

Science and Engineering Beyond Moore's Law

This paper describes Moore's law for CMOS technology, examines its limits, and considers some of the possible future pathways for both CMOS and successor technologies with objective of encouraging some radical rethinking for the development of possible future information processing technologies.

By RALPH K. CAVIN, III, Life Fellow IEEE, PAOLO LUGLI, Fellow IEEE, AND VICTOR V. ZHIRNOV

ABSTRACT | In this paper, the historical effects and benefits of Moore's law for semiconductor technologies are reviewed, and it is offered that the rapid learning curve obtained to the benefit of society by feature size scaling might be continued in several different ways. The problem is that as features approach the range of a few nanometers, electron-based devices depart radically from the ideal switch and, in fact, become very leaky in the off state. It is argued that there are some short-term solutions involving more highly parallel manufacturing, increased design efficiency, and lower cost packaging technologies that could continue the steep learning curve for cost reductions that have historically been achieved via Moore's Law scaling. Another alternative might be to increase chip functionality by integrating devices that offer broadened chip functionality including, e.g., sensors, energy sources, oscillators, etc. A third alternative would be to invent an entirely new information processing state variable based on different physics, using electron spin, magnetic dipoles, photons, etc., to improve the performance and reduce switching energy for devices whose smallest features are on the order of a few nanometers. Each of these alternatives is being actively explored and an overview of each strategy and progress to date is given in the paper. A final alternative offered in the paper is to learn from information processing examples in nature, specifically in living systems. An *E.coli* cell of about one cubic micrometer volume is shown to be an incredibly powerful and

energy-efficient information processor relative to the performance of an end-of-scaling silicon processor of the same volume. The paper concludes by pointing out some of the crucial differences between *E.coli* information processing and conventional approaches with the hope technologies can be invented using the hints offered by biosystems.

KEYWORDS | Beyond CMOS; biological computation; logic; memory; more than Moore; scaling limits

I. INTRODUCTION

One of the remarkable technological achievements of the last 50 years is the integrated circuit. Built upon previous decades of research in materials and solid state electronics, and beginning with the pioneering work of Jack Kilby and Robert Noyce, the capabilities of integrated circuits have grown at an exponential rate. Codified as Moore's law, integrated circuit technology has had and continues to have a transformative impact on society. This paper endeavors to describe Moore's law for complementary metal-oxide-semiconductor (CMOS) technology, examine its limits, consider some of the alternative future pathways for CMOS, and discuss some of the recent proposals for successor CMOS technologies. In the spirit of the editorial guidance for this issue, an analysis of the living cell as an information processor is offered and estimates of its performance are given. For comparison, an equal volume CMOS cell is postulated, equipped with extremely scaled technologies, and performance estimates are generated. Indications are that the living cell is architected and operates in such a way that it is extraordinarily energy efficient relative to the performance of the comparison CMOS cell. This analysis is offered with the hope that it will encourage radical rethinking of possible future information processing technologies.

Manuscript received January 25, 2012; accepted February 15, 2012. Date of publication April 18, 2012; date of current version May 10, 2012.

R. K. Cavin, III and V. V. Zhirnov are with Semiconductor Research Corporation, Research Triangle Park, NC 27709 USA (e-mail: Ralph.Cavin@src.org; Victor.Zhirnov@src.org).

P. Lugli is with Lehrstuhl für Nanoelektronik, Technische Universität München, D-80333 Munich, Germany (e-mail: lugli@nano.ei.tum.de).

Digital Object Identifier: 10.1109/JPROC.2012.2190155

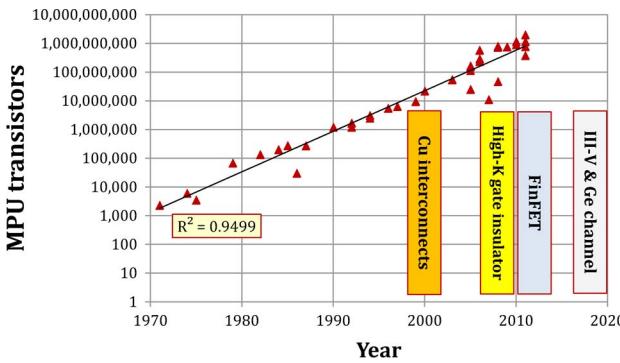


Fig. 1. The number of transistors per microprocessor chip versus time, showing introduction of new enabling technologies.

II. BENEFITS OF SCALING: MOORE'S LAW FOR SEMICONDUCTORS

In 1965, Gordon Moore [1] observed that the number of transistors on a chip could be expected to double annually for at least ten years. At different time points in the ensuing decades, it has appeared that doubling time has varied from 18 months to three years. Overall, however, the chip transistor count has continued to increase, and conversely, the size of each transistor has decreased, at an amazing rate, and Gordon Moore's postulate became known as *Moore's law*. Fig. 1 is a plot of transistor count for a variety of microprocessor chips (listed in Table 1) versus time. Moore was indeed prescient for the transistor count, on average, has doubled approximately every two years.

Table 1 Microprocessor Data Used to Create Fig. 1

Company	Model	Year
Intel	4004	1971
Intel	8080	1974
MOS Technology	6502	1975
Motorola 68000	68000	1979
Intel	286	1982
Motorola	68020	1984
Intel	386DX	1985
ARM	ARM2	1986
Motorola	68030	1987
Motorola	68040	1990
DEC	Alpha 21064 EV4	1992
Intel	486DX	1992
Motorola	68060	1994
Intel	Pentium	1994
Intel	Pentium Pro	1996
IBM - Motorola	PowerPC 750	1997
Intel	Pentium III	1999
AMD	Athlon	2000
AMD	Athlon XP 2500+	2003
Intel	Pentium 4 Extreme Edition	2003
Centaur Technol - VIA Technol	VIA C7	2005
AMD	Athlon FX-57	2005
AMD	Athlon 64 3800+ X2 (Dual core)	2005
IBM	Xbox360 "Xenon" (Triple core)	2005
Sony-Toshiba-IBM	PS3 Cell BE	2006
AMD	Athlon FX-60 (Dual core)	2006
Intel	Core 2 Extreme X6800 (Dual core)	2006
Intel	Core 2 Extreme QX6700 (Quad core)	2006
P.A. Semi	PA6T-1682M	2007
Intel	Core 2 Extreme QX9770 (Quad core)	2008
Intel	Core i7 920 (Quad core)	2008
Intel	Atom N270 (Single core)	2008
AMD	E-350 (Dual core)	2011
AMD	Phenom II X4 940 Black Edition	2009
AMD	Phenom II X6 1100T	2010
Intel	Core i7 Extreme Edition 980X (Hex core)	2010
Intel	Core i7 2600K	2011
Intel	Core i7 875K	2011
AMD	8150 (Eight core)	2011

As feature sizes have decreased, the real density of transistors has correspondingly increased supporting either more functionality for a given chip size or the reduction in chip size to obtain a given level of functionality. The latter benefit enabled the fabrication of more chips per wafer and thus continued cost reduction trends. Cost reductions have also resulted from increased wafer sizes, again allowing the production of more chips/wafer. Also the individual transistor switching time and energy decrease with feature size scaling.

Minimum feature sizes circa 1980 were on the order of 3 μm , while today they are about 32 nm (a 100-fold decrease), giving four orders of magnitude increase in device density. Supply voltages in this time frame have decreased from five to one volt in an effort to reduce power consumption.

It has been the history of the semiconductor industry that, as obstacles are encountered, scientific and engineering solutions are developed to continue the cadence more or less as indicated by Moore's law. In the 1990s, it became evident that scaling was encountering a number of barriers including increasing interconnect power consumption, transistors that were consuming increased power in their OFF state, etc. This led to the search for new material systems and associated processes to sustain the growth in transistor counts that was providing increasing performance and functionality for the electronics and other industries. An example of technology innovation was the introduction of copper interconnects to replace aluminum-based interconnects on chip [2], [3]. Initially, this was viewed as a difficult task since copper diffuses in silicon (Si) and can be detrimental to metal-oxide-semiconductor field-effect transistor (MOSFET) performance. However, barrier systems were developed so that the higher conductivity of copper could be exploited. As another example, the decrease in the gate oxide thickness to a few nanometers was leading to increased gate leakage currents and higher OFF state power consumption. Research led to the incorporation of new gate materials with a higher dielectric constant (e.g., hafnium oxide) so that drive capacitance could be maintained and tunneling reduced [4], [5]. Due to the incompatibility of the high- k dielectric with the traditional polysilicon gate, a new metal gate technology was introduced [5]. In order to increase channel mobility, new strained Si channel technologies were developed [6]. Looking ahead, the pace of innovation continues with, for example, research to determine if higher channel mobility might be achieved by introducing compound III-V channel materials into the Si MOSFET [7]. Even the structure of the MOSFET is under renewed consideration, e.g., the different variations of the multiple-gate FET devices [8], e.g., Trigate, are now being introduced into production. These innovations have continued to provide increases in integrated circuit performance [9].

One indicator of the ultimate performance of an information processor, realized as an interconnected system of

binary switches, is the maximum binary throughput (BIT), that is, the maximum number of on-chip binary transitions per unit time. It is the product of the number of devices M with the clock frequency of the microprocessor f

$$\beta = Mf. \quad (1)$$

(Note that β is an aggregate indicator of technology capability.)

The computational performance of microprocessors μ is often measured in (millions) of instructions per second (IPSs) that can be executed against a standard set of benchmarks. There is a strong correlation between system capability for IPS (μ) and the binary throughput β , as shown in Fig. 5, and to a good approximation

$$\mu = f(\beta) = k\beta^p. \quad (2)$$

For the selected class of microprocessors, $k \sim 0.1$ and $p \sim 0.64$ with a high degree of accuracy (the determination coefficient $R^2 = 0.98$). This strong correlation suggests a possible fundamental law behind the empirical observation.

Fig. 2 also shows an estimated capability of the human brain in μ - β metrics. While it is difficult to quantify the brain operations, there have been several attempts to estimate computational performance of the brain. In [10], an estimate of equivalent binary transitions was made from the analysis of the control function of brain: the equivalent number of binary transitions to support language, deliberate movements, information-controlled functions of the organs, hormone system, etc., resulting in an "effective" binary throughput of the brain $\beta \sim 10^{19}$ b/s. An

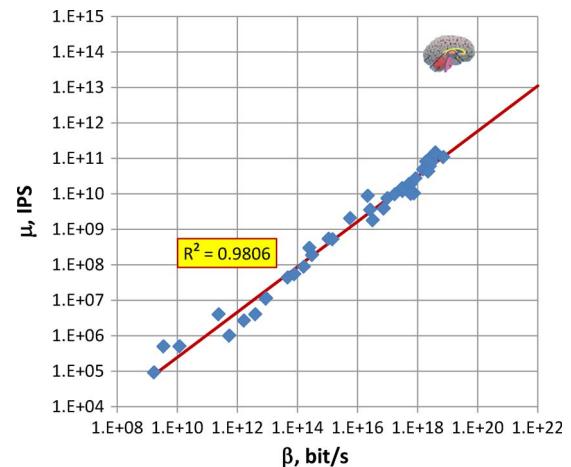


Fig. 2. Benchmark capability μ (instructions per second) as a function of β (bits per second).

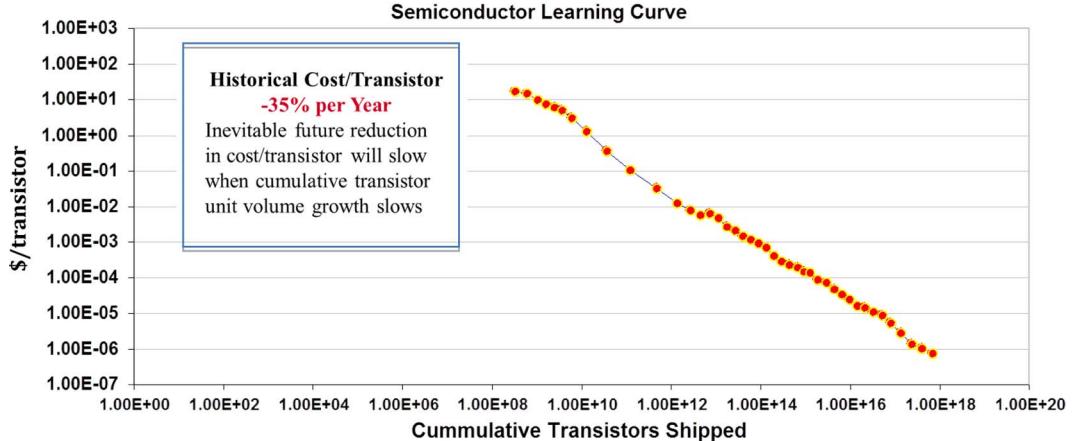


Fig. 3. Transistor cost as a function of the cumulative number of transistors shipped [12].

estimate of the number of equivalent IPSs was made in [11] from the analysis of brain image processing capability resulting in $\mu \sim 10^{14}$ IPSs. It is clear that the brain is not on the microprocessor trajectory in Fig. 2, giving rise to the hope that there may exist alternate technologies and computing architectures offering higher performance (at much lower levels of energy consumption). In the following sections, some of such technologies, complementary and/or alternative to CMOS circuitry, will be discussed in detail.

Without customers for the increased volume of transistors and chips, there would be little incentive to continue to drive scaling at an exponential pace. The \$300 billion dollar semiconductor industry (circa 2011) provides essential components for the much larger electronics devices and systems industries as well as automotive, entertainment, medical device, and other technology-based industries. One effect of the rapidly increasing capability of integrated circuit systems is to support rapid growth in functionality and this has translated into a *feature-driven* market. That is, electronics customers, as a rule, replace their older electronic systems well before they are no longer useful because the newer devices offer a compelling increase in capability. This contrasts with many other industries where purchases are driven by the need for replacement at the end of the useful life of the product. The net effect is that electronic products have been part of a closed cycle where increasing markets have supported the capability of industry to invest in the continuous reduction of semiconductor costs via the introduction of new technologies.

Moore's law is actually connected to a more fundamental premise, known as the learning curve, which relates decreases in the price of product to corresponding increases in the volume of production of that product. The learning curve for semiconductors has shown much less volatility over the years than Moore's law. The learning curve states that the cost per unit decreases by a fixed

percent every time total cumulative volume doubles. This is applicable across a wide range of products but the striking difference for semiconductors is that the rate of decrease in cost per unit for semiconductors has occurred at a much higher rate than for many other industries. Figs. 3 and 4 (courtesy of W. Rhynes of Mentor Graphics [12]) illustrate the learning curve for transistors and for microprocessors.

Note from Fig. 3 that the per-transistor cost decreases by approximately a factor of two for every doubling of transistor production. To put this in a temporal perspective, the average compound annual rate of cost reduction for transistors is on the order of 35% per year. In Fig. 3, the cost per millions of instructions per second (MIPS) for personal computers has shown an even more dramatic rate of cost decrease—a reduction on the order of a factor of nine in cost-per-MIP every two years.

It is interesting to contemplate whether integrated circuits will be able to sustain these steep learning curves indefinitely. There exist formidable technical and economic challenges to doing so and some of these are considered in this paper. A good example is the search for new lithographic tools that can provide efficient and cost-effective patterning for features less than 10 nm in size. There is a focus today on extreme ultraviolet lithography to replace optical lithography but this technology is not yet production ready. At the same time, research in directed self-assembly [13], [14] continues to make good progress and offers the hope for some relief for optical methods. Moreover, there are clear and fundamental limits for the scaling of electron-based devices [15]. In spite of directed research programs that seek to provide alternatives to the MOSFET switch for logic applications where other nonelectron representations for information might be used, such as the Nanoelectronics Research Initiative (www.src.org/program/nri), no compelling replacement options have yet been identified [16], [17]. It appears that

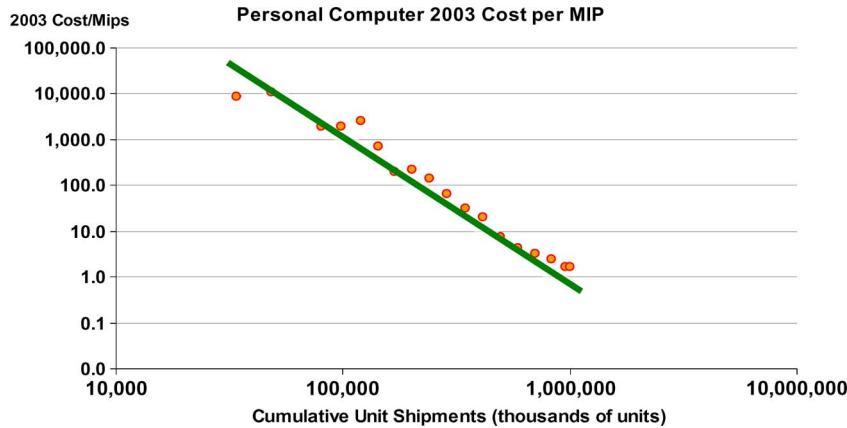


Fig. 4. Personal computer cost (inflation adjusted) per millions of instructions per second versus cumulative units shipped [12].

for many of the proposed alternative devices, their unique properties might be utilized to advantage to achieve a special function when integrated with CMOS technologies. On the other hand, continuing research in alternative memory technologies has resulted in the identification of potential replacements offering the potential for smaller sizes that could meet information processing performance specifications. Indeed, progress in memory technology may foretell changes in memory architectures for information processing and could support an increased focus on data-centric as opposed to logic-centric processing. Nevertheless, what will happen when we reach scaling limits for CMOS-like technologies? Can the learning curve shown above be continued at the same rates as have been sustained to date?

Continuation of the semiconductor learning curve beyond the end of scaling rests on several factors. 1) It is essential to sustain an ever-broadening applications space for integrated circuits since this provides the revenue base for advances in semiconductor technology. 2) As scaling of features becomes more difficult, it will be necessary for advances in design, architectures, 3-D packaging, etc., to play an increased role in cost reduction. 3) *Parallel fabrication* to decrease manufacturing costs per unit transistor needs to be emphasized, e.g., by increasing wafer size [18]. In this paper, possibilities for continuing the semiconductor *virtuous cycle* are explored. The perspective is that the future for semiconductor technologies is very bright, even as we face scaling limits, primarily because the opportunities for integration of new functionalities with CMOS is at an early point and the possibilities for expanding applications incorporating new on-chip physical domains (e.g., mechanical, thermal, chemical, optical) of operation is just beginning. Examples include the integration of sensors that respond to a wide range of stimuli, new architectures that can reason from data leading to integrated systems that can assess and respond, inclusion of devices operating in new physical domains to increase the

energy efficiency and performance of information processing systems, the introduction of 3-D packaging technologies, etc. All of these opportunities will require advances in science and engineering, but this is nature of the semiconductor enterprise.

III. “MORE MOORE”: EXTREMELY SCALED CMOS LOGIC AND MEMORY DEVICES

In order to assess the performance characteristics for extremely scaled transistors and memory cells, it is instructive to consider the generic structure and the physical layout of a transistor (binary switch) and a nonvolatile memory cell. The approach taken is to utilize simple physical and geometrical models to make evident the essential performance characteristics of FETs at the limits of scaling.

A. Electronic Switch (FET)

An energy barrier is used to control electron transport in FETs. The barrier can be formed, e.g., by doping that creates built-in charges in the barrier (channel) region, as shown in Fig. 5(a). The height and the width of the barrier determine essential operational characteristics of transistors such as device size, switching speed, operating voltage, OFF leakage current, etc. In order to control the barrier height, a gate electrode is coupled to the barrier region, separated by a thin layer of gate insulator, e.g., SiO_2 or HfO_2 . When a voltage is applied to the gate, an electric field is created between the gate and the barrier region. This electric field changes the barrier height, thus allowing electrons to pass through the channel. The width of the barrier is defined by device fabrication, and is represented by geometrical characteristics such as channel length L_{ch} or gate length L_g . In the following, it will be assumed $L_{ch} \approx L_g \approx F$, where F is the critical feature size. In order to retain gate control with scaling, it is necessary

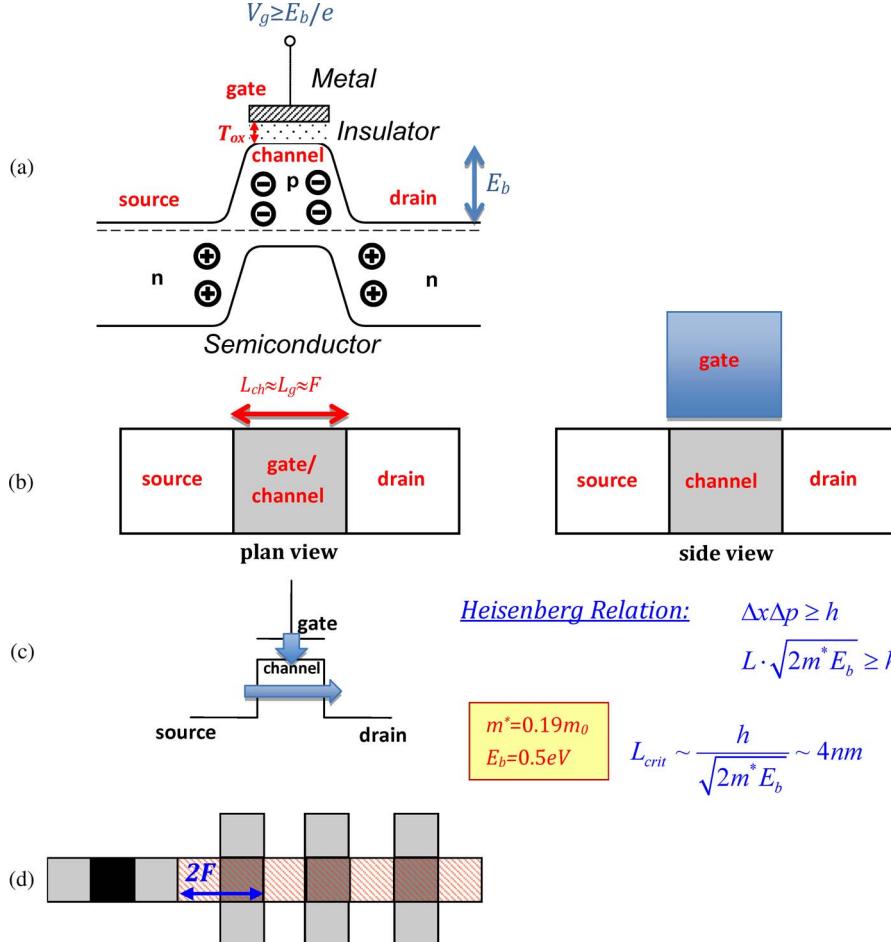


Fig. 5. Semiconductor FET: (a) materials system; (b) generic floorplan; (c) scaling limits; and (d) connected binary switches.

to decrease the gate insulator thickness T_{ox} proportionally to the decrease of the channel length. For an optimized FET structure, a rule of thumb suggests $T_{ox}/L_{ch} \sim 1/30$ [19]. The device platform for modern microelectronics is known as MOSFET.

The barrier representation of a binary switch (e.g., FET) shown in Fig. 5(a) also suggests a generic topology for the ultimately scaled device [Fig. 5(b)]. The 2-D floor plan of a smallest possible binary switch is a $3F \times F$ rectangle consisting of three square “tiles” of the same size F (representing the source, channel, and drain regions of the MOSFET). Further, it can be assumed that the associated insulator and metal layout elements are also composed of tiles of minimum size F . Finally, the metal interconnects, which connect individual devices in more complex logic circuits, can also be represented as a combination of the square tiles, as shown in Fig. 5(b). It is straightforward to show from both topology and physics considerations that in the limiting case, it is useful to consider the size of the interconnect tile as equal to the device tile F . The tiling framework is a useful tool

for circuit/system physical-level explorations of different scenarios of extreme scaling.¹ A detailed treatment of the tiling framework can be found in [20], and two important examples will be considered in Section III-C.

One fundamental issue that limits physical scaling of MOSFETs, and therefore the minimum tile size F , is quantum mechanical tunneling, which dramatically increases the OFF leakage current, as depicted in Fig. 5(c). A simple estimate of the tunneling limit can be made using the Heisenberg relation [see numerical insert in Fig. 5(c)]. For typical parameters for Si FET, this estimate results in a minimum channel length of ~ 4 nm. More detailed calculations yield similar findings—it is argued in a number of studies that tunneling OFF-state leakage becomes overwhelming for $L_{ch} = 4\text{--}7$ nm, and this is sometimes

¹A practical device is somewhat larger than the ideal shown in Fig. 5(b); e.g., wraparound gates, larger area of source and drain to minimize contact resistance, increased gate width to increase ON current, etc. However, the idealized representation shown in Fig. 5(b) will be used in this paper to cast MOSFET technology most favorably for packing density.

Table 2 ITRS Performance Projections for Extremely Scaled High-Performance FETs (2007 ITRS Edition [22])

L_g , nm	EOT, nm	J_g , A/cm ²	I_{sd}/W , $\mu\text{A}/\mu\text{m}$	$I_{leak}=I_g+I_{sd}$, nA	I_{on} , μA	V_{dd}	E_{sw}
10	0.60	2000	0.38	5.80	23.0	0.8	$3.36 \cdot 10^{-18}$
9	0.55	2220	0.44	5.76	23.6	0.8	$2.92 \cdot 10^{-18}$
8	0.55	2500	0.48	5.44	20.3	0.7	$1.89 \cdot 10^{-18}$
7	0.55	2860	0.45	4.55	19.6	0.7	$1.57 \cdot 10^{-18}$
6	0.50	3330	0.47	4.02	16.6	0.7	$1.21 \cdot 10^{-18}$
5.5	0.50	3640	0.43	3.47	14.7	0.65	$9.09 \cdot 10^{-19}$
5	0.50	4000	0.62	4.10	14.0	0.65	$7.65 \cdot 10^{-19}$
4.5	0.50	4440	0.60	3.60	12.5	0.65	$6.50 \cdot 10^{-19}$

cited as the “ultimate” FET [21]. These assessments are consistent with the International Technology Roadmap for Semiconductors (ITRS) [22], which projected the minimal physical gate length in high-performance logic FET to be in the range from 4.5 nm (2007 ITRS) to 5.9 nm (2011 ITRS). Note that the result for the smallest channel length in Fig. 5(c) depends on the mass of the information-bearing particles, e.g., the effective mass of electrons in Si. Heavier particle mass could, in principle, allow for further scaling.

One approach to deal with the severe leakage in a scaled transistor is to develop families of FETs, optimized for specific applications. For example, if highest switching speed is the goal, the smallest channel length and therefore thinnest gate dielectrics are required. As a result, the leakage can be relatively high, which still could be tolerated in some applications. On the other hand, in other applications, such as, e.g., mobile devices, standby power minimization is mandatory. This can be achieved by increasing L_g and T_{ox} , and thus giving up some performance and device density. A set of parameters projected for extremely scaled transistors developed by ITRS is shown in Tables 2 and 3 for high-performance and low standby power transistors, respectively.

B. Nonvolatile Electronic Memory

In memory cells that store electron charge, such as flash, dynamic random-access memory (DRAM), or static random-access memory (SRAM), two distinguishable states 0 and 1 are created by the presence (e.g., state 0) or absence (e.g., state 1) of electrons in a specific location (the charge storage node). In order to prevent losses of the stored charge, the storage node is defined by energy

barriers of sufficient height E_b to retain charge (as shown in Fig. 6). The properties of the barrier, i.e., barrier height E_b and width a , determine the retention time of a memory cell.

In order to obtain a nonvolatile memory cell, sufficiently high barriers must be created to retain the charge for a long period of time. As it was argued in [23], for > 10 y retention, the barrier height E_b must be more than ~ 1.7 eV. High barriers are formed by using layers of insulator (I), which surround a metallic storage node (M). Such an I–M–I structure forms the storage node in the floating gate cell, the basic element of flash memory. The barrier height E_b is a material-specific property [see a table insert in Fig. 6(a)]. As shown in Fig. 6(a), the stored electrons can “leak” from the storage node either over the barrier (if the barrier is not sufficiently high), resulting in leakage current I_{o-b} , or by tunneling through the barrier (if the barrier is not sufficiently wide), thus resulting in leakage current I_T . For long retention (e.g., > 10 y) the theoretical barrier width must be > 5 nm for all known dielectric materials (typically > 7 nm in practical devices). The corresponding practical minimum size of the floating gate cell is ~ 10 nm [23].

The requirement for a large barrier insulator height and thickness also results in a fundamentally high operating voltage, both for WRITE and READ. For example, during the WRITE operation, electrons are injected into the storage node, and this requires operation in the Fowler–Nordheim (F–N) tunneling regime for faster injection. The condition for F–N tunneling is $eV_b > E_b$, i.e., the potential difference across the barrier between the storage node and the external contact must be larger than the barrier height. Since

Table 3 ITRS Performance Projections for Extremely Scaled Low Standby Power FETs (2007 ITRS Edition [22])

L_g , nm	EOT, nm	J_g , A/cm ²	I_{sd} , pA/ μm	$I_{leak}=I_g+I_{sd}$, pA	I_{on} , μA	V_{dd}	E_{sw}
11	0.90	0.27	26.0	0.61	9.85	0.75	$3.22 \cdot 10^{-18}$
10	0.90	0.30	23.9	0.54	9.35	0.75	$2.62 \cdot 10^{-18}$
9	0.80	0.33	33.8	0.57	8.41	0.7	$1.98 \cdot 10^{-18}$
8	0.80	0.38	28.9	0.47	7.57	0.7	$1.65 \cdot 10^{-18}$

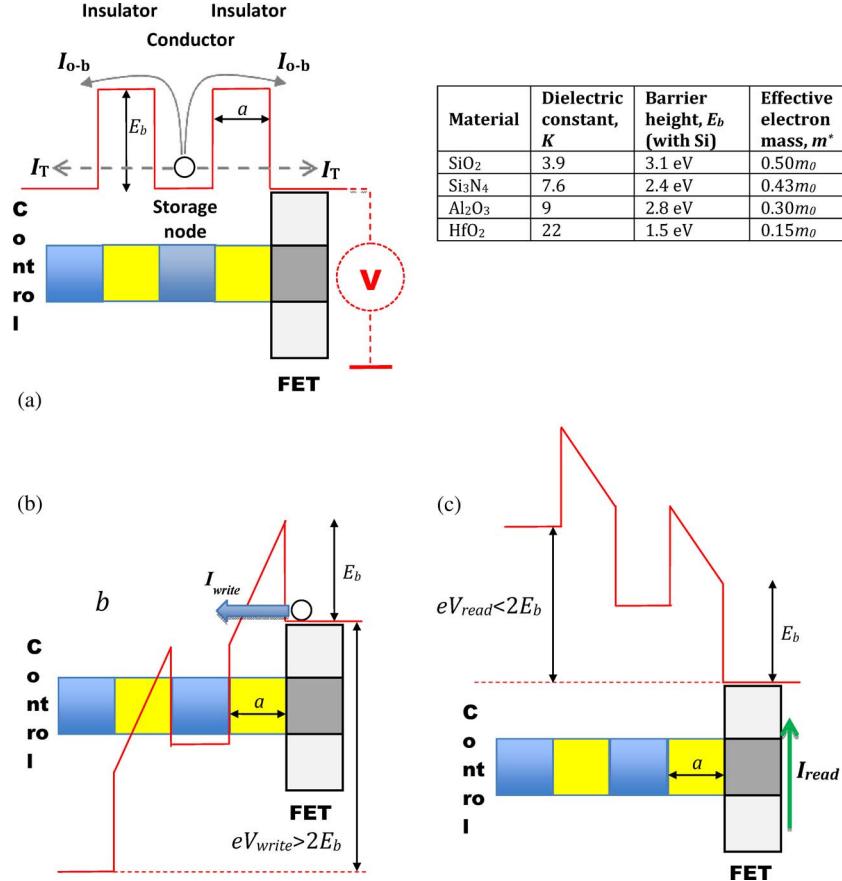


Fig. 6. The two-energy-barrier model for a memory cell: (a) the principle of storage and sensing; (b) WRITE operation; and (c) READ operation.

the storage node is isolated from the external contacts by two barriers (i.e., it is floating), this requires the total WRITE voltage applied to the opposite external contacts of the memory cell to be more than the doubled barrier height: $eV_{\text{write}} > 2E_b$, as shown in Fig. 6(b). (A symmetric barrier structure is assumed.) Thus, the floating gate structure inherently requires high voltage for the WRITE operation: for example, for SiO₂ barriers, $V_{\text{writemin}} > 6$ V, and the WRITE voltage should be $> 10\text{--}15$ V for faster ($\sim \text{ms}\text{--}\mu\text{s}$) operations.

The presence or absence of stored electric charge in the storage can be detected by an electrometer type device [shown schematically in Fig. 6(a)]. The sensing device should be in immediate proximity to the storage node. A FET is commonly used as a sensor, and a complete nonvolatile floating gate memory cell consists of a stack of metallic and insulating layers on the top of a FET channel, as shown in Fig. 6(b). The sensing FET is controlled by the voltage V_{read} applied to an external electrode, the control gate. The source-drain current of the FET depends on the presence or absence of charge in the floating gate, thus the memory state can be sensed by measuring the FET current. The control gate allows modulation of the semiconductor

channel of the FET by external commands, similarly to the logic FET. However, differently from a conventional transistor, the degree of accessibility of the channel from the control gate is rather limited. First, the control gate is physically far from the channel, since the minimal thickness of both top and bottom dielectric layers is large due to the retention requirements, and the minimal thickness of the insulator stack is > 10 nm. Second, the control gate affects the channel only indirectly, as the floating gate lies between the control gate and the channel. Therefore, a large READ voltage must be applied to the control gate for reliable ON/OFF transitions of the sense transistor. The maximum READ voltage is however limited by the condition for the F-N tunneling discussed above, and for the READ operation, it is $eV_{\text{read}} < 2E_b$, for example, for SiO₂ barriers, $V_{\text{readmax}} < 6$ V for a nondisturbing READ. In practice, a typical READ voltage is 4.5–5 V.

C. 2-D and 3-D Layouts of Logic and Memory Circuits

Binary switches in logic circuits will be assumed to be isolated, thus allowing for arbitrary wiring. From the tiling consideration, the most compact layout for an array of isolated devices (assuming at least one tile between each

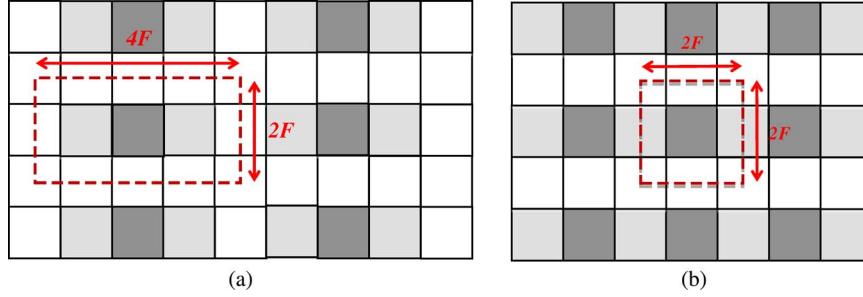


Fig. 7. Representation for maximum device density for (a) logic and (b) memory circuits.

device for insulation) results in maximum packing density of binary switches on a 2-D plane [Fig. 7(a)]

$$n_L = \frac{1}{8F^2}. \quad (3a)$$

In the following, (3a) will be assumed as the device density in the logic layout. Next, interconnects need to be added. To estimate the minimum number of interconnect tiles per device, assume that in a three-terminal device, for each terminal, at least one “contacting” interconnect tile (three total) is needed and one “connecting” interconnect tile (three total) is needed. This results in six interconnect tiles per binary switch. Including the contacting tiles would result in eight interconnect tiles per switch. Thus, the average interconnect length obtained from the tiling consideration is $\langle L \rangle = (6 - 8)F$. This estimate is consistent with the wire-length distribution analysis in practical microprocessors [24]. For this densest arrangement, at least three additional layers of interconnects would be needed.

Memory cells are typically organized in regular X-Y arrays, thus only simple *regular wiring* is needed. Regularly wired memory cells in an array can be connected in series, thereby enabling higher packing density, as shown in Fig. 7(b)

$$n_M = \frac{1}{4F^2}. \quad (3b)$$

[The serial connection of Fig. 7(b) represents a *NAND* array, the typical array architecture of mainstream flash memory products.]

The tiling framework provides a methodology to estimate the *average energy per bit* in an arbitrary logic circuit. As was argued in [25], at the limits of scaling, the energy per tile is nearly the same for both devices and interconnect tiles and approximately equals to the device switching energy (E_{sw} in Tables 2 and 3). For the total number of

tiles k (devices and interconnects), the average switching energy per bit is

$$E_{bit}(k) = \frac{1}{2}k \cdot E_{sw}. \quad (4a)$$

(The factor 1/2 originates from the assumed 50% activity factor.)

Now assuming the average interconnect length $\langle L \rangle = 6F$, the total number of tiles per device $k = 3 + 6 = 9$, then from (4a), we obtain

$$E_{bit} = \frac{9}{2}E_{sw}. \quad (4b)$$

The total dynamic energy consumption by a circuit of N binary switches will be NE_{bit} . Correspondingly, the energy dissipated by transistors themselves (without interconnects) is NE_{sw} . It follows from (4b) that the ratio of transistor energy use to the total energy consumed by a logic circuit constitute about 2/9 or 22% of the total dynamic energy consumption, which is consistent with the energy breakdown analysis in practical microprocessor chips [26].

For memory arrays, due to the regular wiring, in many instances, the properties of interconnecting array wires determine the operational characteristics of the memory system. A given cell in an array is selected (e.g., for *READ* operation) by applying appropriate signals to both interconnect lines, thus charging them. The relatively large operating voltage of flash results in rather large line charging energy $\sim C_{line}V^2$, where C_{line} is the line capacitance. For $F = 10$ nm and a 128×128 array, the line capacitance is $\sim 10^{-14}$ F [27]. If there is a random access *READ* with $V_{read} \sim 5$ V, there results an energy per line access (and therefore for random access operation) of $\sim 10^{-13}$ J, or $\sim 10^{-15}$ J/bit for serial access. For *WRITE* operation with $V_{write} \sim 15$ V, the *WRITE* energy is $\sim 10^{-12}$ J/line. (In practical flash memory devices, the *READ* energy is of the

Table 4 “Ultimate CMOS”: Limiting Density and Energetics

	Logic		Memory
	High performance	Low standby power	
F_{\min} (nm)	4.5	8	10
n_{2D} (cm ⁻²)	$6 \times 10^{11} (1/8F^2)$	$2 \times 10^{11} (1/8F^2)$	$2.5 \times 10^{11} (1/4F^2)$
V (Volts)	0.65	0.7	5 (read) 15 (write)
E_{bit} (J)	2.93×10^{-18}	7.41×10^{-18}	$\sim 10^{-13}$ (read)* $\sim 10^{-12}$ (write)*
P_{leak} (W)	2.34×10^{-9}	3.30×10^{-13}	low
T_{3D}	9F	9F	6F
n_{3D} (cm ⁻³)	$1.5 \times 10^{17} (1/72F^3)$	$2.7 \times 10^{16} (1/72F^3)$	$4.2 \times 10^{16} (1/24F^3)$

*random access operation

order of $10^{-13}\text{--}10^{-11}$ J/bit READ and $10^{-9}\text{--}10^{-10}$ J/bit WRITE [28], [29].)

A model for a tightly integrated 3-D system is useful as a conceptual tool in estimating ultimate performance of CMOS systems. The limits for the 3-D integration can be conceived using the methodology for stacking of 3-D tiles. For example, in logic circuits, the thickness of the FET layer is ~ 3 F (including vertical extension due to gate and 1/2 interlayer insulation from each side) and the interconnect layer thickness 2 F (with insulation). As was mentioned above, for this densest arrangement at least three additional layers of interconnects would be needed. Thus, the resulting thickness of one logic circuit layer is 9 F. In memory circuits, the thickness of one layer can be 6 F, which includes a layer of array grid interconnects and interlayer insulation. Taking into account (3a) and (3b) obtain limiting 3-D density for logic to be $1/72F^3$ and for memory $-1/24F^3$. Table 4 summarizes the essential parameters of CMOS logic and memory devices in the limits of scaling. A question that arises is whether alternative technologies exist that could offer further improvements beyond CMOS. Some examples are considered in Section V.

IV. NOT JUST MOORE (MORE THAN MOORE)

The possibility to extend the functionality of CMOS circuits by integration with other technologies has been referred to as “more than Moore” [30]. An example that has already encountered an extraordinary market success in the last few years is provided by CMOS imagers which can be found in any cell phone camera [31]. There Si photodetectors or phototransistors constitute the optical sensors, which are monolithically integrated on a CMOS chip [32]. Another multifunctional combination within Si technology is provided by the integration of microelectromechanical system (MEMS) devices with CMOS [33]. In this case, hybrid integration is already available at a prototype level; monolithic integration will follow once issues related to thermal mismatch between the two processes are solved. Along the same line, digital micromirror devices

(MDMs), micrometer size mirror realized on top of a CMOS circuit which controls their 3-D movement, are already used in projectors and TV sets [34]. A more ambitious path will be the integration not just of different functions within one material system, but also the integration of different technologies, as, for instance, Si and III-V semiconductors. Several examples can be provided to indicate how relevant such integration would be. As one looks at the fact that interconnection delay is already the bottleneck for the speed of an integrated circuit, it is clear that optical links would provide an obvious solution. Nonlinear optical passive components can be fabricated in Si technology, which could provide guiding, routing, and other optical functionality directly on chip. Light sources though are still a domain of compound semiconductors like GaAs and InP. Thus, the integration of light-emitting diodes (LEDs) or lasers realized with such materials is necessary. Lattice mismatch of the semiconductors, thermal mismatch of the processes, and fabrication compatibility pose serious challenges to integration, even at a hybrid level. Beside optics and radio frequency (RF), the integration with CMOS of other functions such as sensing, biological screening, or information storage can be of great interest. A review of some “more than Moore” devices and a brief discussion of the open challenges are provided below. Such a concept is nicely summarized in the scheme of a possible future chip shown in Fig. 8, which would be based on CMOS technology but would incorporate several other functionalities coming from alternative technologies [35]. The integration can be achieved directly on chip, requiring that the new technologies be fully compatible with CMOS. This is referred to as a “system-on-chip” approach. Alternatively, a 3-D integration is possible, where several chips, possibly realized with different technologies are stacked on top of each other (“system on package”).

A. RF Technologies

Wireless communication has witnessed an unprecedented (and maybe unexpected) development in recent years. It is therefore more and more important to

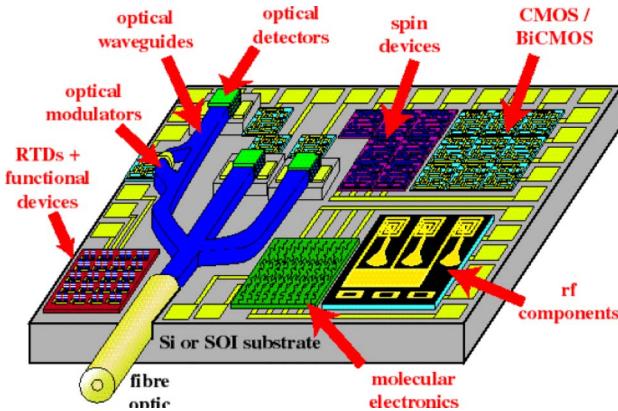


Fig. 8. Illustration of the integration of many technologies on a single CMOS substrate [35].

effectively bridge the two worlds of digital information processing and of RF transmission. If one looks at a modern cellular phone, a variety of specialized modules are present which perform specific tasks. Ideally, one unique chip should incorporate all necessary functionalities. RF functions are provided by antennas, filters, switches, and converters, which in turn use RF transistors, mechanical filters, and other active or passive devices. Many new materials and concepts have been introduced recently, which might have an important impact on RF applications in the future. In addition, they might help in the integration with CMOS technology.

MEMS technology has reached maturity and commercial success in recent years thanks to its application in electronic components for automotive and mobile communication [36]. One of its appealing features is the full compatibility with CMOS technology. By scaling MEMS to the nanoscale, further RF functionalities could be reached (e.g., nanoelectromechanical system (NEMS) resonators). Traditionally, some RF components such as the quartz crystals used in the reference oscillator are kept off-chip. In fact, integration leads to very poor quality factors and temperature instability, mainly due to the poor performance of the integrated inductors and capacitors. The best opportunities for miniaturization and integration of reference oscillators are provided by capacitively transduced microelectromechanical and nanoelectromechanical (MEM/NEM) resonators, which have reached in recent years operating frequencies of several gigahertz. Scaling to nanometer dimensions poses several problems connected, e.g., to fluctuations, friction, and dissipation mechanisms at the nanoscale. Nanostructures have been used and demonstrated in NEMS such as platinum [37] or Si [38] nanowires and carbon nanotubes (CNTs) [39], [40]. They are appealing due to their high stiffness, low density, defect-free structure, and ultrasmall cross section. Recently, graphene material has also attracted considerable attention for these applications [41].

The signal produced by this class of devices is very small and impedance matching can be a problem. One alternative is provided by movable gate transistors, which combine the advantage of a vibrating micro/nanostructure with the large output signal provided by the transistor drain current. Highly scaled versions of the in-plane resonant gate transistors with a front-end process have been reported based on silicon-on-nothing technology [42]. However, the sub-100-nm gaps and 400-nm-thick single crystal resonators suffered from the poor electron mobility. Nevertheless, full CMOS compatibility is guaranteed. An alternative structure has been proposed, the vibrating-body FET, where a combined effect due to modulation of the carrier density and of the piezoresistance in the channel is achieved. Si nanowires are the ideal channel material due to their pronounced piezoelectric properties.

Another appealing candidate for an RF oscillator that is fully CMOS compatible is provided by spin transfer torque effects [43]. In nanosized magnetic multilayer structures, metallic spin valves and magnetic tunnel junctions can drive uniform precession of the free-layer magnetization under an external input (provided by a magnetic field or an electrical current). This precession produces voltage responses that make those magnetic multilayers high-frequency spin torque oscillators. Oscillation frequencies ranging from several hundred megahertz to tens of gigahertz have been demonstrated [44]. The challenges that need to be overcome for practical applications include 1) reaching output powers in the milliwatt range, 2) improving the spectral purity of the oscillator, and 3) realizing auto-oscillating structures, thus eliminating the need for external magnetic fields.

RF mixers are an important building block of an RF front end. One candidate technology is the resonant tunneling diode [45]. Another candidate that should have provided low-noise solutions at RF, the single-electron transistor, could only demonstrate interesting performance at low temperature [46]. In principle, carbon-based devices possess the necessary nonlinearities in electrical characteristics to demodulate an AM signal [47], [48]. The main challenge consists, once again, in the integration of such technologies with CMOS [49].

It could be mentioned here that an alternative technology for RF applications (at least up to few megahertz) has been developed in recent years, based on polymeric devices and circuits. The advantages of such components are the low cost of their fabrication, the independence on the type of substrate used, and the possibility of large area manufacturing. Full organic radio-frequency identification (RFID) circuits including electronics and antennas have been realized. Such circuits include more than 1000 organic thin film transistors (OTFTs) and can operate up to 13.5 MHz [50]. Organic electronics [51] will not compete in speed with CMOS-based solutions, since the low mobility of organic semiconductors (typically below $1 \text{ cm}^2/\text{VS}$) limits the maximum achievable frequency. Nevertheless,

applications such as active matrix backbones for OLED display, ultralow-cost RFID systems, or biocompatible/disposable sensors can be envisaged.

B. Optical Technologies

As mentioned earlier, interconnections have become one of the limiting factors of the speed of integrated circuits. Moving to optical interconnect would greatly enhance the available bandwidth, reduce the heat dissipation on-chip, and assure immunity from electrical interference. Si or Si-compatible optical components have long been demonstrated [52], [53]. Attempts to obtain efficient Si-based light sources have instead not been very successful. Si-based micro/nanostructured devices have been introduced [54]. Considerable interest has been attracted by the demonstration of a Si laser based on the Raman effect [55]. Currently, only optical pumping has been demonstrated. For any realistic application, electrical pumping has to be achieved. The field of passive components (waveguides, filters, connectors) has witnessed considerable advances with the discovery of photonic bandgap structures [56]. Nanotechnology has allowed researchers to realize 3-D, 2-D, or 1-D periodic structures with tailored spectral transmission properties. By inserting defects into an otherwise perfectly symmetric lattice, it is also possible to deflect a light beam over distances of a few nanometers. Thus, an unconventional way to guide and deflect light on a chip can be fabricated with unprecedented properties [57]. Furthermore, by exploiting the properties of surface plasmons in quantum dots structures, the absorption and/or transmission properties of materials and surfaces can be engineered [58]. Being based on Si or on Si-compatible materials, such optical components can be easily integrated with CMOS.

Another optical component which is fully based on Si technology is the CMOS imager, which has witnessed considerable commercial success in recent years mostly due to their use in camera phones. Keys to this success have been features inherent to CMOS technology, such as size, weight, power consumption, mechanical robustness, and price. Two challenges remain on the agenda: the extension of the detectable range, especially to infrared, and scaling of the optical components to keep pace with the CMOS miniaturization.

Typically, CMOS imagers employ a Si photodetector or phototransistor as optical sensor. Spectrally, the sensitivity of such components is limited to the visible range because of the Si energy gap. In order to move into the infrared and far infrared range, which is very attractive for a variety of applications in the fields of security, screening, and environmental monitoring, a possibility is to use small gap semiconductor materials. Photodetectors and imagers using, e.g., InAs or CdTe have been demonstrated and even commercialized [59]. The main problems with such technologies are on the one side the need to cool down the detector in order to have an acceptable noise level and, on

the other hand, the difficulty to integrate III-V and II-VI materials with CMOS. A different approach uses a different concept. Rather than converting the optical signal into an electrical one via creation of electron-hole pairs following photon absorption, one can use the change in some property of the sensing material, for instance, resistance or temperature, under illumination. The most successful component of this type is the bolometer, which can be structured into arrays and driven by CMOS circuitry. A thorough review of infrared detector technologies can be found in [60]. Infrared uncooled cameras based on microbolometers integrated on CMOS are available on the market. They provide acceptable performance but they are quite expensive.

Photodetectors can be realized also using conductive polymers [61]. The advantages of such materials are: 1) the possibility to realize devices, circuits, and systems on large area substrates, which in turn can be of different nature (e.g., glass, plastic, textile, and paper); and 2) the reduced fabrication cost since the material can be processed from solutions. Thus, cheap preparation techniques such as ink jet printing, spin coating, and spray casting can be used [62]. The component that has received great attention not only in the research field but also on the market is the organic light-emitting diode (OLED) [63]. In fact, OLED displays have been introduced in several cellular and smartphones. This is the first time that an organic device, based on conductive polymers, has entered a large volume market. Besides OLEDs, other electronic and optoelectronic organic components have been demonstrated [64]. Organic solar cells based on blends of conducting polymers have also been fabricated with roll-to-roll processes, displaying a conversion efficiency of few percent [65]. Due to the flexibility in fabrication methods, organic photodetectors (OPDs) can be integrated onto CMOS. Inverted structures for OPDs that can directly be fabricated on CMOS as end-of-the-line process have been demonstrated [66]. CMOS imagers with OPD active pixels would guarantee a much larger fill factor with respect to Si photodetectors. In connection to IR imagers, either low gap polymers [67] or hybrid system combining polymers with quantum dots [68] have displayed room temperature sensitivity in this wavelength range. Their use for hybrid CMOS imagers would allow the realization of IR imagers at a cost comparable to conventional CMOS imagers.

Concerning pixel reduction, current technology allows for $2.2\text{-}\mu\text{m}$ pixel pitch, and demonstrations exist for $1.7\text{ }\mu\text{m}$. Pixel size reduction in active pixel sensors is crucial since it leads to higher numbers of pixels at almost constant price. A further reduction is challenging, both due to limited optical capabilities and to signal noise. A large part of today's imager cost is taken by lenses, whose complexity is bound to increase with miniaturization. In order to reverse such a trend: 1) innovative strategies exploiting coherent effects at the nanoscale have to be found, for instance, exploiting plasmonics and photonic

bandgaps; and 2) image correction via on-chip computation will have to be implemented, fully exploiting the capability of the CMOS chip.

C. Sensing Technologies

The computational power and the maturity of CMOS technology can be of great advantage in the sensor field, where environmental parameters have to be determined and corresponding actions undertaken. Since the signals to be sensed are mostly nonelectrical in nature, appropriate transducer elements are needed. Examples of external physical stimuli are mechanical (pressure, motion, vibration), electrical (voltage), thermal (temperature difference), electromagnetic (light), chemical (presence of a particular chemical species), etc. In response to the external stimulus, the transducer generates an electrical signal that is further processed by accompanying circuitry and is used to provide actionable information to the end user. Sensor metrics include sensitivity, selectivity, and repeatability. Nanotechnology can provide adequate solutions in the form of novel materials and structures with high sensitivity and into the Si mainframe technology. Two kinds of transducers are currently receiving considerable attention for sensing: 1-D structures (e.g., nanowires and nanotubes) [69], [70] and NEM devices [71].

Sensors may potentially be everywhere, providing instrumentation for the state of the environment, security, supporting regulation of different processes, etc. It is necessary in many applications for the sensors to communicate their data to a central information collection/decision-making authority. This gives rise to the need to establish sensor communication networks that can be used, for instance, to create autonomic systems that are user-transparent, self-healing, self-configuring, self-optimizing, and self-protecting analogous to many of the functions of the human nervous systems such as control of heart rate, breathing, etc. Of course, sensor networks in general represent an important field of endeavor where issues of configuration, optimum communication protocols, and information carrying capacity are essential concerns [72].

One example of the many application areas for sensors is in fields related to biology. The state of the living system can be monitored by sensing different physical parameters, e.g., chemical, electrical, optical, thermal, magnetic, etc. There are indications that 1-D structures, such as semiconductor nanowires and CNTs, may offer superior sensitivity to planar devices and allow for picomolar detection of biomolecules [73]. An additional attractive feature of 1-D structures is that they might lend themselves to minimally invasive probes to contact or even puncture the cellular membrane, or even to be ingested into the cell itself. This suggests the intriguing possibility of electrically monitoring processes inside the cell [74]. Recently, it has been shown that nanowires and nanotubes can also be used not individually but rather as a conductive film, which can be

used as semitransparent electrodes, sensors, transistors, and in general, for flexible and stretchable electronics [75]–[79]. Such solution-based technology is compatible with CMOS.

In many applications, sensors should operate in a standalone mode at extremely low levels of energy consumption. In some cases, operational energy could be harvested from the sensor environment, e.g., in the form of solar, thermal, electromagnetic, or mechanical energy. Currently, there is continuing progress in miniature energy harvesting devices to support autonomous operations of sensor units [80], [81]. Understanding of maximum performance potential of such energy harvesting devices, given practical size constraints, requires further studies. In parallel with the need to scavenge energy from the environment, there will be an increasing need to store it in batteries or capacitors.

Nanostructure and nanodevices, as well as novel materials, can be decisive in the search for efficient power solution of future electronic systems. Some very promising results have already been obtained. Among them, the so-called third-generation solar cells promise enhanced efficiency and/or reduced costs by using quantum nanostructures or organic semiconductors [82]. CNT or graphene sheets provide ideal solutions for compact, long-lasting miniature super capacitors [83]. Nanowires possess optical, electrical, and thermoelectric properties which can be useful in a number of energy-related applications [84]. In most cases, the technologies and devices just mentioned can be used as standalone technologies or in hybrid systems integrated on CMOS.

V. BEYOND MOORE

A. Terminology and Context

There is an international effort underway to identify an alternative to the CMOS transistor, which within one-to-two decades will no longer submit to feature size and voltage scaling [22]. Many of these alternative devices operate using state variables other than charge and some of them may offer functionalities beyond those of a binary device that could be useful for more complex operations. Indeed, the choice of state variable for a device not only has ramifications for device performance but echoes up the abstraction hierarchy to impact device-to-device communication, achievable chip complexity, and ultimately system performance capability. The dependency is depicted in Fig. 9. The symbols in Fig. 9 have the following meaning:

- L_{sw} the smallest device (switch) feature, e.g., gate length in CMOS;
- t_{sw} device switching time, i.e., time required to change state;
- E_{sw} the energy required to change the device state (switching energy);

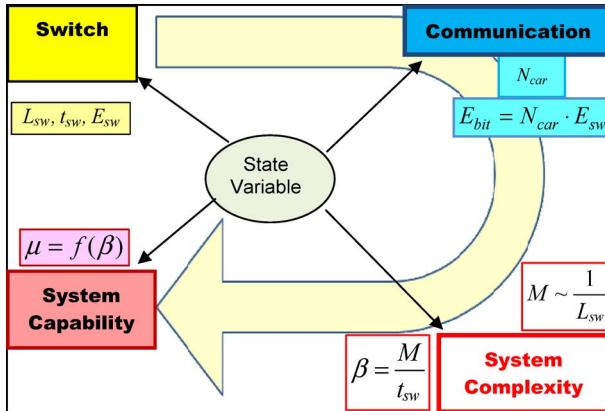


Fig. 9. State variable and different facets of information processing system.

N_{car}	the number of information carriers required to transmit state to downstream devices;
M	the device count, a measure of system complexity;
β	binary information throughput, a measure of technological capability;
μ	instructions per second, a measure of information processor capability.

The state of a binary switch is that minimum set of physical variables that fully describe the system and its response to a given set of control variables. In characterizing the functionality of various candidate devices, it is important to draw a distinction between the physical entities used in their realization and the properties of these entities utilized in the operation of the device, which we

refer to as variables. For example, physical entities might include electrons, atoms, ferromagnetic (FM) domains, etc. Associated with these physical entities are properties such as charge, spin, magnetic dipoles, etc.; the same entity might be used in two different devices, each exploiting a different property of that entity. In the following, “property” is used as a synonym for the word “variable” to agree with conventional usage.

Now each device has input, state, and output variables: for example, the FET utilizes the electron as the physical entity, and the properties of charge are used for the input, output, and state variables. On the other hand, the spinFET utilizes electrons but it is controlled by electron spin, its state is defined by spin, and its output is transferred as charge.

Table 5 provides a tabulation of physical entities and the properties employed by several of the emerging devices. Also, an expanded taxonomy employed by the ITRS Emerging Research Devices Chapter [22] is shown in Fig. 10.

B. Novel Device Examples

1) *III-V, Ge Channel, and Nanowire FET*: It is well known that III-V compound semiconductors are ideal candidates for high-speed devices (several tens of gigahertz), due to their excellent bulk electron (e.g. $33\,000\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ for InAs and $80\,000\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ for InSb) and hole ($1250\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ for InSb and $850\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ for GaSb) mobilities. The integration of GaAs and InP on Si substrates has been long sought but never achieved. Advances in epitaxial techniques have recently offered new perspectives on this challenge. In particular, Sb-based compound

Table 5 Taxonomy for Candidate Information Processing Devices

Device	Entity	Properties		
		Control Variable	State Variable	Output Variable
FET – Novel Materials (III-V, Ge, carbon-based, etc.)	Electron	Charge	Charge	Charge
SpinFET	Electron	Charge	Spin	Charge
Spin-Torque	Electron	Spin	Spin	Charge
Spin-Wave	Electron	Spin Waves	Spin	Charge Photon
Tunneling Transistor	Electron	Charge	Charge	Charge
Molecular switch	Electron or Atoms	Charge	Charge	Charge
NEMS	Atoms	Charge	Position	Charge
Atomic Switch	Atoms	Charge	Position	Electron
Memristor	Atoms	Charge	Charge, Electron	Electron
Magnetic Cellular Automata	FM Domain	Magnetic dipole	Spin	FM Domain
Moving Domain Wall	FM Domain	Magnetic Dipole	Spin	FM Domain
Multi-Ferroic Tunnel Junction	FM Domain	Spin	Charge	Electron
Optical or Plasmonics	Atoms or Electrons	Charge	Optical Density	Photons
Thermal Transistor	Phonons	Thermal Energy	Temperature	Phonons

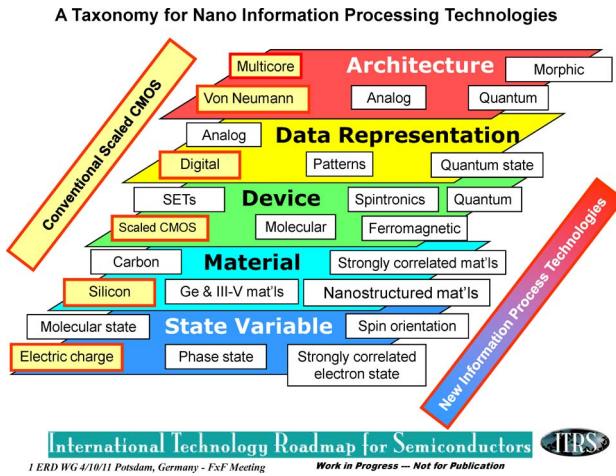


Fig. 10. ITRS taxonomy for information processing nanotechnologies [22].

semiconductors are seen as realistic CMOS channel replacement materials due to the high mobilities for both electrons and holes [85], [86]. A further appealing system is provided by InAs, which can be grown in the form of nanowires directly on Si substrates with excellent material quality [87]. The major challenges include the need for high-quality, high- k gate dielectrics (if MOSFETs are going to be used), damage-free low-resistivity junctions, and heterointegration on a very large-scale integration (VLSI)-compatible Si substrates. Similarly to III-V semiconductors, germanium (Ge) is also a potential channel replacement material because of its excellent bulk electron mobility of $3900 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, almost three times higher than in bulk Si. Unfortunately, the poor quality of the Ge/dielectric has resulted in much lower mobilities in fabricated transistors [88]. Strain engineering of Ge n -channel MOSFETs has also been studied as a performance booster technology and its effectiveness has been demonstrated at a small strain level. An open issue is whether the low electron saturation velocity in Ge will limit the short channel performance of n -channel Ge MOSFETs relative to Si n -channel MOSFETs. In conclusion, III-V compound semiconductor and Ge FETs are considered viable candidates to extend CMOS to the end of the Roadmap.

Nanowire FETs are structures in which the conventional planar MOSFET channel is replaced with a semiconducting nanowire. Such nanowires have been demonstrated with diameters as small as 0.5 nm. They may be composed of a wide variety of materials, including Si, Ge, various III-V compound semiconductors (GaN, AlN, InN, GaP, InP, GaAs, InAs), II-VI materials (CdSe, ZnSe, CdS, ZnS), as well as semiconducting oxides (In_2O_3 , ZnO, TiO_2) [89]. Nanowires can exhibit quantum confinement behavior, i.e., 1-D conduction, that can lead to the reduction of short channel effects and other limitations to the scaling of planar MOSFETs. Vapor-liquid-

solid (VLS) growth mechanism has been used to demonstrate a variety of nanowires, including core-shell and core-multishell heterostructures [90], [91]. Heterogeneous composite nanowire structures have been configured in both core-shell and longitudinally segmented configurations using group IV and compound materials. The longitudinally segmented configurations are grown epitaxially so that the material interfaces are perpendicular to the axis of the nanowire. This allows substantial lattice mismatches without significant defects. Vertical transistors have been fabricated in this manner using Si, InAs, and ZnO, with quite good characteristics [92]–[94]. The small lateral dimension of the nanowires allows their direct growth on lattice-mismatched substrates without the typical problems of dislocations and defects encountered in films. Thus, for instance, InAs of very good morphological quality has been grown directly on Si. Circuit and system functionality of nanowire devices has been demonstrated, including individual CMOS logic gates and other prototype circuit elements [95], [96]. Still a lot of work is needed to minimize parasitic components and achieve the high frequencies which have been predicted.

One of the crucial parameters controlling the power dissipation of CMOS devices is the subthreshold swing. In conventional MOSFETs, the thermal injection of carriers from the source to the channel sets a room temperature limit value of 60 mV/dec.

Tunnel FETs based on a gated p-i-n junction are expected to display an abrupt $I_{\text{on}}/I_{\text{off}}$ transition, thus lowering the subthreshold swing below the intrinsic MOSFET limit [97]. Such improvement is intrinsically connected to the quantum mechanical band-to-band tunneling process [98], which reacts sharply to variation of the gate voltage. High-performance tunnel FETs have been explored using low bandgap materials like Ge [99], SiGe [100], or based on Si nanowires [101] and CNTs [102]. A major challenge is the integration of such materials and structures on advanced Si platforms [103].

A completely different type of switch can be achieved exploiting the mechanical displacement of a solid beam controlled electrostatically to create a conducting path between two electrodes [104]. Such micro/nanoelectromechanical (M/NEM) switch has two major advantages with respect to MOSFETs: negligible leakage and negligible subthreshold swing. Thus, standby energy dissipation as well as dynamic energy consumption can be drastically reduced. The most recent developments suggest that M/NEM switches are attractive for ultralow-power digital logic applications. In addition, it is expected that the energy performance as well as the functional densities can largely improve with scaling. M/NEM switches can be fabricated by top-down approaches using conventional lithography techniques on Si, reaching actuation gaps as small as 15 nm [105]. Alternatively, bottom-up approaches employing CNTs [106] or Si nanowires [107] have been followed. In all cases, the leakage was virtually zero. The

main weakness is switching speed, as the beam requires around 1 ns to move from the off position to the on position. A further challenge for M/NEM switches is the control of the surface forces and the reliability of the contacts.

2) Carbon Electronics: The previous century has been the Silicon Century. The pervasion of electronic and optoelectronic devices in whole sectors of the society has been made possible mostly due to the success of CMOS technology (along the line of Moore's law). The new century might be the Carbon Century. Diamond has very interesting semiconducting properties, for instance, great heat and charge conductivity. Unfortunately, it is very difficult to obtain diamond in crystalline form at wafer level. Twenty years ago, CNTs were discovered. Some of their attributes make them very appealing in view of the miniaturization of electronic components. Despite a huge research effort, CNTs have not yet found a real application in nano and optoelectronics. Part of the problem is the difficulty to control the exact morphology (which in turn determines the CNT electronic properties) in a reliable and reproducible way. Recently, applications for a CNT network have emerged which make such system competitive with polymer materials for large area, low-cost electronics, and optoelectronics. A new carbon-made material has now appeared on the scene, receiving a great deal of attention. Graphene, a 2-D hexagonal grid of carbon atoms, has unique electronic, electrical, optoelectronic, and mechanical properties. It is therefore an appealing candidate for a variety of components like, e.g., transistors, sensors, electrodes, lasers. Although it is too early to forecast the market impact of graphene, the academic and industrial community as well the funding agencies are betting strongly on that novel nanomaterial. In the following, we will briefly discuss some of the important achievements for carbon-based devices and outline the main challenges.

CNT FETs are attractive because of the high mobility of charge carriers, the intrinsically small dimensions, and the possibility of minimizing short channel effects via all-around gate geometry. In the past two years, significant advances have been made in fabricating and characterizing CNT FETs [108], [109]. For instance, transistors with 15-nm channel length displayed no short channel effects and a transconductance of $40 \mu\text{S}$ for a single channel [110]. Frequencies as high as 15 GHz have been reached [111]. Nevertheless, major challenges remain, in particular concerning the ability to control 1) bandgap energy and nanotube chirality with sufficient precision for industrial applications; 2) the positioning of the nanotubes in required locations and directions; 3) the deposition of a gate dielectric; and 4) the formation of low-resistance electrical contacts.

Thanks to the extremely high electron mobilities, graphene is an ideal material for RF transistors [112]–[114]. Very high values of cutoff frequency have been demon-

strated, in excess of 200 GHz [115]. In order to achieve better performances, the quality of the source and drain contacts have to be improved, especially in the top gate configuration. Graphene FETs were first based on exfoliated graphene to form a transistor channel, which offers the highest mobility, but is hardly manufacturable [116]. Recently, epitaxial graphene on SiC substrates and chemical vapor deposition (CVD)-grown graphene on, e.g., copper foils, have been obtained [117], [118]. Back-gated graphene FETs with SiO_2 dielectric were typically shown to have room temperature field-effect mobilities up to around $10\,000 \text{ cm}^2/\text{Vs}$ [119]. Suspended graphene or graphene sheet on flat and inert substrate such as boron nitride can reach mobilities above $100\,000 \text{ cm}^2/\text{Vs}$ at room temperature [120], [121]. In top gate devices, lower mobilities are found, possibly because of a degradation of the channel properties when the gate dielectric is deposited [122]. Due to the peculiar band structure of graphene, electron and hole mobilities are similar. It also displays no energy gap, at least for extended single sheets. One crucial consequence are bipolar transport characteristics, which imply very small $I_{\text{on}}/I_{\text{off}}$ ratios. This is of course a major limitation for digital applications. Several methods to open a bandgap have been proposed, as, for instance, through the use of graphene nanoribbons [123].

3) Memristors: Recently, interest on hysteretic devices has risen in the context of nonvolatile memories. Such devices, named memristors, were pioneered in the work of Chua in the 1970s [124]. There, he indicated the memristor as the missing element, in addition to inductors, resistors, and capacitors, needed for a coherent description of electronic circuits. Much later, other groups rediscovered the definition in connection to nonlinear elements embedded in a crossbar architecture [125].

One possible memristor structure can be based on a polymeric film sandwiched between two metal electrodes [126], [127]. As pointed out earlier in connection to polymeric devices, the main motivation for using such material is the low fabrication cost. On the other hand, scaling has not been widely discussed. Although polymeric resistive memory arrays have been demonstrated, including a 3-D stack of three active layers, the memory operation mechanisms are still unclear [128]. Some research suggests that the changes in resistance could be due to intrinsic molecular mechanisms, charge trapping, or redox/ionic mechanisms [129].

Another type of memristive devices is the so-called “atomic switch,” basically an electrochemical switch based on the diffusion of metal cations and their reduction/oxidation processes to form/dissolve a metallic conductive path [130]. The metal atoms are introduced into the ionic conductive materials from a reversible electrode. The atomic switch was initially developed as a two-terminal device using sulfide materials that were embedded in a crossbar architecture with scalability down to 20 nm [131]. Later, an atomic switch using fully CMOS compatible

materials was developed to enable the formation of these devices in the metal layers of CMOS devices. This configuration resulted in the development of new type of programmable logic device [132]. Three-terminal atomic switches characterized by high I_{on}/I_{off} ratio, low ON-resistance, nonvolatility, and low-power consumption have also been demonstrated [133]. Several operating mechanisms have been proposed, including gate-controlled formation and annihilation of a metal filament, and gate-controlled nucleation of a metal cluster, but no complete understanding of the process currently exists. Switching speed, cyclic endurance, uniformities of the switching bias voltage, and resistances both for the ON-state and the OFF-state should be improved for general usage as a logic device [134].

In a variety of materials, ion migration combined with a redox process can cause a change in resistance of a metal–insulator–metal structure [135]. For instance, for silver electrode, Ag⁺ cations can drift through the insulator in the presence of an applied voltage, forming a highly conductive filament connecting the metal electrodes resulting in the ON-state of the cell. Reversing the applied voltage, an electrochemical dissolution of these filaments takes place, resetting the system into the high-resistance OFF-state [136]. In the case of transition metal oxides, such as TiO₂, the motion of oxygen vacancies is responsible for the change in the cell resistance. In a third class of materials, a unipolar thermochemical mechanism leads to a stoichiometry change due to a current-induced increase of the temperature. In some cases, a formation process is required before the bistable switching can be started. Since the conduction is often of filamentary nature, memories based on this bistable switching process can be scaled to very small feature sizes. The switching speed is limited by the ion transport, typically rather slow. Thus, the distance between the electrodes has to be limited to a few nanometers. Although the microscopic nature of the switching process has yet to be understood in detail, recent experimental demonstrations of scalability, retention, and endurance are encouraging [137].

From an architectural point of view, memristive devices could be coupled with two-terminal select devices in order to build passive memory arrays (crossbars) [138]. The general requirements for such two-terminal switches are sufficient ON-currents at proper bias to support READ and WRITE operations and sufficient ON/OFF ratio to enable selection even in the absence of a transistor. These specifications are quite challenging and severely limit the maximum size of a crossbar array [139]. Currently, two approaches to integrating a two-terminal select device with storage node are being pursued. The first approach integrates the external select device in series with the storage element in a multilayer stack. The second approach uses a storage element with inherent nonlinear properties. The simplest realizations of two-terminal memory select devices use semiconductor diode structures, possibly in a

back-to-back configuration for bipolar memory cells. Alternatively, a selector exhibiting resistive switching behavior could be used. That is, the selector works on the same principle as the restore element, the main difference being that it can be volatile. One possible device is based on a metal–insulator transition and exhibits a high resistance for voltage below a given value. As an example, a VO₂-based device has been demonstrated as a select device for NiO_x resistive random access memory (RRAM) element [140]. The main challenge for switch-type select devices is to identify the right material and the switching mechanism to achieve the required reliability, drive current density, and I_{on}/I_{off} ratio.

In addition to memories, it has been suggested that logic gates can also be built using memristors [141]. Furthermore, neuromorphic architectures based on memristive crossbars have been investigated [142], [143].

4) Molecular Electronics: One approach to beyond CMOS electronics is based on the use of single conductive molecules [144], [145]. Due to their intrinsically small size and the possibility to use self-assembling techniques, single molecules could be an alternative to Si nanostructures for nonvolatile memories, diodes, or switches [146], [147]. In fact, when properly functionalized, single molecules can display nonlinear electrical characteristics and, in some cases, hysteresis [148]. In a molecular memory, data are stored by applying an external voltage that causes a transition of the molecule into one of two possible conduction states. Data are read by measuring resistance changes in the molecular cell. The concept emphasizes extreme scaling; in principle, one bit of information can be stored in the space of a single molecule, namely, few nanometers. Computing with molecules as circuit building blocks is an exciting concept with several desirable advantages over conventional circuit elements. Because of their small size, very dense circuits could be built and bottom-up self-assembly of molecules in complex structures could be applied. However, major challenges still exist. First, the very nature of the molecular conduction and molecular switching has not been fully understood. The role of the metallic leads is not clear and parasitic effects due to the environment could appear which might determine the transport characteristics of a molecular device. In any case, prototypical molecular memories have been built, which show remarkable endurance and reproducibility [149], [150]. At an architectural level, both molecular quantum cellular automata (QCA) and crossbar structures have been investigated [151], [152].

5) Magnetic Components: Electronic systems combining computing and storage capabilities could be realized based on magnetic structures. Magnetic random-access memories (RAMs) [153] are a mature technology with some products already on the market. The control of single spins of either atoms or electrons has also been proven a

promising new way to achieve electronic functionalities. The possibility to build logic circuits with magnetic nanostructures has been demonstrated at a prototypical level. There, a novel architecture based on field coupling [called magnetic quantum cellular automaton (MQCA)] is adopted [154], where the spatial arrangement of coupled nanomagnets can be used to build logic functions and complete circuits. In the following, we will briefly describe some of the suggested magnetic components.

In spin transistors, the current is controlled by the magnetization configuration of the ferromagnetic electrodes or by the spin direction of the carriers [155]. Thus, feature could lead to low-power circuit architectures that are inaccessible to ordinary CMOS circuits. Recently, an experimental demonstration of spin FET was reported [156], [157]. Oscillatory spin signals controlled by a gate voltage were observed implying spin precession of spin-polarized carriers in the channel. However, the origin of the observed spin signals is not yet clear. Spin MOSFETs using ferromagnetic electrodes have also been proposed but not yet demonstrated [158].

Spin wave devices (SWDs) are a type of magnetic logic exploiting collective spin oscillation (spin waves) for information transmission and processing [159]. The spin waves are generated in a magnetoelectric cell which is driven by external voltage pulses. Such a cell also acts as detector and storage element. The information is encoded into the initial phase of the spin wave. Spin waves propagate through spin wave buses and interfere at the points of junction constructively or destructively, depending on the relative phase. The result of computation can be stored in the magnetization or converted into the voltage pulse by the output magnetoelectric cells. The primary expected advantages of SWDs are: 1) the ability to utilize phase in addition to amplitude for building logic devices with a fewer number of elements than required for transistor-based approach; 2) nonvolatile magnetic logic circuits; and 3) parallel data processing on multiple frequencies at the same device structure by exploiting each frequency as a distinct information channel. Prototypes operating at room temperature and at gigahertz frequency have been demonstrated.

In nanomagnetic devices, binary information can be encoded in the magnetization state. Fringing field interactions between neighboring nanomagnets can be used to perform Boolean logic operations [160]. A functionally complete logic set based on nanomagnets has been demonstrated [161]. In addition, nanomagnetic devices have nonlinear response characteristics, the output of one device is capable of driving another, power amplification (or gain) is present, and dataflow directionality can be obtained. Nanomagnet logic (NML) has therefore a great potential for low-power applications. A clock modulates the energy barriers between magnetization states in an NML circuit. Recently, experimental demonstrations of individual island switching as well as the reevaluation of

NML lines and gates with CMOS-compatible clock structures have been reported [162]. Furthermore, NML appears to be scalable to the ultimate limit of using individual atomic spins. Whether a circuit ultimately exhibits reliable and deterministic switching is a function of how it is clocked—and requires additional study.

Field coupling via magnetic interaction belongs to a novel class of architectures called MQCA [154]. A cellular automaton (CA) is an array of cells, organized in a regular grid [163], [164]. Each cell can be in one of a finite number of states from a predefined state set, which is usually a set of integers. The state of each cell is updated according to transition rules, which determine the cell's next state from its current state as well as from the states of the neighboring cells. The functionality of each cell is defined by the transition rules of the CA. Typically, each cell encodes one bit into a single electrical or magnetic dipole. The cell-to-cell communication is guaranteed by magnetic interaction between neighboring dipoles. A QCA architecture has some appealing features: its regular structure has the potential for manufacturing methods that can deliver huge numbers of cells in a cost-effective way. Top-down as well as bottom-up manufacturing methods can be used. Furthermore, the design of a cell can be relatively simple as compared to that of a microprocessor unit, so design efforts are greatly reduced. Wires are completely unnecessary since the cells can interact with their neighboring cells through some physical mechanism. Thus, interconnection delay and power dissipation through interconnects are avoided. Clearly, QCAs also have some drawbacks and challenges. For instance, input and output of data to cells with nanometer dimensions may be difficult. Clocking the cells requires additional wires or external inputs. Speed might be a limiting factor. Room temperature operation has to be assured for any realistic application, which, up to now, has only been demonstrated for magnetic QCAs.

A concept that combines spin-controlled devices and nanomagnetic logic has been proposed recently [165]. In the all-spin logic (ASL), the information stored in the nanomagnets propagates as spin current in spin coherent channels. Recent advancements have shown that a combination of spintronics and magnetics can provide a low-power alternative to charge-based information processing. Key elements of ASL are the spin injection into metals and semiconductors from magnetic contacts and the switching of magnets by injected spins. Major challenges to be overcome are room temperature operation and a further improvement of the energy-delay product. It should be mentioned that ASL could also provide a natural implementation for biomimetic systems with architectures that are radically different from the standard von Neumann architecture.

6) *New Architectures for Beyond CMOS:* Research on architectures that exploit the properties of the devices described in this section is at an early stage of development.

Many different possibilities exist, two of which are CA and neural-inspired networks. CA typically use a form of nearest neighbor communication and they can be shown to be universal. Theoretical and experimental quantification of CA performance relative to the von Neumann architecture remains an open question [164]. Neural networks take a different approach by seeking to emulate structures in the brain and these have been studied for decades. So far it appears that neural networks can offer advantages for special classes of problems. There are indications that memristors in crossbar arrays may be able to emulate neural behavior. In the next section, a perspective on architectures for computation inspired by the operation of living cells is provided.

VI. BIOLOGICAL COMPUTATION: LIVING CELL EXAMPLE

A. A Basis for Quantitative Comparisons

The reliance of CMOS and many other proposed information technologies on electron charge to support their operations places them at risk as features scale downward into the few nanometer regime. Not only does tunneling become detrimental to performance, but also smaller features usually make the devices more susceptible to minute, manufacturing-induced, variations in material structure and composition. It has been said that the creativity of nature far exceeds that of humans and it seems reasonable to seek inspiration for new information processing technologies from this source. In that which follows, it is argued that the living cell can be viewed as an information processor that is extraordinarily efficient in the execution of its functions. The living cell is, in a sense, a universal constructor as suggested by von Neumann, which is capable of creating copies of itself [166], [167]. The model that is used in the following is the *E.coli* cell which has dimensions on the order of 1 μm and which has been heavily studied so that quantitative estimates of its complexity, performance, and energy efficiency are available. Given this, it is important to point out that many of the mysteries of cell operation are yet unresolved and are the focus of continuing investigations.

In order to provide a benchmark for *E.coli* cell operation, we first extrapolate the capabilities of a 1- μm scale CMOS information processor when end-of-scaling CMOS technology is utilized. Favorable assumptions for the 1- μm CMOS cell are offered including the stipulation that no volume is required for energy storage and for communication. A development is then offered, from available data, of the information processing capability of the *E.coli* cell. It is argued that the information processing capabilities of the *E.coli* cell far exceed that of the 1- μm CMOS cell and inferences are drawn suggestive of directions for future information processing technologies. The terminology *in silico* is used to refer to the semiconductor benchmark

cell and the term *in carbo* is used to refer to the *E.coli* cell in the following.

B. Bio- μ Cell Information Processor

Are there example information processing systems now extant from which inspiration might be drawn for new technologies? It has been recognized that individual cells, the smallest units of living matter, possess amazing computational capabilities, and are indeed the smallest known information processors [168]. As is argued in a number of studies, individual living cells, such as bacteria, have the attributes of a Turing machine, capable of a general-purpose computation [168]–[170]. It can also be viewed as *universal constructor* in the sense of von Neumann because it manufactures copies of itself, thus a computer making computers [169].

Just how does the cell go about implementing its information processing system? The cell is a very complex organism and any brief attempt to describe its operations is bound to be inadequate. A vastly oversimplified view of cellular processes is presented below.

A cell's primary functions can be described as follows.

- 1) **Reproduction:** making cells by acquiring/processing information from internal storage (DNA) and utilizing the structural building blocks and energy from the nutrients.

The reproduction task requires a massive information processing effort, a crude estimate of which is made in Section VII-A. In short, elementary structural building blocks (22 amino acids and five nucleotides) need to be synthesized or acquired, and then utilized to form functional building blocks, which include different proteins, RNA, and DNA molecules. Finally, all building blocks need to be properly placed within cell's volume for assembly. A special cell-cycle control mechanism regulates the sequence, timing, etc., of the cell assembly process.

- 2) **Adaptation for survival:** Acquiring/processing information from external stimuli with feedback from DNA.

Single-cell organisms, such as *E.coli* bacteria, could not survive without the ability to sense the environment and adapt to its changes (positive or negative). For example, in response to the external presence of specific nutrients, particular proteins are produced within the cell to facilitate the uptake and digestion of those nutrients. In the absence of nutrients, the cell can switch to a resting mode, where the reproductive process is inhibited. In addition, single-cell organisms can respond to a variety of external stimuli such as temperature, light, presence of toxic chemicals, magnetic field, etc. Many single-cell organisms also possess motility organs (e.g., flagellae in case of *E. coli*).

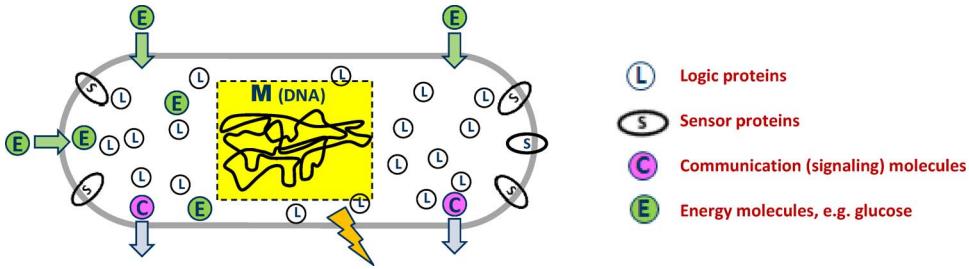


Fig. 11. unicellular organism as information processor.

3) **Extracellular communication:** Sending and receiving signals to coordinate community behavior.

Many unicellular organisms communicate to each other by the release and detection of special signal molecules. Cells use chemical signaling to detect population density and to exchange information about the local environment. Cell-to-cell communication coordinates the behavior of a cell population to increase access to nutrients, provide for collective defense, or enable the community to escape in case of threats to its survival.

In the following, we offer a simple estimate of single-cell computational capabilities based on two different approaches. A bottom-up approach counts the cell hardware, i.e., the number of memory and logic elements in the *in carbo* processor. The top-down approach deals with total amount of computation needed to implement operations to assemble a new cell.

C. Cell Hardware

Fig. 11 shows a cartoon of a cell as information processor. It contains a localized long-term memory block M (DNA molecule), a number of short-term memory and logic units L (different protein and RNA molecules), (input) sensors S to monitor both outside environment and the cell interior (extracellular and intracellular receptor proteins), and two output units: the ribosomes, where new structural building blocks for reproduction are “printed,” and signaling units that “wirelessly” connect to neighboring cells by sending signal molecules.

The cell hardware is made from three types of macromolecules: proteins, DNA, and RNA. Table 6 presents a summary of the statistics and functions of these molecules in *E.coli* cell. A description of essential features of different parts of the cell hardware is given below.

1) **Logic Hardware:** Many proteins (Fig. 12) in living cells have as their primary function the transfer and processing of information, and are therefore regarded as logic elements of the *in carbo* processor [171]–[174]. In fact, as recent studies indicate, the proportion of components devoted to computational networks increases with the com-

plexity of the cell, and are absolutely dominant in humans [173]. Proteins can alter their 3-D structural shapes (conformation) in response to external stimuli, and different conformations can represent different logic states. These nanomechanical changes form a state variable, sometimes called *conformon* [175]. The essential functions of the protein devices are determined by their conformational states. A simple example of the “binary” conformational change is the ion channel protein, which is embedded in a cell’s membrane and acts as a gate for ions, and can be opened or closed depending on command from either internal or external sources, e.g., light, pressure, chemical signal, etc. Different nanomechanical conformations of these protein devices are recognized by other elements of the *in carbo* cell circuit by a process based on selective affinity of certain biomolecules with given conformational states. Molecular recognition implemented with conformons plays a fundamental role in the communication of information packages within the processor, and it facilitates targeted interactions between different elements, e.g., protein–protein, protein–DNA, RNA–ribosome, etc.

The protein *conformons* control all processes in the cell, such as sensing, signaling, information retrieval, etc. Some examples will be given in the next section.

2) **Memory Hardware:** All data about structure and operation of a living cell are stored in the long DNA molecule. DNA coding uses a base-4 (quaternary) system. The information is encoded digitally by using four different molecular fragments, called nucleobases, to represent a state: adenine (A), cytosine (C), guanine (G), and thymine (T). The four molecular state symbols are attached in series to a flexible “tape” or a “backbone” made of sugar and phosphate groups. The complete DNA unit consists of two complementary “tapes” forming the so-called double helix. Each state symbol (base) on the first tape forms a pair (base pair) with a complementary state symbol on the second tape: adenine forms a pair with thymine, while cytosine forms a pair with guanine. Information content in each tape is identical, but is written with different (complimentary) sequences of symbols. Thus, the base pair (bp) is a natural unit of information stored in DNA. One bp equals to two bits of binary information and corresponds to

Table 6 Essential Parameters of the *E.coli* Molecular Processor

Devices		
<i>In Carbo 'device' count</i>		<i>Function</i>
DNA size	4.6×10^6 bp=9.6 Mbit	Nonvolatile memory
Number of RNA/cell	222,000	Memory interface
Number of cytoplasmic proteins	1,000,000	1) Logic processor 2) Signal processor 3) Metabolic functions*
Number of ribosomal proteins	900,000	I/O interface
Number of logic 'devices' (Proteins and RNA)	>1,000,000	1) Logic processor 2) Signal processor 3) Memory interface 4) I/O interface
Number of all proteins	3,600,000	1) Logic processor 2) Signal processor 3) Metabolic functions* 4) Structural functions*
Timing		
Time for cell replication	40min=2400s	
Energetics		
Number of glucose/cell	200,000-400,000	
Number of ATP/cell	500,000-3,000,000	
Total energy stored	$\sim 2 \times 10^{-12}$ J	
Power dissipation	1.4×10^{-13} W	

*proteins with metabolic and structural functions, excluded from the device count

approximately 0.34 nm of length along the tape, as shown in Fig. 13.

Some examples of DNA storage capacity (genome size) are given in Table 7. Note that the storage density of molecular DNA memory is ~ 10 Mb/ μm^3 or 10^{19} b/cm 3 , which is much denser than the density limits for the electronic long-term memory evaluated in Section III-B. Also it is interesting to note that the single-cell organism *Amoeba Dubia* stores a huge amount of information (1.34 Tb), compared to ~ 6 Gb stored in the human genome.

a) *READ from long-term memory*: The access to the DNA memory is facilitated by special protein devices, which form memory interface circuitry. For example, the DNA-binding proteins (forming transcription factor complex) act as gates to the specific snippets of DNA, and they therefore represent a mechanism to retrieve information from a specific DNA address. The cell's signaling network regulates the state of the DNA-gating proteins that determine when and where a DNA snippet (a gene) is activated, thus address specification. Next, the information is retrieved from the specified address by a special device formed by the RNA polymerase protein. This protein acts as a memory read head moving along the specified snippets of

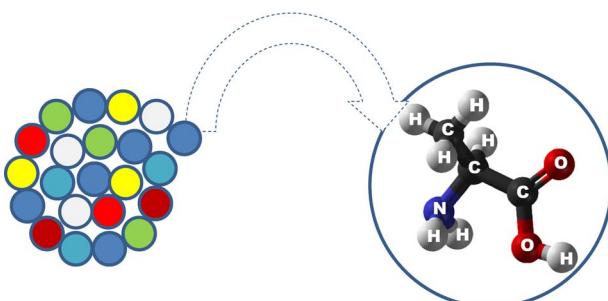


Fig. 12. Protein molecule formed from different amino acids (shown as circles of different colors).

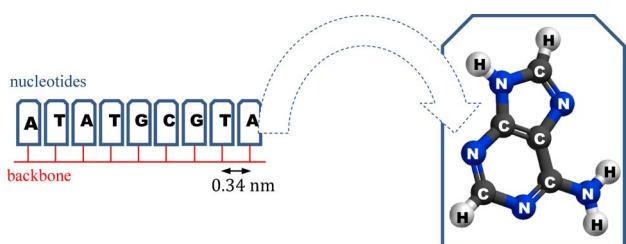


Fig. 13. A fragment of DNA molecule formed from four different nucleotides.

Table 7 DNA Storage Capacity (Genome Size) for Several Representative Cellular Organisms

	Cell	Volume, μm^3	Genome size
<i>Cyanobacteria</i>	<i>E.coli</i>	1	$4.6 \times 10^6 \text{ bp} = 9.6 \text{ Mbit}$
<i>Bacteria</i>	<i>P. marinus</i>	0.1	$1.75 \times 10^6 \text{ bp} = 3.5 \text{ Mbit}$
<i>Protozoa</i>	<i>Amoeba dubia</i>	$\sim 10^7$	$670 \times 10^9 \text{ bp} = 1.34 \text{ Tbit}$
<i>Human cells</i>	(average parameters)	~ 1000	$\sim 3 \times 10^9 \text{ bp} = 6 \text{ Gbit}$

DNA and copying its information by synthesizing pieces of messenger RNA (mRNA) molecules. The mRNA then transfers the information to the sites of protein synthesis, the ribosomes. RNA is thus part of the memory interface, with functions of the memory buffer (DRAM, SRAM) and the information package which facilitates interactions between memory, logic, and input/outputs (I/Os). Note that the information can be retrieved from different parts of DNA memory simultaneously, thus this is a highly parallel process with many RNA more or less simultaneously accessing snippets of the DNA and transferring this information to the ribosomes to make proteins.

Different parts of the DNA memory of the cell are continuously accessed to support its operation. One example is signal transduction, which is the DNA-controlled process of cellular response to external stimuli.

b) *Writing to long-term memory*: The view that DNA is a read-only memory has undergone a dramatic change in recent years. The copying of the parental DNA to the offspring, called vertical gene transfer, is the basis for inheritance and until recently was regarded as the only or at least the vastly predominant mechanism for transferring the genetic information. There is, however, an alternative mechanism for information transfer, which is lateral gene transfer. This can happen: 1) by a direct uptake (swallowing) of naked DNA from the cell environment; 2) by a virus; and 3) by direct physical contact between two cells. Fragments of DNA, imported from outside, can be integrated into the host DNA, and thus new information is written in the memory unit. Until the advent of the genome-sequencing era, a prevailing opinion among the research community was that lateral gene transfer was a rare and insignificant event. Currently, it is recognized that in prokaryotic, e.g., bacterial cells, lateral transfer is the predominant form of genetic variation and is one of the primary driving forces for bacterial evolution. In fact, the scale of lateral gene transfer can be very large: for example, two different strains of *E.coli* differ more radically in their genetic information than all mammals.

VII. QUANTITATIVE ESTIMATES FOR THE BIO- μ CELL AND THE Si- μ CELL

A. The Bio- μ Cell

The overall information content of a material system consists of information about the system's composition and

shape [176]. For example, if a case of von Neumann universal constructor is considered, i.e., a computer with the task of controlled assembly of the structure (e.g., another computer) from building blocks, a certain amount of information must be processed, which is related to the complexity of materials system. For each step, the computer must: 1) select the appropriate category of the building blocks; and 2) calculate x -, y -, z -coordinates of the position for each of the building blocks. If there are N different building blocks (in the case of ultimate bottom-up construction, these building blocks could be atoms composing a material structure), information content of selection in step 1 is

$$I_s = \log_2 N \text{ (bit)} \quad (5)$$

and the information of the xyz-positioning is

$$I_{xyz} = 3n \quad (6)$$

where n is the lengths of a binary number representing each coordinate. In this estimate, $n = 32$ b will be used, which is sufficient for representing numbers with practically arbitrary precision ("floating point" format).

Thus, if the total number of the building blocks in a material structure is K , the total information processed in assembly is

$$I_K = K(\log_2 N + 3n). \quad (7)$$

Now, consider the task assembling of living cell of *E.coli* bacterium from individual atoms. The elemental composition of the bacterial cell is known with high accuracy and is shown in Table 8 [177]. The cell is mainly composed of ten different atoms with the total number of $\sim 3 \times 10^{10}$ atoms. Thus, using (7), we obtain

$$I_{\text{cell}} \sim 3 \times 10^{10} \times (\log_2 10 + 3 \times 32) \sim 3 \times 10^{12} \text{ bit.}$$

This result is remarkably close to the experimental estimates of the informational content of bacterial cells

Table 8 Elemental Composition of *E.coli* [177]

Element	% of dry weight	N_{at}
C	50	8×10^9
O	20	2×10^9
N	14	2×10^9
H	8	1×10^{10}
P	3	2×10^8
S	1	6×10^7
K	1	5×10^7
Mg	0.5	4×10^7
Ca	0.5	2×10^7
Fe	0.2	6×10^6
Total	98.2%	3×10^{10}

based on microcalorimetric measurements which range from 10^{11} to 10^{13} bits per cell. In the following, it is assumed that $I_{cell} \sim 10^{11}$ bit, i.e., the conservative estimate is used.

B. The Si- μ Cell

For a system-level comparison between extremely scaled Si-based technology and carbon-based computational elements in biosystems, consider a hypothetical computer that is realized in a cube of $1\text{-}\mu\text{m}$ in size (the volume of the bio- μ cell). Such computer, later referred to as Si- μ cell, must contain logic circuitry and non-volatile memory to store program. Suppose further all components of the computer are to be implemented in ultimately scaled Si technology summarized in Table 4. In the following, 3-D-stacked logic and memory circuit layers will be used to fill the $1\text{-}\mu\text{m}^3$ volume. (In Table 4, the thickness of one layer in the stack is assumed to be 9F for logic and 6F for memory.) The corresponding densest conceivable 3-D arrangement of FETs is 1.5×10^{17} transistors/cm³, and thus $1\text{-}\mu\text{m}^3$ volume could contain up to 150 000 logic transistors. For nonvolatile memory, the densest 3-D stack of NAND layers is 4.2×10^{16} b/cm³, or 42 kb of memory in $1\text{-}\mu\text{m}^3$. Comparing to the bio- μ cell in Table 6, the Si- μ cell can contain $\sim 10\times$ less logic elements and more than $100\times$ less memory. (Note that this estimate was made even before partitioning of the $1\text{-}\mu\text{m}^3$ volume between logic and memory and not including an energy source.)

Next, according to Table 4, the OFF-state leakage power in the ultimate FET circuit is ~ 2.34 nW per transistor, thus $\sim 357\text{ }\mu\text{W}$ of total static power dissipation in a system of 152 000 FETs. This results in catastrophic heat densities in the $1\text{-}\mu\text{m}^3$ cube: $q = 357\text{ }\mu\text{W}/6\text{ }\mu\text{m}^2 = 5800\text{ W/cm}^2$. This is almost equal to the heat density at the Sun's surface ($\sim 6000\text{ W/cm}^2$, as shown in Table 8). Clearly, such Si- μ cell computer cannot exist. Therefore, for such a

system, larger scale devices or/and smaller device count must be used.

What is the smallest device count that could suffice for the Si- μ cell? von Neumann has argued that the minimum logic circuit complexity required to implement general-purpose computing is of the order of a few hundred devices [178]. In an attempt for a more accurate estimate of the *von Neumann threshold*, a model 1-bit minimal Turing machine (MTM) has been constructed with total device count of about 320 binary switches/transistors and requires 8-b instruction words for its operation [25]. In the following, it will be assumed that the logic processor of the Si- μ cell is implemented by an MTM. This also allows the maximization of the amount of memory in Si- μ cell (which is much less than the bio- μ cell).

Suppose the MTM is implemented within the Si- μ cell using FETs with $L_g \sim 4.5\text{ nm}$ with the parameters listed in Table 4. The remainder of the $1\text{-}\mu\text{m}$ cube is available for memory.

Implementation of each of the MTM instructions requires a minimum of three sequential operations/cycles [25]. On average, $\sim 50\%$ of transistors are active during each cycle, thus ~ 160 switching events per cycle or ~ 500 switching events per instruction. Since execution of one instruction results in one output bit, the ratio of the total transistor switchings (raw bits) to the output bits is 1/500. Therefore, in order to generate 10^{11} output bits, the typical outcome of biological computation, at least 5×10^{13} raw bits must be processed in the MTM. It takes 3×10^{11} MTM cycles to complete the computational task (i.e., three MTM cycles per one output bit). If it is required that these events occur over 2400 s (to match the bio- μ cell), the cycle time $t_{cycle} = 8\text{ ns}$. This appears to be easily achievable by CMOS technology. The total switching energy and power per MTM cycle are

$$E_{cycle} = N \cdot E_{bit} = 320 \times 2.93 \times 10^{-18} \\ = 9.36 \times 10^{-16} \text{ J} \quad (8a)$$

[note that E_{bit} in (8a) corresponds to the 50% activity factor (4b)] and

$$P_{active} = \frac{E_{cycle}}{t_{cycle}} = \frac{9.36 \times 10^{-16}}{8 \times 10^{-9}} = 1.17 \times 10^{-7} \text{ W.} \quad (8b)$$

(There is also leakage power consumption and this is approximately 749 nW.)

Next, the energy consumed by the memory access needs also be taken into account. At each cycle, an 8-b instruction must be read from the memory block. Assuming a serial READ (typical for NAND memory) with

Table 9 Energetics of Si- μ cell Implemented With Ultimate High-Performance CMOS Technology

	Logic	Memory
F_{min} (nm)	4.5	10
N	320	40,000
E_{bit} (J/bit)	2.93×10^{-18}	$\sim 10^{-15}$ (read)*
E_{cycle} (J)	9.36×10^{-16}	$\sim 10^{-13}$ (read)
t_{cycle} (s)	8×10^{-9}	
P_{active} (W)	1.17×10^{-7}	1.25×10^{-5}
P_{leak} (W)	7.49×10^{-7}	assumed low
P_{total} (W)	1.34×10^{-5}	
E_{total} (J)	3.21×10^{-2}	
Q_{active} (W/cm ²)	1.95	208
Q_{leak} (W/cm ²)	12.48	assumed low
Q_{total} (W/cm ²)	223	

only one line in memory array charged, the energy for reading eight serial bits is close to 10^{-13} J

$$P_{M_{cycle}} = \frac{E_M}{t_{cycle}} \sim \frac{10^{-13} \text{ J}}{8 \times 10^{-9} \text{ s}} = 1.25 \times 10^{-5} \text{ W.}$$

A summary of energetics of Si- μ cell implemented with ultimate high-performance CMOS and also for low standby power technologies is given in Tables 9 and 10. As follows from the tables, the Si- μ cell cannot operate in this mode due excessive heat generation. Therefore, the cycle time has to be increased to reduce the power dissipation. Note that the predominant source of power consumption in both cases is a consequence of charging memory access lines.

How much heat could be tolerated by a Si- μ cell computer? Table 11 provides some reference numbers for several model heat generators along with heat removal capabilities for different cooling techniques. If it is postulated that only passive cooling can be used for the Si- μ cell

Table 10 Energetics of Si- μ cell Implemented With Ultimate Low-Standby Power CMOS Technology

	Logic	Memory
F_{min} (nm)	8	10
N	320	38,700
E_{bit} (J/bit)	7.41×10^{-18}	$\sim 10^{-15}$ (read)*
E_{cycle} (J)	2.37×10^{-15}	$\sim 10^{-13}$ (read)
t_{cycle} (s)	8×10^{-9}	
P_{active} (W)	2.96×10^{-7}	1.25×10^{-5}
P_{leak} (W)	1.06×10^{-10}	assumed low
P_{total} (W)	1.28×10^{-5}	
E_{total} (J)	3.07×10^{-2}	
Q_{active} (W/cm ²)	4.94	208
Q_{leak} (W/cm ²)	1.76×10^{-3}	assumed low
Q_{total} (W/cm ²)	213	
*serial access assuming a 128×128 array		

(i.e., no additional space overheads), the maximum heat flux through the walls of the cube must be $< 1 \text{ W/cm}^2$ (\sim max. free water convection cooling rate).

If only passive air or water cooling is used, the max heat flux should be $< 1 \text{ W/cm}^2$, thus total power dissipation $< 6 \times 10^{-8} \text{ W}$, which limits the cycle time to be $> 1.70 \mu\text{s}$. For this, the total time needed to emulate the bio- μ cell task (i.e., equivalent of 10^{11} output bits) will be 510 000 s, which is more than $200\times$ larger than time needed for the bio- μ cell.

As follows from the above, the bio- μ cell outperforms the Si- μ cell in all respects. A summary of the comparison between the two μ cells is presented in Section VII-C and some of the implications are discussed.

C. Comparisons and Implications

As is clear from the previous section, the Si- μ cell fundamentally cannot match the bio- μ cell in the density of memory and logic elements, or operational speed, or operational energy. A core challenge is the MTM requirement for a large number of memory accesses per output bit. The above analysis suggests that there is much to be learned from the designs of nature and they may provide hints as to how future technologies could evolve. Fig. 14 provides a brief summary of comparative data. As follows from the analysis, the number of functional elements in the bio- μ cell is extraordinary and far exceeds foreseeable device densities of semiconductors. This may be at least in part due to different mass of information carriers: As was argued in Section III, the smallest size for both memory and logic devices depends on the mass of the information-bearing particles, e.g., the smallest barrier width in Si devices is $\sim 5 \text{ nm}$ due to low effective mass of electrons in Si. Heavier particle mass, in principle, allows for smaller device size, which seems to be realized in the *in carbo* logic and memory elements. Emerging technologies discussed in Section V-B3, such as memristors, atomic switch, and redox memory, also use heavier mass particles, and their potential for very dense circuits needs to be further explored.

As was shown in the previous section, memory access is the most severe limiting factor of Si- μ cell. Not only is there simply not enough nonvolatile memory bits, but also access to them to support computations takes too much energy. In larger scale computers, this problem is easily circumvented by initial massive serial readout from the nonvolatile media (e.g., hard disk drives or flash memory) and buffering these data in low-energy SRAM or DRAM. However, at the scale of the 1- μm cube, there is no space for the buffers, and just direct access to nonvolatile memory was assumed for the Si- μ cell operation. Another related observation is that organizing solid state memory in crossbar arrays, while an elegant solution is at larger scale, also contributes to excessive energy dissipation. In this regard, access to the DNA memory can be viewed as similar to access to hard disc drives [179], [180]. It could be argued that at least in theory, the serial access principle of hard disc drives might be a better solution for

Table 11 Cooling Capabilities for Air and Water and Examples of Representative Heat Generating Systems

	Heat transfer coefficient, W/(cm ² ·K)	Heat flux, W/(cm ²)
<i>Cooling techniques</i>		
Free convection, air	(5-25)×10 ⁻⁴	0.05-0.25 ($\Delta T=100\text{K}$)
Free convection, water	(20-100)×10 ⁻⁴	0.2-1 ($\Delta T=100\text{K}$)
Forced convection, air	(10-500)×10 ⁻⁴	0.1-5 ($\Delta T=100\text{K}$)
Forced convection, water	(100-15,000)×10 ⁻⁴	1-150 ($\Delta T=100\text{K}$)
<i>Reference examples</i>		
Human body	—	0.01
Hot plate	—	10
Microprocessor chip	—	20-60
Sun's surface	—	6000

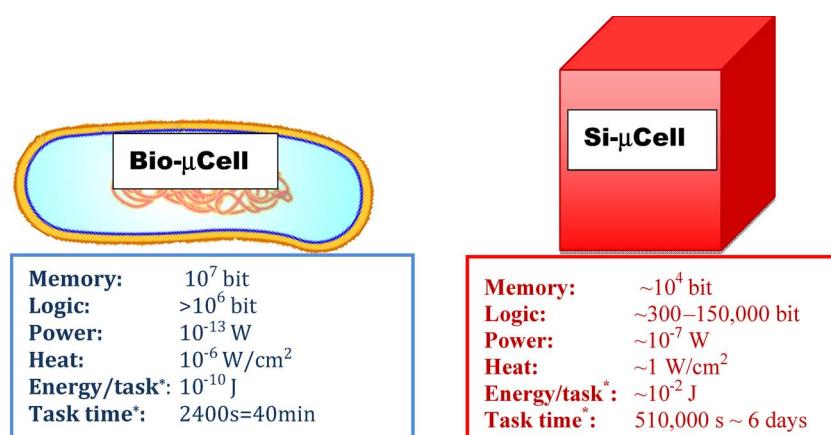
low-energy systems (of course, in practice, the mechanical overheads significantly add to the total energy consumption of hard disk drives).

The architectural organization of computation in *in carbo* systems appears to be much more efficient than for Si computers. As was mentioned in Section II, the biological processors, such as brain, are not on the computational trajectory for Si microprocessors in Fig. 2, suggesting that there may exist alternate technologies and computing architectures offering higher performance (at much lower levels of energy consumption). One key factor here is that basic algorithms need to work in very few steps [181]. Indeed, it appears that the bio- μ cell utilizes fine-grained and massive parallelism per instruction, i.e., by sending out into the cytoplasm multiple copies of DNA instructions by RNA messengers. Also, the bio- μ cell utilizes undirected thermally driven motion of the mRNA molecules to achieve connectivity to the ribosomes in the cytoplasm. A correct transfer to an appropriate ribosome is achieved by electrostatic attraction arising from conformations, i.e., from the specific shapes of the transmitted and recipi-

ent molecular structures. In contrast, data transfer in electrical circuits follows predetermined routes that require an expenditure of energy. Whereas electrical circuits utilize a controllable energy barrier whose operation requires an expenditure of energy and whose physical extent is determined by electron tunneling considerations, it is not clear that there is a similar use of energy barriers in the ribosome's execution of RNA instructions.

As a side remark, although not emphasized in this study, the bio- μ cell incorporates within its volume the capability to incorporate and transform materials from its environment into energy yielding molecules in a form accessible to its processes. Such additional processes for energy transformation were not included in the analyses of the Si- μ cube.

Finally, in 1959, Feynman [182] gave a presentation in which he suggested the possibility of building computers whose dimensions were "submicroscopic." Although the progress of CMOS technology has been extraordinary, submicroscopic computers remain outside our grasp. As has been indicated above, nature appears to have successfully addressed the submicroscopic design challenge.



*Equivalent to 10^{11} output bits

Fig. 14. Comparison of significant parameters of the bio- μ cell and the Si- μ cell.

VIII. SUMMARY

Feature size scaling has enabled a very steep learning curve for CMOS technology that has helped to create a feature-driven marketplace. Although there are compelling physical arguments that physical scaling must end for CMOS, it appears that the benefits of Moore's law will continue for some time, aided by the advent of new materials, processes, and device structures. Very likely, the application space for CMOS technology will continue to grow rapidly as new functionalities are combined with more traditional information processing and communication capabilities.

At the same time, there is intense research underway to find alternatives to CMOS technology that have the potential to extend the benefits of Moore's law scaling for decades into the future. It was pointed out that there are many options at this time, but there is no one-for-one substitute for CMOS technology yet available. Replacement options may eventually be identified, but it appears that a likely scenario is that this research will yield devices with functionalities that can be integrated with CMOS technology to provide unique capabilities or to replace

CMOS modules with special-purpose structures based on the novel devices.

It also may be that dramatic improvements in information processing technologies will result from a radical rethinking of both architectures and supporting technologies. A comparative analysis between the bio- μ cell and the Si- μ cell was offered to stimulate thinking about alternative scenarios. As the bio- μ cell goes about its complex task of creating a copy of itself, it does so using fine-grained processes, devices, and architectures that are completely different and much more energy efficient than existing CMOS/von Neumann paradigms. Perhaps, the design of nature's information processors can inspire radical breakthroughs in inorganic information processing.

There is substantial momentum to sustain Moore's law for many more decades because of the benefits that it accrues to society. The challenges that lie before us to achieve this are substantial but so is the creativity of scientists and engineers. Although the road ahead is not well marked, there are many indications that there are no insurmountable barriers that would deny progress in information processing technologies for the foreseeable future. ■

REFERENCES

- [1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114–117, 1965.
- [2] S. P. Murarka and S. W. Hymes, "Copper metallization for ULSI and beyond," *Critical Rev. Solid State Mater. Sci.*, vol. 20, pp. 87–124, 1995.
- [3] R. H. Havemann and J. A. Hutchby, "High-performance interconnects: An integration overview," *Proc. IEEE*, vol. 89, no. 5, pp. 586–601, May 2001.
- [4] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-k gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, vol. 89, 5243, 2001.
- [5] C. H. Jan, P. Bai, S. Biswas, M. Buehler, Z. P. Chen, G. Curello, S. Gannavaram, W. Hafez, J. He, J. Hicks, U. Jalan, N. Lazo, J. Lin, N. Lindert, C. Litteken, M. Jones, M. Kang, K. Komeyli, A. Mezhiba, S. Naskar, S. Olson, J. Park, R. Parker, L. Pei, I. Post, N. Pradhan, C. Prasad, M. Prince, J. Rizk, G. Sacks, H. Tashiro, D. Towner, C. Tsai, Y. Wang, L. Yang, J. Y. Yeh, J. Yip, and K. Mistry, "A 45 nm low power system-on-chip technology with dual gate (logic and I/O) high-k/metal gate strained silicon transistors," in *Proc. Int. Electron Devices Meeting*, 2008, pp. 637–640.
- [6] S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-process-induced strained-Si: Extending the CMOS roadmap," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1010–1020, May 2006.
- [7] M. Mayberry, "Progress towards the merger of compound semiconductors and silicon," *Solid State Technol.*, vol. 52, pp. 16–17, 2009.
- [8] K. J. Kuhn, "Moore's crystal ball: Device physics and technology past the 15 nm generation," *Microelectron. Eng.*, vol. 88, pp. 1044–1049, 2011.
- [9] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Baeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J. P. Schoellkopf, E. Pereira, R. Ferrant, and H. Mingam, "Innovative materials, devices, and CMOS technologies for low-power mobile multimedia," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 96–130, Jan. 2008.
- [10] W. Gitt, "Information—The 3rd fundamental quantity," *Siemens Rev.*, vol. 56, pp. 36–41, 1989.
- [11] H. Moravec, "When will computer hardware match the human brain?" *J. Evol. Technol.*, vol. 1, pp. 1–12, 1998.
- [12] W. Rhynes, "Keynote: 'Less of Moore,'" presented at the TECHCON, Austin, TX, Sep. 14, 2009.
- [13] D. J. C. Herr, "Directed block copolymer self-assembly for nanoelectronics fabrication," *J. Mater. Res.*, vol. 26, pp. 122–139, 2011.
- [14] C. Bencher, J. Smith, L. Y. Miao, C. Cai, Y. M. Chen, J. Y. Cheng, D. P. Sanders, M. Tjio, H. D. Truong, S. Holmes, and W. D. Hinsberg, "Self-assembly patterning for sub-15 nm half-pitch: A transition from lab to fab," *Proc. SPIE*, vol. 7970, 79700F, 2011.
- [15] V. V. Zhirnov, R. K. Cavin, J. A. Hutchby, and G. I. Bourianoff, "Limits to binary logic switch scaling—A Gedanken model," *Proc. IEEE*, vol. 91, no. 11, pp. 1934–1939, Nov. 2003.
- [16] T. N. Theis and P. M. Solomon, "In quest of the 'Next Switch': Prospects for greatly reduced power dissipation in a successor to the silicon field-effect transistor," *Proc. IEEE*, vol. 98, no. 12, pp. 2005–2014, Dec. 2010.
- [17] K. Bernstein, R. K. Cavin, W. Porod, A. Seabaugh, and J. Welser, "Device and architecture outlook for beyond CMOS switches," *Proc. IEEE*, vol. 98, no. 12, pp. 2169–2184, Dec. 2010.
- [18] M. Mayberry, "Keynote: 'A look forward,'" presented at the TECHCON, Austin, TX, Sep. 13, 2010.
- [19] T. Skotnicki and F. Baeuf, "Optimal scaling methodologies and transistor performance," in *High-K Gate Dielectric Materials for VLSI MOSFET Applications*, H. Huff and D. Gilmer, Eds. New York: Springer-Verlag, 2005, pp. 143–194.
- [20] S. Shankar, V. Zhirnov, and R. Cavin, "Computation from devices to system level thermodynamics," *ECS Trans.*, vol. 25, no. 7, pp. 421–431, 2009.
- [21] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Documaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, "Silicon CMOS devices beyond scaling," *IBM J. Res. Develop.*, vol. 59, pp. 339–361, 2006.
- [22] International Technology Roadmap for Semiconductors (ITRS). [Online]. Available: www.itrs.net
- [23] V. Zhirnov and T. Mikolajick, "Chapter 26: Flash memories," in *Nanoelectronics and Information Technology*, R. Waser, Ed. New York: Wiley, 2012.
- [24] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, no. 4, pp. 619–635, Apr. 1995.
- [25] R. Cavin, W. Joyner, and T. Noll, "Chapter 22: Performance estimates for microprocessors: At technology limits and in practice," in *Nanoelectronics and Information Technology*, R. Waser, Ed. New York: Wiley, 2012.
- [26] G. Chandra, P. Kapur, and K. C. Saraswat, "Scaling trends for the on chip power dissipation," in *Proc. 5th Int. Interconnect Technol. Conf.*, Burlingame, CA, Jun. 3–5, 2002, pp. 170–172.
- [27] V. V. Zhirnov, R. K. Cavin, S. Menzel, E. Linn, S. Schmelzer, D. Bräuhäus, C. Schindler, and R. Waser, "Memory devices: Energy-space-time tradeoffs," *Proc. IEEE*, vol. 98, no. 12, pp. 2185–2200, Dec. 2010.

- [28] L. M. Grupp, A. M. Caulfield, J. Coburn, S. Swanson, E. Yaakobi, P. H. Siegel, and J. K. Wolf, "Characterizing flash memory: Anomalies, observations, and applications," in Proc. *MICRO*, New York, Dec. 12–16, 2009, pp. 24–33.
- [29] N. Derhacobian, S. C. Hollmer, N. Gilbert, and M. N. Kozicki, "Power and energy perspectives of nonvolatile memory technologies," *Proc. IEEE*, vol. 98, no. 2, pp. 283–298, Feb. 2010.
- [30] W. Arden, M. Brillouët, P. Cogez, M. Graef, B. Huizing, and R. Mahnkopf, *More-than-Moore*, White Paper. [Online]. Available: <http://www.itrs.net/Links/2010ITRS-IRC-ITRS-MtM-v2%203.pdf>
- [31] C. B. Les, "Image sensor market: Changing, but moving upward," *Photon. Spectra*, vol. 43, pp. 27–28, 2009.
- [32] A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuit Device Mag.*, vol. 21, no. 3, pp. 6–20, May–Jun. 2005.
- [33] G. K. Fedder, R. T. Howe, T. K. Liu, and E. P. Quevy, "Technologies for cofabricating MEMS and electronics," *Proc. IEEE*, vol. 96, no. 2, pp. 306–322, Feb. 2008.
- [34] C. Liao and J. Tsai, "The evolution of MEMS display," *IEEE Trans. Ind. Electron.*, vol. 56, no. 4, pp. 1057–1065, Apr. 2009.
- [35] EU Commision, *Technology Roadmap for Nanoelectronics*, 2001. [Online]. Available: <ftp://ftp.cordis.europa.eu/pub/esprit/docs/melmanr.pdf>
- [36] A. Hussain, J. Hone, H. W. Postma, X. M. H. Huang, T. Drake, M. Narbic, A. Scherer, and M. L. Roukes, "Nanowire-based very-high-frequency electromechanical resonator," *Appl. Phys. Lett.*, vol. 83, no. 6, 1240, 2003.
- [37] G. M. Rebeiz, *RF MEMS: Theory, Design and Technology*. New York: Wiley, 2003.
- [38] X. L. Feng, R. He, P. Yang, and M. L. Roukes, "Very high frequency silicon nanowire electromechanical resonators," *Nano Lett.*, vol. 7, no. 7, pp. 1953–1959, 2007.
- [39] V. Sazonova, Y. Yaish, H. Ustunel, D. Roundy, T. A. Arlas, and P. L. McEuen, "A tunable carbon nanotube electromechanical oscillator," *Nature*, vol. 431, pp. 284–287, 2004.
- [40] H. B. Peng, C. W. Chang, S. Aloni, T. D. Yuzvinsky, and A. Zettl, "Ultrahigh frequency nanotube resonators," *Phys. Rev. Lett.*, vol. 97, 087203, 2006.
- [41] J. S. Bunch, A. M. van der Zande, S. S. Verbridge, I. W. Frank, D. M. Tanenbaum, J. M. Parpia, H. G. Craighead, and P. L. McEuen, "Electromechanical resonators from graphene sheets," *Science*, vol. 315, pp. 490–493, 2007.
- [42] C. Durand, F. Casset, P. Renaux, N. Abele, B. Legrand, D. Renaud, E. Ollier, P. Ancey, A. M. Ionescu, and L. Buchaillot, "In-plane silicon-on-nothing nanometer-scale resonant suspended gate MOSFET for In-IC integration perspectives," *IEEE Electron Device Lett.*, vol. 29, no. 5, pp. 494–496, May 2008.
- [43] J. A. Katine and E. E. Fullerton, "Device implications of spin-transfer torques," *J. Magn. Magn. Mater.*, vol. 320, pp. 1217–1226, 2008.
- [44] P. Villard, U. Ebels, D. Houssameddine, J. Katine, D. Mauri, B. Delaet, P. Vincent, M.-C. Cyrille, B. Viala, J.-P. Michel, J. Prouvée, and F. Badets, "A GHz spintronic-based RF oscillator," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 214–223, Jan. 2010.
- [45] I. Magrini and A. C. G. Manes, "A low local oscillator power K-band mixer based on tunneling diodes," *Microw. Opt. Technol. Lett.*, vol. 51, no. 4, pp. 1140–1143, 2009.
- [46] R. Knobel, C. S. Yung, and A. N. Cleland, "Single-electron transistor as a radio-frequency mixer," *Appl. Phys. Lett.*, vol. 81, no. 3, pp. 532–534, 2002.
- [47] H. Wang, A. Hsu, J. Wu, K. Jing, and T. Palacios, "Graphene-based ambipolar RF mixers," *IEEE Electron Device Lett.*, vol. 31, no. 9, pp. 906–908, Sep. 2010.
- [48] C. Rutherford and P. Burke, "Carbon nanotube radio," *Nano Lett.*, vol. 7, no. 11, pp. 3296–3299, 2007.
- [49] N. Rouhi, D. Jain, and P. J. Burke, "Nanoscale devices for large-scale applications," *IEEE Microw. Mag.*, vol. 11, no. 7, pp. 72–80, Dec. 2010.
- [50] E. Cantatore, T. C. T. Geuns, G. H. Gelinck, E. van Veenendaal, A. F. A. Gruijthuijsen, L. Schrijnemakers, S. Drews, and D. M. de Leeuw, "A 13.56-MHz RFID system based on organic transponders," *IEEE J. Solid-State Circuit*, vol. 42, no. 1, pp. 94–92, Jan. 2007.
- [51] J. Lewis, J. Burroughes, Y. Ohmori, and K. S. Narayan, "Organic electronics," *Proc. IEEE*, vol. 97, no. 9, pp. 1555–1557, Sep. 2009.
- [52] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, *Nature Photon.*, vol. 4, pp. 518–526, 2010.
- [53] J. Leuthold, C. Koos, and W. Freude, "Nonlinear silicon photonics," *Nature Photon.*, vol. 4, pp. 535–544, 2010.
- [54] D. Liang and J. E. Bowers, "Recent progress in lasers on silicon," *Nature Photon.*, vol. 4, pp. 511–517, 2010.
- [55] H. Rong, S. Xu, Y. Kuo, V. Sih, O. Cohen, O. Raday, and M. Paniccia, "Low-threshold continuous-wave Raman silicon laser," *Nature Photon.*, vol. 1, pp. 232–237, 2007.
- [56] C. M. Soukoulis and M. Wegener, "Past achievements and future challenges in the development of three-dimensional photonic metamaterials," *Nature Photon.*, vol. 5, pp. 523–530, 2011.
- [57] M. Lipson, "Guiding, modulating, and emitting light on Silicon-challenges and opportunities," *J. Lightw. Technol.*, vol. 23, no. 12, pp. 4222–4238, Dec. 2005.
- [58] J. A. Dionne, L. A. Sweatlock, M. T. Sheldon, A. P. Alivisatos, and H. A. Atwater, "Silicon-based plasmonics for on-chip photonics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 295–306, Jan./Feb. 2010.
- [59] G. Konstantatos and E. H. Sargent, "Nanostructured materials for photon detection," *Nature Photon.*, vol. 5, pp. 391–400, 2010.
- [60] A. Rogalski, "Infrared detectors: Status and trends," *Progr. Quantum Electron.*, vol. 27, no. 2–3, pp. 59–210, 2003.
- [61] S. F. Tedde, J. Kern, T. Sterzl, J. Fürst, P. Lugli, and O. Hayden, "Fully spray coated organic photodiodes," *Nano Lett.*, vol. 9, no. 3, pp. 980–983, 2009.
- [62] A. Abdellah, B. Fabel, P. Lugli, and G. Scarpa, "Spray deposition of organic semiconducting thin-films: Towards the fabrication of arbitrary shaped organic electronic devices," *Organic Electron.*, vol. 11, no. 6, pp. 1031–1038, 2010.
- [63] J. H. Burroughes, D. D. C. Bradley, A. R. Brown, R. N. Marks, K. Mackay, R. H. Friend, R. L. Burn, and A. B. Holmes, "Light-emitting diodes based on conjugated polymers," *Nature*, vol. 347, pp. 539–541, 1990.
- [64] A. C. Arias, J. D. MacKenzie, I. McCulloch, J. Rivnay, and A. Salleo, "Materials and applications for large area electronics: Solution-based approaches," *Chem. Rev.*, vol. 110, no. 1, pp. 3–24, 2010.
- [65] S. Günes, H. Neugebauer, and N. S. Sariciftci, "Conjugated polymer-based organic solar cells," *Chem. Rev.*, vol. 107, no. 4, pp. 1324–1338, 2007.
- [66] D. Baierl, B. Fabel, P. Lugli, and G. Scarpa, "Efficient indium-tin-oxide (ITO) free top-absorbing organic photodetector with highly transparent polymer top electrode," *Organic Electron.*, vol. 12, no. 10, pp. 1669–1673, 2011.
- [67] T. Agostinelli, M. Caiaroni, D. Natali, M. Sampietro, G. Dassa, E. V. Canesi, C. Bertarelli, G. Zerbini, J. Cabanillas-Gonzalez, S. De Silvestri, and G. Lanzani, "A planar organic near infrared light detector based on bulk heterojunction of a heteroquaterphenoquinone and poly[2-methoxy-5-(2'-ethyl-hexyloxy)-1,4-phenylene vinylene]," *J. Appl. Phys.*, vol. 104, no. 11, pp. 114 508–114 513, 2008.
- [68] T. Rauch, M. Böberl, S. F. Tedde, J. Fürst, M. V. Kovalenko, G. Hesser, U. Lemmer, W. Heiss, and O. Hayden, "Near-infrared imaging with quantum-dot-sensitized organic photodiodes," *Nature Photon.*, vol. 3, pp. 332–336, 2009.
- [69] X.-L. Huang and Y.-K., "Chemical sensors based on nanostructured materials," *Sens. Actuators B*, vol. 122, pp. 659–671, 2007.
- [70] B. L. Allen, P. D. Kichambarre, and A. Star, "Carbon nanotube field-effect-transistor-based biosensors," *Adv. Mater.*, vol. 19, pp. 1439–1451, 2007.
- [71] M. Li, H. X. Tang, and M. L. Roukes, "Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications," *Nature Nanotechnol.*, vol. 2, pp. 114–120, 2007.
- [72] N. M. Freris, H. Kowshik, and P. R. Kumar, "Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation," *Proc. IEEE*, vol. 98, no. 11, pp. 1828–1846, Nov. 2010.
- [73] G. Zheng, X. Gao, and C. M. Lieber, "Frequency domain detection of biomolecules using silicon nanowire biosensors," *Nano Lett.*, vol. 10, pp. 3179–3183, 2010.
- [74] I. Heller, W. T. T. Small, S. G. Lemay, and C. Dekker, "Probing macrophage activity with carbon-nanotube sensors," *Small*, vol. 5, pp. 2528–2532, 2009.
- [75] J. A. Rogers, T. Someya, and Y. Huang, "Materials and mechanics for stretchable electronics," *Science*, vol. 327, pp. 1603–1607, 2010.
- [76] L. Hu, D. S. Hecht, and G. Grüner, "Carbon nanotube thin films: Fabrication, properties, and applications," *Chem. Rev.*, vol. 110, no. 10, pp. 5790–5844, 2010.
- [77] Q. Cao and J. A. Rogers, "Ultrathin films of single-walled carbon nanotubes for electronics and sensors: A review of fundamental and applied aspects," *Adv. Mater.*, vol. 20, pp. 29–53, 2008.
- [78] N. Rouhi, D. Jain, and P. J. Burke, "High-performance semiconducting nanotube inks: Progress and prospects,"

- ACS Nano, vol. 5, no. 11, pp. 8471–8487, 2011.
- [79] S. M. Goetz, C. M. Erlen, H. Grothe, W. Wolf, P. Lugli, and G. Scarpa, “Organic field-effect transistors for biosensing applications,” *Organic Electron.*, vol. 10, no. 4, pp. 573–580, 2009.
- [80] J. A. Paradiso and T. Starner, “Energy scavenging for mobile and wireless electronics,” *Perv. Comput.*, vol. 4, pp. 18–27, 2005.
- [81] J. Olivo, S. Carrara, and G. De Micheli, “Energy harvesting and remote powering for implantable biosensors,” *IEEE Sensors J.*, vol. 11, no. 7, pp. 1573–1586, Jul. 2011.
- [82] C. Deibel, V. Dyakonov, and C. J. Brabec, “Organic bulk-heterojunction solar cells,” *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 6, pp. 1517–1527, Nov./Dec. 2010.
- [83] M. D. Stoller, S. Park, Y. Zhu, J. An, and R. S. Ruoff, “Graphene-based ultracapacitors,” *Nano Lett.*, vol. 8, no. 10, pp. 3498–3502, 2008.
- [84] A. I. Hochbaum and P. Yang, “Semiconductor nanowires for energy conversion,” *Chem. Rev.*, vol. 110, no. 1, pp. 527–546, 2010.
- [85] T. Ashley, M. T. Emeny, D. G. Hayes, K. P. Hilton, R. Jeffries, J. O. Maclean, S. J. Smith, A. W.-H. Tang, D. J. Wallis, and P. J. Webber, “High-performance InSb based quantum well field effect transistors for low-power dissipation applications,” in *IEEE Int. Electron Devices Meeting Tech. Dig.*, 2009, DOI: 10.1109/IEDM.2009.5424207.
- [86] L. Xia, J. B. Boos, B. R. Bennett, M. G. Ancona, and J. A. del Alamo, “Hole mobility enhancement in InGaSb quantum-well field-effect transistors,” *Appl. Phys. Lett.*, vol. 98, 053505, 2011.
- [87] S. Hertenberger, D. Rudolph, M. Bichler, J. Finley, G. Abstreiter, and G. Koblmüller, “Growth kinetics in position-controlled and catalyst-free InAs nanowire arrays on Si (111) grown by selective area molecular beam epitaxy,” *J. Appl. Phys.*, vol. 108, 2010, DOI: 10.1063/1.3525610.
- [88] D. Kuzum, T. Krishnamohan, A. Nainani, Y. Sun, P. A. Panetta, H.-S. P. Wong, and K. C. Saraswat, “High-mobility Ge N-MOSFETs and mobility degradation mechanisms,” *IEEE Trans. Electron Devices*, vol. 59, no. 1, pp. 59–66, Jan. 2011.
- [89] H. Yan and P. Yang, “Semiconductor nanowires: Functional building blocks for nanotechnology,” in *The Chemistry of Nanostructured Materials*, P. Yang, Ed. River Edge, NJ: World Scientific, 2004.
- [90] J. Xiang, W. Lu, Y. Hu, Y. Wu, H. Yan, and C. M. Lieber, “Ge/Si nanowire heterostructures as high-performance field-effect transistors,” *Nature*, vol. 441, pp. 489–493, 2006.
- [91] W. Lu and C. M. Lieber, “Nanoelectronics from the bottom up,” *Nature Mater.*, vol. 6, pp. 841–850, Nov. 2007.
- [92] W. M. Weber, L. Geelhaar, A. P. Graham, E. Unger, G. S. Duesberg, M. Liebau, W. Pamler, C. Chéze, H. Riechert, P. Lugli, and F. Kreupl, “Silicon-nanowire transistors with intruded nickel-silicide contacts,” *Nano Lett.*, vol. 6, pp. 2660–2666, 2006.
- [93] L.-E. Wernersson, C. Thelander, E. Lind, and L. Samuelson, “III-V nanowires—Extending a narrowing road,” *Proc. IEEE*, vol. 98, no. 12, pp. 2047–2060, Dec. 2010.
- [94] H. T. Ng, J. Han, T. Yamada, P. Nguyen, Y. P. Chen, and M. Meyyappan, “Single crystal nanowire vertical surround-gate field-effect transistor,” *Nano Lett.*, vol. 4, no. 7, pp. 1247–1252, 2004.
- [95] W. Lu, P. Xie, and C. M. Lieber, “Nanowire transistor performance limits and applications,” *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 2859–2876, Nov. 2008.
- [96] B. A. Sheriff, D. Wang, J. R. Heath, and J. N. Kurtin, “Complementary symmetry nanowire logic circuits: Experimental demonstrations and in silico optimizations,” *ACS Nano*, vol. 2, no. 9, pp. 1789–1798D, 2008.
- [97] Q. Zhang, W. Zhao, and A. Seabaugh, “Low-subthreshold-swing tunnel transistors,” *IEEE Electron Device Lett.*, vol. 27, no. 4, pp. 297–300, Apr. 2006.
- [98] S. O. Koswatta, M. S. Lundstrom, and D. E. Nikonorov, “Performance comparison between p-i-n tunneling transistors and conventional MOSFETs,” *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 456–465, Mar. 2009.
- [99] O. M. Nayfeh, C. N. Chleirigh, J. Hennessy, L. Gomez, J. L. Hoyt, and D. A. Antoniadis, “Design of tunneling field-effect transistors using strained-silicon/strained-germanium type-II staggered heterojunctions,” *IEEE Electron Device Lett.*, vol. 29, no. 9, pp. 1074–1077, Sep. 2008.
- [100] K. K. Bhuwalka, J. Schulze, and I. Eisele, “Performance enhancement of vertical tunnel field-effect transistor with SiGe in the $\delta p+$ layer,” *Jpn. J. Appl. Phys.*, vol. 43, no. 7A, pp. 4073–4078, Jul. 2004.
- [101] K. Boucart, W. Riess, and A. M. Ionescu, “Lateral strain profile as key technology booster for all-silicon tunnel FETs,” *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 656–658, Jun. 2009.
- [102] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, “Band-to-band tunneling in carbon nanotube field-effect transistors,” *Phys. Rev. Lett.*, vol. 93, no. 19, pp. 196805-1–196805-4, 2004.
- [103] A. C. Seabaugh and Q. Zhang, “Low-voltage tunnel transistors for beyond CMOS logic,” *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, Dec. 2010.
- [104] K. Akarvardar, D. Elata, R. Parsa, G. C. Wan, K. Yoo, J. Provine, P. Peumans, R. T. Howe, and H.-S. P. Wong, “Design considerations for complementary nanoelectromechanical logic gates,” in *IEEE Int. Electron Devices Meeting Tech. Dig.*, 2007, pp. 299–302.
- [105] W. W. Jang, J. O. Lee, J.-B. Yoon, M.-S. Kim, J.-M. Lee, S.-M. Kim, K.-H. Cho, D.-W. Kim, D. Park, and W.-S. Lee, “Fabrication and characterization of a nanoelectromechanical switch with 15-nm-thick suspension air gap,” *Appl. Phys. Lett.*, vol. 92, no. 10, 2008, DOI: 10.1063/1.2892659.
- [106] S. Fujita, K. Nomura, K. Abe, and T. H. Lee, “3-D nanoarchitectures with carbon nanotube mechanical switches for future on-chip network beyond CMOS architecture,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 11, pp. 2472–2479, Nov. 2007.
- [107] Li, S.-M. Koo, M. D. Edelstein, J. S. Suehle, and C. A. Richter, “Silicon nanowire electromechanical switches for logic device application,” *Nanotechnology*, vol. 18, 315202, 2007.
- [108] L. Ding, A. Tselev, J. Y. Wang, D. N. Yuan, H. B. Chu, T. P. McNicholas, Y. Li, and J. Liu, “Selective growth of well-aligned semiconducting single-walled carbon nanotubes,” *Nano Lett.*, vol. 9, pp. 800–805, 2009.
- [109] Z. X. Wang, H. L. Xu, Z. Y. Zhang, S. Wang, L. Ding, Q. S. Zeng, L. J. Yang, T. A. Pei, X. L. Liang, M. Gao, and L. M. Peng, “Growth and performance of yttrium oxide as an ideal high-kappa gate dielectric for carbon-based electronic,” *Nano Lett.*, vol. 10, pp. 2024–2030, 2010.
- [110] A. D. Franklin and Z. H. Chen, “Length scaling of carbon nanotube transistors,” *Nature Nanotechnol.*, vol. 5, pp. 858–862, 2010.
- [111] L. Nougaret, H. Happy, G. Dambrine, V. Derycke, J. P. Bourgoain, A. A. Green, and M. C. Hersam, “80 GHz field-effect transistors produced using high purity semiconducting single-walled carbon nanotubes,” *Appl. Phys. Lett.*, vol. 94, 243505, 2009.
- [112] L. Liao, Y.-C. Lin, M. Bao, R. Cheng, J. Bai, Y. Liu, Y. Qu, K. L. Wang, Y. Huang, and X. Duan, “High-speed graphene transistors with a self-aligned nanowire gate,” *Nature*, vol. 467, pp. 305–308, 2010.
- [113] D. Waldmann, J. Jobst, F. Speck, T. Seyller, M. Krieger, and H. B. Weber, “Bottom-gated epitaxial graphene,” *Nature Mater.*, vol. 10, pp. 357–360, 2011.
- [114] I. Meric, M. Y. Han, A. F. Young, B. Ozilmez, P. Kim, and K. L. Shepard, “Current saturation in zero-bandgap, top-gated graphene field-effect transistors,” *Nature Nanotechnol.*, vol. 3, pp. 654–659, 2008.
- [115] X. S. Li, W. W. Cai, J. H. An, S. Kim, J. Nah, D. X. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc, S. K. Banerjee, L. Colombo, and R. S. Ruoff, “Large-area synthesis of high-quality and uniform graphene films on copper foils science,” *Science*, vol. 324, pp. 1312–1314, 2009.
- [116] M. C. Lemme, T. J. Echtermeyer, M. Baus, and H. Kurz, “A graphene field-effect device,” *IEEE Electron Device Lett.*, vol. 28, no. 4, pp. 282–284, Apr. 2007.
- [117] J. Kedzierski, P.-L. Hsu, P. Healey, P. W. Wyatt, C. L. Keast, M. Sprinkle, C. Berger, and W. A. de Heer, “Epitaxial graphene transistors on SiC substrates,” *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2078–2085, Aug. 2008.
- [118] S. Bae, H. F. Kim, Y. Lee, X. F. Xu, J.-S. Park, Y. Zheng, J. Balakrishnan, T. Lei, H. R. Kim, Y. I. Song, Y.-J. Kim, K. S. Kim, B. Özilmez, J.-H. Ahn, B. H. Hong, and S. Iijima, *Nature Nanotechnol.*, vol. 5, pp. 574–578, 2010.
- [119] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, “Electric field effect in atomically thin carbon films,” *Science*, vol. 306, pp. 666–669, 2004.
- [120] X. Du, I. Skachko, A. Barker, and E. Y. Andrei, “Approaching ballistic transport in suspended graphene,” *Nature Nanotechnol.*, vol. 3, pp. 491–495, 2008.
- [121] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, and J. Hone, “Boron nitride substrates for high-quality graphene electronics,” *Nature Nanotechnol.*, vol. 5, pp. 722–726, 2010.
- [122] X. S. Li, C. W. Magnuson, A. Venugopal, J. H. An, J. W. Suk, B. Y. Han, M. Borysiak, W. W. Cai, A. Velamakanni, Y. W. Zhu, L. F. Fu, E. M. Vogel, E. Voelkl, L. Colombo, and R. S. Ruoff, “Graphene films with large domain size by a two-step chemical vapor deposition process,” *Nano Lett.*, vol. 10, pp. 4328–4334, 2010.

- [123] X. L. Li, X. R. Wang, L. Zhang, S. W. Lee, and H. J. Dai, "Chemically derived, ultrasmooth graphene nanoribbon semiconductors," *Science*, vol. 319, pp. 1229–1232, 2008.
- [124] L. O. Chua, "Memristor—The missing circuit element," *IEEE Trans. Circuit Theory*, vol. CT-18, no. 5, pp. 507–519, Sep. 1971.
- [125] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, pp. 80–83, 2008.
- [126] C. Scott and L. D. Bozano, "Nonvolatile memory elements based on organic materials," *Adv. Mater.*, vol. 19, pp. 1452–1463, 2007.
- [127] Heremans, G. H. Gelinck, R. Müller, K.-J. Baeg, D.-J. Kim, and Y.-Y. Noh, "Polymer and organic nonvolatile memory devices," *Chem. Mater.*, vol. 23, pp. 341–358, 2011.
- [128] S. Song, B. Cho, T.-W. Kim, Y. Ji, M. Jo, G. Wang, M. Choe, Y. H. Kahng, H. Hwang, and T. Lee, "Three-dimensional integration of organic resistive memory devices," *Adv. Mater.*, vol. 22, pp. 5048–5052, 2010.
- [129] T. Lee, L.-Z. Yu, and H.-C. Chen, "Memory bistable mechanisms of organic memory devices," *Appl. Phys. Lett.*, vol. 97, 043301, 2010.
- [130] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Mater.*, vol. 6, pp. 833–840, 2007.
- [131] T. Sakamoto, H. Sunamura, H. Kawaura, T. Hasegawa, T. Nakayama, and M. Aono, "Nanometer-scale switches using copper sulfide," *Appl. Phys. Lett.*, vol. 82, pp. 3032–3034, 2003.
- [132] S. Kaeriyama, T. Sakamoto, H. Sunamura, M. Mizuno, H. Kawaura, T. Hasegawa, K. Terabe, T. Nakayama, and M. Aono, "A nonvolatile programmable solid-electrolyte nanometer switch," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 168–176, Jan. 2005.
- [133] T. Sakamoto, N. Iguchi, and M. Aono, "Nonvolatile triode switch using electrochemical reaction in copper sulfide," *Appl. Phys. Lett.*, vol. 96, 2010, DOI: 10.1063/1.3457861.
- [134] I. Valov, R. Waser, J. R. Jameson, and M. N. Kozicki, "Electrochemical metalization memories—fundamentals, applications, prospects," *Nanotechnology*, vol. 22, 254003, 2011.
- [135] R. Waser, R. Dittman, G. Staikov, and K. Szot, "Redox-based resistive switching memories—Nanoionics mechanisms, prospects, and challenges," *Adv. Mater.*, vol. 21, pp. 2632–2663, 2009.
- [136] H. Akinaga and H. Shima, "Resistive random access memory (ReRAM) based on metal oxides," *Proc. IEEE*, vol. 98, no. 12, pp. 2237–3225, Dec. 2010.
- [137] M.-J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U.-I. Chung, I.-K. Yoo, and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures," *Nature Mater.*, vol. 10, pp. 625–630, 2011.
- [138] R. Rosezin, E. Linn, L. Nielen, C. Kugeler, R. Bruchhaus, and R. Waser, "Integrated complementary resistive switches for passive high-density nanocross arrays," *Electron Device Lett.*, vol. 32, no. 2, pp. 191–193, Feb. 2011.
- [139] M. J. Lee, Y. Park, D.-S. Suh, E.-H. Lee, S. Seo, D.-C. Kim, R. Jung, B.-S. Kang, S.-E. Ahn, C. B. Lee, D. H. Seo, Y.-K. Cha, I.-K. Yoo, J.-S. Kim, and B. H. Park, "Two series oxide resistors applicable to high speed and high density nonvolatile memory," *Adv. Mater.*, vol. 19, pp. 3919–3923, 2007.
- [140] V. V. Zhirnov, R. Meade, R. K. Cavin, and G. Sandhu, "Scaling limits of resistive memories," *Nanotechnology*, vol. 22, 254027, 2011.
- [141] Q. Xia, W. Robinett, M. W. Cumbie, N. Banerjee, T. J. Cardinali, J. J. Yang, W. Wu, X. Li, W. M. Tong, D. B. Strukov, G. S. Snider, G. Medeiros-Ribeiro, and R. S. Williams, "Memristor-CMOS hybrid integrated circuits for reconfigurable logic," *Nano Lett.*, vol. 9, no. 10, pp. 3640–3645, 2009.
- [142] D. Kuzum, R. G. D. Jeyasingh, B. Lee, and H.-S. P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," *Nano Lett.*, 2011, DOI: 10.1021/nl101040y.
- [143] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.*, vol. 10, no. 4, pp. 1297–1301, 2010.
- [144] H. Song, M. A. Reed, and T. Lee, "Single molecule electronic devices," *Adv. Mater.*, vol. 23, pp. 1583–1608, 2011.
- [145] M. A. Reed, J. Chen, A. M. Rawlett, D. W. Price, and J. M. Tour, "Molecular random access memory cell," *Appl. Phys. Lett.*, vol. 78, no. 23, pp. 3735–3737, 2001.
- [146] M. Tour, L. Cheng, D. P. Nackashi, Y. X. Yao, A. K. Flatt, S. K. St Angelo, T. E. Mallouk, and P. D. Franzon, "NanoCell electronic memories," *J. Amer. Chem. Soc.*, vol. 125, pp. 13 279–13 283, 2003.
- [147] C. N. Lau, D. R. Stewart, R. S. Williams, and M. Bockrath, "Direct observation of nanoscale switching centers in metal/molecule/metal structures," *Nano Lett.*, vol. 4, pp. 569–657, 2004.
- [148] J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin, E. Delonno, Y. Luo, B. A. Sheriff, K. Xu, Y. Shik Shin, H.-R. Tseng, J. F. Stoddart, and J. R. Heath, "A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimetre," *Nature*, vol. 445, pp. 414–441, 2007.
- [149] T. Pro, J. Buckley, K. Huang, A. Calborean, M. Gely, G. Delapierre, G. Ghibaudo, F. Duclairoir, J.-C. Marchon, E. Jalaguier, P. Maldivi, B. De Salvo, and S. Deleonibus, "Investigation of hybrid molecular/silicon memories with redox-active molecules acting as storage media," *IEEE Trans. Nanotechnol.*, vol. 8, no. 2, pp. 204–212, Mar. 2009.
- [150] E. Lörtscher, J. W. Ciszek, J. Tour, and H. Riel, "Reversible and controllable switching of a single-molecule junction," *Small*, vol. 2, no. 8–9, pp. 973–977, 2006.
- [151] G. Csaba and P. Lugli, "Read-out design rules for molecular crossbar architectures," *IEEE Trans. Nanotechnol.*, vol. 8, no. 3, pp. 369–374, May 2009.
- [152] C. S. Lent, B. Isaksen, and M. Lieberman, "Molecular quantum-dot cellular automata," *J. Amer. Chem. Soc.*, vol. 125, no. 4, pp. 1056–1063, 2003.
- [153] J. Akerman, "Towards a universal memory," *Science*, vol. 308, pp. 508–510, 1721.
- [154] A. Orlov, A. Imre, G. Csaba, L. Ji, W. Porod, and G. H. Bernstein, "Magnetic quantum-dot cellular automata: Recent developments and prospects," *J. Nanoelectron. Optoelectron.*, vol. 3, pp. 55–68, 2008.
- [155] S. Sugahara and J. Nitta, "Spin-transistor electronics: An overview and outlook," *Proc. IEEE*, vol. 98, no. 12, pp. 2124–2154, Dec. 2010.
- [156] S. P. Dash, S. Sharma, R. S. Patel, M. P. de Jong, and R. Jansen, "Electrical creation of spin polarization in silicon at room temperature," *Nature*, vol. 462, pp. 491–494, 2009.
- [157] C. H. Li, O. M. J. van't Erve, and B. T. Jonker, "Electrical injection and detection of spin accumulation in silicon at 500 K with magnetic metal/silicon dioxide contacts," *Nature Commun.*, vol. 2, 2011, article 245.
- [158] H. C. Koo, J. H. Kwon, J. Eom, J. Chang, S. H. Han, and M. Johnson, "Control of spin precession in a spin-injected field effect transistor," *Science*, vol. 325, pp. 1515–1518, Sep. 2009.
- [159] A. Khiton, D. E. Nikonorov, and K. L. Wang, "Magnetolectric spin wave amplifier for spin wave logic circuits," *J. Appl. Phys.*, vol. 106, 2009, DOI: 10.1063/1.3267152.
- [160] A. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod, "Majority logic gate for magnetic quantum-dot cellular automata," *Science*, vol. 311, no. 5758, pp. 205–208, 2006.
- [161] E. Varga, A. Orlov, M. T. Niemier, X. S. Hu, G. H. Bernstein, and W. Porod, "Experimental demonstration of fanout for nanomagnetic logic," *IEEE Trans. Nanotechnol.*, vol. 9, no. 6, pp. 668–670, Dec. 2010.
- [162] M. T. Alam, M. J. Siddiq, G. H. Bernstein, M. Niemier, W. Porod, and X. S. Hu, "On-chip clocking for nanomagnet logic devices," *IEEE Trans. Nanotechnol.*, vol. 9, no. 3, pp. 348–351, May 2010.
- [163] G. L. Snider, A. O. Orlov, I. Amlani, X. Zuo, G. H. Bernstein, C. S. Lent, J. L. Merz, and W. Porod, "Quantum-dot cellular automata: Review and recent experiments," *J. Appl. Phys.*, vol. 85, no. 8, pp. 4283–4285, 1999.
- [164] V. Zhirnov, R. Cavin, G. Leeming, and K. Galatsis, "An assessment of integrated digital cellular automata architectures," *IEEE Computer*, vol. 41, no. 1, pp. 38–44, Jan. 2008.
- [165] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotechnol.*, vol. 5, pp. 266–270, 2010.
- [166] N. G. Cooper, "From Turing and von Neumann to the present," *Los Alamos Sci.*, pp. 22–27, Fall, 1983.
- [167] J. von Neumann, *Theory of Self-Reproducing Automata*. Chicago, IL: Univ. Illinois Press, 1966.
- [168] S. Ji, "The cell as the smallest DNA-based molecular computer," *Biosystems*, vol. 52, pp. 123–133, 1999.
- [169] A. Danchin, "Bacteria as computer making computers," *FEMS Microbiol. Rev.*, vol. 33, pp. 3–26, 2009.
- [170] C. T. Fernando, A. M. L. Liekens, L. E. H. Bingle, C. Beck, T. Lenser, D. J. Stekel, and J. E. Rowe, "Molecular circuits for associative learning in single-celled organisms," *J. Roy. Soc. Interface*, vol. 6, pp. 463–469, 2009.
- [171] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, pp. 307–312, 1995.

- [172] N. Ramakrishnan, U. S. Bhalla, and J. J. Tyson, "Computing with proteins," *Computer*, vol. 42, pp. 47–56, 2009.
- [173] L. F. Agnati, D. Guidolin, C. Carone, M. Dam, S. Genedani, and K. Fuxe, "Understanding neuronal cellular network architecture," *Brain Res. Rev.*, vol. 58, pp. 379–399, 2008.
- [174] A. Wagner, "From bit to it: How a complex metabolic network transforms information into living matter," *BMC Syst. Biol.*, vol. 1, 2007, DOI: 10.1186/1752-0509-1-33.
- [175] S. Ji, "Free energy and information content of Conformons in protein and DNA," *Biosystems*, vol. 54, pp. 107–130, 2000.
- [176] R. U. Ayres, *Information, Entropy and Progress*. New York: AIP, 1994.
- [177] University of Wisconsin—Madison, *The Microbial World*. [Online]. Available: <http://textbookofbacteriology.net/themicrobialworld/nutgro.html>
- [178] J. von Neumann, *The Computer and the Brain*. New Haven, CT: Yale Univ. Press, 1959.
- [179] G. Bate, "Bits and genes: A comparison of the natural storage of information in DNA and digital magnetic recording," *IEEE Trans. Magn.*, vol. MAG-14, no. 5, pp. 964–965, Sep. 1978.
- [180] D. J. D'Onofrio and G. An, "A comparative approach for the investigation of biological information processing: An examination of the structure and function of computer hard drives and DNA," *Theor. Biol. Med. Model.*, vol. 7, pp. 3–20, 2010.
- [181] L. G. Valiant, "A quantitative theory of neural computation," *Biol. Cybern.*, vol. 95, no. 3, pp. 205–211, 2006.
- [182] R. P. Feynman, *Miniatrization*, D. H. Gilbert, Ed. New York: Reinhold, 1961, pp. 282–296.

ABOUT THE AUTHORS

Ralph K. Cavin, III (Life Fellow, IEEE) received the B.S.E.E. and M.S.E.E. degrees from Mississippi State University, Starkville, in 1961 and 1962, respectively, and the Ph.D. degree in electrical engineering from Auburn University, Auburn, AL, in 1968.

He was Senior Engineer at the Martin-Marietta Company, Orlando, FL, from 1962 to 1965. In 1968, he joined the faculty of the Department of Electrical Engineering, Texas A&M University, College Station, obtaining the rank of Full Professor. In 1983, he joined the Semiconductor Research Corporation, Triangle Park, NC, as Director of Design Sciences. He became Head of the Department of Electrical and Computer Engineering from 1989 to 1994 and Dean of Engineering at North Carolina State University, Raleigh, from 1994 to 1995. He served as the Semiconductor Research Corporation Vice President for Research Operations from 1996 to 2007 and is currently the SRC Chief Scientist. His technical interests span very large-scale integration (VLSI) design, advanced information processing technologies, semiconductor device and technology limits, and control and signal processing. He has authored or coauthored over 100 refereed technical papers and contributions to books. He has served as an advisor to a number of government, industrial, and academic institutions.



Paolo Lugli (Fellow, IEEE) graduated in physics from the University of Modena, Modena, Italy, in 1979. He received the M.Sci. and Ph.D. degrees in electrical engineering from Colorado State University, Fort Collins, in 1982 and 1985, respectively.

In 1985, he joined the Physics Department, University of Modena, as a Research Associate. From 1988 to 1993, he was an Associate Professor of "Solid State Physics" at the "Engineering Faculty" of the 2nd University of Rome "Tor Vergata," Rome, Italy, where in 1993, he was appointed as Full Professor of "Optoelectronics." In 2002, he joined the Technische Universität München, Munich, Germany, where he was appointed Head of the newly



created Institute for Nanoelectronics. His current research interests involve, besides nano imprint lithography, the modeling, fabrication, and characterization of organic devices for electronics and optoelectronics applications, the design of circuits and architectures for nanostructures and nanodevices, the numerical simulation of microwave semiconductor devices, and the theoretical study of transport processes in nanostructures. He is the author of more than 250 scientific papers and the coauthor of the books *The Monte Carlo Modelling for Semiconductor Device Simulations* (New York: Springer-Verlag, 1989) and *High Speed Optical Communications* (Norwell, MA: Kluwer, 1999).

Dr. Lugli served as General Chairman of the IEEE International Conference on Nanotechnology held in Munich, Germany, in 2004. He is a member of the German "National Academy of Science and Engineering" (ACATECH).

Victor V. Zhirnov received the M.S. degree in applied physics from the Ural Polytechnic Institute, Ekaterinburg, Russia, in 1989 and the Ph.D. in solid state electronics and microelectronics from the Institute of Physics and Technology, Moscow, Russia, in 1992.

He is the Director of Special Projects at the Semiconductor Research Corporation, Triangle Park, NC, which he joined in 2004. He holds adjunct faculty position at North Carolina State University, Raleigh, and has served as an advisor to a number of government, industrial, and academic institutions. From 1992 to 1998, he was a Senior Scientist at the Institute of Crystallography, Russian Academy of Science, Moscow, Russia. From 1998 to 2004, he was a Research Professor at North Carolina State University. His research interests include nanoelectronics devices and systems, properties of materials at the nanoscale, bio-inspired electronic systems, etc. He has authored and coauthored over 100 technical papers and contributions to books.

Dr. Zhirnov is the Chair of the Emerging Research Device (ERD) Working Group for the International Technology Roadmap for Semiconductors (ITRS).

