Carbon nanotube electronics and photonics

Phaedon Avouris

Their geometry, mechanical flexibility, and unique charge-transport properties make carbon nanotubes ideally suited to supplant silicon in the next generation of FETs.

Phaedon Avouris is an IBM fellow and manager of nanometer-scale science and technology at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York.

Since the demonstration of the first transistors in 1947, computing electronics has been based mostly on one material, silicon. The emergence of MOSFETs in the 1960s and the development of microlithography and other fabrication techniques have allowed the continuous miniaturization of silicon devices. The resulting increases in density and switching speed and the reduced cost have made today's information age possible. Unfortunately, for various technological and even fundamental physical reasons, that scaling process is reaching its useful limits. In response, researchers are exploring ideas for logic devices based on either different operating principles or different materials whose properties can sustain the scaling to yet smaller size and greater performance.

Among the alternative technologies, carbon-based electronics is particularly significant. It is based on the 1991 discovery of carbon nanotubes and the more recent studies of individual graphite layers, so-called graphenes. A single-walled CNT can be thought of as a ribbon of graphene rolled into a seamless, one-atom-thick cylinder. The ribbon's width determines the CNT's diameter (typically 1-2 nm), and the direction in which the graphene is rolled defines the CNT chirality. Moreover, most of the CNT's electrical properties can be deduced from those of graphene after additional boundary conditions are introduced. Both materials display outstanding electrical properties such as either ballistic transport or diffusive transport with long mean free paths. And both are mechanically strong, flexible, and thermally conductive. But unlike graphene, semiconducting nanotubes have fairly large bandgaps that allow them to produce large on/off current ratios when fashioned into logic devices such as an FET.

The structure of traditional MOSFETs is described in the box on page 35. In a CNT FET, the silicon channel is replaced by an individual single-walled CNT (see figure 1); the first such transistors were independently reported in 1998 from groups at Delft University of Technology and IBM's Thomas J. Watson Research Center (see the article by Cees Dekker in PHYSICS TODAY, May 1999, page 22).

For the last several decades optoelectronics has been dominated by III–V semiconductors, particularly gallium arsenide and its alloys. In 2002, Rice University researchers observed photoluminescence from CNTs and the following year, IBM researchers reported electroluminescence. Those later developments opened the possibility of basing both computing electronics and optoelectronics on the same material; for recent reviews, see reference 3.

Nanotube transistors

A CNT has an atomic and electronic structure that gives it unique advantages as an FET channel. Its small, 1- to 2-nm diameter enhances the gate's ability to control the potential of the channel, particularly when the gate is configured to wrap around the CNT. The strong coupling makes the CNT the ultimate thin-body semiconductor system and allows one to shrink an FET in size yet still avoid the dreaded "short-channel effects," in which the gate field basically loses control of the device.

That all bonds in the CNT are satisfied and the surface is atomically smooth also has important implications. Absent are the scattering of electrons by surface states and the roughness that plagues conventional FETs, especially at high gate voltages. The electrical bandgap of a semiconducting CNT is inversely proportional to the tube diameter, which allows such tubes to be used in various different applications. As discussed later, the key advantage is the CNT's high charge mobility and its low capacitance. For typical performance characteristics as an FET channel, see figure 2.

The "simple" substitution of a silicon channel with a CNT dramatically alters the detailed physics of the device. The CNT's quasi-one-dimensional character, strong electron confinement, nanometer width, and strong covalent bonding drastically affect the channel's electrical- and thermal-transport properties. The chemistry of the metal–carbon contacts, the interaction of insulators with CNTs, and the sensitivity of those one-atom-thick structures to environmental influences all raise fascinating challenges and opportunities.

Carrier scattering

A key reason to use CNTs in electronics is their excellent transport properties. The elastic scattering of charge carriers is weak, and the carrier mean free paths are long—on the order of a micrometer for pure CNTs. Inelastic phonon-scattering processes therefore tend to dominate. The strength of the scattering depends on the energy of the carriers. At low temperatures and low bias, only acoustic phonons can couple to electrons, which gives rise to an inverse temperature dependence of the carrier mobility in semiconducting CNTs. The weak temperature dependence in quasi-one dimension is due to phase-space restrictions.
The principle of a MOSFET

The basic structure of a MOSFET involves a channel made of a semiconductor, typically silicon, connected to two electrodes: a source and drain. An insulating thin film, usually silicon dioxide, separates the channel, source, and drain from a third electrode called the gate (G), as pictured schematically. By applying a voltage $V_g$ between the gate and source, the conductance of the semiconducting channel can be modulated. Charge carriers (electrons or holes) traveling between source and drain encounter a material- and structure-dependent energy barrier in the bulk of the semiconductor.

The barrier, shown in the energy profile as a function of the channel position, lies in the conduction band for an electron (the charge carrier in an n-type semiconductor) or in the valence band for a hole (the carrier in a p-type semiconductor). The electric field generated by the biased gate, depending on its direction and strength, lowers or raises the barrier and thus changes the conductivity. For example, a positive gate voltage lowers the barrier for an electron and a negative gate voltage raises it. The drain–source voltage $V_{ds}$ provides the driving force that moves carriers across the channel.

The scaling of a MOSFET involves reducing its dimensions and supply voltage by the same factor $\alpha$. The speed of the device thus increases by $\alpha$ and the density of devices, when packed together, by $\alpha^2$, while the power density remains constant.

A number of limitations arise as dimensions reach the nanometer regime. Two variables are already close to their scaling limits. One of them is the thickness of the gate insulator, typically $\text{SiO}_2$, which is currently about 1 nm. Were that thickness to shrink much more, the leakage current from quantum tunneling would undermine the switching action of the FET. The leakage current also contributes to a major problem—excess power dissipation.

Another parameter that is difficult to scale down is the operating voltage $V$ usually about 1 V. The dynamic switching energy of a device is given by $\frac{1}{2}(C_{dso} + C_s)V^2$, where $C_{dso}$ and $C_s$ are the device and wiring capacitance contributions, respectively. If $V$ is not reduced as devices shrink in size and are more closely packed, the dissipated power density will increase. High temperatures then develop locally and decrease both the performance and lifetime of devices. Ways to reduce $V$ and the total capacitance are, therefore, being vigorously pursued.

To maintain the basic principle of the FET—that is, the control of transport in the channel using gate bias—one also needs to maintain a certain relationship between the screening length $\lambda$—the scale over which the electric field is screened out in the semiconductor—and the length of the gate, $L_g$. Specifically, $L_g$ should be a few times $\lambda$, which depends on the channel and oxide-insulator thicknesses $d_c$ and $d_o$, respectively, and their dielectric constants, $\varepsilon_c$ and $\varepsilon_o$. In a planar FET, $\lambda = (d_c d_o \varepsilon_c \varepsilon_o)^{1/2}$. Therefore, thin-device structures with thin, high-$\varepsilon_o$ dielectrics are desired. To increase the effective coupling of the gate and channel, a geometry in which the metal gate wraps around the channel is optimal. In that cylindrical geometry, $\lambda$ is shorter and depends less on $d_o$. Consequently, devices can be made shorter, and the scaling process can continue.

and the absence of low-angle scattering.

As a result of energy- and momentum-conservation requirements and the mismatch between the phonon sound velocity and electron band velocity, only a small fraction of phonons can effectively participate in the scattering in CNTs. The consequence is a very high low-electric-field mobility (typically about 10 000 cm$^2$/Vs), and values as high as 100 000 cm$^2$/Vs have been reported, even at room temperatures, by the University of Maryland's Michael Fuhrer. That is unlike other materials such as III–V semiconductors, whose charge mobilities are high at low temperatures but substantially degraded at room temperature.

Unlike acoustic-phonon scattering, optical-phonon scattering is very strong in CNTs. The phonons contract and elongate the carbon–carbon bonds, thus strongly modulating the electronic structure and coupling the electrons and phonons. At ambient temperatures, however, the phonon population is negligible, and electrons must have energies larger than about 160 meV to emit such a phonon. That can only be achieved under high voltage-bias conditions.

Unusual, nonequilibrium phonon distributions can be produced by current flow in nanotubes at those high biases due to differences in electron–phonon coupling among the various phonon modes, the large disparity in phonon frequencies between acoustic and optical modes, and the poor thermal interface between the CNT and the insulating substrate. As the carrier energy increases further, other inelastic processes take place that involve CNT excited states, which are discussed later.

Current injection

The semiconducting CNTs forming FET channels are connected to metal electrodes, typically palladium, gold, titanium, or aluminum. The differences in the metal and CNT work functions lead to charge transfer at their interface. The resulting interface dipoles produce a potential-energy...
Figure 1. A single-walled carbon nanotube, shown in this scanning tunneling microscope image (a), can form the basis of an FET. (b) In a top-gated configuration, a single nanotube is sandwiched between two insulating layers, hafnium oxide and silicon dioxide in this case. Palladium metal electrodes serve as source and drain. (Adapted from A. Javey et al., Nano Lett. 4, 1319, 2004.) (c) In an alternate configuration, an insulating layer and metal gate are wrapped around each nanotube to optimize the gate-channel coupling. The nanotube segments away from the gate are doped. (Adapted from Ph. Avouris et al., ref. 3.)

barrier, the so-called Schottky barrier. The alignment of the Fermi levels of the metal and the CNT, and therefore the SB height, depend on their respective work functions, the CNT bandgap, and chemical-bonding details at the interface.7

An SB exists at each end of an FET channel. When one of the two barriers is much higher than the other—their sum must equal the bandgap—the FET operates as a unipolar device, meaning it transports only one type of carrier, electrons or holes. For example, a metal such as Pd with a high work function could be used to form a nearly barrierless contact for holes. But electron injection would then experience the maximum barrier with a value close to the bandgap. Ali Javey and colleagues at Stanford University used Pd as a metal in contact with 2-nm-diameter CNTs and observed quasi-ballistic transport in short, p-type CNT FETs.8

In general, the current strongly depends not only on the type of metal but also on the diameter of the CNT. Moreover, because carrier transport through the metal–CNT interface is dominated by quantum mechanical tunneling through the SB, the barrier’s thickness becomes critical. But thanks to the enhanced screening provided by a metal gate near the thin oxide layer, both SBs can be made thin enough to allow the injection of either electrons, holes, or both simultaneously.7 That type of transistor is called ambipolar. The insets of figure 2 show how the application of different gate voltages triggers the injection of electrons or holes. The SB thickness can be modified by doping the CNT near the contacts or by using CNT FETs with multiple gates.

The 1D character and single-atomic-layer structure of CNTs make them susceptible to environmental perturbations. Intrinsic screening is weak, and trapped charges in a gate insulator shift electrical characteristics such as the FET threshold voltage. Specific interactions such as charge trans-
fer to or from the substrate or the absorption of species on the CNT, the metal electrodes, or the substrate can alter the character of the transport, the doping, and the resistance.\(^3\) Indeed, such effects on the electrical properties of CNTs provide the basis for various sensor applications.

**Implementation and performance**

Early CNT FETs were simple structures composed of a long CNT channel that is van der Waals–bonded to metal contacts and back-gated using the Si substrate itself.\(^3,4\) Although functional, they had rather poor electrical characteristics. Shalom Wind and colleagues improved on those characteristics by embedding the CNTs in the metal electrodes and using top metal gates and thin gate insulators. Javey, Hongjie Dai, and coworkers further enhanced the channel–gate coupling and performance by using high-dielectric insulators such as zirconium oxide.\(^6\) A number of different device structures have since evolved, among them multiple-gate devices\(^5\) and wrap-around-gate configurations fabricated jointly by IBM and Harvard University.

The on/off current ratio in a CNT FET is generally high: \(10^{10}\), ideal for logic devices. Another important parameter is its transconductance \(g_m\) — the change in drain–source current \(I_{ds}\) divided by the change in the gate voltage \(V_{gs}\) at a constant drain–source voltage. For example, a top-gated FET with a high-dielectric insulator and a CNT diameter of about 1.7 nm can reach a \(g_m\) of about 30 microsiemens.\(^\text{6}\) Theoretical limit is close to 155 \(\mu S\), which leaves a lot of room for improvement.

Another important characteristic of FET operation is what's known as subthreshold swing, \(S = \frac{dV_{gs}}{d \log I_{ds}}\), which describes the voltage needed to produce an order of magnitude change in the current. Conventional FETs require a change in the channel potential of at least 60 mV at 300 K to produce that change, and the minimum subthreshold swing imposes a lower limit on the operating voltage and hence on the power dissipation in conventional FETs. Researchers have achieved that thermodynamic limit in CNT FETs by suppressing SBs in double-gate devices\(^9\) or by doping the contacts. More important, work at the Watson Research Center has shown that the limit can be exceeded. By taking advantage of band-to-band tunneling, Joerg Appenzeller and colleagues showed that the electron distribution can be filtered to remove high-energy electrons and reduce the subthreshold swing.\(^3\)

High-speed devices, of course, are a key goal of the whole effort. Consider a typical measure of the frequency response of conventional transistors, the cut-off frequency \(f_T = \frac{g_m}{2\pi C_{gs}}\). Using 5 aF as a typical gate–source capacitance \(C_{gs}\) and \(g_m = 10–20 \mu S\) for a 100-nm-length CNT FET, we find a cut-off frequency in the range of 1 THz.\(^2\) Direct measurement of such high frequencies is difficult because of the mismatch in impedances of CNT FETs and the measurement systems. Indirect measurements on single CNT FETs have shown that their performance is currently limited by the parasitic capacitance of the circuits.

CNTs can also help alleviate the power dissipation problem that limits the scaling of MOSFETs. To dynamically switch a device takes an energy of \(\frac{1}{2}(C_{sd} + C_{ds})V^2\), where \(C_{sd}\) and \(C_{ds}\) are the device and wiring capacitance contributions, respectively. A primary advantage of CNT FETs over Si MOSFETs is their much lower capacitance values, roughly 0.05 aF/nm. Careful design, though, has to ensure that parasitic capacitance and \(C_{sd}\) typically 0.2 aF/nm, do not overwhelm \(C_{sd}\) as they might, for example, when the wires are long. The power dissipation can also be lowered by reducing the supply voltage \(V_{dd}\) (power \(= fV_{dd}^2\), where \(f\) is the operating frequency). However, \(V_{dd}\) cannot be arbitrarily reduced. The performance, \(\tau = 1/f\), degrades with decreasing \(V_{dd}\). As a result, there is an optimum \(V_{dd}\) for a given \(f\).

**Integrated circuits**

The fabrication of CNT-based devices has advanced beyond single transistors to include logic gates and more complex structures such as ring oscillators.\(^3\) The oscillators are composed of an odd number of pairs of p- and n-type FETs made by appropriate doping of the CNTs. Controlled doping in nanoscale devices is difficult, however, and fluctuations in...
the number and position of the dopants can have a profound effect on device performance.

Fortunately, the ambipolar behavior of an undoped CNT can be used to implement CMOS architecture. A CMOS logic device is formed using a pair of p- and n-type transistors that operate under conditions in which one FET is turned on and the other is off. The ambipolar characteristics of a CNT FET provide a pair of p- and n-type branches. But the leakage branch of one type has a current level comparable to the main branch of the other. Controlling the threshold voltage in an ambipolar device is therefore key to achieving the p- and n-type characteristics suitable for CMOS logic. One can control the threshold voltage by tuning the work function of the metal gate in each FET; the gate work function can be viewed as an extra voltage source added to $V_{gs}$. When the metals are properly selected, their different work functions shift the characteristics of p- and n-type branches to generate a distinguishable on state in one CNT FET and an off state in the other.

That approach was employed by IBM's Zhihong Chen and coworkers to fabricate the most complex electronic structure so far based on a single CNT. Figure 3 shows a scanning electron microscope image of a five-stage ring oscillator in which Pd was chosen as the metal gate for the p-type FET and Al for the n-type FET.

Despite the outstanding performance of single-CNT FETs, their practical implementation faces a number of hurdles. The principal one is the heterogeneity of as-prepared CNT mixtures. Synthetic techniques have emerged in recent years for producing ever-smaller CNT diameter and chirality distributions.

One electronic technology that bypasses the heterogeneity problem to some extent is that of CNT thin-film transistors (TFTs), introduced by Eric Snow and coworkers in 2003.

A film of randomly oriented CNTs is deposited on an insulator and electrodes are patterned on top of it. Ensemble averaging is expected to correct for the heterogeneity of the film. The presence of metallic CNTs in the mixture poses a problem, however, since they tend to short out the FET channel.

To avoid the problem, the channel is made much longer than the average length of the CNTs so that no single tube can bridge the source–drain gap. Charge transport occurs along "percolation paths," and switching involves numerous tube–tube junctions. The technique achieves reasonably high on/off current ratios, but at the cost of increasing the channel resistance and consequently producing low drive currents. The carrier mobility is also reduced drastically, although it remains higher than that of typical organic TFTs. Furthermore, the nanotubes can be deposited at low temperatures on a variety of substrates for various applications, one of them pictured in figure 4.

One approach to improving the performance of TFTs is to use purely semiconducting CNTs in a film. Another is to use highly oriented CNTs, which reduces the number of high-resistance tube–tube junctions that are so prevalent in randomly oriented CNTs. Researchers at the Watson Research Center and Northwestern University have recently implemented those improvements in their films, and now both high driving currents (up to milliamps) and fairly high on/off current ratios (roughly $10^4$) can be achieved simultaneously. The mobilities are high for TFTs, but they are still significantly lower than those of single-CNT devices.

Analysis of those results suggests that at least two problems remain. A distribution of tube diameters, and thus bandgaps, still exists in the films. That leads to a spread in the threshold voltage for switching the current in the different tubes. Moreover, high currents require a high CNT density, but closely spaced nanotubes tend to interact and screen the gate field.

**Excited states**

Studies of the excited states of CNTs have contributed greatly to our understanding of the electronic structure and interactions of CNTs and opened the door to new applications.

Early on, the electronic excitations of CNTs were discussed, in the context of the single-particle model, as transitions between corresponding pairs of so-called van Hove singularities in the ID density of states that produce free electron–hole pairs. However, as pointed out by Tsuneya Ando while at the University of Tokyo in 1997, electron–electron interactions should be strong in a confined 1D system and the interactions should form bound states, known as excitons, below the free-particle continuum. First-principle calculations by Steven Louie and coworkers at the University of California, Berkeley, showed that exciton binding energies in nanotubes are unusually large (on the order of an eV). Excitons carry most of the oscillator strength and should, therefore, dominate the absorption spectra (see figure 5).

The predictions were confirmed by two-photon excitation spectroscopy by Tony Heinz's group at Columbia University and Christian Thomsen's group at the Technical University of Berlin. The exciton binding energies of semiconducting CNTs depend inversely on their diameter and, unlike the case of bulk semiconductors, also depend on the screening provided by the medium in which the CNTs are embedded.

Experimental studies of the excited states of nanotubes did not start in earnest until 2002 when Bruce Weisman and coworkers at Rice observed photoluminescence from CNTs wrapped by surfactants and dispersed individually in solu-
tion. Previous attempts were hampered by aggregation effects and the coexistence of metallic and semiconducting CNTs, with the former acting as luminescence quenchers. Once the conditions for luminescence were established, the field grew rapidly. A key application has been the identification of the structure of emitting CNTs. By monitoring the luminescence intensity as a function of laser frequency, researchers can determine the excitation spectra of individual tubes and deduce their structure.

Following the observation of photoluminescence from CNTs, the decay of their excited states became an active research field. Measured room-temperature fluorescence lifetimes from dispersions of CNTs are in the range of 10–100 ps, although theory predicts CNT radiative lifetimes of 1–10 ns. Reported fluorescence quantum yields, Q, are also low, but there are significant uncertainties in their values. However, despite the lack of precise Q values, it is clear that fast non-radiative decay processes operate in CNTs.

Another result of the strong electron-electron interactions in CNTs is that the degeneracy present in the single-particle states is partially removed, which results in new excitonic states. The transition to the lowest energy state was shown by theory and experiment to be "dark"—that is, dipole forbidden. Trapping in the dark exciton state might be responsible for the observed low yield. However, the energy separation between dark and bright excitons determined by Junichiro Kono and coworkers at Rice is too small to account for the low Q. Recent theoretical work at the Watson Research Center predicts two fast CNT decay channels: multiphonon decay of trapped excitons and a new, phonon-assisted electronic decay pathway. The latter is made possible by doping the CNT through charge-transfer interactions with its environment or by external fields.

The photoluminescence statistics of individual CNTs are also interesting. A recent study by Atac Imamoglu and coworkers at ETH Zürich found evidence of photon antibunching, a discovery that may lead to the use of CNTs as single-photon sources and in quantum cryptography.

**Electroluminescence**

From a photonics-applications perspective, a key finding was that light emission from CNTs can be obtained by passing a current through them. As noted previously, CNT FETs tend to be ambipolar, so electron and hole currents can flow simultaneously. When those electrons and holes recombine, a fraction of the recombination events emit a photon and thus give rise to electroluminescence. EL is widely used to produce solid-state light sources such as LEDs.

Conventionally, EL is generated near the interface between a hole-doped and an electron-doped material. But in an ambipolar CNT, doping is unnecessary, which simplifies LED fabrication. Even more interesting, EL from a long CNT does not originate from a fixed point but can actually be translated at will by applying an appropriate bias at the gate; the gate bias determines the position where electron–hole recombination takes place in the intrinsic material. The energy E of the EL is the same as that of photoluminescence of the CNT and can thus be tuned by using tubes of different diameters d since E is proportional to 1/d.

Another mechanism by which a current can excite EL involves the presence of hot carriers in CNTs. In a unipolar CNT at high bias, the carriers can achieve high enough energies and, because of the strong Coulomb interactions among them, can induce impact excitation and ionization. In bulk solids, ionization is the principal process induced, but in the quasi-1D CNTs with high exciton binding energies, neutral excitations dominate. Typically, the emission is localized at defect sites, where voltage drops provide enough energy to the carriers to exceed the excitation threshold. More important, one can choose the location of the emission by introducing along the CNT channel a voltage drop—for example, by locally modifying the coupling between the gate and the CNT.

Electroluminescence from CNTs is broad and nondirectional. But most optoelectronics applications, such as optical interconnections, require a narrow, directional light beam. IBM's Fengnian Xia and coworkers recently achieved that effect by confining a single CNT in a planar photonic half-wavelength microcavity. In that configuration, the fact that different diameter nanotubes in a sample have slightly different emission energies becomes less important, because the resonant wavelength is determined by the structure of the cavity itself.

Photoconductivity, essentially the reverse of EL, occurs when optical radiation produces free electrons and holes in a material. The process was first observed in CNTs at the Watson Research Center in 2003, but since then many such studies have appeared. The same CNT FET device can be used as a transistor, photoswitch, light emitter, or photodetector.
In an open-circuit configuration, internal electric fields can separate electrons and holes and generate a photovoltage. Photovoltage imaging microscopy can provide images of electric fields along CNTs such as those at SBs or charged defects.

**On the horizon**

With their unique properties, semiconducting nanotubes can form the basis for extremely small, fast, and flexible transistor circuits. CNT devices are molecular devices; therefore, simple and inexpensive self-assembly-based fabrication may be possible. Transistors with chemically modified CNTs are already being used as sensitive and selective chemical and biological sensors. CNT-based light sources and detectors may separate electrons and holes and generate a photovoltaic effect. Photovoltage imaging microscopy can provide images of electric fields along CNTs such as those at SBs or charged defects. And researchers who investigate self-assembly of CNT structures, and there is no reliable way at present to directly produce any single type of CNT needed in a large integrated system. Promising signs suggest that may change. Already, separation of a single type of CNT from mixtures has been successful. And researchers who investigate self-assembly of CNT devices are reporting encouraging results.

The evolution cannot be characterized as an incremental improvement of existing technology as embodied in the scaling of silicon devices. It involves fundamental changes in materials, processing, assembly techniques, and physical principles. The development of CNT technology depends, as does everything else in industry, on cost-benefit analysis and unanticipated breakthroughs. Already, however, the study of CNTs has provided us with unique insights into the physics and chemistry of nanoscale materials.

**References**
