INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

EMERGING RESEARCH DEVICES

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

Scope ......................................................................................................................................................... 1

Difficult Challenges ................................................................................................................................. 2
  Introduction ........................................................................................................................................ 2
  Device Technologies ............................................................................................................................. 2
  Materials Technologies ......................................................................................................................... 3

Nano-information Processing Taxonomy .................................................................................................. 3

Emerging Research Devices .................................................................................................................... 4
  Memory Devices ................................................................................................................................. 4
    Memory Taxonomy ............................................................................................................................ 5
    Memory Devices—Definition and Discussion of Table Entries .......................................................... 12
  Logic and Alternative Information Processing Devices ....................................................................... 15
    Logic Devices .................................................................................................................................. 16
    Logic Devices—Definition and Discussion of Table Entries ............................................................. 18
    Alternative Information Processing Devices .................................................................................... 21
    Alternative Information Processing Devices—Definition and Discussion of Table Entries .......... 22

Emerging Research Architectures .......................................................................................................... 26
  Introduction ...................................................................................................................................... 26
  Trends in CMOS Many-core Architectures ......................................................................................... 26
  Heterogeneous Multi-core Architectures ............................................................................................. 27
  Possible Architectural Opportunities for Emerging Research Devices ............................................. 29

Emerging Memory and Logic Devices—A Critical Assessment .......................................................... 30
  Introduction .................................................................................................................................... 30
  Technologies beyond CMOS .............................................................................................................. 31
    Overall Technology Requirements and Relevance Criteria ............................................................ 31
    Electronic Charge-based Nanoscale Devices .................................................................................. 32
    Alternate Computational-State-Variable Nanoscale Devices .......................................................... 32
  Potential Performance Assessment for Memory and Logic Devices ................................................. 33
    Methodology .................................................................................................................................. 33
    Results .......................................................................................................................................... 34

Fundamental Guiding Principles— “Beyond CMOS” Information Processing ...................................... 44
  Introduction .................................................................................................................................... 44
  Guiding Principles ............................................................................................................................ 44
    Computational State Variable(s) other than Solely Electron Charge ............................................. 44
    Non-thermal Equilibrium Systems ................................................................................................. 44
    Novel Energy Transfer Interactions ............................................................................................... 44
    Nanoscale Thermal Management .................................................................................................. 44
    Sub-lithographic Manufacturing Process ......................................................................................... 44
    Emerging Architectures ................................................................................................................... 44

Endnotes ................................................................................................................................................. 45
### LIST OF FIGURES

| Figure ERD1 | A Taxonomy for Emerging Research Information Processing Devices .......... 4 |
| Figure ERD2a | Technology Performance Evaluation for Engineered Tunnel Barrier Memory ........................................ 37 |
| Figure ERD2b | Technology Performance Evaluation for Fuse/Antifuse Memory .................................................... 37 |
| Figure ERD2c | Technology Performance Evaluation for Nano Mechanical Memory .................................................... 38 |
| Figure ERD2d | Technology Performance Evaluation for Electronic Effects Memory .................................................... 38 |
| Figure ERD2e | Technology Performance Evaluation for Ionic Memory ................................................................. 39 |
| Figure ERD2f | Technology Performance Evaluation for Ferroelectric FET Memory .................................................. 39 |
| Figure ERD2g | Technology Performance Evaluation for Macromolecular Memory ............................................... 40 |
| Figure ERD2h | Technology Performance Evaluation for Molecular Memory ......................................................... 40 |
| Figure ERD3a | Technology Performance Evaluation for 1D Structures (CNTs and NWs) Logic Devices .................. 41 |
| Figure ERD3b | Technology Performance Evaluation for Channel Replacement Materials Logic Devices .................. 41 |
| Figure ERD3c | Technology Performance Evaluation for Single-Electron Transistors Logic Devices ....................... 42 |
| Figure ERD3d | Technology Performance Evaluation for Molecular Logic Devices ................................................. 42 |
| Figure ERD3e | Technology Performance Evaluation for Ferromagnetic Logic Devices ........................................... 43 |
| Figure ERD3f | Technology Performance Evaluation for Spin Transistors Logic Devices ........................................ 43 |

### LIST OF TABLES

| Table ERD1 | Emerging Research Devices Difficult Challenges ................................................................. 3 |
| Table ERD2 | Memory Taxonomy .................................................................................................................. 6 |
| Table ERD3 | Current Baseline and Prototypical Memory Technologies ....................................................... 7 |
| Table ERD4 | Transition Table for Emerging Research Memory Devices .................................................... 8 |
| Table ERD5a | Emerging Research Capacitance-based Memory Devices—Demonstrated and Projected Parameters .......... 9 |
| Table ERD5b | Emerging Research Resistance-based Memory Devices—Demonstrated and Projected Parameters ........... 11 |
| Table ERD6 | Transition Table for Emerging Research Logic Devices ........................................................ 16 |
| Table ERD7a | Emerging Research Logic Devices—Demonstrated and Projected Parameters .................................... 17 |
| Table ERD7b | Alternative Information Processing Devices ............................................................................... 22 |
| Table ERD8 | Emerging Research Architectures .......................................................................................... 30 |
| Table ERD9 | Potential Evaluation for Emerging Research Memory Devices .................................................. 35 |
| Table ERD10 | Potential Evaluation for Emerging Research Logic Devices ..................................................... 36 |
EMERGING RESEARCH DEVICES

SCOPE

Defined and driven by the relentless cadence of CMOS scaling, information technology is increasingly expanding into and indeed enabling new applications across a variety of markets. Dimensional scaling of CMOS will continue at least for the time horizon of this roadmap, and “equivalent” or performance scaling of CMOS will likely continue beyond this horizon. Consequently, CMOS certainly will provide a platform processing technology for sometime beyond the end of dimensional scaling, exploiting the notion that the ultimately scaled MOSFET is a nearly ideal electronic charge-based device. Emerging new information processing technologies heterogeneously integrated with CMOS processor cores and offering unique non-Boolean specialized functions may extend functional scaling of the CMOS platform by providing more efficient solutions to current applications and enabling new applications.

This possibility motivates an expansion in scope and content of the Emerging Research Devices (ERD) chapter for 2007. The chapter evaluates emerging new research technologies for memory and information processing or logic devices and nanoarchitectures applied to two different approaches to realizing integrated electronic functions. One is heterogeneous integration of these new technologies with the CMOS platform, i.e., “enhanced CMOS” or “Functional Diversification”. The second addresses the exciting but daunting challenge to invent one or more fundamentally new approaches to information and signal processing. This theme will require discovery and exploitation of a new means to physically represent, process, store, and transport information via new materials, process, device, nanoarchitecture, and systems innovations. In addition, the Emerging Research Materials section introduced in the 2005 ERD chapter has been expanded substantially to become a new chapter. Materials research issues related to emerging devices are summarized in this ERD chapter and treated more fully in the new Emerging Research Materials chapter.

A primary goal of this chapter is to stimulate invention and research leading to one or more new concepts to extend functional scaling of information processing substantially beyond “ultimately scaled” CMOS. This goal is accomplished by addressing the two technology-defining domains identified above—extending the CMOS platform via heterogeneous integration of new technologies and, later, via developing new technological and nanoarchitectural concepts to invent a new information processing paradigm.

The intent is two-fold. First is to gather in one place substantive, alternative concepts for memory and information processing devices and information processing nanoarchitectures that would, if successful, substantially extend the Roadmap beyond CMOS. As such, this discussion will provide a window into these candidate approaches. Second is to provide a balanced, critical assessment of these emerging new device technologies.

Further, a brief section is included to propose a set of fundamental principles that will likely govern successful extension of information processing technology substantially beyond that attainable solely with ultimately scaled CMOS. This broadened chapter, therefore, provides an industry perspective on emerging new device technologies and serves as a bridge between bulk CMOS and the realm of microelectronics beyond the end of CMOS dimensional and equivalent performance scaling.

The discussion is divided into the following three categories: 1) memory devices, 2) information processing or logic devices, and 3) information processing nanoarchitectures. The discussions provide some detail regarding their operation principles, advantages, challenges, maturity, and current and projected performance. Also included is a new device and architectural focus combining emerging research devices offering specialized, unique functions as heterogeneous core processors integrated with a CMOS platform technology. The goal is to exploit specialized functions provided by emerging devices operating in a parallelized architecture with CMOS processors to either realize current systems functions more efficiently or new systems functions. This represents the nearer term focus of the ERD chapter, with the longer term focus remaining on discovery of an alternate information processing technology to eventually replace digital CMOS.

---

1 Information processing refers to the input, transmission, storage, manipulation or processing, and output of data. The scope of the ERD Chapter is restricted to data or information manipulation, transmission, and storage.
2 Emerging Research Devices

As in previous editions of the ERD Chapter, this chapter includes “transition tables.” The purpose of the transition tables is twofold. The first is to track technologies that have appeared in or have been removed from the 2005 tables and so provide a very short explanation of the reason for this change. The second purpose is to identify technologies that are considered important but do not meet the criteria for full inclusion into the more detailed tables. These technologies may be expected to become more or less visible in future editions of the roadmap and hence the name “transition table.” In this way these transition tables provide the disposition of technologies transitioned from 2005 to 2007. It also gives a preview of new technologies that may be included in the 2009 edition and contains entries that will be tracked for possible future inclusion. Finally, inclusion of a concept in this chapter does not in any way constitute advocacy or endorsement of that concept. Conversely, not including a particular concept does not in any way constitute rejection of that approach.

DIFFICULT CHALLENGES

INTRODUCTION

The semiconductor industry is facing two classes of difficult challenges related to extending integrated circuit technology to and beyond the end of CMOS dimensional scaling. One set relates to extending CMOS beyond its ultimately scaled density and functionality by integrating, for example, a new high speed, dense, and low power memory technology on the CMOS platform. Another class of challenges is to extend information processing substantially beyond that attainable by CMOS alone using an innovative combination of new devices and architectural approaches for extending CMOS and, eventually, inventing a new information processing platform technology. Difficult challenges are presented in Table ERD1.

DEVICE TECHNOLOGIES

Difficult challenges related to emerging research devices are divided into those related to memory technologies and those related to information processing or logic devices. 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 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 to and beyond 16 nm. One approach to sustaining performance gains as CMOS scaling matures in the next decade is to replace the strained silicon MOSFET channel with an alternate material offering a higher quasi-ballistic carrier velocity and higher mobility than strained silicon. Candidate materials include strained Ge, SiGe, a variety of III-V compound semiconductors, and graphene. Introduction of non-silicon materials into the channel of otherwise silicon MOSFET (i.e., onto a silicon substrate) is fraught with several very difficult challenges. These challenges include fabrication of high-quality (i.e., defect free) channel materials on non-lattice matched silicon, minimization of band-to-band tunneling in narrow bandgap channel materials, elimination of Fermi level pinning on III-V and Ge surfaces, and fabrication of high-κ gate dielectrics on the passivated channel materials. Additional challenges are to sustain the 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 source/drain (S/D) doping concentrations.

A longer-term challenge is invention and reduction to practice of a new 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 solution. Solutions to this challenge beyond the end of CMOS scaling also may lead to new opportunities for such an emerging research device technology to eventually replace the CMOS gate as a new information processing primitive element.
## Materials Technologies

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

## Nano-Information Processing Taxonomy

Information processing to accomplish a specific system function, in general, requires several different interactive layers of technologies. A comprehensive top-down list of these layers begins with the required application or system function, leading to system architecture, micro- or nanoarchitecture, circuits, devices, and materials. As shown in Figure ERD1 below, a different bottom-up representation of this hierarchy begins with the lowest physical layer represented by a computational state variable and ends with the highest layer represented by a nanoarchitecture. In this more schematic representation, focused on generic information processing at the device/circuit level, a fundamental unit of information (e.g., a bit) is represented by a computational state variable, for example, the position of a bead in the ancient Abacus calculator or the charge or voltage state of a node capacitance in CMOS logic. A device provides the physical means of representing and manipulating a computational state variable among its two or more allowed discrete states. 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, is the various materials and processes necessary to fabricate the required device structure. The data representation is how the computational state variable is encoded by the assemblage of devices to process the data. Two of the most common examples of data representation are binary digital and continuous or analog signaling. The architecture plane encompasses three sub classes of this

<table>
<thead>
<tr>
<th>Difficult Challenges ≥ 22 nm</th>
<th>Summary of Issues and opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale high-speed, dense, embeddable, volatile and non-volatile memory technologies to and beyond 22 nm</td>
<td>SRAM and FLASH scaling will reach definite limits within the next several years (see PIDS chapter for Difficult Challenges). These are driving the need for new memory technologies to replace SRAM and FLASH memories. Identify the most promising technical approach(es) to obtain electrically accessible, high-speed, high-density, low-power, (preferably) embeddable volatile and non-volatile RAM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difficult Challenges &lt;22 nm</th>
<th>Summary of Issues and opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale CMOS to and beyond the 16 nm technology generation.</td>
<td>Develop new materials to replace silicon as an alternate channel to increase the saturation velocity and maximum drain current in MOSFETs while minimizing leakage currents and power dissipation for technology scaled to 16 nm and beyond. Candidate materials include Ge, SiGe, III-V compound semiconductors, and graphene. Develop 1D (nanowire or nanotube) structures to scale MOSFETs and CMOS gates beyond the 16 nm technology generation. Develop means to control the variability of critical dimensions and statistical distributions (e.g., gate length, channel thickness, S/D doping concentrations, etc.)</td>
</tr>
<tr>
<td>Extend ultimately scaled CMOS as a platform technology into new domains of application.</td>
<td>Discover and reduce to practice new device technologies and a primitive-level architecture to provide special purpose optimized functional cores heterogeneously integrable with silicon CMOS.</td>
</tr>
<tr>
<td>Continue functional scaling of information processing technology substantially beyond that attainable by ultimately scaled CMOS.</td>
<td>Invent and develop a new information processing technology eventually to replace CMOS. Ensure that a new information processing technology is compatible with the new memory technology discussed above; i.e., the logic technology must also provide the access function in a new memory technology. Bridge a knowledge gap that exists between materials behaviors and device functions.</td>
</tr>
</tbody>
</table>
4 Emerging Research Devices

Taxonomy: 1) nanoarchitecture or the physical arrangement or assemblage of devices to form higher level functional primitives to represent and enable execution of 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.

EMERGING RESEARCH DEVICES

MEMORY DEVICES

The memory technologies tabulated in this section are a representative sample of published research efforts (circa 2005-2007) selected to describe some attractive alternative approaches. Historically, very few memory research options yield practical memory devices. Existing research efforts are exploring a variety of basic memory mechanisms. These mechanisms include electronic charge isolated by surrounding dielectrics; remnant polarization on a ferroelectric gate dielectric; and resistance change caused by a variety of phenomena. Table ERD2 is an organization or taxonomy of the existing and emerging memory technologies into four categories. A strong 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 a CMOS platform technology. A goal is to present the end user with a device that behaves similar to the familiar silicon memory chip.
Because each of these new approaches attempts to mimic and improve on the capabilities of a present day memory technology, key performance parameters are provided in Table ERD3 for existing baseline and prototypical memory technologies. These parameters provide relevant benchmarks against which the current and projected performance of each new research memory technology may be compared.

The Emerging Research Memory technology entries in the current version of the roadmap differ in several respects from the 2005 edition. These changes in technology entries dropped and added to this section are captured in the Transition Table for Emerging Research Memory Devices (Table ERD4). The changes are: 1) drop nanofloating gate memory; 2) replace insulator resistance change memory with 3) fuse/anti-fuse memory, 4) ionic memory, and 5) electronic effects memory, and lastly 6) add an entry for nanomechanical memory. The reasons and motivations for these changes are given in Table ERD4. Also the table titled “Emerging Research Memory Devices” was split in two parts to indicate demonstrated and projected parameters: Table ERD5a: Capacitance-based Memory Devices and Table ERD5b: Resistance-based Memory devices.

This section is organized around a set of eight technology entries shown in the column headers of Tables ERD5a and ERD5b. These entries were selected using a systematic survey of the literature to determine the areas of greatest worldwide research activity. Each technology entry listed has several sub-categories of devices that are grouped together to simplify the discussion. Key parameters associated with the technologies are listed in the table. For each parameter, three numbers for performance are given that indicate: 1) minimum performance, satisfactory for practical application, 2) theoretically predicted performance values based on calculations and early experimental demonstrations, 3) up-to-date experimental values of these performance parameters reported in the cited technical references.

The last row in Tables ERD5a and ERD5b contains the number of papers on the particular device technology published in the last two years. It is meant to be a gauge of the amount of research activity currently taking place in the research community and it is a primary metric that determines which of the candidate devices are included in this table. The tables have been extensively cited and details may be found in the indicated references. The text associated with the table gives a brief summary of the operating principles of each device and as well as significant issues that are not captured in the table.

**MEMORY TAXONOMY**

Table ERD2 provides a simple way to categorize memory technologies. In this scheme, equivalent functional elements that make up a cell are identified. For example, the familiar DRAM cell that consists of an access transistor and a capacitor storage node is labeled as a 1T1C technology. Other technologies such as MRAM where data are stored as the spin state in a magnetic material can be represented as a 1T1R technology. Here the resistance “R” indicates that the cell readout is accomplished by sensing the current through the cell. The utility of this form of classification reflects the trend to simplify cells (i.e., reduce cell area) by reducing the number of equivalent elements to a minimum. Thus, early in the development of a given technology it is common to see multi-transistor multi-x (x equals capacitor or resistor) cells. As learning progresses, the structures are scaled down to a producible 1Tx form. The near ideal arrangement is to incorporate the data storage element directly into the transistor structure such that a 1T cell is achieved. In ultra-dense nanoelectronic memory arrays, instead of the transistor “T.” a two terminal non-linear diode-like element may be used with a resistive memory element. Such structure is represented as 1DIR technology.

An important property that differentiates emerging memory technologies is whether data can be retained when power is not present. Nonvolatile memory offers essential use advantages, and the degree to which non-volatility exists is measured in terms of the length of time that data can be expected to be retained. Volatile memories also have a characteristic retention time that can vary from milliseconds to (for practical purposes) the length of time that power remains on.
## Table ERD2 Memory Taxonomy

<table>
<thead>
<tr>
<th>Cell Element</th>
<th>Type</th>
<th>Non-volatility</th>
<th>Retention Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRAM</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>Phase change memory</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>Polymer memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Molecular memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Nanomechanical memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Fuse/antifuse memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Ionic memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Electronic effects memory</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>1T1R or 1D1R [A]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRAM</td>
<td>Volatile</td>
<td>~ seconds</td>
<td></td>
</tr>
<tr>
<td>FeRAM [B]</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>1T [A]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB DRAM [A]</td>
<td>Volatile</td>
<td>&lt; seconds</td>
<td></td>
</tr>
<tr>
<td>Flash memory</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>SONOS</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>Nano floating gate memory</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>Engineered tunnel barrier memory</td>
<td>Nonvolatile</td>
<td>&gt; 10 years</td>
<td></td>
</tr>
<tr>
<td>FeFET memory [A]</td>
<td>Nonvolatile</td>
<td>&gt; years</td>
<td></td>
</tr>
<tr>
<td>Multiple T [A]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRAM</td>
<td>Volatile</td>
<td>large</td>
<td></td>
</tr>
<tr>
<td>STTM [C]</td>
<td>Volatile</td>
<td>small</td>
<td></td>
</tr>
</tbody>
</table>

Notes for Table ERD2:

[A] 1T1R—1 transistor–1 resistor  1D1R—1 diode–1 resistor  1T1C—1 transistor–1 capacitor  1T—1 transistor  FB DRAM—floating body  DRAM  FeFET—ferroelectric FET  Multiple T—multiple transistor

[B] FeRAM—ferroelectric RAM with one ferroelectric transistor and one ferroelectric capacitor

### Table ERD3: Current Baseline and Prototypical Memory Technologies

<table>
<thead>
<tr>
<th>Storage Mechanism</th>
<th>DRAM</th>
<th>SRAM</th>
<th>Floating Gate</th>
<th>Trapping Charge</th>
<th>FeRAM</th>
<th>MRAM</th>
<th>PCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Technologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stand-alone [A]</td>
<td>Embedded [C]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge on a capacitor</td>
<td>Interlocked state of logic gates</td>
<td>Charge on floating gate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prototypical Technologies [E]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stand-alone [A]</td>
<td>Embedded [C]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge on a capacitor</td>
<td>Interlocked state of logic gates</td>
<td>Charge on floating gate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature size F, nm</th>
<th>2007</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>6E^2</td>
<td>12F^2</td>
</tr>
<tr>
<td>2022</td>
<td>6E^2</td>
<td>12F^2</td>
</tr>
<tr>
<td>Read Time</td>
<td>&lt;10 ns</td>
<td>1 ns</td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td>10 ns</td>
</tr>
<tr>
<td>W/E Time</td>
<td>&lt;10 ns</td>
<td>0.7 ns</td>
</tr>
<tr>
<td>2007</td>
<td>1/0.1 ms</td>
<td>20ns/20ns[HI]</td>
</tr>
<tr>
<td>Retention Time</td>
<td>64 ms</td>
<td>64 ms</td>
</tr>
<tr>
<td>2007</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
</tr>
<tr>
<td>2022</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
</tr>
<tr>
<td>Write Cycles</td>
<td>7.5</td>
<td>2.5</td>
</tr>
<tr>
<td>2007</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Read Operating Voltage (V)</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Write Energy (J/bit)</td>
<td>5E-15</td>
<td>5E-15</td>
</tr>
<tr>
<td>2007</td>
<td>1E-14</td>
<td>&gt;1E-14 [F]</td>
</tr>
<tr>
<td>Comments</td>
<td>Multiple-bit potential</td>
<td>Multiple-bit potential</td>
</tr>
<tr>
<td>Notes for Table ERD3:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **A** 2007 ITRS PIDS chapter.
- **B** Estimated as E~0.5*CV^2 for C=25fF, Vc=0.65 Volts (in 2007) and Vc=0.35 Volts in 2022 (energy to refresh is not included).
- **C** See the Embedded Memory Requirements table in the System Drivers chapter.
- **D** SRAM memory state is preserved so long as voltage is applied.
- **E** Embedded applications (see the Embedded Memory Requirements table in the System Drivers chapter).
- **F** Lower bound for Fowler Nordheim write/erase.
- **G** Trapping charge memories in PIDS chapter include SONOS, and a number of engineered barrier concepts, some of which are described in Table ERD3a.
- **L** Estimated as E~0.5*q*A*V for q=10.9 μC/cm^2, A=0.33μm^2, Vc=1.5 Volts (in 2007) and q=30 μC/cm^2, A=0.069μm^2, Vc=0.7 Volts (in 2022).
8 Emerging Research Devices


[Q] Estimated as E~0.5*I^2*R*tw for I=235 μA, R=3.54E3 Ohm, tw=50 ns (in 2007) and I=13 μA, R=3.54E4 Ohm, <50 ns (in 2022).

<table>
<thead>
<tr>
<th>Table ERD4</th>
<th>Transition Table for Emerging Research Memory Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN/OUT (Table ERD5)</td>
<td>Reason for IN/OUT</td>
</tr>
<tr>
<td>Nanofloating Gate Memory</td>
<td>OUT</td>
</tr>
<tr>
<td>Insulator Resistance Change Memory</td>
<td>OUT</td>
</tr>
<tr>
<td>Fuse/Antifuse Memory</td>
<td>IN</td>
</tr>
<tr>
<td>Ionic Memory</td>
<td>IN</td>
</tr>
<tr>
<td>Electronic Effects Memory</td>
<td>IN</td>
</tr>
<tr>
<td>Nanomechanical Memory</td>
<td>IN</td>
</tr>
</tbody>
</table>
### Table ERD5a  Emerging Research Capacitance-based Memory Devices—Demonstrated and Projected Parameters

<table>
<thead>
<tr>
<th></th>
<th>Engineered tunnel barrier memory</th>
<th>Ferroelectric FET memory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Storage Mechanism</strong></td>
<td>Charge on floating gate</td>
<td>Remnant polarization on a ferroelectric gate dielectric</td>
</tr>
<tr>
<td><strong>Cell Elements</strong></td>
<td>1T</td>
<td>1T</td>
</tr>
<tr>
<td><strong>Device Types</strong></td>
<td>FG FET with graded/multilayer gate insulator</td>
<td>FET with FE gate insulator</td>
</tr>
<tr>
<td><strong>Feature size F</strong></td>
<td>Minimum required: &lt;65 nm</td>
<td>Minimum required: &lt;65 nm</td>
</tr>
<tr>
<td></td>
<td>Best projected: 10 nm [A]</td>
<td>Best projected: 22 nm [I]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 130/90 nm [B]</td>
<td>Demonstrated: -2 μm [J]</td>
</tr>
<tr>
<td><strong>Cell Area</strong></td>
<td>Minimum required: 10 F²</td>
<td>Minimum required: 8F²/4F² [A, B]</td>
</tr>
<tr>
<td></td>
<td>Best projected: 8F²/4F² [A, F]</td>
<td>Best projected: 8F²/4F² [I]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 6F² [B, 4F² [C]</td>
<td>Data not available</td>
</tr>
<tr>
<td><strong>Read Time</strong></td>
<td>Minimum required: &lt;15 ns</td>
<td>Minimum required: &lt;15 ns</td>
</tr>
<tr>
<td></td>
<td>Best projected: 2.5 ns</td>
<td>Best projected: 2.5 ns</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 20 ns [D]</td>
<td>Demonstrated: 20 ns [H]</td>
</tr>
<tr>
<td><strong>W/E time</strong></td>
<td>Minimum required: 1 μs/10 ns</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected: 1 ns @ 9V [A]</td>
<td>Best projected: 2.5 ns [B]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: ~1 μs @11V [B]</td>
<td>Demonstrated: 20 ns [K]</td>
</tr>
<tr>
<td><strong>Retention Time</strong></td>
<td>Minimum required: &gt;10 y</td>
<td>Minimum required: &gt;10 y</td>
</tr>
<tr>
<td></td>
<td>Best projected: &gt;10 y</td>
<td>Best projected: &gt;1 y</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: &gt;10 y [B]</td>
<td>Demonstrated: &gt;30 days [L, M]</td>
</tr>
<tr>
<td><strong>Write Cycles</strong></td>
<td>Minimum required: &gt;1E5</td>
<td>Minimum required: &gt;1E5</td>
</tr>
<tr>
<td></td>
<td>Best projected: &gt;3E16</td>
<td>Best projected: &gt;3E16</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 1E5 [G]</td>
<td>Demonstrated: 1E12</td>
</tr>
<tr>
<td><strong>Write Operating Voltage (V)</strong></td>
<td>Minimum required: Application dependent</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected: &gt;3 V [E]</td>
<td>Best projected: &lt;0.9 V [I]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 6.5 [B]</td>
<td>Demonstrated: ±6</td>
</tr>
<tr>
<td><strong>Read Operating Voltage (V)</strong></td>
<td>Minimum required: 2.5</td>
<td>Minimum required: 2.5</td>
</tr>
<tr>
<td></td>
<td>Best projected: 0.7</td>
<td>Best projected: 0.7</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: 2.5 [D]</td>
<td>Demonstrated: 2.5 [D]</td>
</tr>
<tr>
<td><strong>Write Energy (J/Bit)</strong></td>
<td>Minimum required: Application dependent</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected: &gt;1E-15</td>
<td>Best projected: 2E-15 [N]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated: Data not available</td>
<td>Demonstrated: Data not available</td>
</tr>
<tr>
<td><strong>Comments</strong></td>
<td>Potential for multi-bit/cell storage</td>
<td>Potential for non-destructive readout</td>
</tr>
<tr>
<td><strong>Research activity [O]</strong></td>
<td>25</td>
<td>48</td>
</tr>
</tbody>
</table>
10 Emerging Research Devices

Notes for Table ERD5a:

[D] Based on floating gate and SONOS data (see Table ERD3). The read voltage and read time of all IT memory devices are expected to be similar.
[E] Based on minimum barrier height of 1.5 eV for nonvolatile charge retention.
[N] Calculated based on the parameters of scaled ferroelectric capacitor projected in Ref. [I].
[O] The number of referred articles in technical journals that appeared in the Science Citation Index database for 7/1/2005–7/1/2007.
### Table ERD5b: Emerging Research Resistance-based Memory Devices—Demonstrated and Projected Parameters

<table>
<thead>
<tr>
<th>Storage Mechanism</th>
<th>Nanomechanical Memory</th>
<th>Fuse/Antifuse Memory</th>
<th>Ionic Memory</th>
<th>Electronic Effects Memory</th>
<th>Macromolecular Memory</th>
<th>Molecular Memories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Elements</td>
<td>1T1R or 1DIR</td>
<td>1T1R or 1DIR</td>
<td>1T1R or 1DIR</td>
<td>1T1R or 1DIR</td>
<td>1T1R or 1DIR</td>
<td>1T1R or 1DIR</td>
</tr>
<tr>
<td>Device Types</td>
<td>1) nanobridge/cantilever 2) telescoping CNT 3) Nanoparticle</td>
<td>M -I-M (e.g., Pt/NbO/Pt)</td>
<td>1) cation migration 2) anion migration</td>
<td>1) Charge trapping 2) Mott transition 3) FE barrier effects</td>
<td>M-I-M (nc)-I-M</td>
<td>Bi-stable switch</td>
</tr>
<tr>
<td>Feature size F</td>
<td>Min. required &lt;65 nm</td>
<td>&lt;65 nm</td>
<td>&lt;65 nm</td>
<td>&lt;65 nm</td>
<td>&lt;65 nm</td>
<td>&lt;65 nm</td>
</tr>
<tr>
<td></td>
<td>Best projected 5-10 nm [B]</td>
<td>5-10 nm</td>
<td>5-10 nm</td>
<td>5-10 nm</td>
<td>5-10 nm</td>
<td>5 nm [U]</td>
</tr>
<tr>
<td>Cell Area</td>
<td>Min. required 10F²</td>
<td>10 F²</td>
<td>10 F²</td>
<td>10 F²</td>
<td>10 F²</td>
<td>10 F²</td>
</tr>
<tr>
<td></td>
<td>Best projected 5F²</td>
<td>8/5F² [H]</td>
<td>8/5F² [H]</td>
<td>8/5F² [H]</td>
<td>8/5F² [H]</td>
<td>5F²</td>
</tr>
<tr>
<td></td>
<td>Demonstrated Data not available</td>
<td>Data not available</td>
<td>Data not available</td>
<td>Data not available</td>
<td>Data not available</td>
<td>Data not available</td>
</tr>
<tr>
<td>Read Time</td>
<td>Min. required &lt;15 ns</td>
<td>&lt;15 ns</td>
<td>&lt;15 ns</td>
<td>&lt;15 ns</td>
<td>&lt;15 ns</td>
<td>&lt;15 ns</td>
</tr>
<tr>
<td></td>
<td>Best projected &lt;3 ns</td>
<td>&lt;10 ns</td>
<td>&lt;10 ns</td>
<td>&lt;10 ns</td>
<td>&lt;10 ns</td>
<td>&lt;10 ns</td>
</tr>
<tr>
<td></td>
<td>Demonstrated 3 ns [C]</td>
<td>Data not available</td>
<td>&lt;50 ns [G]</td>
<td>Data not available</td>
<td>~10 ns [S]</td>
<td>Data not available</td>
</tr>
<tr>
<td>W/E time</td>
<td>Min. required Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected &lt;1 ns [AB]</td>
<td>&lt;10 ns</td>
<td>&lt;20 ns [P]</td>
<td>&lt;20 ns [M]</td>
<td>&lt;10 ns</td>
<td>&lt;40 ns [U]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated 3 ns [C]</td>
<td>10 ns/5 μs [E]</td>
<td>&lt;50 ns [G]</td>
<td>100 ns [M]</td>
<td>10 ns [S]</td>
<td>0.2 μs [V]</td>
</tr>
<tr>
<td>Retention Time</td>
<td>Min. required &gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
</tr>
<tr>
<td></td>
<td>Best projected &gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
<td>&gt;10 y</td>
<td>Not known</td>
<td>Not known</td>
</tr>
<tr>
<td>Write Cycles</td>
<td>Min. required &gt;1E5</td>
<td>&gt;1E5</td>
<td>&gt;1E5</td>
<td>&gt;1E5</td>
<td>&gt;1E5</td>
<td>&gt;1E5</td>
</tr>
<tr>
<td></td>
<td>Best projected &gt;3E16</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
<td>&gt;3E16</td>
</tr>
<tr>
<td>Write operating voltage (V)</td>
<td>Min. required Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected Not known [D]</td>
<td>0.5/1</td>
<td>&lt;0.5 V [O]</td>
<td>&lt;3 V</td>
<td>&lt;1 V [R]</td>
<td>80 mV[Y]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated 1.5 V [A]</td>
<td>0.5/1 [E]</td>
<td>+0.6/-0.2 V [G]</td>
<td>3-5 V [L,M]</td>
<td>~2 [S]</td>
<td>~1.5 V [W]</td>
</tr>
<tr>
<td>Read operating voltage (V)</td>
<td>Min. required 2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Best projected 0.7</td>
<td>&lt;5</td>
<td>&lt;0.2 V [O]</td>
<td>0.7</td>
<td>0.7</td>
<td>0.3 [U]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated 1.5 V [A]</td>
<td>0.4 [E]</td>
<td>0.15 V[G]</td>
<td>0.7 V [L]</td>
<td>1 V [S]</td>
<td>0.5 V [W]</td>
</tr>
<tr>
<td>Write energy (J/bit)</td>
<td>Min. required Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
<td>Application dependent</td>
</tr>
<tr>
<td></td>
<td>Best projected Not known [D]</td>
<td>Not known</td>
<td>1E-15 [J]</td>
<td>&lt;1E-10</td>
<td>Not known</td>
<td>2E-19 [Y]</td>
</tr>
<tr>
<td>Comments</td>
<td>Inverse voltage scaling presents a problem</td>
<td>Potential for multi-bit storage</td>
<td>2 Mbit prototype chip demonstrated [G]</td>
<td>Potential for multi-bit storage</td>
<td>Low read voltage presents a problem</td>
<td>160 Kbit prototype chip demonstrated [V]</td>
</tr>
<tr>
<td>Research activity [Z]</td>
<td>22</td>
<td>30</td>
<td>47</td>
<td>44</td>
<td>77</td>
<td>90</td>
</tr>
</tbody>
</table>

Notes for Table ERD5b:


[C] www.nantero.com

Emerging Research Devices


[F] Estimated based on experimental data reported in Ref. [E]: $E \sim 0.5*V_0*I_0*t$, for $V_0=1$ Volt, $I_0=0.5 \mu A$, $t_0=10$ ns.


[I] Estimated based on experimental data reported in Ref. [G]: $E \sim 0.5*V_0*I_0*t$, for $V_0=0.6$ Volt, $I_0=10 \mu A$, $t_0=50$ ns.

[J] Estimated based on experimental data reported in Ref. [G]: $E \sim 0.5*V_0*I_0*t$, for $V_0=0.2$ Volts, $R_0=2E5 \Omega$, $t_0=10$ ns.

[K] Obtained in ref. [G] from elevated temperature accelerated data retention measurements over 30 h.


[Q] Electrochemical cell potentials control the write voltage. In appropriate combinations, 0.5 V will leave some safety margin. Read voltages will be significantly smaller.


[T] Estimated based on experimental data reported in Ref. [S]: $E \sim 0.5*V_0*I_0*t$, for $V_0=2$ Volts, $I_0=10 \mu A$, $t_0=10$ ns.


[2] The number of referred articles in technical journals that appeared in the Science Citation Index database for 7/1/2005~7/1/2007.

MEMORY DEVICES—DEFINITION AND DISCUSSION OF TABLE ENTRIES

A. Capacitance-based (Table ERD5a)

Engineered Tunnel Barrier Memory—The engineered tunnel barrier memory includes the concept of floating gate memory with graded or multilayer barrier insulator to improve write/erase properties. The graded (e.g., “crested”) barrier floating gate memory uses a stack of insulating materials to create a special shape of barrier enabling effective tunneling into and out of the storage node. This concept is very attractive, however the realization of a graded bandgap tunnel barrier fabricated using layered dielectrics is very difficult to achieve. This requires new dielectric materials with graded bandgap and graded dielectric constant. The concept of graded bandgap dielectric materials is analogous to the graded bandgap III-V heterostructures. It is noteworthy that a graded Al$_{1-x}$Ga$_x$As/GaAs structure was used to fabricate a graded charge injection barrier, the experimentally demonstrated predecessor of the graded injection barrier memory concept discussed above. The crested tunnel barrier stack structure Si$_x$N$_4$/SiO$_2$/Si$_x$N$_4$ was experimentally investigated and an improvement of its NVM characteristics was reported. Also, there were recent studies of Al$_2$O$_3$ and HfO$_2$ layers as a possible candidate for the graded tunneling barrier.

In the VARIOT memory, a dual layer dielectric stack with two different dielectric constants is used. This structure allows a high tunneling current at a relatively low applied voltage while providing good data retention. An engineered tunneling barrier allows the voltage necessary to program or erase the memory cell to be lowered. Stacks consisting of SiO$_2$ and HfO$_2$ or Al$_2$O$_3$ have been experimentally studied, and they demonstrated lower voltage programming by tunneling and ten years of data retention.

In another concept, the band-gap engineered SONOS memory (BE-SONOS), an ultrathin SiO$_2$/Si$_x$N$_4$/SiO$_2$ barrier is used to replace the tunnel oxide. Such a barrier creates a necessary asymmetry between erase and store modes.

Ferroelectric FET Memory—Conventional 1T1C ferroelectric RAM (FeRAM) requires switchable polarization charges in the same order as charges stored on a DRAM cell capacitor. This amount of required charge leads to the necessity of 3D folded ferroelectric capacitor and limits the scalability due to the capacitor area requirements. In contrast, if the...
ferroelectric capacitor is integrated into the gate stack of a FET, the ferroelectric polarization directly affects charges in the channel and leads to a defined shift of the input characteristics of the FET. This 1T memory device is called ferroelectric FET (FeFET)\textsuperscript{16}. At the channel interface, a high quality insulator will still be required to guarantee a low interface state density. Hence, the device realistically shows a metal-ferroelectric-insulator-semiconductor (MFIS) gate stack. Sometimes, another metal layer is introduced between the ferroelectric and the insulator (MFMIS). The FeFET device scales as a MOSFET. However, scaling is projected to end approximately with 22 nm, because the insulation layer becomes too thin and the properties of the ferroelectric with respect to thickness dependence of the coercive field will not allow further reduction.\textsuperscript{14} In the last decade, many attempts have been made to fabricate FeFET-based nonvolatile memories. The major challenge is the long-term reliability related to the ferroelectric-semiconductor interface. Recently, significant improvements in the retention time have been reported.\textsuperscript{12, 13} Other reliability issues such as fatigue and imprint are not yet thoroughly studied and understood.

**B. Resistance-based** (Table ERD5b)

Resistance-based memories consist of a nanoelectromechanical memory and of a range of metal-insulator-metal (MIM) systems, which show electrical pulse induced resistance change effects. The conceivable mechanisms of the resistive switching in MIM systems often consist of combination of physical and/or chemical effects. The mechanisms can be grouped according to a dominant contribution by thermal effects, by ionic effects, or by electronic effects\textsuperscript{14}. This section will first discuss nanoelectromechanical memory followed by discussion of three metal-insulator-metal systems.

**Nanoelectromechanical Memory (NEMM)—**The NEMM is based on a bi-stable nanoelectromechanical switch. In this concept, mechanical digital signals are represented by displacements of solid nano elements (e.g., nanowires, nanorods or nanoparticles), which results in closing or opening an electrical circuit. The original concept of NEMM was a carbon nanotube (CNT) cross-bar memory\textsuperscript{15}. Each memory element is based on a suspended crossed CNT. A cross-bar array of CNTs forms mechanically bi-stable, electrostatically-switchable device elements at each cross point, and the memory state is read out as the junction resistance. Several different modifications of NEMMS are currently being explored, including: CNT bridge\textsuperscript{16}, CNT cantilevers\textsuperscript{17}, and silicon cantilevers.\textsuperscript{18} Recently, a nanoelectromechanical device incorporating the nanocrystalline silicon (nc-Si) dots has been proposed.\textsuperscript{19} The nc-Si dots are embedded as charge storage centers in a mechanically bistable floating gate formed in a MOSFET structure. In addition to nonvolatile RAM, a nanoelectromechanical shift register memory has recently been proposed.\textsuperscript{20} A difficult challenge of NEMM is reliable fabrication of dense memory arrays due to material and fabrication issues, such as lack of control of CNT parameters.

**Unipolar Fuse/Antifuse Memory**—A typical resistive switching phenomena in a MIM structure based on a thermal effect is manifest in a unipolar characteristic. It is initiated by a voltage-induced partial dielectric breakdown in which the material in a discharge filament is significantly modified due to Joule heating. Due to the current compliance, only a weak conductive filament with a controlled resistance is formed. This filament may be composed of the electrode metal transported into the insulator, carbon from residual organics\textsuperscript{21}, or decomposed insulator material such as sub-oxides.\textsuperscript{22} During the reset transition, this conductive filament is disrupted thermally again because of high power density in the order of $10^{12}$ W/cm\textsuperscript{3} generated locally. This mechanism is referred to as the fuse–antifuse type. One candidate out of many is NiO, first reported in the 1960s.\textsuperscript{23} Recently, the filamentary nature of the conductive path in the ON-state has been confirmed for NiO\textsuperscript{24} and TiO\textsubscript{2}\textsuperscript{25}. Pt/NiO/Pt thin film cells have been successfully integrated into CMOS technology to demonstrate nonvolatile memory operation.\textsuperscript{26} A critical parameter for this unipolar switching effect seems to be the value of the current compliance. Note that while thermal effects seem to dominate in the unipolar resistive switching, there are indications that electrochemical processes also play a role.\textsuperscript{27, 28, 29} Recent studies of copper oxide resistive switching showed very strong electrode dependence.\textsuperscript{30} Switching current, and therefore the write energy, is relatively high in this type of memory. Another issue is relatively large dispersion of memory switching parameters. Recently, progress on both issues was demonstrated. Smaller switching current densities were reported\textsuperscript{31}, and the dispersion of all memory switching parameters was minimized by using thin IrO\textsubscript{2} layers between NiO and electrodes.\textsuperscript{32}

**Ionic Memory**—The memory operation is based on redox processes involving the electrode material or the insulator material, or both. The insulator is required to exhibit ionic conductivity. The material class is comprised of oxides, higher chalcogenide (including glasses), semiconductors, as well as organic compounds including polymers. One variant is based on the cation transport, the cathodic reduction and the growth of metallic filaments. Primarily Ag and Cu-based systems have been successfully realized in demonstrator cells.\textsuperscript{33, 34} A filament thus formed connecting opposite electrodes leads to a low-resistance state, while oxidation dissolves the filament and restores the high-resistance state. Other variants originate from ion (e.g., oxygen ion) transport and redox reactions that introduce an electronic conduction within the insulator material itself.\textsuperscript{35, 36} In most cases, a formation process is required before the bi-stable switching can be started. Often, the conduction is of filamentary nature. If this effect can be controlled, memories based on this bi-stable switching process can be scaled to very small feature sizes. The switching speed is limited by the ion transport. If the active distance
that is relevant for the redox controlled bi-stable switching is small (in the < 10 nm regime) the switching time can be as low as a few nanoseconds. Precise predictions are not yet possible, because the details of the mechanism of the reported phenomena are still unknown.

Electronic Effects Memory—The electronics effects memory includes three different mechanisms: 1) charge injection and trapping, 2) Mott transition, and 3) ferroelectric polarization effects:

1. Charge injection and trapping can be a cause for the changes in resistance. In the charge-trap model (Simmons-Verderber theory), charges are injected by Fowler-Nordheim tunneling at high electric fields and subsequently trapped at sites such as defects or metal nanoparticles within the insulator. This modifies the electrostatic barrier character of the MIM structure and, hence, the resistance of the structure. In a modified model, trapping at interface states is discussed to affect the adjacent Schottky barrier at various metal/semiconducting perovskite interfaces. Another example is Si nanowire memory, where the resistance of a polycrystalline Si nanowire is modulated with a small number of charges trapped at the grain boundaries within the wire. While this approach suffers lack of reproducibility, the elimination of the bulk MOSFET allows a significant reduction of the cell size. A 128 Mbit memory based on silicon nanowires has been demonstrated. A materials issue that requires further investigation is cycling fatigue–defect formation during switch cycles. The formation of defects may limit both the lifetime and the dimensional scaling due statistical uncertainty of defect density distribution.

2. In the Mott transition memory charge injection induces transition from strongly correlated to weakly correlated electrons, resulting in an insulator-metal transition. The Mott transition mechanism was reported for perovskite-type oxides such as (Pr, Ca)MnO$_3$ (PCMO), $SrTiO_3$:Cr and Ag/CoO$_2$/LCMO heterostructures. Recently, Mott transition resistive switching was also reported for the Pt/TiO$_2$/TiN/Pt system. A critical issue of this type of device is the sensitivity of the behavior of correlated electrons to small changes in parameters, including charge density, strain, disorder, and local chemical composition. Thus, precise control of physical and chemical structure of the material and interfaces is crucial.

3. The ferroelectric polarization can modify the tunneling properties of ultrathin films or modify the Schottky-type space-charge layer in adjacent semiconducting layers, resulting in ferroelectric resistive switching. At present, the ferroelectric origin of the observed switching phenomena was not definitely confirmed.

Macromolecular Memory—Macromolecular memory, sometimes referred to as polymer or organic memory, consists of a memory element, which is a thin-film of organic material with embedded metal components. The embedded metal components could be thin metal layer, metal nanoclusters, or metal ions in organometallic materials CuTCNQ and AgTCNQ (TCNQ=7,7,8,8-tetracyano-p-quinodimethane). A memory device based on C$_{60}$ molecules embedded in polymer film was recently reported. All these structures can exhibit two states of different conductivities, at the same applied voltage. The WRITE operation is performed by applying a voltage pulse to the structure, which results in reversible switching between a low-resistance and a high-resistance state. After transition occurs, the device remains in one of two states after turning off the power. The ERASE operation is performed by application of a reverse voltage pulse.

Experimental results suggest that the embedded metal layer plays a critical role in bistable $I–V$ characteristics of the Macromolecular Memory element. The memory operation mechanism is still unclear. It does not appear to be associated with the formation of conducting filaments, as in the case of the fuse/antifuse or ionic memories. Some researchers suggested that the mechanism of bistability could be explained by the Simmons-Verderber theory, according to which the changes in resistance are due to trapping the charge in the discrete metal nanocrystals, indicating that this type of memories falls into the class of electronic effect memories.

Molecular Memory—Molecular memory is a broad term encompassing different proposals for using individual molecules as building blocks of memory cells in which one bit of information can be stored in the space of a molecule. One experimentally demonstrated approach is based on rapid reversible change of effective conductance of a molecule attached between two electrodes controlled by applied voltage. In this molecular memory data are stored by applying an external voltage that causes the transition of the molecule into one of two possible conduction states. Data is read by measuring resistance changes in the molecular cell. There are also concepts for combining molecular components with current memory technology, such as DRAM and floating gate memory. The mechanism of conductivity switching in molecules is not completely understood. Several of the earlier reported experimental results on electron transport through molecules were found to be due to formation of metal filaments along the molecule attached between two metal electrodes. Hence, the intrinsic behavior of molecular switches may often be masked by other effects. The next step namely molecular interconnects between functionally active molecules is viewed as a long-term research goal. The knowledge base for molecular electronics needs further fundamental work.
LOGIC AND ALTERNATIVE INFORMATION PROCESSING DEVICES

Three previous editions of the ERD Logic section have evaluated alternative logic technology entries in terms of their potential to displace scaled CMOS devices in high performance general purpose computing. The conclusion reached in these previous editions was that none of the alternative technologies surveyed had a high potential for displacing the scaled CMOS devices on the ITRS roadmap scheduled for the end of the next decade. Therefore, in light of the huge amount of research activity taking place in alternative technologies around the world, the natural question to ask is whether there are some useful information processing functions other than general purpose Boolean logic where the particular physical characteristics of an emerging technology could offer an advantage relative to and combined with scaled CMOS. This edition of the Emerging Research Device chapter addresses this and related questions.

Also, as in previous sections of the ERD chapter, this section includes a transition table, Table ERD6 as shown below. The transition table provides the disposition of technologies transitioned into and out of Table ERD7a from 2005 to 2007. It also gives a preview of new technologies that may be included in the 2009 edition. Table ERD6 shown below contains four entries that will be tracked for possible future inclusion.

Any analysis of projected performance of alternative devices for non-Boolean applications is intimately related to the associated architectural configuration. As noted in the Emerging Research Architecture section, the current industry trend towards heterogeneous asymmetric multicore processors is consistent with the idea that future systems could support dedicated coprocessors utilizing novel devices for specialized applications. These dedicated coprocessors and accelerators would be integrated as one or more cores dedicated to specific operations in an otherwise conventional general purpose CMOS-based system. Specific examples might include an image recognition or speech recognition coprocessor, a Bayesian inference engine for data mining, or an associative memory unit for synthesis applications.

The Emerging Research Device Logic section is therefore expanded to include consideration of novel devices for both general purpose, Boolean logic as well as special purpose applications such as recognition, mining and synthesis. To this end, the section is organized around two tables rather than one. Table ERD7a is very similar to the logic table in previous editions of the ERD chapter. It contains parameters for the alternative technology entries relevant to evaluating performance of each technology for general purpose Boolean logic. The motivation for doing this is to provide a convenient scorecard to chart research progress in the alternative technology in the prior two years. The second table, Table ERD7b, evaluates some of the same novel devices but for new applications that the System Drivers chapter predicts to be of major importance over the next decade. It is well understood that the novel applications will require novel architectures but they may well also open the door for novel devices with response functions quite different from those of CMOS devices.
Table ERD6  Transition Table for Emerging Research Logic Devices

<table>
<thead>
<tr>
<th>IN/OUT</th>
<th>Reason for IN/OUT</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid Single Flux Quanta (RSFQ)</td>
<td>OUT</td>
<td>RSFQ devices, systems and circuits have been developed, prototyped, and fabricated. They could become an important technology if the correct market driver emerges. Design and fabrication lines for RSFQ systems exist. Cryogenic operation, cost and material integration issues limit application space.</td>
</tr>
<tr>
<td>CMOS extension-III-V channel replacements</td>
<td>IN</td>
<td>Low bandgap, compound III-V semiconductors can potentially improve transistor performance. Research on compound III-V semiconductors on Si substrates has increased significantly over the last 2 years.</td>
</tr>
<tr>
<td>Impact Ionization MOS</td>
<td>Possible Future</td>
<td>Simulation results showing very low sub threshold slopes indicate potential for low power operation. Reliability remains an issue may be included in future editions.</td>
</tr>
<tr>
<td>Nano Electro Mechanical Systems (NEMS)</td>
<td>Possible Future</td>
<td>Potential for ultra low leakage device based on nano relay operation. Issues associated with stiction, speed, active power and reliability are being studied—may be included in future editions.</td>
</tr>
<tr>
<td>Lateral interband tunneling transistor</td>
<td>Possible Future</td>
<td>Potential to utilize gate modulated interband tunneling to reduce subthreshold slope. May be included in future editions.</td>
</tr>
<tr>
<td>Floating gate MOS devices</td>
<td>Possible Future</td>
<td>Devices with nanocrystals embedded in gate allow circuits with tuneable thresholds. Potential for low power circuits. May be included in future editions.</td>
</tr>
</tbody>
</table>

**LOGIC DEVICES**

The emerging logic technology entries included in this edition of the roadmap differ slightly from previous versions. The differences are summarized below. The parameters listed in the logic table are those relevant to high performance general purpose computing. All the quantitative parameters have separate entries for best projected value and best demonstrated value and the values for each are referenced.

1. Extensions to CMOS—Low dimensional structures previously included carbon nanotube FETs, nanowire FETs, and nanowire heterostructures. This edition will also include devices using nanoribbon graphene.
2. Extensions to CMOS—MOSFETs including high mobility Ge and III-V compound semiconductor layers as channel replacement materials.
3. Single electron devices discussion has similar scope to previous editions.
4. Molecular devices has similar scope to previous editions with primary focus on molecule on CMOS architecture (CMOL) concept.
5. Ferromagnetic logic devices are based on collective magnetic effects associated with the magnetic polarity of a nanodomain.
6. Spin devices are based on spin dynamics of one or a few electrons, defects, or nuclei.
<table>
<thead>
<tr>
<th>Device</th>
<th>FET [A]</th>
<th>1D structures</th>
<th>Channel replacement</th>
<th>SET</th>
<th>Molecular</th>
<th>Ferromagnetic logic</th>
<th>Spin transistor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Typical example devices</strong></td>
<td>Si CMOS</td>
<td>CNT FET NW FET NW hetero-structures Nanoribbon transistors with graphene</td>
<td>III-V compound semiconductor and Ge channel replacement</td>
<td>SET</td>
<td>Crossbar latch Molecular transistor Molecular QCA</td>
<td>Moving domain wall M: QCA</td>
<td>Spin Gain transistor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spin FET</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spin Torque Transistor</td>
</tr>
<tr>
<td><strong>Cell Size (spatial pitch) [B]</strong></td>
<td>Projected</td>
<td>100 nm</td>
<td>100 nm [D]</td>
<td>300 nm [I]</td>
<td>40 nm [O]</td>
<td>10 nm [U]</td>
<td>140 nm [Y]</td>
</tr>
<tr>
<td><strong>Density (device/cm²)</strong></td>
<td>Projected</td>
<td>1E10</td>
<td>4.5E9</td>
<td>6.1E9</td>
<td>6E10</td>
<td>1E12</td>
<td>5E9</td>
</tr>
<tr>
<td></td>
<td>Demonstrated</td>
<td>2.8E8</td>
<td>7E7</td>
<td>3.5E7</td>
<td>~2E9</td>
<td>2E7</td>
<td>1.6E9</td>
</tr>
<tr>
<td><strong>Switch Speed</strong></td>
<td>Projected</td>
<td>12 THz</td>
<td>6.3 THz [F]</td>
<td>&gt;1 THz</td>
<td>10 THz [O]</td>
<td>1 THz [W]</td>
<td>1 GHz [Y]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated</td>
<td>1.5 THz</td>
<td>200 MHz [G]</td>
<td>&gt;300 GHz</td>
<td>2 THz [R]</td>
<td>100 Hz [V]</td>
<td>30 Hz [Z, AA]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated</td>
<td>5.6 GHz</td>
<td>220 Hz [H]</td>
<td>Data not available</td>
<td>1 MHz [P]</td>
<td>100 Hz [V]</td>
<td>30 Hz [Z]</td>
</tr>
<tr>
<td><strong>Switching Energy, J</strong></td>
<td>Projected</td>
<td>3E-18</td>
<td>3E-18</td>
<td>3.00E-18</td>
<td>1×10⁻¹⁰ [O] [&gt;1.5×10⁻¹¹] [S]</td>
<td>5E-17 [X]</td>
<td>~1E-17 [Z]</td>
</tr>
<tr>
<td></td>
<td>Demonstrated</td>
<td>1E-16</td>
<td>1E-11 [H]</td>
<td>1E-16 [J]</td>
<td>8×10⁻¹⁷ [T] [&gt;1.3×10⁻¹⁴][S]</td>
<td>3E-7 [V]</td>
<td>6E-18 [AA]</td>
</tr>
<tr>
<td><strong>Binary Throughput, GBit/ns/cm²</strong></td>
<td>Projected</td>
<td>238</td>
<td>238</td>
<td>61</td>
<td>1000</td>
<td>5E-2</td>
<td>Not known</td>
</tr>
<tr>
<td></td>
<td>Demonstrated</td>
<td>1.6</td>
<td>1E-8</td>
<td>Data not available</td>
<td>2E-4</td>
<td>2E-9</td>
<td>5E-8</td>
</tr>
<tr>
<td><strong>Operational Temperature</strong></td>
<td>RT</td>
<td>RT</td>
<td>RT</td>
<td>RT [M, N]</td>
<td>RT</td>
<td>RT</td>
<td>RT</td>
</tr>
<tr>
<td><strong>Materials System</strong></td>
<td>Si</td>
<td>CNT, Si, Ge, III-V, In₂O₃, ZnO, TiO₂, SiC, InGaAs, InAs, InSb</td>
<td>III-V, Si, Ge, Organic molecules</td>
<td>Ferromagnetic alloys</td>
<td>Si, III-V, complex metals oxides</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Research Activity [AD]</strong></td>
<td>379</td>
<td>62</td>
<td>91</td>
<td>244</td>
<td>32</td>
<td>122</td>
<td></td>
</tr>
</tbody>
</table>
18 Emerging Research Devices

Notes for Table ERD7a:

[B] The effective dimension that one transistor occupies on the MPU chip floor space. For CMOS MPU chips, the relation between cell size and Lg holds approximately constant by scaling: cell size = 20Lg.
[C] Lg = 5 nm.
[D] Size and circuit speed scaling of these structures is the same as the scaling of MOSFETs.
[R] In notation [O] above, the reported number of 2 THz for “intrinsic speed” of an experimental SET was derived from capacitance measurements, and not from experimental time-dependent characteristics.
[S] The value in the [ ] is the value that includes cooling energy. If an ideal Carnot refrigerator is used for cooling to the operation temperature Tc, the total switching energy

\[
c_{sw} = \frac{300}{Tc} \times E_c
\]

total switching energy

\[ E_c = \frac{300}{Tc} \]

\[ E_c \]

where \( E_c \) is the net switching energy, when cooling energy is not taken into account.

[AD] The number of articles in technical journals that appeared in the Science Citation Index database for July 2005–July 2007.

LOGIC DEVICES—DEFINITION AND DISCUSSION OF TABLE ENTRIES

Extensions to CMOS: Low-dimensional Structures—There are many efforts currently underway to extend traditional CMOS devices. Two of the primary approaches involve strategies to replace the FET channel with novel high carrier mobility materials. Some of the materials being considered are not semiconducting in their bulk form and display a semiconducting band structure only under quantum confinement. The three principle types of quantum-confined structures in the research community today are carbon nanotubes, nanowires, and graphene nanoribbons. Recent research activity in nanowire (NW) and carbon nanotube (CNT) devices can in general be divided into the three general categories of (1) experimental growth and assembly, (2) CNT and NW device fabrication and characterization, (3) CNT and NW circuits and integration.

Important progress has been the demonstration of various avenues for gaining better control over the chirality of nanotube materials by means of adjusting the plasma parameters of the PECVD chamber and catalyst engineering respectively.
An alternative approach for controlling the nanotube characteristics is a selective etch of metallic nanotubes from assembled films. Such methods can yield up to 96% semiconducting tubes. While individual nanotubes have been selectively doped with N and P type dopants, doping of nanotubes arrays has not yet been demonstrated. Progress has been made in understanding the role of contacts as a source of 1/f noise.

Important progress has also been made toward better characterization of miniaturized CNT structures involving a novel measurement technique for directly obtaining both quantum and electrostatic capacitances of nanotubes and nanowires. As a result, mobility can now be deduced directly from electrical measurements of the devices, confirming the high carrier mobility values that were previously estimated for these materials. The high frequency operation limit of nanotubes and nanowires still remains to be explored. While a number of techniques have been demonstrated for indirect RF characterization, the parasitic capacitances and resistances associated with these high impedance devices have prevented direct RF characterization. Recent advancements in controlling the assembly of 1D materials has been made by various groups demonstrating different strategies for assembling parallel arrays of nanotubes and nanowires on substrates, with an average pitch approaching 50 nm or less. However, better control in the pitch is necessary and requires further innovation in the assembly and fabrication technology. Importantly, devices based on the assembled nanowire parallel arrays have been characterized and deviations of electrical properties have been correlated with deviation in nanotubes properties. Deviations of the order of ~15% in the ON current are typical. Reducing and controlling the variation of the electrical properties is of particular importance for circuit and system integration of nanomaterials. Refer to the 2007 ITRS Emerging Research Materials chapter.

Heterogeneous composite nanowire structures have been configured in both core-shell and longitudinally segmented configurations using group IV and III-V materials. The longitudinally segmented configurations are epitaxially grown so that the material interfaces are perpendicular to the axis of the nanowire. This allows significant lattice mismatches without significant defects. Vertical InAs transistors have been fabricated in this manner with quite good characteristics. Core-shell gate-all-around configurations in general display excellent gate control and short channel effects. Interconnects for all vertical transistor configurations are more complex than for bulk transistors. Several groups have demonstrated circuit functionality of nanotubes and nanowire devices including a CMOS ring oscillator that was successfully fabricated on one single carbon nanotube, exhibiting ~13 MHz operation. The performance of the test circuit was limited by the device capacitance and was far from the THz operation regime that is theoretically predicted.

Two-dimensional graphene films are well known to behave as high mobility zero bandgap semiconductors with carrier mobilities approaching \(10^4\) cm\(^2\)/V sec. When patterned to sufficiently small ribbon widths, the graphene ribbons begin to display a finite band gap resulting from quantum confinement effects leading to a great deal of interest in graphene as a interesting channel replacement material in FET structures. Key material properties and issues are discussed more fully in the Emerging Research Materials chapter.

Extensions to CMOS: High Mobility Channel Replacement Materials—Some compound III-V narrow bandgap semiconductors have high electron mobility both in bulk and in thin film forms and offer potential performance advantages to Si-based CMOS devices if process integration and fabrication issues can be resolved. Primary materials being considered are InSb, InGaAs, InAs, InAsSb, and Ge, but many other combinations are of interest. Specific research issues being considered include defect mitigation in III-V epitaxy on silicon; high-k gate dielectrics; scalable enhancement mode device architectures; Fermi-level pining at dielectric/semiconductor interfaces; low-hole mobility in III-V compound semiconductors; scalable self-aligned structures; and many others. The literature on III-V materials and devices is extensive and while a thorough review is beyond the scope of this section, a good review summarizes the current status of the technology.

Single-electron Transistors (SETs)—SETs are three-terminal switching devices that convey electrons from source to drain one at a time by a tunneling mechanism. Potentially, SETs can be applied to general purpose Boolean logic but significant circuit and architecture innovation will be required. In such applications, they can potentially deliver high device density and power efficiency at good speed if the issues of the large threshold voltage variation and charging parasitic capacitances can be solved.

New applications and architectures that exploit the unique functionality of room temperature operating SET circuits have been developed, especially by monolithic integration of SETs with FET circuits to complement the conventional Si CMOS performance. Representative examples include SET/CMOS hybrid multi-value logic circuits, multiband filtering circuits, analog pattern matching circuits (discussed more fully in the next section), associative recognition tasks, and others, in which characteristic Coulomb blockade oscillations of SETs are typically utilized to reduce the number of devices. Note that certain aspects of the circuit performance, especially the room temperature operation, already
20  Emerging Research Devices

exceed the theoretical evaluation of the logic gate parameters for 2 nm SETs. These devices have a theoretically estimated maximum operation temperature T~20 K, integration density n~10^{11} cm^{-2}, and speed of the order of 1 GHz. However, large threshold voltage variation continues to impede realization of a large scale SET circuits, making it difficult for SETs to compete directly with CMOS devices used to implement Boolean logic operations. Engineering breakthroughs are needed to eliminate the size and background charge fluctuations among SETs necessary to suppress the threshold voltage variations.

The majority of the SET circuits demonstrated to date employ so called “voltage state logic” where a bit is represented by the voltage of capacitor charged by many electrons. The problem of the low fan-out for this scheme can be overcome by reducing the capacitance and/or by combining with conventional FET circuits. Truly single-electron approaches, representing a bit by a single electron (“bit state logic”) and the use of a single electron as a source of random number generations, have been limited to laboratory demonstrations. The problem of the limited fan-out, which is caused by using only a single electron in the truly single-electron devices, may be solved by innovative circuit designs such as the binary-decision-diagram. Discussion of the materials’ issues related to SETs is presented in the new Emerging Research Materials chapter.

Molecular Devices—The potential of molecular devices for general purpose computing is based on high densities, a large variety of molecular characteristics, uniformity of molecular characteristics, self assembly, and very low power. The functionality of molecular elements in electronic circuits arises either from their ability to store electronic charge and function as Coulomb blockade devices or as variable resistive elements that depend on the configuration state of the molecule. In the past two years, significant research progress has been made in improving the desirable characteristics of the molecules themselves as well as a great deal of progress in theoretical studies of potential architectural configurations that would utilize the molecular elements. However, very significant problems in molecular synthesis, device fabrication, circuit fabrication, reliability, and contacting remain and the gulf between the projected parameters and observed parameters is quite large.

The use of molecules as programmable diodes (i.e., latching switches) is the core technology underlying most of the concepts for future applications. These fall into three categories including filament formation and dissolution, charge trapping or storage, or configurational change. Reproducibility and repeatability of experimental measurements of speed, resistivity, power dissipation, and reliability typically vary by orders of magnitude between approaches and experimental groups. However, experimental repeatability within certain experimental groups has shown significant improvement over the last few years.

There has been a great deal of research effort devoted to developing architectural concepts for hybrid CMOS/nanoelectronic systems. These are based on conventional CMOS devices connected to nanowire arrays with molecular elements functioning as programmable diodes as discussed above and are generically referred to as CMOL systems. Simulations of CMOL systems have been applied to a variety of applications including FPGAs, image recognition, and crossbar memories and predict very attractive performance potential relative to scaled CMOS systems assuming the core technological issues can be resolved. At the present time, there have been no successful demonstrations of hybrid CMOS/nano device circuits.

Ferromagnetic Devices—Ferromagnetic logic devices are a class of alternative logic devices that use the local magnetization orientation of a domain of ferromagnetic material to store the computational state. In the nomenclature adopted here, ferromagnetic devices are distinct from spin devices, which are based on the individual dynamics of one or a few charge carriers to store and manipulate computational state or on spin-dependent electron transport between a source and a drain of a transistor structure. Ferromagnetic devices have the potential of being non-volatile and radiation hard, which is derived from the properties of the ferromagnetic materials themselves. While many ferromagnetic metals have Curie temperatures well above room temperature, the Currie temperatures of most ferromagnetic semiconductors are still limited to well below room temperature.

Ferromagnetic logic devices can be fabricated with ferromagnetic, metallic wires patterned to form Boolean logic devices. Their potential was recognized when the propagation of domain wall boundaries separating magnetic nanodomains reached a velocity of several hundred meters per second. This discovery led to geometric realization of NOT gates, AND gates, fanout structures, cross-over structures, and shift registers using the domain wall movement driven by the external magnetic field.

More recently, it was shown that single-domain wall propagation in a magnetic nanowire could be driven at room temperature by current flow through metallic nanowires. That discovery was followed by the current-induced domain-
 Characteristics of domain wall logic devices depend on the nanoscopic mechanisms of the current-induced domain wall movement; whether by transfer of electron momentum or of angular momentum. Basic studies probing such details can lead to new device structures, e.g., use of spin polarized currents to move domain walls in the specially shaped metal for the memory and amplifier applications. Proposals to exploit the nonvolatile nature of ferromagnetic elements as latches and registers have been proposed for applications in self-check pointing microprocessors and FPGAs.

**Single Spin Devices**—Single spin devices include concepts and devices that use individual spins to store a logical bit as well as devices that use spin dependent transport phenomena to help modulate the drain current of a FET. The original concept envisioned spins of localized electrons confined in an array of quantum dots as logic bits. The orientation of adjacent spins was coupled through the quantum exchange interaction. Schemes with electrostatic lateral confinement have been demonstrated, but they are too sensitive to variations and lead to bit size larger than a few microns. Layouts implementing AND and OR gates have been proposed. Alternatively, the circuit can be based on the majority logic gates, and applied to spins in quantum dots rather than nanomagnets. Neighbor bits are switched due to exchange interaction between electrons (that is spin-dependent). The strength of this interaction can be tuned by applying voltage and thus changing the overlap of the evanescent tails of electron wave functions in the two adjacent quantum dots. The signal in the spin circuit is passed from one neighbor bit to another and propagates in a manner similar to a spin wave in a ferromagnetic wire. A major problem of this class of logic is the presence of back action of downstream bits on the upstream ones. One of the solutions is clocking operation such that the calculation is interrupted and reset before the reflected spin wave reaches the input bits.

An important advantage of this class of logic is that it need not rely on a potential barrier to separate the two distinct logic states. For a system near thermal equilibrium, the height of this barrier must be more than the energy of thermal fluctuations and results in the von Neumann-Landauer limit of switching energy per bit of kTln2. Unlike other classes of logic, the relaxation rate of spin states theoretically can be much slower than the switching rate. The spin bits could then be switched many times before they return to thermal equilibrium. It was theoretically suggested that energy for switching a bit could be substantially smaller than the kTln2 limit. A related concept is based on varying (e.g., via spin-orbit interaction) the matrix element corresponding to coupling between the two spin states. This work also claims a theoretical limit of less than kTln2 switching energy. It should be noted, however, that a large external magnetic field may be required to control the state by an external stimulus that will result in high energy consumption.

Spin MOSFET devices are defined as novel metal-oxide-semiconductor FETs consisting of a MOS gate structure and ferromagnetic or half-metallic-ferromagnet (HMF) contacts for the source and drain. In principle spin MOSFETs have large magnetocurrent ratio, high transconductance, high gain (voltage, current and/or power), small power-delay product and small off-current. Spin injection from ferromagnetic metal contacts into semiconductor channels significantly limits performance of such devices. Recently, advanced spin transistors have been fabricated in epitaxially grown, single crystal, GaMnAs structures demonstrating good on/off switching ability and clear current amplification at T=2.6 K.

**ALTERNATIVE INFORMATION PROCESSING DEVICES**

The previous Logic Devices section considered emerging research devices in the context of the suitability for general purpose logic operations. The context assumed von-Neumann style computing where information is represented in a binary form and algorithmic operations are decomposed into a sequential application of the Boolean logic operators. The Logic Devices section is a continuation of the organizational structure used in the previous editions of the Emerging Research Device chapter.

This section on alternative information processing devices represents a significant departure from previous editions. As mentioned previously, it is motivated by the observation that some of the emerging research devices may have unique physical response characteristics better suited for applications involving recognition, mining and synthesis than general purpose computation. The current industry trend toward heterogeneous multicore systems, as noted in the architecture section, will, in principle, allow inclusion of such devices into a hybrid system architecture combining special purpose processors containing novel devices with more conventional, general purpose processors integrated onto a silicon CMOS platform.
The characteristics of present day CMOS devices have been optimized to give a strongly non-linear bi-stable response function that maximizes the $I_{on}/I_{off}$ ratio and minimizes the sub-threshold slope. These characteristics are very well suited to binary logic operations and in some sense, the CMOS device may be thought of as the natural device for binary Boolean logic. However, the physics of some of the alternative devices being investigated are quite different from the physics of CMOS devices and yield non-linear response characteristics quite different from the strongly bi-stable response of CMOS devices. Examples include response functions with a peaked structure (either in voltage or frequency), response functions with more than two stable states, periodic response functions and sigmoidal response functions with tunable slopes and curvatures. Each of these response functions may be associated with some “natural” application or algorithmic application. Some of those applications will be discussed here.

The other factor relevant to consideration of alternative information processing devices is that some, if not most, future information processing will be done on information where the state variable is something other than electronic charge. This information includes optical images, image sequences, speech, and data sets derived from physical sensors. This data or the signal is usually analog in nature. If the processing of analog signals is to be done in general purpose computers, data conversions must be counted as part of the information processing costs. In some of these cases, it may be more efficient to process the data in its original analog representation rather than convert everything to a digital (i.e., binary, electrical) representation and use a general purpose processor. In other cases, especially those in which total power or total energy is strongly constrained, a hybrid analog/digital system may be optimal.

### Table ERD7b  Alternative Information Processing Devices

<table>
<thead>
<tr>
<th>State Variable</th>
<th>Response Function</th>
<th>Class—Example</th>
<th>Architecture</th>
<th>Application</th>
<th>Comments</th>
<th>Status</th>
<th>Material Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge</td>
<td>Negative differential resistance</td>
<td>Mobile</td>
<td>Heterogeneous</td>
<td>Elements in hybrid magneto electric circuits</td>
<td>Additional functionality</td>
<td>Demo</td>
<td>Stray charge</td>
</tr>
<tr>
<td>Dielectric and magnetic domain polarization</td>
<td>Four resistive states</td>
<td>Multi-ferroic tunnel junction</td>
<td>Morphic</td>
<td>Analog pattern matching</td>
<td>Density, functionality</td>
<td>Demo</td>
<td>Demo</td>
</tr>
<tr>
<td>Charge</td>
<td>Staircase I/V from Coulomb blockade</td>
<td>Voltage tunable transfer function</td>
<td>Heterogeneous, morphic</td>
<td>Associative processing , NP complete</td>
<td>Density, cost functionality</td>
<td>Demo</td>
<td>Demo</td>
</tr>
<tr>
<td>Molecular conformation</td>
<td>Hysteric</td>
<td>CMOL, cross bar latch</td>
<td>MQCA, morphic</td>
<td>Elements in hybrid magneto-electric circuits</td>
<td>Radiation hard, environmental rugged</td>
<td>Demo</td>
<td>Demo</td>
</tr>
<tr>
<td>Ferromagnetic polarization</td>
<td>Nonlinear</td>
<td>Amplifiers, buses, switches</td>
<td>Morphic</td>
<td>Microwave power, tunable rectifiers</td>
<td>RF functionality</td>
<td>Demo</td>
<td>Simulation</td>
</tr>
<tr>
<td>Precession frequency</td>
<td>Nonlinear</td>
<td>Spin torque oscillator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ALTERNATIVE INFORMATION PROCESSING DEVICES—DEFINITION AND DISCUSSION OF TABLE ENTRIES**

The entries in the Alternative Information Processing Device table include novel devices that may prove useful for various information processing tasks other than high performance, general purpose computing. Some of the more specialized tasks include associative processing, communication, multivalued logic and ferromagnetic elements for nonvolatility and radiation hardness and error tolerance. In general, these may require a functional organization other than von Neumann architecture. The task or application and architectural configuration are noted in the table and linked to the entries in the Emerging Research Architecture section.
Resonant Tunnel Device\textsuperscript{133, 134}—Resonant tunnel devices (RTD) are widely recognized as being inherently high-speed devices and having a load curve characterized by a region of negative differential resistance. There have been many attempts to adapt the RTD structure to conventional Boolean logic gates with very limited success. These efforts usually involved integrating pairs of RTDs with a CMOS gate to achieve bi-stable operation. Several different implementations of such combinations in various circuit configurations were well described in the 2005 edition of the Emerging Research Devices chapter of the ITRS.

Recent research efforts have focused on adapting the inherent characteristics of resonant tunneling diodes (high speed, negative differential resistance (NDR) in novel ways that may be more closely matched to their inherent device characteristics. RTDs combined with an array of self-assembled quantum dots in a CNN connection pattern can be applied very effectively to image processing tasks such as edge detection, image recognition, and noise reduction.\textsuperscript{135} The basic logic cell of the proposed structure consists of two RTDs connected in series through a quantum dot. The local interconnections between the nanocells are achieved via tunneling in the quantum dot layer.

If two RTDs are connected in series with opposing polarities, they have two stable operating points and can switch between the two stable points. This results in a monostable-bistable transition logic element (MOBILE) configuration.\textsuperscript{136} Recent work\textsuperscript{137} has demonstrated the application of MOBILE circuits to ultra high speed A to D converters. MOBILES have also been applied to high speed (80Gb/s) multivalued quantizers and chaos generators useful for Monte Carlo simulations. Utilization of MOBILE type structures may be applied to multi-valued threshold circuits and multi-threshold gates in general.\textsuperscript{138}

Adding a control terminal to RTDs extends their usability to a variety of applications. This approach has been used to build resonant tunneling transistors (RTT).\textsuperscript{139} RTTs have a negative transconductance that can be used in several logic circuits, e.g., in XOR gate with only one transistor.\textsuperscript{140}

A number of recent works explore spin-polarized resonant tunneling, which could be useful for application in spintronic devices.\textsuperscript{141, 142, 143} Another potential niche application for RTDs is in photodetectors for detection of single photons with low dark current count rates and high efficiency.\textsuperscript{144}

Overall, the resonant tunneling devices may be useful for information processing applications other than Boolean logic that require high speed but need low dynamic range and low peak currents. However, they have serious issues to overcome. One is they must be scalable to lateral dimensions substantially less than 1 μm, below which their total current-voltage characteristic at this time is dominated by undesirable surface leakage current. Second the precise control of the layer thicknesses and properties of the RTDs may require use of commercial molecular beam epitaxy to achieve the required control.

Multiferroic Tunnel Junctions (MFTJs)—Ferroelectric materials exhibit a stable and switchable electrical polarization that is manifested in the form of cooperative atomic displacements. Similarly, ferromagnetic materials exhibit a stable and switchable magnetization that arises through the quantum mechanical phenomenon of exchange. There are very few “multiferroic” materials that exhibit both of these properties simultaneously. A subset of multiferroic materials demonstrates a dynamic “magnetoelectric” coupling of magnetic and electrical properties such that a change in the dielectric orientation of the material induces a ferromagnetic change and visa versa. These materials are discussed at length in the Emerging Materials chapter and offer significant potential for use in future hybrid magneto electronic circuits.

One of the first implementations of a device based on the multiferroic properties of a material is a multiferroic tunnel junction.\textsuperscript{145} In principle, it is possible to encode information independently in electric polarization and magnetization to obtain four different logic states. Direct electrical readout can be achieved if a multiferroic material is used as the tunnel barrier in a magnetic tunnel junction. Thin films of ferromagnetic-ferroelectric La_{0.1}Bi_{0.9}MnO_{3} (LBMO) that retain both ferroic properties down to a thickness of only 2 nm have been demonstrated. The films are then used as spin-filtering tunnel barriers the magnetization and electric polarization of which can be switched independently. In that case, the tunnel current across the structure is controlled by both the magnetic and ferroelectric configuration of the barrier, which gives rise to four distinct resistance states.

A device with four independent internal states could have obvious applications in multi-valued logic circuits, specifically four-state logic circuits. They could also be used as the storage element in dense memory circuits where effectively four bits are stored in one element.\textsuperscript{146}

Single Electron Transistors (SETs)—SETs suffer from low noise immunity and limited fan-out relative to conventional CMOS devices when used for conventional Boolean logic gates. However, the non-linear current-voltage characteristics
Molecular Devices—Molecular devices are characterized by individual or groups of molecules whose state (physical conformation or chemical state) may be changed by appropriate stimulation. Often these transitions are mediated by a bistable potential, with one of the states being thermodynamically stable and the other kinetically stable with a barrier between these states. The significant potential of molecular devices arises from significant densities that can be attained; the ability to synthesize molecules with different characteristics; the ability to self-assemble (for example, bilayer of surfactant molecules acting as a membrane into which ion channels can be integrated similar to biological cells); very low the ability to reversibly mediate reaction pathways via the environment (example pH); and the ability to change state or chemical state) may be changed by appropriate stimulation. Often these transitions are mediated by a bi-stable potential, with one of the states being thermodynamically stable and the other kinetically stable with a barrier between these states. The significant potential of molecular devices arises from significant densities that can be attained; the ability to synthetize molecules with different characteristics; the ability to self-assemble (for example, bilayer of surfactant molecules acting as a membrane into which ion channels can be integrated similar to biological cells); very low power; the ability to reversibly mediate reaction pathways via the environment (example pH); and the ability to change state via electrical, optical, or chemical means. Compared to chemical stimulation, photo-chemical and electrochemical stimulation can be switched on and off easily and more rapidly.

Conventional logic functions have been conceived with molecules and polymers (macromolecule), but other applications are considered in this section. Molecular schemes for combinatorial logic have been identified and neural type systems have been envisaged. In addition to switching, molecular Brownian motion and attachment have been used for computing. A DNA computer using DNA molecules utilizing self-assembly to perform computational steps in test-tubes can solve the “Traveling Salesman” problem significantly faster than a powerful computer.

Solution-based approaches have been conceived and fabricated that are capable of playing a game of Tic-Tac-Toe against a human opponent. In general, solution-based computing works because chemicals that react in ways that simulate logic functions and the output may be sensed spectroscopically. In this scenario, a molecular automaton, called MAYA, encodes all possible states of the tic-tac-toe game board as a particular deoxyribizome that react in a set of wells to simulate moves. The automaton is a Boolean network of deoxyribozymes that incorporates 23 molecular-scale logic gates and one constitutively active deoxyribozyme arrayed in nine wells corresponding to the game board. Significant research efforts are being devoted to implementing such approaches with nano fluidics in order to increase the speed and decrease the quantity of reagents needed.

Molecular Cascades seek to bridge the gap between physical chemistry and computer architectures. Temporal logic is used to characterize molecular interactions and specify the behavior of logic gates. Model-checking techniques are used for the exploration of structures behaviorally equivalent to the logic gates. A complete library of combinational logic gates has been designed using a particular molecular system.
Another unique role of electrical molecular devices is in CMOL systems that integrate the best features of CMOS with those of molecular devices with 2-terminal molecular devices self-assembled on a nanowire crossbar framework. This has potential for FPGA applications as well as neuromorphic networks for advanced information processing such as pattern recognition.

The key challenges for molecular devices include the ability to electrically stimulate and measure response or the state. Electrical signal communication between molecules is challenging. In some systems, protons have been used to communicate signals. Optical signal communication is also being investigated. Tunneling transport between molecular wires and devices is actively being researched and may be a viable option. Currently, spectroscopic analysis appears to be one of the ways of identifying molecular state.

**Ferromagnetic Devices**—One of the motivations for using the magnetic polarization of a ferromagnetic nanodomain as a state variable is that the individual spins of the charge carriers act in concert with each other and have a single degree of freedom, thereby reducing its noise sensitivity and possibly reducing the switching energy. It is of interest to consider technologies that rely of the collective ferromagnetic behavior of groups of charge carriers that are manipulated en mass. However, there may be a scaling penalty related to such a collective ferromagnetic system. In contrast, the next section will deal with alternative devices that rely on manipulating one or a few charge carriers and in particular, the magnetic precession induced by spin torque transfer.

Any future information processing technology that uses magnetic polarization of a ferromagnetic nanodomain as an alternative state variable will need to inject, transport, manipulate, store, amplify, and interface these ferromagnetic states. A rather broad set of ferromagnetic processing devices is beginning to emerge. The functionality of these devices varies greatly but falls into the three broad categories of amplification, switching, and communication; some of these will be discussed in this section.

Magnetic amplifiers were developed in the 1940s and were considered to provide highly dependable operation in extreme conditions. The amplification depended on the non-linear characteristics of the saturable magnetic material used in the devices and AC power gain was accomplished by extracting power from DC bias fields. More recently, arrays of magnetic quantum dots have been shown capable to extract power from an applied clocking field in a similar fashion to achieve power gain.

Magnetic amplifiers designed to amplify or restore a magnetic signal have been proposed in Mn:GaAs and Mn:Ge. Both concepts operate by triggering a spontaneous ferromagnetic transition in a bipolar-like heterostructure by modulating the charge density in a dilute magnetic semiconductor (DMS) channel. The goal is to develop an element for future magnetic circuits that could restore magnetic state to overcome losses inherent in any physical system. Fabrication efforts are underway to experimentally demonstrate the operation of this device.

Spin communication structures that rely on propagation of coherent spin waves without electron current flow have been designed, simulated, fabricated, and measured. Signals are excited and propagated in a Fe ferromagnetic film and the propagation length is measured as a function of frequency and applied field. Ferromagnetic spin waves with a frequency of a few GHz were clearly detected.

Magnetic switches based on asymmetric nanorings have been designed and tested. These are based on competition between the exchange energy and the magnetostatic energy in nanomagnets. The relative probability of exciting different magnetic modes and hence the interaction energy in symmetric nanorings is dictated by the ring geometry and cannot be altered after fabrication. Asymmetric nanorings, however, allow tuning the asymmetry electrically, acting as a gate to modulate the interaction.

Self check-pointing architectures have been proposed that use hybrid Hall Effect devices that combine ferromagnetic elements with semiconductor structures to provide high-performance nonvolatile storage that can be tightly integrated with logic. The proposed microprocessor uses the magnetoelectronic devices to “snapshot” the state of the currently executing program at regular intervals, providing protection against power failures and enabling rapid context switching. Novel ferromagnetic devices similar to those mentioned in this section represent important steps towards a viable magnetic technology based on collective magnetic effects in ferromagnetic materials.

**Frequency Coherent Spin Devices**—Since 1996 when the current-driven excitation of magnetic multilayers was predicted, several fundamental properties have been discovered in layered magnetic structures, generating a wealth of exciting physics. In this regard, the most remarkable phenomena are the antiferromagnetic exchange coupling between magnetic layers separated by a nonmagnetic metal layer, the related giant magnetoresistance; the oscillatory behavior of...
Emerging Research Devices

the exchange coupling; and the electron tunneling through ferromagnetic metal layers separated by an insulating film. Several research groups have been working on understanding and utilizing the detailed nonlinear interactions involved in these phenomena to create new classes of nanoresonators and nanooscillators. These include spin valves, magnetic tunnel junctions (MTJs), spin torque oscillators, and tunable magnetic rectifiers.

The fundamental discovery was that a spin-polarized electrical current can apply a large torque to a ferromagnet through direct transfer of spin angular momentum without needing external magnetic fields. Although there is no mechanical motion, a simple magnetic-multilayer structure acts like a nanoscale motor—it converts energy from a DC electrical current into high-frequency magnetic rotations that might be applied in new devices including microwave sources and resonators. The first direct observation of the microwave radiation suggested the presence of both coherent and incoherent microwave sources in the magnetic structures that were subsequently explained by theory and modeling. Recent work experimentally demonstrated spontaneous phase locking of individual resonators that have the potential to generate useful amounts of microwave power (on the order of 1 µW) directly from a single simple nanoscale device. Shaping the nanostructure in order to excite specific magnetic modes can lead to devices optimized to utilize specific current pulse structures or microwave characteristics.

In addition to spin torque-driven oscillators, a great deal of research is devoted to MTJs for use as switch elements in spin torque memories. These rely on the giant magnetoresistive effect for their functionality and use spin torque coupling to manipulate the polarization of the free layer. Magneto resistance changes up to 500% have been observed, but relatively large drive currents remain a problem for full-scale integration of these devices into magnetic memories. However, other spin torque devices that have been proposed and fabricated include tunable RF rectifiers that select one single frequency component. Spin torque coupling in engineered magnetic layer nanostructures excite specific magnetic modes and the spontaneous, coherent, and nonlinear coupling of those modes. These new classes of nanodevices may be anticipated given the richness and complexity of the spin torque coupling in magnetic layer structures.

Emerging Research Architectures

Introduction

Information processor architecture will be subject to change as the level of integration is increased and as reduced feature sizes allow great increases in device performance. The possible introduction of non-conventional active devices further complicates architectural design, for example, by requiring that much more attention be given to designs that provide error correction due to a higher percentage of devices that may not be functional. The mission of the Emerging Research Architectures working-group is to explore device and architectural trends to give visibility to architectural options and attempt to establish criteria for weighing alternative approaches.

In this section, the term “architecture” refers to a functional arrangement on a single chip of interconnected devices that includes embedded computational components. Of particular interest are architectures utilizing, for special purposes, novel devices other than CMOS to perform unique functions. Implicit in this formulation is the assumption that these devices can be integrated with CMOS structures that will continue to perform many of the generic computational and I/O functions required of the implementation.

Trends in CMOS Many-core Architectures

Many of the expected trends in CMOS applications, and hence in CMOS architectures are discussed in the System Drivers chapter of the 2007 ITRS. That discussion is not repeated here; however, it must be said that CMOS provides a fertile ground for new applications that is far from exhausted. For example, the emergence of many-core (symmetric and asymmetric) architectures is an established industry trend. High-end microprocessor architecture is moving to a multi-core format. Dual-core products are currently available commercially and quad-core chips are entering the marketplace. Indeed, an eighty-core experimental chip has been recently announced. Several other companies are now producing multi-core like devices that some are calling “next generation FPGAs.” Specifically, they are implementing field programmable object arrays technology, consisting of arrays of “objects” that are simple processors and other support objects such as memory. Likewise, many ASIC/embedded SoC systems are taking on a multi-core like configuration. These many-core architectures utilize the complexity obtained from scaled CMOS while obtaining more equitable use of on-chip devices and at the same time mitigating heat management and reliability problems. It is estimated that there is headroom for perhaps an order-of-magnitude improvement, relative to single-core processors, in these performance metrics as more elementary processors are added. The difficult problem is to utilize this many-core capability to gain
algorithmic advantage across the general class of computational problems. Finally, CMOS technology offers many stand-alone opportunities for innovation as well including memory systems, power supplies for low-power applications, imaging, identification tags, and many more.

Many-core architectures require an interconnect system that can be either fixed or flexible depending on the intended applications. Flexible connectivity is obtained via implementation of a switching fabric. The performance of the interconnect system drives architectural choices; for example, the globally asynchronous, locally synchronous (GALS) architecture accommodates interconnect delays for clock and data signals by sustaining synchronous operation only in local processors. More generally, various kinds of switching systems and associated controllers have been proposed for the switching fabric and there appear also to be opportunities for emerging research devices in the implementation of the switching fabric.

**HETEROGENEOUS MULTI-CORE ARCHITECTURES**

How does one characterize and use the computational structures made possible by new kinds of emerging devices? It does not appear likely that the current set of emerging research logic devices will provide enhanced general-purpose computation capabilities. One scenario is that some elements of computation that could be done by traditional CMOS could be undertaken by emerging research devices because their implementation is more cost-effective and/or offers higher performance. In many cases, emerging research devices will need to be augmented by traditional CMOS, to allow, among other things, an interface to traditional computation and I/O.

An example of such a hybrid technology is CMOL, where nanogrids of (ideally) single molecules are fabricated on top of traditional CMOS. The grids are rotated so that there is a natural self-alignment and more or less complete coverage of CMOS contacts to nanogrid contacts. CMOS provides external communication as well as current drive and signal restoration. The molecular grids provide ultra-dense programmable interconnect capability that could enhance the basic CMOS circuits.

The molecular cross-bar architecture is another approach to developing a hybrid chip with CMOS/molecular electronics. Although still in their infancy, molecular switches and nanowire interconnect technology can, at least in theory, provide up to $10^x$ improvement in area density and an estimated $10^{11}$ bits/cm$^2$ using conventional programmable logic array (PLA) architectures.

The basic computation topology for the cross-bar architecture is a parameterized arrangement of memory tiles, each consisting of a high-resolution nanowire crossbar array with CMOS row/column multiplexers/de-multiplexers and CMOS combinatorial logic enabling tolerance to defects and faults. To achieve robust PLA architectures in the presence of device defects, a variety of techniques including protective codes with simple matrix implementation on crossbars, defect tolerant decoders, and reconfigurable switches are used.

Despite progress in molecular electronics, a worrisome prospect of the proposed architectures is the high line and junction resistances of nanowires, yielding large RC time constants with maximum estimated operating frequencies of around 1GHz. While nanocrossbar densities may provide some advantages, the speed limitations cannot be overlooked and will require improvement to be competitive with scaled CMOS technologies. In other words, parallelism rather than speed is how performance improvement will likely be achieved with molecular scale computing. Consequently there needs to be a focus on architectures that support significant parallelism and applications that are fundamentally parallel.

Although there are a number of architecture issues that need to be resolved, both high and low end are converging on general multi-processor like structures consisting of independent processors, generally with some local memory, and sophisticated on-chip interconnect. High-end chips use homogeneous cores, but at the lower end, the various modules tend to be heterogeneous with specialized devices added for application-specific computation. It is likely that the high-end chips will also eventually include application or function-specific heterogeneous cores.

The move to heterogeneous, multi-core architectures provides a convenient platform to incrementally incorporate specialized heterogeneous cores based on new technologies, for example, emerging research devices. These cores would execute specialized functions, but at a significantly better cost-performance than special-function CMOS-only cores.

Viewed in this way, the analysis and projection of emerging research device usage is more straightforward and allows the necessary architecture research to proceed concurrently. The ITRS ERD ITWG proposes that within this framework, emerging research devices should be evaluated according to the following set of criteria:
28 Emerging Research Devices

- **Utility**—Does the emerging research device-based processor implement a useful function?
- **Cost/Performance**—What are the cost and performance advantages of the proposed function for an emerging research device implementation? Costs include per device manufacturing costs, on-going power requirements, and performance (including speed) and capacity.
- **“Core-ness”**—Is it configurable as a specialized core in a multi-core NoC (Network on Chip) platform?
- **Access**—Can the emerging research device configuration interface easily with existing CMOS for I/O?

Another important advantage to this approach is that it cleanly merges with another important initiative in computer architecture research, Research Accelerator for Multiple Processors (RAMP). RAMP is a multi-university effort to create a standardized FPGA-based environment for multi-processor computer architecture research.\(^{180}\)

Although there still are a number of complex architectural problems, the biggest problem is the usage model. Some applications map cleanly to parallel machines and can take advantage of automatic or semi-automatic partitioning and mapping software, but there is no automatic way to take general programs and transparently create high-quality parallel versions. Taking advantage of parallelism to move to greater levels of performance will not be as easy as it was to take advantage of increasing clock rates and deeper pipelines, which were mostly transparent to the programmer.\(^{181}\) The functionality of the heterogeneous architectures could be enhanced by the use of some of the emerging research logic devices (see the ERD Logic section). For example, “check-point architectures” could be implemented with ferromagnetic logic devices. Ferromagnetic devices are non-volatile, and therefore their computational state would not be lost by a sudden power failure.

**Morphic Computational Architectures**

The word *morphic* literally means “having a specified form or shape.” It is usually used as a suffix, for example, neuromorphic referring to neuron-like structures in the brain. In the context of this discussion, we use *morphic* to refer to architectures adapted to effectively address a particular problem set, often gaining their inspiration from biological or scientific computational paradigms. It appears that this class of structures is a particularly fertile area for the application of novel devices and even for the extended application of CMOS technologies. This section does not address either sensor or transducer technologies since both are ancillary to the central information processing focus of ITRS. The ERD ITWG acknowledges the central importance of these technologies but recognize that a proper treatment is not within the scope of this chapter.

As a general rule, morphic architectures have the capability to 1) convert data from a physical domain to electrical signals; 2) filter the signals to remove noise components; 3) extract features from the filtered signals; 4) characterize the data based on the extracted features, and 5) store/communicate or act on the characterization of the data. At each of these levels, there are opportunities for utilization of components that are not strictly derived from CMOS technology, but that may benefit from the manufacturing prowess of this technology, e.g., sensors.

The use of primitives in computation that arise from the physics of the computing devices can lead to very efficient signal processing.\(^{182}\) The *basis functions* in analog computing, derived from the physical relationships that result from combining passive and active semiconductor elements, can be optimally matched to the required information processing problem. Examples include filtering, Fourier transform, feature extraction, compression, addition, subtraction, and others. Computation of this form is most efficient for low-resolution analog computing and when the cost of analog processing can be balanced with the costs of A/D conversion, communication, digital post-processing, and power dissipation. This cost balance between analog and digital computing results in a bio-inspired hybrid computer, in that the effects of noise accumulation inherent to analog systems is compensated by periodically restoring the analog values using digital elements of computation.\(^{183}\) Alternatively, there are indications\(^{184}\) that if data is encoded in the time-separation of pulses rather than by analog signal levels, more robust immunity to noise can be achieved.

*Biologically Inspired Hybrid Computation*—Biologically inspired hybrid computation appears to be well suited for tasks requiring processing of visual and auditory information. With more than 130 billion neurons and 1–10 quadrillion synapses, the brain is a densely interconnected information processor with global systems structure and localized computational principles.\(^{185}\) The brain also exhibits three key features: (a) adaptive modification of the processing architecture, (b) spatio-temporal pattern memory and (c) memory prediction. Plasticity or adaptability in the brain is key to embedding learned information via experience by altering structural changes in connectivity. The cortical neurons employ spatio-temporal patterns for neural processing, which do not rely only on intricate neuronal logic gate structures.
Neural processing also utilizes sparseness of connectivity, selectivity of connectivity, and a hierarchy of network connectivity. It has been suggested that local computation employs a probabilistic Bayesian approach.\textsuperscript{186}

One hypothesis is that intelligent systems consist primarily of locally connected, hierarchical, Bayesian modules, which is an example of a general property of locally structured (sparse) systems. The layered hierarchy allows them to efficiently capture higher-order, highly structured probabilities and to perform efficient inference over those probabilities. Furthermore this is general purpose functionality in the sense that it can be applied to a wide array of sensing, control and intelligent computing problems.\textsuperscript{187}

Brain connectivity comes to embed the prior probability structure of the world that a particular system is processing. For example, if it is the visual cortex, then the regularities of spatial autocorrelation, co-linearity, and co-circularity will be embedded in the connectivity of the cortex. The system then uses this information about prior probability, together with current signals, to compute the statistically optimal posterior probability either of the stimulus that caused the signals, or the motor movement that is going to be made. In this way, the system uses its memory of the way the world is to predict what to represent or to do. This is what is meant by memory prediction. For example, recent results in motor neuroscience and visual psychophysics suggest quite strongly that the behavior of the system cannot be accounted for without reference to this sort of memory-based prediction. In other research in brain-like computing, opportunities for nanodevices in associative memory processors (AMP) have been explored.\textsuperscript{188} AMP architectures can be implemented with floating gate field effect transistors (FG-FET) or single-electron transistors. In other bio-inspired implementations, multi-input and multi-output devices could be useful, such as multi-ferroic tunnel junctions (MFTJ) and spin gain transistors (see ERD Logic section).

Cellular Nonlinear Networks (CNNs)—Cellular nonlinear networks are a class of computational structures that, for some applications such as signal processing, may offer performance and power consumption advantages relative to digital information processing systems.\textsuperscript{189} A CNN is typically organized as a connected array of elementary processing elements with problem-dependent layout geometry and connection networks. The processors in a CNN may operate in combined signal domains and are usually characterized by a few continuously adjustable parameters. CNNs may implement an iterative computational mode and they may have sensor arrays co-located with the processors as in image processing/recognition applications. A specific example of CNNs are the cellular sensory wave computational architectures that use a blend of analog and digital technologies to realize local programmable processors that are co-located with sensors in a sensor array. These systems may offer superior power/performance characteristics relative to high performance central computer implementations.\textsuperscript{190} Typically CNNs provide front-end extraction of features for use in back-end knowledge processing.

**POSSIBLE ARCHITECTURAL OPPORTUNITIES FOR EMERGING RESEARCH DEVICES**

In this chapter, several devices have been identified that have been proposed by the research community for use in information processing. How might these devices be most profitably employed in information processing applications on a CMOS platform? This is a difficult question since the operational characteristics of these devices have not yet solidified to the point where detailed assessments can be made. In Table ERD8, an assessment is offered on possible application arenas for these novel devices in the context of special functions where they might offer a performance advantage relative to the CMOS technology where they would be embedded.
Table ERD8    Emerging Research Architectures

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Implementation</th>
<th>Computational Elements</th>
<th>Network</th>
<th>Application</th>
<th>Research Activity [D]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneous Many-Core</td>
<td>Symmetric cores</td>
<td>CMOS</td>
<td>Irregular/Fixed</td>
<td>Synthesis/GPP [A]</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>Asymmetric cores</td>
<td>CMOS</td>
<td>Irregular/Fixed</td>
<td>Synthesis/GPP</td>
<td></td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>CMOL</td>
<td>CMOS+Molecular switches</td>
<td>Irregular/Fixed</td>
<td>Synthesis/GPP</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Molecular cross-bar</td>
<td>Molecular switches</td>
<td>Regular/Flexible</td>
<td>Synthesis/GPP</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Check-point</td>
<td>CMOS+ferromagnetic logic</td>
<td>Irregular/Fixed</td>
<td>Synthesis/GPP</td>
<td>3</td>
</tr>
<tr>
<td>Morphic</td>
<td>CNN</td>
<td>CMOS+sensors</td>
<td>Regular/Flexible</td>
<td>Recognition/Vision [B]</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>AMP</td>
<td>FG-FET, SET</td>
<td>Irregular/Fixed</td>
<td>Recognition/Vision</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Bio-inspired</td>
<td>MFTD, Spin-gain transistor</td>
<td>Mixed</td>
<td>Recognition Mining</td>
<td>35[E]</td>
</tr>
</tbody>
</table>

CMOL—Molecule on CMOS Architecture
CNN—Cellular Nonlinear Network
AMP—Associative Memory Processor
FG-FET—Floating Gate Field Effect Transistor
GPP—general purpose processor
MFTD—Multiferroic Tunnel Diode
SET—Single Electron Transistor

Notes for Table ERD8:
[A] Synthesis—The ability to explore new scenarios by constructing new instances of a model.
[B] Recognition—Machine learning techniques that examine data and construct models for the data.
[C] Mining—The capability to find a model in a large volume of data.
[D] The number of refereed articles in technical journals that appeared in the Science Citation Index Database from July 1, 2005–July 1, 2007.
[E] Not including CNN and AMP.

Table ERD8 indicates the possible application opportunities for CMOS integration of various emerging research devices. It is useful to speculate on the relative performance attributes of the three classes of architectures considered—homogeneous multi-core, heterogeneous multi-core, and morphic. It appears that algorithmic gains for homogeneous digital multi-core systems must await the development of a general-purpose programming environment and that technology gains will be limited by memory management and access issues. The most optimistic projections for performance gains are linear in the number of processors.

Heterogeneous digital architectures offer the capability to include specialized processors and on-chip memory, for example, digital signal processing, linear algebra processor, I/O processors, accelerators, etc., in combination with a general purpose processor (GPP). It is difficult to offer potential performance enhancements for heterogeneous multi-core systems. It is conceivable that orders of magnitude performance gains can be obtained for applications that directly align with embedded special purpose processors.

Morphic architectures embrace a broad class of mixed-signal systems that are focused on a particular application and that draw inspiration for their structure from the application. In some cases, processing is carried out in the analog domain, offering orders of magnitude improvement in performance and power dissipation, albeit with reduced accuracy. As an example, biologically inspired inference networks for cognition may yield to a partial analog implementation and provide substantial gains in performance relative to their digital counterparts.

EMERGING MEMORY AND LOGIC DEVICES—A CRITICAL ASSESSMENT

INTRODUCTION

While the role of nanoscale devices in meeting future computing and communications applications is not clear at this point, undoubtedly there will be many applications that could benefit from the terascale level of integration that such devices might offer. As discussed in the previous sections, these devices will encompass a broad range of fabrication methodologies and functional modalities. As suggested in the Logic and Architectures sections of this chapter, these
Emerging research devices will likely be first integrated with CMOS to supplement the maturing CMOS platform technology extending it to new applications. For example, the terminal characteristics of a new device, governed by its physics, might naturally provide the function of an accelerator block, thereby replacing a complex digital block to realize substantial gains in power dissipation, performance or density.

Conversely, there are significant limitations that arise with nanoscale devices that will impact their usefulness. In particular, their near-term applications will require nanoscale devices to be functionally and technologically compatible with silicon CMOS. In the longer term, electronic charge-based nanoscale devices may be supplemented with one or more new information processing technologies using a new logic “computational state variable” or means of representing a unit of information (a bit). Possibly, in the longer term, a new information processing technology will become the dominant technology, eventually replacing CMOS as the platform.

The purpose of this section is to introduce a set of overall technology requirements and evaluation or relevance criteria and second, based on these criteria, to offer an assessment of the potential of each emerging research technology entry considered in this chapter to perform one of two complementary functions—1) eventually replace CMOS with a highly scalable, high performance, low power binary Boolean logic switch, and 2) provide a memory or storage technology capable of scaling either volatile and/or nonvolatile memory technology beyond 22 nm. This critical review only assesses the long term potential of each emerging research device (memory and logic) technology to replace the ultimately scaled CMOS logic or a memory device technology in the semiconductor industry’s on-going pursuit of increased functional density and performance. This review does NOT judge the potential of any proposed emerging research device technology to supplement or complement CMOS in this or any other fashion.

Assessing the long-range potential of emerging new device and information processing technologies at best is quite difficult, yet providing such an objective and balanced judgment is important. Effective allocation of limited resources requires the semiconductor industry and the research community to consider the long-term potential and advantages offered by a new device technology compared to the projected performance of fully scaled CMOS or of related memory technologies. The intent is to provide such a technically grounded, objective benchmarking for current emerging research device technologies.

Additionally, electronic charge-based approaches will be discussed in this section separately from those approaches proposing use of a new means for “computational state variable” and data representation. This separate discussion addresses an important question related to new charge-based information processing approaches concerning the fundamental limits of an elemental switch (size, energy, speed, etc.).

TECHNOLOGIES BEYOND CMOS

OVERALL TECHNOLOGY REQUIREMENTS AND RELEVANCE CRITERIA

[A] Scalability—First and foremost the major incentive for developing and investing in a new information processing technology is to exploit a new domain for scaling information processing functional density and throughput per Joule substantially beyond that attainable by ultimately-scaled CMOS. Silicon-based CMOS has provided several decades of scaling of MOSFET densities. The goal of a new information processing technology is to replicate this success by providing additional decades of functional and information throughput rate scaling using a new technology. In other words, it should be possible to articulate a Moore’s law for the proposed technology over additional decades.

[B] Performance—Future performance metrics will be very similar to current performance metrics. They are cost, size, and speed. A future information processing technology must continue to provide (at least) incremental improvements in these parameters beyond those attainable by ultimately scaled CMOS technology. In addition, nanodevices that implement both logic and memory in the same device would revolutionize circuit and nanoarchitecture implementations.

[C] Energy Efficiency—Energy efficiency appears likely to be the limiting factor of any beyond CMOS device using electronic charge or electric current as a state variable. It also appears likely that it will be a dominant criterion in determining the ultimate applicability of alternate state variable devices. Clock speed versus density trade-offs for electron transport devices will dictate that for future technology generations, clock speed will need to be decreased for very high densities or conversely, density will need to be decreased for very high clock speeds. Nanoscale electron transport devices will best suit implementations that rely on the efficient use of parallel processing to minimize energy dissipation more than on fast switching.

[D1] OFF/ON or “1/0” Ratio (Memory Devices)—The OFF/ON ratio of a memory device is the ratio of the access resistance of a memory storage element in the OFF state to its access resistance in the ON state. For non-volatile memories, the OFF/ON ratio represents the ratio between leakage current of an unselected memory cell to the read current.
of a selected cell. This definition will apply to new memory technologies so long as they have a selection device. In cross-point memories, a very large OFF/ON ratio is required to minimize power dissipation and maintain adequate read signal margin.

[D2] Gain (Logic Devices)—The gain of nanodevices is an important limitation for current combinatorial logic where gate fan-outs require significant drive current and low voltages make gates more noise sensitive. New logic and low-fan-out memory circuit approaches will be needed to use most of these devices for computing applications. Signal regeneration for large circuits may need to be accomplished by integration with CMOS. In the near-term integratibility of nanodevices with silicon CMOS is a key requirement due to the need for signal restoration for many logic implementations and to be compatible with the established technology and market base. This integration will be necessary at all levels from design tools and circuits to process technology.

[E] Operational Reliability—Operational reliability is the ability of the memory and logic devices to operate reliably within their operational error tolerance given in their performance specifications. The error rate of all nanoscale devices and circuits is a major concern. These errors arise from the difficulty providing highly precise dimensional control needed to fabricate the devices and also from interference from the local environment, such as spurious background charges in SETs. Large-scale error detection and correction schemes will need to be a central theme of any architecture and implementations that use nanoscale devices.

[F] Operational Temperature—Nanodevices must be able to operate close to room temperature environment for most practical applications with sufficient tolerance for higher temperature (e.g., 100°C) operation internal to the device structure.

[G] CMOS Technological Compatibility—The semiconductor industry has been based for the last 40 years on incremental scaling of device dimensions to achieve performance gains. The principle economic benefit of such an approach is it allows the industry to fully apply previous technology investments to future products. Any alternative technology as a goal should utilize the tremendous investment in infrastructure to the highest degree possible.

[H] CMOS Architectural Compatibility—This criterion is motivated by the same set of concerns that motivate the CMOS technological compatibility, namely the ability to utilize the existing CMOS infrastructure. Architectural compatibly is defined in terms of the logic system and the data representation used by the alternative technology. CMOS utilizes Boolean logic and a binary data representation and ideally, the alternative technology would need to do so as well.

Electronic Charge-based Nanoscale Devices

An important issue regarding emerging electronic 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. The 2007 ITRS projects the scaling of CMOS to 11 nm by 2022. This generation represents a physical gate length for a MPU/ASIC device of ~5 nm with an average power dissipation of approximately 100 W/cm². A recent analysis concludes that the fundamental limit of scaling an electronic charge-based switch is only a factor of 3× smaller than the physical gate length of a CMOS MOSFET in 2022. Furthermore the density of these switches is limited by maximum allowable power dissipation of approximately 100 W/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 closely reach the theoretical limits of scaling for charge-based devices. Consequently, application of emerging electronic charge-based logic technologies, such as 1D structures (nanowires and nanotubes) may be best suited for use as a replacement of the silicon channel in an otherwise silicon-based MOSFET technology infrastructure. In other words, use of novel device structures for electronic charge-based switches to develop a completely new information processing technology, including binary switches, memory elements, interconnects (local and global) may not be justified to obtain a relatively modest maximum of 3× scaling in size or speed. This conclusion is particularly true since the device density is limited by power dissipation and not by the size of the binary switch. The corollary of this observation is that the search for alternative logic devices should embrace the concept of using computational state variables other than electronic charge.

Alternate Computational-State-Variable Nanoscale Devices

In this context, the term “computational state variable” refers to the notion of the finite state machine introduced by Turing in 1930s. The idea is that there are numerous ways to represent, manipulate, and store computational information or logic state. The earliest example of a finite state storage device was the abacus, which represents numerical data by the position of beads on a string. In this example, the computational state variable is simply a physical position, and the operator accomplishes readout by looking at the abacus. The operator's fingers physically move the beads to perform the
Emerging Research Devices 33

data manipulations. Early core memories used the orientation of magnetic dipoles to store state. Similarly, paper tapes and
punch cards used the presence or absence of holes to store the state of the computational variable. Several possible new
computational state variables include: magnetic dipole (e.g., electron or nuclear spin state), molecular state, phase state,
strongly correlated electron state, quantum qubit, photon polarization, etc. The question is: can a new computational state
variable together with its physical representation be realized that will scale information processing technology additional
decades in terms of functional density, speed and power similar to that provided by CMOS over the past 40 years? This is
the question addressed in this Critical Assessment.

**POTENTIAL PERFORMANCE ASSESSMENT FOR MEMORY AND LOGIC DEVICES**

The long-term potential performance is critically reviewed for each new memory and logic research device technology
discussed in this chapter eventually to replace an ultimately-scaled CMOS logic or memory device technology to sustain
functional scaling. This analysis does not pertain to proposed emerging research device technologies addressed in the
section entitled Alternative Information Processing Devices.

**METHODOLOGY**

Nanoscale devices in the “Beyond CMOS-scaling” domain span multiple applications, computational state variables, and
technologies. A set of relevance or evaluation criteria, defined above in the section entitled “Overall Technology
Requirements and Relevance Criteria,” are used to parameterize the extent to which a given “Beyond CMOS” technology
is applicable to information processing applications.

Each beyond-CMOS-scaling emerging research nanoscale memory and logic device technology is evaluated against each
Relevance Criterion according to a single factor. For logic, this factor relates to the *projected potential performance* of a
nanoscale device technology, assuming its successful development to maturity, for each Relevance Criterion, *compared to that
for silicon CMOS scaled to the end of the Roadmap at 11 nm*. For memory, this factor relates the *projected potential performance*
of each nanoscale memory device technology, assuming its successful development to maturity, for each Relevance Criterion,
*compared to the existing memory technology the new memory technology would displace*. Performance potential is assigned a
value from 1–3, with “3” substantially exceeding CMOS at 11 nm, and “1” substantially inferior to CMOS or, again, a
comparable existing memory technology. The Relevance Criteria are defined in the section above entitled “Overall Technology
Requirements and Relevance Criteria”. This evaluation is determined by a poll of the ERD Working Group members composed
of individuals representing a broad range of technical backgrounds and expertise.

*Logic—Individual Potential for Emerging Research Logic Devices
Related to each Technology Relevance Criterion*

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Substantially exceeds ultimately scaled CMOS</td>
</tr>
<tr>
<td></td>
<td>* or is compatible with CMOS architecture</td>
</tr>
<tr>
<td></td>
<td>** or is monolithically integrable with CMOS wafer technology</td>
</tr>
<tr>
<td></td>
<td>*** or is compatible with CMOS operating temperature</td>
</tr>
<tr>
<td></td>
<td>(i.e., Substantially Better than Silicon CMOS Logic)</td>
</tr>
<tr>
<td>2</td>
<td>Comparable to ultimately scaled CMOS</td>
</tr>
<tr>
<td></td>
<td>* or can be integrated with CMOS architecture with some difficulty</td>
</tr>
<tr>
<td></td>
<td>** or is functionally integrable (easily) with CMOS wafer technology</td>
</tr>
<tr>
<td></td>
<td>*** or requires a modest cooling technology, T ≥ 77K</td>
</tr>
<tr>
<td></td>
<td>(i.e., Comparable to Silicon CMOS Logic)</td>
</tr>
<tr>
<td>1</td>
<td>Substantially (2×) inferior to ultimately scaled CMOS</td>
</tr>
<tr>
<td></td>
<td>* or can not be integrated with CMOS architecture</td>
</tr>
<tr>
<td></td>
<td>** or is not integrable with CMOS wafer technology</td>
</tr>
<tr>
<td></td>
<td>*** or requires very aggressive cooling technology, T &lt; 77K</td>
</tr>
<tr>
<td></td>
<td>(i.e., Substantially Worse than Silicon CMOS Logic)</td>
</tr>
</tbody>
</table>
Memory—Individual Potential for Emerging Research Memory Devices

Related to each Technology Relevance Criterion

<table>
<thead>
<tr>
<th>Potential</th>
<th>Description</th>
</tr>
</thead>
</table>
| 3 | Substantially exceeds the appropriate Baseline Memory Technology
  * or is compatible with CMOS wafer technology
  ** or is monolithically integrable with CMOS wafer technology
  *** or is compatible with CMOS operating temperature (i.e., Substantially Better than Silicon Baseline Memory Technology) |
| 2 | Comparable to the appropriate Baseline Memory Technology
  * or can be integrated with CMOS architecture with some difficulty
  ** or is functionally integrable (easily) with CMOS wafer technology
  *** or requires a modest cooling technology, T ≥ 77K (i.e., Comparable to Silicon Baseline Memory Technology) |
| 1 | Substantially (2x) inferior to the appropriate Baseline Memory Technology
  * or can not be integrated with CMOS architecture
  ** or is not integrable with CMOS wafer technology
  *** or requires very aggressive cooling technology, T < 77K (i.e., Substantially Worse than Silicon Baseline Memory Technology) |

Overall Potential Assessment (OPA) = Potential Summed over the Eight Relevance Criteria for each Technology Entry

Maximum Overall Potential Assessment (OPA) = 24
Minimum Overall Potential Assessment (OPA) = 8

Overall Potential Assessment for Technology Entries

<table>
<thead>
<tr>
<th>Potential</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential for the Technology Entry is projected to be significantly better than silicon CMOS or baseline memory (compared using the Technology Relevance Criteria) (OPA &gt; 20)</td>
<td></td>
</tr>
<tr>
<td>Potential for the Technology Entry is projected to be slightly better than silicon CMOS or baseline memory (compared using the Technology Relevance Criteria) (OPA = &gt;18–20)</td>
<td></td>
</tr>
<tr>
<td>Potential for the Technology Entry is projected to be slightly less than silicon CMOS or baseline memory (compared using the Technology Relevance Criteria) (OPA = &gt;16–18)</td>
<td></td>
</tr>
<tr>
<td>Potential for the Technology Entry is projected to be significantly less than silicon CMOS or baseline memory (compared using the Technology Relevance Criteria) (OPA ≤ 16)</td>
<td></td>
</tr>
</tbody>
</table>

RESULTS

Tables ERD9 and ERD10 summarize the results of the critical review. Again, the purpose is to evaluate the potential of several emerging research memory and logic technologies discussed in this chapter eventually to replace fully scaled, mature CMOS enabling additional decades scaling of information processing technology. The color scale is given in the table above entitled “Overall Potential Assessment for Technology Entries.” The color represents the overall assessment for each emerging research memory and logic technology. White indicates the ERD Working Group’s judgment of a relatively high potential for a fully matured technology to excel compared to CMOS for logic or compared to the current memory technology to be replaced. Conversely, red indicates a relatively low potential. Green and yellow provide additional granularity from a moderately high potential (green) to a lower assessment for potential (yellow). On a scale of 1–3, the numbers given in each box are the average of the responses for that technology/relevance criterion received from members of the ERD Working Group. The error bars indicate the average response ± the standard deviation. Assignment of the relative ratings for each Technology Entry for memory and for logic is the collective judgment of the ERD Working Group and is intended to be somewhat prescriptive and not prescriptive. These ratings taken together with the numerical tables and descriptive text are intended to provide the reader with the ERD Working Group’s perspective on each Technology Entry following two years of conducting several workshops, reviewing the literature, and engaging in lively discussions within the Working Group. This evaluation is illustrated in further detail for each Memory Technology in Figures ERD2 through ERD7 and for each Logic Technology in Figures ERD8 through ERD15.
Two new technology candidates for memory applications are identified as promising. These are the engineered tunnel barrier memory and the fuse/antifuse memory. For logic, technology entries related to the long-term scaling of CMOS—specifically 1D structures—are thought to be promising. Channel replacement materials also have potential, albeit with higher risk. Conversely, no candidate technology for “beyond CMOS” logic applications is viewed to be very attractive. Therefore research is needed to identify new applications for these technologies, for example possible ways they might complement and extend the functionality of the CMOS platform technology.

Table ERD9 Potential Evaluation for Emerging Research Memory Devices

<table>
<thead>
<tr>
<th>Technology</th>
<th>Scalability</th>
<th>Performance</th>
<th>Energy Efficiency</th>
<th>Off/On ratio</th>
<th>Operational Reliability</th>
<th>Operational Temperature***</th>
<th>CMOS Technological Compatibility**</th>
<th>CMOS Architectural Compatibility*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineered Tunnel Barrier Memory</td>
<td>2.4</td>
<td>2.3</td>
<td>2.2</td>
<td>2.0</td>
<td>2.2</td>
<td>2.7</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Fuse/Anti-fuse Memory</td>
<td>2.6</td>
<td>1.9</td>
<td>2.0</td>
<td>2.2</td>
<td>1.8</td>
<td>2.8</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Nano Mechanical Memory</td>
<td>1.7</td>
<td>1.9</td>
<td>2.4</td>
<td>2.5</td>
<td>1.9</td>
<td>2.9</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Electronic Effects Memory</td>
<td>2.3</td>
<td>2.2</td>
<td>2.3</td>
<td>2.1</td>
<td>2.0</td>
<td>2.4</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Ionic Memory</td>
<td>2.6</td>
<td>2.0</td>
<td>2.4</td>
<td>2.1</td>
<td>1.7</td>
<td>2.5</td>
<td>2.1</td>
<td>2.5</td>
</tr>
<tr>
<td>Ferroelectric FET Memory</td>
<td>1.8</td>
<td>2.0</td>
<td>1.9</td>
<td>2.1</td>
<td>1.7</td>
<td>2.6</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Macromolecular Memory</td>
<td>2.1</td>
<td>1.8</td>
<td>2.1</td>
<td>1.8</td>
<td>1.4</td>
<td>2.2</td>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Molecular Memory</td>
<td>2.4</td>
<td>1.7</td>
<td>2.4</td>
<td>1.4</td>
<td>1.3</td>
<td>2.2</td>
<td>1.8</td>
<td>1.9</td>
</tr>
</tbody>
</table>
## Table ERD10  Potential Evaluation for Emerging Research Logic Devices

<table>
<thead>
<tr>
<th></th>
<th>Scalability</th>
<th>Performance</th>
<th>Energy Efficiency</th>
<th>Gain</th>
<th>Operational Reliability</th>
<th>Operational Temperature***</th>
<th>CMOS Technological Compatibility**</th>
<th>CMOS Architectural Compatibility*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1D Structures</strong></td>
<td>2.4</td>
<td>2.2</td>
<td>2.5</td>
<td>2.3</td>
<td>2.0</td>
<td>2.5</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Channel Replacement</strong></td>
<td>2.0</td>
<td>2.9</td>
<td>2.3</td>
<td>2.4</td>
<td>1.9</td>
<td>2.3</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Materials</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Single Electron Transistors</strong></td>
<td>2.4</td>
<td>1.1</td>
<td>2.3</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Molecular Devices</strong></td>
<td>2.5</td>
<td>1.5</td>
<td>2.2</td>
<td>1.5</td>
<td>1.3</td>
<td>1.8</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ferromagnetic Devices</strong></td>
<td>1.2</td>
<td>1.3</td>
<td>1.7</td>
<td>1.5</td>
<td>2.0</td>
<td>2.1</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spin Transistors</strong></td>
<td>1.7</td>
<td>1.4</td>
<td>2.3</td>
<td>1.7</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Engineered Tunnel Barrier Memory

Figure ERD2a  Technology Performance Evaluation for Engineered Tunnel Barrier Memory

Fuse/Anti-fuse Memory

Figure ERD2b  Technology Performance Evaluation for Fuse/Antifuse Memory
Figure ERD2c  Technology Performance Evaluation for Nano Mechanical Memory

Figure ERD2d  Technology Performance Evaluation for Electronic Effects Memory
Ionic Memory

Figure ERD2e  Technology Performance Evaluation for Ionic Memory

Ferroelectric FET Memory

Figure ERD2f  Technology Performance Evaluation for Ferroelectric FET Memory
40 Emerging Research Devices

THE INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS: 2007
Emerging Research Devices

1D Structures

Figure ERD3a Technology Performance Evaluation for 1D Structures (CNTs and NWs) Logic Devices

Channel Replacement Materials

Figure ERD3b Technology Performance Evaluation for Channel Replacement Materials Logic Devices
Emerging Research Devices

Figure ERD3c Technology Performance Evaluation for Single-Electron Transistors Logic Devices

Molecular Devices

Figure ERD3d Technology Performance Evaluation for Molecular Logic Devices
Emerging Research Devices

Ferromagnetic Devices

Figure ERD3f  Technology Performance Evaluation for Ferromagnetic Logic Devices

Spin Transistors

Figure ERD3e  Technology Performance Evaluation for Spin Transistors Logic Devices
FUNDAMENTAL GUIDING PRINCIPLES—
“BEYOND CMOS” INFORMATION PROCESSING

INTRODUCTION

In considering the many disparate new approaches proposed to provide order of magnitude scaling of information processing beyond that attainable with ultimately scaled CMOS, the Emerging Research Devices Working Group proposes the following comprehensive set of guiding principles. We believe these “Guiding Principles” are necessary for a new “Beyond CMOS” information processing technology to dramatically enhance scaling of functional density and performance while simultaneously reducing the energy dissipated per functional operation. Further this new technology would be realized using a highly manufacturable fabrication process.

GUIDING PRINCIPLES

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 of alternative state variable devices to ultimately scaled CMOS should be made as early in the program as possible to down-select and identify key trade-offs.

NON-THERMAL EQUILIBRIUM SYSTEMS

These are non-thermal equilibrium systems that serve to reduce the perturbations of stored information energy in the system caused by thermal interactions with the environment. This function can be accomplished by systems that perform all computational processing functions in a time short compared to the system’s energy relaxation time. Thermal fluctuations will require energy barriers of order $10kT$ to prevent random fluctuations of computational state in any bistable-switching device where $k_B$ is Boltzmann’s constant and $T$ is the effective temperature. One path to low energy, room temperature switching is to find systems that can operate out of thermal equilibrium with the phonon bath so the effective temperature $T$ for the system is less than the general environment. Nuclear spin is a naturally occurring example of such a system.

NOVEL ENERGY TRANSFER INTERACTIONS

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

NANOSCALE THERMAL MANAGEMENT

This might be accomplished by manipulating lattice phonons for constructive energy transport and heat removal. These would include phonon stop band structures for local energy redistribution and structures for non-isotropic heat transport

SUB-LITHOGRAPHIC MANUFACTURING PROCESS

One example of this principle is directed self-assembly of complex structures composed of nanoscale building blocks This requirement is essential to fabricate blocks including quantum dots, semiconductor nanocrystals, metallic nanocrystals, and resonant cavities (metacrystals) in a bulk material capable of supporting the quantum interactions described above (e.g., complex metal oxides). 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.

EMERGING ARCHITECTURES

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

MEMORY

46 Emerging Research Devices


THE INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS: 2007


LOGIC


Koichi Maezawa. “Resonant Tunneling Diodes and Their Application to High-Speed Circuits.” CSIC 2005 Digest, IEEE.

Emerging Research Devices


155 Joseph Carmonan et. al . “From molecular interactions to gates: a systematic approach.” ICCAD ‘06, November 5-9, 2006, San Jose, CA.


166 S. M. Rezende, F. M. de Aguiar, M.A. Lucena, and A. Azevedo. “Magnon Excitation by Spin Injection in Thin Fe/Cr/Fe Films.” PRL, Volume 84, Number 18, May 1, 2000.


ARCHITECTURES

175 R. Bashirulla, et al. “Perspectives on Nano-Architectures.” SRC working paper (2006). This analysis concluded that an on-chip architecture consisting of many of light-weight processors instead of one heavy-weight processor have reduced energy error constraints, and therefore can operate at lower voltage without sacrifices in reliability.