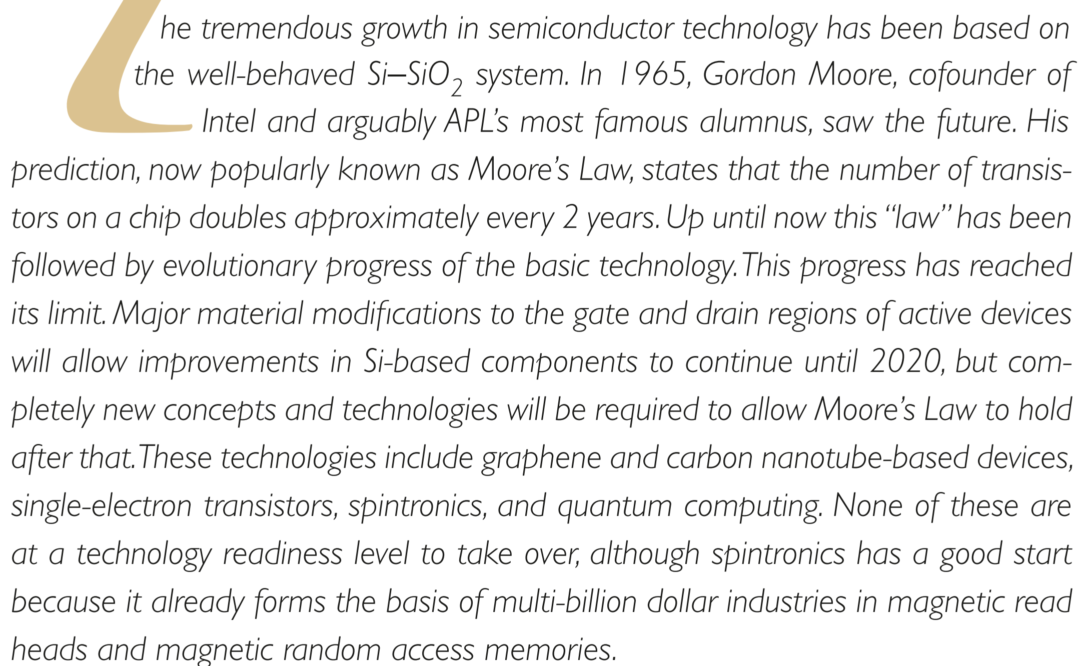
Semiconductor Devices: Moore Marches On

Dennis K. Wickenden



The tremendous growth in semiconductor technology has been based on the well-behaved Si–SiO_2 system. In 1965, Gordon Moore, cofounder of Intel and arguably APL's most famous alumnus, saw the future. His prediction, now popularly known as Moore's Law, states that the number of transistors on a chip doubles approximately every 2 years. Up until now this "law" has been followed by evolutionary progress of the basic technology. This progress has reached its limit. Major material modifications to the gate and drain regions of active devices will allow improvements in Si-based components to continue until 2020, but completely new concepts and technologies will be required to allow Moore's Law to hold after that. These technologies include graphene and carbon nanotube-based devices, single-electron transistors, spintronics, and quantum computing. None of these are at a technology readiness level to take over, although spintronics has a good start because it already forms the basis of multi-billion dollar industries in magnetic read heads and magnetic random access memories.

INTRODUCTION

The first known semiconductor devices were based on "cat's whiskers" and followed the discovery in 1874 by Karl Braun of the rectification of current by a metal point in contact with a galena (lead sulfide) crystal.[1] This discovery was expanded upon in 1907, when Captain Henry Round, a personal assistant to Guglielmo Marconi, while working with carborundum (silicon carbide) crystals, noticed that some of them emitted light while current was passing through them.[2] This is the first known report of the observation of a light-emitting diode in action. These detectors were very finicky to operate, requiring the operator to move a small phosphor bronze or tungsten filament (the whisker) around the surface of a suitable crystal until it suddenly started working. Then, over a period of a few hours or days, the cat's whisker would slowly stop working, and the process would have to be repeated. At the time, their operation was a complete mystery. This remained the case until

Walter Schottky formulated the theory of metal-semiconductor junctions.[3]

The operation of a surface field-effect transistor (FET; forerunner of today's MOSFET) was first proposed in the 1930s by Lilienfeld.[4] These devices were based on compound semiconductors such as copper sulfide and the oxides of copper, lead, and vanadium. Their characteristics were very irreproducible, most probably because of the impurities and defect densities in the material. During World War II, Bell Telephone Laboratories had a contract to produce very pure germanium crystals for Schottky barrier mixer diodes used in early microwave radar receivers. In 1947, William Shockley, John Bardeen, and Walter Brattain[5] succeeded in building the first practical point-contact transistor with this material. The announcement set off a frenzy of activity worldwide. It was soon realized that germanium had some practical problems (in particular, the devices had the bad habit of not working at high temperatures). It was thought that silicon, which is above germanium in the periodic table, might be a better candidate, and work started on developing this. However, it took several years to produce pure enough material, and it was not until 1954 that Gordon Teal, of Texas Instruments, succeeded in producing the first silicon transistor.[6]

The next breakthrough came in 1958 when Jack Kilby, also working for Texas Instruments, demonstrated the first "integrated" circuit, which consisted of a simple oscillator circuit composed of a single germanium transistor and a few feedback resistors and capacitors. In that same year, Robert Noyce, of Fairchild Semiconductor, developed the monolithic integrated circuit (IC) in silicon. Noyce's colleague, Jean Hoerni, then took the idea a step further and put a collector, base, and emitter all on one plane on the substrate surface. Finally, in 1960, Kahng and Atalla of Bell Telephone Laboratories reported the first demonstration of an Si–SiO_2 metal-oxide-semiconductor (MOS) transistor.[7] Thus, practical planar junction and FETs were born, and with them, the modern semiconductor industry. This industry has blossomed into one of the wonders of the modern age that has allowed progress on all fronts and has become a pervasive part of every aspect of modern postindustrial society.

SILICON DEVICES

The physics and technology of semiconductor devices are well covered in the literature.[8,9] In principle, such devices are independent of the semiconducting material used to fabricate them. However, silicon stands alone as the giant in the field and has been responsible for the explosive growth in the semiconductor device field. Its dominance is attributable to several important material and technological properties,[10] which include the following:

- Its band gap of 1.12 eV is small enough to allow for reasonably small built-in voltages and yet large enough to prevent significant thermal generation of minority carriers in reverse-biased *p*–*n* junctions.
- It is one of the easiest materials to dope. The usual donor (P, As) and acceptor (B) atoms have solubilities well in excess of 10^{20} atoms per cm^{-3}, allowing very high carrier densities and low-resistivity material to be obtained. In addition, these dopants have very low diffusion coefficients at temperatures <800°C.
- Its high thermal stability is compatible with requirements for efficient device processing and subsequent stable device operation.
- The electron and hole mobilities of pure silicon at room temperature are 1450 and 450 $cm^2 \cdot V^{-1} \cdot s^{-1}$, respectively; this is more than adequate for most high-speed devices, which are limited by the transit time across either a base or channel.
- Its indirect gap makes it relatively hard for minority carriers to be lost to recombination by themselves or by deep traps introduced by chemical impurities or crystalline defects. This results in minority carrier lifetimes in excess of 100 μs, which corresponds to electron diffusion lengths of >0.5 μm.
- Its thermal conductivity of 1.5 $W \cdot K^{-1} \cdot cm^{-1}$ is the highest of all commercial semiconductors. This is an important consideration for power devices and modern-day high-density ICs.

In addition to the above advantageous properties of bulk silicon, possibly the most important characteristic is the availability of a high-quality thermally grown oxide (SiO_2) and an almost perfect Si–SiO_2 interface. The latter results in surface state densities as low as 10^{10} cm^{-2}, which is comparable to the atomic bond density and allows the ready control of the depletion layer in MOS devices. Furthermore, thermal SiO_2 has a high dielectric breakdown strength (10^7 $V \cdot cm^{-1}$), which allows gate voltages of a few volts to be applied across

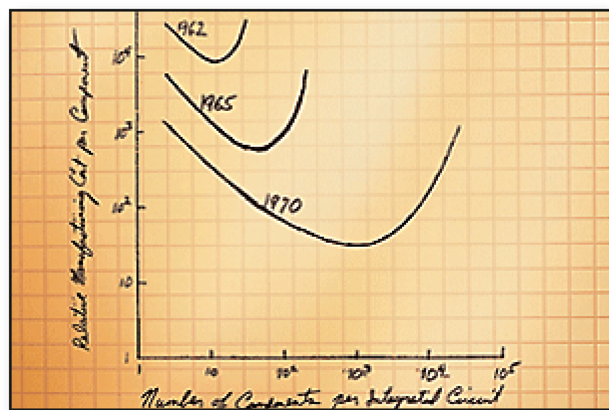


Figure 1. Gordon Moore's original graph from 1965. (Reproduced with permission from Intel, http://www.intel.com/technology/mooreslaw/index.htm.)

very thin layers. Another key behavior in SiO_2 is that the diffusion coefficients of common dopants are much lower than in Si itself. As a result, the oxide acts as a diffusion barrier, and locally doped areas of silicon can be produced by oxidizing its surface, etching windows in the oxide, and diffusing or implanting in *p*-type or *n*-type dopants from gaseous sources into the bulk. These processes are the basis of planar technology, which more than any other factor has contributed to the expansion of the semiconductor industry.

In summary, the miracle of semiconductor technology is firmly founded on the excellent properties of the Si–SiO_2 system. Because the noise characteristics are generally inferior to bipolar transistors, silicon MOSFETs are the dominant devices for most complex digital circuits. Bipolar transistors are used for most analog devices and circuits, power devices, and digital circuits requiring very high-speed operation.

MOORE'S LAW

In 1965, Gordon Moore, Intel cofounder and arguably APL's most famous alumnus, plotted the past and immediate future progress in circuit complexities. His original graph is reproduced in Fig. 1.[11] From these data, he predicted that the number of transistors on a chip would double approximately every 2 years. His prediction, popularly known as Moore's Law, holds true to this day. This sustained progress in circuit complexity has been fueled by the intense competition in the semiconductor industry and the concomitant drive for faster performance and cost-effective processing. Such progress has caused the industry to become one of the wonders of the modern technological age. It has allowed progress on all fronts, and its products have become a pervasive staple to every aspect of society. Two metrics of its evolutionary progress are minimum feature and wafer sizes. The first allows more devices to be packed into the same area, and the latter allows more circuits to be processed per wafer. Graphs of the minimum feature size (in nanometers) and silicon wafer diameter (in millimeters) developments over time are shown in Fig. 2. The bulk of current technology is based on a 90-nm process using 300-mm-diameter wafers. A photograph of two as-grown silicon ingots used to produce the 300-mm-diameter wafers, together with a view of the Czochralski growers, is shown in Fig. 3. Each is ~1.5 m long (excluding the tapered regions) and weighs ~275 kg. This is to be compared with a 10-cm-long, 25-mm-diameter crystal grown by the author in the late 1960s.

The sustained applicability of Moore's Law as represented by the development of the different Intel microprocessors is reproduced in Fig. 4. This figure starts with the 4004 processor and ends with the Dual-Core Itanium 2 processor. Over the same period, the price of the average transistor in a microprocessor has dropped from $1 per unit to less than $0.00000001 per unit, which does not come about without some cost penalty. Capital equipment charges have been rising drastically as the specifications for producing smaller and smaller devices over larger areas has increased. For example, Intel has recently announced that it is currently building two factories in green-field sites that will use the 45-nm process at a cost of over $3 billion each and retooling another from the now-obsolete 180-nm process to 45-nm manufacturing at a cost of approximately $1.5 billion.

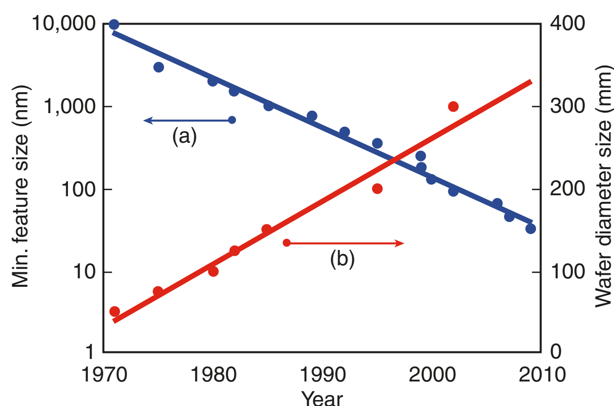


Figure 2. Minimum feature size (blue) (a) and wafer diameter (red) (b) in semiconductor processing as a function of time. The curves are guides to the eye.

Figure 3. Photograph of as-grown 300-mm-diameter silicon ingots with a Czochralski puller in the background. (Reproduced with permission from MEMC Electronic Materials, Inc.)

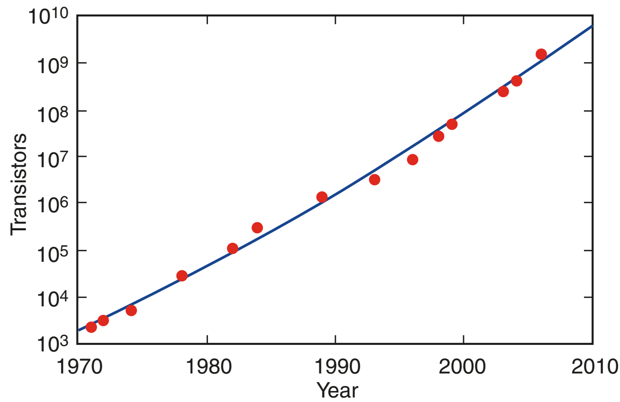


Figure 4. Moore's Law as applied to Intel processors. The curve is a guide to the eye.

APL had a semiconductor device processing line in the late 1970s based on 2-inch-diameter substrates, which was more than adequate to meet APL's custom IC requirements at that time. However, as the technology continued to expand and older equipment was not supported by the vendors, it proved more and more difficult to keep the facility online. The cost to upgrade the equipment to 4-inch wafer processing in 1985 (probably less than $500,000 in 1985 dollars) proved prohibitive, and in-house front-end processing ceased. This did not mean the end of custom IC design and fabrication for APL programs. To cover the costs of these expensive upgrades, many manufacturers opened their lines to foundry services that enabled the small-volume user to produce limited numbers of wafers at a time. On an even smaller scale, MOSIS continues to provide a low-cost prototyping and small-volume production service for very large-scale-integration (VLSI) circuit development (http://www.mosis.org/). Since 1981, MOSIS has successfully fabricated >50,000 circuit designs for commercial firms, government agencies, and research and educational institutions around the world. MOSIS provides designers with a single interface to the constantly changing technologies of the semiconductor industry with mask generation, wafer fabrication, and device packaging being contracted to leading industry vendors. MOSIS also keeps the cost of fabricating prototype quantities low by aggregating multiple designs onto one mask set, which allows customers to share overhead costs associated with mask making, wafer fabrication, and assembly. MOSIS offers minimum (e.g., 40 die) and medium (500 die, 2000 die, etc.) quantities within the regularly scheduled multiproject runs.

THE END OF THE Si–SiO_2 SYSTEM

Almost every year since the early 1970s, people have predicted that the limits of performance of silicon devices would soon be reached and extolled the virtues of devices based on compound semiconductors such as GaAs and InP. Every year, silicon process engineers and scientists and equipment vendors have proved them wrong. However, it is now recognized that the limits of the Si–SiO_2-based MOSFET technology have finally been reached with the 65-nm process. Here, the SiO_2 gate is only 1.2 nm thick, which corresponds to a thickness of five monolayers of material. This is thin enough that some current leakage is inevitable and, as such, represents a major source of power dissipation. Hence, large microprocessors run very hot, and their performances are compromised. Thermal management has become a major part of system design. Any further scaling down of the SiO_2 gate thickness would lead to completely unacceptable gate leakage. However, it has been a challenge to find a suitable replacement that has good compatibility with both the silicon substrate and the polysilicon gate electrode. The effort to find a replacement material has gone on for >10 years. While hafnium and zirconium oxides have good dielectric properties, neither of these compounds is compatible with the polysilicon material used for gate electrodes. Both pin the Fermi level at the dielectric/polysilicon interface, and their inherent polarization characteristics result in phonon-assisted scattering of carriers in the channel. Recently, both Intel[12] and IBM have announced solutions to the problem and are ramping up production of their 45-nm processes. Both organizations have reportedly identified a new hafnium-based "high-*k*" material to replace the MOSFET's SiO_2 gate dielectric, as well as new metals to replace the polysilicon gate electrode of NMOS and PMOS devices. Cross-sections of these new gate structures are illustrated in Fig. 5. The polysilicon is added to the top of the new metal electrodes to make the new technology "invisible" to standard processing. These new materials reduce gate leakage by >100-fold, while delivering record transistor performance. They also ensure that process evolution will lead to the development of 32-nm and 22-nm processes and that Moore's Law will march on into the 2020s. According to Moore himself, "The implementation of high-*k* and metal materials marks the biggest change in transistor technology since the introduction of polysilicon gate MOS transistors in the late 1960s."[13]

Silicon dioxide also is being replaced as the inter-metal dielectric. As transistors become smaller, the connections between them can have a greater influence on the circuit performance than the devices themselves. Interconnect performance is determined by the product of the resistance of the metal lines and the capacitance of the inter-metal dielectric. The resistance is being lowered by replacing aluminum lines with copper ones, and the capacitance is being reduced by replacing the SiO_2 insulator with low-*k* materials. Various approaches are currently being pursued to lower the dielectric constant. These approaches include doping SiO_2 with fluorine to produce fluorinated silica glass, developing methods to

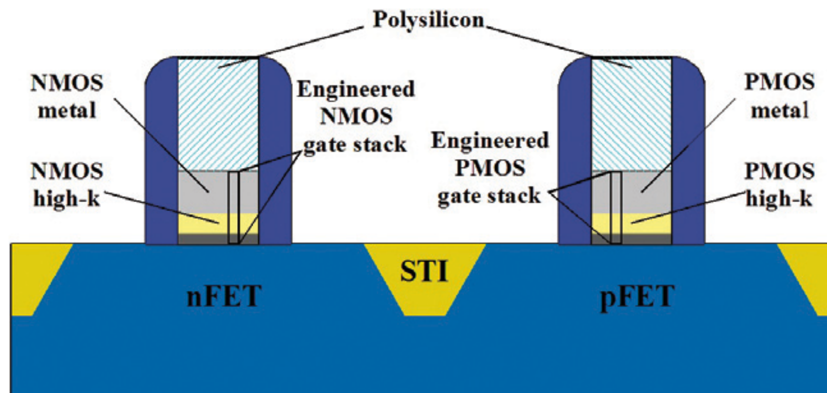


Figure 5. Cross-sections of Intel-engineered high-*k* dielectrics and metal gate stacks. (Reproduced with permission from Intel, http://www.intel.com/technology/index.htm?iid=tech_sil+45nm.) STI, shallow trench isolation.

create large voids or pores in the SiO_2, and developing spin-on organic dielectrics.

Just as the performance of MOSFETs is limited by gate leakage, the performance of bipolar junction transistors is limited by the injection of holes at the emitter-base junction. This limitation becomes nontrivial as the base width is reduced for high-frequency device applications, as became apparent in the mid- to late 1990s when demand for high-performance RF devices was picking up in cellular and other wireless applications; thus, silicon engineers started to address the problem. As a result, heterojunction base transistors, in which the base material is an SiGe alloy, were developed.[14] SiGe has a lower energy gap than Si and, as a result, presents a larger barrier for the injection of holes at the emitter-base junction and allows thinner bases to be used. Because SiGe alloys naturally have a larger lattice constant than Si, care must be taken in ensuring that the deposited layer is pseudomorphic. As such, the SiGe is forced to have the same lattice constant as the silicon substrate, is highly strained, and has an even lower energy gap. If it is thicker than a critical thickness, it will relax to its natural crystal structure and generate deleterious misfit dislocations at the interface. The SiGe base can either have a constant composition or be graded across the base to create an accelerating electric field for the minority carriers moving across the base.[15] Thanks to SiGe's substantial performance benefits, it has quickly become the technology of choice for both wireless ICs and low-power RF chips. There is a slight cost penalty involved for standard processing because of the extra low-temperature epitaxial deposition of the SiGe. Current devices now demonstrate f_T/f_{max} numbers well above 200 GHz and associated gain values >10 dB.[16] The 130-nm IBM SiGe process is available through MOSIS.

FUTURE TECHNOLOGIES

The quest for faster and cheaper devices has been satisfied by the continuous miniaturization of silicon-based transistors. As outlined above, this device scaling and performance enhancement cannot continue forever, and the end may finally be in sight for silicon-based technologies. The need to continue the improvement in performance has led to a worldwide effort in developing possible alternative device technologies. Some of these approaches involve continuing the miniaturization route by using alternative semiconducting material systems such as carbon nanotubes (CNTs), graphene layers or single-electron transistors (SETs), while others are moving toward completely different concepts as in spintronics and quantum optics.

CNT and Graphene Devices

Carbon, like silicon and germanium, is a Group IV element that exists in several allotropic forms. The cubic form, diamond, is a well-known wide band-gap semiconductor. Graphite is composed of weakly coupled layers of sp^2-bonded carbon atoms in a honeycomb arrangement. This weak coupling produces a small valence and conduction band overlap of ~40 meV at the *K* and *K*′ points, which makes graphite a semi-metal.[17] This overlap is removed in single layers of graphite (graphene), and the density of states is zero where the bands touch, which makes graphene a zero-gap semiconductor. The single-particle band structure of graphene results in electrons and holes having zero effective mass and a velocity that is ~300 times slower than that of light. The linear dispersion relationship also means that quasi-particles in graphene display properties quite different from those observed in conventional three-dimensional materials, which have parabolic dispersion relationships. For example, graphene displays an anomalous quantum Hall effect and half-integer quantization of the Hall conductivity. The quantum Hall effect in graphene can be observed even at room temperature.

Graphene layers are usually produced by exfoliation from bulk highly oriented pyrolytic graphite (HOPG) by a simple but effective procedure. A freshly cleaved surface of an HOPG crystal is rubbed against another surface, leaving a variety of flakes attached to it. Virtually any solid surface is suitable, but usually one suited to semiconductor processing is used. This rubbing process can be described as similar to "drawing by [using] chalk on a blackboard."[18] Individual flakes contain different numbers of graphite layers, but all samples apparently produce discrete areas of graphene large enough for further study. These areas are easily identified by phase-contrast microscopy and subsequent atomic force microscopy analysis. Figure 6 shows a schematic diagram of the structure of a graphene layer.

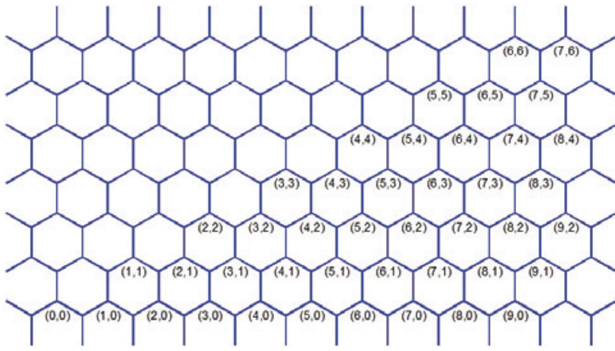


Figure 6. Schematic diagram of the structure of graphene. Nanotubes are formed by rolling a sheet of graphene along a chiral vector, C_h, defined from the point (0,0) to one labeled (*n*,*m*).

CNTs are formed by rolling a sheet of graphene along a chiral vector, C_h, defined from the point (0,0) to one labeled (*n*,*m*). Thus, to produce a nanotube with the indices (8,4), the sheet is rolled up so that the atom labeled (0,0) is superimposed on the one labeled (8,4). There are three distinct ways in which a graphene sheet can be rolled into a tube, as indicated in Fig. 7, which shows end-on views of nanotubes with indices (8,8), (8,0), and (8,4). The first two of these, known as "armchair" and "zigzag" because of the arrangement of hexagons around the circumference, have high degrees of symmetry. The third class of tube, which is the most common, is known as chiral, meaning that it can exist in two mirror-related forms. Figure 7 shows so-called single-walled CNTs (SWCNTs). Multi-walled CNTs consist of bundles of concentric SWCNTs.

The periodic boundary conditions around the circumference of a nanotube require that the component of the momentum along the circumference, $k_\perp$, is quantized: $C_h k_\perp = 2\pi\nu$, where ν is a non-zero integer. On the other hand, electron motion along the length of the tube is free, and $k_{||}$ is a continuous variable. The quantization of $k_\perp$ leads to the formation of a set of discrete energy sub-bands for each nanotube. The relation of these sub-bands to the band structure of graphene determines the electronic structure of the nanotube. If the sub-bands pass through the *K* or *K*′ points, the nanotube is a metal; if they do not, the nanotube is a semiconductor. Specifically, armchair tubes are always metallic, and (*n*,*m*) nanotubes with $n - m = 3j$, where $j = 1, 2, 3 \ldots$, are nearly metallic with a small, curvature-induced gap that has a $1/d^2$ dependence. Tubes with $n - m \neq 3j$ are semiconductors.

These favorable characteristics are the driving force behind the extensive efforts worldwide to develop CNT-based electronics. The structure is similar to conventional devices in that the source and drain are at either end of the nanotube and conduction along the tube is controlled by a gate electrode along part of the tube but isolated from it by an insulating layer.

CNTs are generally grown by chemical vapor deposition. Being a three-dimensional structure, it is difficult, if not impossible, to nucleate them on a two-dimension lattice as in normal epitaxial growth. It is therefore necessary to "seed" the substrate on which the nanotubes are to be deposited with small iron-containing particles. These particles act as nucleation sites, and pyrolysis of, for example, methane at 900°C allows the production of a host of nanotubes. In general, there is no control of either the orientation to the substrate or the chirality of the individual nanotubes. The resultant random mixture of metallic and semiconducting nanotubes is examined in a scanning electron microscope (SEM), and any potentially useful ones are metallized *in situ* by using electron-beam lithography. This type of processing, although suitable for research, does not lend itself easily to production techniques.

Recently, a novel growth technique has been developed that allows for the production of dense, perfectly aligned arrays of SWCNTs. These features, together with the ability of these devices to provide both *p*-type and *n*-type operation with minimal process modification, and their compatibility with a range of substrates suggest that these approaches have some promise for realistic SWCNT-based electronic and optoelectronic technologies.[19] The technique involves the use of S- or AT-cut quartz wafers that are annealed in air at 900°C for long periods of time to produce atomic steps on the $(01\bar{1}0)$ planes. An array of fine lines of iron then are delineated on such a substrate that when heated in air to 550°C form iron oxide nanoparticles bound to the quartz steps. The particles serve as the catalytic seeds for chemical vapor deposition growth of the SWCNTs, which preferentially propagate along the steps in the $(2\bar{1}\bar{1}0)$ direction via an unspecified weak bonding scheme. More than 99.9% of the SWCNTs were parallel, to within 0.01°, with perfectly linear configurations. The simplest method to process the material is to define source and drain electrodes on the SWCNT/quartz substrates in regions between the catalyst stripes. Etching SWCNTs outside of the channel region, spin-casting a uniform epoxy gate dielectric,

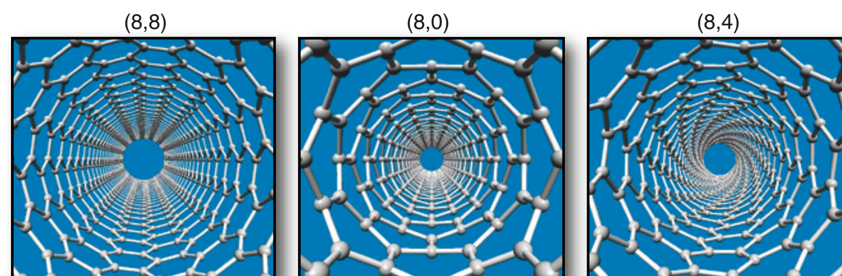


Figure 7. Illustration of the three types of SWCNTs with indices (8,8) (left), (8,0) (center), and (8,4) (right).

and defining top gate electrodes aligned to the channel regions yielded arrays of electrically isolated transistors. An SEM image of the channel region of such a device is shown in Fig. 8.

This technique has tremendous promise. Quartz undergoes a phase change at just below 600°C. To prevent cracking in the quartz, it is necessary to cool the samples slowly (<5°C•min^{-1}). Once the SWCNT arrays are processed into devices, it is relatively easy to transfer them to more conventional substrates.[20] In addition, there still is no control of the chirality of the SWCNTs, and each device consists of a mixture of metallic and semiconducting nanotubes. As produced, the presence of the metallic tubes results in the transistors' on/off ratios being small. However, these can be improved by several orders of magnitude by pinching off the semiconducting tubes to render them nonconducting, heating the metallic tubes by passing suitably high currents through them, and burning them off.

What is the future? CNTs have already provided us with an ideal model system to study electrical and optical phenomena on the nanometer scale. One-dimensional material with their exotic properties, long the realm of theoretical studies, is now open to experimentation. Graphene is a novel, covalent two-dimensional system that already has been found to exhibit a number of unique phenomena. There is no doubt that in the future we will continue to obtain new information on the physics of the nanoscale through the study of nanotubes and graphene. Nanotube and graphene research also is teaching us how to handle and process nanomaterials and develop nanotechnology in general. Nanotubes offer the potential of very fast (terahertz) transistors, ultimately scaled logic devices, and simpler and cheaper self-assembly-based fabrication. In addition, transistors with properly functionalized CNTs can and are already being used as sensitive and selective chemical sensors and biosensors. CNT-based nano-light sources and detectors may allow intrachip optical communications and individual molecule-level spectroscopy. The excellent electrical conduction of metallic CNTs may eventually allow the development of electronic systems where both active devices and interconnects are based on the same material: CNTs. Further integration to include optics could lead to a unified electronic–optoelectronic technology.[17]

Single-Electron Transistors

The switching action of even the smallest MOSFET is achieved by turning the channel region on or off by the voltage applied to the gate. The device operation can be described by using classical solid-state physics. However, when the dimensions of the active regions are shrunk further, the electrons are confined to a small volume and quantum mechanical effects become significant. In the limit, when these electrons communicate with the electrodes by tunneling, one has a device that can turn on and off every time a single electron is added to it. Various structures have been made in the past two decades in which electrons are confined to small volumes in metals or semiconductors. Perhaps not surprisingly, there is a deep analogy between such confined electrons and atoms. Whereas natural atoms are studied by adding, removing, or exciting electrons with light, these artificial atoms typically have such small energy scales that they are best studied by measuring the voltage and current resulting from tunneling between the artificial atom and nearby electrodes.[21]

A schematic of one kind of SET is shown in Fig. 9. It consists of a semiconductor, in this case GaAs, separated from metal electrodes by a thin layer of AlGaAs. The AlGaAs is doped with Si, which donates electrons to the two-dimensional electron gas (2DEG) at the

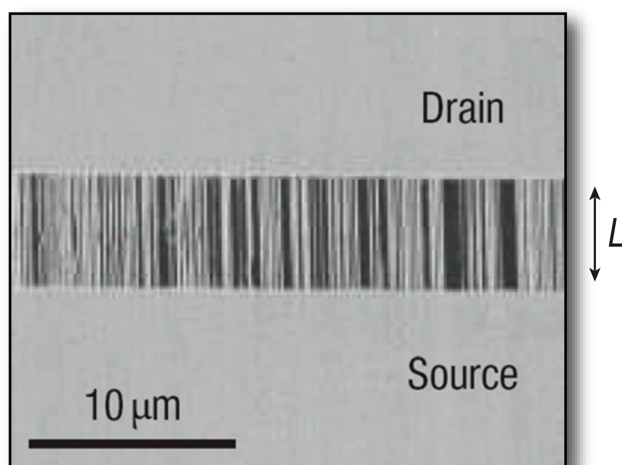


Figure 8. SEM image of the channel region of an SWCNT transistor. The distance between the source and drain electrodes defines the channel length (*L*). (Reproduced with permission from Ref. 19, © 2007, Macmillan Publishers Ltd.)

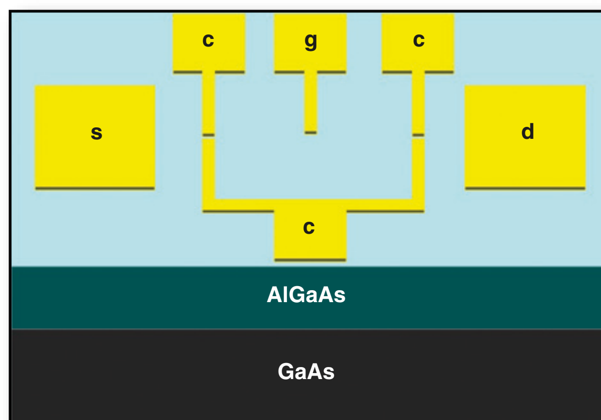


Figure 9. Schematic diagram of an AlGaAs/GaAs-based SET. The source, drain, and gate are designated "s," "d," and "g," respectively. The electrodes labeled "c" confine the electrons to a small volume (adapted from Ref. 21).

AlGaAs/GaAs interface. The 2DEG is confined perpendicular to the GaAs/AlGaAs interface and is further confined in the other two directions with additional confinement electrodes. The lateral dimensions of the confined electrons are in the 10- to 100-nm range. The voltage between the source and drain allow electrons to flow in the circuit through the 2DEG. The gate voltage controls the potential of the electrons confined in the resultant potential well. When this voltage is increased, the potential minimum in which the electrons are trapped becomes deeper and increases the capacity of the trapped electrons in the well. However, unlike a conventional transistor, in which the charge increases continuously, the charge in the trap increases in discrete steps, which is reflected in the conductance between source and drain.

Figure 10 shows the conductance as a function of gate voltage for a typical SET.[22] The conductance is measured by applying a very small voltage between the drain and source, small enough that the current is in the linear regime. As seen in the figure, the conductance increases and decreases by several orders of magnitude almost periodically in the gate voltage. A calculation of the capacitance between the gate electrode and the droplet of confined electrons shows that the voltage between two peaks or two valleys is just that necessary to add one electron to the droplet. Hence, the device is called a "single-electron transistor."

There are strong efforts around the world to make the artificial atoms in SETs smaller in order to raise the temperature at which charge quantization can be observed. In fact, the observation of rudimentary room-temperature characteristics with 8-nm tungsten islands deposited by a focused ion-beam deposition technique has been published recently.[23] Current SET fabrication techniques produce devices on a "one-off" basis. Much more fundamental engineering development work is required before SET technology can be considered for practical applications. It is important to realize that, as SETs get smaller, all of their energy scales will be larger, and so it is likely that potentially important new phenomena will be reported.

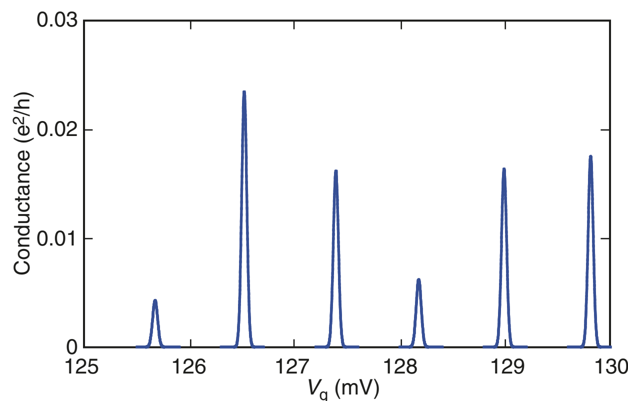


Figure 10. Conductance as a function of gate voltage for a typical SET (adapted from Ref. 21).

Spintronics

Spintronics is an emerging technology based on the quantum spin properties of electrons rather than on the charge properties. When the intrinsic spin of an electron is measured, it is found in one of two spin states, which are denoted as spin "up" or spin "down." Spintronics was born in 1988 when France's Albert Fert and Germany's Peter Grünberg independently discovered the giant magnetoresistance (GMR) effect.[24,25] In this effect, very weak changes in magnetism generate larger changes in electrical resistance. It is observed in a multilayer sample consisting of two different ferroelectric layers. The magnetic axes of the two ferromagnetics can be either parallel or anti-parallel to each other, depending on the strength of an applied magnetic field. The magnetic axis of the electron connected with its spin lines up with this axis. When all of the layers are aligned in the same direction, say "up," electrons with the same alignment pass through the material easily, whereas those aligned "down" encounter a barrier. Importantly, when the layers are organized in an alternating "up–down" alignment, all electrons encounter resistance, independent of their spin. The net effect is an increase in resistance that is much bigger than anything seen before (hence the term "giant"). This discovery earned Grünberg and Fert the 2007 Nobel Prize in Physics.

The most important spintronic device based on the GMR effect is the spin valve. It consists basically of a three-layer stack of ferromagnetic/metal/ferromagnetic, as shown in Fig. 11a. In the actual device an additional layer of an anti-ferromagnetic is used to pin the magnetic axis of one of the ferromagnetic layers by exchange anisotropy so that the other is free to respond to the external magnetic field, as depicted in Fig. 11b. IBM introduced such a device as the read head in computer hard drives in 1997. Advances in spin-valve technology are the key to the explosive increase in the capacity of mass-storage media. Significant improvement in the performance of spin valves was achieved by replacing the metal layer in a spin-valve structure with an ultra-thin insulating layer of amorphous aluminum oxide[26] so that carrier transport occurs by tunneling. Such magnetic tunnel junctions (MTJ) had a magnetoresistance ratio (MR) [defined as the ratio $(R_{AP} - R_P)/R_P$, where R_{AP} is the resistance when the ferromagnetic layers are antiparallel and R_P is the resistance when the ferromagnetic layers are parallel] of 18% at room temperature. More recently, MRs of ~200% have been obtained with single crystal MgO(001) barriers and up to 500% with CoFeB/MgO/CoFeB MTJs.[27] This giant tunneling magnetoresistance is attributable to spin-dependent tunneling and will be incorporated into the next generation of read heads.

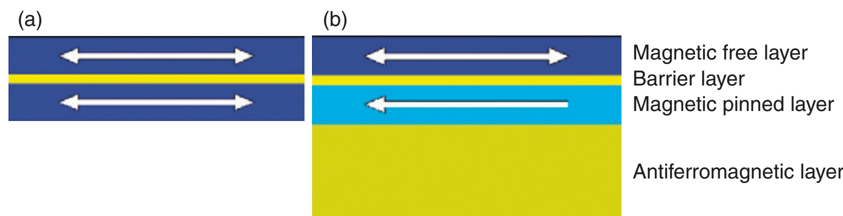


Figure 11. Basic layer structure of a spin valve without (a) and with (b) an anti-ferromagnetic pinning layer.

The great success in metallic magnetic spin valves and tunnel junctions for magnetic field sensors has led to the desire to integrate spins directly with semiconductors. Magnetic random access memory (MRAM) involves incorporating a spin valve or MTJ with CMOS technologies.[28] A FET is used to determine the resistance state of the spin valve or MTJ to determine whether it is either aligned or anti-aligned. Because the magnetic state will remain without any applied power, this is a nonvolatile memory element. The major challenges in MRAM technology have been controlling the magnetic states of multiple magnetic element devices and reducing the current in the write lines required to create a sufficient magnetic field to switch the state of a magnetic bit. However, it is already possible to produce uniform and reproducible MTJ structures on 200- or 300-mm-diameter silicon substrates in a large-scale preproduction environment.[29] Such is the power of a large multi-billion-dollar commercial market.

The core element of electronics is the transistor with its amplification, which relies on extremely pure semiconductor materials in which electrical conduction can be controlled by manipulating the carrier density using electric fields supplied by a gate.[30] Spintronics based on semiconductors is therefore an active and promising research field, with several approaches being explored in parallel. Spin-valve transistors with a metallic base comprising a spin valve have been successfully fabricated and characterized. It has become clear that the metallic base has too little transmission to support amplification, which may limit its applicability to magnetic field sensors or alternative magnetic memory elements.

The largest impact that semiconductor spintronic devices is likely to have is in the development of novel devices with new functionalities. The last few years have demonstrated that there are currently no fundamental barriers for developing novel spintronic devices. However, electrical detection of spins in semiconductors is still not completely resolved. Devices are being envisaged with only the spin being manipulated without motion of the charge. The limit to CMOS technology is power dissipation attributable to transport of charge. Although the actual devices that have been foreseen must still be developed, they have the potential to be ultra-low-power replacements for CMOS. Spintronics is on the semiconductor road map beyond CMOS.[30]

Quantum Computing

Another technology that is a candidate to further the performance of electronic devices and systems is based on quantum computing, in which the quantum bits, or "qubits," of information are represented by the quantum state of single photons. For example, the logical value 0 can be represented by a horizontally polarized photon, and the logical value 1 can be represented by a vertically polarized photon. Alternatively, 0 and 1 could be represented by the presence of a single photon in one of two optical fibers. This technology is currently being developed by APL researchers and their collaborators and has been fully described in previous *Johns Hopkins APL Technical Digest* articles.[31,32] Thus, it will not be discussed further. However, a photograph of the experimental apparatus used to demonstrate the first quantum controlled-NOT logic gate for single-photon qubits is shown in Fig. 12. The experimental components included a mode-locked Ti-sapphire laser, many meters of single-mode fiber components, parametric down-conversion photon sources, and low-noise single-photon detectors. A considerable amount of technological development is required in order to bring quantum-computing systems to the marketplace.

CONCLUSIONS

The tremendous growth in semiconductor technology has been based on the well-behaved Si–SiO_2 system and on Moore's Law, which states that the number of transistors on a chip doubles approximately every 2 years. Up until now, this "law" has been followed by evolutionary progress of the basic technology. This progress has finally reached its limiting steps. Modifications to the material characteristics of the basic structures will allow Moore's Law to continue for a while, but Si-based components

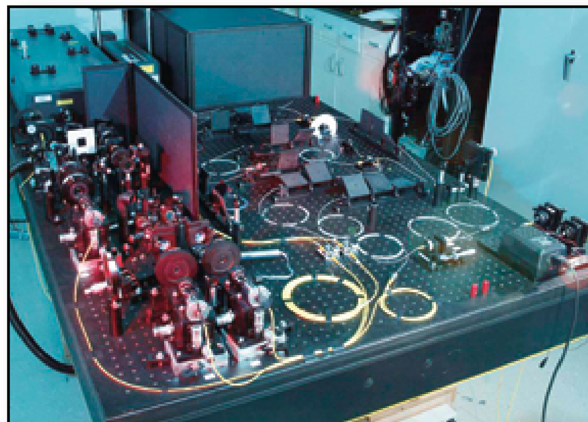


Figure 12. Experimental apparatus used to demonstrate the first quantum controlled-NOT logic gate for single-photon qubits.

should reach their ultimate limits around 2020. The need to continue to improve performance still further has led to a worldwide effort to develop possible alternative device technologies. Some of these approaches involve continuing the miniaturization route by using alternative semiconducting material systems such as CNTs, graphene layers, or SETs, whereas others are moving toward completely different concepts (e.g., spintronics and quantum computing).

REFERENCES

[1]Braun, F., "Über die Sromleitung durch Schwefelmetalle," *Ann. Phys. Chem.* **153**, 556–563 (1874).
[2]Round, H. J., "A Note on Carborundum," *Electr. World* **19**, 309 (1907).
[3]Schottky, W., "Halbleiter Theorie der Sperrschicht," *Naturwissenschaften* **26**, 843–844 (1938).
[4]Lilienfeld, J. E., *Method and Apparatus for Controlling Electric Currents*, U.S. Patent 1,745,175 (1930).
[5]Bardeen, J., and Brattain, W., "Physical Principles Involved in Transistor Action," in *Bell System Tech. J.* **28**(2), 239–277 (April 1949).
[6]PBS Online, "The First Silicon Transistor" in *Transistorized!*, ScienCentral, Inc./The American Institute of Physics, New York, http://www.pbs.org/transistor/science/events/silicont1.html (1999).
[7]Kahng, D., and Atalla, M. M., "Silicon–Silicon Dioxide Field Induced Surface Devices," in *Proc. IRE–AIEE Solid-State Device Research Conf.*, Carnegie Institute of Technology, Pittsburgh, PA (1960).
[8]Sze, S. M., and Ng, K. K., *Physics of Semiconductor Devices*, 3rd Ed., Wiley, Hoboken, NJ (2007).
[9]Sah, C.-T., *Fundamentals of Solid-State Electronics*, World Scientific, Hackensack, NJ (1991).
[10]Wallace, R. H., "Key Electrical Devices," Chap. 6, in *Electronic Materials from Silicon to Organics*, L. S. Mill and J. B. Mullin (eds.), Plenum, New York, pp. 47–65 (1991).
[11]Moore, G. E., "Cramming More Components onto Integrated Circuits," *Electronics* **38**(8), 114–117 (1965).
[12]Intel, *Silicon R&D Pipeline*, http://www.intel.com/technology/silicon/research.htm (accessed 6 Feb 2007).
[13]Intel, *Intel's Transistor Technology Breakthrough Represents Biggest Change to Computer Chips in 40 Years*, http://www.intel.com/pressroom/archive/releases/20070128comp.htm (accessed 6 Feb 2007).
[14]Cressler, J. D., "SiGe HBT Technology: A New Contender for Si-Based RF and Microwave Circuit Applications," *IEEE Trans. Microwave Theory Tech.* **46**(5), 572–589 (1998).
[15]Jain, S. C., Decoutere, S., Willander, M., and Maes, H. E., "SiGe HBTs for Application in BiCMOS Technology: I. Stability, Reliability and Material Parameters," *Semiconductor Sci. Tech.* **16**(6), R51–R65 (2001).
[16]Gaucher, B., Floyd, B., Reynolds, S., Pfeiffer, U., Grzyb, J., et al., "Silicon Germanium Based Millimetre-Wave ICs for Gbps Wireless Communications and Radar Systems," *Semiconductor Sci. Tech.* **22**(1), S236–S243 (2007).
[17]Avouris, P., Chen, Z., and Perebeinos, V., "Carbon-Based Electronics," *Nat. Nanotechnol.* **2**(10), 605–615 (2007).
[18]Novoselov, K. S., Jiang, D., Schedin, F., Booth, T. J., Khotkevich, V. V., et al., "Two-Dimensional Atomic Crystals," *Proc. Natl. Acad. Sci. USA* **102**(30), 10451–10453 (2005).
[19]Kang, S. J., Kocabas, C., Ozel, T., Shim, M., Pimparkar, N., et al., "High-Performance Electronics Using Dense, Perfectly Aligned Arrays of Single-Walled Carbon Nanotubes," *Nat. Nanotechnol.* **2**(4), 230–236 (2007).
[20]Hur, S.-H., Park, O. O., and Rogers, J. A., "Extreme Bendability of Single-Walled Carbon Nanotube Networks Transferred from High-Temperature Growth Substrates to Plastic and Their Use in Thin-Film Transistors," *Appl. Phys. Lett.* **86**(24), 243502 (2005).
[21]Kastner, M. A., "The Single Electron Transistor and Artificial Atoms," *Ann. Phys.* **9**(11–12), 885–894 (2000).
[22]Meirav, U., Kastner, M. A., and Wind, S. J., "Single-Electron Charging and Periodic Conductance Resonances in GaAs Nanostructures," *Phys. Rev. Lett.* **65**(6), 771–774 (1990).
[23]Karre, P. S. K., Bergstrom, P. L., Mallick, G., and Karna, S. P., "Room Temperature Operational Single Electron Transistor Fabricated by Focused Ion Beam Deposition," *J. Appl. Phys.* **102**(2), 024316 (2007).
[24]Baibich, M. N., Broto, J. M., Fert, A., Nguyen Van Dau, F., Petroff, F., et al., "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," *Phys. Rev. Lett.* **61**(21), 2472–2475 (1988).
[25]Saurenbach, F., Walz, U., Hinchey, L., Grünberg, P., and Zinn, W., "Static and Dynamic Magnetic Properties of Fe-Cr-Layered Structures with Antiferromagnetic Interlayer Exchange," *J. Appl. Phys.* **63**(8), 3473–3475 (1988).
[26]Miyazaki, T., and Tezuka, N., "Giant Magnetic Tunneling Effect in Fe/Al_2O_3/Fe Junction," *J. Magn. Magn. Mater.* **139**(3), L231–L234 (1995).
[27]Lee, Y. M., Hayakawa, J., Ikeda, S., Matsukura, F., and Ohno, H., "Effect of Electrode Composition on the Tunnel Magnetoresistance of Pseudo-Spin-Valve Magnetic Tunnel Junction with a MgO Tunnel Barrier," *Appl. Phys. Lett.* **90**(21), 212507 (2007).
[28]Yuasa, S., and Djayaprawira, D. D., "Giant Tunnel Magnetoresistance in Magnetic Tunnel Junctions with a Crystalline MgO(001) Barrier," *J. Phys. D. Appl. Phys.* **40**(21), R337–R354 (2007).
[29]Jansen, R., "The Spin-Valve Transistor: A Review and Outlook," *J. Phys. D. Appl. Phys.* **36**(19), R289–R308 (2003).
[30]Palmstrøm, C., "Spintronics," *IEEE Signal Process. Mag.* **23**(6), 96–94 (2006).
[31]Franson, J. D., and Jacobs, B. C., "Quantum Computing," *Johns Hopkins APL Tech. Dig.* **18**(2), 188–192 (1997).
[32]Pittman, T. B., Jacobs, B. C., and Franson, J. D., "Quantum Computing Using Linear Optics," *Johns Hopkins APL Tech. Dig.* **25**(2), 84–90 (2004).

The Author

Dennis K. Wickenden is a member of the Principal Professional Staff in the Sensor Science Group of the APL Milton S. Eisenhower Research Center. He received his B.Sc. and Associateship of the Royal College of Science (A.R.C.S.) degrees in 1962 and his Ph.D. and Diploma of the Imperial College (D.I.C.) degrees in 1965, all in physics from the Imperial College of Science and Technology (University of London). After a postdoctoral fellowship at the University of Notre Dame, he was employed at the GEC Hirst Research Centre in England to focus on the growth, characterization, and application of a variety of compound semiconductors, including gallium nitride, zinc sulfide, and gallium arsenide. In 1985, he moved to Portland, Oregon, to serve as senior scientist at Crystal Specialties, developing epitaxial growth equipment for the semiconductor industry. Dr. Wickenden joined APL in 1987, and his current research interests include the development of microelectromechanical systems (MEMS) sensors, metamaterials, and frequency-selective surfaces. He is author or coauthor of more than 100 papers, holds 11 patents, and was elected a Fellow of the Institute of Physics in 1973. His e-mail address is dennis.wickenden@jhuapl.edu.

Dennis K. Wickenden