

Device and Architecture Outlook for Beyond CMOS Switches

Many new devices that are being studied as replacements for CMOS are discussed in this paper; early results for benchmarking and performance comparison are presented for some of the devices.

By KERRY BERNSTEIN, *Fellow IEEE*, RALPH K. CAVIN, III, *Life Fellow IEEE*, WOLFGANG POROD, *Fellow IEEE*, ALAN SEABAUGH, *Fellow IEEE*, AND JEFF WELSER, *Senior Member IEEE*

ABSTRACT | Sooner or later, fundamental limitations destine complementary metal-oxide-semiconductor (CMOS) scaling to a conclusion. A number of unique switches have been proposed as replacements, many of which do not even use electron charge as the state variable. Instead, these nanoscale structures pass tokens in the spin, excitonic, photonic, magnetic, quantum, or even heat domains. Emergent physical behaviors and idiosyncrasies of these novel switches can complement the execution of specific algorithms or workloads by enabling quite unique architectures. Ultimately, exploiting these unusual responses will extend throughput in high-performance computing. Alternative tokens also require new transport mechanisms to replace the conventional chip wire interconnect schemes of charge-based computing. New intrinsic limits to scaling in post-CMOS technologies are likely to be bounded ultimately by thermodynamic entropy and Shannon noise.

KEYWORDS | Nanoarchitectures; nanomagnet logic; post-complementary metal-oxide-semiconductor (CMOS); pseudospin; quantum-dot cellular-automata architectures (QCAs); quantum-dot cellular automata; spin; tunnel field-effect transistor (TFET); tunneling

I. MOTIVATION

It has been estimated that information technology (IT) producing and intensive IT-using industries currently account for over a quarter of the U.S. Gross Domestic Product (GDP), and drive 50% of this country's economic growth [1]. The unprecedented growth of the IT industry has largely been due to the exponential increase in the performance of the semiconductor chips that are at the heart of all modern electronics. The key component on these chips is the complementary metal-oxide-semiconductor (CMOS) field-effect transistor (FET), and the ability to scale these devices to ever-smaller dimensions has been the primary driver of this increased performance. For over 30 years, the industry has been able to pack twice as many FETs onto a chip every 18–24 months, in what has come to be known as “Moore's law” [2]. This has resulted in an exponential increase in the information processing capability per unit area on the chip—or more importantly, per dollar. This has meant not only that existing chip-based products get faster and/or cheaper each year, but also has expanded the number of products that use semiconductor chips to increase functionality, from toasters to cell phones to supercomputers.

The rules for FET scaling that have enabled this revolution were outlined by Dennard *et al.* in the early

Manuscript received February 3, 2010; accepted May 24, 2010. Date of publication October 4, 2010; date of current version November 19, 2010. This work was supported in part by the Semiconductor Research Corporation, Nanoelectronics Research Institute (SRC-NRI) including the member companies the National Institute of Standards and Technology (NIST) and the National Science Foundation (NSF).

K. Bernstein is with IBM T. J. Watson Research Center, Yorktown Heights, NY USA (e-mail: kbernstein@us.ibm.com).

R. K. Cavin, III is with the Semiconductor Research Corporation, Durham, NC 27707 USA (e-mail: cavin@src.org).

W. Porod and **A. Seabaugh** are with the University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: porod@nd.edu; seabaugh.1@nd.edu).

J. Welsler is with the IBM Almaden Research Center, San Jose, CA 95120-6099 USA and with the Semiconductor Research Corporation, Durham, NC 27707 USA (e-mail: jeff.welsler@src.org).

Digital Object Identifier: 10.1109/JPROC.2010.2066530

1970s [3]. The key insight was that if all of the critical dimensions of the FET, along with the operating voltage, were reduced by the same factor, the speed of the FET would go up while the area and power would go down, so that the power density remained constant. However, in recent chip generations, our ability to scale the voltage has become limited due to the difficulty to maintain performance. While it seems clear that CMOS can continue to scale in size for at least another ten years, our ability to continue to achieve the full historical benefits of scaling is being limited, as we are forced to trade off between transistor density and speed to mitigate the power density increase. Moreover, the power concern is not unique to silicon (the semiconductor used in CMOS FETs) but would in fact apply to a FET in any material, including exotic options such as carbon nanotubes or organic molecules. While changing materials might improve FET operation for a generation or two (which certainly might be worthwhile), a new device needs to be found to allow long-term continued scaling—where scaling should be understood in the most generic sense of increasing computational performance (function) per unit area (dollar) in each subsequent generation.

The International Technology Roadmap for Semiconductors (ITRS) Emerging Research Device Technical Working Group began to study the challenge presented by power density for future scaling in the early 2000s. It is interesting to note that this quest for a new digital switch is not unprecedented: the first solid state transistors, based on bipolar technology, were developed when vacuum tubes and mechanical switches were reaching similar power constraints in the late 1940s, and that the current FET replaced bipolar transistors in the majority of semiconductor applications in the late 1980s for the same reason. To structure a well-defined program to develop a new switch, the Semiconductor Research Corporation (SRC) and the National Science Foundation (NSF) jointly organized a set of industry–academia–government workshops [4]–[6]. In parallel, the Technology Strategy Committee of the Semiconductor Industry Association (SIA) also conducted several workshops whose objective was to identify research initiatives to advance integrated circuit technology beyond currently identified scaling limits. These activities ultimately defined 13 research vectors considered to be important components of the search for the next switch, with the first five of these vectors considered to be crucial for a research program: 1) computational state vectors, other than charge; 2) nonequilibrium systems; 3) novel, noncharge data transfer mechanisms; 4) nanoscale phonon engineering for thermal management; and 5) directed self-assembly of such structures.

The SIA chartered a new research program to pursue these vectors in 2005. Managed by the SRC, the Nanoelectronics Research Initiative (NRI) has the mission to demonstrate novel computing devices capable of replacing the CMOS FET as a logic switch in the 2020 timeframe.

These devices should show significant advantage over ultimate FETs in power, performance, density, and/or cost to enable the semiconductor industry to extend the historical cost and performance trends for information technology.

To meet these goals, the NRI has focused research on devices utilizing new computational state variables and switching mechanisms. In addition, the NRI is interested in new interconnect technologies and novel circuits and architectures, including nonequilibrium systems, for exploiting these devices, as well as improved nanoscale thermal management and novel materials and fabrication methods for these structures and circuits. Finally, it is desirable that these technologies be capable of integrating with CMOS, to allow exploitation of their potentially complementary functionality in heterogeneous systems and to enable a smooth transition to a new scaling path.

The NRI member companies comprise many of the leading U.S. semiconductor companies (AMD/GlobalFoundries, IBM, Intel, Micron, and Texas Instruments), which partner with both federal agencies and state governments to sponsor research at U.S. universities. The program currently funds over 30 universities in 20 states (Fig. 1) using two different research models. The bulk of the NRI research takes place in four multi-university, virtual centers (Fig. 2) funded by industry, the National Institute of Standards and Technology (NIST), and the lead state and local governments. Each of these centers focuses on a different approach to finding a post-CMOS logic switch.

- Western Institute of Nanoelectronics (WIN), University of California at Los Angeles (UCLA; Dir: Prof. Kang Wang): focuses on spintronics and related phenomena, including materials, device structures, and interconnects, for logic applications.
- Institute for Nanoelectronics Discovery and Exploration (INDEX), State University of New York (SUNY), Albany (Dir: Prof. Alain Kaloyeros): focuses on new phenomena for logic devices, organized in centers of competency around excitonic, quantum-dot spin, magnetic, and graphene devices, with emphasis on fabrication and characterization.
- SouthWest Academy for Nanoelectronics (SWAN), University of Texas, Austin (Dir: Prof. Sanjay Banerjee): focuses on graphene, integrating projects on theory, material fabrication, device structures, and metrology, as well as work on magnetic materials, pseudospintronics, magnetic and multiferroic materials, and plasmonics.
- Midwest Institute for Nanoelectronics Discovery (MIND), University of Notre Dame, Notre Dame (Dir: Prof. Alan Seabaugh): focuses on tunneling, nanomagnetism, and nonequilibrium phenomena for energy efficient devices and architectures, as well as thermal phonon management.

In addition, NRI is jointly funding 18 projects with the National Science Foundation (NSF) at 15 existing NSF

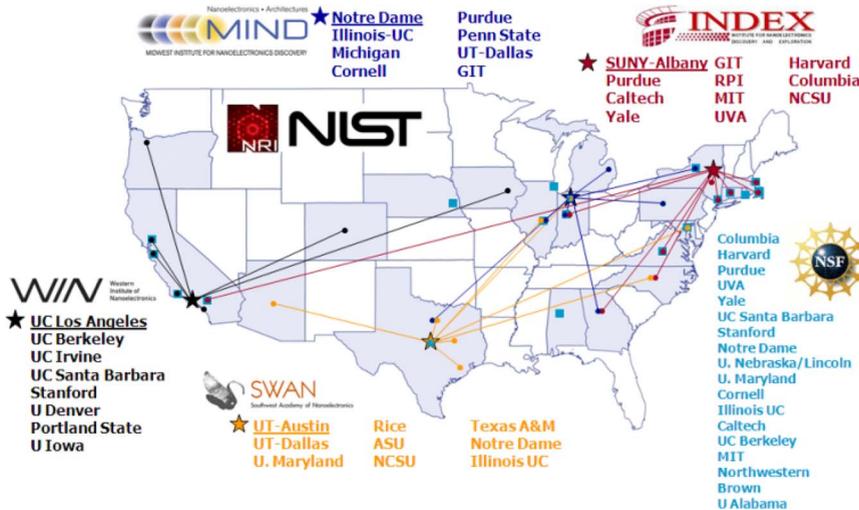


Fig. 1. NRI Research Programs, including NRI-NIST Centers and projects at NSF Nanoscience Centers.

Nanoscience Centers across the country: the Nanoscale Science and Engineering Centers (NSECs), the Materials Research Science and Engineering Centers (MRSECs), and the Network for Computational Nanotechnology (NCN).¹

Given the complexity of finding a new phenomenon capable of being exploited to perform logic, and the broad range of disciplines required, the NRI vision has been to foster a goal-oriented, basic-science research program—one that gives the researchers freedom to explore a large range of novel base-research ideas, guided by the final goal of finding a new device. A key aspect of this has been having technical experts from the member companies and NIST working directly with the university researchers, to give the academics insight into the practical challenges facing the industry while simultaneously facilitating transfer of any promising emerging ideas back to the member company labs for further development.

Over the past four years, NRI has ramped up very quickly and several emerging device ideas are starting to show promise. It has become clear that a method for benchmarking these ideas is needed, and the NRI industry and academic researchers recently embarked on this effort. Unlike the benchmarking of one CMOS technology against another, which is fairly straightforward and involves a known set of agreed upon parameters for measurement, benchmarking NRI devices is an exercise often requiring comparisons of apples to oranges. Many of the devices operate on very different physical principles and may perform computation utilizing unique architectures, so it requires looking at not just the device but also the circuit implementation and in some cases even the specific application or computation algorithm being implemented.

¹More information on all of the research programs, with links to specific projects and research centers, is available on the NRI website: <http://nri.src.org>.

Hence, our goal is to find a quantitative set of metrics that can be used to contrast the devices and architectures on a relatively even playing field. At this time, the metrics are not sufficient to judge a device as “good” or “bad,” but are being used to stimulate continued innovation by the NRI researchers, by highlighting both the favorable attributes and the technical challenges.

The benchmarking effort described here is part of an evolving process undertaken to guide the NRI over the next several years. This paper provides an overview of the approach, as well as a snapshot of the results with the intent to encourage more researchers to consider the “grand challenge” of finding a new device to extend beyond the limits of CMOS.

II. POST-CMOS ALTERNATIVES

Ideally, such a new post-CMOS switch would function as a drop-in replacement for CMOS, but the new switch technology might also complement conventional CMOS FETs in

WIN	INDEX	SWAN	MIND
Western Institute of Nanoelectronics	Institute for Nanoelectronics Discovery & Exploration	SouthWest Academy for Nanoelectronics	Midwest Institute for Nanoelectronics Discovery
UCLA, UCSB, UC-Irvine, Berkeley, Stanford, U Denver, Iowa, Portland State	SUNY-Albany, GIT, RPI, Harvard, MIT, Purdue, Yale, Columbia, Caltech, NCSU, UVA	UT-Austin, UT-Dallas, TX A&M, Rice, ASU, Notre Dame, Maryland, NCSU, Illinois-UC	Notre Dame, Purdue, Illinois-UC, Penn State, Michigan, UT-Dallas, Cornell, GIT
Spin devices Spin circuits Benchmarks & metrics Spin Metrology	Novel state-variable devices Fabrication & Self-assembly Modeling & Architecture Theory & Simulation Roadmap Metrology	Logic devices with new state-variables Materials & structures Nanoscale thermal management Interconnect & Arch Nanoscale characterization	Graphene devices: Thermal, Tunnel, & Spin Interband Tunnel Devices Non-equilibrium Systems Model / Measurement Nanoarchitecture

Fig. 2. NRI-NIST Research Centers.

hybrid structures, and leverage the significant existing technology infrastructure. Alternatively, the CMOS successor technology might take advantage of the idiosyncrasies of the new switch, and implement certain functionality in different ways than done today. This search for a new post-CMOS switch will require a rethinking of how information is represented, how information is manipulated, and what logic functions can naturally be realized in a given technology.

The basic issue here is the need to represent (binary) information by some physical property, which readily allows the manipulation of this information according to rules given by logic. This physical property also has to allow the storage and transmission of information. Charge, a basic property of matter, has been most successful for this purpose since the dawn of the information age, but the power dissipation associated with the flow of charge (i.e., current) in CMOS-based chips is reaching unacceptable levels.

The amount of power dissipation is fundamentally related to three basic issues: 1) the physical property used to represent the information (the data “token”); 2) the way this physical property is being used to manipulate information (switch + transport mechanism); and 3) the way logic functionality is achieved (architecture). Current CMOS technology is based on the following basic choices, which necessarily entail certain consequences for chip performance.

- 1) Token: FET switches use electron charge to capture and transfer information. The absence or presence of charge enables the representation of binary information.
- 2) Transport mechanism: In a FET, the flow of charges is controlled by a (voltage) barrier that regulates over-the-barrier transport. Thermodynamics dictates that the turn-on of a FET requires 60 mV/decade increase in current, and, as a consequence, this sets a limit on operating voltage for a given ON-to-OFF current ratio.
- 3) Architecture: FETs, as binary switches, naturally represent Boolean logic functions. As such, inversion, AND, and OR functions can readily be realized in CMOS. Other functions are built from these logic primitives, and processors are based on the von Neumann stored-program architecture.

The challenge of the Semiconductor Research Corporation, Nanoelectronics Research Institute (SRC-NRI) is to identify new tokens, new transport mechanisms, and possibly new architectures, to identify a successor switch. The research portfolio of the NRI includes both noncharge-based tokens such as spin (either single or collective), pseudospin, magnetization, excitons, and plasmons, as well as charge itself, albeit with tunneling-based or other switching mechanisms.

Graphene has received much attention in recent years because of its linear (photon-like) dispersion characteristic

[7]. Furthermore, 2-D graphene sheets can be patterned into quasi-1-D graphene nanoribbons, where the energy gap due to the lateral confinement can be controlled by lithographic means [8]. Graphene bilayers, i.e., two sheets of graphene in close proximity, are predicted to have unusual transport characteristics. Specifically, interactions between the charge carriers in the two adjacent layers lead to correlations, which can be described as an excitonic pseudospin state [9]. Theory predicts that exciton condensation might be quite robust in suitably engineered graphene bilayers [10], and such a Bose condensate would exhibit transport characteristics of interest for low-power switching. Based on this new material system and transport mechanism, a new device concept, the bilayer pseudospin FET is under investigation [11].

Spin is another fundamental property of matter that can be used to represent information. Spin represents the quantized angular momentum of an electron (or nuclear particle), which also leads to a magnetic moment, and thus is the origin of the magnetic properties of matter. Two notable manifestations of individual spins of interest for devices are strongly coupled spins (magnetic domains) and correlated spins (spin waves).

For individual spins, an applied magnetic field leads, due to the Zeeman effect, to two distinguishable states (spin “up” and “down”), which can be used to represent binary information. A similar level splitting also results, due to the Rashba effect, in the vicinity of heterointerfaces in the presence of an electric field and spin-orbit coupling. The latter effect [12] is the basis for several spin-FET device proposals. What these proposals have in common is that the spin token is transported by electrons along with their charge. Spin-FETs of this type face the same limitations encountered by conventional charge-token FETs. Where individual spins might have an advantage is in the switching between the two basic states, since switching of spins is different than moving charges. Other challenges of using individual spins to represent information include 1) the writing and reading of this information, and 2) that individual spins do not readily lend themselves to gain, a critical limitation for logic applications.

Magnetic domains result from the strong coupling of atomic magnetic moments, mediated by the sea of conduction electrons through quantum-mechanical exchange coupling. Information can readily and reliably be encoded in the magnetization directions of ferromagnetic domains, and this phenomenon is extensively used for data-storage applications. In addition, it has been recognized early on that magnetic phenomena can also be used for logic [13], and one of the early computers, the Elliott 803, used magnetic cores not only for data storage, but also for logic. In recent years, patterned magnetic islands, which are sufficiently small to support only one magnetic domain, have been proposed as magnetic switches for logic [14], [15]. In this proposal, the individual single-domain nanomagnets interact through their physical dipole

interactions; logic functionality has been demonstrated [16]. The use of direct physical interactions to achieve connectivity naturally maps such magnetization-token switches onto locally interconnected quantum-dot cellular-automata architectures (QCAs) [17], [18]. However, it has also been proposed to implement random combinatorial logic circuits using these interacting nanomagnets. Nanomagnet logic (NML) holds the promise of an all-magnetic information processing system that combines memory and logic.

Spin waves (magnons) are collective oscillations of spins in a magnetic material, and they have coherence lengths of several microns at room temperature. The phase of the spin wave can be used as the information token [20], and logic functionality can be achieved through wave interference [21]. Since such spin waves do not exhibit gain, external circuitry needs to be added to provide signal restoration.

Wave phenomena have also been proposed as information tokens in arrays of optically coupled plasmonic particles [22]. Surface plasmons are coupled vibrations of electrons and the electromagnetic field, formed at the interface of thin metals and dielectrics, and at resonance (which sensitively depends upon the size and the shape of the metallic particle) strong field enhancement exists which may lead to physical coupling between neighboring particles. Within the NRI, these phenomena have been proposed for a terahertz plasmon–polariton switch [23].

Temperature (or a phonon) has been proposed as an information token. Such thermal circuits are based on microheaters and thermometers as building blocks. It has been demonstrated experimentally that asymmetrically patterned graphene can be engineered to achieve thermal rectification, and negative differential thermal conductivity has been observed in these structures [24]. This work is in its early stages, and further research is needed to evaluate its potential for logic.

While several post-CMOS information tokens and transport mechanisms are under intense study, there has been less progress on alternative architectures. Exceptions include a binary decision tree architecture based on single-electron devices [25], and locally interconnected QCA structures for NML [26]. It appears that opportunities exist for novel architectures to take advantage of the idiosyncrasies of novel switches, and this work will receive increased emphasis as the NRI moves forward. To confine the scope of this study however, new switches were compared on the binary “level playing field.”

What has become apparent so far is the interesting possibility of merging logic with memory. While the emphasis of the NRI is on a post-CMOS logic switch, several of the emerging research devices could also function as memory. Embedded nonvolatile memory could enable check pointing (without the need to write to disk) and offer the potential of instant-ON processors. Such a merging of logic with memory might also open possibilities for processor-in-memory and logic-in-memory architectures.

III. NEW SWITCHES AND MECHANISMS FOR LOGIC

The devices included in this study are a subset of those being studied in the NRI and are listed in Table 1. As a point of reference, the first entry is a metal–oxide–semiconductor FET (MOSFET) with a 15-nm gate length simulated by Augustine *et al.* [27] using the analytic model of Khakifirooz and Antoniadis [28]. This model provides a self-consistent prediction for an advanced CMOS technology node. The Purdue Emerging Technology Emulator (PETE) [27] was used to obtain circuit performance estimates for the MOSFET and the graphene-based tunnel FET (TFET) examined in the study. PETE accepts numerical inputs for drain current per micron gate width versus gate voltage and drain voltage, and gate capacitance per micron versus gate voltage, and uses this input to compute the performance and power consumption of an inverter with a fan-out of 1, a two-input NAND gate with a fan-out of 1, a two-input XOR gate with a fan-out of 1, ten-stage NAND/NOR chains, a ring oscillator and an 8-b ripple carry adder.

The range of device concepts included in this study is beyond what can be adequately described in a short space. Fortunately, there is a recent review of spin and molecule-based devices [29], however many new proposals have been made in the last year and some are not yet published. Several of these new concepts will be introduced here.

Spinwave logic devices (Spinwave) refers to devices like the one demonstrated by Khitun *et al.* [30], [31] consisting of surface inputs and output wires on a SiO₂/NiFe bilayer. Currents in the input wires generate magnetic fields perpendicular to the magnetization of the NiFe layer. The input magnetic field launches spin waves in the NiFe that interfere to perform logic. Detection is made with a current loop. There is no gain mechanisms for the device shown in Fig. 3; methods for incorporating gain into the device are in development [32].

Nanomagnets encode a binary logic state in the magnetization direction of a thin film ferromagnet. Chains of patterned nanomagnets in a magnetic quantum-dot cellular array architecture (MQCA) are used to both transmit the information and to perform logic. Majority gates have been demonstrated [16]; clocked logic gates are now in development, as shown in Fig. 4. New concepts for nanomagnet switching in a cellular architecture are being explored in a concept labeled reconfigurable array of magnetic automata (RAMA) [33]. In RAMA, multiferroics are used to null nanomagnet pillars and magnetocapacitance is used to sense magnetic polarization.

TFETs use electric field gating of interband tunnel currents to enable low supply voltages and sub-60-mV/decade subthreshold swing [35]. Fig. 5 shows a graphene nanoribbon TFET schematic and energy band diagram in the ON- and OFF-states. In the OFF-state, the gate depletes the channel and suppresses interband tunneling. In the ON-state, with positive gate bias, interband tunneling in the source is enabled. Simulated *n*- and *p*-channel TFET

Table 1 Devices in Benchmarking Project

Devices Benchmarked	Short	Device	State variable	Input	Output	Clock	Reference
15 nm CMOS	CMOS HP	CMOS $LG = 15$ nm	Q	V	V	–	Augustine 2009
	CMOS LP						
Excitonic FET	ExFET	excitonic FET	Q	V	V		J. Appenzeller, Purdue, unpublished
Magnetic tunnel junction MTJ logic switch	MTJ Logic Switch	Co/MgO/NiFe GMR or Co/Cu/NiFe or TMR	magnetic polarization	I	I	–	C. Ross, MIT unpublished
All-spin logic	All-Spin Logic	semiconductor spin channel with nanomagnet input/output	electron spin	I	I	I	B. Behin-Aein 2010
Graphene PN Junction	Graphene PN Junction	field-controlled graphene p - n junction	electron wave phase	V	V	–	J.-U. Lee, SUNY Albany unpublished
Electronic Ratchet	Electronic Ratchet	backgated graphene structured-nanoribbon	Q	V	V	V	M. Stan, unpublished
Graphene thermal logic	Thermal	graphene thermal transistor 10 nm x 10 nm	T	T	T	–	Y. P. Chen, Purdue, unpublished
Binary decision diagram architecture	BDD Arch	generic	Q	V	V	V	S. Datta and V. Narayanan, Penn
Nanomagnet logic	NML	permalloy nanomagnet chains 40 x 60 x 60 nm	magnetic polarization	magnetic polarization	magnetic polarization	I	A. Dingler 2009
Graphene tunnel FET	Tunnel FET	graphene nanoribbon TFET $w = 3$ nm, $LG = 20$ nm	Q	V	V	–	Q. Zhang 2008
InAs tunnel FET	InAs TFET	nanowire TFET 7 nm diameter, $LG = 30$ nm	Q	V	V	–	Y. Lu, Notre Dame, unpublished
e-Struct. Modulation Transistor	e-Struct. Modulation Trans.	edge-gated graphene nanoribbon	Q	V	V	–	Raza, Univ. Iowa unpublished
Reconfigurable array of magnetic automata	RAMA	gated multiferroic CoFeO/BiFeO(BaTiO) composite	magnetic polarization	V	V	V	S. Wolf, unpublished
Bilayer pseudospin FET	BiSFET	graphene/insulator/graphene bilayer pseudospin FET	pseudospin	V	V	V	S. Banerjee 2009
Resonant-enhanced-injection FET	RIEFET	III-V heterojunction resonant tunneling FET	Q	V	V	–	L. F. Register, Univ. TX Austin, unpublished
Heterobarrier tunnel FET	HeTFET	III-V heterojunctions TFET	Q	V	V	–	L. F. Register, Univ. TX Austin, unpublished
Spin wave phase logic	Spin Wave	spin wave 3-input majority gate in NiFe	spinwave phase	V	V	–	A. Khitun 2008b
Magnetic tunnel junction spin torque transfer	MTj/STT	Spin torque logic with magnetic tunnel junctions and CMOS	magnetic polarization	V	V	–	D. Markovic, UCLA unpublished
Spin torque amplifier	Spin Torque Amplifiers	MgO tunnel junction	magnetic polarization	V	V	–	A. Krivorotov, UCLA unpublished

current-per-unit-gate-width characteristics are shown in Fig. 5 including parasitic capacitance and access resistance. A second embodiment of the interband TFET examined in this study uses a heterobarrier (HeTFET), outlined in Fig. 6. Field control of resonant tunneling is represented by a third transistor, labeled RIEFET (resonant-injection-enhanced) and described in [36]. The Datta–Das spin FET [12] based on graphene (graphene SpinFET) is also listed in Table 1.

Another new transistor concept, which utilizes tunneling, is the bilayer pseudospin FET (BiSFET) previously discussed. In the BiSFET, two metal oxide gates sandwich two separately contacted graphene monolayers, which are themselves separated by a tunnel oxide (Fig. 7). Under certain gate conditions, an exciton condensate forms between the graphene layers leading to the possibility of a collective many-body current between the two layers [37]. The circuit operation of the BiSFET discussed here was implemented in SPICE and analyzed by Banerjee *et al.* [38].

Another new transistor concept using collective many-body effects has recently been proposed by Appenzeller [39]. This low-subthreshold-swing transistor is illustrated in Fig. 8. Coulomb interaction between electrons in an n -type branch and holes in a p -type branch enable exciton binding under certain gate bias conditions. As a function of gate bias the channel current switches between a conductive ON-state to a nonconductive OFF-state when the conditions for formation of a collective excitonic condensate are satisfied.

All-spin logic (ASL) is a magnetic-spin-based logic approach [40] in which nanomagnets are used to store the state, information is communicated between magnets by spin currents, and spin torque is used to determine the output magnetization state. The device concept is outlined in Fig. 9.

Single-electron transistors (SETs) used in a binary decision diagram (BDD) architecture [41] have been simulated by Datta and Narayanan using the Monte Carlo simulator SIMON to provide a projection for this computing paradigm.

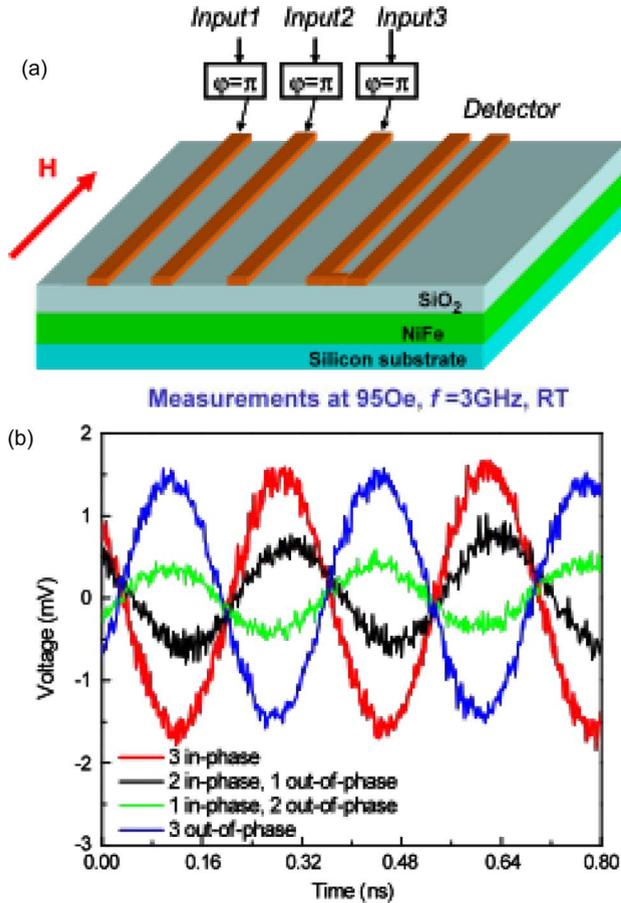


Fig. 3. (a) Spin wave three-input majority logic gate. Currents in the input lines create magnetic fields to generate spin waves in the NiFe transport layer. Information is coded in the phase of the spin wave (0 and π phases correspond to logic states 0 and 1 , respectively). The result of interference is detected by a surface loop conductor. (b) Experimental data showing the output inductive voltage for different phase combinations.

Three-terminal thermal logic gates in which the gate temperature controls the heat flux have been proposed by Wang *et al.* [42] and are now being developed in graphene by Chen *et al.* [43]. Performance estimates for this technology (referred to herein as “graphene thermal logic”) have been provided by Y. P. Chen, Purdue University, West Lafayette, IN.

Switch and gate concepts which utilize excitons, magnetic rings, Veselago focusing [44], spin torque, multiferroics [45], and electric-field coupling to single donor spins in semiconductors [46], are also part of the study, but have not reached a stage where circuit performance can be estimated.

IV. BENCHMARKING AND PERFORMANCE

For this study, quantitative architectural benchmarks were developed as extensions of the ITRS Emerging Research

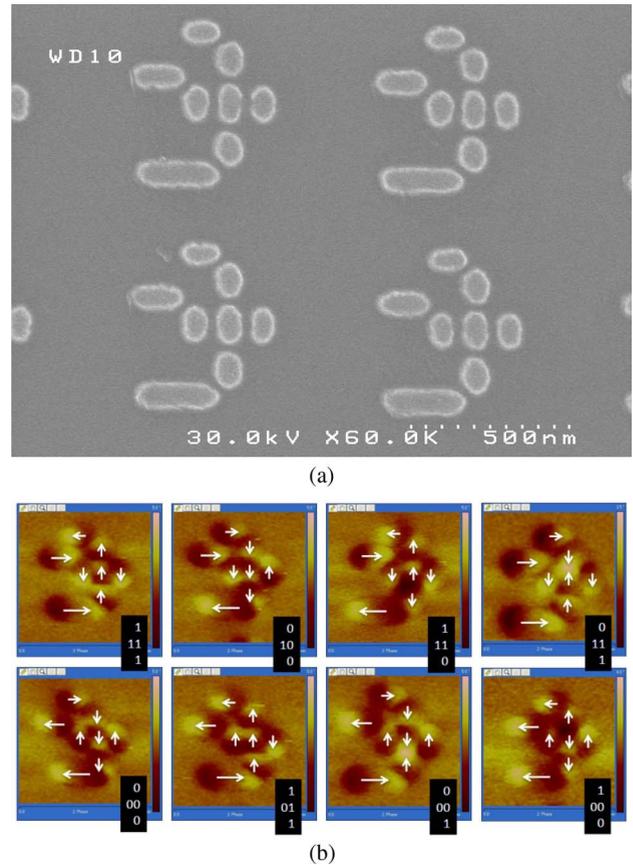


Fig. 4. Scanning electron micrographs of Nanomagnet logic quantum cellular automata (NML) (a) SEM photo of NAND2 (b) magnetic force microscopy (MFM) of NAND2.

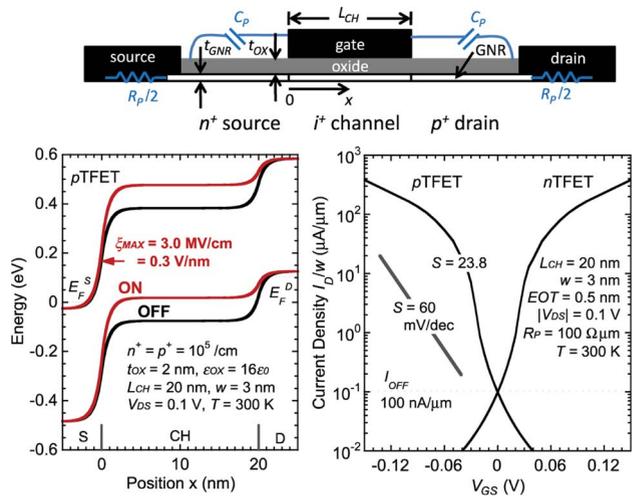


Fig. 5. Graphene nanoribbon TFET: (a) cross section including (b) energy band diagrams and (c) simulated drain current per unit gate width versus gate-to-source voltage.

Device (ERD) tabulation [47]. While the ERD metrics provided insight into fundamental parametrics of proposed devices (i.e., delay per switch, power per switch, area per device), the architectural extensions to this table attempt to anticipate the effectiveness of those switches in realizing specific higher order logic functions. Qualitative entries evaluated additional implementation concerns not captured in the quantitative measures, i.e., Is the technology compatible with CMOS? Does the switch require clocking? Is the switch scalable? Fig. 10 shows the final set of composite benchmarks selected. Thirteen independent device structures were assessed using data provided by the principal investigators. The device structures represented in this study are shown in Table 1, along with key characteristics. It is noteworthy that the tokens of information passed by these devices are magnetic polarization, electron spin, electron-wave phase, electron condensation, and spinwave phase. This study is perhaps the first attempt at comparing devices with disparate state variables using common figures of merit. The level of projection from the investigators varied; some provided theoretical estimates, others used models validated by experiment, some included parasitics, others did not. Efforts were taken to standardize the metrics, but inevitably, some variability in the interpretation of the specification can be anticipated.

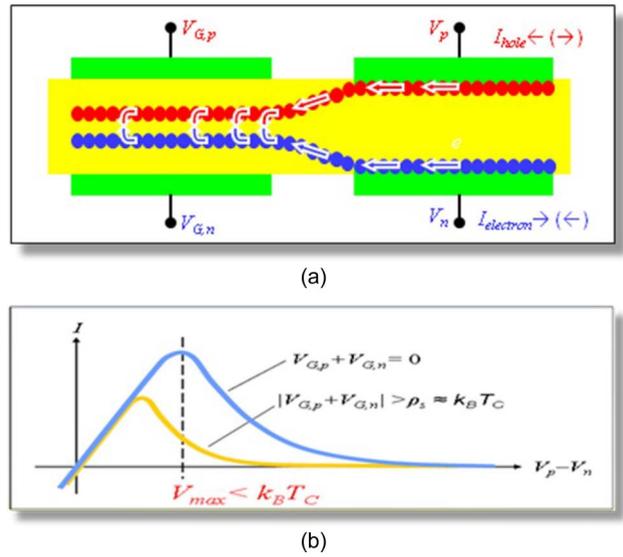


Fig. 7. BisFET consists of two graphene layers separated by a tunnel oxide: (a) schematic and (b) expected transistor characteristics.

Primary higher level logic functions evaluated included 8- and 32-b adders: a two-input NAND gate driving the input of an identical gate on its output (NAND2FO1),

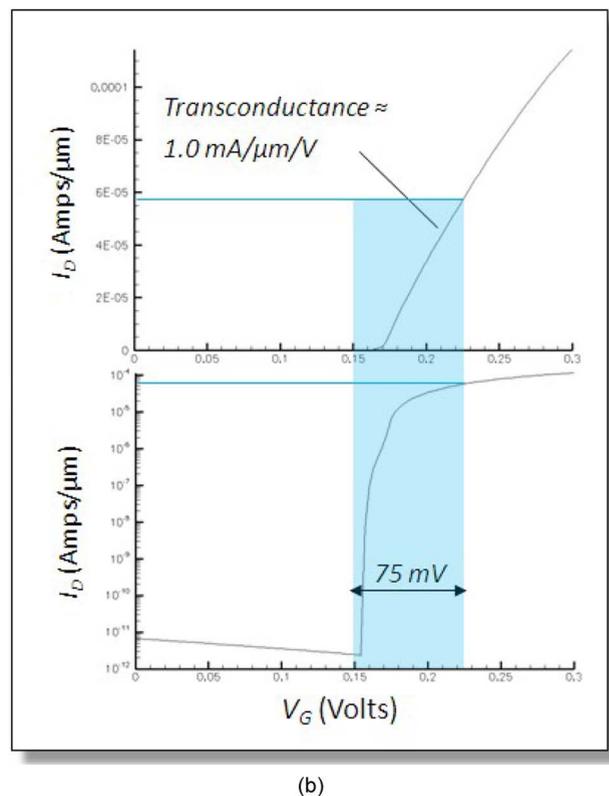
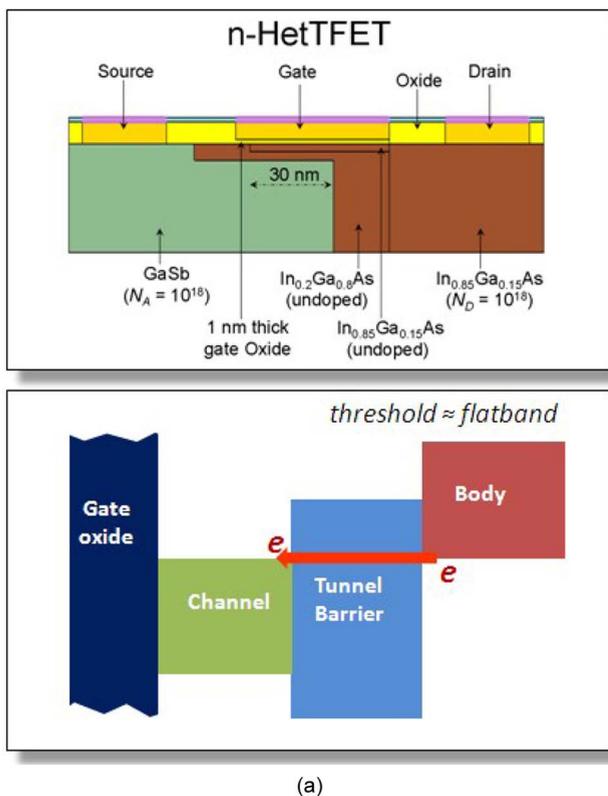


Fig. 6. HetFET: (a) schematic and energy band diagrams and (b) linear and log plots of Sentaurus device simulated transfer characteristics (L. F. Register, University of Texas at Austin, currently in review).

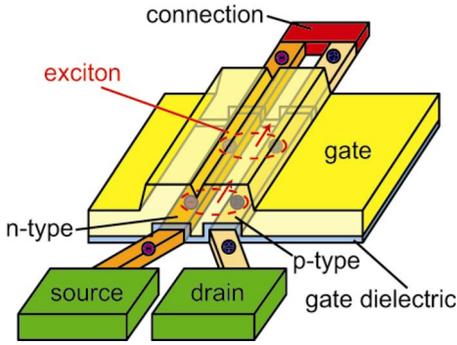


Fig. 8. Schematic layout of the excitonic field-effect transistor (ExFET) (J. Appenzeller, Purdue University unpublished).

and an inverter driving four identical inverters on its output (INVFO4). Fig. 11 provides the delay, energy, and area, respectively, of the median data from the study. Taken as a representation of the status of post-CMOS device development, the median delay of higher level logic functions is at least one decade slower than that of CMOS. It is appropriate to expect then that architectures leveraging parallelism rather than device performance will become of increasing value as investigators seek to replace CMOS devices. Median energy per function reductions of at least 10X and median area reductions of approximately 2X illustrate the promise that replacement switches already offer. Horowitz *et al.* observed that the product of energy and delay of logic operations serves as a useful tool in assessing overall effectiveness [48]. Energy delay product (EDP) for the four studied logic functions was calculated and plotted in Fig. 12. Low EDP values are generally associated with more effective structures.

Delay, power, and area for both simple and complex logic were also examined, referencing their particular information token, to determine if specific state variables offer consistently superior solutions. Fig. 13 shows a plot of the NAND2FO1, with each data point annotated for the specific state variable in use. The preferred characteristics

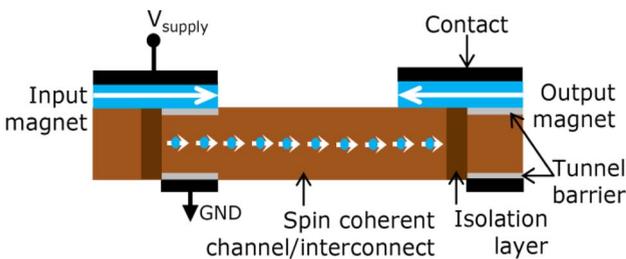


Fig. 9. All-spin logic concept [40] in which input and output states are set in nanomagnet inputs and outputs and the states are flipped using spin currents. Clocking, used to null the output magnet prior to setting the output state, is not shown.

for future switches place these devices in the far low corner of the plot. It is of little surprise that charge-based, evolutionary device proposals initially show the most promise, as the industry since its inception has depended upon charge to describe information. Fig. 14, plotting the more complex adder response, reveals however that while charge-based structures remain among the fastest, their area in complex functions is far from the best.

The study also indicated that the energy conundrum existing in CMOS appears to extend into new devices. Specifically, relationships in delay and noise immunity associated with energy existing in CMOS are observed to continue in alternative venues. Fig. 15 reports the estimated noise immunity of proposed replacement switches, plotted against the device’s energy per transition. Noise immunity is classically defined by the inset schematic as the signal margin remaining after subtracting the highest possible arriving “low” input signal strength from the highest possible input still interpreted by the circuit as a “low” signal. Noise immunity erosion eventually becomes indistinguishable from quantum-mechanical Heisenberg uncertainty as energy and area per switch drops, and presents a lower information limit to switch definition [49]. The next section addresses fundamental thermodynamics in more detail.

A number of significant conclusions may be drawn from the design data compiled from the study. 1) MOSFET successors have not yet surpassed CMOS in both circuit energy and delay, for Boolean applications. Continued device development may change this. 2) In alternative

Communication	
Communication Metrics	
Area of die/host accessible in 1 delay	
No. of Switches accessible in one delay	
Sq BW / Unit Area (Channels x Freq)X x (Channels x Freq)Y	
Sq Communication Channels (Nx x Ny) per unit area	
(Accessible Area within 1 Sw.Delay) / (Area of 1 Switch)	
Memory Element Delay / Logic Element Delay	
Logic Metrics	
32 Bit Adder	
Inverter with FO4	
NAND2 FO1	
Normalized Noise Immunity	
Normalized Logical Effort	
Compute Density (MIPS / no. of devices)	
PETE1	
PETE2	
Qualitative Benchmarks / Descriptions	
CMOS Compatibility	
Clocking Infrastructure and Locality	
Memory Requirements and Compatibility	
Scalability	
Reconfigurability or Library Dimension	
Logic Execution Technique	
Specific Logic Functions performed well	
Useful Specific Physical Behaviors	
Equivalent Accelerated Logic Function	
Logic Function accelerated in new Switch hardware	
Logic Function accelerated in CMOS hardware	
Logic Function expressed in CMOS software in CPU (unaccelerated)	
Improvement Ratio	

Fig. 10. Quantitative and qualitative architectural metrics used in study.

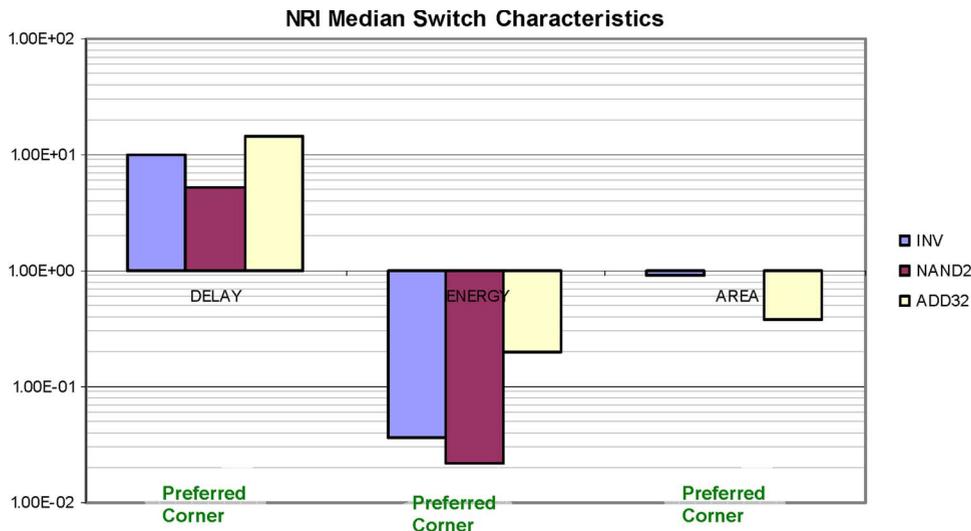


Fig. 11. Median delay, energy, and area of proposed devices, normalized to ITRS 15-nm CMOS. (Based on principal investigators' data.)

architectures, new devices may be superior. Replacement organizations may be neuromorphic, asymmetric-core, Bayesian, cellular nonlinear, or CMOL in nature. 3) Transport mechanisms will have a profound challenge communicating new information tokens. (This topic will be explored in more detail in Section VI.) 4) Post-CMOS architectures will need to accommodate parallel computation to enable improvement of delay and energy simultaneously. 5) The devices that performed well in this first assessment are predominantly charge-based, evolutionary three-terminal devices. Given the prevalence of charge-based computing, this is not a surprise, suggesting evolutionary devices may precede development of truly revolutionary structures. 6) The low-voltage, energy/delay tradeoff conundrum in CMOS continues, and most likely will ultimately define a lower limit for computing efficiency (see Section V). 7) Patterning, precise control of

layer deposition, material purity, dopant placement, alignment precision, etc., will remain a challenge in emerging devices.

V. SYSTEM-LEVEL BENCHMARKS

Many of the novel devices and physical phenomenon in the NRI are at the conceptual stage, i.e., working prototypes do not yet exist and the circuit-level models are based on theoretical devices. Interconnect system models are also in an early state for some of the devices. The performance extrapolations offered for simple logic circuits are indicative of the potential of candidate technologies but should be considered as preliminary, pending further refinement of the models and examination of system-level applications. Benchmarking of these digital devices at the system level is now at a formative stage and it is important that a

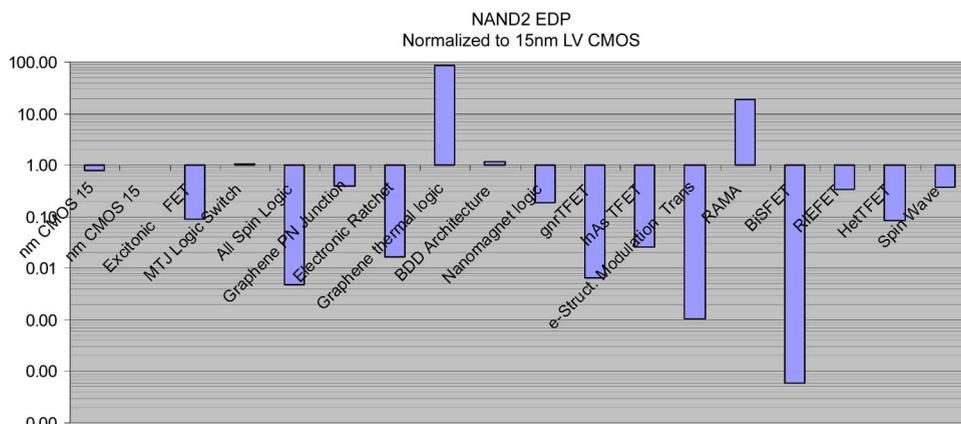


Fig. 12. EDP for proposed devices, normalized to ITRS 15-nm CMOS. (Raw data provided by principal investigators.)

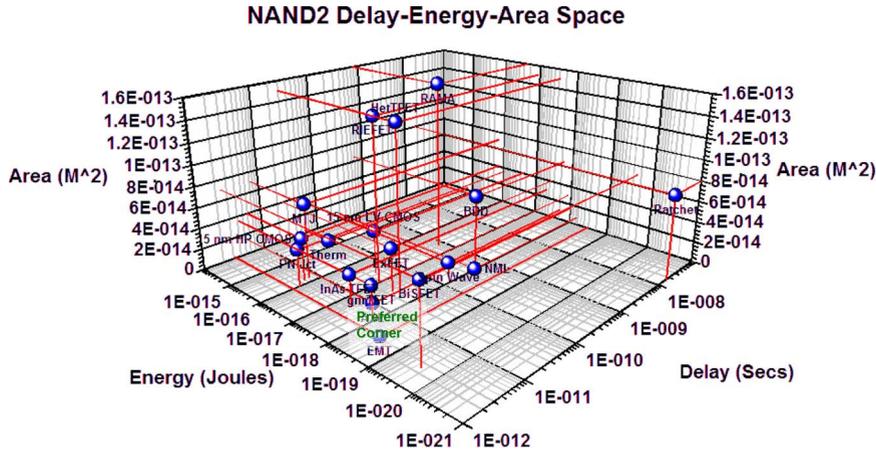


Fig. 13. Delay, energy, and area design space for new switches expressing the NAND2. Both CMOS and TFET estimates include an additional parasitic interconnect capacitances of 1 fF loading each transistor. (Data provided by principal investigators.)

standard system-level model be defined against which to judge the impact of the various proposed technologies. Since much of the work reported in this paper is at the basic research level, it is too early to consider benchmarking at the complexity of a microprocessor.

As an example of system-level benchmarking, in [50], a basic four-instruction, 1-b microprocessor was selected to evaluate performance limits using electronic switches operating at the limit of $k_B T \ln(2)$ joules per switching event for each device. (k_B is Boltzman’s constant and T is temperature in Kelvin.) The switching time per device was chosen to be the Heisenberg time, i.e., $\tau_S = \hbar / (k_B T \ln(2))$, where \hbar is Planck’s constant. The interconnects for the 1-b microprocessor were developed using a few-electron probabilistic model for which energy dissipation was modeled as $k_B T$ per unit gate length. Based on an analysis of the average interconnect length in microprocessors, an

interconnect length of six gate lengths was associated with each device required to realize the microprocessor. Approximately 300 switches were needed to realize a functional processor and it was assumed that the duty cycle for each switch was 50%. Finally, execution of each of the four instructions was assumed to be equally probable. Using a gate length of 1.5 nm, the estimated area required for the 1-b microprocessor was 75 nm × 75 nm and a performance of approximately 10⁵ million instructions per second was projected. Although the estimated power consumption of the microprocessor was very small, the power density estimates were on the order of 10 kW/cm², likely exceeding known heat removal capability.

The work on performance limits described in [50] was motivated by the desire to compare achievable computational efficiency of devices and interconnects at the limits of electron-based information technology with those

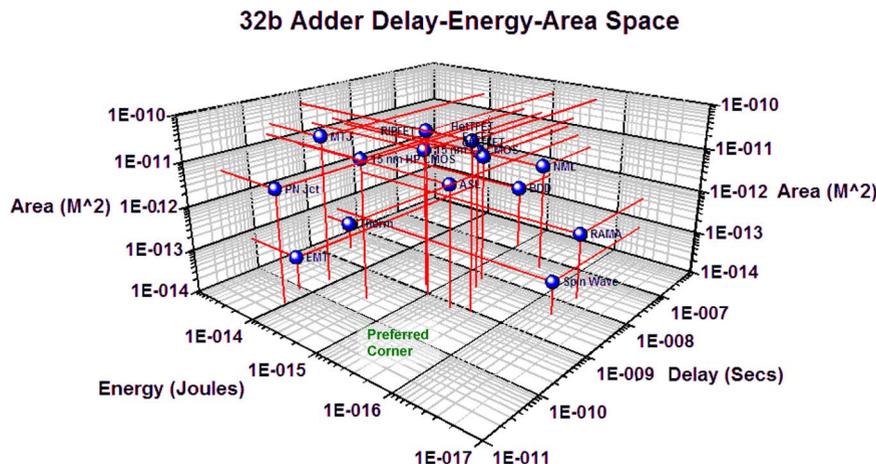


Fig. 14. Delay, energy, and area design space for new devices expressing the 32-b adder. Both CMOS and TFET estimates include an additional parasitic interconnect capacitance of 1 fF loading each transistor. (Data provided by principal investigators.)

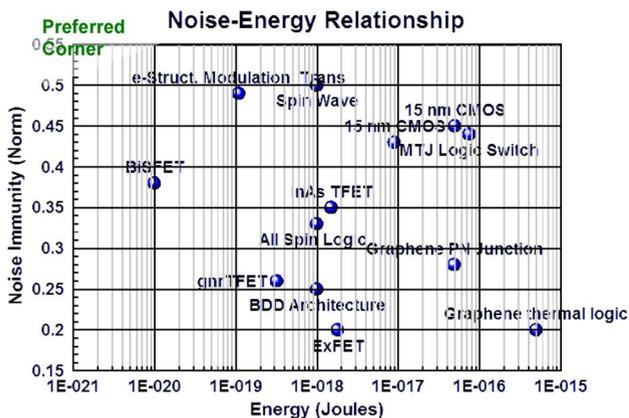


Fig. 15. Noise immunity relationship to switch energy in emerging switches. (Data provided by principal investigators.)

achievable by the brain. Consequently, the system model was used more to establish performance bounds than to offer precise estimates of system performance. For the NRI system benchmarking effort, a similar 1-b microprocessor might serve as a common base system for comparison of various candidate digital switches that are emerging. It will be necessary to describe the required functional behavior of the microprocessor and the constraints on the design that is to be offered. For example, a footprint constraint could be given for the layout area; a power density limit could be set for the layout, and a probability distribution prescribed for execution of the instruction set. A detailed layout for the 1-b microprocessor would be required. The specific topology of the detailed layout would be open but it would be necessary to define the interfaces of the microprocessor to external devices. Since the devices in [50] were chosen to operate at the limits of their reliability, and hence were error-prone, little attention was paid to overall system reliability. In the system benchmarking study however, it is important that the devices and interconnect systems be chosen such that the system operates with a prescribed computation reliability. A careful formulation of this elementary system benchmark will be necessary for a fruitful benchmarking study.

Hybrid CMOS systems could provide an effective utilization of the novel logic devices. In this case, conversion circuitry between domains will need to be considered. Emerging NRI technologies may find applications in domains other than digital processing. At this point, the definition of a benchmark reference system for nondigital applications is premature.

Finally, it should be noted that the vulnerability of each of the proposals to process-induced delay variation, defects, and new failure modes was not assessed in this study. These characteristics are of course critical to successful implementation, and will need to be examined for switches of interest.

VI. GOING FORWARD

The evaluation of proposed replacement switches makes it apparent that adjunct technologies and key concepts must also be considered in order to effectively extend compute power performance. In this closing section, a few of these issues are outlined.

“Span of control” provides a means of relating the relative delay of a switch to its area and to the delay of the transport mechanism it uses to communicate with other switches. Matzke and Bosshart opened the discussion of this issue in the late 1990s [51]. Fig. 16 from that work shows the trend at the time for clock locality, and how higher clock speeds would effectively prevent the entire chip from being reachable within one clock cycle. In part, the movement to multicore processors is in response to this issue. Within the delay of, say, one switch of a particular device type, the number of subsequent devices that a given device can touch or fan out to is a function of 1) the delay of the switch, 2) the propagation delay of the transport mechanism, and 3) the area of the switches. Fig. 17 is a contemporary attempt to convolve these dependencies. In the plot, the delay per switch is plotted on the x-axis against the area per switch on the y-axis. The accessible number of switches is determined by the area per switch (represented by the sizing of the data point circles), as well as by the transport delay of the token conducted by the interconnect, (which is implicit to the data). To support future massive parallelism, it is desirable to implement a technology located as far to the upper left region of this plot as possible. The selection of a future technology depends as much on the effectiveness of the interconnect as it does on the power-performance of the switch.

The impact of interconnect extends also to the total number of inputs and outputs. Rent’s rule is an empirical formula used to relate the number of input/output pins an

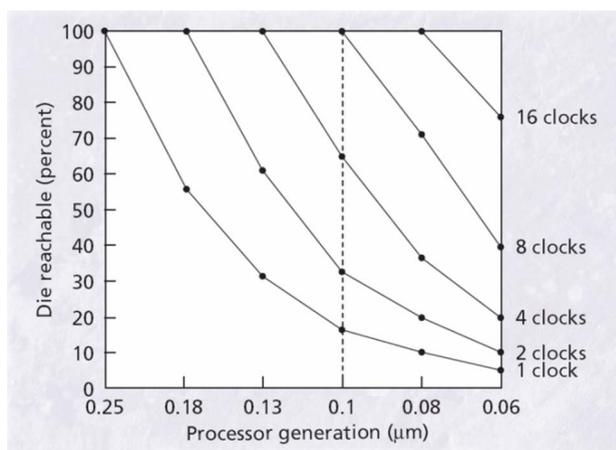


Fig. 16. Span of control, as described by Matzke and Bosshart in 1997 [48].

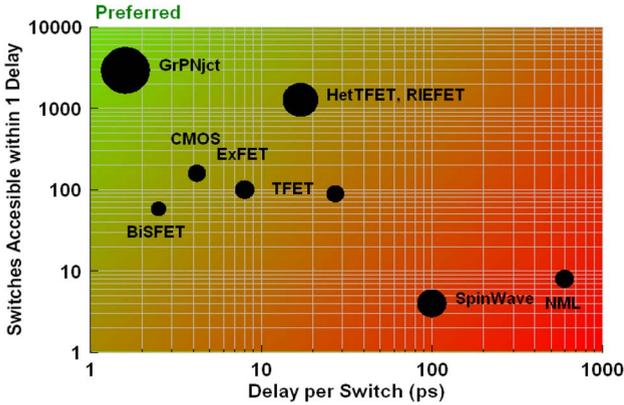


Fig. 17. Transport impact on switch delay, size, and area of control. Circle size is logarithmically proportional to physically accessible area in one delay. (Data provided by principal investigators.)

assembled MOS chip needs to the number of gates or devices used to accomplish the function [52]. Simply stated

$$T = K(N_g)^p \tag{1}$$

where T is the number of terminals, K is Rent's constant, N_g is the number of gates, and p is the Rent exponent. Rent's constants T and p are empirically derived for specific circuit topologies. Rules of thumb are essential to design, and in new switches it will be essential to determine if a Rent's rule variant will continue to describe system requirements.

Another CMOS realm construct, which can be borrowed to describe the effectiveness of nanoscale replacement devices, is that of logical effort (LE). Coined by Ivan Sutherland at Sun Microsystems [53], LE quantitatively captures the effect of circuit topology and device physics upon the ability to produce output, or more concisely, how good a circuit realized in a technology is at evaluating logic. Within CMOS, LE is quoted in reference to an inverter of the same generation. The algorithm computes the number of times worse a given circuit is at driving an output load, compared to a simple inverter with the same amount of input capacitance, by calculating the ratio of a given circuit input's capacitance to that of an inverter delivering the same output current. Proposed switches vary in the tolerance, parasitic overhead, and idiosyncrasy they will encounter. It will be essential to quickly regain LE insights in the emerging devices being proposed. LE values for simple and complex logic circuits built in proposed structures are shown in Fig. 18. The reader is directed to [53] for a more through explanation of this figure of merit.

A final approach with utility in comparing proposed devices, and in discovering potential applications for them is associated with matching the physics and function. By finding idiosyncrasy of the particular device, which complements the desired logic function, it is hoped that further study will reveal specific applications for new devices. A number of existing hardware accelerators currently realized in CMOS, such as encryption, compression, or H.264 filtering, may be improved by pairing them with specific switches, which behave physically in a complementary fashion to the algorithm. Practically speaking, new switches most likely will initially supplement

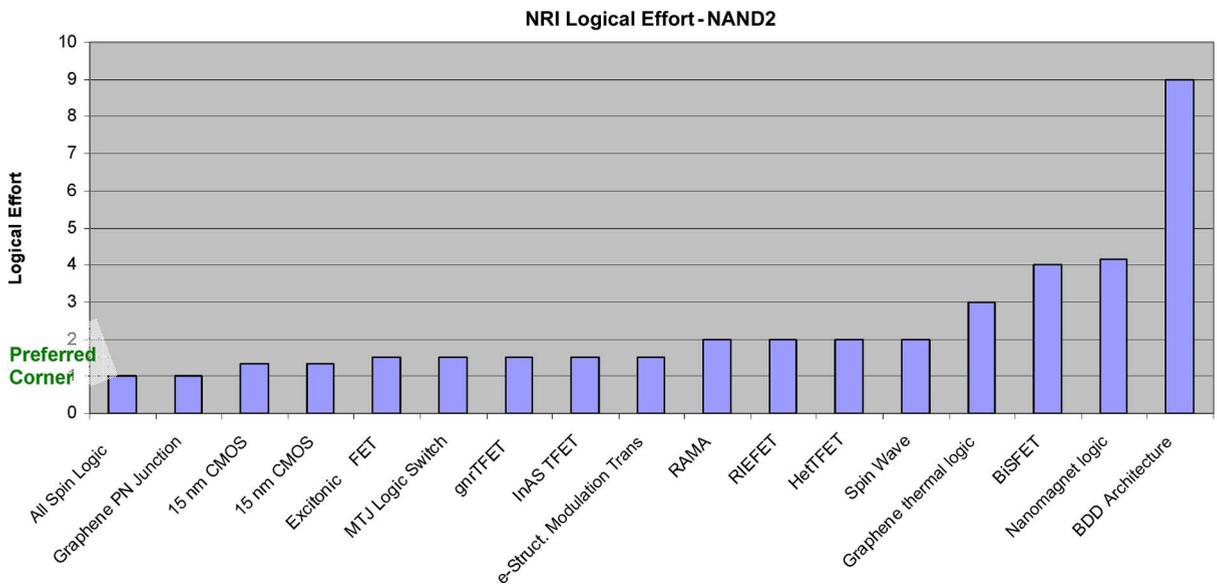


Fig. 18. Estimated logical effort for new switches. (Data provided by principal investigators.)

CMOS, and so necessarily will need to be compatible with CMOS processing (potentially through novel 3-D packaging) and operation. To that end, it is quite likely that first uses for post-CMOS devices will be found in hardware accelerators. In the longer term, these devices may support architectures, which extend the current von Neumann paradigm; or they can lead us into new machine organizations inspired by alternative venues, perhaps cellular or neural in nature.

The semiconductor industry is moving closer to establishing the means of envisioning the structure of next-generation computing machines. With insights from this study, here are some of the considerations.

- 1) The present composition of contemporary high-performance microprocessors is known—the total device count, the number of random logic gates, memory array size, etc.
- 2) Using inference from this study, one can realistically estimate the delay, power, and area of each circuit built in a given new switch. Maximum fan-out, minimum noise immunity, and worst case power density projections form new rules of thumb that can be used to anticipate a workable design point for logic resources mapped into a new switch.
- 3) From ongoing work on the communication transport mechanisms required by various tokens, the delay, power and length distribution profile of the interconnect may be also estimated.

Therefore, it is reasonable to anticipate that a path will soon exist for mapping an entire existing high-performance microprocessor design into proposed new state variables and technologies. These approximations will provide the first credible glimpse into what conventional computing will look like after CMOS; alternative architectures supported by new switches, however, may enable even more profound advances. It is with a sense of purpose and urgency that this work proceeds in universities through the U.S., supported by the Nanoscale Research Initiative. ■

Acknowledgment

The authors would like to thank NRI participants who provided the inputs used this study: M. Baldo and C. Ross (Massachusetts Institute of Technology), S. Koswatta (IBM), S. Datta and V. Narayanan (Pennsylvania State University), C. Augustine, B. Behin-Aein, Y. Chen, S. Datta, K. Roy, and P. Ye (Purdue University), J.-U. Lee (SUNY Albany), J. Bokor and S. Salahuddin (University of California at Berkeley), A. Khitun and D. Markovic (University of California at Los Angeles), M. Flatte (University of Iowa), S. Kurtz, M. Niemier, X. S. Hu, and Q. Zhang (University of Notre Dame), and L. F. Register (University of Texas at Austin). They would also like to thank the anonymous reviewers of this paper for their insightful suggestions.

REFERENCES

- [1] D. Jorgenson, "Moore's law and the emergence of the new economy," Semiconductor Industry Association, Washington, DC, 2005 Annu. Rep., 2005.
- [2] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.
- [3] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design for ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SC-9, no. 5, pp. 256–268, Oct. 1974.
- [4] R. K. Cavin and V. V. Zhirnov, "Silicon nanoelectronics and beyond: Reflections from a semiconductor industry-government workshop," *J. Nanoparticle Res.*, vol. 8, pp. 137–147, 2004.
- [5] R. K. Cavin, V. V. Zhirnov, G. I. Bourianoff, J. A. Hutchby, D. J. C. Herr, H. H. Hosack, W. H. Joyner, and T. A. Wooldridge, "A long-term view of research targets in nanoelectronics," *J. Nanoparticle Res.*, vol. 7, pp. 573–586, 2005.
- [6] R. K. Cavin, V. V. Zhirnov, D. J. C. Herr, A. Avila, and J. Hutchby, "Research directions and challenges in nanoelectronics," *J. Nanoparticle Res.*, vol. 8, pp. 841–858, 2006.
- [7] A. Geim and K. S. Novoselov, "The rise of graphene," *Nature Mater.*, vol. 6, pp. 183–191, 2007.
- [8] M. Y. Han, B. Ozyilmaz, Y. Zhang, and P. Kim, *Phys. Rev. Lett.*, vol. 98, 206805, 2007.
- [9] H. Min, G. Borghi, M. Polini, and A. H. MacDonald, "Pseudospin magnetism in Graphene," *Phys. Rev. B*, vol. 77, 041407(R), 2008.
- [10] H. Min, R. Bistritzer, J.-J. Su, and A. H. MacDonald, "Room temperature superfluidity in graphene bilayers," *Phys. Rev. B*, vol. 78, 121401, 2008.
- [11] S. K. Banerjee, L. F. Register, E. Tutuc, D. Reddy, and A. H. MacDonald, "Bilayer pseudoSpin Field-Effect Transistor (BiSFET): A proposed new logic device," *IEEE Electron Device Lett.*, vol. 30, no. 2, pp. 158–200, Feb. 2009.
- [12] S. Datta and B. Das, "Electronic analog of the electro-optic modulator," *Appl. Phys. Lett.*, vol. 56, no. 7, pp. 665–667, Feb. 1990.
- [13] H. W. Gschwind, *Design of Digital Computers*. New York: Springer-Verlag, 1967.
- [14] R. Cowburn and M. Welland, "Room-temperature magnetic quantum cellular automata," *Science*, vol. 287, no. 5457, pp. 1466–1468, 2000.
- [15] G. Csaba, A. Imre, G. H. Bernstein, W. Porod, and V. Metlushko, "Nanocomputing by field-coupled nanomagnets," *IEEE Trans. Nanotechnol.*, vol. 1, no. 4, pp. 209–213, Dec. 2002.
- [16] A. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod, "Majority logic gate for magnetic quantum-dot cellular automata," *Science*, vol. 311, pp. 205–208, Jan. 2006.
- [17] C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, "Quantum cellular automata," *Nanotechnology*, vol. 4, pp. 49–57, 1993.
- [18] G. Csaba, W. Porod, and A. I. Csurgay, "A computing architecture composed of field-coupled single-domain nanomagnets clocked by magnetic fields," *Int. J. Circuit Theory Appl.*, vol. 31, pp. 67–82, 2003.
- [19] D. B. Carlton, N. C. Emlay, E. Tuchfeld, and J. Bokor, "Simulation studies of nanomagnet-based logic architecture," *Nano Lett.*, vol. 8, pp. 4173–4178, Dec. 2008.
- [20] M. P. Kostylev, A. A. Serga, T. Schneider, B. Leven, and B. Hillebrands, "Spin-wave logical gates," *Appl. Phys. Lett.*, vol. 87, pp. 153501-1–153501-3, 2005.
- [21] A. Khitun and K. Wang, "Nano scale computational architectures with spin wave bus," *Superlattices Microstruct.*, vol. 38, pp. 184–200, 2005.
- [22] S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. G. Requicha, B. E. Koel, and H. A. Atwater, "Plasmonics—A route to nanoscale optical devices," *Adv. Mater.*, vol. 13, p. 1501, 2001.
- [23] K. Song and P. Mazumder, "Active tera hertz (THz) spoof surface plasmon polariton (SPP) switch comprising the perfect conductor meta-material," in *Proc. IEEE Nanotechnol. Conf.*, Genoa, Italy, Jul. 2009, pp. 98–101.
- [24] J. Hu, X. Ruan, and Y. P. Chen, "Thermal conductivity and thermal rectification in graphene nanoribbons: A molecular dynamics study," *Nano Lett.*, vol. 9, no. 7, pp. 2730–2735, 2009, DOI: 10.1021/nl901231s.
- [25] V. Saripalli, V. Narayanan, and S. Datta, "Ultra low energy binary decision diagram circuits using few electron transistors," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 20.

- Berlin, Germany: Springer-Verlag, 2009, pp. 200–209.
- [26] W. Porod et al., “Nanomagnetic logic,” in *Proc. 1st Berkeley Symp. Energy-Efficient Syst.*, Berkeley, CA, Jun. 2009.
- [27] C. Augustine, A. Raychowdhury, Y. Gao, M. Lundstrom, and K. Roy, “PETE: A device/circuit analysis framework for evaluation and comparison of charge based emerging devices,” in *Proc. Int. Soc. Qual. Electron. Design*, pp. 80–85.
- [28] A. Khakifirooz and D. A. Antoniadis, “MOSFET performance scaling—Part I: Historical trends,” *IEEE Trans. Electron Devices*, vol. 55, no. 6, pp. 1391–1400, Jun. 2008.
- [29] K. Galatsis, A. Khitun, R. Ostroumov, K. L. Wang, W. R. Dichtel, E. Plummer, J. F. Stoddart, J. L. Zink, J. Y. Lee, Y.-H. Xie, and K. W. Kim, “Alternate state variables for emerging nanoelectronic devices,” *IEEE Trans. Nanotechnol.*, vol. 8, no. 1, pp. 66–75, Jan. 2009.
- [30] A. Khitun, M. Bao, J.-Y. Lee, K. L. Wang, D. W. Lee, S. X. Wang, and I. V. Roshchin, “Inductively coupled circuits with spin wave bus for information processing,” *J. Nanoelectron. Optoelectron.*, vol. 3, no. 1, pp. 24–34, 2008.
- [31] A. Khitun, M. Bao, and K. L. Wang, “Spin wave magnetic nanofabric: A new approach to spin-based logic circuitry,” *IEEE Trans. Magn.*, vol. 44, no. 9, pp. 2141–2152, Sep. 2008.
- [32] A. Khitun, D. E. Nikonov, and K. L. Wang, “Magnetoelectric spin wave amplifier for spin wave logic circuits,” *J. Appl. Phys.*, vol. 106, 123909, 2009.
- [33] S. Wolf, J. Lu, M. Stan, E. Chen, and D. Treger, “The promise of nanomagnetism and spintronics for future logic and universal memory,” *Proc. IEEE*, 2010.
- [34] M. T. Alam, M. J. Siddiq, G. H. Bernstein, M. Niemier, W. Porod, and X. S. Hu, “On-chip clocking for nanomagnetic logic devices *IEEE Trans. Nanotechnol.*, 2010.
- [35] Q. Zhang, T. Fang, H. Xing, A. Seabaugh, and D. Jena, “Graphene nanoribbon tunnel transistors,” *IEEE Electron Device Lett.*, vol. 29, no. 12, pp. 1344–1346, Dec. 2008.
- [36] H. Chen, L. F. Register, and S. K. Banerjee, “Resonant injection enhanced field effect transistor for low voltage switching: Concept and quantum transport simulation,” in *Proc. Int. Conf. Simul. Semicond. Processes Devices*, San Diego, CA, Sep. 9–11, 2009, DOI: 10.1109/SISPAD.2009.5290247.
- [37] J.-J. Su and A. H. MacDonald, “How to make a bilayer exciton condensate flow,” *Nature Phys.*, vol. 4, pp. 799–802, Oct. 2008.
- [38] S. K. Banerjee, L. F. Register, E. Tutuc, D. Reddy, and A. H. MacDonald, “Bilayer pseudospin field-effect transistor (BiSFET): A proposed new logic device,” *IEEE Electron Device Lett.*, vol. 30, no. 2, pp. 158–160, Feb. 2009.
- [39] J. Appenzeller, *NRI INDEX Annu. Rev.*, Sep. 25, 2009.
- [40] 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.
- [41] S. B. Akers, “Binary decision diagrams,” *IEEE Trans. Comput.*, vol. C-27, no. 6, pp. 509–516, Jun. 1978.
- [42] L. Wang and B. Li, “Thermal logic gates: Computation with phonons,” *Phys. Rev. Lett.*, vol. 99, 177208, Oct. 2007.
- [43] J. Hu, X. Ruan, and Y. P. Chen, “Thermal conductivity and thermal rectification in graphene nanoribbons: A molecular dynamics study,” *Nano Lett.*, vol. 9, no. 7, pp. 2730–2735, 2009.
- [44] V. V. Chelano, V. Fal’ko, and B. Altshuler, “The focusing of electron flow and a Veselago lens in graphene p-n junctions,” *Science*, vol. 315, pp. 1252–1255, Mar. 2007.
- [45] L. W. Martin, Y.-H. Chu, Q. Zian, R. Ramesh, S.-J. Han, S. X. Wang, M. Warusawithana, and D. G. Schlom, “Room temperature exchange bias and spin valves based on BiFeO₃/SrRuO₃/SrTiO₃/Si (001) heterostructures,” *Appl. Phys. Lett.*, vol. 91, 172513, 2007.
- [46] A. De, C. E. Pryor, and M. E. Flatte, “Electric-field control of a hydrogenic donor’s spin in a semiconductor,” *Phys. Rev. Lett.*, vol. 102, 017603, Jan. 2009.
- [47] International Technology Roadmap for Semiconductors, 2008 update. [Online]. Available: http://www.itrs.net/Links/2008ITRS/Update/2008_Update.pdf
- [48] M. Horowitz, T. Indermaur, and R. Gonzalez, “Low-power digital design,” in *Proc. Symp. Low Power Electr.*, Oct. 1994, pp. 8–11.
- [49] M. P. Frank, “The physical limits of computing,” *Comput. Sci. Eng.*, vol. 4, no. 3, pp. 16–26, May/Jun. 2002.
- [50] V. V. Zhirnov and R. K. Cavin, III, “Scaling beyond CMOS: Turing-Heisenberg rapprochement,” in *Proc. ESSDERC-ESSCIRC*, Athens, Greece, Sep. 14–18, 2009, DOI: 10.1109/ESSCIRC.2009.5325930.
- [51] D. Matzke, “Will physical scalability sabotage performance gains?” *IEEE Computer*, vol. 30, no. 9, pp. 37–39, Sep. 1997.
- [52] B. S. Landman and R. L. Russo, “On a pin versus block relationship for partitions of logic graphs,” *IEEE Trans. Comput.*, vol. C-20, no. 12, pp. 1469–1479, Dec. 1971.
- [53] I. Sutherland et al., *Logical Effort: Design Fast CMOS Circuits*, 1st ed. San Mateo, CA: Morgan Kaufmann, Feb. 1999, ISBN: 10:1558605576.

ABOUT THE AUTHORS

Kerry Bernstein (Fellow, IEEE) received the B.S. degree in electrical engineering from Washington University in St. Louis, St. Louis, MO, in 1978.

He is a Research Staff Member at the IBM T.J. Watson Research Center, Yorktown Heights, NY, and has been with IBM for 31 years. He holds 105 U.S. Patents, and is a coauthor of three college textbooks and approximately 50 papers on high-speed logic. His research interests are in the areas of emerging nanodevice/circuit architectures for future high-performance computing; 3-D electronic integration; validating the design integrity of fabricated integrated circuits; and neuro-morphic computing. He is a staff instructor of Computational Neuroscience at the Research update in Neuroscience for Neurosurgeons (RUNN), Woods Hole, MA, and holds the rank of Major in the Vermont State Guard.



Ralph K. Cavin, III (Life Fellow, IEEE) received the B.S.E.E. and M.S.E.E. degrees from Mississippi State University, Mississippi State, in 1961 and 1962, respectively, and the Ph.D. degree 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, Durham, 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 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. He has authored or coauthored over 100 refereed technical papers and contributions to books. 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 served as an advisor to a number of government, industrial, and academic institutions.



Wolfgang Porod (Fellow, IEEE) received the M.S. and Ph.D. degrees in physics from the University of Graz, Graz, Austria, in 1979 and 1981, respectively.

Currently, he is Frank M. Freimann Professor of Electrical Engineering at the University of Notre Dame, Notre Dame, IN. After appointments as a Postdoctoral Fellow at Colorado State University and as a Senior Research Analyst at Arizona State University, he joined the University of Notre Dame in 1986 as an Associate Professor. He now also serves as the Director of Notre Dame's Center for Nano Science and Technology. He has authored some 300 publications and presentations. His research interests are in the area of nanoelectronics, with an emphasis on new circuit concepts for novel devices.

Dr. Porod has served as the Vice President for Publications for the IEEE Nanotechnology Council (2002-2003), and he was appointed an Associate Editor for the IEEE TRANSACTIONS ON NANOTECHNOLOGY (2001-2005). He has been active on several committees, in organizing Special Sessions and Tutorials, and as a speaker in IEEE Distinguished Lecturer Programs.



Jeff Welser (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1995. His graduate work was focused on utilizing strained-Si and SiGe materials for FET devices.

He joined IBM's Research Division at the T.J. Watson Research Center. Since joining IBM, he has worked on a variety of novel devices, including nanocrystal and quantum-dot memories, vertical-FET DRAM, and Si-based optical detectors, and eventually took over managing the Novel Silicon Device group at Watson. He was also working at the time as an Adjunct Professor at Columbia University, teaching semiconductor device physics. In 2000, he took a technical staff assignment to the Sr. VP of IBM Technology Group, and then joined the Microelectronics division in 2001, as Project Manager for the high-performance CMOS device design groups. In May 2003, he was named Director of high-performance SOI and BEOL technology development, in addition to his continuing work as the IBM Management Committee Member for the Sony, Toshiba, and AMD development alliances. In late 2003, he returned to the Research Division as the Director Next Generation Technology Components. He worked on the Next Generation Computing project, looking at technology, hardware, and software components for systems in the 2008-2012 timeframe, and in 2005, the group moved into IBM Systems and Technology Group to focus on developing early system prototypes. In 2006, he returned to the Research Division and was named the Director of the Nanoelectronics Research Initiative (NRI). In this position, on assignment to the Semiconductor Research Corporation (SRC), he directs university-based research on future nanoscale logic devices to replace the CMOS transistor in the 2020 timeframe and is based at the IBM Almaden Research Center, San Jose, CA.



Alan Seabaugh (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1985.

His current research interests include tunneling and low-power devices and circuits, nanofabrication, and energy scavenging. He joined the University of Notre Dame, Notre Dame, IN, in August of 1999 as Professor of Electrical Engineering following positions at the National Bureau of Standards (1979-1986), Texas Instruments Incorporated (1986-1997), and Raytheon Systems Company (1997-1999). He was named TI Distinguished Member of Technical Staff in 1997 and Raytheon Senior Fellow in 1999. He has authored/coauthored more than 200 papers and holds 22 U.S. patents. He received outstanding teacher awards from the University of Texas at Dallas and the University of Notre Dame in 1990 and 2001, respectively. He is Director of the Semiconductor Research Corporation, Nanoelectronics Research Institute (SRC-NRI) Midwest Institute for Nanoelectronics Discovery (MIND) and Associate Director of the Notre Dame Center for Nano Science and Technology.



Dr. Seabaugh is a member of the American Physical Society.