et al.*, "A novel true random number generator based on a stochastic diffusive memristor," *Nature Communications,* vol. 8, 2017.
[1140] Z. Wang, S. Joshi, S. E. Saveliev, H. Jiang, R. Midya, P. Lin*, et al.*, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature Materials,* vol. 16, pp. 101–108, 09/26/online 2016.
[1141] Y. H. Tseng, C.-E. Huang, C.-H. Kuo, Y.-D. Chih, and C. J. Lin, "High density and ultra-small cell size of contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits," *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1–4: IEEE
[1142] C.-Y. Huang, W. C. Shen, Y.-H. Tseng, Y.-C. King, and C.-J. Lin, "A contact-resistive random-access-memory-based true random number generator," *IEEE Electron Device Letters,* vol. 33, no. 8, pp. 1108–1110, 2012
[1143] M. Park, J. C. Rodgers, and D. P. Lathrop, "True random number generation using CMOS Boolean chaotic oscillator," *Microelectronics Journal* 46, 1364 (2015).
[1144] Russek, S. E. et al. in 2016 IEEE International Conference on Rebooting Computing (ICRC). 1-5.
[1145] Schneider, M. L. *et al.* Ultralow power artificial synapses using nanotextured magnetic Josephson junctions. *Science Advances* **4**, e1701329, doi:10.1126/sciadv.1701329 (2018).
[1146] Yamanashi, Y., Umeda, K. & Yoshikawa, N. Pseudo Sigmoid Function Generator for a Superconductive Neural Network. IEEE Transactions on Applied Superconductivity 23, 1701004-1701004, doi:10.1109/TASC.2012.2228531 (2013).
[1147] Onomi, T., Kondo, T. & Nakajima, K. Implementation of high-speed single flux-quantum up/down counter for the neural computation using stochastic logic. *IEEE Transactions on Applied Superconductivity* **19**, 626-629 (2009).
[1148] D. E. Nikonov and I. A. Young, "Overview of Beyond-CMOS Devices and a Uniform Methodology for Their Benchmarking," *Proceedings of the IEEE,* vol. 101, no. 12, pp. 2498–2533, 2013
[1149] D. E. Nikonov and I. A. Young, "Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits," *Exploratory Solid-State Computational Devices and Circuits, IEEE Journal on,* vol. 1, pp. 3–11, 2015
[1150] C. Pan and A. Naeemi, "An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits,* vol. 3, pp. 101–110, 2017
[1151] A. Iyengar, K. Ramclam, and S. Ghosh, "DWM-PUF: A low-overhead, memory-based security primitive," *Hardware-Oriented Security and Trust (HOST), 2014 IEEE International Symposium on*, 2014, pp. 154–159: IEEE
[1152] G. S. Rose, N. McDonald, Y. Lok-Kwong, and B. Wysocki, "A write-time based memristive PUF for hardware security applications," *Computer-Aided Design (ICCAD), 2013 IEEE/ACM International Conference on*, 2013, pp. 830–833
[1153] S. T. C. Konigsmark, L. K. Hwang, C. Deming, and M. D. F. Wong, "CNPUF: A Carbon Nanotube-based Physically Unclonable Function for secure low-energy hardware design," *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*, 2014, pp. 73–78
[1154] J. Rajendran et al., "Nano meets security: Exploring nanoelectronic devices for security applications," *Proceedings of the IEEE,* vol. 103, no. 5, pp. 829–849, 2015
[1155] G. E. Suh, C. W. O'Donnell, I. Sachdev, and S. Devadas, "Design and implementation of the AEGIS single-chip secure processor using physical random functions," *ACM SIGARCH Computer Architecture News*, 2005, vol. 33, no. 2, pp. 25–36: IEEE Computer Society.
[1156] J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "Physical unclonable functions and public-key crypto for FPGA IP protection," *Field Programmable Logic and Applications, 2007. FPL 2007. International Conference on*, 2007, pp. 189–195: IEEE.
[1157] T. Marukame, T. Tanamoto, and Y. Mitani, "Extracting physically unclonable function from spin transfer switching characteristics in magnetic tunnel junctions," *IEEE Transactions on Magnetics,* vol. 50, no. 11, pp. 1–4, 2014.
[1158] J. Das, K. Scott, S. Rajaram, D. Burgett, and S. Bhanja, "MRAM PUF: A novel geometry based magnetic PUF with integrated CMOS," *IEEE Transactions on Nanotechnology,* vol. 14, no. 3, pp. 436–443, 2015.
[1159] L. Zhang, Z. H. Kong, C.-H. Chang, A. Cabrini, and G. Torelli, "Exploiting process variations and programming sensitivity of phase change memory for reconfigurable physical unclonable functions," *IEEE Transactions on Information Forensics and Security,* vol. 9, no. 6, pp. 921–932, 2014.
[1160] Y. Pang, H. Wu, B. Gao, N. Deng, D. Wu, R. Liu, S. Yu, A. Chen, and H. Qian, "Optimization of RRAM-Based Physical Unclonable Function With a Novel Differential Read-Out Method," *IEEE Electron. Dev. Lett*. 38(2), 168-171 (2017).
[1161] Y. Pang, H. Wu, B. Gao, D. Wu, A. Chen, and H. Qian, "A Novel PUF Against Machine Learning Attack: Implementation on a 16 Mb RRAM Chip," *IEDM* 290-293 (2017)
[1162] U. Rührmair, C. Jaeger, C. Hilgers, M. Algasinger, G. Csaba, and M. Stutzmann, "Security applications of diodes with unique current-voltage characteristics," *International Conference on Financial Cryptography and Data Security*, 2010, pp. 328–335: Springer.
[1163] D. P. C. Dimitrakopoulos, and J. Smith, "Authentication using graphene based devices as physical unclonable functions," 2004. Available: http://www.google.com/patents/US20140159040
[1164] S. C. Konigsmark, L. K. Hwang, D. Chen, and M. D. Wong, "Cnpuf: A carbon nanotube-based physically unclonable function for secure low-energy hardware design," *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*, 2014, pp. 73–78: IEEE

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1165] G. S. Rose, N. McDonald, L.-K. Yan, and B. Wysocki, "A write-time based memristive PUF for hardware security applications," *Proceedings of the International Conference on Computer-Aided Design*, 2013, pp. 830–833: IEEE Press
[1166] P. Koeberl, Ü. Kocabaş, and A.-R. Sadeghi, "Memristor PUFs: a new generation of memory-based physically unclonable functions," *Proceedings of the Conference on Design, Automation and Test in Europe*, 2013, pp. 428–431: EDA Consortium
[1167] Y. H. Tseng, C.-E. Huang, C.-H. Kuo, Y.-D. Chih, and C. J. Lin, "High density and ultra-small cell size of contact ReRAM (CR-RAM) in 90 nm CMOS logic technology and circuits," *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1–4: IEEE
[1168] C.-Y. Huang, W. C. Shen, Y.-H. Tseng, Y.-C. King, and C.-J. Lin, "A contact-resistive random-access-memory-based true random number generator," *IEEE Electron Device Letters,* vol. 33, no. 8, pp. 1108–1110, 2012
[1169] A. Colli, S. Pisana, A. Fasoli, J. Robertson, and A. C. Ferrari, "Electronic transport in ambipolar silicon nanowires," *physica status solidi (b),* vol. 244, no. 11, pp. 4161–4164, 2007
[1170] R. Martel et al., "Ambipolar Electrical Transport in Semiconducting Single-Wall Carbon Nanotubes," *Physical Review Letters,* vol. 87, no. 25, p. 256805, 12/03/ 2001
[1171] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nat Mater,* 10.1038/nmat1849 vol. 6, no. 3, pp. 183–191, 03//print 2007
[1172] L. Yu-Ming, J. Appenzeller, J. Knoch, and P. Avouris, "High-performance carbon nanotube field-effect transistor with tunable polarities," *Nanotechnology, IEEE Transactions on,* vol. 4, no. 5, pp. 481–489, 2005
[1173] N. Harada, K. Yagi, S. Sato, and N. Yokoyama, "A polarity-controllable graphene inverter," *Applied Physics Letters,* vol. 96, no. 1, p. 012102, 2010
[1174] A. Heinzig, S. Slesazeck, F. Kreupl, T. Mikolajick, and W. M. Weber, "Reconfigurable Silicon Nanowire Transistors," *Nano Letters,* vol. 12, no. 1, pp. 119–124, 2012/01/11 2012
[1175] S. Das and J. Appenzeller, "WSe2 field effect transistors with enhanced ambipolar characteristics," *Applied Physics Letters,* vol. 103, no. 10, p. 103501, 2013
[1176] M. De Marchi et al., "Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire FETs," *Electron Devices Meeting (IEDM), 2012 IEEE International*, 2012, pp. 8.4.1–8.4.4
[1177] T. Vasen, "Investigation of III-V Tunneling Field-Effect Transistors," A Dissertation submitted to the University of Notre Dame, 2014.
[1178] Y. Bi *et al.*, "Emerging Technology Based Design of Primitives for Hardware Security," *ACM Journal on Emerging Technologies in Computing Systems (JETC),* vol. 13, no. 1, Article 3, 2016
[1179] B. Yu, P. E. Gaillardon, X. S. Hu, M. Niemier, Y. Jiann-Shiun, and J. Yier, "Leveraging Emerging Technology for Hardware Security - Case Study on Silicon Nanowire FETs and Graphene SymFETs," *Test Symposium (ATS), 2014 IEEE 23rd Asian*, 2014, pp. 342–347
[1180] A. Chen, X. S. Hu, Y. Jin, M. Niemier, and X. Yin, "Using Emerging Technologies for Hardware Security Beyond PUFs," *Design, Automation, and Test in Europe (DATE),* Dresden, Germany, 2016
[1181] J. Rajendran *et al.*, "Fault Analysis-Based Logic Encryption," *Computers, IEEE Transactions on,* vol. 64, no. 2, pp. 410–424, 2015
[1182] Y. Bi *et al.*, "Emerging Technology Based Design of Primitives for Hardware Security," *ACM Journal on Emerging Technologies in Computing Systems (JETC),* vol. 13, no. 1, Article 3, 2016
[1183] B. Yu, P. E. Gaillardon, X. S. Hu, M. Niemier, Y. Jiann-Shiun, and J. Yier, "Leveraging Emerging Technology for Hardware Security - Case Study on Silicon Nanowire FETs and Graphene SymFETs," *Test Symposium (ATS), 2014 IEEE 23rd Asian*, 2014, pp. 342–347
[1184] Y. Bi, K. Shamsi, J.-S. Yuan, F.-X. Standaert, and Y. Jin, "Leverage emerging technologies for dpa-resilient block cipher design," *Proceedings of the 2016 Conference on Design, Automation & Test in Europe*, 2016, pp. 1538–1543: EDA Consortium
[1185] Y. Bi, K. Shamsi, J.-S. Yuan, Y. Jin, M. Niemier, and X. S. Hu, "Tunnel fet current mode logic for dpa-resilient circuit designs," *IEEE Transactions on Emerging Topics in Computing,* vol. 5, no. 3, pp. 340–352, 2017
[1186] A. C. Seabaugh and Z. Qin, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE,* vol. 98, no. 12, pp. 2095–2110, 2010
[1187] H. Lu and A. Seabaugh, "Tunnel field-effect transistors: State-of-the-art," *IEEE Journal of the Electron Devices Society,* vol. 2, no. 4, pp. 44–49, 2014
[1188] L. Britnell et al., "Resonant tunnelling and negative differential conductance in graphene transistors," *Nat Commun,* Article vol. 4, p. 1794, 04/30/online 2013
[1189] MishchenkoA et al., "Twist-controlled resonant tunnelling in graphene/boron nitride/graphene heterostructures," *Nat Nano,* Letter vol. 9, no. 10, pp. 808–813, 10//print 2014
[1190] M. O. Li, D. Esseni, J. J. Nahas, D. Jena, and H. G. Xing, "Two-Dimensional Heterojunction Interlayer Tunneling Field Effect Transistors (Thin-TFETs)," *Electron Devices Society, IEEE Journal of the,* vol. 3, no. 3, pp. 200–207, 2015
[1191] H. Mamiya, A. Miyaji, and H. Morimoto, "Efficient Countermeasures against RPA, DPA, and SPA," *Cryptographic Hardware and Embedded Systems – CHES 2004*, vol. 3156, M. Joye and J.-J. Quisquater, Eds. (Lecture Notes in Computer Science: Springer Berlin Heidelberg, 2004, pp. 343–356
[1192] B. Sedighi, J. J. Nahas, M. Niemier, and X. S. Hu, "Boolean circuit design using emerging tunneling devices," *Computer Design (ICCD), 2014 32nd IEEE International Conference on*, 2014, pp. 355–360
[1193] P. Kocher, "Design and validation strategies for obtaining assurance in countermeasures to power analysis and related attacks," NIST Physical Security Workshop, September 26-29 (2005).
[1194] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," *Annual International Cryptology Conference*, 1999, pp. 388–397: Springer

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1195] M.-L. Akkar and C. Giraud, "An implementation of DES and AES, secure against some attacks," *International Workshop on Cryptographic Hardware and Embedded Systems*, 2001, pp. 309–318: Springer

[1196] K. Tiri, M. Akmal, and I. Verbauwhede, "A dynamic and differential CMOS logic with signal independent power consumption to withstand differential power analysis on smart cards," *Solid-State Circuits Conference, 2002. ESSCIRC 2002. Proceedings of the 28th European*, 2002, pp. 403–406: IEEE

[1197] Y. Bi, K. Shamsi, J.-S. Yuan, Y. Jin, M. Niemier, and X. S. Hu, "Tunnel fet current mode logic for dpa-resilient circuit designs," *IEEE Transactions on Emerging Topics in Computing,* vol. 5, no. 3, pp. 340–352, 2017

[1198] C. De Canniere, O. Dunkelman, and M. Knežević, "KATAN and KTANTAN—a family of small and efficient hardware-oriented block ciphers," *Cryptographic Hardware and Embedded Systems-CHES 2009*: Springer, 2009, pp. 272–288

[1199] A. Cevrero, F. Regazzoni, M. Schwander, S. Badel, P. Ienne, and Y. Leblebici, "Power-gated mos current mode logic (pg-mcml): A power aware dpa-resistant standard cell library," *Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE*, 2011, pp. 1014–1019: IEEE

[1200] D. Gall, "The search for the most conductive metal for narrow interconnect lines" J. Apple. Phys. 127, 050901 (2020): https://doi.org/10.1063/1.5133671

[1201] M.S. Fuhrer, M.T. Edmonds, D. Culcer, M. Nadeem, X. Wang, N. Medhekar, Y. Yin, J and .H Cole, "Proposal for a Negative Capacitance Topological Quantum Field-Effect Transistor", in 2021 IEEE International Electron Devices Meeting (IEDM), 2021, pp. 38.2.1-38.2.4.

[1202] F. Zuo, P. Panda, M. Kotiuga, J. Li, M. Kang, C. Mazzoli, H. Zhou, A. Barbour, S. Wilkins, B. Narayanan, M. Cherukara, Z. Zhang, S. K. R. S. Sankaranarayanan, R. Comin, K. M. Rabe, K. Roy and S. Ramanathan, "Habituation based synaptic plasticity and organismic learning in a quantum perovskite", Nature Communications 8, 240 (2017): https://doi.org/10.1038/s41467-017-00248-6

[1203] S.-J. Kim, K. Ohkoda, M. Aono, H. Shima, M. Takahashi, Y. Naitoh, and H. Akinaga, "Reinforcement Learning System Comprising Resistive Analog Neuromorphic Devices", IEEE International Reliability Physics Symposium (2019): 10.1109/IRPS.2019.8720428

[1204] Z. Wang, C. Li, W. Song, M. Rao, D. Belkin, Y. Li, P. Yan, H. Jiang, P. Lin, M. Hu, J. P. Strachan, N. Ge, M. Barnell, Q. Wu, A. G. Barto, Q. Qiu, R. S. Williams, Q. Xia 1 and J. J. Yang, "Reinforcement learning with analogue memristor arrays", Nature Electronics 2, 115 (2019): https://doi.org/10.1038/s41928-019-0221-6

[1205] T. Machida, Y. Sun, S. Pyon, S. Takeda, Y. Kohsaka, T. Hanaguri, T. Sasagawa, and T. Tamegai, "Zero-energy vortex bound state in the superconducting topological surface state of Fe (Se,Te)", Nature Materials 18, 811–815 (2019): https://doi.org/10.1038/s41563-019-0397-1

[1206] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions", Nature 573, 390-393 (2019): https://doi.org/10.1038/s41586-019-1557-9

[1207] Z. Wang, H. Wu, G. W. Burr, C. S. Hwang, K. L. Wang, Q. Xia, and J. J. Yang, "Resistive switching materials for information processing", Nat. Rev. Mater. 5, 173–195 (2020). https://doi.org/10.1038/s41578-019-0159-3

[1208] Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H.-S. Philip Wong, and M, Lanza, "Electronic synapses made of layered two-dimensional materials", Nat. Electron. 1, 458–465 (2018): https://doi.org/10.1038/s41928-018-0118-9

[1209] S. Chen, M. R. Mahmoodi, Y. Shi, C. Mahata, B. Yuan, X. Liang, C. Wen, F. Hui, D. Akinwande, D. B. Strukov, and M. Lanza, "Wafer-scale integration of two-dimensional materials in high-density memristive crossbar arrays for artificial neural networks", Nat. Electron. 3, 638–645 (2020): https://doi.org/10.1038/s41928-020-00473-w

[1210] K. Sugawara,H. Shima, M. Takahashi, Y. Naitoh, H. Suga, and H. Akinaga, "Low-Frequency-Noise Spectroscopy of TaOx-based Resistive Switching Memory", Adv. Electron.Mater., 2100758, 2021: https://doi.org/10.1002/aelm.202100758

[1211] S.-W. Hwang, J.-K. Song, X. Huang, H. Cheng, S.-K. Kang, B. H. Kim , J.-H. Kim, S. Yu , Y. Huang, and J. A. Rogers, "High-Performance Biodegradable/Transient Electronics on Biodegradable Polymers", Adv. Mater. 26, 3905–3911 (2014): https://doi.org/10.1002/adma.201306050

[1212] H. Cheng and V. Vepachedu, "Recent development of transient electronics", Theoretical and Applied Mechanics Letters 6, 21–31 (2016): https://doi.org/10.1016/j.taml.2015.11.012

[1213] K. K. Fu, Z. Wang, J. Dai, M. Carter, and L. Hu, "Transient Electronics: Materials and Devices" Chem. Mater. 28, 3527−3539 (2016): https://doi.org/10.1021/acs.chemmater.5b04931

[1214] Y. Gao, Y. Zhang, X. Wang, K. Sim, J. Liu, J. Chen, X. Feng, H. Xu, and C. Yu, "Moisture-triggered physically transient electronics", Sci. Adv. 3, e1701222 (2017): https://doi.org/10.1126/sciadv.1701222

[1215] C. Arima, A. Suzuki, A. Kassai, T. Tsuruoka, and T. Hasegawa, "Development of a molecular gap-type atomic switch and its stochastic operation", J. Appl. Phys. 124, 152114 (2018): https://doi.org/10.1063/1.5037657

[1216] H. Sato, H. Shima, T. Nokami, T. Itoh, Y. Honma, Y. Naitoh, H. Akinaga, and K. Kinoshita, "Memristors with controllable data volatility by loading metal ion-added ionic liquids", Front. Nanotechnol. 3, 660563 (2021): https://doi.org/10.3389/fnano.2021.660563

[1217] F. Oba and Y. Kumagai, "Design and exploration of semiconductors from first principles: A review of recent advances", Appl. Phys. Express 11, 060101 (2018): https://doi.org/10.7567/APEX.11.060101

[1218] A. Ueda, Y. Zhang, N. Sano, H. Imamura and Y. Iwasa, "Ambipolar device simulation based on the drift-diffusion model in ion-gated transition metal dichalcogenide transistors", npj Comput Mater 6, 24 (2020): https://doi.org/10.1038/s41524-020-0293-x

[1219] E. Lidrikis, et al., "Coupling Length Scales for Multiscale Atomistics-Continuum Simulations: Atomistically induced stress distributions in Si/Si3N4 nanopixels", Phys. Rev. Lett. 87, 086104 (2001): https://doi.org/10.1103/PhysRevLett.87.086104

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.

[1220] M. Andersson, J. Yuan, and B. Sundén, "Review on modeling development for multiscale chemical reactions coupled transport phenomena in solid oxide fuel cells", Applied Energy 87, 1461-1476 (2010): https://doi.org/10.1016/j.apenergy.2009.11.013
[1221] H. Kobayashi et al., "Hopping and band mobilities of pentacene, rubrene, and 2,7-dioctyl[1]benzothieno[3,2-b][1]benzothiophene (C8-BTBT) from first principle calculations", J. Chem. Phys. 139, 014707 (2013): https://doi.org/10.1063/1.4812389
[1222] H. Kobayashi, and Y. Tokita, "Modeling of hole transport across grain boundaries in organic semiconductors for mesoscale simulations", Appl. Phys. Exp. 8, 051602 (2015): https://doi.org/10.7567/APEX.8.051602
[1223] H. Kobayashi et al., "Carrier mobility in mesoscale heterogeneous organic materials: Effects of crystallinity and anisotropy on efficient charge transport", Appl. Phys. Lett. 111, 033301 (2017): https://doi.org/10.1063/1.4995243
[1224] R. Liu et al., Integrating Materials and Manufacturing Innovation 4, 192 (2015): https://doi.org/10.1186/s40192-015-0042-z
[1225] 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.
[1226] Doug Matzke, "Will Physical Scalability Sabotage Performance Gains?" *IEEE Computer*, Volume 30, Issue 9, September, 1997, Page: 37 – 39.
[1227] 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.
[1228] 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.
[1229] D.E. Nikonov and I.A. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," *IEDM Tech. Dig*., pp. 573–576, Dec. 2012.
[1230] 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).
[1231] 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.
[1232] C. Pan and A. Naeemi, "An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits," *IEEE J. Exploratory Solid-State Computational Devices and Circuits*, vol. 3, pp 101-110, December 2017.
[1233] I. Sutherland et al., *Logical Effort: Design Fast CMOS Circuits*, 1st ed. San Mateo, CA: Morgan Kaufmann, Feb. 1999, ISBN: 10:1558605576

COPYRIGHT © 2022 IEEE. ALL RIGHTS RESERVED.