



INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS

INTERNATIONAL
ROADMAP
FOR
DEVICES AND SYSTEMS

2020 EDITION

BEYOND CMOS

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

© 2020 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.

Table of Contents

Acknowledgments	vi
1. Introduction	1
1.1. Scope of Beyond-CMOS Focus Team	1
1.2. Difficult Challenges	2
1.3. Nano-information Processing Taxonomy.....	4
2. Emerging Memory Devices	4
2.1. Memory Taxonomy	5
2.2. Emerging Memory Devices.....	6
2.3. Memory Selector Device.....	17
2.4. Storage Class Memory	20
3. Emerging Logic and Alternative Information Processing Devices.....	27
3.1. Taxonomy	27
3.2. Devices for CMOS Extension	28
3.3. Beyond-CMOS Devices: Charge-Based.....	32
3.4. Beyond-CMOS Devices: Alternative Information Processing.....	36
4. Emerging Device-Architecture Interaction.....	41
4.1. Introduction	41
4.2. Analog Computing	43
4.3. Probabilistic Circuits.....	57
4.4. Reversible Computing.....	59
4.5. Device-Architecture Interaction: Conclusions/Recommendations.....	64
5. Beyond CMOS Devices for More-Than-Moore Applications.....	64
5.1. Emerging Devices for Security Applications.....	64
6. Emerging Materials Integration	68
6.1. Introduction and Scope.....	68
6.2. Challenges	69
6.3. Technology Requirements and Potential Solutions.....	70
6.4. Emerging/Disruptive Concepts and Technologies.....	73
6.5. Conclusions and Recommendations	74
7. Assessment	74
7.1. Introduction	74
7.2. NRI Beyond-CMOS Benchmarking	75
7.3. Archive of ITRS ERD Survey-Based Assessment.....	77
8. Summary	80
9. Endnotes/References.....	81

List of Figures

Figure BC1.1	Relationship of More Moore, Beyond CMOS, and Novel Computing Paradigms and Applications (Courtesy of Japan beyond-CMOS group).....	1
Figure BC2.1	Taxonomy of Emerging Memory Devices.....	6
Figure BC2.2.	Taxonomy of Memory Select Devices	17
Figure BC2.3	Comparison of Performance of Different Memory Technologies.....	23
Figure BC3.1	Taxonomy of Options for Emerging Logic Devices	27
Figure BC4.1.	Conventional vs. Alternative Computing Paradigms	42
Figure BC4.2	A Resistive Memory Crossbar $Ax = b$ Solver is Illustrated.....	45
Figure BC4.3	A Spiking CNN for Gesture Recognition with Local Learning.....	47
Figure BC4.4	One Generalized Instantiation of a Photonic MVM unit, with Wavelength Multiplexed Inputs and Outputs and a Coupler-based Tunable Array. Reproduced from ⁹³¹	54
Figure BC4.5	Energy Dissipation per Stage vs. Frequency in an Adiabatic CMOS Shift Register	61
Figure BC4.6	Energy dissipation of RQFP and irreversible AQFP 1-b full adders	62
Figure EMI1	Emerging Material Integration Promotes the Advancement of Existing Technologies	69
Figure EMI2	An Example of the Role of Machine Learning in the Multiscale Simulation.....	74
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 ¹¹⁵⁶	76
Figure BC7.2	(a) Survey of Emerging Memory Devices and (b) Survey of Emerging Logic Devices in 2014 ERD Emerging Logic Workshop (Albuquerque, NM)	78
Figure BC7.3	Comparison of Emerging Memory Devices Based on 2013 Critical Review	78
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	79

List of Tables

Table BC1.1	Beyond CMOS Difficult Challenges.....	3
Table BC2.1	Emerging Research Memory Devices—Demonstrated and Projected Parameters	5
Table BC2.2	Experimental Demonstrations of Vertical Transistors in Memory Arrays	18
Table BC2.3	Benchmark Select Device Parameters.....	18
Table BC2.4a	Experimentally Demonstrated Two-terminal Memory Select Devices.....	20
Table BC2.4b	Experimentally Demonstrated Self-selecting Memory Devices (self-rectifying)	20
Table BC2.5	Target Device and System Specifications for SCM.....	22
Table BC2.6	Potential of Current Prototypical and Emerging Research Memory Candidates for SCM Applications	22
Table BC2.7.	Likely Desirable Properties of M (Memory) Type and S (Storage) Type Storage Class Memories.....	26
Table BC3.1a	MOSFETS: Extending MOSFETs to End of Roadmap	27
Table BC3.1b	Charge-based Beyond CMOS: Non-conventional FETs and Other Charge-based Information Carrier Devices	27
Table BC3.1c	Alternative Information Processing Devices	28
Table BC4.1	Metrics for Analog Capacitive Vector-Matrix Multiply (VMM) ICs	54
Table BC4.2	Comparison of Some Electrical Oscillators for Computing.....	56
Table EMI1	Near-term Difficult Challenges.....	69
Table EMI2	Long-term Difficult Challenges	70
Table EMI3	Materials for Transistor Scaling and Integration	71
Table EMI4	Materials for Lithography and Patterning.....	71
Table EMI5	Interconnect Materials	71
Table EMI6	Heterogeneous Integration, Assembly and Packaging Materials	72
Table EMI7	Emerging Research Materials Needs for Outside System Connectivity.....	72
Table EMI8	Emerging Materials for Memory	72
Table EMI9	Emerging Materials for Memory Select	72
Table EMI10	Emerging Materials for Advanced and Beyond-CMOS Logic Devices.....	72
Table EMI11	Spin Devices versus Materials	73
Table EMI12	Spin Material Requirements and Properties.....	73
Table EMI13	Metrology Needs and Challenges for Emerging Research Materials.....	73
Table EMI14	Modeling and Simulation	74
Table EMI15	Summary of Potentially Disruptive Emerging Research Materials Application Opportunities	74

ACKNOWLEDGMENTS

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

BEYOND CMOS

1. INTRODUCTION

1.1. SCOPE OF BEYOND-CMOS FOCUS TEAM

Dimensional and functional scaling¹ of CMOS is driving information processing² technology into a broadening spectrum of new applications. Scaling has enabled many of these applications through increased performance and complexity. As dimensional scaling of CMOS will eventually approach fundamental limits, several new information processing devices and microarchitectures for both existing and new functions are being explored to extend the historical integrated circuit scaling cadence. This is driving interest in new devices for information processing and memory, new technologies for heterogeneous integration of multiple functions, and new paradigms for system architecture. This chapter, therefore, provides an IRDS perspective on emerging research device technologies and serves as a bridge between conventional CMOS and the realm of nanoelectronics beyond the end of CMOS scaling.

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, e.g., hardware security.

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

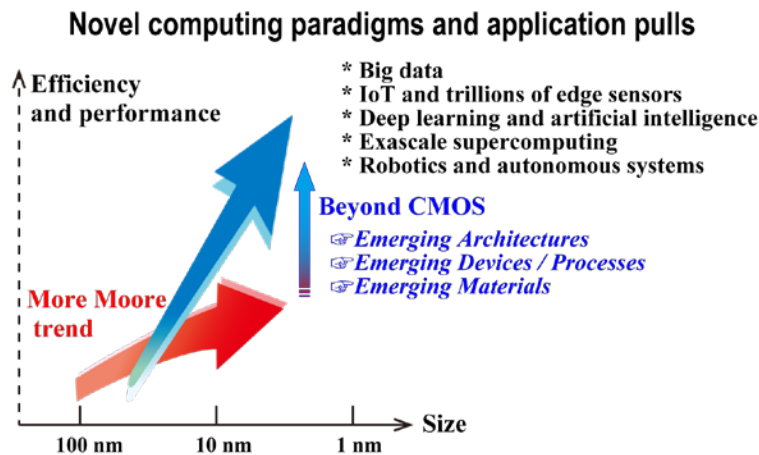


Figure BC1.1 Relationship of More Moore, Beyond CMOS, and Novel Computing Paradigms and Applications (Courtesy of Japan beyond-CMOS group)

The chapter is intended to provide an objective, informative resource for the constituent nanoelectronics communities pursuing: 1) research, 2) tool development, 3) funding support, and 4) investment. These communities include universities, research institutes, and industrial research laboratories; tool suppliers; research funding agencies; and the semiconductor industry. The potential and maturity of each emerging research device and architecture technology are reviewed and assessed to identify the

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.

2 Introduction

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. The former IRDS Emerging Research Materials (ERM) chapter is rolled into this chapter as section 6 renamed to “emerging materials integration”. 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 an MPU to execute programs is limited by interaction between the processor and the memory, and scaling does not automatically solve this problem. The current evolutionary solution is to increase MPU cache memory, thereby increasing the floor space that SRAM occupies on an MPU chip. This trend eventually leads to a decrease of the net information throughput. In addition to auxiliary circuitry to maintain stored data, volatility of semiconductor memory requires external storage media with slow access (e.g., magnetic hard drives, optical CD, etc.). Therefore, development of electrically accessible non-volatile memory with high speed and high density would initiate a revolution in computer architecture. This development would provide a significant increase in information throughput beyond the traditional benefits of scaling when fully realized for nanoscale CMOS devices.

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

The industry is now addressing the increasing importance of a new trend of functional diversification, where added value to devices is provided by incorporating functionalities that do not necessarily scale according to “Moore’s Law”. In this chapter, an “Beyond-CMOS devices for More-than-Moore Applications” section covers unconventional applications of existing and novel technologies. The section currently covers emerging devices for hardware security and will be expanded in future update.

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

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.

4 Emerging Memory Devices

1.2.3. MATERIALS TECHNOLOGIES

The most difficult challenge for Beyond CMOS is to deliver materials with controlled properties that will enable operation of emerging research devices in high density at the nanometer scale. To improve control of material properties for high-density devices, research on materials synthesis must be integrated with work on new and improved metrology and modeling. These important objectives are addressed in the “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 2017 – 2019) describing alternative approaches to established memory technologies.³ The scope of this section also includes updated subsections addressing the “Select Device” required for a crossbar memory application and an updated treatment of “Storage Class Memory” (including Solid State Disks).

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-

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.

dimensional array, and it is essential to consider the performance of memory cells in the context of this array architecture. A memory cell in such an array can be viewed as being composed of two fundamental components: the ‘storage node’ and the ‘select device’, the latter of which allows a given memory cell in an array to be addressed for read or write. Both components impact scaling limits for memory. For several emerging resistance-based memories, the storage node can, in principle, be scaled down below 10 nm,¹ and the memory density will be limited by the select device. Planar transistors (e.g. FET or BJT) are typically used as select devices. In a two-dimensional layout using in-plane select FETs the cell layout area is $A_{\text{cell}}=(6-8)F^2$. In order to reach the highest possible 2-D memory density of $4F^2$, a vertical select transistor can be used. Table 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.

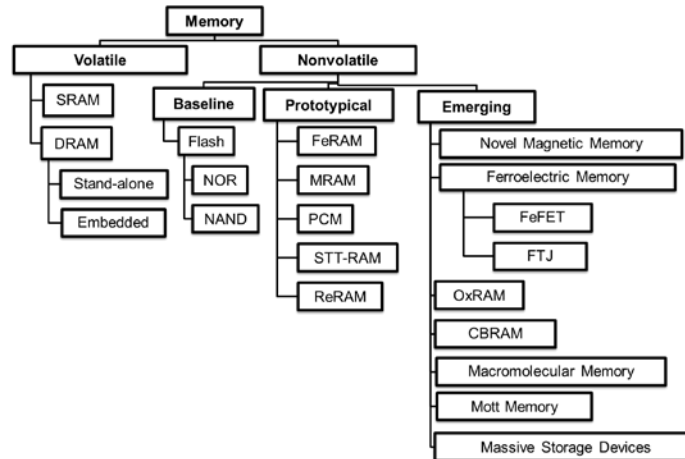


Figure BC2.1 Taxonomy of Emerging Memory Devices

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 design⁸. 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 architecture¹³ 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/cm²,^{14,15} compared with 3-10 ns switching of STT-MRAM at a current density of 7 MA/cm².^{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 effect¹⁸, and spin-momentum locking from topological protected electronic states¹⁹. Most experimental work has discovered a damping-like SOT in the in-plane transverse 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/cm² has been demonstrated in a 100 x 400 nm² type X MTJ, which will lead to 51- μ A, 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/cm² in 30 x 190 nm² 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/cm² of a 60 nm perpendicular MTJ with endurance up to 10¹¹ 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 alloys²¹, 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 Y²⁸. 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 MTJ³⁰. 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 coupling³⁵, exchange bias²⁴, and dipolar external magnetic field³⁶. 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 symmetry³⁷ 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) effect⁴⁰ 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 3d orbitals. Because the interfacial perpendicular magnetic anisotropy (PMA) originates from the Fe 3d and O 2p orbitals hybridization, this change of electron

8 Emerging Memory Devices

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/cm² 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/cm² 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 nm² to 50 x 120 nm² in-plane MTJ using 2-ns pulse with a current density of 12 MA/cm² with a high endurance of 10¹³ write cycles.⁴⁸ Another work shows 5-ns 62 μ A SOT current writing (VCMA bias of 1.2 V) of 30 x 80 nm² in-plane MTJ with WER <10⁻⁸ and endurance over 10¹² 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 ~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 SOT⁴⁷ and STT⁵⁸ 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 which relate to redox processes in the MIM cell. In these three ReRAM classes, there is a competition between thermal and electrochemical driving forces involved in the switching mechanism. Two major types of ReRAM exist: i) those based on metal oxide (OxRAM), which involve oxygen 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 endurance⁶⁷ 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 materials⁷³ (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). TaO_x and HfO_x 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 HfO_x device with a 10 nm dimension in 2011⁷⁴, scaling to a smaller size has been achieved by employing a sidewall electrode in a $1 \times 3 \text{ nm}^2$ 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 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) endurance⁷⁹ 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 - 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/ HfO_2 /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 PrCaMnO_3 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

10 Emerging Memory Devices

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.

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 (perovskites) 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 ($\text{HfO}_2/\text{A}_2\text{O}_3$) 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 reference⁹³ 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 electrodes⁹⁴). Only in symmetric structures (e.g. Pt/HfO₂/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

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 paper¹⁰³ shows an endurance of over 10⁶ 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 μm²), switching voltage of less than 3V, RESET current of less than 60 μA, endurance > 10⁶ cycles, and short SET and RESET times of 500 ns and 100 μs, respectively. One paper¹⁰⁷ 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 paper¹⁰⁸ 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, is a device which utilizes electrochemical control of nano-scale quantities of metal in thin dielectric films or solid electrolytes to perform the resistive switching operation.¹⁰⁹ The basic CBRAM cell is a metal–ion conductor–metal (MIM) system consisting of an electrode made of an electrochemically active material such as Ag, Cu or Ni, an electrochemically inert electrode such as W, Ta, or Pt, and a thin film of solid electrolyte sandwiched between both electrodes.¹¹⁰ Large, non-volatile resistance changes are caused by the oxidation and reduction of the metal ions by the application of low bias voltages. Key attributes are low voltage, low current, rapid write and erase, good retention and endurance, and the ability for the storage cells to be physically scaled to a few tens of nm. The material class for the dielectric film or the solid electrolyte is comprised of oxides, higher chalcogenides (including glasses), semiconductors, as well as organic compounds including polymers.

CBRAM is a strong emerging memory candidate primarily due to scalability (~10 nm),¹¹¹ ultra-low energy operations due to fast read, write and erase times, and low voltage requirements.¹¹² Maturity of the CBRAM technology development can be assessed

12 Emerging Memory Devices

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,¹¹³ embedded NVM,¹¹⁴ and serial interface non-volatile memory replacement.¹¹⁵ In 2012, a CBRAM-based serial NVM replacement product became commercially available.¹¹⁶ In 2014, a 16 GB CBRAM array based on a CuTe CBRAM cell was demonstrated.^{117,118} Such efforts are critical to identify core technology challenges¹¹⁹ and fundamental materials and mechanisms.¹²⁰ Novel applications such as reconfigurable switch¹²¹ and synaptic elements in Neuromorphic systems¹²² 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 is also released using Cu conduction bridge in a new polymer solid electrolyte.¹²³

As with other filamentary ReRAM technologies, CBRAM is challenged by bit level variability,¹¹⁹ the random nature of reliability failure such as retention or endurance, and random telegraph noise potentially contributing to read disturbs.¹²⁴ Such issues require large populations of bits to be studied, which suggests collaboration between universities and industry may be beneficial. Focus on fundamental understanding and simultaneously addressing some mitigation path such as error correction schemes, redundancy and algorithm development would enable closing the technology gap.

Engineering hurdles include the availability and integration of new materials used in CBRAM at advanced process nodes especially when there could be issues with compatibility of thermal budgets and process tooling. The availability of integrated array level information suggests that some of these challenges are being resolved in the recent years.^{115,121} Active participation from semiconductor equipment vendors and material suppliers would assist in overcoming manufacturing hurdles rapidly.

2.2.4. MACROMOLECULAR (POLYMER) MEMORY

Macromolecular memory is a category of memory which focuses 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 2005.¹²⁵ Around the same time, memory devices using gold nano-particles and 8-hydroxyquinoline, dispersed in a polystyrene matrix, were also demonstrated.¹²⁶ 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.^{127,128} Programming at very low power (70 nW) has been realized.¹²⁸ 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).¹²⁹ Downscaling of polymer resistive memory cell to the 100 nm length scale has been reported.¹³⁰ 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.¹³¹ For amorphous carbon, downscaling to nanometer sized cells has been published (1×10^3 nm²).¹³² Using carbon nanotubes as macromolecular electrodes and aluminum oxide as interlayer, isolated, non-volatile, rewriteable memory cells with an active area of essentially 36 nm² have been achieved, requiring a switching power less than 100 nW, with estimated switching energies below 10 fJ per bit.¹³³ 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.^{134,135} The recent and past studies show resistive,¹³⁶ capacitive (charge storage, based on electric dipole formation)^{137,138} and ferroelectric behavior¹³⁹ 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 memory¹⁴⁰ and could benefit enormously from development of new macromolecular polymeric materials with combined ferroelectric switching and semiconducting structural units. Finally, a number of macromolecular memories involve oxide layers. Here mutually beneficial interaction with the (research) community on metal oxide ReRAM switching could spring, because at the macromolecular / oxide interface trap states can be engineered by tuning the electron levels of the macromolecular material.

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

It is not possible to replace silicon-based memory devices 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 HfO₂ 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 .¹⁴¹ 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,^{142,143} but rather due to their low E_c (SBT: 10-100 kV/cm, PZT: ~50 kV/cm, summarized in¹⁴⁴) 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-HfO₂.¹⁴⁵ This CMOS-compatible material innovation enabled the demonstration of a FeFET

14 Emerging Memory Devices

technology scaled to the 28 nm node utilizing a conventional HKMG technology and is already used in high volume production.¹⁴⁶ The close resemblance of the HKMG transistor and the FE-HfO₂-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 insulator¹⁴⁷. The capacitive voltage divider is therefore directly responsible for the retention loss during stand-by as well as for the gate voltage distribution and the corresponding charge injection during write operations. This retention- and endurance-critical distribution of the electric field within the gate stack may be optimized by choosing the insulator capacitance as high as possible and the ferroelectric capacitance as low as possible. In the perovskite-based FeFET this is achieved by utilizing high-k buffer layers and is additionally fostered by the unavoidably large physical thickness of the perovskite ferroelectrics.^{4, 148} In the case of the aggressively scaled FE-HfO₂-based FeFET, the small thickness of the ferroelectric is compensated by the comparably low permittivity of HfO₂, the possibility to use ultra-thin interfacial layers, and by the depolarization resilience of the high E_c.^{144, 149} 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 10¹² switching cycles have been demonstrated for FE-HfO₂-^{145, 146, 150, 151} as well as for perovskite-based FeFETs.^{141, 152, 153} In the case of cycling endurance, however, the high E_c of FE-HfO₂ 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 10⁵ cycles observed in FE-HfO₂-based FeFETs with ultra-thin interfacial layer enabling excellent data retention.¹⁵⁴ 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.¹⁵⁰ 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)¹⁴¹, a reduced trap density at the interfaces,¹⁵⁴ an optimized capacitive voltage divider by area scaling in the MF₂MISFET approach¹⁵⁵ or the realization of a MFSFET device by implementing recent breakthroughs in the epitaxial growth of FE-HfO₂.¹⁵⁶ Despite promising results obtained for perovskite-based FeFET devices implemented into 64Kb NAND-Arrays at a feature size of 5 μm ¹⁵³, 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 HfO₂ 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. ^{157, 158, 159}. Recently, 64 kb and 32 Mb FeFET arrays were demonstrated in the 28 nm¹⁶⁰ and the 22 nm FD-SOI CMOS platform,¹⁶¹ 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 ZrO₂ thin film capacitors may allude to new way of addressing and potentially solving some of these challenges in FeFETs.¹⁶²

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 (¹⁶³ and references therein). However, giant TER of $> 10^4$ has most recently been reported in a super-tetragonal BiFeO₃ based FTJ by Yamada et al.¹⁶⁴ A similarly high TER was demonstrated by Wen and co-workers¹⁶⁵ for a BaTiO₃ 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 BiFeO₃ 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 (> 10¹⁴) 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.¹⁶⁶ 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 4x10⁶ endurance cycles and extrapolated data retention of 10 years at room temperature for a BiFeO₃-based FTJ.¹⁶⁷ In the context of retention, it should be noted that despite improved TER, the newly proposed MFS-FTJ structure will give rise to a depolarization field, which will most likely degrade memory retention in a similar manner as described for the FeFET in Section 2.2.5.1. The highly energy efficient electric field switching, common to all ferroelectric memories, enables fast (10 ns¹⁶⁸) and low voltage (1.4 V¹⁶⁷) switching in FTJ devices and results in a minimal power consumption during write operation (1.4 fJ/bit, calculation based on the device characteristics given¹⁶⁹). 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.¹⁷⁰ Simply based on the lateral dimensions of the FTJ element (assuming unlimited scalability of the selector), scaling below 50 nm²,¹⁶⁹ based on PFM data,¹⁷¹ 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 1.4x10⁵ A/cm² current density at 300 nm feature size has been achieved by Bruno et al.¹⁷² 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.¹⁷³

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. ¹⁷⁴. 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 (¹⁷⁵ 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 HfO₂ and its doped variant¹⁵¹ as well as recent breakthroughs in its epitaxial growth¹⁵⁶ might yield great potential for the manufacturability of competitive FTJs. Experimental demonstrations of FTJs based on doped variant of HfO₂ were recently reported in Refs. ^{176,177,178}.

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 10²⁴ bits, and the high estimate is ~10²⁹ bits¹⁷⁹ (these estimates are based on research by Hilbert and Lopez¹⁸⁰). 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 10²⁴ bits, for which >10⁹ kg of silicon Flash memory would be needed.¹⁷⁹ Thus, a few tens of kilograms of DNA could meet all of the world's storage needs for centuries to come.

16 Emerging Memory Devices

A number of recent studies have shown that DNA can support scalable, random-access and error-free information storage.^{181,182,183} 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.¹⁸⁴ The DNA-stored files are compatible with mainstream digital format, and large-scale DNA storage up to 200 MB has been demonstrated.¹⁸⁵ 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 – $10^7\times$) projected ultimate electronic memory densities. Also, in the living cell, the memory read/write operations occur at high speed ($<100\ \mu\text{s/bit}$) and require very low energy of $\sim 10^{-17}$ J/bit or 10^{-11} W/GB.¹⁸⁶ DNA can store information stably at room temperature for hundreds of years with zero power requirements, making it an excellent candidate for large-scale archival storage.¹⁸⁶ 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.¹⁸⁷ One of the goals for research efforts is to demonstrate miniaturized, on-chip integrated DNA storage. New methods for DNA synthesis and sequencing are key components for these developments.

Two major categories of technical challenges remain:

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

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

2.2.7. MOTT MEMORY

Mott memory is a metal/insulator/metal capacitor structure consisting of a correlated electron insulator (or Mott insulator). Correlated electron insulators often show the electronic phase transition accompanied by a drastic change in their resistivity under external stimuli such as temperature, magnetic field, electric field, and light. Mott memory exploits this electronic phase transition (called Mott metal-to-insulator transition or Mott transition¹⁸⁸) 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.¹⁸⁹ 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.¹⁸⁸ In addition, Mott memory based on the Mott transition involving a large number of carriers (more than $10^{22}\ \text{cm}^{-3}$) 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 $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$.¹⁹⁰ After this demonstration, two-terminal devices such as switches and memories have been intensively studied using such correlated electron oxides as $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$,^{191,192} VO_2 ,^{193,194,195} SmNiO_3 ,¹⁹⁶ NiO ,^{197,198} Ca_2RuO_4 ,¹⁹⁹ and NbO_2 ,^{200,201} and using Mott-insulator chalcogenides of AM_4X_8 ($\text{A}=\text{Ga, Ge}$; $\text{M}=\text{V, Nb, Ta}$; $\text{X}=\text{S, Se}$).^{202,203,204,205} 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).²⁰⁶

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.¹⁹⁶ 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.²⁰⁷ 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 $110 \times 110\ \text{nm}^2$ 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 NbO₂-Mott memristors have been evaluated to be less than 2.3 ns, of the order of 100 fJ, and $>10^9$, respectively.^{200,201} The programing and read voltages reported so far are <2 V and <0.2 V, respectively.¹⁹⁸ The non-volatile resistive switching of AM₄X₈ single crystals was induced by the electric field of less than 10 kV/cm.^{202,203,204,205} This suggests that if the device consisting of a 10-nm-thick AM₄X₈ 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 AM₄X₈ and θ -(BEDT-TTF)₂X, their retention characteristics are not elucidated in detail.^{202,203,204,205,206} In addition, the NbO₂-Mott memristors and VO₂-based devices are volatile switch.^{193,194,195,200,201} 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.¹⁸⁸ 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 transition¹⁸⁹ and the formation of conductive filament generated by local Mott transition.^{202,204,205} However, a thorough understanding of the mechanism has not been achieved yet. Therefore, the elucidation of detailed mechanism is also a major research challenge.

2.3. MEMORY SELECTOR DEVICE

The *capacity* (or *density*) is one of the most important parameters for memory systems. In a typical memory system, memory devices (cells) are connected to form an array. A memory cell in an array can be viewed as being composed of two components: the ‘*storage node*’, which is usually characterized by an element with switchable states, and the ‘*selector*’, which allows the storage node to be selectively addressed for read and write. Both components impact scaling limits of memory. It should be noted that for several advanced concepts of resistance-based memories, the storage node could in principle be scaled down below 10 nm,²⁰⁸ and the memory density is often limited by the selector devices. Thus, the selector device represents a serious bottleneck for emerging memory scaling to 10 nm and beyond.

The most commonly used memory selector devices are transistors (e.g., FET or BJT), as in DRAM, FRAM, *etc.* Flash memory is an example of a storage node (floating gate) and a selector (transistor) combined in one device. Planar transistors typically have the footprint around $(6-8)F^2$. In order to reach the highest possible 2D memory density of $4F^2$, a vertical transistor selector needs to be used. However, transistors as selector devices are generally unsuitable for 3D memory architectures. Two-terminal selector devices are preferred for scalability and can be used in crossbar memory arrays to achieve $4F^2$ footprint.^{209,210} 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).²¹¹ Volatile switches can also be used as selector devices. Figure BC2.2 shows a taxonomy of memory selector devices. In addition to external selector devices, some storage elements may have inherent self-selecting properties (e.g., intrinsic nonlinearity or self-rectification), which may enable functional crossbar arrays without external selectors.

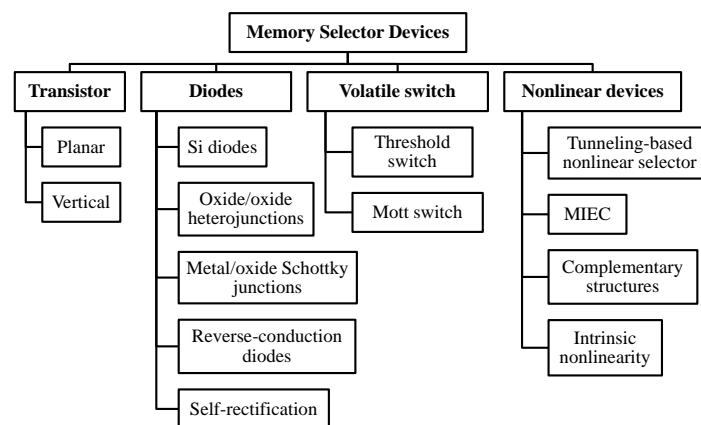


Figure BC2.2. Taxonomy of Memory Select Devices

2.3.1. VERTICAL TRANSISTORS

Several examples of experimental demonstrations of vertical transistors used as selectors in memory arrays are presented in Table BC2.3. While a vertical transistor selector allows for the highest planar array density ($4F^2$), it is challenging to integrate a transistor selector into stacked 3D memory. For example, to avoid thermal stress on the memory elements on the existing layers, the processing temperature of the vertical transistor in 3D stacks must be low. Also, making contact to the third terminal (gate) of vertical transistors constitutes an additional integration challenge, which usually results in cells size larger than $(4F^2)$;²¹² although true $4F^2$ arrays can, in principle, be implemented with 3-terminal selector devices.²¹³

Table BC2.2 *Experimental Demonstrations of Vertical Transistors in Memory Arrays*

2.3.2. DIODE SELECTOR DEVICES

General requirements for two-terminal selector devices are sufficient ON currents at proper bias to support read and write operations and sufficient ON/OFF ratio to enable selection. The minimum ON current required for fast read operation is $\sim 1 \mu\text{A}$ (Table BC2.4). The required ON/OFF ratio depends on the size of the memory block, $m \times m$: for example using a standard scheme of array biasing the required ON/OFF ratio should be in the range of 10^7 – 10^8 for $m=10^3$ – 10^4 , in order to minimize the ‘sneak’ currents.²¹⁴ These specifications are quite challenging, and the experimentally demonstrated selector devices can rarely meet them. Thus, selector devices are becoming a critical challenge of emerging memory. It should be noted that different application targets of resistance-based memories also impact selector device requirements.

Table BC2.3 *Benchmark Select Device Parameters*

The simplest realization of diode selectors uses semiconductor diodes, such as a *pn*-junction diode, Schottky diode, or heterojunction diode. Such devices are suitable for a unipolar memory cell. For bipolar memory cells, selectors with bi-directional switching are needed. Proposed examples include Zener diodes,²¹⁵ BARITT diodes,²¹⁶ and reverse breakdown Schottky diodes.²¹⁷

2.3.2.1. SI DIODE SELECTOR DEVICES

Both single-crystal Si²¹⁸ and poly-Si^{219,220,221} diodes have been developed as selector devices for PCM arrays. To provide high ON current, the contact resistivity needs to be reduced to $<10^{-7} \Omega \text{ cm}^2$, which was achieved by engineering the metal electrodes and electrode-Si interface.²¹⁹ A short-time annealing technique helps to reduce the OFF current and enlarge the ON/OFF ratio. Poly-Si technology can achieve ON current density of 10^7 A/cm^2 (at $\sim 1.8\text{V}$) and an ON/OFF ratio of 10^8 . It is believed that Si diodes can be scaled beyond 20 nm or 10 nm. Poly-Si diode selector devices have been integrated in PCM crossbar arrays, 3D vertical chain-cell type PCM,²²⁰ and a 1 Gb PCM test chip.²²¹ A major challenge of Si diodes is the high processing temperature (above 1000°C) required to crystallize Si to reduce contact resistivity and OFF current.

2.3.2.2. OXIDE DIODE SELECTOR DEVICES

Oxide-based heterojunction^{222,223,224,225} or Schottky junction^{228,229,230,231} diodes may be fabricated at lower temperatures and used as selector devices. They are particularly suitable for oxide-based RRAM devices. A p-NiO_x/n-TiO_x diode has demonstrated a rectification ratio of 10^5 at $\pm 3\text{V}$ and ON current density of $5 \times 10^3 \text{ A/cm}^2$ (at $\sim 2.5\text{V}$).²²² A p-CuO_x/n-InZnO_x diode achieved higher ON current density of 10^4 A/cm^2 (at $\sim 1.3\text{V}$) and was integrated with NiO_x RRAM in a 2-layer 8×8 crossbar array^{223,224} and with Al₂O₃ antifuse in a one-time-programmable (OTP) memory.²²⁵ Si substrates can be used as a part of heterojunction diodes as demonstrated in n-ZnO/p-Si²²⁶ and n-Ge-nanowire/p-Si diodes.²²⁷ In TiO_x-based diodes with Pt electrodes, temperature-dependent current-voltage (I-V) characteristics confirm a Schottky barrier of $\sim 0.55 \text{ eV}$ at the TiO_x/Pt interface.²²⁸ The rectification ratio is $\sim 1.6 \times 10^4$ at $\pm 1\text{V}$ but ON current density is low ($\sim 13 \text{ A/cm}^2$) due to large size. A Pt/TiO₂/Ti diode with Pt as the Schottky contact and Ti the ohmic contact achieved higher rectification ratio of $10^7 - 10^9$ at $\pm 1\text{V}$.²²⁹ Another demonstration of Pt/TiO₂/Ti Schottky diodes improved ON current density to $\sim 3 \times 10^5 \text{ A/cm}^2$ at 2 V on a $4 \mu\text{m}^2$ area.²³⁰ Measurement showed that current is not uniform across the diode area, probably due to edge leakage. Therefore, current density is higher at smaller diode size. An Ag/n-ZnO Schottky diode with non-alloyed Ti/Au ohmic contact demonstrated a rectification ratio of 10^5 and forward current density over 10^4 A/cm^2 at 2 V.²³¹ In addition to oxide Schottky diodes, Si Schottky diodes are also utilized as selector devices, e.g., Al/p-Si.²³² The ON current of oxide-based heterojunction diodes is often limited by both contact resistance and density of states of the oxide materials.

2.3.3. VOLATILE SWITCH AS SELECTOR DEVICES

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

2.3.3.1. MOTT SWITCH

This device is based on metal-insulator transition (i.e., Mott transition) and exhibits a low resistance above a critical electric field (threshold voltage, V_{th}). It recovers to a high-resistance state if the voltage is below a hold voltage (V_{hold}). If the electronic conditions that triggered Mott transition can relax within the memory device operation time scale, the Mott transition device is essentially a volatile resistive switch and can be utilized as a selector device. A VO_2 -based device has been demonstrated as a selector device for NiO_x RRAM element.²³³ However, the feasibility of Mott-transition switches as selector devices still needs further research. It should be noted that VO_2 undergoes a phase transition to the metallic state at temperature around 68°C, which restricts its operation temperatures and limits practical applications of a VO_2 selector as current specifications require operational temperature of 85°C. Suitable Mott materials with higher transition temperatures need to be investigated. Metal insulator transitions at ~ 130°C, and electrically driven switching were observed in thin films of $SmNiO_3$.²³⁴

2.3.3.2. THRESHOLD SWITCH

This type of device is based on the threshold-switching effect observed in thin-film MIM structures caused by electronic charge injection. Significant resistance reduction occurs at V_{th} , and this low-resistance state quickly recovers to the original high-resistance state when the applied voltage falls below V_{hold} . It was reported that chalcogenide-based threshold switches could be used as access devices in PCM arrays.²³⁵ Niobium oxide is found to possess both memory switching and threshold switching properties at different compositions, based on which hybrid memory (W/bi-layer- NbO_x /Pt) was demonstrated in a 1kb array.²³⁶ NbO_x -based selectors have also been integrated with TiO_x/TaO_x based RRAM in crossbar arrays at 5xnm node.²³⁷ In Si-As-Te ternary alloy, the composition (controlled by the sputtering power during deposition) determines the emergence of threshold switching.²³⁸ Both V_{th} and V_{hold} vary with composition, which may provide a method to optimize the selector device operation window. Another threshold switch device based on chalcogenide AsTeGeSiN was shown to be scalable to 30nm with current density exceeding 10MA/cm² and endurance over 10⁸ cycles.^{239,240} It was integrated with TaO_x -based RRAM devices. An unavoidable finite delay time was found in this selector device due to intrinsic properties of the chalcogenide material, which may limit the selector speed. Another doped-chalcogenide (material undisclosed) based selector demonstrates low V_{hold} (0.2 V), large ON/OFF ratio (>10⁷), fast speed (<10 ns), long endurance (>10⁹ cycles) and good thermal stability (180°C).²⁴¹ A so-called “FAST” (Field Assisted Superliner Threshold) selector was recently reported, with abrupt switching (<5 mV/dec), high ON/OFF ratio (10⁷), and long endurance (10⁸ cycles).²⁴² Unlike other threshold switch selectors, this device does not exhibit recovery to OFF-state at certain V_{hold} , which appears more like a nonlinear selector. A 4Mb 1S1R crossbar RRAM array has been demonstrated based on this selector.

2.3.4. NONLINEAR SELECTOR DEVICES

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

2.3.4.1. NONLINEAR SELECTOR DEVICES

Nonlinearity in device characteristics can be introduced with non-ohmic transport mechanisms, e.g., tunneling. A Ni/ TiO_2 /Ni nonlinear selector device is integrated with HfO_2 -RRAM to demonstrate a 1S1R memory structure.²⁴³ Another selector device with Pt/ TiO_2 /TiN structure is combined with a bi-layer Pt/ TiO_{2-x} / TiO_2 /W RRAM for a functional memory device.²⁴⁴ A so-called “varistor” selector device is based on a sandwiched $TaO_x/TiO_2/TaO_x$ structure.²⁴⁵ It was found that the substitution of Ti^{4+} in TiO_2 by Ta^{5+} ions increases the conductivity of the initially insulating TiO_2 layer. The ON-current of nonlinear selector devices can be modulated by oxide thickness and oxidation conditions. Another multi-oxide stack ($Ta_2O_5/TaO_x/TiO_x$) based selector also leverages various interface engineering techniques to obtain high ON-current density (>10⁷ A/cm²), high nonlinearity ratio (~10⁴), and low OFF-current (<100 nA).²⁴⁶ It is integrated with CBRAM in a 1 kb crossbar array. A back-to-back diode structure, $n^+/p/n^+$ poly-Si, is also suggested as a selector device, where the middle p-layer is fully depleted and a drain induced barrier lowering (DIBL) effect causes exponential current increase with applied voltage.²⁴⁷

2.3.4.2. MIEC SWITCH

The device is made from Cu-containing “mixed ionic and electronic conduction” (MIEC) materials sandwiched between an inert top electrode (TE) (e.g., TiN, W) and a bottom electrode (BE). Negative voltage applied on TE pulls Cu^+ in MIEC away from the BE and creates vacancies near the BE. The hole and vacancy concentrations depend exponentially on the applied voltage. Symmetrical diode-like I-V characteristics are achieved with two inert electrodes. Large fraction of mobile Cu^+ enables high

20 Emerging Memory Devices

current density ($> \text{tens of MA/cm}^2$).²⁴⁸ Endurance above 10^8 cycles has been demonstrated on MIEC devices in small arrays.²⁴⁹ The MIEC selector devices were also integrated with PCM in a 512 kb testing array using 180nm CMOS process.²⁵⁰ The scalability of MIEC select devices was tested to below 30nm in diameter and below 12nm in thickness.²⁵¹

2.3.4.3. COMPLEMENTARY RESISTIVE SWITCHES

A complementary resistive switch (CRS) provides a self-selecting memory by connecting two bipolar RRAM devices anti-serially.²⁵² It may be considered as “constructed nonlinearity”. Both states “0” and “1” have high resistance in CRS, which helps to minimize leakage through sneak paths. In either state, one of the two RRAMs is in LRS and the other in HRS. When reading a “1” state, the HRS device is switched to LRS and both devices end up in LRS. When reading a “0” state, no switching occurs and CRS remains in HRS. Notice that the reading operation is destructive, although non-destructive readout method was also proposed.²⁵³ CRS has been demonstrated in different resistive switching devices, e.g., Cu/SiO₂/Pt bipolar resistive switches,²⁵⁴ amorphous carbon-based RRAM,²⁵⁵ TaO_x-based RRAM,²⁵⁶ multi-layer TiO_x device,²⁵⁷ HfO_x RRAM,²⁵⁸ ZrO_x/HfO_x bi-layer RRAM,²⁵⁹ Cu/TaO₂ atomic switch,²⁶⁰ Nb₂O_{5-x}/NbO_y RRAM,²⁶¹ etc.

Table BC2.5a summarizes experimentally demonstrated parameters of some two-terminal select devices, including diodes, volatile switches, and nonlinear devices. Table BC2.5b summarizes the parameters of some reported self-rectifying memories. It should be emphasized that these summary tables can only capture a snapshot of selector device characteristics; however, the functionality of these devices depends on their actual voltage in arrays with random data patterns and the balance between selectors and storage elements. Therefore, these parameter tables should only be used for illustration purpose, not for rigorous benchmark or assessment.

It remains a great challenge for the demonstrated selector devices to meet all the requirements in Table BC2.4. For scaled two-terminal select devices, two fundamental challenges are *contact resistance*²¹⁹ and *lateral depletion effects*.^{262,263} Very high doping concentration is needed to minimize both effects. However, high doping concentrations result in increased reverse bias currents in classical diode structures and therefore reduced $I_{\text{on}}/I_{\text{off}}$ ratio. For switch-type selector devices the main challenges are identifying the right material and the switching mechanism to achieve the required drive current density, $I_{\text{on}}/I_{\text{off}}$ ratio, and reliability.

Table BC2.4a Experimentally Demonstrated Two-terminal Memory Select Devices

Table BC2.4b Experimentally Demonstrated Self-selecting Memory Devices (self-rectifying)

2.4. STORAGE CLASS MEMORY

2.4.1. STORAGE CLASS MEMORY DEVICES

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

Conventionally, magnetic hard-disk drives are used for non-volatile data storage. The cost of HDD storage in \$/GB is extremely low and continues to decrease. Although the bandwidth with which contiguous data can be streamed is high, the poor random-access time of HDDs limits the maximum number of I/O requests per second (IOPs). In addition, HDDs have relatively high energy consumption, a large form factor, and are subject to mechanical reliability failures in ways that solid state technologies are not. Despite these issues, the sheer number and growth in HDD shipments per year (380,000 Petabytes in 2012, growing at 32% per year) means that magnetic disk storage is highly unlikely to be “replaced” by solid-state drives at any time in the foreseeable future.²⁶⁴

Non-volatile semiconductor memory in the form of NAND Flash has become a widely-used alternative storage technology, offering faster access times, smaller size and lower energy consumption when compared to HDD. However, there are several serious limitations of NAND Flash for storage applications, such as poor endurance (10^3 – 10^5 erase cycles), only modest retention (typically 10 years on a new device, but only 1 year at the end of rated endurance lifetime), long erase time (~ms), and high operating voltage (~15 V). Another difficult challenge of NAND Flash SSD is posed by its page/block-based architecture. By not allowing for direct overwrite of data, sophisticated procedures for garbage collection, wear-leveling and bulk erase are required. This in turn requires additional computation – which reduces performance and increases cost and power because of the need for a local processor, RAM, and logic – as well as over-provisioning of the SSD which further increases cost per effective user-bit of data.²⁶⁵

Although Flash memory technology continues to project for further density scaling, inherent performance characteristics such as read, write and erase latencies have been nearly constant for more than a decade.²⁶⁶ While the introduction of multi-level cell (MLC) Flash devices extended Flash memory capacities by a small integral factor (2–4), the combination of scaling and MLC have resulted in the degradation of both retention time and endurance, two parameters critical for storage applications. The migration of NAND Flash into the vertical dimension above the silicon has continued this trend of improving bit density (and thus cost-per-bit) while maintaining or in some cases, even slightly degrading the latency, retention, and endurance characteristics of present-day NAND Flash.

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

2.4.1.2. WHAT IS STORAGE CLASS MEMORY?

Storage-class memory (SCM) describes a device category that combines the benefits of solid-state memory, such as high performance and robustness, with the archival capabilities and low cost of conventional hard-disk magnetic storage.^{267,268} Such a device requires a non-volatile memory (NVM) technology that could be manufactured at a very low cost per bit.

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

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

2.4.1.2.1. STORAGE-TYPE SCM

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

Note however that such system cost reduction can come from other sources than the raw cost of the device technology: a slightly-higher-cost NVM technology that enabled a simple, low-cost SSD by eliminating or simplifying costly and /or performance-degrading overhead components would achieve the same overall goal. If the cost per bit could be driven low enough through ultrahigh memory density, ultimately such an S-SCM device could potentially replace magnetic hard-disk drives in enterprise storage server systems as well as in mobile computers (subject to the same issues mentioned above in terms of needing numerous IC fabs to ship the many petabytes of HDD delivered to those markets²⁶⁴).

2.4.1.2.2. MEMORY-TYPE SCM

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

Particularly critical for M-SCM will be device endurance, since the time available for wear-leveling, error-correction, and other similar techniques is limited. The volatile portion of the memory hierarchy will have effectively infinite endurance compared to any of the non-volatile memory candidates that could become an M-SCM. Even if device endurance can be pushed well over 10⁹

22 Emerging Memory Devices

cycles, it is quite likely that the role of M-SCM will need to be carefully engineered within a cascaded-cache or other Hybrid Memory approach.²⁶⁹ That said, M-SCM offers a host of new opportunities to system designers, opening up the possibility of programming with truly persistent data, committing critical transactions to M-SCM rather than to HDD, and performing commit-in-place database operations.

2.4.1.3. TARGET SPECIFICATIONS FOR SCM

Since the density and cost requirements of SCM transcend the straightforward scaling application of Moore's Law, additional techniques will be needed to achieve the ultrahigh memory densities and extremely low cost demanded by SCM, such as 1) 3-D integration of multiple layers of memory, currently implemented commercially for write-once solid-state memory,²⁷⁰ and/or 2) Multiple level cell (MLC) techniques.

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

In order to make overall system power usage (as shown in Table BC2.6) competitive with NAND Flash and HDD, and since faster I/O interfaces can be expected to consume considerable power, the device-level power requirements must be extremely minimal. This is particularly important since low latency is necessary but not sufficient for enabling high bandwidth – high parallelism is also required. This in turn mandates a sufficiently low power per bit access, both in terms of peripheral circuitry and device-level write and read power requirements. Finally, standby power should be made extremely low, offering opportunities for significant system power savings without loss of performance through rapid switching between active and standby states.

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

Also shown in Table BC2.6 are the target specifications for M-type SCM devices. Given the faster latency target (which enables coherent access through a memory controller), the program-erase cycle endurance must be higher, so that the overall non-volatile memory system can offer a sufficiently large lifetime before needing replacement or upgrade. Although some studies have shown that a device endurance of 10^7 cycles is sufficient to enable device lifetimes on the order of 3–10 years,²⁷¹ we anticipate that the need for sufficient engineering margin would suggest a minimum cycle endurance of 10^9 cycles. While such endurance levels support the use of M-class SCM in memory support roles, significantly higher endurance values (e.g., $1e12$ or $1e14$ cycles) would allow M-class SCM to be used in more varied memory applications, where the total number of memory accesses may become very large.

Table BC2.5 *Target Device and System Specifications for SCM*

Table BC2.6 *Potential of Current Prototypical and Emerging Research Memory Candidates for SCM Applications*

2.4.1.4. FIRST SCM PRODUCTS REACH THE MARKET

In July 2015, 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.^{272,273} (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.^{274,275} The particular non-volatile memory device was not specified, other than that it depends on bulk changes of resistance,²⁷⁶ nor was the nonlinear access device described. Speculation based on patent searches and job solicitations suggest that the technology may be a combination of some variant of phase change memory and an Ovonic Threshold Switching access device.²⁷⁶

While the details of the devices involved may still be uncertain, the projected array specifications and the target applications are, for all intents and purposes, indistinguishable from those described above for S-type Storage Class Memory. Thus we can consider 3D-Xpoint as the first commercial implementation of the Storage Class Memory concept first described in 2008.^{267,268} Furthermore, in a later presentation, a second, “Performance-focused” form of 3D-Xpoint memory was described as being under active development.²⁷⁷ Compared to the initial “Cost-focused” form of 3D-Xpoint memory, this variant is said to offer even faster latencies and higher endurance (Figure BC2.3). Thus, this second variant of 3D-Xpoint is somewhat similar to M-type SCM as described above, both in terms of its specifications and in terms of potential applications. The only major difference, as observed in Figure BC2.3, is the strong similarity between the expected volatility of Performance-focused 3D-Xpoint and the known volatility of DRAM. In contrast, one of the benefits of M-type SCM was supposed to be its non-volatility. Ideally, retention of data for perhaps 1-3 weeks would permit successful recovery of server data, even after a power outage due to a natural disaster or other major event.

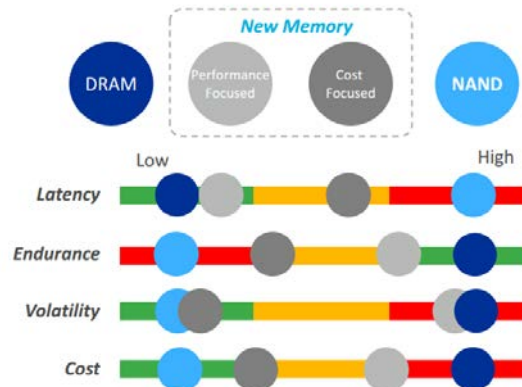


Figure BC2.3 Comparison of Performance of Different Memory Technologies

2.4.2. STORAGE CLASS MEMORY ARCHITECTURES

2.4.2.1. INTRODUCTION

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

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

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

2.4.2.2. CHALLENGES IN MEMORY SYSTEMS

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

24 Emerging Memory Devices

Historically, roughly one-third of the power in a large computer system is consumed in the memory sub-system.²⁷⁸ Some portion of this is refresh power, required by the volatile nature of DRAM. As a result, modern data servers consume considerable power even when operating at low utilization rates. For example, Google²⁷⁹ has reported that servers are typically operating at over 50% of their peak power consumption even at very low utilization rates. The requirement for rapid transition to full operation precludes using a hibernate mode. As a result, a persistent memory that did not require constant refresh would be valuable.

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

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

Total cost of ownership is influenced by cost-to-purchase, cost-to-maintain, and system lifetime. Current cost-to-purchase trends are that Hard Disk Drives (HDD) cost roughly an order of magnitude less per bit than Flash memory, which in turn costs almost an order of magnitude less per bit than DRAM. However, cost-to-purchase is not the only consideration. It is anticipated that S-class SCM will consume considerably less power than HDD (both directly and in terms of required cooling), and will take up considerably less floorspace. One early projection what that by 2020, if the main storage system of a data center is still built solely from HDD, the target performance of 8.4 G-SIO/s could consume as much as 93 MW and require 98,568 square feet of floor space.²⁸⁰ In contrast, the improved performance of emerging memories could supply this performance with only 4 kW and 12 square feet. Given the cost of energy, this differential can easily shift the total cost advantage to emerging memory, away from HDD, even if a cost per bit differential still exists.

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

The emergence of SCM leads to the need to resolve issues beyond the device level, including software organization, wear leveling management, and error management. Because of the inherent speed in SCMs, software can easily limit the system performance. Some of these changes have already been initiated by the advent of SSDs. All types of IO software – from the filesystem, through the operating system and up to applications – had to be redesigned in order to best leverage both SSDs and then SCMs. The number of software interactions was reduced, and disk-centric features were removed. Inefficiencies buried deep within conventional software were accounting for anywhere from 70% to 94% of the total IO latency.²⁸¹ It is likely to be valuable to give application software direct access to the SCM interface, although this can then require additional considerations to protect the SCM device from malicious software. This direct access has not occurred for SSDs, with manufacturers applying numerous undocumented operations between the input data and the raw storage. (One example is the inversion of entire data blocks, based on the number of 0s or 1s to be stored.) However, this approach is not typically used in current operating systems that use some form of File Address Table as an intermediate index mechanism.

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

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

Error management is a broader problem than just wear leveling. While DRAM has traditionally benefited from simple methods such as ECC and EDCs, Flash with its large page sizes and slow accesses can afford more sophisticated algorithms such as LDPC. Unfortunately, SCM will need more error correction than DRAM but will need faster error correction than Flash, especially for M-class SCMs. This is an open area for research. Some possible options include exploring codes that exploit specifics of error

patterns, such as Tensor codes, and the use of in-situ scrub,²⁸² where accumulated errors are periodically eliminated so that one or two-bit error correction can remain sufficient.

2.4.2.3. EMERGING MEMORY ARCHITECTURES FOR M-CLASS SCM

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

The interface is likely to be a word-addressable bus, treating the entire memory system as one flat address space. Since the cost of adapting to new memory interfaces is sizeable, an interface standard that could support multiple generations of M-SCM devices would be highly preferred. Many systems (such as in automobiles) might be deployed for a long time, so any new standard should be backward-compatible. Such a standard should be compatible to DRAM interfaces (though with simpler control commands) and should reuse existing controllers and PHY (physical layers), as well as power supplies, as much as possible. It should be power efficient, e.g. supporting small page sizes, and should support future directions, such as 3D Master/slave configurations. The M-SCM device should indicate when writes have been completed successfully. Finally, an M-SCM standard should support multiple data rates, such as a DDR-like speed for the DRAM and a slower rate for the NVRAM.²⁸³

While wear-leveling in a block-based architecture requires significant overhead to track the number of writes to each block, simple techniques such as “Start-Gap” Wear-Leveling are available for direct-byte-access memories such as PCM (Phase Change Memory).²⁸⁴ In this technique, a pair of registers are used to identify the location of the start point and an empty gap within a region of memory. After some threshold number of write accesses, the gap register is moved through the region, with the start register incrementing each time the gap register passes through the entire region. Additional considerations can be added to defend against detrimental attacks intended to intentionally wear out the memory.

With such techniques, even an M-class SCM that is markedly slower than DRAM can offer improved performance by increasing available capacity and by reducing the occurrence of costly cache misses.²⁸⁵ With proper caching, a carefully-designed M-SCM system could potentially even match DRAM performance despite its lower device latency.²⁸⁶ The presence of a small DRAM cache helps keep the slower speed of the M-class SCM from affecting overall system performance in many common workloads. Even with an endurance of $1e7$ cycles, the system lifetime has been shown to be on the order of 3 years. Techniques for reducing the write traffic back to the SCM device can help improve this by as much as a factor of 3 under realistic workloads.

Direct replacement of DRAM with a slightly slower M-class SCM has also been considered, for the particular example of STT-MRAM.²⁸⁷ Since individual byte-level writes to STT-MRAM consume more power than in DRAM, a direct replacement is not competitive in terms of energy or performance. However, by re-architecting the interaction between the output buffer and the STT-MRAM, unnecessary writes back to the NVM can be eliminated, producing a sizeable energy improvement at almost no loss in performance. However, the use of write buffers means that the device must be able to complete all writes back to non-volatile memory in the event of power loss. Integrating PCM into the mobile environment, together with a redesigned memory management controller, is predicted to deliver a six times improvement in speed and also extends the memory lifetime six times.²⁸⁸

Caches are intended to ensure that frequently-needed data is located near to the processor, in nearby, low-latency memory. In storage architectures, “hot” or frequently-accessed data is identified and then moved to faster tiers of storage. However, as the number of tiers or caches increases, a significant amount of time and energy is being spent moving data. An alternative approach is to completely rethink the hardware/software interface. By organizing the computational system around the data, data is not brought to the processor but instead processing is performed in proximity to the stored data. One such emerging data-centric chip architectures termed “Nanostores”²⁸⁹ was predicted to offer 10–60x improvements in energy efficiency.²⁹⁰

Given the slower than expected deployment of scaled emerging devices for M-class SCM, several projects have employed large amounts of DRAM as a surrogate for and M-class SCM. Though Bresniker et.al. describe a computer enabled by a large non-volatile “Universal memory,”²⁹¹ press reports indicate that early commercial machines will be built with large amounts of DRAM. They point out that an NVM version allows “occasionally-on computing” but could have the downside that new types of bugs might appear as OSs and programs effectively run indefinitely and cannot “re-create their memory state representations each time they start”. New applications and algorithms might emerge as a result of keeping data in perpetuity, or at least for long periods of time.

2.4.2.4. EMERGING MEMORY ARCHITECTURES FOR S-CLASS SCM

S (Storage) type SCMs are intended to replace or supplement hard-disk drives as main storage, much like current Flash-based SSDs, but with even more IOPs (I/O operations per second). Their main advantage will be speed, avoiding the seek time penalty

26 Emerging Memory Devices

of main drives. However, to succeed, their total cost of ownership needs to approach that of HDDs. Research issues include whether the SCM serves as a disk cache or is directly managed, how load leveling is implemented while retaining a sufficiently fast and flexible interface, how error correction is implemented, and identifying the optimal mix of fast-yet-expensive and slow-yet-inexpensive storage technologies. The effective performance of Flash SSD, itself slower than S-SCM, has been strongly affected by interface performance. For instance, the SATA (Serial Advanced Technology Attachment) interface was originally designed for HDD, and was commonly used for early SSD devices despite not being optimized for Flash SSD.²⁹²

One possible introduction of these new memory devices to the market would be as *hybrid* solid-state discs, where the new memory technology complements the traditional Flash memory to boost the SSD performance. Experimental implementations of FeRAM/Flash²⁹³ and PCRAM/Flash²⁹⁴ have been explored. It was shown that the PCRAM/Flash hybrid improves SSD operations by decreasing the energy consumption and increasing the lifetime of Flash memory.

Additional open questions for S-SCM include storage management, interface, and architectural integration, such as whether such a system should be treated like a fast disk drive or as a managed extension of main memory. To date, disk-like systems built using non-volatile memories have had disk-like interfaces, with fixed-sized blocks and a translation layer used to obtain block addresses. However, since the file system also performs a table lookup, some portion of SCM performance is sacrificed. In addition, non-NAND-Flash SCMs have randomly accessible bits and do not need to be organized as fixed-size blocks.²⁹⁵

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

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

Another issue for SCM-based systems will be addressing the asymmetry between read and write in devices such as PCM or other emerging non-volatile memories.²⁹⁶ Such asymmetry can affect the ordering and atomicity of writes and needs to be considered in system or algorithm design.²⁹⁷ Atomicity is critical for operations such as database transactions properties, so that either all of a series of related database operations occur, or none of them occur.

Longer write latencies, in technologies such as PCM, can be compensated by techniques such as data comparison writes,²⁹⁸ partial writes,²⁹⁹ or specialized algorithms/structures that trade writes for reads.^{300,301} (These last set of techniques can also help reduce endurance problems.) Write ordering and atomicity problems can be finessed by hardware primitives. These can either be existing hardware primitives – cache modes (e.g., write-back, write-combining), memory barriers, cache line flush^{302,303,304} – or newly-proposed hardware primitives, such as atomic 8-byte writes and epoch barriers.^{305,306}

Even first-generation PCM chips, although implemented without a DRAM cache, compare favorably with state-of-the-art SSDs implemented with NAND Flash, particularly for small (<2 KB) writes and for reads of all sizes.³⁰⁷ The CPU overhead per input-output operation is also greatly reduced. Another observation for even first-generation PCM chips is that while the average read latency is similar to NAND Flash, the worst-case latency outliers for NAND Flash can be many orders of magnitude slower than the worst-case PCM access. This is particularly important considering that such S-class SCM systems will typically be used to increase system performance by improving the delivery of urgently-needed “hot” data.

Another new software consideration for both S- and M-class SCM is the increased importance of avoiding memory corruption, either through memory leaks, pointer errors, or other issues related to memory allocation and deallocation.³⁰⁸ Since part of the memory system is now non-volatile, such issues are now pervasive and may be difficult to detect and remove without affecting stored user data.

General libraries and programming interfaces – such as NV-heaps,³⁰⁹ Mnemosyne,³¹⁰ NVMalloc,³¹¹ and recovery and durable structures³¹² – have been proposed to expose SCM as a persistent heap and thus ease its adoption. Schemes for filesystem support have been developed to transparently utilize as byte-addressable persistent storage, including Intel’s PMFS,³¹³ BPFS, FRASH,³¹⁴ ConquestFS,³¹⁵ and SCMFS.³¹⁶

Table BC2.7. Likely Desirable Properties of M (Memory) Type and S (Storage) Type Storage Class Memories

3. EMERGING LOGIC AND ALTERNATIVE INFORMATION PROCESSING DEVICES

3.1. TAXONOMY

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

An expanded focus on systems requires also that one evaluate the applicability of these novel devices for new types of systems and architectures, above and beyond the hierarchical influence of devices via small core circuits or functional building blocks. Historically, this evaluation has been conducted in this section in terms of possible applications for alternative information processing devices (*i.e.*, those very unlike CMOS transistors). Beginning with the 2017 Edition of the IRDS, the focus pivoted further toward architectures and applications that depart significantly from the device→circuit→logic gate→functional block→system paradigm.

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

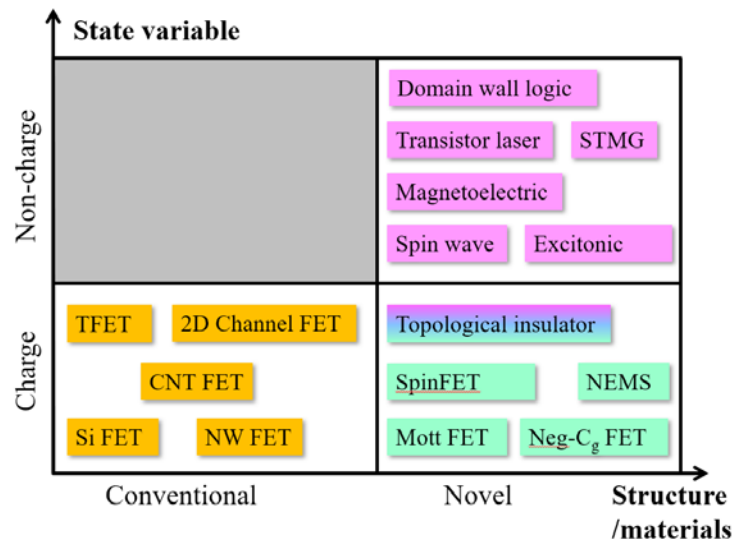


Figure BC3.1 Taxonomy of Options for Emerging Logic Devices

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

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

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

Table BC3.1c *Alternative Information Processing Devices*

3.2. DEVICES FOR CMOS EXTENSION

3.2.1. CARBON NANOTUBE FETS

For many researchers, the search for an ideal semiconductor to be used in FETs succeeded when single-walled carbon nanotubes (CNTs) were first shown to yield promising devices 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.

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,³¹⁷ (2) detailing the impact of contact scalability in CNT FETs,³¹⁸ (3) maintaining performance as the channel length is scaled down to 9 nm without observing short channel effects,³¹⁹ (4) fabricating complementary gate-all-around FETs,³²⁰ (5) fabricating an FET with an intrinsic fT of 153 GHz,³²¹ (6) fabricating CMOS inverters and pass-transistor logic operating at 0.4 V with a non-doped CNT,³²² (7) fabricating a carbon nanotube computer composed of 178 FETs,³²³ (8) progress towards reducing the variability in CNT FETs,³²⁴ (9) understanding origins of hysteresis,³²⁵ and (10) fabricating CNT FETs with ON-current of 0.5 mA/μm.³²⁶

In addition to improvements at the device level, continuous progress has been achieved toward overcoming the dominant material challenges,³²⁷ including the need to achieve purified and sorted semiconducting CNTs with a relatively uniform diameter distribution and then position the CNTs into aligned, closely packed arrays with consistent pitch. With a target purity of 99.9999% semiconducting CNTs and placement density of >125 CNTs/μm (<8 nm pitch), much work still remains. However, it is important to note that progress continues to be steady and without fundamental obstacles barring these goals from being realized. There remains a need for further research toward improving other device-level aspects, including further reduction of contact effects at small contact lengths, demonstrated reduction in variability, improved control of gate dielectric interfaces and properties, and the experimental study of devices and circuits fabricated using the most scaled and relevant device structures and materials. In short, much work remains for CNT FETs, but they have some of the most substantial (and already demonstrated) potential in high-performance, low-voltage, sub-10 nm scaled transistor applications.

3.2.2. NANOWIRE AND NANOSHEET FETS

Nanowire field-effect transistors are structures in which the conventional planar MOSFET channel is replaced with a semiconducting nanowire. Such nanowires have been demonstrated with diameters as small as 0.5 nm.³²⁸ They may be composed of a wide variety of materials, including silicon, germanium, various III-V compound semiconductors (III-As, III-Sb, III-P, and III-N), II-VI materials (CdSe, ZnSe, CdS, ZnS), as well as semiconducting oxides (In₂O₃, ZnO, TiO₂), etc.³²⁹ Importantly, at low diameters, these nanowires exhibit quantum confinement behavior, i.e., 1-D ballistic conduction,³³⁰ and match well with the gate-all-around structure that permits the reduction of short channel effects and other limitations to the scaling of planar MOSFETs.

Important progress has been made in the fabrication of semiconducting nanowires for use as FET channels, for which there are two principal formation methods. The first method is top-down, by which semiconducting channels are formed through lithography and etch. The second approach is bottom-up growth including catalyzed vapor-liquid-solid (VLS) growth^{331,332} and template-assisted selective epitaxy (TASE).³³³ These mechanisms have been used to realize a variety of nanowire geometries, including core-shell and core-multishell heterostructures.^{334,335} Although the top-down methods have been prevailing in the industrial manufacturing, the bottom-up methods have been developed to a point that is worth serious consideration for practical use, given their advantage in heterogeneous integration and convenience for in situ passivation through heterostructures. Nanowire gate-all-around geometry is of interest primarily due to its superb gate electrostatics that can allow further gate-length scaling.

Vertical nanowire transistors have been fabricated in this manner using Si,³³⁶ InAs,^{337,338} and ZnO.³³⁹ Core-shell gate-all-around configurations display excellent gate control and few short channel effects. Circuit and system functionality of nanowire devices has been demonstrated, including vertical InAs MOSFETs with 103 GHz switching speed,³⁴⁰ a down-conversion mixer based on

vertical InAs transistors that showed cut-off frequency of 2 GHz,³⁴¹ and extended, programmable arrays (“tiles”) of non-volatile nanowire-based Flash memory that are used to build circuits such as full-adder, full-subtractor, multiplexer, demultiplexer, clocked D-latch and finite state machines.^{342,343} Planar VLS III-As nanowires perfectly aligned in-plane have enabled multigate HEMTs to enhance high frequency performance.^{344,345}

Despite the promising results that are mostly obtained from academic research labs, bottom-up nanowire transistors still face significant challenges before commercialization is feasible. There is an urgent need to improve device yield and uniformity, as well as position registry if the nanowires are to be transferred to a different substrate.³⁴⁶ On the other hand, the potential of bottom-up nanowire transistors for monolithic core-shell architecture and the inherently relaxed lattice matching requirements for heterogeneous integration have not been completely exploited for low power and high speed applications.³⁴⁷ For top-down fabricated nanowires, with the demonstration of standalone transistors, CMOS integration, and a functional ring oscillator, vertically stacked planar Si nanowires hold enormous potential to succeed FinFETs in sub-5nm technology nodes.^{348,349,350}

3.2.3. 2D MATERIAL CHANNEL FETs

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

Graphene has especially attracted attention as a channel material due to its extremely high mobility since the year of 2004, when a report on monolayer graphene prepared by exfoliation of graphite crystals was published.³⁵¹ The report showed that the field-effect mobility of graphene on SiO₂ was as high as ~10,000 cm²/Vs. It was then predicted that the room-temperature mobility of graphene on SiO₂ would be limited to ~40,000 cm²/Vs due to scattering by surface phonons of the SiO₂ substrate.³⁵² In fact, much higher field effect mobility was obtained using suspended graphene. Values as high as 120,000 cm²/Vs and 1,000,000 cm²/Vs at 240 K and liquid-helium temperature were obtained.^{353,354} Recently, hexagonal boron nitride (hBN), an inert and flat material, has been used as a substrate for graphene-channel transistors, and it has been shown that the field effect mobility of such devices can exceed 100,000 cm²/Vs at room temperature near the charge neutrality point.^{355,356} Furthermore, a device in which graphene was sandwiched between two hBN flakes and edge-contacted with two metal electrodes exhibited mobility even higher than the above.³⁵⁷ The results indicate that hBN can be excellent passivation film for graphene devices.

Graphene can also be obtained on SiC surface by annealing a SiC crystal at high temperatures (often called “epitaxial graphene”). The annealing temperatures range from 1200°C to 2000°C depending on the annealing environment.^{358,359} Chemical vapor deposition of graphene on metal foil or film has also been demonstrated.^{360,361,362,363} Typical growth temperatures are around 1000°C. It has been shown that monolayer graphene is preferentially formed on Cu foil or film,^{360,362,364} while multi-layer graphene can be formed on Ni, Co, and Fe catalyst.^{361,363,365} The quality of epitaxial graphene and CVD graphene is now as good as that of exfoliated graphene.^{366,367}

Efforts have been made to fabricate graphene-channel transistors, where fabrication processes such as doping and contacting to graphene channels have been developed.³⁶⁸ However, since graphene does not have a bandgap, such transistors cannot have an ON/OFF ratio high enough for digital applications. Several approaches have been proposed to open a bandgap of graphene. One of the two major approaches is to apply an electric field perpendicular to AB-stacked bilayer graphene.^{369,370,371,372,373} Experimentally, a transport gap of 130 meV was obtained at an electrical displacement of 2.2 V/nm, providing an ON/OFF ratio of ~100 at room temperature.³⁷² This ON/OFF ratio is probably the largest for this approach so far, which is, however, not high enough for logic applications. In fact, it has been pointed out that a small stacking fault of AB-stacked bilayer graphene can increase the off current,³⁷⁴ which is a serious problem.

A promising approach to form a bandgap in graphene is to make it narrow, that is, to form a graphene nanoribbon (GNR).^{375,376,377,378,379,380,381,382} In fact, simulations using a first-principles many-electron Green’s function approach within the GW approximation have predicted that bandgaps can be as large as ~5 eV depending on their widths for armchair-edged GNRs (AGNRs).^{382,383} Formation of GNRs was first attempted by using electron beam lithography and etching.³⁷⁶ Carrier transport through such a GNR was also investigated. An energy gap of ~200 meV was obtained for a GNR with a width of 15 nm. Devices with multiple GNRs with a sub-10 nm half-pitch were fabricated using patterning with directed self-assembly of block copolymers.³⁸⁴ The transport characteristics of such top-down GNRs were poor, however. This is mainly because the edges of such GNRs were not well controlled, probably with a lot of defects.^{385,386} Recently, however, attempts to form GNRs with controlled edges have been made using bottom-up approaches. In fact, Cai et al. demonstrated the growth of armchair-edged GNRs (AGNRs) from 10,10’-dibromo-9,9’-bianthryl precursors.³⁸⁷ In their approach, precursor molecules are deposited onto a

clean Au(111) surface by vacuum evaporation in ultra-high vacuum. The substrate is then heated to 200°C to remove Br from the precursors and to connect them with each other at the Br-removed points, forming polymers. By further heating the substrate to 400°C, the polymers were cyclodehydrogenated to form AGNRs with a uniform width. The AGNR formed is referred to as 7AGNR, because it has seven dimer lines in the width direction. The band gap of 7AGNR is 3.7-3.8 eV according to the simulations above^{382,383} and agrees with an experimentally obtained bandgap (~2.3 eV) considering image-charge corrections by Au substrate.³⁸⁸ Now several types of GNRs have been formed using similar approaches with different precursors.³⁸⁹ As for AGNRs with a smaller bandgap, 9AGRs with a theoretical bandgap of about 2.2-2.3 eV have been obtained.³⁵³ Formation of 13AGNRs with a theoretical bandgap of about 2.3-2.5 eV was also demonstrated although the GNRs were rather short, typically less than 10 nm in this case. The successful formation of atomically precise AGNRs paves a way for their application to transistor channels.

Performance of GNR-channel transistors have been predicted by numerical simulations.^{390,391,392} It was shown that a transistor with multiple-GNR channels (width: 1.47 nm, pitch: 3.47 nm) with a channel length of 15 nm exhibited an on-current exceeding 1 mA/μm with ON/OFF ratio larger than 10^5 and a subthreshold swing of 64 mV at a drain voltage of 0.1 V⁴⁰. Transistors using 7AGNRs as channels were fabricated and evaluated experimentally. The performance of transistors was, however, poor with a very low on-current and an ON/OFF ratio of 3.6×10^3 at a drain voltage of 1V.³⁹³ The small on-current was attributed to large Schottky barriers at the source and drain contacts caused by the large bandgap of 7AGNR. Use of 9AGNRs and 13AGNRs with smaller bandgaps actually improved the transistor performance. In fact, transistors using 9GNRs as channel exhibited ON/OFF ratios as high as 105 and on-current of 1 μA at a drain voltage of 1V, although the number of GNRs in each transistor is unclear.³⁹⁴ The performance is not yet as good as a counterpart using carbon nanotubes (CNTs)³⁹⁵ but expected to improve further by, for example, covering GNRs with hBN and realizing better contacts between GNRs and source/drain electrodes.

New principle devices using GNRs have also been proposed. One is a tunneling field-effect transistor (TFET). A higher on-current than that of Si TFET has been predicted.³⁹⁶ Use of strained graphene as a channel can also realize tunneling-like transport, according to simulations.³⁹⁷ A Klein-tunneling-based device has also been proposed.³⁹⁸ Graphene can offer possibilities for employing novel switching mechanism for future electronics.

Transition metal dichalcogenides (TMDCs) are another 2D material attracting attention. TMDCs have the chemical formula of MX_2 , where M is a transition metal element and X is a chalcogen. They can be metallic, half-metallic, semiconducting, or superconducting depending on their compositions. Molybdenum disulfide, MoS_2 , is probably the most popular semiconducting TMDC, whose single layer was isolated for electrical measurements in 2005.³⁹⁹ Electrical properties of semiconducting TMDCs depend on the number of layers due to quantum confinement effects and changes in symmetry. For example, single-layer MoS_2 has a direct band gap of 1.9 eV, while bulk MoS_2 has an indirect bandgap of 1.2 eV.⁴⁰⁰ The bandgaps also vary with the compositions. For example, first principles calculations performed using the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional show that MoS_2 , MoSe_2 , MoTe_2 , WS_2 , WSe_2 , and WTe_2 have bandgaps of 2.02 eV, 1.72 eV, 1.28 eV, 1.98 eV, 1.63 eV, and 1.03 eV, respectively.⁴⁰¹

Transistors with TMDCs as a channel material have been demonstrated. Kis et al. fabricated a top-gate MoS_2 -channel transistor, demonstrating a large ON/OFF ratio ($\sim 10^8$) and good subthreshold swing (74 mV/decade).⁴⁰² Exfoliated single-layer MoS_2 was used in this experiment. However, field-effect mobility was relatively low (60-70 cm^2/Vs).⁴⁰³ In fact, transport in single-layer TMDC is severely affected by the environment, often degrading its electrical property. It has actually been demonstrated that mobility of TMDC increases with thickness.^{404,405} This is because influences of charged impurities in substrate decrease as the thickness increases. As for MoS_2 , field-effect mobility as high as 480 cm^2/Vs was obtained for a 47-nm thick MoS_2 channel.⁴⁰⁶ MoS_2 is naturally electron-doped, which is possibly caused by chalcogen vacancies, while WSe_2 is hole-doped.⁴⁰⁷ It is actually still difficult to control doping level of TMDCs, which is a major challenge to realize TMDC-based electronics. Incidentally, Desai et al. fabricated MoS_2 -channel transistor with 1-nm gate lengths by using a CNT as a gate.⁴⁰⁸ The ON/OFF ratio and subthreshold swing were $\sim 10^6$ and 65 mV/decade, respectively. This is probably the shortest channel transistor ever made.

New principles devices using TMDCs have also been demonstrated. Sarkar et al. fabricated a tunneling FET (TFET) by placing n-type MoS_2 on top of p-type Ge, forming a p-n junction.⁴⁰⁹ The TFET was gated by using a solid polymer electrolyte. The subthreshold swing had an average of 31.1mV/decade for four decades of drain current at room temperature. Britnell et al. prepared stacking structures consisting of graphene/ MoS_2 /graphene and graphene/hBN/graphene, demonstrating a switching device based on tunneling phenomena.⁴¹⁰ Although the tunneling devices introduced here can offer steep-slope switching properties, on-current is relatively low, which is an important issue to address for real-world applications. Fabrication of a lateral heterojunction of TMDCs, such as MoSe_2 and WSe_2 , have also been demonstrated,^{411,412} and a p-n junction by such a heterojunction has been made.^{413,414} Valleytronics based on carriers located in two inequivalent valleys in the wave number space also attract attention.^{415,416} New principles devices using TMDCs, as introduced here, may have good opportunities in future electronics.

Growth technology of TMDCs is also in progress. In fact, the synthesis of TMDC film dates back to the 1980's, when the growth was performed with van der Waals epitaxy.⁴¹⁷ More recently, Lee et al. demonstrated CVD growth of MoS₂ using MoO₃ and S powder as precursors.^{418,419} The growth temperature was 650°C. Single-crystal monolayer MoS₂ flakes were successfully obtained. This method was further improved and single-crystal MoS₂ flakes as large as 120 μm in lateral size were obtained.⁴²⁰ There have been a quite a few reports using similar methods for synthesizing TMDCs, including MoSe₂,⁴²¹ WS₂,^{418,422} and WSe₂.⁴²³ 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 MoS₂ by MOCVD using Mo(CO)₆ and (C₂H₅)₂S as precursors.⁴²⁴ The electron mobility of MoS₂ in this case was 30 cm²/Vs at room temperature. In general, mobility of CVD MoS₂ is lower than that of exfoliated MoS₂, which is still an important issue to address. Furthermore, MoS₂ deposition by using sputtering is also in progress.^{425,426,427} The sputtering method is scalable, but it is still difficult to obtain film with a quality as high as that by MOCVD.

3.2.4. 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.^{428,429,430} 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.^{431,432,433,434} The thickness of the tunneling barrier is controlled by changing the electric field in the tunneling junction. Alternatively, it is possible use energy filtering or density of states switching. If the conduction and valence band do not overlap at the tunneling junction, no current can flow. Once they do overlap, current can flow. Simulations have predicted arbitrarily steep subthreshold swings when relying on density of states switching as the current is abruptly cutoff when the conduction band and valence band no longer overlap.⁴³⁰ If phonons or short channel lengths are accounted for, simulated subthreshold swings on the order of 20–30 mV/decade are typical.⁴³⁵ 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.⁴³⁶ The best experimental results to date have relied on a combination of thermal switching and density of states switching.⁴³⁷ So far, the experimental results are far worse than simulated device characteristics. The review by Lu and Seabaugh⁴²⁸ 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,^{433,438,439,440,441,442} strained SiGe,⁴⁴³ Si/Ge,⁴⁴⁴ and strained Ge.⁴⁴⁵ Nanowire III-V TFETs have shown even steeper swings. An InP/GaAs heterojunction⁴⁴⁶ has shown 30 mV/decade at 1 pA/μm. The steepest result ever reported is in a Si/InAs heterojunction⁴⁴⁷ 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.⁴⁴⁸ Recently, a promising InAs/GaAsSb/GaSb nanowire heterostructure TFET with a 48 mV/decade SS at 67 nA/μm and an I₆₀ (current at 60 mV/decade) of 0.31 μA/μm was demonstrated.⁴³⁷

Researchers attempting to achieve higher on-current TFETs have traditionally relied on reducing the effective mass by using III-V's and reducing effective bandgap by using a heterostructure. While this has increased the current, the subthreshold swings and off currents have gotten worse. The increase in off-state current and subthreshold swing needs to be decoupled from the increase in on-state current. Unfortunately, this is a fundamental tradeoff when modulating the thickness of the tunneling barrier:⁴²⁹ 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.⁴⁴⁹ 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.⁴⁵⁰ However, the electrically measured joint DOS in diodes has generally indicated a steepness >90 mV/decade.⁴⁵¹ 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.⁴³⁶

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)^{452,453,454} and spontaneous and Auger generation.⁴⁵⁵ Calculations based upon these mechanisms show that these significantly degrade the subthreshold swing and increase the leakage currents but do not prevent TFETs from achieving sub-60 mV/decade subthreshold swing. Material defects and gate interface traps make these effects worse and result in worse band edges.

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

3.3. BEYOND-CMOS DEVICES: CHARGE-BASED

3.3.1. SPIN FET AND SPIN MOSFET TRANSISTORS

Spin-transistors are classified as “non-conventional charge-based extended CMOS devices,”⁴⁵⁶ and can be further divided into two categories: the spin-FETs proposed by Datta and Das⁴⁵⁷ and spin-MOSFETs proposed by Sugahara and Tanaka.⁴⁵⁸ 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.^{456,459} 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.^{459,460,461,462,463,464,465}

The full operation suggested for spin-FETs^{459,466} and spin-MOSFETs^{459,467,468,469,470} 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.^{471,472,473} Lots of theories^{474,475,476,477} 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.

Recently, 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.^{478,479,480,481} Also, by using a back-gated device structure with FM/MgO tunnel-barrier contacts,^{468,469,470} 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. In particular, electrical spin injection, transport, and detection in SCs without using insulator tunnel barriers have been demonstrated in lateral spin-valve devices with Heusler alloy/SC Schottky-tunnel-barrier contacts.^{482,483,484,485,486} To reduce the value of RA at the source and drain structures in spin-MOSFETs, the delta-doping of dopant impurities near the Heusler alloy/Ge Schottky-tunnel-barrier⁴⁸⁷ has been demonstrated, leading to the room-temperature spin transport⁴⁸⁸ including local magnetoresistance effect in Ge-based lateral devices with RA values of less than $0.2 \text{ k}\Omega\mu\text{m}^2$.⁴⁸⁹ Unfortunately, the MR ratio at room temperature in the Ge-based lateral devices is still much low ($< 0.001\%$) because of an imperfection of the Heusler alloy/Ge quality.⁴⁸⁹ For enhancing the MR ratio in spin-MOSFET structures, it is essential to optimize the formation of high-quality Heusler alloy/SC Schottky-tunnel contacts and to reduce the channel length between source and drain contacts.

Alternative approaches for realizing spin-MOSFETs have been proposed.^{459,464,467,490,491} Pseudo-spin-MOSFETs are circuits that reproduce the functions of spin-MOSFETs using an ordinary MOSFET and a magnetic tunnel junction (MTJ) which is connected to the MOSFET in a negative feedback configuration. Although pseudo-spin-MOSFETs offer the same functionality as spin transistors, such as the ability to drive variable current, pseudo-spin-MOSFETs have larger resistance than spin-FETs or spin-MOSFETs.

For spin manipulation in SCs, channel materials with a strong spin-orbit interaction, such as InGaAs, InAs and InSb, are required⁴⁵⁹ in order to sufficiently induce the Rashba spin-orbit interaction by an electric gate voltage. Using InAs⁴⁹² and InGaAs⁴⁹³ 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.

3.3.2. NEGATIVE GATE CAPACITANCE FET

Salahuddin and Datta originally proposed⁴⁹⁴ 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 device⁴⁹⁵ 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)/SiO₂ organic ferroelectric gate stack, was reported⁴⁹⁶ in 2008, and subsequently⁴⁹⁷ 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.⁴⁹⁸ Essentially, it was demonstrated that in a bi-layer of dielectric Strontium Titanate (SrTiO₃: STO) and Lead Zirconate Titanate (Pb_xZr_{1-x}TiO₃: 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.⁴⁹⁹ 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 HfO₂ by constituents such as Zr, Al or Si.⁵⁰⁰ Using this doped Hf based ALD ferroelectric, a number of experiments have demonstrated the negative capacitance effect.^{501,502,503} For example, by using HfZrO₂ as a gate dielectric, sub-60 mV/decade STS was demonstrated⁵⁰³ 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.⁵⁰⁴

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**.⁵⁰⁵ 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 bipolar 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*).⁵⁰⁶ Additional advantages of mechanical devices include robust operation across a wide temperature range, down to cryogenic temperatures,⁵⁰⁷ and immunity to ionizing radiation. NEM switches can be monolithically integrated with CMOS circuitry by a modular post-CMOS fabrication process with relatively low thermal budget. Since state-of-the-art CMOS technology incorporates air-gapped interconnects,⁵⁰⁸ NEM switches can be implemented using multiple interconnect layers formed in the back-end-of-line (BEOL) process.⁵⁰⁹ Potential applications for hybrid NEMS-CMOS technology include CMOS power gating,⁵¹⁰ configuration of FPGAs,⁵¹¹ non-volatile back-up storage of information in SRAM and CAM cells,⁵⁰⁹ and energy-efficient, fast and reconfigurable look-up tables.⁵¹²

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.⁵¹³ 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.⁵¹⁴ 3-terminal devices and corresponding logic gates also were demonstrated.⁵¹⁵ Piezoelectric materials have been incorporated in NEM devices to enable sub-1V switching⁵¹⁶.

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.⁵⁰⁶ 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.⁵¹⁷ 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.⁵¹⁸ 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.^{519,520} 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)⁵¹⁶. A variety of relay-based computational and memory building blocks have been experimentally demonstrated to date.^{521,522}

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.⁵²³ 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.⁵²⁴ Recently hybrid CMOS-NEMS circuits have been demonstrated, with non-volatile NEM switches operating at the same V_{DD} as for the CMOS circuitry.⁵²⁵ 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 functionality⁵²⁶.

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.⁵²⁷ 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 polymers during the course of device operation. Alternative contacting electrode materials such as Ru/RuO₂ 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.⁵²⁸ 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.⁵²⁹ Most recently, operation of NEM switches and integrated circuits at temperatures down to 4K were demonstrated for the first time.⁵³⁰ 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.

In conclusion, the negligible OFF-state current and ultra-low-voltage operation capability of NEM switches make them compelling 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.

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.^{531,532} 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.^{533,534} 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.⁵³² 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).⁵³⁵ The phase transition time constant in VO_2 materials is in sub-picosecond range determined by optical pump-probe methods.⁵³⁶ 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 .⁵³⁷ VO_2 Mott channels have been experimentally studied with thin film devices and the field effect has been demonstrated in preliminary device structures.^{538, 539, 540} Recently, prototypes of VO_2 transistors using ionic liquid gating have shown larger ON/OFF ratio at room temperature than with solid gate dielectrics like hafnia.^{541, 542} The conductance modulation happens at a slow speed, however, due to the large charging time constants.⁵⁴³ 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.^{543, 544, 545} 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 transistors⁵⁴⁶ and circuits⁵⁴⁷ 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 devices⁵⁴⁸ and steep-slope transistors.⁵⁴⁹ As a Mott device with purely electrostatic modulation in the solid-state base,⁵⁵⁰ the VO_2 -FET with a high-k oxide/organic hybrid dielectric gate has been proposed.^{551, 552} 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.⁵⁵³ Nanofabrication engineering is considered to possess the possibility to enhance the performance of Mott FET. 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.⁵⁵⁴

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.⁵³³ Theoretical studies on the channel/dielectric interfacial electronic band structure are needed for the modeling of subthreshold behaviors of Mott FETs. Understanding the electronic transition mechanisms while de-coupling from structural Peierls (lattice) distortions is also of interest and important in the context of energy dissipation for switching.

While the electric field-induced transitions are typically explored with Mott FET, nanoscale thermal switches with Mott materials could also be of substantial interest. Recent simulation studies of “ON and “OFF” times for nanoscale two-terminal VO_2 switches indicate the possibility of sub-ns switching speeds in ultra-thin device elements in the vicinity of room temperature.^{555, 556} Such devices could also be of interest to Mott memory.⁵⁵⁷ One can, in a broader sense, visualize such correlated electron systems as ‘threshold materials’ wherein the conducting state can be rapidly switched by a slight external perturbation, and hence lead to applications in electron devices. Electronically driven transitions in perovskite-structured oxides such as rare-earth nickelates^{558, 559} or cobaltates⁵⁶⁰ 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.^{561, 562, 563, 564, 565} SmNiO_3 , with its metal-insulator transition temperature near 130°C 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.⁵⁶⁶ 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.⁵⁶⁷ 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 which 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.^{568,569,570,571} 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.

Systematic searches of materials databases have found that topological materials are commonplace, representing a significant fraction of all known materials.^{572,573,574} Two-dimensional topological insulators have been realized with very large bandgaps of 360 meV (Na_3Bi ⁵⁷⁵) and 800 meV (bismuthine⁵⁷⁶), 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.^{577,578,579,580,581,582,583,584,585,586} 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,⁵⁷⁷ staggered sublattice potential,^{577,579,585} inversion symmetry breaking,⁵⁸⁰ or Stark effect.^{581,582,583,584,575} Electric field effect switching has been proposed in a number of materials including graphene,⁵⁷⁷ monolayer transition metal dichalcogenides in 1T’ phase,⁵⁸¹ SnTe and $\text{Pb}_{1-x}\text{Sn}_x\text{Se}(\text{Te})$,⁵⁸⁰ topological semimetals such as Cd_2As_3 and Na_3Bi ,^{582,575} and phosphorene.⁵⁸³ Electric-field switching of the topological state has been demonstrated in at least one material Na_3Bi .⁵⁷⁵ Other transistor proposals have focused on electric-field switching of tunneling between topological edges,⁵⁸⁷ or using strong disorder to produce an off state in a bulk conducting topological insulator.⁵⁸⁶ Topology may also be controlled by magnetic fields,^{588,589} strain,⁵⁹⁰ temperature,⁵⁹¹ or time-dependent electric fields,⁵⁹² hence topological devices have the prospect of adding additional new functionality for “More than Moore” devices.

The deep connection between topological insulators and spin-orbit coupling 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 polarized spin current,⁵⁷⁷ 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,⁵⁹³ opening new possibilities for spin transistors and non-volatile random-access memory. 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 heterostructures⁵⁹⁴ 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.^{595,596,597}

Numerous fundamental research challenges remain before topological transistors may become useful. Topological switching via electric field effect is a fundamentally different mechanism of operation compared to the traditional MOSFET. At present it is unclear what limits the subthreshold swing and therefore the operating voltage of a topological transistor, and how this depends on the specific mechanism of electric field switching. Detailed models and device simulations for topological transistors are needed, as well as experiments on electric-field switching of topological materials in realistic device geometries. It is likely that new topological materials are needed that optimize large bandgaps and sensitive electric field-driven switching, as well as are amenable to device fabrication and processing. Realistic theoretical models will be needed to guide materials discovery and device design.

3.4. BEYOND-CMOS DEVICES: ALTERNATIVE INFORMATION PROCESSING

3.4.1. SPIN WAVE DEVICE

Spin Wave Device (SWD) is a type of magnetic logic devices exploiting collective spin oscillation (spin waves) for information transmission and processing.^{598,599,600,601,602} 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,^{603,604} (ii) magnetic waveguides - spin wave buses for spin wave signal propagation between the magneto-electric cells,⁶⁰⁵ (iii) magnetic junctions to couple two or several waveguides,⁶⁰⁶ (iv) spin wave amplifiers,⁶⁰⁷ (v) phase shifters to control the phase of the propagating spin waves,⁶⁰⁸ and (vi) spin wave phase error correctors.⁶⁰⁹ 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,⁵⁹⁸ SWD concept has evolved in different ways encompassing volatile⁶¹⁰ and non-volatile,⁶¹¹ Boolean^{611,612} and non-Boolean,⁶¹³ single-frequency and multi-frequency circuits.⁶¹⁴ The primary expected advantage of SWD over Si CMOS are the following: (i) the ability to utilize phase in addition to amplitude for building logic devices with a fewer number of elements than required for transistor-based approach; (ii) power consumption minimization by exploiting the intrinsically low energy (μeV) of spin waves in ferromagnets as well as built-in non-volatile magnetic memory and magnetic reconfigurability⁶¹⁵, 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.⁶¹⁶ It is based on ferromagnetic $\text{Ni}_{81}\text{Fe}_{19}$ structure, operates within 1-3 GHz frequency range, and exhibits signal-to-noise ratio of approximately 10 at room temperature.⁶¹⁶ The internal delay of SWD is defined by the spin wave group velocity (e.g. 3×10^6 cm/s in $\text{Ni}_{81}\text{Fe}_{19}$ 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).⁶¹⁷ The later translates in ultra-low power consumption (e.g. 1aJ per multiferroic switching). Recent experimental demonstration⁶⁰⁴ of parametric spin wave excitation by voltage-controlled magnetic anisotropy (VCMA) is another promising route towards energy-efficient generation of spin waves.

Recently, a new type of SWD magnonic holographic memory (MHM), was proposed.⁶¹³ 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.⁶¹⁸ 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.⁶¹⁹ 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 $1\text{Tb}/\text{cm}^2$) 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.^{2,613}

There are several important milestones crucial for further SWD development: (i) nanomagnet switching by spin waves;⁶²⁰ (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.⁶¹¹ The success of the SWD will also depend on the ability to restore/amplify spin waves (e.g. by multiferroic elements or spin torque oscillators). Another recent direction of research is antiferromagnetic SWD⁶²¹ that potentially offer a thousand-time increase in operation speed due to THz scale frequencies of antiferromagnetic spin waves.

3.4.2. 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,⁶²² diodes,⁶²³ and CCD⁶²⁴ was demonstrated. IX condensate and coherent IX currents⁶²⁵ and, recently, long range coherent IX spin currents⁶²⁶ were observed at low temperatures. New heterostructures based on single-atomic-layers of transition-metal dichalcogenide (TMD) for room-temperature IX circuits were proposed.⁶²⁷ Recent progress includes the realization of IXs at room temperature in TMD heterostructures,⁶²⁸ discovery of IX coherent spin currents in GaAs heterostructures⁶²⁹, development of first split-gate device for IXs, the basic mesoscopic device⁶³⁰, development of advanced platform for high-mobility excitonic devices⁶³¹, development of TMD heterostructures with small IX linewidth and finding charged IXs, indirect trions. At present, the studies are focused on the development of basic concepts of excitonic devices at low temperatures using GaAs heterostructures and development of room-temperature excitonic devices using TMD heterostructures.

3.4.3. TRANSISTOR LASER

The Light-Emitting Transistor (LET)⁶³² and Transistor Laser (TL)^{633,634} 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).⁶³⁵ 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.⁶³⁶ 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.^{634,637} When used to realize conventional logic architectures, for example NOR gates,⁶³⁸ 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 ($200\ \mu\text{m} \times 1\ \mu\text{m}$) have been modulated to 22 Gbps and have shown a measured bandwidth of 10.4 GHz.⁶³⁹ Relative intensity noise (RIN) as low as -151 dB/Hz has also been demonstrated, showing an approximately 28 dB improvement over diode lasers.⁶⁴⁰ 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.⁶⁴¹ 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.⁶⁴²

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.⁶⁴³ 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.⁶⁴⁴

3.4.4. MAGNETOELECTRIC LOGIC

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.^{645,646,647,648,649,650} There are also some benchmarking efforts,^{647,648,649,650} where there has been an effort to compare the most competitive magneto-electric devices with CMOS.^{647,648,649,650} The result is that it is now increasingly clear that magneto-electric field effect transistors are more likely to be competitive or surpass CMOS^{647,648,649,650} than the earliest magneto-electric device concepts were based on a magnetic tunnel junction structure.^{651,652} The disadvantage, with the magneto-electric magnetic tunnel junction device structure, is that while much faster than many spintronics devices, there are long delay times in device operation, due to the slow speed of switching a ferromagnet,⁶⁴⁷ and the predicted fidelity is likely low, making it difficult to cascade from one device to the next.

Other magneto-electric devices, like the composite-input magnetoelectric-based logic technology (CoMET)⁶⁵³ and a similar magneto-electric device structure, but using spin-orbit coupling, have also been proposed.⁶⁵⁴ The concern is that both these device concepts will have long delay times, due to the slow switching speed of the ferromagnetic layer, and in the case of the CoMET device, the additional complications of the slow speed of ferromagnetic domain wall propagation. are limited by the switching speed of the ferroelectric and domain wall motion.^{653,654} The basis of these devices is an input switches a ferroelectric material, in contact with a ferromagnet with in-plane magnetic anisotropy placed on top of an intra-gate ferromagnet interconnect with perpendicular magnetic anisotropy. The input voltage nucleates a domain wall while a current is used to drive the domain wall to the output end of the device. Also using spin orbit coupling, but now explicitly also using a magneto-electric layer for electrical control of exchange bias of a laterally scaled spin valve is the nonvolatile magneto-electric spin-orbit (MESO) logic,⁶⁵⁵ but the delay time is limited again by the switching speed of the ferromagnetic layer, although 50 ps switching speeds as short as have been estimated.⁶⁵⁵

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),^{647,648,649,650} greater temperature stability (operational to 400 K or more⁶⁵⁶), 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.^{645,646,647,648,649,650}

Obviously, the semiconductor channel will only work if the very thin, so the boundary polarization of chromia^{657,658} effectively polarizes the semiconductor channel.^{645,646,647} The 2D semiconductors of the trichalcogenide class of quasi one-dimensional semiconductors, such as TiS₃, HfSe₃,⁶⁵⁹ as well as InP, have the potential to be scaled to transistor widths below 10 nm suggesting there is a plausible route forward.⁶⁴⁷

The anti-ferromagnet spin-orbit read (AFSOR) logic device structure has interesting advantages: the potential for high and sharp voltage ‘turn-on’; inherent non-volatility of magnetic state variables; absence of switching currents; large on/off ratios; and multistate logic and memory applications. The design will provide reliable room-temperature operation with large on/off ratios ($>10^7$) well beyond what can be achieved using magnetic tunnel junctions.^{646,647,660} Again, the core idea is the use of the boundary polarization of the magneto-electric to spin polarize or partly spin polarize a very thin semiconductor, ideally a 2D material, with very large spin orbit coupling.

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

The silicon CMOS majority gate (left) requires 13 components. The ME-spinFET majority gate requires, including clocking, of 6 components. This represents an area improvement of over 50%, assuming similar size transistors.⁶⁴⁷ 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.⁶⁴⁷

A majority gate, in the form of a single transistor like device, has also been proposed and modeled,⁶⁴⁸ but depends on the device scaled to dimensions smaller than the natural antiferromagnetic domain size of about 1 to 3 μm .⁶⁵⁸

There is a variant where inversion symmetry is not as strictly broken, that leads to a nonvolatile spintronics version of multiplexer logic (MUX).^{646,647,660} 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.^{646,647} There is a source-drain voltage and current difference, between the two FM source contacts, due to the spin-Hall effect when spin-orbit coupling is present. This output voltage can be modulated by the gate or gates, which influences the spin-orbit interaction in the channel especially when it is both top and bottom gated especially. The spin-Hall voltage in the device can be increased by using different FMs in the source and drain. To increase the spin fidelity of current injection at the source end, one could add a suitable tunnel junction layer (basically a 1nm oxide layer) between the magnetic source and the 2D semiconductor channel. This latter modification would result in diminished source-drain currents though. Again, there is a reduction 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.⁶⁴⁷

Magneto-electric coupling can be used to excite parametric resonance of magnetization by an electric field.⁶⁶¹ This has been considered for the development of spin wave devices based on voltage controlled magnetic anisotropy in ferromagnetic nanowires.⁶⁶² This in turn, in effect, becomes a spin wave field effect transistor. The threshold voltage for parametric excitation 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 remain absent. That said, the magneto-electric transistor has far fewer challenges to implementation than the magneto-electric magnetic tunnel junction, so, not surprisingly, there is a shifting of development effort toward those and related devices for both memory and logic.⁶⁴⁷ New device concepts are being considered as well. These device concepts have advantages like reducing the number of leakage pathways to just one and require only one clock cycle. The actual demonstration of such devices appears almost "in hand" suggesting that experimental evidence of promise is not that far away.

3.4.5. 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 ferromagnetic patterned wire,^{663,664,665} 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.

The domain wall track is composed of heavy metal/ferromagnet/oxide thin films patterned into a wire shape. 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 ferromagnetic track.

The ends of the ferromagnetic track can be exchange-biased in opposing directions using antiferromagnets, and/or an additional electrode can be used to ensure a single domain wall is created 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).

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.^{666,667,668,669,670} 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,⁶⁶⁴ and initial benchmarking was performed up to a full adder simulation.^{671,672,673} A SPICE model has been developed to enable larger scale circuit simulations.⁶⁷⁴ Single device operation and three-device circuit operation has been shown in experimental prototypes,^{665,675} 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.⁶⁶⁸

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 devices, it comes with increased extrinsic DW pinning effects and thermal noise.⁶⁷⁶ 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. Recent 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.⁶⁷⁷

The device has recently been extended by many in the community to non-Boolean logic applications in neuromorphic circuits,^{678,679,680,681,682} including spike-timing-dependent-plasticity synapses^{683,684} and leaky, integrate, and fire neurons.^{685,686}

Major challenges exist in understanding and improving the device-to-device and cycle-to-cycle variation of the devices, which arise 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.⁶⁷⁷ Better understanding of experimental viability of the technology is needed, given TMR variability and constraints and scaling-induced variability and errors.⁶⁸⁷ 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.

3.4.6. 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.⁶⁸⁸ 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),^{689,690,691,692} spin-wave majority gate (SWMG)^{693,694,695} and domain-wall or spin-torque majority gate (STMG).^{688,696,697,698}

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).^{688,696,697,698}

Much of the work on STMGs have been in micromagnetic simulations. 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.^{696,697} 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.⁶⁹⁸ The STMG operates at a smaller drive current compared to its CMOS counterpart, but with slower switching.⁶⁹⁸

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.⁶⁹⁹ Hence, implementation with PMA are more stable and energy efficient.⁶⁹⁶

Other types of implementation include a five-input majority gate with lateral spin valve,⁷⁰⁰ and a fabricated 3-input majority gate in which the input and output nanomagnetic logic devices are field-coupled together rather than monolithically integrated.⁷⁰¹

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.

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

Analog computing (§4.2) – This broad category of non-digital computing paradigms uses physical phenomena to do the computing and includes the following:

- **Crossbar Based Computing Architectures** (§4.2.1): Several different computational kernels can be built on analog crossbars or memory arrays including matrix-vector multiplication (§4.2.1.1), outer product updates (§4.2.1.2), field programmable analog arrays (§4.2.1.3), crossbar based matrix solvers (§4.2.1.4), and ternary content addressable memory (§4.2.1.5). These crossbars perform low-precision matrix operations in parallel, by processing analog data directly at each memory element.
- **Neuro-Inspired Computing** (§4.2.2): These computing architectures take direct inspiration from the brain to develop more efficient systems. Local learning rules (§4.2.2.1) are a neuro-inspired method of training neural networks that keeps all communication local. Hyperdimensional computing (§4.2.2.2) is a cognitive computing model based on the high-dimensional properties of neural circuits in the brain. Spiking neural networks (§4.2.2.3) minimize communication energy by using sparse spike-based communications.

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 BC4.1. Conventional vs. Alternative Computing Paradigms

Note: 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.

- **Computing with Dynamical Systems** (§4.2.3): Analog dynamical systems can be used to solve a variety of problems. Optimization problems can be solved using simulated annealing (§4.2.3.1) or coupled-oscillator-based approaches (§4.2.3.2). Dynamical systems can be used to encode associative memories (§4.2.3.3) or to solve differential equations (§4.2.3.4). Chaotic logic can theoretically enable sub- kT computing (§4.2.3.5).
- **Analog Memory and Compute Devices** (§4.2.4): There are many different types of analog devices that form the building blocks of the computational kernels above. These devices include ReRAM, phase change memory, ion insertion redox transistors, floating gate transistors, capacitive or charge based analog devices, single flux quantum devices, photonic devices, magnetic devices, and more.

Probabilistic/stochastic circuits (§4.3) – Devices and circuits that produce truly nondeterministic or random outputs at the hardware level may be useful for accelerating probabilistic algorithms such as Monte Carlo or simulated annealing, for generating secure cryptographic keys, and for other applications.

Reversible (adiabatic and/or ballistic) computing (§4.4) – Computing paradigms that approach *logical and physical reversibility* offer the potential to greatly exceed the energy efficiency of all other approaches for general-purpose digital computation. While devices for reversible computing may perform fairly conventional functions (such as switching or oscillating), they should be optimized to utilize 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.^{702,703} 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 in order 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.^{704,705}

Quantum computing – Quantum computing⁷⁰⁶ offers the potential to carry out exponentially more efficient algorithms for a variety of specialized problem classes.⁷⁰⁷ 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.⁷⁰⁸ Since the 2018 edition, IRDS has been

beginning to address quantum computing in a new chapter titled *Cryogenic Electronics & Quantum Information Processing* (CEQIP). We will not address quantum computing further in the present chapter.

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.

4.2. 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.⁷⁰⁹ Additionally, new architectures can be used for ultra-low-power co-processors for conventional CMOS designs. A key challenge is that analog signals are typically low-precision, with energy and latencies increasing exponentially with higher bit precision. Fortunately, many machine learning and other applications are being developed that can tolerate such lower precision computation.

In this subsection, we review a number of recent examples of analog computational technologies, starting at the architectural level with (in §4.2.1) some currently popular neural-inspired architectures that also have broader applications in linear algebra. Then §4.2.2 broadens the scope to other neural-inspired computing approaches, and §4.2.3 looks at an even broader variety of approaches to “physical computing” with analog dynamical systems. Finally, §4.2.4 reviews device technologies that have some utility in the context of one more of the various analog architectures.

4.2.1. CROSSBAR-BASED COMPUTING ARCHITECTURES

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.⁷¹⁰ Because a digital memory must individually access each memory cell and move the data to a separate computation unit, digital systems consume more energy and incur longer latencies. Computing on larger crossbars/matrices allows for any analog overhead to be averaged out over many matrix elements. Any two- or three-terminal device that features a modifiable internal physical state variable (which might be, for example, a variable resistance, a variable capacitance, a stored charge, or a stored magnetic field) that modifies the device’s behavior can be used as a building block for analog operations. Several different types of crossbar architectures are summarized in the following sections.

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

MVM and VMM are key computational kernels underlying many different algorithms. There are several approaches to accelerating these kernels. Any programmable resistor such as a two-terminal resistive memory or a three-terminal floating gate cell can be used. Alternatively, a capacitive MVM can be designed by adding charge from capacitive memory elements.⁷¹¹

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

Analog VMMs have been used for experimental demonstration of threshold logic,⁷¹² compressed sensing initial filtering,⁷¹³ robotic navigation and control,⁷¹⁴ adaptive filtering,⁷¹⁵ Fourier transforms⁷¹⁶ and more. Additionally, Analog VMM techniques have been used for ultra-low power classification and neural networks.⁷¹⁷

4.2.1.1.1. 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.⁷¹⁸ 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.

44 Emerging Device-Architecture Interaction

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

If high-resistance memory elements ($R_{on} = 100 \text{ M}\Omega$) 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.⁷¹⁹ The latency for a 1024×1024 MVM will only be 384 ns or 11 ns for 8-bit or 2-bit A/Ds respectively. This is over $100\times$ better than an optimized SRAM-based accelerator, which would require 2,700 fJ and 4,000 ns for 8 bits. The area per weight for the 8-bit A/D ReRAM accelerator is $0.05 \mu\text{m}^2$, $16\times$ better than the $0.8 \mu\text{m}^2$ needed for an SRAM accelerator. The energy and latency are dominated by the A/D circuitry and not by the crossbar itself, with the A/D converters and digital circuitry occupying $10\times$ the area of the ReRAM array itself.

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

A key design choice is the bit precision of the inputs and outputs to the crossbar. The fewer 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.⁷²⁰ Encoding inputs in variable length pulses requires longer reads and an integrator to sum the resulting current. Digital encoding applies each bit of the input sequentially and then combines the result digitally. For digital encoding, the usefulness of 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.⁷²¹

The precision with which each resistor needs to be programmed depends on the application. For neural network inference, the algorithm can be robust to read noise of up to 5% of the weight range or more.⁷²² To extend the precision of computation beyond the limits of reliable programming, a technique called bit slicing can be used.⁷²³ 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,⁷²³ neural network inference,^{724,725} graph analytics,⁷²⁶ and scientific computing.⁷²⁷

4.2.1.2. OUTER PRODUCT UPDATE (OPU)

Analog resistive memory crossbars can also perform a parallel write or an outer product, rank 1 update where all the weights are incremented by the outer product of a vector applied to the rows and a different vector applied to the columns. This is a key kernel for many learning algorithms such as backpropagation⁷²² and sparse coding.^{718,728} Row inputs are encoded in time and column inputs are encoded in either time⁷²⁸ or voltage.⁷²²

When a VMM and MVM are combined on the same crossbar, extremely efficient learning accelerators can be designed,^{719,729,730} with the potential to be $100\text{--}1000\times$ more energy efficient and faster than an optimized digital ReRAM or SRAM based accelerator.⁷¹⁹

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

Several devices have been examined for neural network training, including phase change memory,⁷²⁹ resistive memory^{719,731} and novel lithium-based devices.^{732,733} Currently no devices meet all the ideal requirements for training (high resistance $>10 \text{ M}\Omega$, low write noise $<0.4\%$, low write asymmetry⁷²²). Nevertheless, algorithmic approaches such as periodic carry,⁷³⁴ Local Gains,⁷³⁵ Tiki-Taka,⁷³⁶ or the inclusion of semi-volatile capacitor-on-gate devices⁷³⁷ 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^{729,738,739} 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 k Ω , 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.

4.2.1.3. 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.^{740,741,742} 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 the resulting computational capability.⁷⁴³ FPAAs include aspects of digital computation, such as FPGA blocks or shift registers or microprocessors, to complete the full end-to-end configurable system.⁷⁴³

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.⁷⁴⁴ 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.^{745,746} 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.⁷⁴⁷ These techniques have been utilized by a number of students in university courses.^{748,749} FPAA based VMMs can be scaled to small geometry (e.g. 40 nm and smaller) and operated at RF frequencies.^{750,751} FPAAs have been used for command-word recognition in less than 23 μ W with standard digital interfaces.⁷⁴² The full classification results in less than 1 μ J per classification (or inference), which has 1000 \times improvement over similar digital neuromorphic solutions requiring roughly 1 mJ or higher for just an inference.⁷⁵²

4.2.1.4. 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 BC4.2(a).^{753,754} 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.

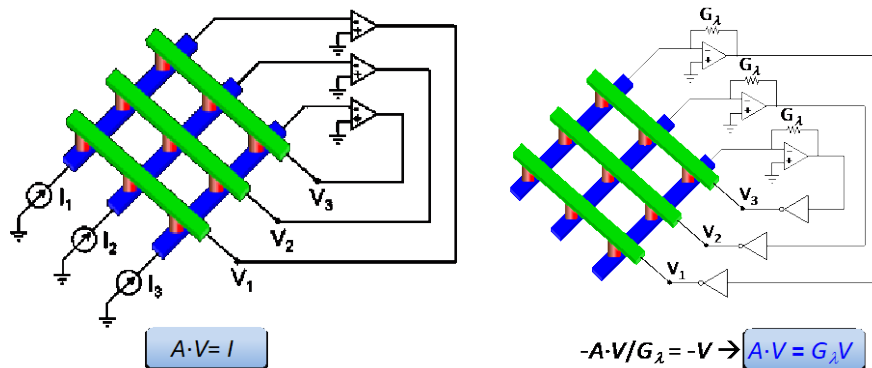


Figure BC4.2 A Resistive Memory Crossbar $Ax = b$ Solver is Illustrated

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

Similarly, eigenvectors of a matrix A can be calculated according to the circuit in Figure BC5.2(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 Fig. 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 μ s, depending on the poles of the analogue feedback circuit.⁷⁵⁵ The same scheme can be extended to one-shot learning by linear/logistic regression.⁷⁵⁶

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.⁷⁵⁷ 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.

4.2.1.5. TERNARY CONTENT ADDRESSABLE MEMORY (TCAM)

Similar to resistive crossbar arrays, content addressable memory (CAM) arrays operate in a highly parallel manner but have a distinct operation which compares a given search word to a set of stored words in parallel within the circuit, and returns the index of the matched entry. This CAM array compare operation takes only one or a few clock cycles, enabling massively parallel high-throughput and low-latency lookups. In addition, Ternary CAMs (TCAMs) have the additional functionality of storing a “don’t care” or “X” value which acts as a wildcard to match regardless of input, enabling compression of stored CAM entries. While very powerful, today’s CAMs based on SRAM suffer from high cost, area and power,⁷⁵⁸ limiting their modern usage to networking (QOS/ACL/IP routing lookup tables) and applications which must trade-off power and cost for high throughput and speed.

To address this gap, a significant body of work has developed TCAM circuits using ReRAM or non-volatile memory (NVM) devices to replace the large and power-hungry SRAM storage devices in conventional TCAM circuits for reduced power and area, and potential for larger array sizes. Work on design, simulation and physical silicon tapeouts have been conducted for phase change memory, metal-oxide memristors, and spin-transfer torque devices.^{759,760,761,762,763,764,765} While several designs have drastically reduced the transistor count per TCAM cell (conventional 16T reduced to 4T, 3T, 2T versions), the limited ON-OFF ratio of NVM and the relatively high device conductance has led to larger power consumption than desired originating from high DC currents during search operations and design work is still underway to optimize area (*i.e.* transistor count), power and device requirements (ON-OFF ratio, write voltage and conductance state stability). Key challenges for the practical realization of TCAM arrays remain, such as the need for highly reliable devices, error correction schemes, scaling of array sizes and balance of devices requirements (ON-OFF ratio, write voltages) and latency-power optimization for different use cases. For example, smaller ON-OFF ratios can lead to longer latency but may be beneficial for certain applications. Despite these challenges, recent work in this area is promising with competitive reported values of 0.5 fJ/bit/search,⁷⁵⁹ and as NVM technologies become more available at commercial foundries, further performance improvements and tuning of NVM CAM circuits are anticipated.

In addition, as increasing work on NVM CAM circuits progresses sufficiently to lower the (power/area) barrier to widespread usage, the development of NVM TCAM circuits as a new in-memory computation primitive presents great potential in future computing systems. 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. While resistive arrays enable in-situ vector-matrix multiplication acceleration for applications in areas such as machine learning, analog signal processing, and scientific computing, CAM circuits provide an orthogonal functionality. Recent work in this area has proposed using CAMs as in-memory computation blocks for applications such as associative processing,^{766,767,768} approximate computing,⁷⁶⁹ spiking NNs,⁷⁷⁰ string matching,⁷⁷¹ and regular expression matching finite state machines.⁷⁷² Further application areas are expected to emerge as NVM CAM circuits mature to make their use advantageous compared to GPU and FPGA approaches.

4.2.2. 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 §4.2.1, analog crossbars can be used as building blocks for conventional neural networks to give significant improvements in energy, latency and area over digital accelerators. For accelerators specialized to a particular algorithm, various analog neurons can be used to process the crossbar outputs and avoid the need for and high cost of analog-to-digital conversion.

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

4.2.2.1. LOCAL LEARNING RULES

Recent breakthroughs suggest that local, approximate gradient descent learning is compatible with Spiking Neural Networks (SNNs) implemented in neuromorphic hardware.^{773,774,775,776,777,778} Spiking neural networks can be viewed as a type of recurrent neural network, where activities are binary and recurrence is both due to connections and the leak of the neurons.⁷⁷³ This analogy enables the transfer of algorithms of deep learning into local synaptic plasticity dynamics in SNNs. 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.

Credit assignment for (deep) hidden parameters in SNNs under tight constraints of information locality is a very challenging issue. All states other than spikes, such as neurotransmitter concentrations, synaptic states, and membrane potentials are local to the neuron. If a non-local quantity is required for a computational process (e.g. learning), a communication channel is required to convey it.⁷⁷⁹ The cost of this channel is non-negligible regardless of the substrate: axons (white matter) take up most of the volume in the brain. Similarly, in modern computers, data movement is the major energy requirement.⁷⁸⁰

One strategy is to train a SNN using local loss functions and the three-factor rule. An example of a loss function that is local to the layer of neurons⁷⁷⁴ is shown in Fig. BC4.3 in diamonds. Local loss functions can be either designed,^{781,782} trained using Back-Propagation (BP)⁷⁸³ or evolved.⁷⁷⁷

Using this approach, it is possible to perform learning in crossbar arrays using temporal dynamics of SNNs in an error-triggered fashion.⁷⁸⁴ Furthermore, the mathematical derivation shows that circuits used for inference and training dynamics can be shared, which renders the learning circuits insensitive to the mismatch in the peripheral circuits. Using error-triggered learning, the number of updates can be reduced a hundredfold compared to the standard rule while achieving performances that are on par with the state-of-the-art.

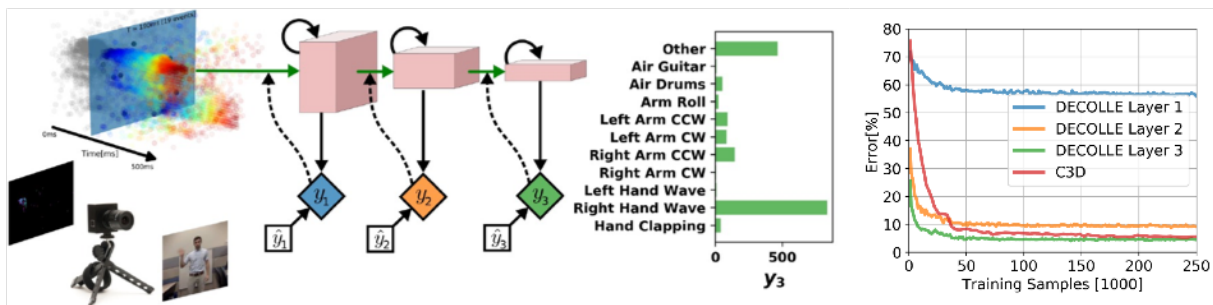


Figure BC4.3 A Spiking CNN for Gesture Recognition with Local Learning

Note: (Left) Deep Continuous Local Learning (DECOLLE) network for gesture recognition. DECOLLE is a forward trained spiking Convolutional Neural Network (CNN) using local loss functions. The network consisted of three convolutional layers with max-pooling. A local classifier (colored diamond) is attached to every layer via a fixed random projection, i.e. each layer is trained to classify the gesture. The fixed random projection enables the network to find increasingly disentangled representations as the data flows through the hierarchy. DECOLLE is fed with 1-ms integer frames recorded from a neuromorphic vision sensor. (Right) Classification Error for the DvsGesture task during learning for all local errors associated with the convolutional layers. C3D is a 3D CNN commonly used for sequence learning.⁷⁸⁵ See Ref. 774 for details.

4.2.2.2. HYPERDIMENSIONAL COMPUTING

Hyperdimensional (HD) computing is a cognitive computing model based on the high-dimensional properties of neural circuits in the brain.⁷⁸⁶ Information is encoded into high-dimensional distributed representations in the form of high-dimensional vectors or *hypervectors*.^{786,787} A hypervector distributes information uniformly across all of its dimensions, resulting in a distributed or

holographic representation.^{788,789} 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.⁷⁸⁹

Computation is performed by systematically transforming hypervectors into other hypervectors (i.e. *closure* is satisfied) using operations such as multiplication, addition, and permutation. Due to the distributed nature of hypervectors, these operations become local to individual dimensions (or bits, in the case of binary hypervectors). Multiplication, for example, becomes element-wise multiplication and can be used to “bind” hypervectors together into a single hypervector representing a key-value pair.^{789,790,791} Multiplying by the multiplicative inverse of the key hypervector will “unbind” the key hypervector from its corresponding value hypervector. Similarly, other operations such as addition and permutation can be used to construct sets and sequences respectively.^{789,790,791} These simple operations together enable the construction of more complex compositional representations, which are ultimately stored in an *autoassociative* memory for item “cleanup” as hypervectors become noisy during computation.^{789,790,791,786}

Constructing a physical HD computer will require innovations at all levels of the computing stack but could result in an energy efficient “thinking” machine.^{792,793,794,795,796} The robustness to noise and bit errors provided by the high-dimensional distributed representations reduces requirements on signal-to-noise ratio. This enables lower supply voltages and greater tolerance for device variability than in conventional computers.^{787,792,793,794} Furthermore, HD computing operations are local to individual digits or bits and are, consequently, highly parallelizable.^{789,790,795,796} The trade-off is large word sizes, which necessitate *processing-in-memory*.^{792,793,794,795,796} Fortunately, HD computing algorithms are geared towards *one-shot* or *few-shot* learning^{787,792,797,798} and do not require frequent weight updates. As a result, hypervectors may lie dormant for extended periods of time, allowing for a greater portion of inactive components for reduced power consumption.

4.2.2.3. SPIKING-BASED NEURAL NETWORKS (SNNS)

Like the brain which 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,⁷⁹⁹ spin torque based devices,⁸⁰⁰ stochastic phase change neurons,⁸⁰¹ superconducting neurons,⁸⁰² 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.

4.2.3. COMPUTING WITH DYNAMICAL SYSTEMS

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

Although some of the below methods target NP-hard problems, it’s important to note that, to date, no general physical computing method (including analog and quantum approaches) has yet been clearly demonstrated to be capable of solving NP-hard problems without requiring exponential physical resources (energy and/or time) to be invested in the physical process performing the computation. The prevailing belief among computational complexity theorists⁸⁰³ is that solving NP-hard problems efficiently would require uncovering new physics (i.e., beyond standard quantum mechanics).

4.2.3.1. 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.⁸⁰⁴ 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 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.⁸⁰⁵

Another important application for combinatorial optimization, specifically graph coloring, was described⁸⁰⁶ and developed theoretically with support from experimental demonstrations using relaxation oscillators based on phase-change IMT materials. An architecture based on non-repeating phase relations⁸⁰⁷ between fabricated CMOS oscillators tries to emulate stochastic local search (SLS) for constraint satisfaction problems.

4.2.3.2. 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,^{808,809,810,811,812} optical,^{813,814,815} and electromechanical⁸¹⁶ 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 combinatorial optimization problem can be converted into an equivalent Ising problem that is programmed onto and solved by the machine. The energy minima of such networks correspond to the solutions of an NP-hard combinatorial optimization problem and can model other NP-hard problems as well.⁸¹²

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. They can thus rapidly sample the local minima of a problem,⁸⁰⁹ and can potentially also be used to arrive at solutions close to the global minimum.^{808,811} The sampling speed is determined fundamentally by the oscillator frequency. 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.^{810,813} Energy and delay benchmarking remain a future step.

Of the proposed coupled oscillator optimization schemes, the systems which 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.^{808,811} In these networks, the oscillators take on a discrete nature by synchronizing to a second-harmonic master oscillator through a circuit nonlinearity. The nonlinear oscillator properties can

be easily provided by semiconductor components such as diodes, MOS capacitors, or amplifiers. In fact, phase-based Boolean computing using such nonlinear oscillator circuits was proposed decades ago.⁸¹⁷

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. In particular, variability in the distribution delay of the second-harmonic master oscillator by more than a half-cycle are detrimental: this will be a significant problem in large networks. To this end, architectures have been proposed that separate the oscillators in time rather than space,⁸¹⁴ but this comes at the loss of continuous-time communication among the oscillators, leading to slower convergence.

Related constraint satisfaction systems have been built.⁸¹⁸ An interesting approach based on memory co-processors was introduced as Memcomputing.⁸¹⁹ Useful insights can also be obtained by looking into dynamical systems like iterated maps,⁸²⁰ and 0-1 continuous reformulations of discrete optimization problems.⁸²¹

4.2.3.3. ASSOCIATIVE MEMORIES

Hopfield networks are attractor networks proposed for associative memories⁸²² 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.^{823,824,825}

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.^{826,827} Five decimal digits, '1'–'5', are associated with the other five digits, '6'–'0'. Hebbian learning is used for storing patterns.⁸²⁸ 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.⁸²⁹

4.2.3.3.1. CELLULAR NEURAL NETWORKS

The cellular neural network⁸³⁰ (CeNN) is a non-Boolean computing architecture that contains an array of computing cells that are connected to nearby cells. Since interconnects are major limitations in modern VLSI systems, CeNN systems take advantage of the local communication and encounter fewer constraints imposed by interconnects. The CeNN is a brain-inspired computing architecture that relies on neurons to integrate the incoming currents. The accumulated and activated output signal drives nearby neurons through weighted synapses. CeNNs can be used to create associative memories for voice and image recognition.

CMOS based CeNNs can be implemented by analog circuits using operational amplifiers and operational transconductance amplifiers (OTAs) as neurons and synapses, respectively.^{831,832} Some recent work has also investigated CeNN using beyond-CMOS charge-based devices, such as TFETs, to potentially improve energy efficiency^{833,834} thanks to their steep subthreshold slope and low operating voltage.

Using novel devices such as all-spin logic (ASL),⁸³⁵ charge-coupled spin logic (CSL),⁸³⁶ and domain wall logic (mLogic)⁸³⁷ 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.⁸³⁸ 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.

4.2.3.4. DIFFERENTIAL EQUATION MODELING

Analog computation aligns well with the solution of differential equations, both Ordinary Differential Equations (ODE) and Partial Differential Equations (PDE).^{839,840} A key limitation of digital ODE and PDE solution accuracy is the need to discretize time and amplitude. Continuous time analog solutions eliminate this. Analog summation by charge or current is also ideal and not subject to round off error. Similarly, analog integration can be ideal and is typically performed as a current sum on a capacitor.

The result of the final computation will have some noise distortion, but the noise is added at the end of the resulting computation and does not affect the core numerics.

Analog computation builds on an analog numerical analysis techniques,⁸⁴⁰ analog algorithm complexity theory,⁸⁴¹ and analog algorithm abstraction theory,⁸⁴² where these directions provide guidance on the relative numerical performance, as well as relative architectural performance, for designs abstracted at multiple levels of representation, respectively. Analog solution of differential equation computes using real valued quantities, often computing over continuous amplitude and time.⁸⁴³

Analog solution of PDEs often utilize spatially coupled ODEs^{844,845,846} where the physical system is continuous in space, but practically the parameters change in discrete points with a finite granularity of parameter resolution setting, as well as output measurement capability. Often differential equations are solved using programmable and reconfigurable techniques,^{847,848,849} enabling utilization of many coupled ODEs or PDEs operating over a range of room-temperature conditions. Differential equations have also been mapped to cellular automata-based networks.^{850,851,852}

Solving an application in analog such as the optimization problems in Section 4.2.3.2, can be thought of as solving a differential equation. These analog solutions involve a transformation between the physical system to be computed to the analog computing substrate. Efficient transformation to an analog computing system can result in a 1000× or larger computational efficiency improvement compared to similar digital approaches.⁸⁴⁰ Differential equations utilize superposition over the linear operating region of the particular physics being used.⁸⁴³ Computational complexity when solving applications using differential equations is still unclear. Although it seems likely that real-valued analog solutions provide an improved computational ability over standard, discrete-valued Turing machines,⁸⁴³ as of this writing, such capabilities have not been convincingly experimentally demonstrated to date and is an active research area. As a result, physical algorithms, such as the recently proposed ODEs to solve the 3-SAT problem^{853,854,855} must be developed and verified only through physical hardware, and not discrete simulation or analysis of physical algorithms.

4.2.3.5. SUB-KT CHAOTIC LOGIC AND CHAOS COMPUTING

Shannon's noisy channel coding theorem⁸⁵⁶ shows that one can reliably communicate information on a channel subject to noise (at a sufficiently low bitrate) even when the transmitted signal power is below the noise floor (*i.e.*, at a signal-to-noise ratio of less than 1). Moreover, any computational process can be viewed as just a special case of a communication channel, namely, one that simply happens to transform the encoded data in transit—since the derivation of Shannon's theorem relies solely on counting distinguishable signals, and nothing about how the signals are being counted in Shannon's argument precludes the encoded data from being transformed as it passes through the channel. This observation suggests that performing reliable computation utilizing signal energies (that is, energies associated with the 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^{857,858,859} 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 (§4.2.3.1). 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 (§4.4) 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 possible that faster variations on this or similar techniques may be found with further investigation.

An earlier, more extensively-developed proposal that is similar to the chaotic logic concept is called *chaos computing*.⁸⁶⁰

4.2.4. ANALOG MEMORY AND COMPUTE DEVICES

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

4.2.4.1. RERAM

ReRAM memory is very dense and can be integrated in the back end of the line, avoiding the use of transistor area for the memory. Several ReRAM crossbars have been demonstrated for inference.^{861,862,863} A key challenge is maintaining good analog properties

52 Emerging Device-Architecture Interaction

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.^{864,865} ReRAM for training remains a challenge due to the non-ideal electrical characteristics of synaptic devices.

4.2.4.2. 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.⁸⁶⁶ 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.⁸⁶⁷ Careful choice of pulse condition can stretch this procedure out to many hundreds of pulses.⁸⁶⁸ Some recent work has shown some evidence for gradual increases in resistance with multiple successive pulses,⁸⁶⁹ 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 (see §4.2.1.2 on Outer-Product Update above).⁷²⁹ 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,^{870,871} 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.^{872,873,874} 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.⁸⁷⁵ This is because any highly-predictable signal loss can be compensated by signal amplification, at least until background noise becomes strongly amplified.

4.2.4.3. ION-INSERTION REDOX TRANSISTOR:

Ion-insertion redox transistors (IIRT) have recently emerged as promising candidates for analog memory.^{876,733} IIRTs are 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. IIRT 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 IIRT 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, IIRT represents a class of devices with dynamic dopant control. For analog memory, a major advantage of IIRT 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.⁷³² The storage capacity can be several orders of magnitude larger than 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 $\sim 5,000$ C/cm³ and polymer-based electrodes reported at ~ 50 C/cm³ which could provide as much as 50 pF/ μ m² for a 10nm thick channel.

Additionally, IIRTs may offer promise for low voltage digital logic.⁸⁷⁷ Dynamic doping can lead to sharp metal insulator transitions, *e.g.* due to correlated electron effects in redox-active oxides,⁸⁷⁸ and may result in abrupt low voltage switching.

4.2.4.4. FLOATING GATE

Floating gate synapses (so-called “synaptic transistors”) were first developed in 1994.⁸⁷⁹ They are modified EEPROM devices which can be fabricated in a standard CMOS process and programmed to within 0.2% accuracy.⁸⁸⁰ A number of sophisticated systems have been developed based on the arrays of synaptic transistors.^{881,882,883} 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 size⁸⁸⁴, 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.^{885,886} One representative example is Embedded SuperFlash (ESF) memory from Silicon Storage Technology, Inc.⁸⁸⁷ The areas of the modified arrays of the ESF1⁸⁸⁵ and ESF3⁸⁸⁶ NOR flash memories, with the latter technology scalable to $F = 28$ nm, are close to $120F^2$ and may be further reduced to $\sim 40F^2$. (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.⁸⁸⁵ 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.⁸⁸⁵ 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.^{888,889}

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 $100 \text{ nm} \times 100 \text{ nm}$ area in 32nm process.⁸⁹⁰ Commercial NAND manufacturers have shown devices at 15 nm⁸⁹¹ and 19 nm.^{892, 893} 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 $30 \text{ nm} \times 30 \text{ nm}$ 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.⁸⁹⁴ The number of layers is projected to further increase to 512 to enable 10 Tb/in² density,⁸⁹⁵ 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 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.⁸⁹⁶ 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.⁸⁹⁷

4.2.4.5. 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.⁸⁹⁸ 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 $100,000 \times$ smaller than the decay constant (*e.g.*, 20 ms decay constant and 200 ns per training example).⁸⁹⁹ 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.^{898,899} 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.⁷³⁷ Later, it was shown that this same technique could suppress fixed device variabilities in other kinds of lower-significance conductances, including PCM devices.⁹⁰⁰ 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.⁷³⁷

4.2.4.6. 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

54 Emerging Device-Architecture Interaction

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^{901,902,903} 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.⁹⁰³

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,⁹⁰⁴ while low-precision versions have achieved efficiencies at the level of 100 fJ/MAC.⁹⁰⁵ Comparison of key metrics with the state-of-the-art in analog capacitive VMM ICs^{906,907,908,909,904} is given in Table BC4.2 above.

Table BC4.1 Metrics for Analog Capacitive Vector-Matrix Multiply (VMM) ICs

4.2.4.7. 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 junction⁹¹⁰ or with a clustered magnetic Josephson junction.⁹¹¹ 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.⁹¹² On the superconducting device side, high speed microprocessors and communication systems have been developed, based mostly on superconducting tunnel junctions.⁹¹³

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.⁹¹¹ 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.

4.2.4.8. 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 ^{914,915,916}). However, their main advantage comes from addressing MAC operations directly. Here, we will look at the 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.

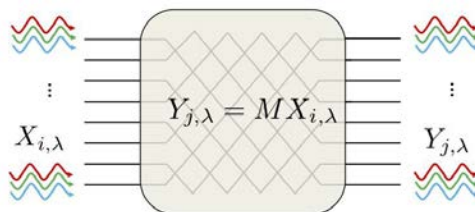


Figure BC4.4 One Generalized Instantiation of a Photonic MVM unit, with Wavelength Multiplexed Inputs and Outputs and a Coupler-based Tunable Array. Reproduced from⁹³¹.

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 increase the compute density of the approach. Classic examples include arrays of resonator weight banks^{915,917,918} or Mach Zehnder interferometers⁹¹⁶. The most important metrics are *energy efficiency* (energy/MAC), *throughput per unit area* ($\text{MACs}/\text{s}/\text{mm}^2$), *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 arrays⁹¹⁹ or SIMD units,⁹²⁰ although there are some other architectures that use aspects of both.⁹²¹ 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/mm², 0.5–1 GMVM/s, and 1–2 μs , respectively. In contrast, photonics MVM units could perform in range 2–10 fJ/MAC, 50 TMACs/mm², 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^{922,923} or devices with an enhanced electro-optic coefficient (dn/dT , $d\alpha/dT$) such that heaters are not as necessary.^{924,925} 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^{926,927} 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,⁹²⁸ photonic devices in state-of-the-art silicon foundries can be designed with low scattering,⁹²⁹ while detectors such as avalanche photodiodes,⁹³⁰ 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.⁹³¹: single-digit aJ/MAC efficiencies, and 100s of PMACs/s/mm² 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 μm), but perform better for larger arrays (distances approximately > 100 μm)⁹³¹. 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,⁹³¹ 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 operations⁹¹⁶ can conserve the variation of the input and output signals, in contrast to other approaches such as resistive crossbar arrays,⁷¹⁸ 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 ($> 10\text{k}$ components) are still early in early development—analogue photonic systems must grapple with the challenge of addressing yield, variability, precision, and tunability. Nonetheless, the total cost to produce a photonic chip package at high volume is dipping below one hundred dollars, and it is expected that the trend will continue⁹³². The orders of magnitude advantages offered by photonics, and its potential for HVM scalability, makes it a viable inroad for the breakthrough performance and innovation required by artificial intelligence algorithms in the years to come.

4.2.4.9. MAGNETIC NEURAL NETWORK DEVICES

Several types of magnetic circuits can be used to implement neural networks including spin-diffusion-based devices,⁹³³ charge-coupled spin logic (CSL),⁹³⁴ and domain wall logic (mLogic).⁹³⁵ The use of these different devices has been benchmarked.⁹³⁶

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 K_u for PMA magnets, large spin Hall angle θ for SHE materials, and small critical depinning current for domain wall magnets.

4.2.4.10. DEVICE TECHNOLOGIES FOR COUPLED OSCILLATOR SYSTEMS

It is challenging to build compact, low-power oscillators that can also be coupled together to give predictable phase or frequency dynamics. Standard digital ring oscillators based on CMOS inverter feedback loops, as well as the typical transistor-driven LC oscillators used in RF designs, are both less than ideal, due to device nonlinearities in the digital regime and the high biasing currents used in linear-regime analog small-signal oscillators, as well as relatively large oscillator sizes in both cases. Thus, various “Beyond CMOS” oscillator technologies have been explored. Such technologies can use and manipulate the charge, spin, or quantum properties of electrons, or use photons. Important examples include spin-torque,^{937,938,939} insulator-metal-transition,⁹⁴⁰ optical,^{815,941} and quantum.⁹⁴² Memristor-based oscillators have also been constructed.^{943,944,945}

Non-silicon electrical oscillators include two important kinds which are currently being developed. One prominent effort is the use of spin torque oscillators (STOs) coupled with using spin diffusion currents, or electrical signals, for providing a computational platform for machine learning, spiking neural networks, and others.^{937–939,815,946,947,948} However, the high current densities of STOs and the limited range of spin diffusion currents continue to pose serious challenges in creating coupled networks of such oscillators. Optical oscillators have been studied^{941,949} and used for computing,⁸¹⁵ but challenges include bulky components, difficult interfacing between the electrical and optical domains, and lack of programmability to enable an optical computing apparatus. Another promising non-silicon technology for very compact oscillators is the IMT (insulator-metal transition) material-based oscillator technology.^{950,951} 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^{952,953,954,955,956} but the focus has been to build high frequency and low power individual oscillators, as opposed to the demonstration of coupled systems of oscillators, or the generation of interesting dynamics for computing. A comparison of some computing-focused electrical oscillators is shown in Table BC4.3 below.

Table BC4.2 Comparison of Some Electrical Oscillators for Computing

4.2.4.11. ANALOG NEURON FUNCTIONAL BLOCK

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

1. Spatial and temporal integration of input signals.
2. Stochastic firing of an action potential after a threshold is reached.
3. Relaxation to the resting potential if the time delay between excitations is too long.^{957,958,959}
4. Non-linear activation functions. In many neural networks, after a linear multiply accumulate is performed on a set of inputs, or in a hidden layer, a non-linear activation or transfer function is applied. The presence of this function prevents the network from mathematically collapsing into a single linear equation, which helps improve the computational capabilities as the number of layers increases. Common functions include sigmoid, tanh and rectified linear (ReLU). In an analog network it is desirable that this function be performed by a single device.

These analog blocks are often with other building blocks such as crossbars, §4.2.1, to accelerate neural networks, to build dynamical systems, §4.2.3, create energy minimization circuits, §4.2.3.1 and enable probabilistic circuits, §4.3. Stochastic neurons/ devices are discussed in §4.3 rather than here. Using an analog neuron allows analog data to be processed directly, obviating the need for time- and energy-consuming ADC/DAC in the circuits.

Recently, Mott memristors,⁹⁶⁰ phase change based memristive switches,⁹⁶¹ and chalcogenide threshold switches⁹⁶² 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. Device candidates to perform the transfer function include STT-MRAM.⁹⁶³

Ionic diffusion dynamics or electrical instabilities enable a single memristive device to perform analog functions like resistance tuning or pulse generation after accumulating input,^{961,964} which typically requires a large number of CMOS transistors. These analog functions are critical in emulating synaptic and neuronal behavior. Taking into account the simple structure and scalability of a two-terminal memristor, both the complexity of a circuit and the area consumed to build a neuron network will be much less compared to a CMOS circuit. Being a passive device, the memristor offers stimulation-dependent electric conductance. However, the passive nature results in a lack of power to maintain sustainable neural signal propagation in a network if passive synaptic blocks are employed. Additional power sources are mandatory for neural networks with multiple layers.

The benchmarks used to evaluate electronic spiking neurons generally measure the energy per operation, the fabrication cost, or the chip area of the integrated functional block, and the fidelity to the desired neuron function (e.g. integrate and fire). High reliability and low variation of devices are two key factors for the viability of a neuron technology. Device failure will require a lot more circuitry for error detection and correction.⁹⁶⁵ Large variation increases the difficulty for designing peripheral circuits and degrades the adaptability of the block.

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

4.2.5. ANALOG COMPUTING—CLOSING REMARKS

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

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

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

4.3. PROBABILISTIC CIRCUITS

Traditionally, conventional computational processes are designed to be deterministic, with computational results determined by the machine’s initial state and inputs. Nevertheless, computations that are *intentionally* designed to behave randomly or stochastically, even at the level of individual bit-operations, are of interest and can have many useful applications such as simulated annealing, Monte Carlo simulation, machine learning or Boltzmann machines, randomized algorithms⁹⁶⁶ 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.^{967,968,969,970}

Obtaining randomness in traditional CMOS is difficult and typically relies on a pseudo-random number generator. This requires a large circuit block and significant computational effort to obtain high quality random numbers. Several new devices have been proposed to obtain true randomness as discussed in §4.3.1. These allow for a random bit to be generated with a single device. Chaotic devices can be used to turn poor quality randomness into high quality random numbers. Novel architectures such as probabilistic p-logic (§4.3.2) or a traveling salesman solver (§4.2.3.1) can be designed using the new true random number generators.

4.3.1. DEVICE TECHNOLOGIES FOR RANDOM BIT GENERATION

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

4.3.1.1. 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,⁹⁷¹ 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 reported⁹⁷² to be 1,000% and can presumably be larger. There is currently intense research activity in half-metallic ferromagnets to increase TMR.

4.3.1.2. SINGLE-ELECTRON BIPOLAR AVALANCHE TRANSISTOR (SEBAT)

The single-electron bipolar avalanche transistor (SEBAT) is a novel Geiger-mode avalanche bipolar transistor structure.^{973,974} 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.

4.3.1.3. MEMRISTORS/RESISTIVE RAM

The intrinsic variability of memristive switching, particularly the switching delay time of memristors, can be a good source of stochasticity.^{975,976,977} Such stochasticity originates from the ionic dynamics within the memristors.⁹⁷⁸

4.3.1.4. CONTACT-RESISTIVE RANDOM ACCESS MEMORY (CRRAM)

CRRAM can be used for random number generation.^{975, 979} 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.⁹⁸⁰ 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.

4.3.1.5. CMOS

There are different ways to obtain random number generators (RNG) in CMOS using different physical noise sources, one being “jitter” in ring oscillators.⁹⁸¹ These TRNGs can be tuned into tunable random number generators as required but require significant amounts of area when compared to single device alternatives.

4.3.1.6. 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 $\Phi_0 = 2 \times 10^{-15}$ 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.⁹⁸²

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.⁹⁸³ Circuits based on these devices have many promising potential applications. For example, stochastic Josephson junctions have been shown to make effective pseudo-sigmoid generators,⁹⁸⁴ and stochastic Josephson junction spiking can perform the neural accumulate operation at speeds up to 70 GHz.⁹⁸⁵

4.3.2. 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.^{986,987} The authors explored how p -bits can be compactly realized by leveraging existing Magnetoresistive RAM (MRAM) technology⁹⁸⁸ and showed different applications of p -circuits including image recognition (inference),⁹⁸⁹ combinatorial optimization,⁹⁹⁰ Bayesian networks,⁹⁹¹ and an enhanced type of Boolean logic that allows invertible operation.⁹⁸⁷ More recently, potential applications have been extended to include emulation of a class of quantum systems⁹⁹² and

on-chip learning for stochastic neural networks.⁹⁹³ 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.⁹⁹⁴

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.^{987,988} 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. Even though the p-bit concept is hardware agnostic and digital implementations of invertible logic have been realized,^{995,996} the main advantage of an MRAM-based p-bit comes from its compact, low-power implementation of a complex functionality (tunable randomness) compared to digital implementations.⁹⁹⁴ 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. Secondly, 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.⁹⁹⁷ Recent breakthroughs in modern MRAM industry has 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.⁹⁹⁸

4.4. REVERSIBLE COMPUTING

Besides analog computing (§4.2) and probabilistic computing (§4.3), a third dimension along which we may explore departures from the conventional computing paradigm, is *reversible computing*.⁹⁹⁹ 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 be brought closer to zero asymptotically as the technology continues to be improved.

In 1961, Rolf Landauer of IBM argued¹⁰⁰⁰ 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.^{702,1001}

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.^{702,1002} 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.⁷⁰²

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. Bennett¹⁰⁰³ 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,¹⁰⁰⁴ 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 resource overheads can be outweighed by the achievable energy savings, and total cost may be reduced compared to an irreversible design. In the long run, reversible computing is the *only* physically possible path by which the amount of general digital computation that can be performed per unit energy (and cost!) might continue to be increased indefinitely, without any known fundamental limit.

In existing *adiabatic* implementations of reversible computing in today's device technologies (see §§4.4.1–4.4.2 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 t_d 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

$$E_{\text{diss}} \cdot t_d \cong c_E,$$

where c_E is the constant DdP, which we may also call the *energy coefficient* of the technology. However, there is no proof that this same linear tradeoff relation extends to *all* possible implementation technologies for reversible computing, and in fact, recent results suggest that an exponential downscaling of energy dissipation with delay may sometimes be possible when quantum effects are leveraged.¹⁰⁰⁵ Further, 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,¹⁰⁰⁶ 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.¹⁰⁰⁷ 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.

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^{1008,1009,1010,1011,1012,1013,1014,1015,1016,1005} and ballistic^{1017,1018,1019,1020,1021,1022,1023,1024,1025,1026,1027,1028,1029,1030,1031} 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 §4.4.4 for further discussion.) There is not yet any “magic bullet” physical implementation strategy that automatically addresses all the many possible energy-loss mechanisms in a complete system all at once. Logical reversibility (when suitably generalized⁷⁰³) 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 at least *a* way that general digital computing can move forward. 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.¹⁰³² Further, even some of the existing reversible superconducting logic styles (such as RQFP^{1033,1034,1035,1036,1037,1038} and nSQUID^{1039,1040,1041} 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.^{704,705} 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 §4.4.2 below for further discussion of superconducting reversible computing technologies.

Reversible computing can also be potentially usefully combined with probabilistic computing (§4.3); 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.⁷⁰² 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.⁷⁰² 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 (irreversible) transformations can be permitted together with compensating one-to-many (nondeterministic) transformations in a computation,¹⁰⁴² 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 §4.2.4.6.¹⁰⁴³ Further, fundamental physics is reversible at the microscale, which suggests that a sufficiently carefully-engineered analog computer might be made to approach macroscopic reversibility, and that its energy efficiency might be increased without limit as its technology is further refined. The degrees of freedom utilized for the analog physical computation would likely have to be very well-isolated from the system's thermal degrees of freedom, and the usual tendency for complex dynamical systems to devolve towards chaotic behavior would have to be suppressed in some way, or else made into a useful feature of the computational process. The previously mentioned work on chaotic logic (§4.2.3.4) suggests one potential technique for harnessing the chaotic analog behavior of conservative dynamical systems usefully for computational purposes, but many other, more sophisticated methods may be possible.

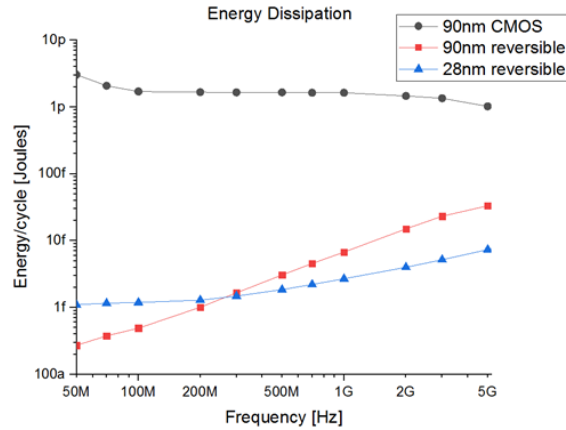


Figure BC4.5 Energy Dissipation per Stage vs. Frequency in an Adiabatic CMOS Shift Register

4.4.1. 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*,⁷⁰⁵ 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 $c_E \propto C^2V^2R$.

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.^{1011,1044,1045,1046,1047,1048} These methods did not gain widespread traction at the time, perhaps 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 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.

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),^{1013,1049} *positive feedback logic* (PFAL),^{1050,1051} and *efficient charge recovery logic* (ECRL).^{1052,1053} 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.^{1054,1055} And general-purpose computing has been pursued in the design of adiabatic microprocessors.^{1056,1057} The simulation results shown in Fig. BC4.5¹⁰⁵⁷ 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.

4.4.2. 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.^{1008–1009,1014,1033–1040} 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)¹⁰⁰⁸ and *quantum flux parametron* (QFP)^{1058,1009} were proposed in early studies. Both approaches were based conceptually on the abstract physical model of adiabatic digital operations introduced by Landauer,^{1000,1003} in which the potential energy function of the digital element is transformed adiabatically between single-well and double-well configurations in the course of the 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,¹⁰⁵⁹ and its energy dissipation was estimated to be a few kT per reversible bit-operation.¹⁰³⁹

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 operation¹⁰¹⁴ and eliminating the junction shunt resistance.¹⁰⁶⁰ The energy dissipation of this improved adiabatic QFP (AQFP) was investigated numerically, taking thermal noise into account,¹⁰⁶¹ and found to be well below kT with a low error rate.¹⁰³³ 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.¹⁰⁶² The energy dissipation per operation of an (irreversible) AQFP 8-b carry-lookahead adder was experimentally evaluated to be 24 kT per Josephson junction.¹⁰⁶³

The first demonstration of logically and physically reversible operation of superconducting logic was performed using a newer reversible QFP (RQFP) design style.¹⁰³⁴ 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

$$(y_0, y_1, y_2) = (\text{MAJ}(\bar{x}_0, x_1, x_2), \text{MAJ}(x_0, \bar{x}_1, x_2), \text{MAJ}(x_0, x_1, \bar{x}_2))$$

where $\text{MAJ}(i, j, k) = (i \wedge j) \vee (j \wedge k) \vee (k \wedge i)$. 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.¹⁰³⁵ Using RQFP gates, the functionality of a 1-bit reversible full adder was demonstrated, and its energy dissipation was numerically calculated, while accounting for thermal noise. Fig. BC4.6 shows the simulation results for the energy dissipation of the reversible and irreversible full adders as a function of the frequency of the driving clock.¹⁰⁶⁴ 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.

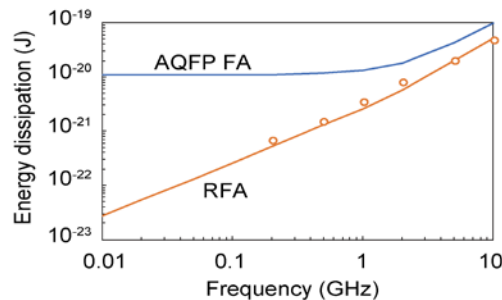


Figure BC4.6 Energy dissipation of RQFP and irreversible AQFP 1-b full adders

Note: Figure is from Ref. 1064. 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.

4.4.3. OTHER REVERSIBLE TECHNOLOGY CONCEPTS

Beyond the adiabatic semiconductor/superconductor technologies discussed in §§4.4.1–4.4.2 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.^{1065,1066,1010,1067} 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 $\sim 4 \times 10^{-26}$ J per operation at room temperature at speeds of 100 MHz.^{1015,1016} This is roughly 74,000 \times lower energy than the Landauer limit at 300 K, and roughly $10^6 \times$ 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.

Back in the electronic domain, since the 1990s, a number of adiabatic reversible computing concepts based on *single-electron devices* have been proposed.^{1012,1068,1005,1069} 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.^{1070,1071} 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 an interesting concept for *adiabatic capacitive logic*.¹⁰⁷²

In addition to the various adiabatic approaches to reversible computing, there are also a number of *ballistic* reversible computing concepts.^{1017–1031} These are based on a rather different picture of the basic physical mechanism of reversible computing than the adiabatic approach suggested by Landauer.¹⁰⁰⁰ 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,¹⁰¹⁷ 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, 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.^{1019–1031} 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)¹⁰⁷³ or *long Josephson junctions* (LJJs).¹⁰⁷⁴ 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,^{1023–1027} and the *asynchronous ballistic* approach which has been in development since 2016 at Sandia.^{1028–1031}

4.4.4. 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:

- 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, $D(d)$. 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?
- 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),¹⁰⁷⁵ topological invariants, dynamical variations of the quantum Zeno effect (QZE)^{1076,1077} 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 $D(d)$ at useful operating speeds while still being inexpensively manufacturable. New 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.

- 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.* through complementary signaling. And if bulk semiconductors are used, this adds another level of challenge relating to time-varying loads during transitions, since device capacitances are more voltage-dependent if depletion regions are not structurally constrained.
- At higher levels, many advances in areas such as reversible architectures, EDA tool enhancements to support reversible design styles, reversible algorithms and so forth still need to be developed. As useful reversible computing hardware technologies emerge and develop, systems engineering practice will need to evolve to best leverage the opportunities and tradeoffs offered by reversible design. However, all these R&D areas remain in their infancy at this time.

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 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 styles 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.

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,^{1078,1079,1080} 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^{1081,1082,1083} have been suggested as a pathway to a PUF design. (More detailed reviews are also available.¹⁰⁸⁴) 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,¹⁰⁸⁴ to securely extract software,¹⁰⁸⁵ in trusted Field Programmable Gate Arrays (FPGAs),¹⁰⁸⁶ and for encrypted storage.¹⁰⁸⁵ Post-CMOS devices also find utility as random number generators (RNGs) that may be employed for secure communication channels (e.g., to generate session keys¹⁰⁸⁴). 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 studied¹⁰⁷⁸ 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,¹⁰⁸⁴ variations in the required write time in spin torque transfer random access memory (STT-RAM) was proposed to create a domain wall memory PUF.¹⁰⁸¹ Other structures based on magnetic tunnel junctions have also been proposed.^{1087,1088} Variations in write times have also been exploited to produce unique responses in phase change memory (PCM) arrays.¹⁰⁸⁹ The variability of ReRAM presents a natural opportunity for PUF implementation, and array demonstration has been reported.^{1090,1091} At the array-level, variations in diode resistivity have also been used to derive challenge/response pairs from crossbar structures.¹⁰⁹² PUFs based on graphene¹⁰⁹³ and carbon nanotubes have also been proposed/considered.¹⁰⁹⁴

As a more representative case study, prior work¹⁰⁸⁴ considers an array structure based on process variation in memristors^{1095,1096} to create a PUF structure (referred to as *NanoPUF*¹⁰⁸⁴). 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 consist 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,¹⁰⁹⁵ the pulse’s duration and amplitude may also be varied.¹⁰⁹⁶

5.1.3. RANDOM NUMBER GENERATORS (RNGs) AND EMERGING TECHNOLOGIES

The inherent randomness in emerging devices can also be used to generate random numbers.¹⁰⁸⁴ As a representative case study, prior work¹⁰⁸⁴ explores an approach based on contact-resistive random access memory (CRRAM).¹⁰⁹⁷ (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.¹⁰⁹⁸)

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.¹⁰⁹⁸ Indeed, RNGs based on emerging devices¹⁰⁹⁸ can successfully pass randomness tests such as those provided by the National Institute of Standards and Technology (NIST).

As random number are derived from current passing through filaments, memristors, PCM, and RRAM devices can also be leveraged to build similar RNGs.¹⁰⁸⁴

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 (45nm and below), the superposition of n-type and p-type carriers is observable under normal bias conditions. The ambipolarity phenomenon exists in various materials such as silicon,¹⁰⁹⁹ carbon nanotubes¹¹⁰⁰ and graphene.¹¹⁰¹ By controlling ambipolarity, device polarity can be adjusted/tuned post-deployment. Transistors with a configurable polarity – e.g., carbon nanotubes,¹¹⁰² graphene,¹¹⁰³ silicon nanowires (SiNWs),¹¹⁰⁴ and transition metal dichalcogenides (TMDs)¹¹⁰⁵ – 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.¹¹⁰⁶

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.¹¹⁰⁷ 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,¹¹⁰⁷ 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.^{1108,1109} 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.¹¹¹⁰ 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.¹¹¹⁰

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

Camouflaging Layout: Split manufacturing and IC camouflaging are used to secure the CMOS fabrication process, albeit with high overhead and decreased circuit reliability. With CMOS camouflaging layouts, both power and area would increase significantly in order to achieve high levels of protection.¹¹¹¹ A CMOS camouflaging layout that can function either as an XOR, NAND or NOR gate requires at least 12 transistors. Emerging technologies help reduce the area overhead. Recent work has demonstrated that only 4 SiNW FETs with tunable polarity are required to build a camouflaging layout that can perform NAND, NOR, XOR or XNOR functionality.^{1112,1113} 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.¹¹¹⁰ 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 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.^{1114,1115} 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.¹¹¹⁶ 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.^{1116,1117}

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.¹¹⁰⁸ Such behavior has been observed experimentally.^{1118,1119} 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.¹¹²⁰

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.¹¹²¹ 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.^{1108,1109} 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,¹¹⁰⁹ 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.¹¹¹⁰ (SymFET-based Boolean logic is also possible.¹¹²²)

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 DPA¹¹²⁴, 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,¹¹²³ e.g., frequently changing the keys to prevent the attacker from collecting enough power traces¹¹²⁴ or using masking bits during the internal stages to limit information leakage.¹¹²⁵

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.¹¹²⁶ 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.¹¹²⁷ 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.¹¹²⁸

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-box¹¹²⁹.

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 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.

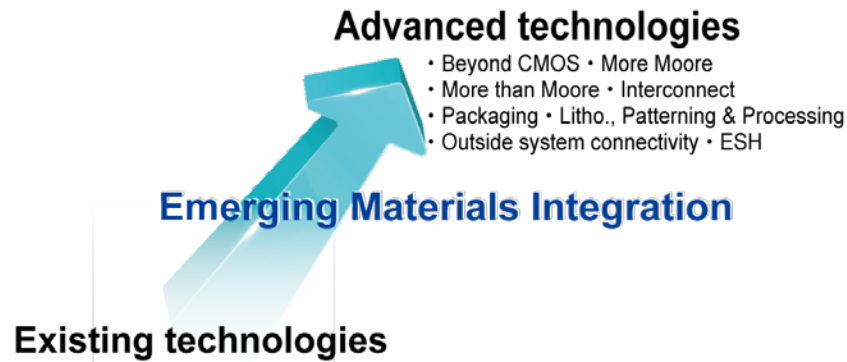


Figure EM11 Emerging Material Integration Promotes the Advancement of Existing Technologies

6.2. CHALLENGES

6.2.1. NEAR-TERM CHALLENGES

Table EM11 Near-term Difficult Challenges

<i>Near-Term Challenges: 2019-2024</i>	<i>Description</i>
<i>Materials and processes that achieve performance and power scaling of lateral fin- and nanowire FETs (Si, SiGe, Ge, III-V).</i>	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 V _{th} control and S/D formation in Si as well as alternate materials.
<i>Materials and processes that improve copper interconnect resistance and reliability</i>	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.
<i>Materials and processes for continued scaling of DRAM/SRAM and embedded NVM</i>	Low temperature materials for high performance vertical transistor memory select structures. High-k, low leakage DRAM dielectrics. Processes for stacking of 3D flash.

70 Emerging Materials Integration

<i>Materials and processes that extend lithography to sub-10 nm dimensions with reproducible properties</i>	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 through pattern rectification and pattern density multiplication.
<i>Materials for heterogeneous integration of multi-chip, multi-function packages.</i>	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.

6.2.1. LONG-TERM CHALLENGES

Table EMI2 Long-term Difficult Challenges

Long-term Challenges: 2025-3032	Description
<i>Materials and processes that achieve 3D monolithic and vertical integration of high mobility and steep subthreshold transistors</i>	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.
<i>Materials and processes that replace copper interconnects with improved reliability and electromagnetic performance at the nanoscale</i>	Synthesis or assembly of CNTs in predefined locations and directions with controlled diameters, chirality and site-density. Carbon and collective excitations. Novel interlayer dielectrics: MOF (Metal Organic Framework) and COF (Carbon Organic Framework). Metals with less size effects such as silicides.
<i>Materials and processes for charge-based and non-charge-based beyond CMOS logic that replaces or extends CMOS</i>	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 than kT at room temperature, ability to modulate bandgap efficiently with electric field. BiSFET heterostructures: achieving exciton condensation at room temperature.
<i>Materials and processes for emerging memory and select devices to replace DRAM/NVM.</i>	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.
<i>Materials and processes that enable monolithically 3D integrated complex functionality including thermal and yield challenges</i>	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 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 193nm 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. 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.

Table EMI5 Interconnect Materials

72 Emerging Materials Integration

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 EMI 9 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 (NEMS) 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. 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.

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^{1130,1131,1132,1133,1134,1135}.

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 most 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, as well as platforms that enable simultaneous measurement of complex nanoscopic properties, and modeling of probe-sample interactions. 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**

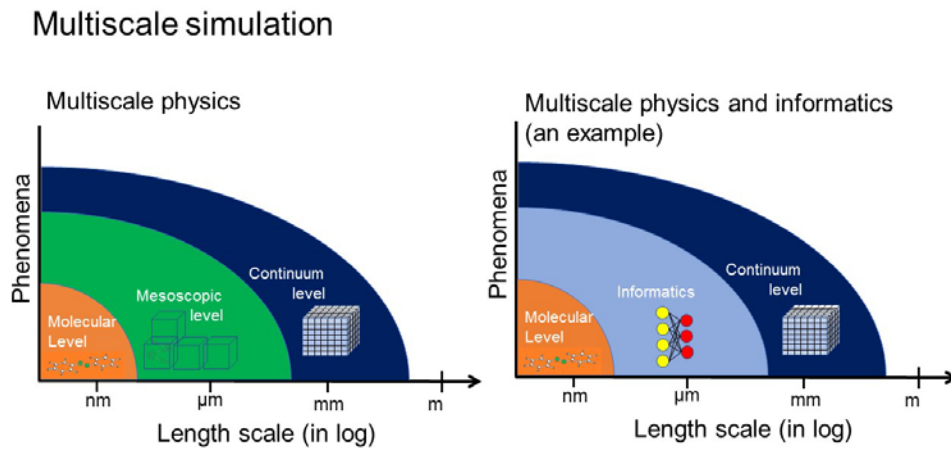
As mentioned at the beginning of the chapter, new system integrators, from mobile to data centers to the Internet of Everything, have appeared with new and complex technology requirements. These application domains require a highly interdisciplinary set of expertise, e.g. electrical and mechanical engineering; 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^{1136,1137,1138,1139}. The conductance change with the controllable decay characteristics was demonstrated in molecular gap atomic switches^{1140,1141}.

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. 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, that gives understanding of elastic behavior of crystals¹¹⁴². Meanwhile, a performance of fuel cell can be derived from macro-model originated from the rate equation of chemical reactions of individual molecules¹¹⁴³. Multiscale view of carrier transport in organic devices were treated by considering molecular orbital levels to mesoscopic scale of electron-hopping^{1144,1145,1146}. Beside the direct approach on multiscale physics, aid of machine learning can bridge the difference scales phenomena in composite material¹¹⁴⁷. These approaches will accelerate the simulation, while proper modeling with understanding of natural physics in difference scales of time and space are still necessary. 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 assess beyond-CMOS devices considered in this chapter against current CMOS technologies. Two methods of assessments have been reported previously in the ITRS ERD chapter: a "quantitative emerging device benchmarking" conducted by the Nanoelectronics Research Initiative (NRI) and a "survey-based assessment" conducted by the ERD working group.

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 5nm CMOS applications. This "beyond CMOS" chapter will update the quantitative benchmarking section with the latest NRI results.

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

An important issue regarding emerging charge-based nanoelectronic switch elements is related to the fundamental limits to the scaling of these new devices, and how they compare with CMOS technology at its projected end of scaling. An analysis¹¹⁴⁸

concludes that the fundamental limit of scaling an electronic charge-based switch is only a factor of 3 times smaller than the physical gate length of a silicon MOSFET in 2024. Furthermore, the density of these switches is limited by a maximum allowable power dissipation of approximately 100W/cm², and not by their size. The conclusion of this work is that MOSFET technology scaled to its practical limit in terms of size and power density will asymptotically reach the theoretical limits of scaling for charge-based devices.

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

7.1.1. ARCHITECTURAL REQUIREMENTS FOR A COMPETITIVE LOGIC DEVICE

The circuit designer and architect depend on the logic switch to exhibit specific desired characteristics in order to insure successful realization of a wide range of applications. These characteristics,¹¹⁴⁹ 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 switch¹¹⁵⁰)

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-security¹¹⁵¹ are particularly valuable.

7.2. NRI BEYOND-CMOS BENCHMARKING

The Nanoelectronics Research Initiative (<https://www.src.org/program/nri/>) has been benchmarking several diverse beyond-CMOS technologies, trying to balance the need for quantitative metrics to assess a new device concept's potential with the need to allow device research to progress in new directions which might not lend themselves to existing metrics.^{1152, 1153, 1154} Several of the more promising NRI devices have been described in detail in the Logic and Emerging Information Processing Device Section. Some results of this benchmark study were included in 2015 ITRS ERD chapter.

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

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 device¹¹⁵⁵ since power density is a primary concern for future CMOS scaling. The power-speed tradeoffs commonly observed in CMOS are also extended into the emerging devices. It is also important to understand the impact of transport delay for the different information tokens these devices employ. Communication with many non-charge tokens can be significantly slower than moving charge, although this may be balanced in some cases with lower energy for transport. The combination of the new balance between switch speed, switch area, and interconnect speed can lead to advantages in the span of control for a given technology. For some of the technologies (*e.g.*, nanomagnetic logic), there is no strong distinction between the switch and the interconnect, indicating the need for novel architecture to exploit unique attributes of these technologies.

A simplified 32-bit arithmetic logic unit (ALU) was built from these devices to evaluate their performance and the result is summarized in Figure BC7.1(a)¹¹⁵⁶. While tunneling devices (*e.g.*, TFET) show limited advantages over CMOS in terms of energy-delay product, most beyond-CMOS devices are inferior to CMOS in energy and/or delay. For example, the majority of spintronic devices are slower than CMOS and also show no energy advantage.

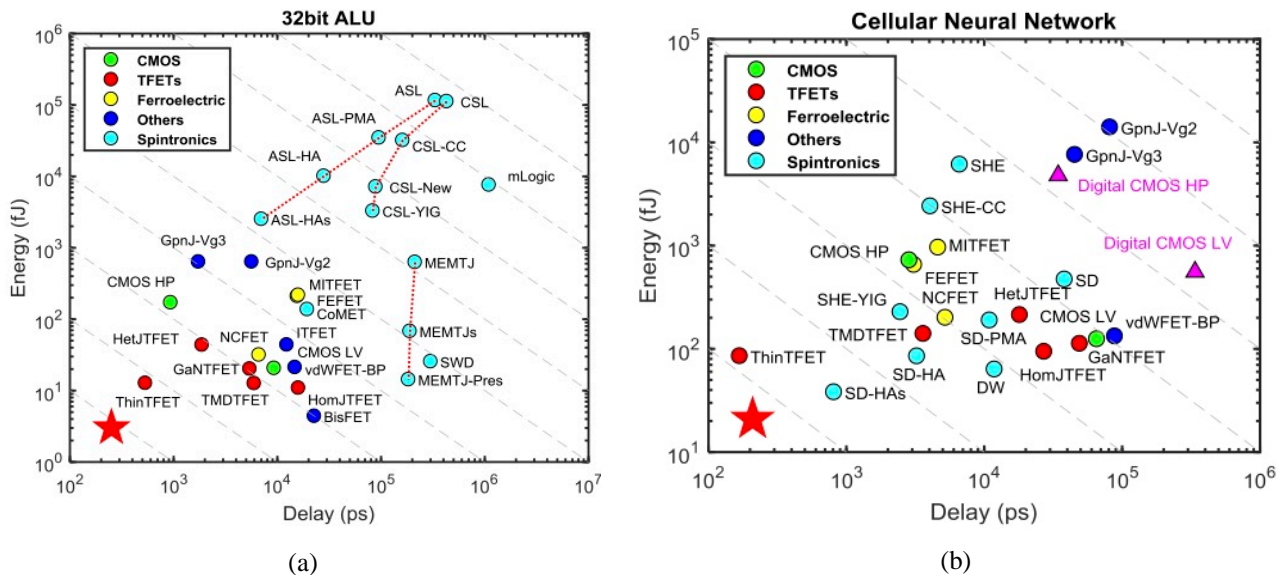


Figure BC7.1 (a) Energy versus Delay of a 32-bit ALU for a Variety of Charge- and Spin-based Devices; (b) Energy versus Delay per Memory Association Operation Using Cellular Neural Network (CNN) for a Variety of Charge- and Spin-based Devices¹¹⁵⁶

At the architecture level, the ability to speculate on how these devices will perform is still in its infancy. While the ultimate goal is to compare at a very high level – *e.g.*, how many MIPS can be produced for 100 mW in 1 mm²? – 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”¹¹⁵⁷ 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.¹¹⁵⁶ 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 benchmark studies and in the observations made during recent studies of beyond-CMOS replacement switches¹¹⁵⁸. A few noteworthy concepts:

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

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

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

7.3. ARCHIVE OF ITRS ERD SURVEY-BASED ASSESSMENT

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

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).

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

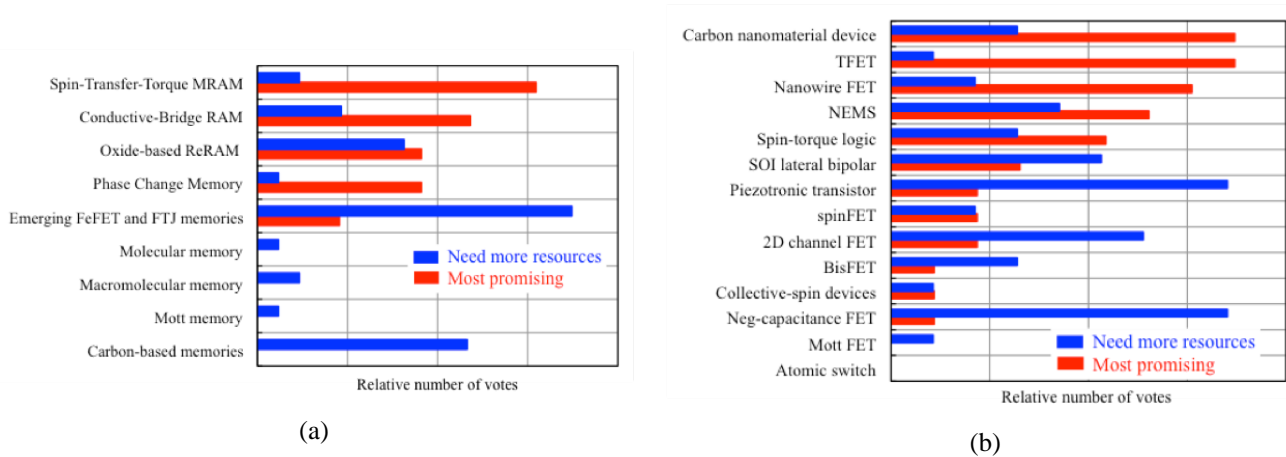


Figure BC7.2 (a) Survey of Emerging Memory Devices and (b) Survey of Emerging Logic Devices in 2014 ERD Emerging Logic Workshop (Albuquerque, NM)

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

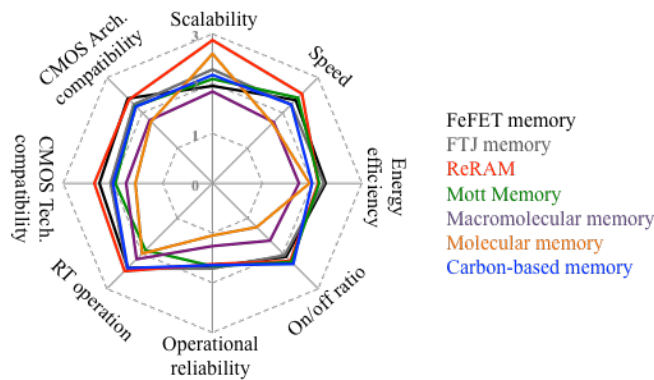


Figure BC7.3 Comparison of Emerging Memory Devices Based on 2013 Critical Review

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 current silicon memory technology which the new memory would displace. Performance potential for each criterion is assigned a value from 1–3, with “3” substantially exceeding ultimately-scaled CMOS, and “1” substantially inferior to CMOS or, again, a comparable existing memory technology. This evaluation is determined by a survey of the ERD Working Group members composed of individuals representing a broad range of technical backgrounds and expertise. Details of the assessment values are also included in 2013 ERD chapter.

Although this survey-based critical review has been conducted in ERD for several versions and has been widely cited in literatures, the decreasing number of votes of some less popular devices has raised concerns about the accuracy of some of the results. Figures BC7.3 and EBC7.4 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.) and several new technologies are added in this chapter (e.g., novel magnetic memory, transistor laser, etc.).

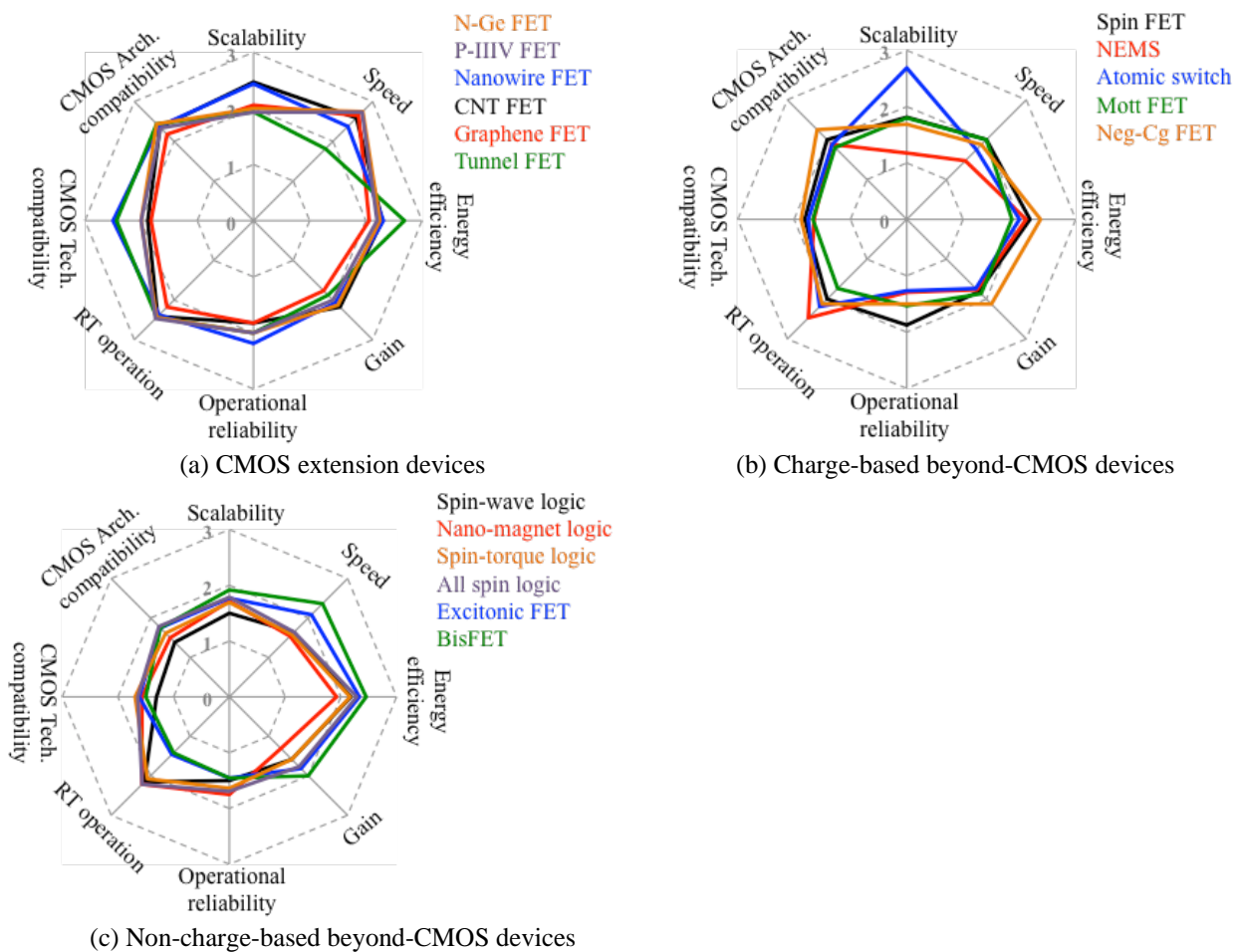


Figure BC7.4 Comparison of Emerging Logic Devices Based on 2013 ITRS ERD Critical Review: (a) CMOS Extension Devices; (b) Charge-based Beyond-CMOS Devices; (c) Non-charge-based Beyond-CMOS Devices

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

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 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.

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 μ 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. Kozyrakakis, 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-xPt_x," *Physical Review Applied*, vol. 10, Sep 6 2018.
- 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.
- 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. Koziol-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.*, "10x10nm² 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.
- 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.*, "10x10nm²Hf/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-protect ive 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. Gilbeto, 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 mm² 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/Pr_{0.7}Ca_{0.3}MnO₃ 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.
- 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 Y.Y. 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 WO_x 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 Valov, Iliia, *et al.* "Electrochemical metallization memories—fundamentals, applications, prospects." *Nanotechnology* 22.25 (2011): 254003.
- ¹¹¹ 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
- ¹¹² 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.
- ¹¹³ K. Tsutsui, "ReRAM for Fast Storage Application," Presented at August 2012 Flash Memory Summit Santa Clara, CA available at http://www.flashmemorysummit.com/English/Collaterals/Proceedings/2012/20120822_S203C_Tsutsui.pdf
- ¹¹⁴ Altis Semiconductor Press release on embedded CBRAM available at <http://www.altissemiconductor.com/en/index.php/about-altis/media-center/press-releases-menu/174-altis-ecbram>
- ¹¹⁵ Gopalan, C., *et al.* "Demonstration of conductive bridging random access memory (CBRAM) in logic CMOS process." *Solid-State Electronics* 58.1 (2011): 54-61.
- ¹¹⁶ Adesto Technologies, <http://www.adestotech.com/cbram>
- ¹¹⁷ 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.
- ¹¹⁸ 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.
- ¹¹⁹ Prall, Kirk, *et al.* "An Update on Emerging Memory: Progress to 2Xnm." Memory Workshop (IMW), 2012 4th IEEE International. IEEE, 2012.
- ¹²⁰ Sankaran, Kiroubanand, *et al.* "Modeling of Copper Diffusion in Amorphous Aluminum Oxide in CBRAM Memory Stack." *ECS Transactions* 45.3 (2012): 317-330.
- ¹²¹ 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.
- ¹²² 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.
- ¹²³ 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 optimized driving capability buffer," *Japanese Journal of Applied Physics* 58, SBBB04 (2019), pp. SBBB04-1-6, <https://doi.org/10.7567/1347-4065/aaf899>.
- ¹²⁴ 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.
- ¹²⁵ 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.
- ¹²⁶ 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)
- ¹²⁷ 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.
- ¹²⁸ W. Bai, R. Huang ; Y. Cai ; Y. Tang ; X. Zhang ; Y. Wang "Record Low-Power Organic RRAM With Sub-20-nA Reset Current," *IEEE Electron. Dev. Lett.* 34, 223-225, 2013.
- ¹²⁹ 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.
- ¹³⁰ 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.
- ¹³¹ 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.
- ¹³² 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.
- ¹³³ 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.
- ¹³⁴ P. Siebeneicher, H. Kleemann, K. Leo, and B. Lüssem, "Non-volatile organic memory devices comprising SiO₂ and C₆₀ showing 10⁴ switching cycles," *Appl. Phys. Lett.* 100, 193301 2012.
- ¹³⁵ 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.

- ¹³⁶ 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.
- ¹³⁷ 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.
- ¹³⁸ 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.
- ¹³⁹ 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.
- ¹⁴⁰ N. Tsutsumi, X. Bai, and W. Sakai, "Towards nonvolatile memory devices based on ferroelectric polymers," *AIP Advances* 2, 012154, 2012.
- ¹⁴¹ S.L. Miller and P.J. McWhorter, "Physics of the ferroelectric nonvolatile memory field effect transistor," *J. Appl. Phys.* 72, 5999 1992.
- ¹⁴² 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).
- ¹⁴³ 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).
- ¹⁴⁴ 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).
- ¹⁴⁵ T.S. Böske, 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.
- ¹⁴⁶ J. Müller, E. Yurchuk, T. Schlosser, J. Paul, R. Hoffmann, S. Müller, D. Martin, S. Slesazek, 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.
- ¹⁴⁷ T.P. Ma and J.-P. Han, "Why is nonvolatile ferroelectric memory field-effect transistor still elusive?" *IEEE Electron Device Lett.* 23, 386 (2002).
- ¹⁴⁸ 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).
- ¹⁴⁹ 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.).
- ¹⁵⁰ 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).
- ¹⁵¹ J. Müller, T.S. Böske, 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.
- ¹⁵² W. Zhang, M. Takahashi, and S. Sakai, "Electrical properties of CaxSr_{1-x}Bi₂Ta₂O₉ ferroelectric-gate field-effect transistors," *Semicond. Sci. Technol.* 28, 85003 (2013).
- ¹⁵³ 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).
- ¹⁵⁴ E. Yurchuk, S. Mueller, D. Martin, S. Slesazek, 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.
- ¹⁵⁵ 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).
- ¹⁵⁶ 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).
- ¹⁵⁷ 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
- ¹⁵⁸ M. Pešić, F. P. G. Fengler, L/. Larcher, A. Padovani, T. Schenk, E. D. Grimley, X. Sang, J. M. LeBeau, S. Slesazek, U. Schroeder, and T. Mikolajick, "Physical Mechanisms behind the Field-Cycling Behavior of HfO₂-Based Ferroelectric Capacitors," *Advanced Functional Materials* 26, 4601 (2016).
- ¹⁵⁹ H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, and S. Slesazek, "Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors," *ACS applied materials & interfaces* 9, 3792 (2017).
- ¹⁶⁰ M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Müller, S. Slesazek, 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.
- ¹⁶¹ 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.

- ¹⁶² M. Petic, 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.
- ¹⁶³ V. Garcia and M. Bibes, "Ferroelectric tunnel junctions for information storage and processing," *Nature Communications* 5 (2014).
- ¹⁶⁴ 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).
- ¹⁶⁵ 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).
- ¹⁶⁶ 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).
- ¹⁶⁷ 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).
- ¹⁶⁸ 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).
- ¹⁶⁹ 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).
- ¹⁷⁰ A. Chen, "Emerging memory selector devices," *13th Non-Volatile Memory Technology Symposium (NVMTS)*, Singapore (2013.), pp. 1-5.
- ¹⁷¹ 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).
- ¹⁷² 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
- ¹⁷³ 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).
- ¹⁷⁴ 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).
- ¹⁷⁵ 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).
- ¹⁷⁶ X. Tian, and A. Toriumi, "New opportunity of ferroelectric tunnel junction memory with ultrathin HfO₂-based oxides," *Electron Devices Technology and Manufacturing Conference (EDTM), 2017 IEEE* (pp. 36-64). IEEE.
- ¹⁷⁷ F. Ambriz-Vargas, G. Kolhatkar, M. Broyer, A. Hadj-Youssef, R. Nouar, A. Sarkissian, R. Thomas, C. Gomez-Yañez, M. A. Gauthier, A. Ruediger, "A Complementary Metal Oxide Semiconductor Process-Compatible Ferroelectric Tunnel Junction," *ACS applied materials & interfaces* 9, 13262 (2017).
- ¹⁷⁸ S. Fujii, Y. Kamimuta, T. Ino, Y. Nakasaki, R. Takaiishi, M. Saitoh, "First demonstration and performance improvement of ferroelectric HfO₂-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.
- ¹⁷⁹ Rebooting the IT Revolution: A Call for Action. SIA-SRC Report (2015): <https://www.semiconductors.org/clientuploads/Resourcer/RITR%20WEB%20version%20FINAL.pdf>
- ¹⁸⁰ M. Hilbert and P. Lopez, "The World's Technological Capacity to Store, Communicate, and Compute Information," *Science* 332 (2011) 60
- ¹⁸¹ G. M. Church, Y. Gao, K. Yuan, S. Kosuri, "Next-generation digital information storage in DNA," *Science* 337 (2012) 1628
- ¹⁸² N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipo, E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature* 494 (2013) 77
- ¹⁸³ 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
- ¹⁸⁴ 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
- ¹⁸⁵ J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, K. Strauss, "Toward a DNA-based archival storage system," *IEEE Micro* 37 (2017)
- ¹⁸⁶ V. Zhirnov, R. M. Zadehan, G. S. Sandhu, G. M. Church, W. L. Hughes, "Nucleic acid memory," *Nature Materials* 15 (2016) 366
- ¹⁸⁷ Y. Erlich and D. Zielinski, "DNA Fountain enables a robust and efficient storage architecture," *Science* 355 (2017) 950
- ¹⁸⁸ N. F. Mott, *Metal-Insulator Transitions*, 2nd ed. (Taylor & Francis, London, 1990)
- ¹⁸⁹ T. Oka and N. Nagaosa, "Interfaces of Correlated Electron Systems: Proposed Mechanism for Colossal Electroresistance," *Phys. Rev. Lett.* 95 (2005) 266403
- ¹⁹⁰ A. Asamitsu, Y. Tomioka, H. Kuwahara, and Y. Tokura, "Current switching of resistive states in magnetoresistive manganites," *Nature* 388 (1997) 50
- ¹⁹¹ 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
- ¹⁹² A. Sawa, T. Fujii, M. Kawasaki, and Y. Tokura, "Hysteretic current-voltage characteristics and resistance switching at a rectifying Ti/Pr_{0.7}Ca_{0.3}MnO₃ interface," *Appl. Phys. Lett.* 85 (2004) 4073

- ¹⁹³ 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
- ¹⁹⁴ 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
- ¹⁹⁵ 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
- ¹⁹⁶ 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
- ¹⁹⁷ 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
- ¹⁹⁸ 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
- ¹⁹⁹ F. Nakamura, M. Sakaki, Y. Yamanaka, S. Tamaru, T. Suzuki, and Y. Maeno, "Electric-field-induced metal maintained by current of the Mott insulator Ca₂RuO₄,," *Scientific Reports* 3 (2013) 2536
- ²⁰⁰ 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
- ²⁰¹ M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors,," *Nature Materials* 12 (2013) 114
- ²⁰² 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
- ²⁰³ 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
- ²⁰⁴ 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
- ²⁰⁵ 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
- ²⁰⁶ 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)
- ²⁰⁷ 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).
- ²⁰⁸ 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
- ²⁰⁹ 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).
- ²¹⁰ 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
- ²¹¹ A. Chen, "Nonlinearity and Asymmetry for Device Selection in Cross-bar Memory Arrays,," *IEEE Trans. Electron Dev.*, 62(9), 2857-2864, (2015)
- ²¹² L. Li, K. Lu, B. Rajendran, T. D. Happ, H-L. Lung, C. Lam, and M. Chan, "Driving Device Comparison for Phase-Change Memory,," *IEEE Trans. Electron. Dev.* 58 (2011) 664-671
- ²¹³ U. Gruening-von Schwerin, *Patent* DE 10 2006 040238 A1; *US Patent Application* "Integrated circuit having memory cells and method of manufacture,," US 2009/012758
- ²¹⁴ G. H. Kim, K. M. Kim, J. Y. Seok, H. J. Lee, D-Y. Cho, J. H. Han, and C. S. Hwang, "A theoretical model for Schottky diodes for excluding the sneak current in cross bar array resistive memory,," *Nanotechnology* 21 (2010) 385202
- ²¹⁵ H. Toda, "Three-dimensional programmable resistance memory device with a read/write circuit stacked under a memory cell array,," *Patent* (2009). US7606059
- ²¹⁶ P. Woerlee et al. "Electrical device and method of manufacturing therefore,," *Patent Application* (2005). WO 2005/124787 A2
- ²¹⁷ S. C. Puthentheradam, D. K. Schroder, M. N. Kozicki. "Inherent diode isolation in programmable metallization cell resistive memory elements,," *Appl. Phys. A* 102 (2011) 817-826
- ²¹⁸ J.H. Oh, et al, "Full integration of highly manufacturable 512Mb PRAM based on 90nm technology,," *IEDM Tech. Dig.*, pp. 515–518, Dec. 2006.
- ²¹⁹ Y. Sasago, et al, "Cross-point phase change memory with 4F₂ cell size driven by low-contact-resistivity poly-Si diode,," *Symposium VLSI Tech.*, pp. 24–25, Jun. 2009.
- ²²⁰ M. Kinoshita, et al, "Scalable 3-D vertical chain-cell-type phase-change memory with 4F₂ poly-Si diodes,," *Symposium VLSI Tech.*, pp. 35-36, Jun. 2012.
- ²²¹ S.H. Lee, et al, "Highly Productive PCRAM Technology Platform and Full Chip Operation: Based on 4F₂ (84nm Pitch) Cell Scheme for 1 Gb and Beyond,," *IEDM Tech. Dig.*, pp. 47-50, Dec. 2011.
- ²²² M.J. Lee, et al, "A low-temperature-grown oxide diode as a new switch element for high-density, nonvolatile memories,," *Adv. Mater.*, vol. 19, no. 1, pp. 73–76, Jan. 2007.

- ²²³ M.J. Lee, et al, “2-stack 1D-1R cross-point structure with oxide diodes as switch elements for high density resistance RAM applications,” *IEDM Tech. Dig.*, pp. 771–774, Dec. 2007.
- ²²⁴ M.J. Lee, et al, “Stack friendly all-oxide 3D RRAM using GaInZnO peripheral TFT realized over glass substrates,” *IEDM Tech. Dig.*, Dec. 2008.
- ²²⁵ S.E. Ahn, et al, “Stackable all-oxide-based nonvolatile memory with Al₂O₃ antifuse and p-CuOx/n-InZnOx diode,” *IEEE Electron Dev. Lett.*, vol. 30, no. 5, pp. 550–552, May 2009.
- ²²⁶ Y. Choi, et al, “High current fast switching n-ZnO/p-Si diode,” *J. Phys. D: Appl. Phys.*, vol. 43, pp. 345101-1-4, Aug. 2010.
- ²²⁷ S. Kim, Y. Zhang, J.P. McVittie, H. Jagannathan, Y. Nishi, and H.S.P. Wong, “Integrating phase-change memory cell with Ge nanowire diode for crosspoint memory—experimental demonstration and analysis,” *IEEE Trans. Electron Dev.*, vol. 55, no. 9, pp. 2307–2313, Sep. 2008.
- ²²⁸ Y.C. Shin, et al, “(In,Sn)2O3 /TiO2 /Pt Schottky-type diode switch for the TiO2 resistive switching memory array,” *Appl. Phys. Lett.*, vol. 92, no. 16, pp. 162904-1-3, Apr. 2008.
- ²²⁹ W.Y. Park, et al, “A Pt/TiO2/Ti Schottky-type selection diode for alleviating the sneak current in resistance switching memory arrays,” *Nanotechnology*, vol. 21, no. 19, pp. 195201-1-4, May 2010.
- ²³⁰ G.H. Kim, et al, “Schottky diode with excellent performance for large integration density of crossbar resistive memory,” *Appl. Phys. Lett.*, vol. 100, no. 21, pp. 213508-1-3, May 2012.
- ²³¹ N. Huby, et al, “New selector based on zinc oxide grown by low temperature atomic layer deposition for vertically stacked non-volatile memory devices,” *Microelectronic Eng.*, vol. 85, no. 12, pp. 2442-2444, Dec. 2008.
- ²³² B. Cho, et al, “Rewritable Switching of One Diode–One Resistor Nonvolatile Organic Memory Devices,” *Adv. Mater.*, vol. 22, no. 11, pp. 1228–1232, Mar. 2010.
- ²³³ M. J. Lee et al., “Two Series Oxide Resistors Applicable to High Speed and High Density Nonvolatile Memory,” *Adv. Mater.*, 19, 3919 (2007)
- ²³⁴ 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
- ²³⁵ D. Kau, et al, “A stackable cross point phase change memory,” *2009 IEDM*, 617
- ²³⁶ S. Kim, et al, “Ultrathin (<10 nm) Nb₂O₅/NbO₂ hybrid memory with both memory and selector characteristics for high density 3D vertically stackable RRAM applications,” *Symposium VLSI Tech.*, pp. 155-156, Jun. 2012.
- ²³⁷ W.G. Kim, et al, “NbO₂-based low power and cost effective 1S1R switching for high density cross point ReRAM application,” *VLSI Tech. Sym.*, 138 (2014).
- ²³⁸ J.H. Lee, et al, “Threshold switching in Si-As-Te thin film for the selector device of crossbar resistive memory,” *Appl. Phys. Lett.*, vol. 100, no. 12, pp. 123505-1-4, Mar. 2012.
- ²³⁹ M.J. Lee, et al, “Highly-Scalable Threshold Switching Select Device based on Chalcogenide Glasses for 3D Nanoscaled Memory Arrays,” *IEDM Tech. Dig.*, pp. 33–35, Dec. 2012.
- ²⁴⁰ S. Kim, et al, “Performance of threshold switching in chalcogenide glass for 3D stackable selector,” *VLSI Tech. Sym.*, 240 (2013).
- ²⁴¹ H. Yang, et al, “Novel selector for high density non-volatile memory with ultra-low holding voltage and 10⁷ on/off ratio,” *VLSI Tech. Sym.*, 130 (2015).
- ²⁴² S.H. Jo, et al, “3D-stackable crossbar resistive memory based on field assisted superliner threshold (FAST) selector,” *IEDM*, 160 (2014).
- ²⁴³ J.J. Huang, Y.M. Tseng, C.W. Hsu, and T.H. Hou, “Bipolar nonlinear Ni/TiO₂/Ni selector for 1S1R crossbar array applications,” *IEEE Electron Dev. Lett.*, vol. 32, no. 10, pp. 1427–1429, Oct. 2011.
- ²⁴⁴ J. Shin, et al, “TiO₂-based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application,” *J. Appl. Phys.*, vol. 109, no. 3, pp. 033712-1-4, Feb. 2011.
- ²⁴⁵ W. Lee, et al, “Varistor-type bidirectional switch (JMAX>107A/cm², selectivity~104) for 3D bipolar resistive memory arrays,” *Symposium VLSI Tech.*, pp. 37–38, Jun. 2012.
- ²⁴⁶ J. Woo, et al, “Multi-layer tunnel barrier (Ta₂O₅/TaO_x/TiO₂) engineering for bipolar RRAM selector applications,” *VLSI Tech. Sym.*, 168 (2013).
- ²⁴⁷ Y.H. Song, S.Y. Park, J.M. Lee, H.J. Yang, G.H. Kil, “Bidirectional two-terminal switching device for crossbar array architecture,” *IEEE Electron Dev. Lett.*, vol. 32, no. 8, pp. 1023–1025, Aug. 2011.
- ²⁴⁸ K. Gopalakrishnan, et al, “Highly Scalable Novel Access Device based on Mixed Ionic Electronic Conduction (MIEC) Materials for High Density Phase Change Memory (PCM) Arrays,” *2010 VLSI Symp.*, 205
- ²⁴⁹ R.S. Shenoy, et al. “Endurance and scaling trends of novel access-devices for multi-layer crosspoint memory based on mixed ionic electronic conduction (MIEC) materials,” *Symposium VLSI Tech.*, T5B-1, Jun. 2011.
- ²⁵⁰ G.W. Burr, et al, “Large-scale (512kbit) integration of multilayer-ready access-devices based on mixed-ionic-electronic-conduction (MIEC) at 100% yield,” *Symposium VLSI Tech.*, T5.4, Jun. 2012.
- ²⁵¹ K. Virwani, et al, “Sub-30 nm scaling and high-speed operation of fully-confined access-devices for 3D crosspoint memory based on mixed-ionic-electronic-conduction (MIEC) materials,” *IEDM Tech. Dig.*, pp. 36–39, Dec. 2012.
- ²⁵² E. Linn, R. Rosezin, C. Kugeler, and R. Waser, Complementary resistive switches for passive nanocrossbar memories, *NATURE MATERIALS* 9 (2010) 403–406
- ²⁵³ S. Tappertzhofen, et al., “Capacity based nondestructive readout for complementary resistive switches,” *Nanotech.*, vol. 22, no. 39, pp. 395203-1-7, Sep. 2011.
- ²⁵⁴ R. Rosezin, et al., “Integrated complementary resistive switches for passive high-density nanocrossbar arrays,” *IEEE Electron Dev. Lett.*, vol. 32, no. 2, pp. 191–193, Feb. 2011.
- ²⁵⁵ Y. Chai, et al., “Nanoscale bipolar and complementary resistive switching memory based on amorphous carbon,” *IEEE Trans. Electron Dev.*, vol. 58, no. 11, pp. 3933–3939, Nov. 2011.

- ²⁵⁶ S. Schmelzer, E. Linn, U. Bottger, and R. Waser, "Uniform complementary resistive switching in tantalum oxide using current sweeps," *IEEE Electron Dev. Lett.*, vol. 34, no. 1, pp. 114–116, Jan. 2013.
- ²⁵⁷ Y. C. Bae, et al., "Oxygen ion drift-induced complementary resistive switching in homo TiO_x/TiO_y/TiO_x and hetero TiO_x/TiON/TiO_x triple multilayer frameworks," *Adv. Funct. Materi.*, vol. 22, no. 4, pp. 709–716, Feb. 2012.
- ²⁵⁸ F. Nardi, S. Balatti, S. Larentis, and D. Ielmini, "Complementary switching in metal oxides: toward diode-less crossbar RRAMs," *IEDM Tech. Dig.*, pp. 709–712, Dec. 2011.
- ²⁵⁹ J. Lee, et al., "Diode-less nano-scale ZrO_x/HfO_x RRAM device with excellent switching uniformity and reliability for high-density cross-point memory applications," *IEDM Tech. Dig.*, pp. 452–455, Dec. 2010.
- ²⁶⁰ N. Banno, et al., "Nonvolatile 32 x 32 crossbar atom switch block integrated on a 65-nm CMOS platform," *Symposium VLSI Tech.*, pp. 39–40, Jun. 2012.
- ²⁶¹ X. Liu, et al., "Complementary resistive switching in niobium oxide-based resistive memory devices," *IEEE Electron Dev. Lett.*, vol. 34, no. 2, pp. 235–237, Feb. 2013.
- ²⁶² B. S. Simpkins, M. A. Mastro, C. R. Eddy, R. E. Pehrsson, "Surface depletion effects in semiconducting nanowires," *J. Appl. Phys.* 103 (2008) 104313
- ²⁶³ V. V. Zhirnov, R. Meade, R. K. Cavin, G. Sandhu, "Scaling limits of resistive memories," *Nanotechnology* 22 (2011) 254027
- ²⁶⁴ R. E. Fontana, Jr. G. M. Decad, and S. R. Hetzler, "The Impact of Areal Density and Millions of Square Inches (MSI) of Produced Memory on Petabyte Shipments of TAPE, NAND Flash, and HDD," *MSS&T 2013 Conference Proceedings*, May 2013.
- ²⁶⁵ Y. Deng and J. Zhou, "Architectures and optimization methods of flash memory based storage systems," *J. Syst. Arch.* 57 (2011) 214–227.
- ²⁶⁶ L. M. Grupp, A. M. Caulfield, J. Coburn, S. Swanson, E. Yaakobi, P. H. Siegel, J. K. Wolf "Characterizing Flash Memory: Anomalies, Observations, and Applications," *MICRO'09*, Dec. 12–16, 2009, New York, NY, USA, pp. 24–33
- ²⁶⁷ 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).
- ²⁶⁸ 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).
- ²⁶⁹ M. Franceschini, M. Qureshi, J. Karidis, L. Lastras, A. Bivens, P. Dube, and B. Abali, "Architectural Solutions for Storage-Class Memory in Main Memory," *CMRR Non-volatile Memories Workshop*, April 2010, http://nvmw.eng.ucsd.edu/2010/documents/Franceschini_Michael.pdf
- ²⁷⁰ M. Johnson, A. Al-Shamma, D. Bosch, M. Crowley, M. Farmwald, L. Fasoli, A. Ilkbahar, et al., "512-Mb PROM with a Three-Dimensional Array of Diode/Antifuse Memory Cells," *IEEE J. Solid-State Circ.* 38, No. 11, 1920–1928 (2003).
- ²⁷¹ M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable high performance main memory system using phase-change memory technology," *ISCA '09 – Proceedings of the 36th annual International Symposium on Computer Architecture*, pages 24–33, ACM, (2009).
- ²⁷² www.micron.com/about/innovations/3d-xpoint-technology
- ²⁷³ www.computerworld.com/article/2951869/computer-hardware/intel-and-micron-unveil-3d-xpoint-a-new-class-of-memory.html
- ²⁷⁴ <https://www.micron.com/products/advanced-solutions/3d-xpoint-technology>
- ²⁷⁵ <https://www.intel.com/content/www/us/en/architecture-and-technology/intel-micron-3d-xpoint-webcast.html>
- ²⁷⁶ www.eetimes.com/author.asp?section_id=36&doc_id=1327313
- ²⁷⁷ www.hpcwire.com/2015/08/27/micron-steers-roadmap-through-memory-scaling-course/
- ²⁷⁸ "Final Report, Exascale Study Group: Technology Challenges in Advanced Exascale Systems" (DARPA), 2007.
- ²⁷⁹ L.A. Barroso, and U. Holzle, "The case for Energy-Proportional Computing," *IEEE Computer*, Vol. 40(12), Dec. 2007, pp. 33–37.
- ²⁸⁰ 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).
- ²⁸¹ S. Swanson, "System architecture implications for M/S-class SCMs," *ITRS SCM workshop*, July 2012.
- ²⁸² M. Awasthi, M. Shevgoor, K. Sudan, B. Rajendran, R. Balasubramonian, V. Srinivasan, "Efficient scrub mechanism for error-prone emerging memories," *HPCA 2012*.
- ²⁸³ K. H. Kim, "Memory Interfaces for M-Class SCMs," *ITRS SCM workshop*, July 2012.
- ²⁸⁴ K. Qureshi, J. Karidis, M. Franceschini, V. Srinivasan, L. Lastras, and B. Abali, "Enhancing lifetime and security of pcm-based main memory with start-gap wear leveling," *MICRO 42: Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 14–23, ACM, (2009).
- ²⁸⁵ M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable high performance main memory system using phase-change memory technology," *ISCA '09 - Proceedings of the 36th annual International Symposium on Computer Architecture*, pages 24–33, ACM, (2009).
- ²⁸⁶ A. M. Caulfield, J. Coburn, T. Mollov, A. De, A. Akel, J. He, A. Jagatheesan, "Understanding the impact of emerging non-volatile memories on high-performance, io-intensive computing," *Proc. ACM/IEEE Int. Conf. High Perform. Comput., Netw., Storage Anal.*, 2010, pp. 1–11.
- ²⁸⁷ E. Kultursay, M. Kandemir, A. Sivasubramaniam, and O. Mutlu, "Evaluating STT-RAM as an energy-efficient main memory alternative," *Proceedings of the 2013 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, (2013).
- ²⁸⁸ H. Lee, "High-Performance NAND and PRAM Hybrid Storage Design for Consumer Electronics," *IEEE Trans. Consumer Electronics*, Vol. 56(1), 112–118 (2010).
- ²⁸⁹ P. Ranganathan, "From Microprocessors to Nanostores: Rethinking Data-Centric Systems," *COMPUTER* 44 (2011) 39–48.
- ²⁹⁰ J. Chang, "Data-centric computing and Nanostores," *ITRS SCM workshop*, July 2012.

- ²⁹¹ K.M. Bresniker, S. Singhal and R.S. Williams, “Adapting to thrive in a new economy of memory abundance,” *IEEE Computer*, Dec. 2015, pp. 44–53.
- ²⁹² D. Kim, K. Bang, S-H. Ha, S. Yoon, and E-Y. Chung, “Architecture exploration of high-performance PCs with a solid-state disk,” *IEEE Trans. Comp.* 59 (2010) 879–890
- ²⁹³ J. H. Yoon, E. H. Nam, Y. J. Seong, H. Kim, B. S. Kim, S. L. Min, Y. Cho, “Chameleon: A high performance Flash/FRAM hybrid solid state disk architecture,” *IEEE Comp. Arch. Lett.* 7 (2008) 17-20
- ²⁹⁴ H. G. Lee, “High-Performance NAND and PRAM Hybrid Storage Design for Consumer Electronics,” *IEEE Trans. Consumer Electronics*, Vol. 56(1), 112–118 (2010).
- ²⁹⁵ E. L. Miller, “Object-based interfaces for efficient and portable access to S-class SCMs,” *ITRS SCM workshop*, July 2012.
- ²⁹⁶ H. Zhang, G. Chen, B. C. Ooi, K.-L. Tan, and M. Zhang, “In-Memory Big Data Management and Processing: A Survey,” *IEEE Trans. Knowledge Data. Engr.*, 27(7), 1920–1948 (2015).
- ²⁹⁷ S. Chen, P. B. Gibbons, and S. Nath, “Rethinking database algorithms for phase change memory,” *Proc. CIDR*, 2011, pp. 21–31.
- ²⁹⁸ B.-D. Yang, J.-E. Lee, J.-S. Kim, J. Cho, S.-Y. Lee, and B.-G. Yu, “A low power phase-change random access memory using a data comparison write scheme,” *Proc. IEEE Int. Symp. Circuits Syst.*, 2007, pp. 3014–3017.
- ²⁹⁹ B. C. Lee, E. Ipek, O. Mutlu, and D. Burger, “Architecting phase change memory as a scalable dram alternative,” *Proc. 36th Annu. Int. Symp. Comput. Archit.*, 2009, pp. 2–3.
- ³⁰⁰ S. D. Viglas, “Write-limited sorts and joins for persistent memory,” *Proc. VLDB Endowment*, vol. 7, pp. 413–424, 2014.
- ³⁰¹ S. Chen and Q. Jin, “Persistent b+-trees in non-volatile main memory,” *Proc. VLDB Endowment*, vol. 8, pp. 786–797, 2015. J. Huang, K. Schwan, and M. K. Qureshi, “NVRAM-aware logging in transaction systems,” *Proc. VLDB Endowment*, vol. 8, pp. 389–400, 2014.
- ³⁰² T. Wang and R. Johnson, “Scalable logging through emerging non-volatile memory,” *Proc. VLDB Endowment*, vol. 7, pp. 865–876, 2014.
- ³⁰³ J. Huang, K. Schwan, and M. K. Qureshi, “NVRAM-aware logging in transaction systems,” *Proc. VLDB Endowment*, vol. 8, pp. 389–400, 2014.
- ³⁰⁴ A. Chatzistergiou, M. Cintra, and S. D. Viglas, “REWIND: Recovery write-ahead system for in-memory non-volatile data structures,” *Proc. VLDB Endowment*, vol. 8, pp. 497–508, 2015.
- ³⁰⁵ R. Fang, H.-I. Hsiao, B. He, C. Mohan, and Y. Wang, “High performance database logging using storage class memory,” *Proc. IEEE 27th Int. Conf. Data Eng.*, 2011, pp. 1221–1231.
- ³⁰⁶ J. Condit, E. B. Nightingale, C. Frost, E. Ipek, B. Lee, D. Burger, and D. Coetzee, “Better I/O through byte-addressable, persistent memory,” *Proc. ACM SIGOPS 22nd Symp. Operating Syst. Principles*, 2009, pp. 133–146.
- ³⁰⁷ A. Akel, A. M. Caulfield, T. I. Mollov, R. K. Gupta, and S. Swanson, “Onyx: a prototype phase change memory storage array,” *Hot Storage*, (2011).
- ³⁰⁸ Coburn, A. M. Caulfield, A. Akel, L. M. Grupp, R. K. Gupta, R. Jhala, and S. Swanson, “Nv-Heaps: Making Persistent Objects Fast and Safe with Next-Generation, Non-Volatile Memories,” *ACM Sigplan Notices*, 47(4), 105-117, (2012).
- ³⁰⁹ J. Coburn, A. M. Caulfield, A. Akel, L. M. Grupp, R. K. Gupta, R. Jhala, and S. Swanson, “Nv-Heaps: Making Persistent Objects Fast and Safe with Next-Generation, Non-Volatile Memories,” *ACM Sigplan Notices*, 47(4), 105-117, (2012).
- ³¹⁰ H. Volos, A. J. Tack, and M. M. Swift, “Mnemosyne: Lightweight persistent memory,” *Proc. 16th Int. Conf. Archit. Support Program. Languages Operating Syst.*, 2011, pp. 91–104.
- ³¹¹ I. Moraru, D. G. Andersen, M. Kaminsky, N. Tolia, P. Ranganathan, and N. Binkert, “Consistent, durable, and safe memory management for byte-addressable nonvolatile main memory,” presented at the *Conf. Timely Results Operating Syst. Held in Conjunction with SOSOP*, Farmington, PA, USA, 2013.
- ³¹² S. Venkataraman, N. Tolia, P. Ranganathan, and R. H. Campbell, “Consistent and durable data structures for non-volatile byte addressable memory,” *Proc. 9th USENIX Conf. File Storage Technol.*, 2011, p. 5.
- ³¹³ S. R. Dulloor, S. Kumar, A. Keshavamurthy, P. Lantz, D. Reddy, R. Sankaran, and J. Jackson, “System software for persistent memory,” *Proc. 9th Eur. Conf. Comput. Syst.*, 2014, pp. 15:1–15:15.
- ³¹⁴ J. Jung, Y. Won, E. Kim, H. Shin, and B. Jeon, “FRASH: Exploiting storage class memory in hybrid file system for hierarchical storage,” *ACM Trans. Storage*, vol. 6, pp. 3:1–3:25, 2010.
- ³¹⁵ A.-I. A. Wang, G. Kuenning, P. Reiher, and G. Popek, “The conquest file system: Better performance through a disk/persistent ram hybrid design,” *ACM Trans. Storage*, vol. 2, pp. 309–348, 2006.
- ³¹⁶ X. Wu and A. L. N. Reddy, “SCMFS: A file system for storage class memory,” *Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal.*, 2011, pp. 39:1–39:11.
- ³¹⁷ 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).
- ³¹⁸ 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).
- ³¹⁹ 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).
- ³²⁰ 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).
- ³²¹ 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).
- ³²² 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).

- ³²³ 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).
- ³²⁴ 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).
- ³²⁵ 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).
- ³²⁶ 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).
- ³²⁷ G. S. Tulevski, A. D. Franklin, D. Frank, J. M. Lobe, 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).
- ³²⁸ Ma, D. D., Lee, C. S., Au, F. C., Tong, S. Y., & Lee, S. T. (2003, Mar. 21), "Small-diameter silicon nanowire surfaces," *Science*, 299, 1874-1877.
- ³²⁹ Yan, H., & Yang, P., "Semiconductor nanowires: functional building blocks for nanotechnology," in P. Yang (Ed.), *The Chemistry of Nanostructured Materials*. River Edge, NJ: World Scientific (2004).
- ³³⁰ Beckman, R., Johnston-Halperin, E., Luo, Y., Green, J. E., & Hearh, J. R., "Bridging dimensions: demultiplexing ultrahigh-density nanowire circuits," *Science*, 310(5747), 465-468. (2005, Oct. 21).
- ³³¹ Chuang, S. et al., "Ballistic InAs Nanowire Transistors," *Nano Lett.* (2012). doi:10.1021/nl3040674
- ³³² Cui, Y., Lauhon, L. J., Gudiksen, M. S., Wang, J., & Lieber, C. M., "Diameter-controlled synthesis of single-crystal silicon nanowires," *Appl. Phys. Lett.*, 78(15), 2214-2216. (2001, Apr. 9).
- ³³³ Schmid, H.; Borg, M., Moselund, K., Gignac, L., Breslin, C. M., Bruley, J., Cutaia, D., Riel, H., *Appl. Phys. Lett.* 2015, 106, 233101.
- ³³⁴ Wu, Y., & Yang, P., "Germanium nanowire growth via simple vapor transport," *Chem. Mater.*, 12, 605-607. (2000).
- ³³⁵ Xiang, J., Lu, W., Hu, Y., Wu, Y., Yan, H., & Lieber, C. M., "Ge/Si nanowire heterostructures as high-performance field-effect transistors," *Nature*, 441, 489-493. (2006).
- ³³⁶ Yang, B., Buddharaju, K. D., Teo, S. H., Singh, N., Lo, G. Q., & Kwong, D. L., "Vertical silicon-nanowire formation and gate-all-around MOSFET," *IEEE Elect. Dev. Lett.*, 29(7), 791-794. (2008, Jul.).
- ³³⁷ Wernersson, L.-E., Thelander, C., Lind, E., & Samuelson, L., "III-V nanowires--extending a narrowing road," *Proc. IEEE*, 98(12), 2047-2060. (2010, Dec. 12).
- ³³⁸ Autran, J.-L., & Munteanu, D., "Beyond silicon bulk MOS transistor: new materials, emerging structures and ultimate devices," *Revue de l'Electricite et de l'Electronique*, 4, 25-37. (2007, Apr.).
- ³³⁹ Ng, H. T., Han, J., Yamada, T., Nguyen, P., Chen, Y. P., & Meyyappan, M., "Single crystal nanowire vertical surround-gate field-effect transistor," *Nano Lett.*, 4(7), 1247-1252. (2004).
- ³⁴⁰ Johansson, S., Memisevic, E., Wernersson, L.-E. & Lind, E., "High-Frequency Gate-All-Around Vertical InAs Nanowire MOSFETs on Si Substrates," *Electron Device Lett. IEEE* 35, 518 - 520 (2014).
- ³⁴¹ Berg, M. et al., "InAs nanowire MOSFETs in three-transistor configurations: single balanced RF down-conversion mixers," *Nanotechnology* 25, 485203 (2014).
- ³⁴² Yan, H., Choe, H. S., Nam, S., Hu, Y., Das, S., Klemic, J. F., et al., "Programmable nanowire circuits for nanoprocessors," *Nature*, 470, 240-244. (2011, Feb. 10).
- ³⁴³ Yao, J. et al., "Nanowire nanocomputer as a finite-state machine," *Proc. Natl. Acad. Sci.* 111, 2431-2435 (2014). doi:10.1073/pnas.1323818111
- ³⁴⁴ Fortuna, S. A., Wen, J., Chun, I., & Li, X., "Planar GaAs Nanowires on GaAs (100) Substrates: Self-Aligned, Nearly Twin-Defect Free, and Transfer-Printable," *Nano Lett.* 8, 4421-4427. (2008).
- ³⁴⁵ Miao, X. Chabak, K., Zhang, C., Mohseni, P. K., Walker Jr, D., & Li, X., "High-speed planar GaAs nanowire arrays with fmax > 75 GHz by wafer-scale bottom-up growth," *Nano Lett.* 15 (5), pp 2780-2786. (2015)
- ³⁴⁶ Lu, W., & Lieber, C. M., "Nanoelectronics from the bottom up," *Nature Materials*, 6, 841-850. (2007, Nov.).
- ³⁴⁷ Zhang, C. & Li, X., "III-V Nanowire Transistors for Low-Power Logic Applications: a Review and Outlook," *IEEE Transactions on Electron Devices*, 63(1), 223. (2016)
- ³⁴⁸ Mertens, H. et al., "Gate-All-Around MOSFETs based on Vertically Stacked Horizontal Si Nanowires in a Replacement Metal Gate Process on Bulk Si Substrates," *2016 Symposium on VLSI Technology Digest of Technical Papers*, pp. 158-159 (2016).
- ³⁴⁹ Mertens, H., Ritzenthaler, R., Chasin, A., et al., "Vertically stacked gate-all-around Si nanowire CMOS transistors with dual work function metal gates," *2016 IEEE International Electron Devices Meeting (IEDM)*. doi:10.1109/IEDM.2016.7838456 (2016)
- ³⁵⁰ Mertens, H. et al., "Vertically stacked gate-all-around Si nanowire transistors: Key Process Optimizations and Ring Oscillator Demonstration," *IEDM 37.4.1 - 37.4.4* (2017)
- ³⁵¹ Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D.; Zhang, Y., Dubonos, S. V., Grigorieva, I. V., Firsov, A. A., "Electric field effect in atomically thin carbon films," *Science* 2004, 306 (5696), 666-669.
- ³⁵² Chen, J.-H., Jang, C., Xiao, S., Ishigami, M., Fuhrer, M. S., "Intrinsic and extrinsic performance limits of graphene devices on SiO₂," *Nat Nano* 2008, 3 (4), 206-209.
- ³⁵³ Bolotin, K. I., Sikes, K. J., Jiang, Z., Klima, M., Fudenberg, G., Hone, J., Kim, P., Stormer, H. L., "Ultrahigh electron mobility in suspended graphene," *Solid State Commun.* 2008, 146 (9-10), 351-355.
- ³⁵⁴ Castro, E. V., Ochoa, H., Katsnelson, M. I., Gorbachev, R. V., Elias, D. C., Novoselov, K. S., Geim, A. K., Guinea, F., "Limits on Charge Carrier Mobility in Suspended Graphene due to Flexural Phonons," *Physical Review Letters* 2010, 105 (26), 266601.
- ³⁵⁵ Dean, C. R., Young, A. F., Meric, I., Lee, C., Wang, L., Sorgenfrei, S., Watanabe, K., Taniguchi, T., Kim, P., Shepard, K. L., Hone, J., "Boron nitride substrates for high-quality graphene electronics," *Nature Nanotechnology* 2010, 5 (10), 722-726.

- ³⁵⁶ Mayorov, A. S., Gorbachev, R. V., Morozov, S. V., Britnell, L., Jalil, R., Ponomarenko, L. A., Blake, P., Novoselov, K. S., Watanabe, K., Taniguchi, T., Geim, A. K., "Micrometer-Scale Ballistic Transport in Encapsulated Graphene at Room Temperature" *Nano Letters* 2011, *11* (6), 2396-2399.
- ³⁵⁷ Wang, L., Meric, I., Huang, P. Y., Gao, Q., Gao, Y., Tran, H., Taniguchi, T., Watanabe, K., Campos, L. M., Muller, D. A., Guo, J., Kim, P., Hone, J., Shepard, K. L., Dean, C. R., "One-Dimensional Electrical Contact to a Two-Dimensional Material," *Science* 2013, *342* (6158), 614-617.
- ³⁵⁸ Berger, C., Song, Z., Li, T., Li, X., Ogbazghi, A. Y., Feng, R., Dai, Z., Marchenkov, A. N., Conrad, E. H., First, P. N., de Heer, W. A., "Ultrathin Epitaxial Graphite: 2D Electron Gas Properties and a Route toward Graphene-based Nanoelectronics," *The Journal of Physical Chemistry B* 2004, *108* (52), 19912-19916.
- ³⁵⁹ Emtsev, K. V., Bostwick, A., Horn, K., Jobst, J., Kellogg, G. L., Ley, L., McChesney, J. L., Ohta, T., Reshanov, S. A., Rohrl, J., Rotenberg, E., Schmid, A. K., Waldmann, D., Weber, H. B., Seyller, T., "Towards wafer-size graphene layers by atmospheric pressure graphitization of silicon carbide," *Nat Mater* 2009, *8* (3), 203-207.
- ³⁶⁰ Bae, S., Kim, H., Lee, Y., Xu, X., Park, J.-S., Zheng, Y., Balakrishnan, J., Lei, T., Ri Kim, H., Song, Y. I., Kim, Y.-J., Kim, K. S., Özyilmaz, B., Ahn, J.-H., Hong, B. H., Iijima, S., "Roll-to-roll production of 30-inch graphene films for transparent electrodes," *Nature Nanotechnology* 2010, *5*, 574.
- ³⁶¹ Kondo, D., Sato, S., Yagi, K., Harada, N., Sato, M., Nihei, M., Yokoyama, N., "Low-Temperature Synthesis of Graphene and Fabrication of Top-Gated Field Effect Transistors without Using Transfer Processes," *Applied Physics Express* 2010, *3* (2), 025102.
- ³⁶² Li, X. S., Cai, W. W., An, J. H., Kim, S., Nah, J., Yang, D. X., Piner, R., Velamakanni, A., Jung, I., Tutuc, E., Banerjee, S. K., Colombo, L., Ruoff, R. S., "Large-Area Synthesis of High-Quality and Uniform Graphene Films on Copper Foils," *Science* 2009, *324* (5932), 1312-1314.
- ³⁶³ Yu, Q., Lian, J., Siriponglert, S., Li, H., Chen, Y. P., Pei, S.-S., "Graphene segregated on Ni surfaces and transferred to insulators," *Applied Physics Letters* 2008, *93* (11), 113103.
- ³⁶⁴ Hu, B., Ago, H., Ito, Y., Kawahara, K., Tsuji, M., Magome, E., Sumitani, K., Mizuta, N., Ikeda, K.-i., Mizuno, S., "Epitaxial growth of large-area single-layer graphene over Cu(111)/sapphire by atmospheric pressure CVD," *Carbon* 2014, *50* (1), 57-65.
- ³⁶⁵ Kondo, D., Nakano, H., Zhou, B., Kubota, I., Hayashi, K., Yagi, K., Takahashi, M., Sato, M., Sato, S., Yokoyama, N., "Intercalated multi-layer graphene grown by CVD for LSI interconnects," *2013 IEEE International Interconnect Technology Conference (IITC2013)*, Kyoto, Japan, Kyoto, Japan, 2013, p 190.
- ³⁶⁶ Banszerus, L., Schmitz, M., Engels, S., Dauber, J., Oellers, M., Haupt, F., Watanabe, K., Taniguchi, T., Beschoten, B., Stampfer, C., "Ultrahigh-mobility graphene devices from chemical vapor deposition on reusable copper," *Science Advances* 2015, *1* (6).
- ³⁶⁷ Baringhaus, J., Ruan, M., Edler, F., Tejada, A., Sicot, M., Taleb-Ibrahimi, A., Li, A.-P., Jiang, Z., Conrad, E. H., Berger, C., Tegenkamp, C., de Heer, W. A., "Exceptional ballistic transport in epitaxial graphene nanoribbons," *Nature* 2014, *506*, 349.
- ³⁶⁸ Shintaro, S., "Graphene for nanoelectronics," *Japanese Journal of Applied Physics* 2015, *54* (4), 040102.
- ³⁶⁹ McCann, E., "Asymmetry gap in the electronic band structure of bilayer graphene," *Physical Review B* 2006, *74* (16).
- ³⁷⁰ McCann, E., Fal'ko, V. I., "Landau-Level Degeneracy and Quantum Hall Effect in a Graphite Bilayer," *Physical Review Letters* 2006, *96* (8), 086805.
- ³⁷¹ Oostinga, J. B., Heersche, H. B., Liu, X. L., Morpurgo, A. F., Vandersypen, L. M. K., "Gate-induced insulating state in bilayer graphene devices," *Nature Materials* 2008, *7* (2), 151-157.
- ³⁷² Xia, F. N., Farmer, D. B., Lin, Y. M., Avouris, P., "Graphene Field-Effect Transistors with High On/Off Current Ratio and Large Transport Band Gap at Room Temperature," *Nano Letters* 2010, *10* (2), 715-718.
- ³⁷³ Zhang, Y. B., Tang, T. T., Girit, C., Hao, Z., Martin, M. C., Zettl, A., Crommie, M. F., Shen, Y. R., Wang, F., "Direct observation of a widely tunable bandgap in bilayer graphene," *Nature* 2009, *459* (7248), 820-823.
- ³⁷⁴ Kim, K. S., Walter, A. L., Moreschini, L., Seyller, T., Horn, K., Rotenberg, E., Bostwick, A., "Coexisting massive and massless Dirac fermions in symmetry-broken bilayer graphene," *Nature Materials* 2013, *12*, 887.
- ³⁷⁵ Fujita, M., Wakabayashi, K., Nakada, K., Kusakabe, K., "Peculiar localized state at zigzag graphite edge," *Journal of the Physical Society of Japan* 1996, *65* (7), 1920-1923.
- ³⁷⁶ Han, M. Y., Özyilmaz, B., Zhang, Y., Kim, P., "Energy Band-Gap Engineering of Graphene Nanoribbons," *Physical Review Letters* 2007, *98* (20), 206805.
- ³⁷⁷ Jiao, L. Y., Zhang, L., Wang, X. R., Diankov, G., Dai, H. J., "Narrow graphene nanoribbons from carbon nanotubes," *Nature* 2009, *458* (7240), 877-880.
- ³⁷⁸ Kosynkin, D. V., Higginbotham, A. L., Sinitskii, A., Lomeda, J. R., Dimiev, A., Price, B. K., Tour, J. M., "Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons," *Nature* 2009, *458* (7240), 872-U5.
- ³⁷⁹ Li, X. L., Wang, X. R., Zhang, L., Lee, S. W., Dai, H. J., "Chemically derived, ultrasoft graphene nanoribbon semiconductors," *Science* 2008, *319* (5867), 1229-1232.
- ³⁸⁰ Nakada, K., Fujita, M., Dresselhaus, G., Dresselhaus, M. S., "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Physical Review B* 1996, *54* (24), 17954-17961.
- ³⁸¹ Son, Y.-W., Cohen, M. L., Louie, S. G., "Energy Gaps in Graphene Nanoribbons," *Physical Review Letters* 2006, *97* (21), 216803.
- ³⁸² Yang, L., Park, C.-H., Son, Y.-W., Cohen, M. L., Louie, S. G., "Quasiparticle Energies and Band Gaps in Graphene Nanoribbons," *Physical Review Letters* 2007, *99* (18), 186801.
- ³⁸³ Zhu, X., Su, H., "Scaling of Excitons in Graphene Nanoribbons with Armchair Shaped Edges," *The Journal of Physical Chemistry A* 2011, *115* (43), 11998-12003.
- ³⁸⁴ Liang, X., Wi, S., "Transport Characteristics of Multichannel Transistors Made from Densely Aligned Sub-10 nm Half-Pitch Graphene Nanoribbons," *ACS Nano* 2012, *6* (11), 9700-9710.

- ³⁸⁵ Gallagher, P., Todd, K., “Goldhaber-Gordon, D., Disorder-induced gap behavior in graphene nanoribbons,” *Physical Review B* 2010, *81* (11), 115409.
- ³⁸⁶ Han, M. Y., Brant, J. C., Kim, P., “Electron Transport in Disordered Graphene Nanoribbons,” *Physical Review Letters* 2010, *104* (5), 056801.
- ³⁸⁷ Cai, J. M., Ruffieux, P., Jaafar, R., Bieri, M., Braun, T., Blankenburg, S., Muoth, M., Seitsonen, A. P., Saleh, M., Feng, X. L., Mullen, K., Fasel, R., “Atomically precise bottom-up fabrication of graphene nanoribbons,” *Nature* 2010, *466* (7305), 470-473.
- ³⁸⁸ Ruffieux, P., Cai, J., Plumb, N. C., Patthey, L., Prezzi, D., Ferretti, A., Molinari, E., Feng, X., Müllen, K., Pignedoli, C. A., Fasel, R., “Electronic Structure of Atomically Precise Graphene Nanoribbons,” *Acs Nano* 2012, *6* (8), 6930-6935.
- ³⁸⁹ Talirz, L., Ruffieux, P., Fasel, R., “On-Surface Synthesis of Atomically Precise Graphene Nanoribbons,” *Advanced Materials* 2016, *28* (29), 6222-6231.
- ³⁹⁰ Fiori, G., Iannaccone, G., “Simulation of graphene nanoribbon field-effect transistors,” *IEEE Electron Device Letters* 2007, *28* (8), 760-762.
- ³⁹¹ Harada, N., Sato, S., “Yokoyama, N., Theoretical Investigation of Graphene Nanoribbon Field-Effect Transistors Designed for Digital Applications,” *Japanese Journal of Applied Physics* 2013, *52*, 094301.
- ³⁹² Tsuchiya, H., Ando, H., Sawamoto, S., Maegawa, T., Hara, T., Yao, H., Ogawa, M., “Comparisons of Performance Potentials of Silicon Nanowire and Graphene Nanoribbon MOSFETs Considering First-Principles Bandstructure Effects,” *Electron Devices, IEEE Transactions on* 2010, *57* (2), 406-414.
- ³⁹³ Bennett, P. B., Pedramrazi, Z., Madani, A., Chen, Y.-C., Oteyza, D. G. d., Chen, C., Fischer, F. R., Crommie, M. F., Bokor, J., “Bottom-up graphene nanoribbon field-effect transistors,” *Applied Physics Letters* 2013, *103* (25), 253114.
- ³⁹⁴ Llinas, J. P., Fairbrother, A., Borin Barin, G., Shi, W., Lee, K., Wu, S., Yong Choi, B., Braganza, R., Lear, J., Kau, N., Choi, W., Chen, C., Pedramrazi, Z., Dumsloff, T., Narita, A., Feng, X., Müllen, K., Fischer, F., Zettl, A., Ruffieux, P., Yablonovitch, E., Crommie, M., Fasel, R., Bokor, J., “Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons,” *Nature Communications* 2017, *8* (1), 633.
- ³⁹⁵ Qiu, C., Zhang, Z., Peng, L. M., “Scaling carbon nanotube CMOS FETs towards quantum limit,” *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017, 2017, pp 5.5.1-5.5.4.
- ³⁹⁶ Zhao, P., Chauhan, J., Guo, J., “Computational Study of Tunneling Transistor Based on Graphene Nanoribbon,” *Nano Letters* 2009, *9* (2), 684-688.
- ³⁹⁷ Souma, S., Ueyama, M., Ogawa, M., “Simulation-based design of a strained graphene field effect transistor incorporating the pseudo magnetic field effect,” *Applied Physics Letters* 2014, *104* (21), 213505.
- ³⁹⁸ Quentin, W., Salim, B., David, T., Nguyen, V. H., Gwendal, F., Jean-Marc, B., Philippe, D., Bernard, P., “A Klein-tunneling transistor with ballistic graphene,” *2D Materials* 2014, *1* (1), 011006.
- ³⁹⁹ Novoselov, K. S., Jiang, D., Schedin, F., Booth, T. J., Khotkevich, V. V., Morozov, S. V., Geim, A. K., “Two-dimensional atomic crystals,” *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102* (30), 10451-10453.
- ⁴⁰⁰ 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.
- ⁴⁰¹ 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.
- ⁴⁰² Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., Kis, A., “Single-layer MoS_2 transistors,” *Nature Nanotechnology* 2011, *6*, 147.
- ⁴⁰³ Radisavljevic, B., Kis, A., “Mobility engineering and a metal-insulator transition in monolayer MoS_2 ,” *Nature Materials* 2013, *12*, 815.
- ⁴⁰⁴ Das, S., Appenzeller, J., “Where Does the Current Flow in Two-Dimensional Layered Systems?” *Nano Letters* 2013, *13* (7), 3396-3402.
- ⁴⁰⁵ 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.
- ⁴⁰⁶ Bao, W., Cai, X., Kim, D., Sridhara, K., Fuhrer, M. S., “High mobility ambipolar MoS_2 field-effect transistors: Substrate and dielectric effects,” *Applied Physics Letters* 2013, *102* (4), 042104.
- ⁴⁰⁷ Fang, H., Chuang, S., Chang, T. C., Takei, K., Takahashi, T., Javey, A., “High-Performance Single Layered WSe_2 p-FETs with Chemically Doped Contacts,” *Nano Letters* 2012, *12* (7), 3788-3792.
- ⁴⁰⁸ Desai, S. B., Madhvapathy, S. R., Sachid, A. B., Llinas, J. P., Wang, Q., Ahn, G. H., Pitner, G., Kim, M. J., Bokor, J., Hu, C., Wong, H.-S. P., Javey, A., “ MoS_2 transistors with 1-nanometer gate lengths,” *Science* 2016, *354* (6308), 99-102.
- ⁴⁰⁹ Sarkar, D., Xie, X., Liu, W., Cao, W., Kang, J., Gong, Y., Kraemer, S., Ajayan, P. M., Banerjee, K., “A subthermionic tunnel field-effect transistor with an atomically thin channel,” *Nature* 2015, *526*, 91.
- ⁴¹⁰ Britnell, L., Gorbachev, R. V., Jalil, R., Belle, B. D., Schedin, F., Mishchenko, A., Georgiou, T., Katsnelson, M. I., Eaves, L., Morozov, S. V., Peres, N. M. R., Leist, J., Geim, A. K., Novoselov, K. S., Ponomarenko, L. A., “Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures,” *Science* 2012, *335* (6071), 947-950.
- ⁴¹¹ Huang, C., Wu, S., Sanchez, A. M., Peters, J. J. P., Beanland, R., Ross, J. S., Rivera, P., Yao, W., Cobden, D. H., Xu, X., “Lateral heterojunctions within monolayer $MoSe_2$ - WSe_2 semiconductors,” *Nature Materials* 2014, *13*, 1096.
- ⁴¹² Yoshida, S., Kobayashi, Y., Sakurada, R., Mori, S., Miyata, Y., Mogi, H., Koyama, T., Takeuchi, O., Shigekawa, H., Microscopic basis for the band engineering of $Mo_{1-x}W_xS_2$ -based heterojunction,” *Scientific Reports* 2015, *5*, 14808.
- ⁴¹³ Gong, Y., Lin, J., Wang, X., Shi, G., Lei, S., Lin, Z., Zou, X., Ye, G., Vajtai, R., Yakobson, B. I., Terrones, H., Terrones, M., Tay, Beng K., Lou, J., Pantelides, S. T., Liu, Z., Zhou, W., Ajayan, P. M., “Vertical and in-plane heterostructures from WS_2/MoS_2 monolayers,” *Nature Materials* 2014, *13*, 1135.

- ⁴¹⁴ Li, M.-Y., Shi, Y., Cheng, C.-C., Lu, L.-S., Lin, Y.-C., Tang, H.-L., Tsai, M.-L., Chu, C.-W., Wei, K.-H., He, J.-H., Chang, W.-H., Suenaga, K., Li, L.-J., "Epitaxial growth of a monolayer WSe₂-MoS₂ lateral p-n junction with an atomically sharp interface," *Science* 2015, 349 (6247), 524-528.
- ⁴¹⁵ Mak, K. F., He, K., Shan, J., Heinz, T. F., "Control of valley polarization in monolayer MoS₂ by optical helicity," *Nature Nanotechnology* 2012, 7, 494.
- ⁴¹⁶ Xiao, D., Liu, G.-B., Feng, W., Xu, X., Yao, W., "Coupled Spin and Valley Physics in Monolayers of MoS₂ and Other Group-VI Dichalcogenides," *Physical Review Letters* 2012, 108 (19), 196802.
- ⁴¹⁷ Koma, A., Sunouchi, K., Miyajima, T., "Fabrication and characterization of heterostructures with subnanometer thickness," *Microelectronic Engineering* 1984, 2 (1), 129-136.
- ⁴¹⁸ 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.
- ⁴¹⁹ 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.
- ⁴²⁰ 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 disulfide," *Nature Materials* 2013, 12, 554.
- ⁴²¹ 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.
- ⁴²² 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.
- ⁴²³ 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.
- ⁴²⁴ 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.
- ⁴²⁵ 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.
- ⁴²⁶ 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 MoS₂ film formed by high-temperature sputtering for enhancement-mode nMOSFETs," *Japanese Journal of Applied Physics* 2015, 54 (4S), 04DN08.
- ⁴²⁷ Hayashi, K., Kataoka, M., Jippo, H., Ohfuchi, M., Iwai, T., Sato, S., "Two-dimensional SnS₂ 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.
- ⁴²⁸ 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.
- ⁴²⁹ 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.
- ⁴³⁰ A. C. Seabaugh and Z. Qin, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, pp. 2095-2110, 2010.
- ⁴³¹ 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.
- ⁴³² 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.
- ⁴³³ 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.
- ⁴³⁴ 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.
- ⁴³⁵ 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.
- ⁴³⁶ 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.
- ⁴³⁷ 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 I_{on} = 10 μA/μm for I_{off} = 1 nA/μm at V_{ds} = 0.3 V," *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 19.1.1-19.1.4.
- ⁴³⁸ 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.
- ⁴³⁹ 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.
- ⁴⁴⁰ 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.
- ⁴⁴¹ 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.

- ⁴⁴² 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.
- ⁴⁴³ 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.
- ⁴⁴⁴ 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.
- ⁴⁴⁵ T. Krishnamohan, K. Donghyun, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and $\ll 60\text{mV/dec}$ subthreshold slope," *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 2008, pp. 1–3.
- ⁴⁴⁶ 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.
- ⁴⁴⁷ 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.
- ⁴⁴⁸ 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.
- ⁴⁴⁹ 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.
- ⁴⁵⁰ S. Johnson and T. Tiedje, "Temperature dependence of the Urbach edge in GaAs," *Journal of Applied Physics*, vol. 78, pp. 5609-5613, 1995.
- ⁴⁵¹ 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.
- ⁴⁵² 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.
- ⁴⁵³ 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.
- ⁴⁵⁴ 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.
- ⁴⁵⁵ 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.
- ⁴⁵⁶ IRDS 2018 edition.
- ⁴⁵⁷ S. Datta and B. Das, "Electronic analog of the electro-optic modulator," *Appl. Phys. Lett.* vol. 56, pp.665–667, 1990.
- ⁴⁵⁸ 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.
- ⁴⁵⁹ S. Sugahara and J. Nitta, "Spin-Transistor Electronics: An Overview and Outlook," *Proc. IEEE*, vol. 98, pp. 2124–2154, 2010.
- ⁴⁶⁰ 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.
- ⁴⁶¹ 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.
- ⁴⁶² S. Yamamoto, and S. Sugahara, "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.
- ⁴⁶³ 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.
- ⁴⁶⁴ 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.
- ⁴⁶⁵ 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.
- ⁴⁶⁶ 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.
- ⁴⁶⁷ 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.
- ⁴⁶⁸ 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.
- ⁴⁶⁹ 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.
- ⁴⁷⁰ 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.
- ⁴⁷¹ I. Appelbaum, B. Huang, and D. J. Monsma, "Electronic measurement and control of spin transport in silicon," *Nature*, vol. 447, pp. 295–298, 2007.

- ⁴⁷² X. Lou, C. Adelman, 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.
- ⁴⁷³ 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.
- ⁴⁷⁴ 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.
- ⁴⁷⁵ 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.
- ⁴⁷⁶ 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.
- ⁴⁷⁷ 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.
- ⁴⁷⁸ 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.
- ⁴⁷⁹ 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.
- ⁴⁸⁰ 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.
- ⁴⁸¹ 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.
- ⁴⁸² 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.
- ⁴⁸³ T. Saito, N. Tezuka, M. Matsuura, and S. Sugimoto, "Spin injection, transport, and detection at room temperature in a lateral spin transport device with $\text{Co}_2\text{FeAl}_{0.5}\text{Si}_{0.5}/\text{n-GaAs}$ Schottky tunnel junctions," *Appl. Phys. Express*, vol. 6, pp. 103006/1-4, 2013.
- ⁴⁸⁴ 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.
- ⁴⁸⁵ T. A. Peterson, S. J. Patel, C. C. Geppert, K. D. Christie, A. Rath, D. Pennachio, M. E. Flattè, P. M. Voyles, C. J. Palmstrøm, 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.
- ⁴⁸⁶ 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 $\text{Co}_2\text{FeAl}_{0.5}\text{Si}_{0.5}$ Electrodes", *Phys. Rev. Appl.* vol. 8, pp. 014007/1-9 2017.
- ⁴⁸⁷ 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.
- ⁴⁸⁸ 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.
- ⁴⁸⁹ 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.
- ⁴⁹⁰ Y. Shuto, R. Nakane, W. Wang, H. Sukegawa, S. Yamamoto, M. Tanaka, K. Inomata, and S. Sugahara, "A New Spin-Functional Metal–Oxide–Semiconductor Field-Effect Transistor Based on Magnetic Tunnel Junction Technology: Pseudo-Spin-MOSFET," *Appl. Phys. Express*, vol. 3, pp. 013003/1-3, 2010.
- ⁴⁹¹ R. Nakane, Y. Shuto, H. Sukegawa, Z. C. Wen, S. Yamamoto, S. Mitani, M. Tanaka, K. Inomata, and S. Sugahara, "Fabrication of pseudo-spin-MOSFETs using a multi-project wafer CMOS chip," *Solid-State Elec.*, vol. 102, pp. 52–58, 2014.
- ⁴⁹² 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.
- ⁴⁹³ 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.
- ⁴⁹⁴ 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.
- ⁴⁹⁵ 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.
- ⁴⁹⁶ 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.
- ⁴⁹⁷ 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.
- ⁴⁹⁸ 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.
- ⁴⁹⁹ 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.

- ⁵⁰⁰ 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.
- ⁵⁰¹ 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.
- ⁵⁰² 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.
- ⁵⁰³ Kai-Shin Li et al., "Sub-60 mV-Swing Negative-Capacitance FinFET without Hysteresis," *Proceedings of IEDM*, 2015.
- ⁵⁰⁴ Krivokapic et al., "14nm Ferroelectric FinFET Technology with Steep Subthreshold Slope for Ultra Low Power Applications," *Proc. IEDM*, 2017.
- ⁵⁰⁵ K. Akarvardar et al., "Design considerations for complementary nanoelectromechanical logic gates," *IEEE International Electron Devices Meeting Technical Digest*, pp. 299–302, 2007.
- ⁵⁰⁶ F. Chen et al., "Integrated circuit design with NEM relays," *IEEE/ACM International Conference on Computer-Aided Design*, pp. 750–757, 2008.
- ⁵⁰⁷ 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.
- ⁵⁰⁸ 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.
- ⁵⁰⁹ 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.
- ⁵¹⁰ H. Fariborzi et al., "Analysis and demonstration of MEM-relay power gating," *Custom Integrated Circuits Conference*, 2010.
- ⁵¹¹ 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.
- ⁵¹² 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.
- ⁵¹³ J. O. Lee et al., "A sub-1-volt nanoelectromechanical switching device," *Nature Nanotechnology*, Vol. 8, pp. 36–40, 2013.
- ⁵¹⁴ Feng, X. L., et al. "Low voltage nanoelectromechanical switches based on silicon carbide nanowires." *Nano Letters*, vol.10, no. 8, pp. 2891-2896, 2010.
- ⁵¹⁵ 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.
- ⁵¹⁶ U. Zaghoul 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.
- ⁵¹⁷ R. Nathanael et al., "4-terminal relay technology for complementary logic," *IEEE International Electron Devices Meeting Technical Digest*, pp. 223–226, 2009.
- ⁵¹⁸ T.-J. K. Liu et al., "Prospects for MEM_relay logic switch technology," *IEEE International Electron Devices Meeting Technical Digest*, pp. 424–427, 2010.
- ⁵¹⁹ 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.
- ⁵²⁰ 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.
- ⁵²¹ 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.
- ⁵²² H. Fariborzi et al., "Design and demonstration of micro-electro-mechanical relay multipliers," *IEEE Asian Solid-State Circuits Conference*, 2011.
- ⁵²³ 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.
- ⁵²⁴ L. Hutin et al., "Electromechanical diode cell scaling for high-density nonvolatile memory," *IEEE Transactions on Electron Devices*, Vol. 61, pp. 1382–1387, 2014.
- ⁵²⁵ 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.
- ⁵²⁶ 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.
- ⁵²⁷ Y. Chen et al., "Reliability of MEM relays for zero leakage logic," *SPIE Conference Proceedings* Vol. 8614, 2013.
- ⁵²⁸ 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.
- ⁵²⁹ 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.
- ⁵³⁰ 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.

- ⁵³¹ 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)
- ⁵³² 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)
- ⁵³³ 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.
- ⁵³⁴ H. Akinaga, "Recent advances and future prospects in functional-oxide nanoelectronics: the emerging materials and novel functionalities that are accelerating semiconductor device research and development," *Japanese J. Appl. Phys.*, 52(100001), 2013.
- ⁵³⁵ Z. Yang, C. Ko and S. Ramanathan, "Oxide Electronics Utilizing Ultrafast Metal-Insulator Transitions," *Annual Review of Materials Research* 41, 8.1 (2011).
- ⁵³⁶ 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," *Phys. Rev. Lett.* 87, 237401 (2001)
- ⁵³⁷ S. Hormoz and S. Ramanathan, "Limits on vanadium oxide Mott metal-insulator transition field-effect transistors," *Solid State Electron* 54 (6), 654–659 (2010)
- ⁵³⁸ H. T. Kim, B. G. Chae, D. H. Youn, S. L. Maeng, G. Kim, K. Y. Kang and Y. S. Lim, "Mechanism and observation of Mott transition in VO₂-based two- and three-terminal devices," *New J Phys* 6, 52 (2004)
- ⁵³⁹ G. Stefanovich, A. Pergament and D. Stefanovich, J., "Electrical switching and Mott transition in VO₂," *Phys: Cond Mat*, 12, 8837 (2000)
- ⁵⁴⁰ D. Ruzmetov, G. Gopalakrishnan, C. Ko, V. Narayanamurti and S. Ramanathan, "Three-terminal field effect devices utilizing thin film vanadium oxide as the channel layer," *J Appl Phys* 107 (11), 114516 (2010)
- ⁵⁴¹ Z. Yang, Y. Zhou, and S. Ramanathan, J., "Studies on room-temperature electric-field effect in ionic-liquid gated VO₂ three-terminal devices," *Appl. Phys.* 111, 014506 (2012)
- ⁵⁴² M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa and Y. Tokura, "Collective bulk carrier delocalization driven by electrostatic surface charge accumulation," *Nature* 487, 459–462 (2012)
- ⁵⁴³ Y. Zhou and S. Ramanathan, "Relaxation dynamics of ionic liquid—VO₂ interfaces and influence in electric double-layer transistors," *J. Appl. Phys.* 111, 084508 (2012)
- ⁵⁴⁴ 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)
- ⁵⁴⁵ 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).
- ⁵⁴⁶ J. Shi, S. D. Ha, Y. Zhou, F. Schoofs and S. Ramanathan, "A correlated nickelate synaptic transistor," *Nature Communications*, 4, 2676 (2013)
- ⁵⁴⁷ 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),
- ⁵⁴⁸ M. D. Pickett, G. Medeiros-Ribeiro and R. Stanley Williams, "A scalable neuristor built with Mott memristors," *Nature Materials* 12, 114–117 (2013)
- ⁵⁴⁹ 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)
- ⁵⁵⁰ 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>
- ⁵⁵¹ Tingting Wei, Teruo Kanki Kouhei Fujiwara, Masashi Chikanari and Hidekazu Tanaka, "Electric field-induced transport modulation in VO₂ FETs with high-k oxide/ organic parylene-C hybrid gate dielectric," *Appl. Phys. Lett.* 108 (2016) 053503.
- ⁵⁵² Tingting Wei, Teruo Kanki, Masashi Chikanari, T. Uemura, Tsuyoshi Sekitani, and Hidekazu Tanaka, "Enhancement of electronic transport modulation in single crystalline VO₂ nanowire-based solid-state field-effect transistor," *Scientific Reports*, 7, 17215 (2017), <https://doi.org/10.1038/s41598-017-17468-x>
- ⁵⁵³ 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>
- ⁵⁵⁴ 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>
- ⁵⁵⁵ 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)
- ⁵⁵⁶ 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).
- ⁵⁵⁷ Y. Zhang and S. Ramanathan, "Analysis of "on" and "off" times for thermally driven VO₂ metal-insulator transition nanoscale switching devices," *Solid State Electronics*, August 2011, Vol. 62, pp. 161–164 (2011)
- ⁵⁵⁸ 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)
- ⁵⁵⁹ P. Lacorre, J. B. Torrance, J. Pannetier, A. I. Nazzari, P. W. Wang and T. C. Huang, J., "Synthesis, crystal structure, and properties of metallic PrNiO₃: Comparison with metallic NdNiO₃ and semiconducting SmNiO₃," *Sol. St. Chem.* 91, 225 (1991)

- ⁵⁶⁰ P.-H. Xiang, S. Asanuma, H. Yamada, H. Sato, I. H. Inoue, H. Akoh, A. Sawa, M. Kawasaki, and Y. Iwasa, "Electrolyte-gated SmCoO₃ thin-film transistors exhibiting thickness-dependent large switching ratio at room temperature," *Adv. Mat.*, **25**, 2158–2161, 2013.
- ⁵⁶¹ 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)
- ⁵⁶² S. Asanuma, P.-H. Xiang, H. Yamada, H. Sato, I. H. Inoue, H. Akoh, A. Sawa, K. Ueno, H. Shimotani, H. Yuan, M. Kawasaki, and Y. Iwasa, "Tuning of the metal-insulator transition in electrolyte-gated NdNiO₃ thin films," *Appl. Phys. Lett.* **97**, 142110 (2010)
- ⁵⁶³ 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).
- ⁵⁶⁴ 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)
- ⁵⁶⁵ S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabiny, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett.* **106**, 122102 (2015)
- ⁵⁶⁶ S. Bubel, A. J. Hauser, A. M. Glaudell, T. E. Mates, S. Stemmer and M. L. Chabiny, "The electrochemical impact on electrostatic modulation of the metal-insulator transition in nickelates," *Appl. Phys. Lett.* **106**, 122102 (2015)
- ⁵⁶⁷ 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)
- ⁵⁶⁸ M. Z. Hasan and C. L. Kane, "Colloquium: Topological insulators," *Reviews of Modern Physics* **82**, 3045-3067 (2010).
- ⁵⁶⁹ Hongming Weng, Rui Yu, Xiao Hu, Xi Dai & Zhong Fang, "Quantum anomalous Hall effect and related topological electronic states," *Advances in Physics* **64**, 227-282 (2015).
- ⁵⁷⁰ Yafei Ren, Zhenhua Qiao and Qian Niu, "Topological phases in two-dimensional materials: a review," *Rep. Prog. Phys.* **79**, 066501 (2016).
- ⁵⁷¹ Ching-Kai Chiu, Jeffrey C. Y. Teo, Andreas P. Schnyder and Shinsei Ryu, "Classification of topological quantum matter with symmetries," *Reviews of Modern Physics* **88**, 035005 (2016).
- ⁵⁷² Tiantian Zhang, Yi Jiang, Zhida Song, He Huang, Yuqing He, Zhong Fang, Hongming Weng and Chen Fang, "Catalogue of topological electronic materials," *Nature* **566**, 475–479 (2019).
- ⁵⁷³ M. G. Vergniory, L. Elcoro, Claudia Felser, Nicolas Regnault, B. Andrei Bernevig and Zhijun Wang, "A complete catalogue of high-quality topological materials," *Nature* **566**, 480–485 (2019).
- ⁵⁷⁴ Feng Tang, Hoi Chun Po, Ashvin Vishwanath & Xiangang Wan, "Comprehensive search for topological materials using symmetry indicators," *Nature* **566**, 486–489 (2019).
- ⁵⁷⁵ James L. Collins, Anton Tadich, Weikang Wu, Lidia C. Gomes, Joao N. B. Rodrigues, Chang Liu, Jack Hellerstedt, Hyejin Ryu, Shujie Tang, Sung-Kwan Mo, Shaffique Adam, Shengyuan A. Yang, Michael S. Fuhrer & Mark T. Edmonds, "Electric-field-tuned topological phase transition in ultrathin Na₃Bi," *Nature* **564**, 390–394 (2018).
- ⁵⁷⁶ F. Reis, G. Li, 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* **357**, 287-290 (2017).
- ⁵⁷⁷ C. L. Kane and E. J. Mele, "Z₂ Topological Order and the Quantum Spin Hall Effect," *Phys. Rev. Lett.* **95**, 146802 (2005).
- ⁵⁷⁸ L. Andrew Wray, "Topological transistor," *Nature Physics* **8**, 705–706 (2012).
- ⁵⁷⁹ Motohiko Ezawa, "Quantized conductance and field-effect topological quantum transistor in silicene nanoribbons," *Appl. Phys. Lett.* **102**, 172103 (2013).
- ⁵⁸⁰ Junwei Liu, Timothy H. Hsieh, Peng Wei, Wenhui Duan, Jagadeesh Moodera & Liang Fu, "Spin-filtered edge states with an electrically tunable gap in a two-dimensional topological crystalline insulator," *Nature Materials* **13**, 178–183 (2014).
- ⁵⁸¹ Xiaofeng Qian, Junwei Liu, Liang Fu, Ju Li, "Quantum spin Hall effect in two-dimensional transition metal dichalcogenides," *Science* **346**, 1344-1347 (2014).
- ⁵⁸² Hui Pan, Meimei Wu, Ying Liu & Shengyuan A. Yang, "Electric control of topological phase transitions in Dirac semimetal thin films," *Scientific Reports* **5**, 14639 (2015).
- ⁵⁸³ Qihang Liu, Xiuwen Zhang, L. B. Abdalla, Adalberto Fazzio, Alex Zunger, "Switching a Normal Insulator into a Topological Insulator via Electric Field with Application to Phosphorene," *Nano Lett.* **15**, 1222-1228 (2015).
- ⁵⁸⁴ Zuo Cheng Zhang, Xiao Feng, Jing Wang, Biao Lian, Jinsong Zhang, Cui Zu Chang, Minghua Guo, Yunbo Ou, Yang Feng, Shou-Cheng Zhang, Ke He, Xucun Ma, Qi-Kun Xue & Yayu Wang, "Magnetic quantum phase transition in Cr-doped Bi₂(SexTe_{1-x})₃ driven by the Stark effect," *Nature Nanotechnology* **12**, 953–957 (2017).
- ⁵⁸⁵ Alessandro Molle, Joshua Goldberger, Michel Houssa, Yong Xu, Shou-Cheng Zhang & Deji Akinwande "Buckled two-dimensional Xene sheets," *Nature Materials* **16**, 163–169 (2017).
- ⁵⁸⁶ William G. Vandenberghe & Massimo V. Fischetti, "Imperfect two-dimensional topological insulator field-effect transistors," *Nature Communications* **8**, 14184 (2017).
- ⁵⁸⁷ Yong Xu, Yan-Ru Chen, Jun Wang, Jun-Feng Liu, and Zhongshui Ma, "Quantized Field-Effect Tunneling between Topological Edge or Interface States," *Phys. Rev. Lett.* **123**, 206801 (2019).
- ⁵⁸⁸ A. M. Kadykov, F. Teppe, C. Consejo, L. Viti, M. S. Vitiello, S. S. Krishtopenko, S. Ruffenach, S. V. Morozov, M. Marcinkiewicz, W. Desrat, N. Dyakonova, W. Knap, V. I. Gavrilenko, N. N. Mikhailov, and S. A. Dvoretzky, "Terahertz detection of magnetic field-driven topological phase transition in HgTe-based transistors," *Appl. Phys. Lett.* **107**, 152101 (2015).
- ⁵⁸⁹ Motohiko Ezawa, "Topological Switch between Second-Order Topological Insulators and Topological Crystalline Insulators," *Phys. Rev. Lett.* **121**, 116801 (2018).

- ⁵⁹⁰ Shan Guan, Zhi-Ming Yu, Ying Liu, Gui-Bin Liu, Liang Dong, Yunhao Lu, Yugui Yao & Shengyuan A. Yang, “Artificial gravity field, astrophysical analogues, and topological phase transitions in strained topological semimetals,” *npj Quantum Materials* **2**, 23 (2017).
- ⁵⁹¹ S. S. Krishtopenko, I. Yahnuk, D. B. But, V. I. Gavrilenko, W. Knap, and F. Tepe, “Pressure- and temperature-driven phase transitions in HgTe quantum wells,” *Phys. Rev. B* **94**, 245402 (2016).
- ⁵⁹² Netanel H. Lindner, Gil Refael & Victor Galitski, “Floquet topological insulator in semiconductor quantum wells,” *Nature Physics* **7**, 490–495 (2011).
- ⁵⁹³ Cui-Zu Chang, Jinsong Zhang, Xiao Feng, Jie Shen, Zuocheng Zhang, Minghua Guo, Kang Li, Yunbo Ou, Pang Wei, Li-Li Wang, Zhong-Qing Ji, Yang Feng, Shuaihua Ji, Xi Chen, Jinfeng Jia, Xi Dai, Zhong Fang, Shou-Cheng Zhang, Ke He, Yayu Wang, Li Lu, Xu-Cun Ma, Qi-Kun Xue, “Experimental Observation of the Quantum Anomalous Hall Effect in a Magnetic Topological Insulator,” *Science* **340**, 167–170 (2013).
- ⁵⁹⁴ M. Serlin, C. L. Tschirhart, H. Polshyn, Y. Zhang¹, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, A. F. Young, “Intrinsic quantized anomalous Hall effect in a moiré heterostructure,” *Science* (2019) DOI: 10.1126/science.aay5533
- ⁵⁹⁵ Long Ju, Zhiwen Shi, Nityan Nair, Yinchuan Lv, Chenhao Jin, Jairo Velasco Jr, Claudia Ojeda-Aristizabal, Hans A. Bechtel, Michael C. Martin, Alex Zettl, James Analytis & Feng Wang, “Topological valley transport at bilayer graphene domain walls,” *Nature* **520**, 650–655 (2015).
- ⁵⁹⁶ Jing Li, Ke Wang, Kenton J. McFaul, Zachary Zern, Yafei Ren, Kenji Watanabe, Takashi Taniguchi, Zhenhua Qiao & Jun Zhu, “Gate-controlled topological conducting channels in bilayer graphene,” *Nature Nanotechnology* **11**, 1060–1065 (2016).
- ⁵⁹⁷ Jing Li, Rui-Xing Zhang, Zhenxi Yin, Jianxiao Zhang, Kenji Watanabe, Takashi Taniguchi, Chaoxing Liu, Jun Zhu, “A valley valve and electron beam splitter,” *Science* **362**, 1149–1152 (2018).
- ⁵⁹⁸ A. Khitun and K. Wang, *Superlattices & Microstructures* **38**, 184–200 (2005).
- ⁵⁹⁹ A. Khitun, M. Bao and K. L. Wang, *IEEE Transactions on Magnetics* **44** (9), 2141–2152 (2008).
- ⁶⁰⁰ M. P. Kostylev, A. A. Serga, T. Schneider, B. Leven, and B. Hillebrands, *Applied Physics Letters* **87**, 153501 (2005).
- ⁶⁰¹ K.-S. Lee and S.-K. Kim, *Journal of Applied Physics* **104**, 053909 (2008).
- ⁶⁰² G. Csaba, A. Papp, and W. Porod, *Physics Letters A* **381**, 1471–1476 (2017).
- ⁶⁰³ R. Verba, V. Tiberkevich, I. Krivorotov, and A. Slavin, *Physical Review Applied* **1**, 044006 (2014).
- ⁶⁰⁴ Y.-J. Chen et al., *Nano Letters* **17**, 572 (2017).
- ⁶⁰⁵ S. Dutta et al., *Scientific Reports* **5**, 09861 (2015).
- ⁶⁰⁶ K. Vogt et al., *Nature Communications* **5**, 1–5 (2014).
- ⁶⁰⁷ B. Divinskiy et al., *Advanced Materials* **30**, 1802837 (2018).
- ⁶⁰⁸ J. Han, P. Zhang, J. T. Hou, A. A. Siddiqui and L. Liu, *Science* **366**, 1121–1125 (2019).
- ⁶⁰⁹ R. Verba et al., *Phys. Rev. Applied* **11**, 054040 (2019).
- ⁶¹⁰ A. Khitun et al., *Journal of Nanoelectronics and Optoelectronics* **3** (1), 24–34 (2008).
- ⁶¹¹ A. Khitun and K. L. Wang, *Journal of Applied Physics* **110** (3) (2011).
- ⁶¹² S. Klingler et al., *Applied Physics Letters* **105**, 152410 (2014).
- ⁶¹³ A. Khitun, *Journal of Applied Physics* **113** (16) (2013).
- ⁶¹⁴ A. Khitun, *Journal of Applied Physics* **111** (5) (2012).
- ⁶¹⁵ K. Wagner et al., *Nature Nanotechnology* **11**, 432–436 (2016).
- ⁶¹⁶ 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).
- ⁶¹⁷ S. Cherepov et al., *Applied Physics Letters* **104** (8) (2014).
- ⁶¹⁸ F. Gertz, A. Kozhevnikov, Y. Filimonov and A. Khitun, *Magnetics, IEEE Transactions on Magnetics* **51** (4), 4002905–4400910 (2015).
- ⁶¹⁹ A. Kozhevnikov, F. Gertz, G. Dudko, Y. Filimonov and A. Khitun, *Applied Physics Letters* **106** (14), 142409–142405 (2015).
- ⁶²⁰ Y. Wang et al., *Science* **366**, 1125–1128 (2019).
- ⁶²¹ R. Lebrun et al., *Nature* **561**, 222–225 (2018).
- ⁶²² 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).
- ⁶²³ 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).
- ⁶²⁴ 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).
- ⁶²⁵ 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).
- ⁶²⁶ 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*.
- ⁶²⁷ M.M. Fogler, L.V. Butov, K.S. Novoselov, “Indirect excitons in van der Waals heterostructures,” *Nature Commun.* **5**, 4555 (2014).
- ⁶²⁸ 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*.
- ⁶²⁹ 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).

- ⁶³⁰ 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).
- ⁶³¹ 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.
- ⁶³² 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).
- ⁶³³ 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).
- ⁶³⁴ H.W. Then, M. Feng, and N. Holonyak, Jr., "The Transistor Laser: Theory and Experiment," *Proc. IEEE* 101, pp. 2271, (2013).
- ⁶³⁵ 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).
- ⁶³⁶ 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.
- ⁶³⁷ 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).
- ⁶³⁸ 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).
- ⁶³⁹ 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).
- ⁶⁴⁰ 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).
- ⁶⁴¹ 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.
- ⁶⁴² 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.
- ⁶⁴³ 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>.
- ⁶⁴⁴ 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.
- ⁶⁴⁵ P. A. Dowben, C. Binek, and D. E. Nikonov, in *Nanoscale Silicon Devices*; edited by Shuni Oda and David Ferry; Taylor and Francis (London) 2016, Chapter 11, pp 255-278.
- ⁶⁴⁶ P. A. Dowben, C. Binek, K. Zhang, L. Wang, W.-N. Mei, J. P. Bird, U. Singiseti, X. Hong, K. L. Wang, D. Nikonov, *IEEE J. Exploratory Solid-State Computational Devices and Circuits* 4, 1-9 (2018)
- ⁶⁴⁷ N. Sharma, J. P. Bird, Ch. Binek, P. A. Dowben, D. Nikonov, and A. Marshall, "Evolving Magneto-electric Device Technologies", *Semiconductor Science and Technology* (2019), in the press
- ⁶⁴⁸ C. Pan, A. Naeemi, *IEEE J. Explor. Solid-State Computat.* 4, 69-75 (2018)
- ⁶⁴⁹ N. Sharma, C. Binek, A. Marshall, J. P. Bird, P. A. Dowben and D. Nikonov, *2018 31st IEEE International System-on-Chip Conference (SOCC)*, Arlington, VA, USA, IEEE Explore 146-151 (2018)
- ⁶⁵⁰ N. Sharma, C. Binek, A. Marshall, J. P. Bird, P. A. Dowben, D. Nikonov, "Compact Modeling and Design of Magneto-electric Transistor Devices and Circuits", *IEEE International System-on-Chip Conference (SOCC)*, 2018, p146-151
- ⁶⁵¹ Ch. Binek, B. Doudin, *J. Phys. Condens. Matter* 17, L39-L44 (2005).
- ⁶⁵² M. Bibes, A. Barthélémy, *Nature Mat.* 7, 425-426 (2008).
- ⁶⁵³ M. G. Mankalale, Z. Liang, Z. Zhao, C. H. Kim, J.-P. Wang, and S. S. Sapatnekar, *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* 3, 27 (2017)
- ⁶⁵⁴ M. Kazemi, *Scientific Reports* 7, 15358 (2017)
- ⁶⁵⁵ S. Manipatruni, D. E. Nikonov, R. Ramesh, H. Li, I. A. Young, arXiv.org, (2015, 2017); arXiv:1512.05428
- ⁶⁵⁶ M. Street, Will Echtenkamp, Takashi Komesu, Shi Cao, P. A. Dowben, and Ch. Binek, *Applied Physics Letters* 104, 222402 (2014)
- ⁶⁵⁷ X. He, Y. Wang, N. Wu, A. N. Caruso, E. Vescovo, K. D. Belashchenko, P. A. Dowben, Ch. Binek, *Nature Materials* 9, 579 (2010).
- ⁶⁵⁸ N. Wu, Xi He, A. Wysocki, U. Lanke, T. Komesu, K.D. Belashchenko, Ch. Binek, P.A. Dowben, *Phys. Rev. Lett.* 106, 087202 (2011).
- ⁶⁵⁹ A. Lipatov, M. Loes, H. Lu, J. Dai, P. Patoka, D. S. Muratov, G. Ulrich, B. Kästner, A. Hoehl, G. Ulm, G. G. Batir, X. C. Zeng, E. Rühl, A. Gruverman, P. A. Dowben, A. Sinitskii, *ACS Nano* 12 (2018) 12713–12720
- ⁶⁶⁰ D. E. Nikonov, C. Binek, X. Hong, J. P. Bird, L. Wang, P. A. Dowben, "Anti-Ferromagnetic Magneto-electric Spin-Orbit Read Logic". U.S. Patent Application No.: 62/460,164; filed February 17, 2017
- ⁶⁶¹ Yu-Jin Chen, H. K. Lee, R. Verba, J. A. Katine, I. Barsukov, V. Tiberkevich, J. Xiao A. Slavin, Ilya Krivorotov, *Nano Letters* 17, 572 (2017)
- ⁶⁶² R. Verba, M. Carpentieri., G. Finocchio, V. Tiberkevich, A. Slavin, *Physical Review Applied* 7, 064023 (2017).
- ⁶⁶³ 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).
- ⁶⁶⁴ 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)
- ⁶⁶⁵ J. A. Currivan-Incorvia, S. Siddiqui, S. Dutta, E. R. Everts, 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).

- ⁶⁶⁶ 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).
- ⁶⁶⁷ 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).
- ⁶⁶⁸ 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).
- ⁶⁶⁹ D. M. Bromberg, J. Zhu, L. Pileggi, V. Sokalski, M. Moneck & P. C. Richardson. (12) United States Patent.
- ⁶⁷⁰ 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
- ⁶⁷¹ D. E. Nikonov & I. A. Young. Overview of beyond-CMOS devices and a uniform methodology for their benchmarking. *Proc. IEEE* **101**, 2498–2533 (2013).
- ⁶⁷² 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).
- ⁶⁷³ D. E. Nikonov & I. A. Young. Uniform methodology for benchmarking beyond-CMOS logic devices. *Electron Devices Meet. (IEDM), 2012 IEEE Int.* 24–25 (2012).
- ⁶⁷⁴ 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*
- ⁶⁷⁵ 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).
- ⁶⁷⁶ 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).
- ⁶⁷⁷ 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).
- ⁶⁷⁸ 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).
- ⁶⁷⁹ 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
- ⁶⁸⁰ D. Blaauw & Q. Dong. Binary-threshold, Rectified-linear and Recurrent Neural Networks Built with Spintronic Devices. *SPIN Neuron DAC (1).pdf* (2016).
- ⁶⁸¹ 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
- ⁶⁸² 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).
- ⁶⁸³ 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).
- ⁶⁸⁴ 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).
- ⁶⁸⁵ 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).
- ⁶⁸⁶ 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).
- ⁶⁸⁷ 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).
- ⁶⁸⁸ 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
- ⁶⁸⁹ 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)
- ⁶⁹⁰ Cowburn, R. P. & Welland, M. E. Room temperature magnetic quantum cellular automata. *Science (80-.)*. **287**, 1466–1468 (2000).
- ⁶⁹¹ 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).
- ⁶⁹² 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).
- ⁶⁹³ Fischer, T., Kewenig, M., Bozhko, D. A., Serga, A. A., Syvorotka, I. I., Ciubotaru, F., Adelman, C., Hillebrands, B. & Chumak, A. V. Experimental prototype of a spin-wave majority gate. *Appl. Phys. Lett.* **110**, (2017).
- ⁶⁹⁴ 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).

- ⁶⁹⁵ 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).
- ⁶⁹⁶ 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).
- ⁶⁹⁷ Nikonov, D. E., Bourianoff, G. I. & Ghani, T. Proposal of a spin torque majority gate logic. *IEEE Electron Device Lett.* **32**, 1128–1130 (2011).
- ⁶⁹⁸ 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
- ⁶⁹⁹ 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
- ⁷⁰⁰ 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).
- ⁷⁰¹ Breikreutz, 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).
- ⁷⁰² 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](https://arxiv.org/abs/1901.10327) [cs.ET], 2019.
- ⁷⁰³ 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](https://arxiv.org/abs/1806.10183) [cs.ET], 2018.
- ⁷⁰⁴ 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>.
- ⁷⁰⁵ E. P. DeBenedictis, “New Design Principles for Cold, Scalable Electronics,” technical report, Nov. 2019. http://www.zettaflops.org/CATC/DPfC_52ver4.pdf.
- ⁷⁰⁶ M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Ed., Cambridge University Press, 2011.
- ⁷⁰⁷ NIST, *Quantum Algorithm Zoo*, website, online at: <https://math.nist.gov/quantum/zoo/>. Accessed: Nov. 5, 2018.
- ⁷⁰⁸ 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.
- ⁷⁰⁹ 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.
- ⁷¹⁰ C. Mead, “Neuromorphic electronic systems,” *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct 1990. doi: 10.1109/5.58356
- ⁷¹¹ 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.
- ⁷¹² 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.
- ⁷¹³ 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.
- ⁷¹⁴ 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.
- ⁷¹⁵ 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.
- ⁷¹⁶ S. Suh, A. Basu, C. Schlotmann, 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.
- ⁷¹⁷ 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.
- ⁷¹⁸ 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.
- ⁷¹⁹ M. J. Marinella *et al.*, “Multiscale Co-Design Analysis of Energy, Latency, Area, and Accuracy of a ReRAM Analog Neural Training Accelerator,” *arXiv preprint arXiv:1707.09952*, 2017.
- ⁷²⁰ 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.
- ⁷²¹ 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.
- ⁷²² S. Agarwal *et al.*, “Resistive memory device requirements for a neural algorithm accelerator,” *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 929–938.
- ⁷²³ 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.

- ⁷²⁴ 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.
- ⁷²⁵ 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
- ⁷²⁶ 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
- ⁷²⁷ 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
- ⁷²⁸ 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.
- ⁷²⁹ 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.
- ⁷³⁰ 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.
- ⁷³¹ 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.
- ⁷³² E. J. Fuller *et al.*, "Li-Ion Synaptic Transistor for Low Power Analog Computing," *Advanced Materials*, vol. 29, no. 4, p. 1604310, 2017.
- ⁷³³ 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.
- ⁷³⁴ S. Agarwal *et al.*, "Achieving ideal accuracies in analog neuromorphic computing using periodic carry," *VLSI Technology, 2017 Symposium on*, 2017, pp. T174–T175: IEEE.
- ⁷³⁵ 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.
- ⁷³⁶ Gokmen, Tayfun, and Wilfried Haensch. "Algorithm for Training Neural Networks on Resistive Device Arrays." *arXiv preprint arXiv:1909.07908* (2019).
- ⁷³⁷ 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.
- ⁷³⁸ S. Agarwal *et al.* (2017). *CrossSim*. Available: <http://cross-sim.sandia.gov>.
- ⁷³⁹ T. Gokmen and Y. Vlasov, "Acceleration of deep neural network training with resistive cross-point devices," *arXiv preprint arXiv:1603.07341*, 2016.
- ⁷⁴⁰ 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.
- ⁷⁴¹ 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.
- ⁷⁴² 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.
- ⁷⁴³ Jennifer Hasler, "Large-Scale Field-Programmable Analog Arrays," *IEEE Proceedings*, Dec 2019
- ⁷⁴⁴ 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.
- ⁷⁴⁵ 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.
- ⁷⁴⁶ S. Kim, S. Shah, and J. Hasler, "Calibration of Floating-Gate SoC FPAA System," *IEEE Transactions on VLSI*, 2017.
- ⁷⁴⁷ 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.
- ⁷⁴⁸ 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.
- ⁷⁴⁹ 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.
- ⁷⁵⁰ J. Hasler and H. Wang, "A Fine-Grain FPAA fabric for RF+Baseband," *GOMAC*, March 2015.
- ⁷⁵¹ 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.
- ⁷⁵² 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>
- ⁷⁵³ I. Richter *et al.*, "Memristive Accelerator for Extreme Scale Linear Solvers," *40th Annual GOMACTech Conference*, St. Louis, MO, 2015.
- ⁷⁵⁴ 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).
- ⁷⁵⁵ 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.

- ⁷⁵⁶ Sun, Z., Pedretti, G., Bricalli, A., Ielmini, D. One-step regression and classification with crosspoint resistive memory arrays. *Sci. Adv.* (2019). Accepted for publication
- ⁷⁵⁷ 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.
- ⁷⁵⁸ Pagiamtzis, Kostas and Sheikholeslami, Ali, "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey," *IEEE Journal of Solid-State Circuits* **41**, 712 (2006).
- ⁷⁵⁹ Chang, Meng-Fan, et al. "17.5 A 3T1R nonvolatile TCAM using MLC ReRAM with sub-1ns search time." *2015 IEEE International Solid-State Circuits Conference-(ISSCC) Digest of Technical Papers*. IEEE, 2015
- ⁷⁶⁰ Lin, Chien-Chen, et al. "7.4 A 256b-wordlength ReRAM-based TCAM with 1ns search-time and 14× improvement in wordlength-energy efficiency-density product using 2.5 T1R cell." *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 2016.
- ⁷⁶¹ Li, Jing, et al. "1Mb 0.41 μm 2 2T-2R cell nonvolatile TCAM with two-bit encoding and clocked self-referenced sensing." *2013 Symposium on VLSI Technology*. IEEE, 2013.
- ⁷⁶² L. Zheng, S. Shin, and S.-M. S. Kang, "Memristor-based ternary content addressable memory (mtcam) for data-intensive computing," *Semiconductor Science and Technology*, vol. 29, no. 10, p. 104010, 2014
- ⁷⁶³ Chang, Meng-Fan, et al. "A ReRAM-based 4T2R nonvolatile TCAM using RC-filtered stress-decoupled scheme for frequent-OFF instant-ON search engines used in IoT and big-data processing." *IEEE Journal of Solid-State Circuits* 51.11 (2016): 2786-2798.
- ⁷⁶⁴ Matsunaga, Shinichiro, et al. "Fabrication of a 99%-energy-less nonvolatile multi-functional CAM chip using hierarchical power gating for a massively-parallel full-text-search engine." *2013 Symposium on VLSI Technology*. IEEE, 2013.
- ⁷⁶⁵ Grossi, Alessandro, et al. "Experimental investigation of 4-kb RRAM arrays programming conditions suitable for TCAM." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26.12 (2018): 2599-2607.
- ⁷⁶⁶ 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.
- ⁷⁶⁷ L. Yavits, S. Kvatinisky, A. Morad, and R. Ginosar, "Resistive Associative Processor," *IEEE Computer Architecture Letters*, vol. 14, no. 2, pp. 148–151, 2015.
- ⁷⁶⁸ 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.
- ⁷⁶⁹ 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.
- ⁷⁷⁰ 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
- ⁷⁷¹ 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.
- ⁷⁷² Graves, Catherine E., et al. "Regular Expression Matching with Memristor TCAMs." *2018 IEEE International Conference on Rebooting Computing (ICRC)*. IEEE, 2018.
- ⁷⁷³ E. O. Neftci, H. Mostafa, and F. Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, Nov 2019. doi: 10.1109/MSP.2019.2931595.2
- ⁷⁷⁴ J. Kaiser, H. Mostafa, and E. Neftci. Synaptic plasticity for deep continuous local learning. arXiv preprint arXiv:1812.10766, Nov 2018.
- ⁷⁷⁵ Friedemann Zenke and Surya Ganguli. Superspike: Supervised learning in multi-layer spiking neural networks. arXiv preprint arXiv:1705.11146, 2017.
- ⁷⁷⁶ Timothy P Lillicrap, Daniel Cownden, Douglas B Tweed, and Colin J Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7, 2016.
- ⁷⁷⁷ Guillaume Bellec, Franz Scherr, Elias Hajek, Darjan Salaj, Robert Legenstein, and Wolfgang Maass. Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. arXiv preprint arXiv:1901.09049, 2019.
- ⁷⁷⁸ João Sacramento, Rui Ponte Costa, Yoshua Bengio, and Walter Senn. Dendritic cortical microcircuits approximate the backpropagation algorithm. In *Advances in Neural Information Processing Systems*, pages 8721–8732, 2018.
- ⁷⁷⁹ Pierre Baldi, Peter Sadowski, and Zhiqin Lu. Learning in the machine: The symmetries of the deep learning channel. *Neural Networks*, 95:110–133, 2017.
- ⁷⁸⁰ M. Horowitz. 1.1 computing's energy problem (and what we can do about it). In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 10–14. IEEE, 2014.
- ⁷⁸¹ Hesham Mostafa, Vishwajith Ramesh, and Gert Cauwenberghs. Deep supervised learning using local errors. arXiv preprint arXiv:1711.06756, 2017.
- ⁷⁸² Arild Nøkland and Lars Hiller Eidnes. Training neural networks with local error signals. arXiv preprint arXiv:1901.06656, 2019.
- ⁷⁸³ Max Jaderberg, Wojciech Marian Czarnecki, Simon Osindero, Oriol Vinyals, Alex Graves, and Koray Kavukcuoglu. Decoupled neural interfaces using synthetic gradients. arXiv preprint arXiv:1608.05343, 2016.
- ⁷⁸⁴ M Payvand, M. Fouda, A Eltawil, F Kurdahi, and E. Neftci. Error-triggered three-factor learning dynamics for crossbar arrays. arXiv preprint arXiv:1910.06152, Dec 2019.
- ⁷⁸⁵ Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

- ⁷⁸⁶ Pentti Kanerva. "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." *Cognitive Computation*, 1(2):139-159, Jun 2009.
- ⁷⁸⁷ Justin C. Wong. "Negative capacitance and hyperdimensional computing for unconventional low-power computing." PhD thesis, EECS Department, University of California, Berkeley, Dec 2018.
- ⁷⁸⁸ Geoffrey E. Hinton. "Mapping part-whole hierarchies into connectionist networks." *Artificial Intelligence*, 46(1):47-75, 1990.
- ⁷⁸⁹ Tony Plate. "Holographic reduced representations: Convolution algebra for compositional distributed representations." In Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'91, pages 30-35, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.
- ⁷⁹⁰ Tony A. Plate, Holographic Reduced Representation: Distributed Representation for Cognitive Structures. CSLI Publications, Stanford, CA, USA, 2003.
- ⁷⁹¹ Pentti Kanerva, "Binary spatter-coding of ordered k-tuples" Artificial Neural Networks - ICANN 96, pages 869-873, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.
- ⁷⁹² Abbas Rahimi, Pentti Kanerva, and Jan M. Rabaey, "A robust and energy-efficient classifier using brain-inspired hyperdimensional computing" Proceedings of the 2016 International Symposium on Low Power Electronics and Design, ISLPED '16, pages 64-69, New York, NY, USA, 2016. ACM.
- ⁷⁹³ H. Li, et al, "Hyperdimensional computing with 3d vrram in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition," *2016 IEEE International Electron Devices Meeting (IEDM)*, pages 16.1.1-16.1.4, Dec 2016.
- ⁷⁹⁴ T. F. Wu, et al., "Brain-inspired computing exploiting carbon nanotube fets and resistive ram: Hyperdimensional computing case study," *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 492-494, Feb 2018.
- ⁷⁹⁵ S. Datta, R. A. G. Antonio, A. R. S. Ison, and J. M. Rabaey, "A programmable hyper-dimensional processor architecture for human-centric IOT" *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(3):439-452, Sep. 2019.
- ⁷⁹⁶ Sohun Datta. "A binary, 2048-dim. generic hyper-dimensional processor." Master's thesis, EECS Department, University of California, Berkeley, May 2019.
- ⁷⁹⁷ Aditya Joshi, Johan T. Halseth, and Pentti Kanerva. "Language geometry using random indexing." *Quantum Interaction*, pages 265-274, Cham, 2017. Springer International Publishing.
- ⁷⁹⁸ A. Rahimi, S. Benatti, P. Kanerva, L. Benini, and J. M. Rabaey. "Hyperdimensional biosignal processing: A case study for emg-based hand gesture recognition." *2016 IEEE International Conference on Rebooting Computing (ICRC)*, pages 1-8, Oct 2016.
- ⁷⁹⁹ 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.
- ⁸⁰⁰ Locatelli, N., Cros, V. & Grollier, J. Spin-torque building blocks. *Nature Mater* **13**, 11–20 (2014).
- ⁸⁰¹ Tuma, T., Pantazi, A., Le Gallo, M. *et al.* Stochastic phase-change neurons. *Nature Nanotech* **11**, 693–699 (2016).
- ⁸⁰² 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
- ⁸⁰³ S. Aaronson, *Quantum Computing since Democritus*, Cambridge University Press, 2013.
- ⁸⁰⁴ 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.
- ⁸⁰⁵ S. Kumar, J. P. Strachan and R. S. Williams, "Chaotic dynamics in nanoscale NbO₂ Mott memristors for analogue computing," *Nature*, 548, 318 (2017)
- ⁸⁰⁶ 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.
- ⁸⁰⁷ 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.
- ⁸⁰⁸ T. P. Xiao, "Optoelectronics for refrigeration and analog circuits for combinatorial optimization," Ph.D. dissertation, Chapter 6, University of California, Berkeley, 2019.
- ⁸⁰⁹ T. Wang and J. Roychowdhury, "Oscillator-based Ising machine," arXiv e-prints, 2017. arXiv: 1709.08102
- ⁸¹⁰ 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.
- ⁸¹¹ J. Chou, S. Bramhavar, S. Ghosh, and W. Herzog, "Analog coupled oscillator based weighted Ising machine," arXiv e-prints, 2019. arXiv: 1906.06312
- ⁸¹² A. Lucas, "Ising formulations of many NP problems," *Interdisciplinary Physics*, vol. 2, p. 5, 2014.
- ⁸¹³ 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.
- ⁸¹⁴ T. Inagaki et al, "A coherent Ising machine for 2000-node optimization problems," *Science*, vol. 354, no. 6312, pp. 603-606, 2016.
- ⁸¹⁵ 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.
- ⁸¹⁶ I. Mahboob, H. Okamoto, and H. Yamaguchi, "An electromechanical Ising Hamiltonian," *Science Advances*, vol 2, no. 6, 2016.
- ⁸¹⁷ E. Goto, "The parametron, a digital computing element which utilizes parametric oscillation," *Proceedings of the IRE*, vol. 47, no. 8, pp. 1304-1316, 1959.

- ⁸¹⁸ 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.
- ⁸¹⁹ 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.
- ⁸²⁰ 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.
- ⁸²¹ G. Appa, L. S. Pitsoulis, and H. P. Williams, Eds., *Handbook on modelling for discrete optimization*. New York: Springer, 2006.
- ⁸²² 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.
- ⁸²³ F. C. Hoppensteadt and E. M. Izhikevich, "Oscillatory neurocomputers with dynamic connectivity," *Physical Review Letters*, vol. 82, no. 14, p. 2983, 1999.
- ⁸²⁴ 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.
- ⁸²⁵ 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.
- ⁸²⁶ 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.
- ⁸²⁷ 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.
- ⁸²⁸ 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.
- ⁸²⁹ 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
- ⁸³⁰ 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.
- ⁸³¹ L. Yang, L. Chua, and K. Krieg, "VLSI implementation of cellular neural networks," *Circuits and Systems, IEEE International Symposium on*, 1990, pp. 2425–2427.
- ⁸³² 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.
- ⁸³³ 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.
- ⁸³⁴ 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.
- ⁸³⁵ 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.
- ⁸³⁶ 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.
- ⁸³⁷ 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.
- ⁸³⁸ C. Pan and A. Naemi, "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.
- ⁸³⁹ C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, no. 78, 1990. pp. 1629–1636.
- ⁸⁴⁰ J. Hasler, "Starting Framework for Analog Numerical Analysis," *Journal of Low Power Electronics Applications*, vol. 7, no. 17, June 2017.
- ⁸⁴¹ J. Hasler, "Analog Architecture and Complexity Theory SoC Systems," *Journal of Low Power Electronics Applications*, vol. 9, no. 1, January 2019.
- ⁸⁴² J. Hasler, S. Kim, and A. Natarajan, "Enabling Energy-Efficient Physical Computing through Analog Abstraction and IP Reuse," *Journal of Low Power Electronics Applications*, vol. 8, no. 4, November, 2018.
- ⁸⁴³ Eric Black and Jennifer Hasler, "A Model and Implication of Physical Computing," GOMAC 2020
- ⁸⁴⁴ S. George, J. Hasler, S. Koziol, S. Nease, and S. Ramakrishnan, "Low power dendritic computation for wordspotting," *Journal of Low Power Electronics Applications*, vol. 3, 2013, 73–98.
- ⁸⁴⁵ S. Koziol, R. Wunderlich, J. Hasler, M. Stilman, "Single-Objective Path Planning for Autonomous Robots Using Reconfigurable Analog VLSI," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, pp. 1–14.
- ⁸⁴⁶ S. Koziol, S. Brink, and J. Hasler, "A neuromorphic approach to path planning using a reconfigurable neuron array IC," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 12, pp. 2724–2737, 2014.
- ⁸⁴⁷ Y. Huang, et. al, "Analog computing in a modern context: A linear algebra accelerator case study," *IEEE Micro*, vol. 37, no. 3, 2017.
- ⁸⁴⁸ S. George, et. al, "A Programmable and Configurable Mixed-Mode FPAA SoC," *TVLSI*, vol. 24, no. 6, 2016.

- ⁸⁴⁹ J. Hasler, "Large-Scale Field Programmable Analog Arrays," IEEE Proceedings, March 2020, pp. 1-20.
- ⁸⁵⁰ G. Y. Vichniac, "Simulating physics with cellular automata," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 96–116, Jan. 1984.
- ⁸⁵¹ T. Toffoli, "Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 117–127, Jan. 1984.
- ⁸⁵² T. Toffoli, "CAM: A high-performance cellular-automaton machine," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, pp. 195–204, Jan. 1984.
- ⁸⁵³ A. Vergis, K. Steiglitz, B. Dickinson, "The Complexity of Analog Computation, *ACM Journ. Math and Comp in Simulation*, Vol. 28, no. 2, April 1986. pp. 91 - 113.
- ⁸⁵⁴ M. Ercsey-Ravasz, and Z. Toroczkai, "Optimization hardness as transient chaos in an analog constraint satisfaction," *Nature Physics*, vol. 7, no. 12, 2011.
- ⁸⁵⁵ X. Yin, B. Sedighi, M. Varga, M. Ercsey-Ravasz, Z. Toroczkai, X. S. Hu, "Efficient Analog Circuits for Boolean Satisfiability," *IEEE TVLSI*, 2017.
- ⁸⁵⁶ C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, January 1949.
- ⁸⁵⁷ 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.
- ⁸⁵⁸ 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.
- ⁸⁵⁹ 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>.
- ⁸⁶⁰ Wikipedia, "Chaos Computing," https://en.wikipedia.org/wiki/Chaos_computing, retrieved Dec. 6, 2019.
- ⁸⁶¹ M. Hu *et al.*, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," *2016 53rd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2016, pp. 1–6.
- ⁸⁶² 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.
- ⁸⁶³ 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).
- ⁸⁶⁴ X. Sheng, *et al.* "Low-Conductance and Multilevel CMOS-Integrated Nanoscale Oxide Memristors." *Advanced Electronic Materials* (2019): 1800876
- ⁸⁶⁵ 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).
- ⁸⁶⁶ G. W. Burr, *et al.* "Recent progress in phase-change memory technology," *IEEE J. Emerg. Sel. Top. Circ. Sys.* 6(2), 146-162 (2016).
- ⁸⁶⁷ 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.
- ⁸⁶⁸ 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.
- ⁸⁶⁹ 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.
- ⁸⁷⁰ 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).
- ⁸⁷¹ 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.
- ⁸⁷² 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.
- ⁸⁷³ W. W. Koelmans, *et al.*, "Projected phase-change memory devices," *Nature communications*, (6) 8181 (2015).
- ⁸⁷⁴ 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.
- ⁸⁷⁵ 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*.
- ⁸⁷⁶ E. J. Fuller *et al.*, Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* **364**, 570-574 (2019).
- ⁸⁷⁷ 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).
- ⁸⁷⁸ C. A. Marianetti, G. Kotliar, G. Ceder, A first-order Mott transition in Li_xCoO₂. *Nature Materials* **3**, 627-631 (2004).
- ⁸⁷⁹ 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.
- ⁸⁸⁰ 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.

- ⁸⁸¹ 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.
- ⁸⁸² Chakrabarty S and Cauwenberghs G 2007 Sub-microwatt analog VLSI trainable pattern classifier *IEEE J. Solid-State Circuits* **42** 1169–1179
- ⁸⁸³ 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
- ⁸⁸⁴ Hasler J and Marr H 2013 Finding a roadmap to achieve large neuromorphic hardware systems *Frontiers Neurosci.* **7** 118
- ⁸⁸⁵ Guo X, Merrih 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
- ⁸⁸⁶ X. Guo, F. Merrih 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.
- ⁸⁸⁷ Silicon Storage Technology, Inc. <http://www.sst.com/products-and-services/superflash-technology-products>
- ⁸⁸⁸ 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
- ⁸⁸⁹ 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
- ⁸⁹⁰ 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.
- ⁸⁹¹ 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.
- ⁸⁹² N. Shibata, K. Kanda, T. Hisada, K. Isobe, *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.
- ⁸⁹³ 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.
- ⁸⁹⁴ 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.
- ⁸⁹⁵ 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.
- ⁸⁹⁶ 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 *ArXiv (submitted)*
- ⁸⁹⁷ 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 *ArXiv (submitted)*
- ⁸⁹⁸ 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.
- ⁸⁹⁹ 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.
- ⁹⁰⁰ 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.
- ⁹⁰¹ 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.
- ⁹⁰² 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.
- ⁹⁰³ 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.
- ⁹⁰⁴ 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.
- ⁹⁰⁵ 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.
- ⁹⁰⁶ 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.
- ⁹⁰⁷ 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.
- ⁹⁰⁸ 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.

- ⁹⁰⁹ W. Guo and N. Sun, "A 12b-ENOB 61 μ W noise-shaping SAR ADC with a passive integrator," *42nd IEEE European Solid-State Circuits Conference (ESSCIRC 2016)*, 405-408, 2016.
- ⁹¹⁰ 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.
- ⁹¹¹ 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.
- ⁹¹² 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.
- ⁹¹³ 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.
- ⁹¹⁴ Hamerly, Ryan, *et al.* "Large-scale optical neural networks based on photoelectric multiplication." *Physical Review X* 9.2 (2019): 021032.
- ⁹¹⁵ Tait, Alexander N., *et al.* "Neuromorphic photonic networks using silicon photonic weight banks." *Scientific reports* 7.1 (2017): 7430
- ⁹¹⁶ Shen, Yichen, *et al.* "Deep learning with coherent nanophotonic circuits." *Nature Photonics* 11.7 (2017): 441.
- ⁹¹⁷ Yang, Lin, *et al.* "On-chip CMOS-compatible optical signal processor." *Optics express* 20.12 (2012): 13560-13565.
- ⁹¹⁸ Tait, Alexander N., *et al.* "Microring weight banks." *IEEE Journal of Selected Topics in Quantum Electronics* 22.6 (2016): 312-325.
- ⁹¹⁹ 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
- ⁹²⁰ Papamichael, Jeremy Fowers Kalin Ovtcharov Michael, *et al.* "A Configurable Cloud-Scale DNN Processor for Real-Time AI." (2018).
- ⁹²¹ Nicol, Chris. "A coarse grain reconfigurable array (cgra) for statically scheduled data flow computing." *Wave Computing White Paper* (2017).
- ⁹²² 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.
- ⁹²³ Atabaki, Amir H., *et al.* "Accurate post-fabrication trimming of ultra-compact resonators on silicon." *Optics express* 21.12 (2013): 14139-14145.
- ⁹²⁴ Atabaki, Amir H., *et al.* "Accurate post-fabrication trimming of ultra-compact resonators on silicon." *Optics express* 21.12 (2013): 14139-14145.
- ⁹²⁵ Padmaraju, Kishore, and Keren Bergman. "Resolving the thermal challenges for silicon microring resonator devices." *Nanophotonics* 3.4-5 (2014): 269-281.
- ⁹²⁶ Krishnamoorthy, Ashok V., *et al.* "Exploiting CMOS manufacturing to reduce tuning requirements for resonant optical devices." *IEEE Photonics Journal* 3.3 (2011): 567-579.
- ⁹²⁷ Su, Zhan, *et al.* "Reduced wafer-scale frequency variation in adiabatic microring resonators." *OFC 2014*. IEEE, 2014.
- ⁹²⁸ Mekis, Attila, *et al.* "A grating-coupler-enabled CMOS photonics platform." *IEEE Journal of Selected Topics in Quantum Electronics* 17.3 (2010): 597-608.
- ⁹²⁹ 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.
- ⁹³⁰ Assefa, Solomon, Fengnian Xia, and Yurii A. Vlasov. "Reinventing germanium avalanche photodetector for nanophotonic on-chip optical interconnects." *Nature* 464.7285 (2010): 80.
- ⁹³¹ 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.
- ⁹³² Glick, Madeleine, Lionel C. Kimmerling, and Robert C. Pfahl. "A roadmap for integrated photonics." *Optics and Photonics News* 29.3 (2018): 36-41.
- ⁹³³ 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.
- ⁹³⁴ 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.
- ⁹³⁵ 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.
- ⁹³⁶ C. Pan and A. Naemi, "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.
- ⁹³⁷ 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.
- ⁹³⁸ 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.
- ⁹³⁹ 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.
- ⁹⁴⁰ 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.

⁹⁴¹ 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.
- ⁹⁴² R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. L. Heras, R. Babbush, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, E. Solano, H. Neven, and J. M. Martinis, "Digitized adiabatic quantum computing with a superconducting circuit," *Nature*, vol. 534, no. 7606, pp. 222–226, Jun. 2016.
- ⁹⁴³ 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.
- ⁹⁴⁴ 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.
- ⁹⁴⁵ 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.
- ⁹⁴⁶ 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.
- ⁹⁴⁷ 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.
- ⁹⁴⁸ 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.
- ⁹⁴⁹ L. Maleki, "Sources: The optoelectronic oscillator," *Nature Photonics*, vol. 5, no. 12, pp. 728–730, 2011.
- ⁹⁵⁰ W. A. Vitale, C. F. Moldovan, A. Paone, A. Schüller, and A. M. Ionescu, "Fabrication of CMOS-compatible abrupt electronic switches based on vanadium dioxide," *Microelectronic Engineering*, vol. 145, pp. 117–119, Sep. 2015.
- ⁹⁵¹ H.-T. Kim, B.-G. Chae, D.-H. Youn, S.-L. Maeng, G. Kim, K.-Y. Kang, and Y.-S. Lim, "Mechanism and observation of Mott transition in VO₂-based two- and three-terminal devices," *New Journal of Physics*, vol. 6, no. 1, p. 52, 2004.
- ⁹⁵² 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.
- ⁹⁵³ 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.
- ⁹⁵⁴ 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.
- ⁹⁵⁵ H. Eisele, "State of the art and future of electronic sources at terahertz frequencies," *Electronics Letters*, vol. 46, no. 26, p. S8, 2010.
- ⁹⁵⁶ 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.
- ⁹⁵⁷ J. C. Magee, "Dendritic integration of excitatory synaptic input," *Nature Reviews Neuroscience*, vol. 1, pp. 181–190, 12/print 2000.
- ⁹⁵⁸ H. C. Lai and L. Y. Jan, "The distribution and targeting of neuronal voltage-gated ion channels," *Nature Reviews Neuroscience*, vol. 7, pp. 548–62, Jul 2006.
- ⁹⁵⁹ B. P. Bean, "The action potential in mammalian central neurons," *Nature Reviews Neuroscience*, vol. 8, pp. 451–65, Jun 2007.
- ⁹⁶⁰ 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.
- ⁹⁶¹ 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.
- ⁹⁶² 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.
- ⁹⁶³ A. Sengupta and K. Roy, "Spin-Transfer Torque Magnetic neuron for low power neuromorphic computing," *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney, pp. 1–7, 2015.
- ⁹⁶⁴ S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems," *Nano Letters*, vol. 10, pp. 1297–1301, 2010.
- ⁹⁶⁵ 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.
- ⁹⁶⁶ M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed., Cambridge University Press, 2017.
- ⁹⁶⁷ A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature reviews. Neuroscience*, vol. 9, p. 292, 2008.
- ⁹⁶⁸ 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.
- ⁹⁶⁹ 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.
- ⁹⁷⁰ 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.
- ⁹⁷¹ 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).

- ⁹⁷² 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).
- ⁹⁷³ 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)
- ⁹⁷⁴ 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)
- ⁹⁷⁵ 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
- ⁹⁷⁶ 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.
- ⁹⁷⁷ 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.
- ⁹⁷⁸ 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.
- ⁹⁷⁹ 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
- ⁹⁸⁰ 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
- ⁹⁸¹ M. Park, J. C. Rodgers, and D. P. Lathrop, "True random number generation using CMOS Boolean chaotic oscillator," *Microelectronics Journal* 46, 1364 (2015).
- ⁹⁸² Russek, S. E. et al. in 2016 IEEE International Conference on Rebooting Computing (ICRC). 1-5.
- ⁹⁸³ Schneider, M. L. et al. Ultralow power artificial synapses using nanotextured magnetic Josephson junctions. *Science Advances* 4, e1701329, doi:10.1126/sciadv.1701329 (2018).
- ⁹⁸⁴ 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).
- ⁹⁸⁵ 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).
- ⁹⁸⁶ Behin-Aein, Behtash, Vinh Diep, and Supriyo Datta. "A building block for hardware belief networks." *Scientific reports* 6 (2016): 29893.
- ⁹⁸⁷ Camsari, Kerem Yunus, et al. "Stochastic p-bits for invertible logic." *Physical Review X* 7.3 (2017): 031014.
- ⁹⁸⁸ Camsari, Kerem Yunus, Sayeef Salahuddin, and Supriyo Datta. "Implementing p-bits with embedded MTJ." *IEEE Electron Device Letters* 38.12 (2017): 1767-1770.
- ⁹⁸⁹ 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.
- ⁹⁹⁰ Sutton, Brian, et al. "Intrinsic optimization using stochastic nanomagnets." *Scientific reports* 7 (2017): 44370.
- ⁹⁹¹ Faria, Rafatul, Kerem Y. Camsari, and Supriyo Datta. "Implementing Bayesian networks with embedded stochastic MRAM." *AIP Advances* 8.4 (2018): 045101
- ⁹⁹² 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.
- ⁹⁹³ Kaisers, Jan, et al. "Probabilistic Circuits for Autonomous Learning: A simulation study." arXiv preprint arXiv:1910.06288 (2019)
- ⁹⁹⁴ Borders, William A., et al. "Integer factorization using stochastic magnetic tunnel junctions." *Nature* 573.7774 (2019): 390-393
- ⁹⁹⁵ 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.
- ⁹⁹⁶ 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).
- ⁹⁹⁷ Sutton, Brian, et al. "Autonomous Probabilistic Coprocessing with Petaflips per Second." arXiv preprint arXiv:1907.09664 (2019).
- ⁹⁹⁸ Camsari, Kerem Y., Brian M. Sutton, and Supriyo Datta. "p-bits for probabilistic spin logic." *Applied Physics Reviews* 6.1 (2019): 011305.
- ⁹⁹⁹ 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.
- ¹⁰⁰⁰ 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.
- ¹⁰⁰¹ 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.
- ¹⁰⁰² 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>.
- ¹⁰⁰³ 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.
- ¹⁰⁰⁴ M. P. Frank and M.J. Ammer, "Relativized separation of reversible and irreversible space-time complexity classes," preprint, arxiv:1708.08480 [cs.ET], 2017.

- ¹⁰⁰⁵ S. Pidaparathi and C. Lent, “Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata,” *Journal of Low Power Electronics and Applications* vol. 8, no. 3, 2018.
- ¹⁰⁰⁶ M. P. Frank, “On the interpretation of energy as the rate of quantum computation,” *Quantum Information Processing*, vol. 4, no. 4, 2005.
- ¹⁰⁰⁷ M. P. Frank and T. F. Knight Jr., “Ultimate theoretical models of nanocomputers,” *Nanotechnology* 9, no. 3, 1998.
- ¹⁰⁰⁸ K. Likharev, “Dynamics of some single flux quantum devices: I. Parametric quantron,” *IEEE Trans. Magn.* vol. 13, no. 242, 1977.
- ¹⁰⁰⁹ 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.
- ¹⁰¹⁰ 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>.
- ¹⁰¹¹ 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.
- ¹⁰¹² C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, “Quantum cellular automata,” *Nanotechnology* vol. 4, no. 1, p. 49, 1993.
- ¹⁰¹³ 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.
- ¹⁰¹⁴ 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.
- ¹⁰¹⁵ 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](https://arxiv.org/abs/1801.03534) [cs.ET], 2018.
- ¹⁰¹⁶ 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.
- ¹⁰¹⁷ E. Fredkin and T. Toffoli, “Conservative logic,” *International Journal of Theoretical Physics*, vol. 21, nos. 3–4, pp. 219–253, Apr. 1982.
- ¹⁰¹⁸ A. Adamatzky, ed., *Collision-Based Computing*, Springer, 2002.
- ¹⁰¹⁹ 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.
- ¹⁰²⁰ T. Asai, K. Yamada, and Y. Amemiya, “Single-flux-quantum logic circuits exploiting collision-based fusion gates,” *Physica C*, vol. 468, 2008.
- ¹⁰²¹ 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.
- ¹⁰²² 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.
- ¹⁰²³ K. D. Osborn, “Reversible computation with flux solitons,” *United States Patent 9812836 B1*, Nov. 2017.
- ¹⁰²⁴ 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](https://arxiv.org/abs/1711.04339) [cond-mat.supr-con], 2017.
- ¹⁰²⁵ W. Wustmann and K. D. Osborn, “Reversible Fluxon Logic: Topological particles allow gates beyond the standard adiabatic limit,” preprint, [arXiv:1711.04339](https://arxiv.org/abs/1711.04339) [cond-mat.supr-con], Nov. 2017.
- ¹⁰²⁶ 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](https://arxiv.org/abs/1806.08011) [cond-mat.mes-hall].
- ¹⁰²⁷ 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.ieeeesc.org/sites/ieeeesc.org/files/documents/snf/abstracts/STP652_Final_ISECExtendedAbstract.pdf.
- ¹⁰²⁸ M. P. Frank, “Asynchronous Ballistic Reversible Computing.” in 2017 *IEEE International Conference on Rebooting Computing (ICRC)*, pp. 172–179, 2017.
- ¹⁰²⁹ 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.
- ¹⁰³⁰ 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-et-al_final_v7.pdf.
- ¹⁰³¹ 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>.
- ¹⁰³² 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>.
- ¹⁰³³ N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, “Simulation of sub- $k_B T$ 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.

- ¹⁰³⁴ 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.
- ¹⁰³⁵ 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.
- ¹⁰³⁶ 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.
- ¹⁰³⁷ N. Takeuchi, Y. Yuki, and Y. Nobuyuki, “Recent Progress on Reversible Quantum-Flux-Parametron for Superconductor Reversible Computing,” *IEICE Transactions on Electronics* vol. 101, no. 5, May 2018.
- ¹⁰³⁸ Y. Taiki, T. Naoki, Y. Yamanashi, and N. Yoshikawa, “Design and demonstration of reversible full adders using adiabatic quantum flux parametron logic,” *IEEE Applied Superconductivity Conference*, 2018.
- ¹⁰³⁹ 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.
- ¹⁰⁴⁰ J. Ren, “Physically and Logically Reversible Superconducting Circuit,” Ph.D. dissertation, Stony Brook University, May 2011. <https://dspace.sunyconnect.suny.edu/handle/1951/56102>.
- ¹⁰⁴¹ 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.
- ¹⁰⁴² M. P. Frank, “Approaching the physical limits of computing,” *35th International Symposium on Multiple-Valued Logic (ISMVL’05)*, 2005, pp. 168–185.
- ¹⁰⁴³ 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.
- ¹⁰⁴⁴ 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.
- ¹⁰⁴⁵ 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>
- ¹⁰⁴⁶ C. J. Vieri, “Pendulum—A reversible computer architecture,” M.S. thesis, MIT Dept. of EECS, 1995
- ¹⁰⁴⁷ M. P. Frank, “Reversibility for Efficient Computing,” Ph.D. dissertation, MIT Dept. of EECS, 1999. <https://dspace.mit.edu/handle/1721.1/9464>.
- ¹⁰⁴⁸ C. J. Vieri, “Reversible computer engineering and architecture,” Ph.D. dissertation, MIT Dept. of EECS, 1999. <https://dspace.mit.edu/handle/1721.1/80144>.
- ¹⁰⁴⁹ 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.
- ¹⁰⁵⁰ A. Vetuli, S. Di Pascoli, and L. M. Reyneri, “Positive feedback in adiabatic logic,” *Electronics Letters* vo. 32, no. 20, 1996.
- ¹⁰⁵¹ P. Teichman, “Adiabatic Logic: Future Trend and System Level Perspective,” Springer Science & Business Media, vol. 24, New York: Springer, 2012.
- ¹⁰⁵² Y. Moon and D.-K. Jeong, “Efficient charge recovery logic,” Digest of Technical Papers, Symposium on VLSI Circuits, IEEE, 1995.
- ¹⁰⁵³ 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.
- ¹⁰⁵⁴ 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.
- ¹⁰⁵⁵ 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.
- ¹⁰⁵⁶ 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.
- ¹⁰⁵⁷ 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.
- ¹⁰⁵⁸ K. Loe, E. Goto, “Analysis of flux input and output Josephson pair device,” *IEEE Trans. Magn.*, vol. 21, pp. 884–887, 1985.
- ¹⁰⁵⁹ 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.
- ¹⁰⁶⁰ N. Takeuchi, Y. Yamanashi, N. Yoshikawa, “Energy efficiency of adiabatic superconductor logic,” *Supercond. Sci. Technol.*, vol. 28, art. 015003, 2015.
- ¹⁰⁶¹ N. Takeuchi, Y. Yamanashi, N. Yoshikawa, “Thermodynamic Study of Energy Dissipation in Adiabatic Superconductor Logic,” *Phys. Rev. Applied*, vol. 4, art. 034007, 2015.
- ¹⁰⁶² 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.
- ¹⁰⁶³ N. Takeuchi, T. Yamae, C. L. Ayala, H. Suzuki, N. Yoshikawa, “An adiabatic superconductor 8-bit adder with $24 k_B T$ energy dissipation per junction,” *Appl. Phys. Lett.*, vol. 114, art. 042602, 2019.

- ¹⁰⁶⁴ T. Yamae, N. Takeuchi, N. Yoshikawa, "A reversible full adder using adiabatic superconductor logic," *Supercond. Sci. Technol.*, vol. 32, art. 035005, 2019.
- ¹⁰⁶⁵ 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.
- ¹⁰⁶⁶ 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.
- ¹⁰⁶⁷ K. E. Drexler, *Nanosystems: molecular machinery, manufacturing, and computation*, New York: Wiley, 1992.
- ¹⁰⁶⁸ 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.
- ¹⁰⁶⁹ E. Blair, "Electric-Field Inputs for Molecular Quantum-Dot Cellular Automata Circuits," *IEEE Transactions on Nanotechnology*, vol. 18, pp. 453–460, 2019.
- ¹⁰⁷⁰ 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.
- ¹⁰⁷¹ 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.
- ¹⁰⁷² H. F. G. Pillonnet, S. Hourii, "Adiabatic Capacitive Logic: a paradigm for low-power logic," in *IEEE International Symposium on Circuits and Systems*. IEEE: Baltimore, MD, 2017.
- ¹⁰⁷³ 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.
- ¹⁰⁷⁴ 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.
- ¹⁰⁷⁵ 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.
- ¹⁰⁷⁶ 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.
- ¹⁰⁷⁷ 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.
- ¹⁰⁷⁸ 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
- ¹⁰⁷⁹ 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
- ¹⁰⁸⁰ 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
- ¹⁰⁸¹ 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
- ¹⁰⁸² 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
- ¹⁰⁸³ 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
- ¹⁰⁸⁴ 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
- ¹⁰⁸⁵ 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.
- ¹⁰⁸⁶ 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.
- ¹⁰⁸⁷ 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.
- ¹⁰⁸⁸ 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.
- ¹⁰⁸⁹ 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.
- ¹⁰⁹⁰ 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).
- ¹⁰⁹¹ 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)
- ¹⁰⁹² 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.
- ¹⁰⁹³ D. P. C. Dimitrakopoulos, and J. Smith, "Authentication using graphene based devices as physical unclonable functions," 2004. Available: <http://www.google.com/patents/US20140159040>

- ¹⁰⁹⁴ 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
- ¹⁰⁹⁵ 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
- ¹⁰⁹⁶ 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
- ¹⁰⁹⁷ 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-DRAM) in 90 nm CMOS logic technology and circuits," *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1–4: IEEE
- ¹⁰⁹⁸ 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
- ¹⁰⁹⁹ 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
- ¹¹⁰⁰ 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
- ¹¹⁰¹ 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
- ¹¹⁰² 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
- ¹¹⁰³ N. Harada, K. Yagi, S. Sato, and N. Yokoyama, "A polarity-controllable graphene inverter," *Applied Physics Letters*, vol. 96, no. 1, p. 012102, 2010
- ¹¹⁰⁴ A. Heinzig, S. Slesazek, 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
- ¹¹⁰⁵ S. Das and J. Appenzeller, "WSe₂ field effect transistors with enhanced ambipolar characteristics," *Applied Physics Letters*, vol. 103, no. 10, p. 103501, 2013
- ¹¹⁰⁶ 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
- ¹¹⁰⁷ T. Vasen, "Investigation of III-V Tunneling Field-Effect Transistors," A Dissertation submitted to the University of Notre Dame, 2014.
- ¹¹⁰⁸ 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
- ¹¹⁰⁹ 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
- ¹¹¹⁰ 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
- ¹¹¹¹ J. Rajendran et al., "Fault Analysis-Based Logic Encryption," *Computers, IEEE Transactions on*, vol. 64, no. 2, pp. 410–424, 2015
- ¹¹¹² 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
- ¹¹¹³ 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
- ¹¹¹⁴ 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
- ¹¹¹⁵ 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
- ¹¹¹⁶ 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
- ¹¹¹⁷ 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
- ¹¹¹⁸ L. Britnell et al., "Resonant tunnelling and negative differential conductance in graphene transistors," *Nat Commun*, Article vol. 4, p. 1794, 04/30/online 2013
- ¹¹¹⁹ 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
- ¹¹²⁰ 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
- ¹¹²¹ 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
- ¹¹²² 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
- ¹¹²³ 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).
- ¹¹²⁴ P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," *Annual International Cryptology Conference*, 1999, pp. 388–397: Springer
- ¹¹²⁵ 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

- ¹¹²⁶ 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
- ¹¹²⁷ 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
- ¹¹²⁸ C. De Canniere, O. Dunkelmann, 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
- ¹¹²⁹ A. Cevrero, F. Regazzoni, M. Schwander, S. Badel, P. Jenne, 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
- ¹¹³⁰ 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>
- ¹¹³¹ 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
- ¹¹³² 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>
- ¹¹³³ 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>
- ¹¹³⁴ Y. S. Ang, S. A. Yang, C. Zhang, Z. Ma, and L. K. Ang, "Valleytronics in merging Dirac cones: All-electric-controlled valley filter, valve, and universal reversible logic gate", *Phys. Rev. B* 96, 245410 (2017): <https://doi.org/10.1103/PhysRevB.96.245410>
- ¹¹³⁵ 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>
- ¹¹³⁶ 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>
- ¹¹³⁷ 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>
- ¹¹³⁸ 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>
- ¹¹³⁹ 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>
- ¹¹⁴⁰ 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>
- ¹¹⁴¹ A. Suzuki, T. Tsuruoka, and T. Hasegawa, "Time-Dependent Operations in Molecular Gap Atomic Switches", *Phys. Status Solidi B*, 256, 1900068 (2019): <https://doi.org/10.1002/pssb.201900068>
- ¹¹⁴² E. Lidrikis, et al., *Phys. Rev. Lett.* 87, 086104 (2001).
- ¹¹⁴³ M. Andersson, J. Yuan, and B. Sundén, *Applied Energy* 87, 1461-1476 (2010).
- ¹¹⁴⁴ H. Kobayashi et al., *J. Chem. Phys.* 139, 014707 (2013).
- ¹¹⁴⁵ H. Kobayashi, and Y. Tokita, *Appl. Phys. Exp.* 8, 051602 (2015).
- ¹¹⁴⁶ H. Kobayashi et al., *Appl. Phys. Lett.* 111, 033301 (2017).
- ¹¹⁴⁷ R. Liu et al., *Integrating Materials and Manufacturing Innovation* 4, 192 (2015): <https://doi.org/10.1186/s40192-015-0042-z>
- ¹¹⁴⁸ V. V. Zhirnov, R. K. Cavin, J. A. Hutchby, G. I. Bourianoff, "Limits to Binary Logic Scaling – A Gedanken Model," *Proc. IEEE*, November 2003.
- ¹¹⁴⁹ 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.
- ¹¹⁵⁰ Doug Matzke, "Will Physical Scalability Sabotage Performance Gains?" *IEEE Computer*, Volume 30, Issue 9, September, 1997, Page: 37 – 39.
- ¹¹⁵¹ 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.
- ¹¹⁵² 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.
- ¹¹⁵³ D.E. Nikonov and I.A. Young, "Uniform Methodology for Benchmarking Beyond-CMOS Logic Devices," *IEDM Tech. Dig.*, pp. 573–576, Dec. 2012.
- ¹¹⁵⁴ 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).

-
- ¹¹⁵⁵ 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.
- ¹¹⁵⁶ 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.
- ¹¹⁵⁷ I. Sutherland et al., *Logical Effort: Design Fast CMOS Circuits*, 1st ed. San Mateo, CA: Morgan Kaufmann, Feb. 1999, ISBN: 10:1558605576
- ¹¹⁵⁸ An extremely valuable collection of different approaches to post-CMOS technology can be found in *Proceedings of the IEEE Special Issue - Nanoelectronics Research: Beyond CMOS Information Processing*, ed. G. Bourianoff, M. Brillouët, R. K. Cavin, III, T. Hiramoto, J. A. Hutchby, A. M. Ionescu, and K. Uchida, Volume 98, Issue 12, Dec 2010.