Beyond the conventional transistor

by H.-S. P. Wong

This paper focuses on approaches to continuing CMOS scaling by introducing new device structures and new materials. Starting from an analysis of the sources of improvements in device performance, we present technology options for achieving these performance enhancements. These options include high-dielectric-constant (high-*k*) gate dielectric, metal gate electrode, double-gate FET, and strained-silicon FET. Nanotechnology is examined in the context of continuing the progress in electronic systems enabled by silicon microelectronics technology. The carbon nanotube field-effect transistor is examined as an example of the evaluation process required to identify suitable nanotechnologies for such purposes.

1. Introduction

The semiconductor industry has been so successful in providing continued system performance improvement year after year that the Semiconductor Industry Association (SIA) has been publishing roadmaps for semiconductor technology since 1992. These roadmaps represent a consensus outlook of industry trends, taking history as a guide. The recent roadmaps [1] incorporate participation from the global semiconductor industry, including the United States, Europe, Japan, Korea, and Taiwan. They basically affirm the desire of the industry to continue with Moore's law [2], which is often stated as the doubling of transistor performance and quadrupling of the number of devices on a chip every three years. The phenomenal progress signified by Moore's law has been achieved through scaling of the metal-oxide–semiconductor field-effect transistor (MOSFET) [3, 4] from larger physical dimensions to smaller physical dimensions, thereby gaining speed and density.

Shrinking the conventional MOSFET beyond the 50-nm-technology node requires innovations to circumvent barriers due to the fundamental physics that constrains the conventional MOSFET. The limits most often cited [4–12] are 1) quantum-mechanical tunneling of carriers through the thin gate oxide; 2) quantum-mechanical tunneling of carriers from source to drain, and from drain to the body of the MOSFET; 3) control of the density and location of dopant atoms in the MOSFET channel and source/drain region to provide a high on-off current ratio; and 4) the finite subthreshold slope. These fundamental limits have led to pessimistic predictions of the imminent end of technological progress for the semiconductor industry [4]. On the other hand, the push to scale the conventional MOSFET continues to show remarkable progress [13, 14].

Instead of reiterating the considerations of device scaling limits here, we refer the reader to our previous analyses [8–10] as well as analyses by others in the literature [4–7, 11, 12]. We focus this paper instead on approaches to circumvent or surmount the barriers to device scaling. The organization of this paper is as follows. We first address opportunities for the silicon MOSFET, focusing primarily on approaches that depart from conventional scaling techniques (for example, doping profile control, thin silicon dioxide gate dielectrics, SOI). Topics covered include high-dielectric-constant (high-*k*)

©**Copyright** 2002 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/02/$5.00 © 2002 IBM

Table 1 Device performance improvement opportunities.

Source of improvement	*Parameters affected*	*Method*
Charge density	1. S (inverse subthreshold slope) 2. Q_{inv} at a fixed off-current	1. Double-gate FET. 2. Lowered operating temperature.
Carrier transport	1. Mobility (μ_{eff}) 2. Carrier velocity 3. Ballistic transport	1. Strained silicon. 2. High-mobility and -saturation-velocity materials (e.g., Ge, InGaAs, InP). 3. Reduced mobility degradation factors (e.g., reduced transverse electric field, reduced Coulomb scattering due to dopants, reduced phonon scattering). 4. Shorter channel length. 5. Lowered operating temperature.
Ensuring device scalability to a shorter channel length	1. Generalized scale length (λ). 2. Channel length (L_g)	1. Maintaining good electrostatic control of channel potential (e.g., double-gate FET, ground-plane FET, and ultrathin-body SOI) by controlling the device physical geometry and providing means to terminate drain electric fields. 2. Sharp doping profiles, halo/pocket implants. 3. High gate capacitance (thin gate dielectrics, metal gate electrode) to provide strong gate control of channel potential.
Parasitic resistance	1. R_{ext}	1. Extended/raised source/drain. 2. Low-barrier Schottky contact.
Parasitic capacitance	1. C_{jn} 2. C_{GD}, C_{GS}, C_{GB}	1. SOI. 2. Double-gate FET.

gate dielectric, metal gate electrode, double-gate FET, and strained-silicon FET. The second part of this paper examines the space between conventional microelectronics technology and the more exploratory nanotechnology. Such a wide spectrum of nanotechnologies are being explored today that it is impossible to make even a modest attempt to cover the field. The approach adopted in this paper is to select an example, the carbon nanotube field-effect transistor, to illustrate both the opportunities offered by nanotechnologies and the most important questions that must be answered before such technologies can find practical use. The example is therefore chosen for illustrative purposes rather than an implied suggestion of eventual technological utility.

2. Silicon MOSFET

For digital circuits, a figure of merit for MOSFETs for unloaded circuits is CV/I, where C is the gate capacitance, V is the voltage swing, and I is the current drive of the MOSFET. For loaded circuits, the current drive of the MOSFET is of paramount importance. Historical data indicate that scaling the MOSFET channel length improves circuit speed, as suggested by scaling theory [3]. Reference [15] illustrates data on the CV/I metric from recent literature. The off-current specification for CMOS has been rising rapidly to keep the speed performance high. While 1 nA/μm was the maximum off-current allowed in the late 1990s [8], off-currents in excess of 100 nA/μm are proposed today [13]. This trend obviously cannot continue, since the on-current increases only linearly as off-current increases exponentially in a typical device design tradeoff. Means to mitigate the standby power increase must be found.

Keeping in mind both the CV/I metric and the benefits of a large current drive, we note that device performance may be improved by 1) inducing a larger charge density for a given gate voltage drive; 2) enhancing the carrier transport by improving the mobility, saturation velocity, or ballistic transport; 3) ensuring device scalability to achieve a shorter channel length; and 4) reducing parasitic capacitances and parasitic resistances. **Table 1** summarizes these opportunities and proposed technology options for capitalizing on them. These options generally fall into two categories: new materials and new device structures. In many cases, the introduction of a new material requires the use of a new device structure, or vice versa. Throughout the discussion, we direct attention to areas of device physics and materials science that must be better understood in order to advance the technology.

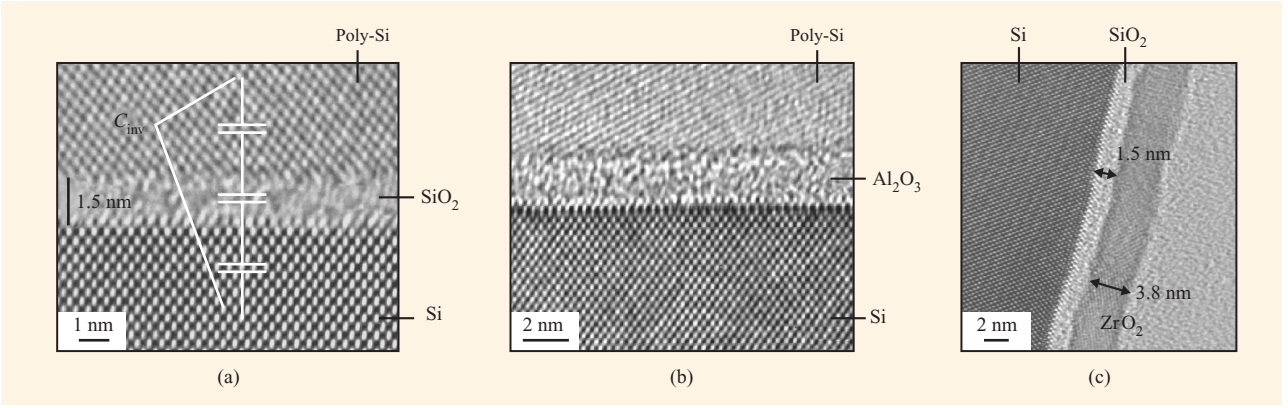


Figure 1

(a) Transmission electron micrograph (TEM) of a conventional silicon dioxide (oxynitride) with a physical thickness of 1.5 nm. (b) TEM of a 2.2-nm Al_2O_3 with an equivalent electrical thickness of 1 nm. (c) TEM of a 3.8-nm ZrO_2 on a 1.5-nm interfacial silicon dioxide. Adapted with permission from Gusev et al. [20]; © 2001 IEEE.

MOSFET gate stack

Continued device scaling requires the continued reduction of the gate dielectric thickness. This requirement arises from two different considerations: controlling the short-channel effect and achieving a high current drive by keeping the amount of charge induced in the channel large as the power-supply voltage decreases. In both cases, to a first approximation, it is the electrical thickness that is important. The electrical thickness at inversion is determined by the series combination of three capacitances in the gate stack: the depletion capacitance of the gate electrode, the capacitance of the gate dielectric, and the capacitance of the inversion layer in the silicon [**Figure 1**, part (a)].

On the other hand, the direct tunneling current through the gate dielectric grows exponentially with decreasing physical thickness of the gate dielectric [16]. This tunneling current has a direct impact on the standby power of the chip and puts a lower limit on unabated reduction of the physical thickness of the gate dielectric. It is likely that tunneling currents arising from silicon dioxides (SiO_2) thinner than 0.8 nm cannot be tolerated, even for high-performance systems [10].

Solutions that reduce the gate tunneling current and gate capacitance degradation due to polysilicon depletion are explored through introduction of new materials: high-dielectric-constant gate dielectrics and metal gate electrodes.

*High-*k *gate dielectric*

A gate dielectric with a dielectric constant (k) substantially higher than that of SiO_2 (k_{ox}) will achieve a smaller equivalent electrical thickness (t_{eq}) than the SiO_2, even with a physical thickness (t_{phys}) larger than that of the SiO_2 (t_{ox}):

$$t_{eq} = \left(\frac{k_{ox}}{k}\right) t_{phys} .$$

Replacing the SiO_2 with a material having a different dielectric constant is not as simple as it may seem. The material bulk and interface properties must be comparable to those of SiO_2, which are remarkably good. Basic material properties such as thermodynamic stability with respect to silicon, stability under thermal conditions relevant to microelectronic fabrication, low diffusion coefficients, and thermal expansion match are some critical examples. In addition, interface traps of the order of a few 10^{10} $cm^{-2}eV^{-1}$ and bulk traps of the order of a few 10^{10} cm^{-2} are common among SiO_2 and the closely related oxynitrides [17, 18]. Charge trapping and reliability for the gate dielectrics are particularly important considerations.

Thermal stability with respect to silicon is an important consideration, since high-temperature anneals are generally employed to activate dopants in the source/drain as well as the polysilicon gate. Although many binary and ternary oxides are predicted to be thermally stable with respect to silicon [19], recent research on high-dielectric-constant gate insulators have focused primarily on binary metal oxides such as Ta_2O_5, TiO_2, ZrO_2, HfO_2, Y_2O_3, La_2O_3, Al_2O_3, and Gd_2O_3 and their silicates [20]. **Table 2** compares the properties of the common high-k gate dielectrics reported in the literature. The dielectric constant of these materials generally ranges from 10 to 40, which is about a factor of 3 to 10 higher than SiO_2. Leakage current reduction from $10^3\times$ to $10^6\times$, in comparison with SiO_2 of the same electrical thickness, is generally achieved experimentally for high-k gate dielectrics [21]. The benefits of using a very-high-dielectric-constant material to simply replace SiO_2 for

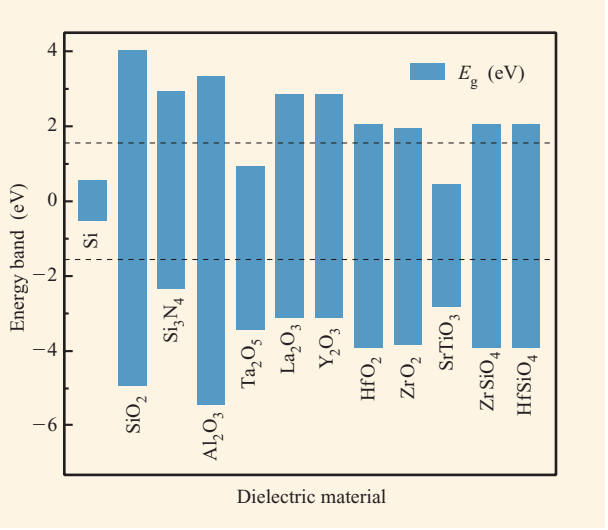


Figure 2

Bandgap and band alignment of high-*k* gate dielectrics with respect to silicon. Data from Robertson [25], with permission. The dashed line represents 1 eV above/below the conduction/valence bands.

the same electrical thickness are limited because of the presence of two-dimensional electric fringing fields from the drain through the physically thicker gate dielectric [10, 22]. The drain fringing field lowers the source-to-channel potential barrier and lowers the threshold voltage in a way similar to the well-known drain-induced barrier lowering (DIBL), in which the drain field modulates the source-to-channel potential barrier via coupling through the silicon substrate. The use of higher-*k* materials must therefore be combined with a concurrent reduction of the electrical thickness.

A large silicon-to-insulator energy barrier height is desirable because the gate direct-tunneling current is exponentially dependent on the (square root of the) barrier height [23]. In addition, hot-carrier emission into the gate insulator is also related to the same barrier height [24]. The high-*k* material should therefore not only have a large bandgap, but also have a band alignment which results in a large barrier height. **Figure 2** illustrates the bandgap and band alignment for several high-*k* gate dielectrics calculated by Robertson [25]. Most high-*k* materials that have other desirable properties do have relatively low band offsets and small bandgaps. Aluminum oxide (Al_2O_3) is probably the only material that has a bandgap and band alignment similar to those of SiO_2.

Figure 1 illustrates examples of thin gate dielectrics: SiO_2, Al_2O_3, and ZrO_2 with an interfacial SiO_2 layer. These dielectrics are only a few atoms thick. The thin dielectric films can be deposited by sputtering, sol–gel, physical vapor deposition (PVD), metallo-organic chemical vapor deposition (MOCVD), or atomic-layer deposition (ALD). Deposition uniformity does not appear to be a significant issue. However, integration of the deposited

Table 2 Selected material and electrical properties of high-*k* gate dielectrics. Data compiled from Robertson [25], Gusev et al. [20], Hubbard and Schlom [19], and other sources.

Dielectric	*Dielectric constant (bulk)*	*Bandgap* (eV)	*Conduction band offset* (eV)	*Leakage current reduction w.r.t. SiO_2*	*Thermal stability w.r.t. silicon (MEIS data)*
Silicon dioxide (SiO_2)	3.9	9	3.5	N/A	>1050°C
Silicon nitride (Si_3N_4)	7	5.3	2.4		>1050°C
Aluminum oxide (Al_2O_3)	~10	8.8	2.8	10^2–10^3×	~1000°C, RTA
Tantulum pentoxide (Ta_2O_5)	25	4.4	0.36		Not thermodynamically stable with silicon
Lanthanum oxide (La_2O_3)	~21	6*	2.3		
Gadolinium oxide (Gd_2O_3)	~12				
Yttrium oxide (Y_2O_3)	~15	6	2.3	10^4–10^5×	Silicate formation
Hafnium oxide (HfO_2)	~20	6	1.5	10^4–10^5×	~950°C
Zirconium oxide (ZrO_2)	~23	5.8	1.4	10^4–10^5×	~900°C
Strontium titanate ($SrTiO_3$)		3.3	−0.1		
Zirconium silicate ($ZrSiO_4$)		6*	1.5		
Hafnium silicate ($HfSiO_4$)		6*	1.5		

*Estimated value.

dielectric with the rest of the device fabrication process requires further research and development in several areas. If a conventional self-aligned polysilicon gate is used, the dielectric film must be able to withstand rapid thermal anneals (RTAs) up to at least 950°C for dopant activation in the polysilicon gate. The typical thermal treatments during a polysilicon gate CMOS process pose potential problems such as formation of silicates and interfacial SiO_2. In addition, diffusion (for example, boron, oxygen) through the gate dielectric is a serious concern. If a metal gate electrode is employed (using a low-temperature process), many of the thermal stability concerns can be relieved.

Figure 3(a) shows the electrical characteristics of an 80-nm polysilicon gate n-FET using Al_2O_3 as the gate dielectric, as reported by Buchanan et al. [26]. This work and that of others (for example, [21]) illustrates some of the obstacles for high-*k* gate dielectrics: 1) There are a significant number of traps and fixed charges in the film (or at the interfaces), leading to flat-band voltage shifts (up to 450 mV) and voltage bias instability; 2) the traps raise questions of reliability as channel hot carriers and carriers from gate tunneling traverse the gate dielectric, resulting in trap generation; and 3) the mobility of carriers in the FET channel is severely degraded (up to a factor of 2) for high-*k* gate dielectrics [**Figure 3(b)**].

The cause of the mobility degradation is not clear at present. Presumably, some of the differences can be attributed to the difficulty of obtaining accurate estimates of the effective electric field due to the charge trapping. Coulomb scattering due to the trapped charge alone cannot explain entirely the mobility degradation observed. Another source of mobility lowering may be found in remote phonon scattering [27]. The static dielectric constant of a high-bandgap high-*k* material derives its high dielectric constant primarily from ionic polarizability, since the large bandgap results in a small electronic polarizability. The ionic polarizability is associated with the "soft" metal–oxygen bonds with low-energy phonons. Fischetti et al. [27] studied the scattering of electrons in the inversion layer by surface optical phonons and suggested that there is generally an inverse relation between surface-optical-phonon-limited mobility and the static dielectric constant: the higher the dielectric constant, the lower the surface-optical-phonon-limited mobility.

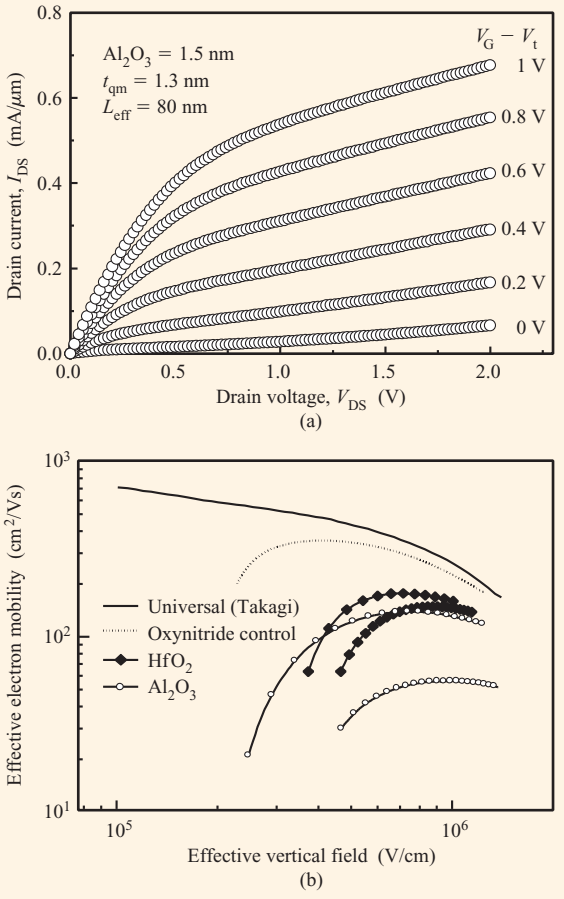


Figure 3

Electrical characteristics of a polysilicon-gated Al_2O_3 n-FET. (a) Drain current vs. drain voltage characteristics of an 80-nm-channel-length n-FET. Reproduced with permission from Buchanan et al. [26]; © 2000 IEEE. (b) Effective electron mobility of long-channel FET compared with the universal mobility curve [60]. Two HfO_2 curves show the effect of surface preparation. The Al_2O_3 curves illustrate the range of mobility for Al_2O_3 gate stacks. Mobility approximately twice as high as that of [26] is achieved due to improved processing. Reproduced with permission from Gusev et al. [21]; © 2001 IEEE.

Metal gate electrode

A metal gate electrode has several advantages compared to the doped polysilicon gate used almost exclusively today. Gate capacitance degradation due to the depletion of the doped polysilicon gate typically accounts for 0.4–0.5 nm of the equivalent-oxide thickness of the total gate capacitance at inversion. This is a substantial amount, considering that a gate equivalent oxide of less than 1.5 nm (at inversion) is required for sub-50-nm CMOS. The thermal instability of most high-*k* gate dielectrics may require the use of a low thermal budget process after the gate dielectric deposition. While junction activation may be performed prior to gate dielectric deposition, the high-temperature gate polysilicon activation step necessarily occurs after the gate dielectric formation. A further potential benefit of metal gate electrodes is the elimination of carrier mobility degradation due to plasmon scattering from the gate electrode. The plasmon frequency

of a highly conductive metal electrode is too high to interact with the carriers in the inversion layer [28].

From a device design point of view, the most important consideration for the gate electrode is the work function of the material. While the polysilicon gate technology has somewhat locked in the gate work functions to values close to the conduction band and the valence band of silicon,[1] the use of a metal gate material opens up the opportunity to choose the work function of the gate and redesign the device to achieve the best combination of work function and channel doping. For bulk or partially depleted SOI, because of the requirements on the threshold voltages and the need to use heavy dopants to control short-channel effects, the most suitable gate work-function values are still close to the conduction and valence bands of silicon. A mid-gap work function results in either a threshold voltage that is too high for high-performance applications, or compromised short-channel effects, since the channel must be counterdoped to bring the threshold voltage down. For double-gate FETs (see the section on double-gate FET electrostatics), because the short-channel effects are controlled by the device geometry, the threshold voltage is determined mainly by the gate work function. Therefore, the choice of the gate electrode is particularly important for the double-gate FET. For example, for symmetric double-gate FETs (SDG), a gate-electrode work function ± 250 mV from mid-gap is suitable. The section on double-gate FET electrostatics expands on this discussion.

While there are plenty of metal choices that may satisfy the work-function requirements [30–32], other device and integration considerations narrow down the choices significantly. The requirements of a low gate-dielectric/silicon interface state density and low gate-dielectric fixed charges imply that a damage-free metal deposition process (e.g., CVD instead of sputtering) is required. At the same time, the deposition process must not introduce impurities (e.g., traces of the CVD precursor materials) into the gate stack. The thermal stability of the metal electrode must at least withstand the thermal anneals required to passivate the silicon/gate-dielectric interface (e.g., forming-gas anneal) after the metal deposition, as well as the thermal processing of the back-end metallization processes. Furthermore, it is desirable to have a low resistivity (at least similar to conventional silicides such as $CoSi_2$ and $TiSi_2$), although this requirement may be relaxed by strapping the gate electrode of the proper work function with a lower-resistivity material on top.

In principle, a single-metal electrode is advantageous, since it avoids many potential problems of alloyed metals such as composition uniformity control and phase separation. On the other hand, alloying provides flexibility in choosing the desired material properties. The gate-electrode work-function issue is further complicated by the fact that the work function measured in vacuum (values reported in most materials data books) is different from the work-function value when the metal is in contact with a dielectric. In general, a dipole forms at the metal/dielectric interface which alters the effective work function of the metal/dielectric combination [33, 34]. The choice of appropriate metal electrode is then also dependent upon the choice of the gate dielectric: SiO_2 or high-*k* material.

One of the promising process integration schemes for metal gate is the replacement-gate technology [35]. In this process, a dummy gate material (e.g., polysilicon) is used for forming the self-aligned gate-to-source/drain structure. Subsequently, the dummy gate material is removed and replaced with the desired gate dielectric and electrode [35]. Alternatively, the metal gate electrode may be etched in a way similar to the polysilicon gate technology. However, the learning curve is long and steep for developing the same (or a better) level of etch selectivity and profile control for the metal gate compared to the polysilicon gate. In addition, thermal stability issues (from the source/drain dopant activation anneal) must be addressed. In either case, if metals with two different work functions are employed for n-FET and p-FET, respectively, the integration of n-FET and p-FET in a CMOS process remains a challenge, since 1) the deposition of the metals for n-FET and p-FET must be done separately, and 2) one must find a way to strap the two different metals in a compact way to connect the n-FET and p-FET gates. It is desirable to circumvent these two requirements and find a way to alter the work function of the metal by some simple means (for example, one that requires only a block mask). Two interesting approaches, yet to be proven through more rigorous examination, have been reported. In the first approach, a single metal (molybdenum) is deposited, and the work function is altered using ion implantation of nitrogen into the metal [36]. It is not clear how the nitrogen influences the work function, and how thermally stable this material is. On the other hand, ion implantation is an attractive process, since it requires only a single metal deposition and a photoresist block mask. In another approach, metals are intermixed to obtain the desired work function [37]. Two metals (Ti and Ni) are sequentially deposited on the gate dielectric, followed by selective etching of the top metal, leaving the bottom metal at desired locations. After thermal annealing, the metal on the top migrates to the metal/gate-dielectric interface and alters the work function locally.

[1] The use of polySiGe gates can tailor the work function near the valence band somewhat, using the dependence of bandgap on the Ge fraction [29].

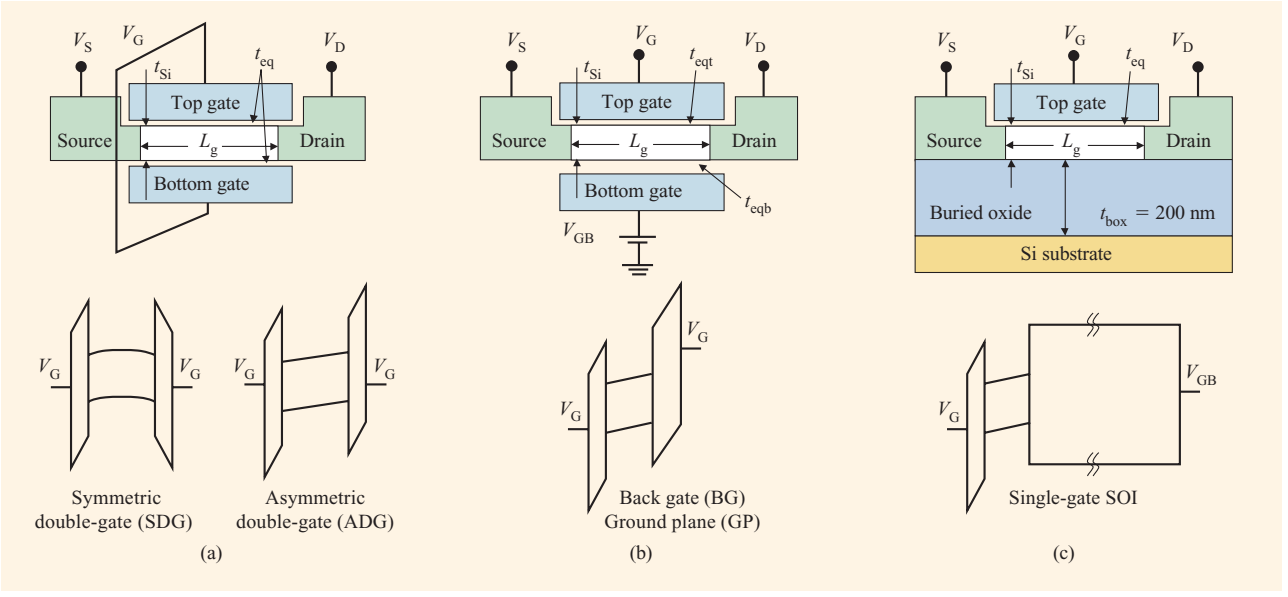


Figure 4

Conceptual device schematics of (a) double-gate, (b) ground-plane (or back-gate FET, BG), and (c) single-gate SOI MOSFET. On-chip biasing of the ground plane is assumed The upper figures are cross-section schematics of the devices; the lower figures illustrate their respective band diagrams. The gate work functions of the top and bottom gates can be the same (symmetric DG FET, or SDG) or different (asymmetric DG FET, or ADG). Adapted with permission from Wong et al. [50]; © 1998 IEEE.

Ultimately scalable FET—the double-gate FET

Device concepts

The double-gate FET (DG FET) shown in **Figure 4**, part (a) was proposed in the early 1980s [38]. The concept has been gradually explored both experimentally and theoretically by many groups [39–46]. The Monte Carlo and drift-diffusion modeling work by Fiegna et al. [41] and Frank et al. [42] clearly showed that a DG FET can be scaled to a very short channel length (25 to 30 nm) while achieving the expected performance derived from scaling. While the early work focused on the better scalability of DG FET, recent work suggests that the scalability advantage may not be as large as previously envisioned [10, 47], although the carrier transport benefits may be substantial. In this section, we first discuss the advantages of DG FET, followed by device design requirements, and conclude with a summary of latest hardware results.

The salient features of the DG FET (Figure 4) are [48] 1) control of short-channel effects by device geometry, as compared to bulk FET, where the short-channel effects are controlled by doping (channel doping and/or halo doping); and 2) a thin silicon channel leading to tight coupling of the gate potential with the channel potential. These features provide potential DG FET advantages that include 1) reduced 2D short-channel effects leading to a shorter allowable channel length compared to bulk FET; 2) a sharper subthreshold slope (60 mV/dec compared to >80 mV/dec for bulk FET) which allows for a larger gate overdrive for the same power supply and the same off-current; and 3) better carrier transport as the channel doping is reduced (in principle, the channel can be undoped). Reduction of channel doping also relieves a key scaling limitation due to the drain-to-body band-to-band tunneling leakage current. A further potential advantage is more current drive (or gate capacitance) per device area; however, this density improvement depends critically on the specific fabrication methods employed and is not intrinsic to the device structure.

The most common mode of operation of the DG FET is to switch the two gates simultaneously. Another use of the two gates is to switch only one gate and apply a bias to the second gate to dynamically alter the threshold voltage of the FET[2] [49, 50]. In this mode of operation, called "ground plane" (GP) or back-gate (BG), the subthreshold slope is determined by the ratio of the switching gate capacitance and the series combination of the channel capacitance and the nonswitching gate capacitance, and is generally worse than the DG FET. A thin gate dielectric at the nonswitching gate reduces the voltage required to

[2] One should note that the threshold voltage adjustment is primarily effective in the reverse-bias condition (raising the threshold voltage) where the back-gated channel remains in depletion and not inverted.

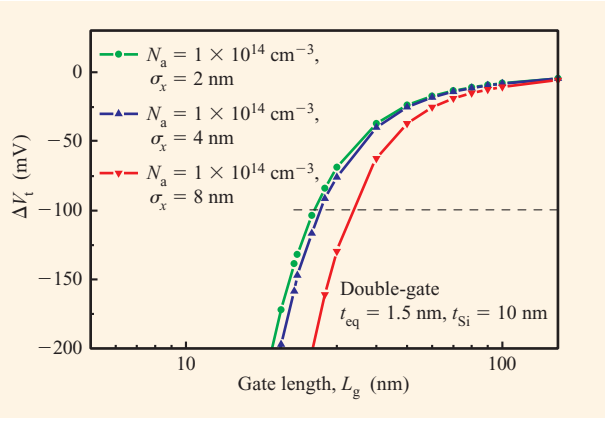


Figure 5

Threshold voltage roll-off characteristics of double-gate FET with different junction gradients, illustrating the importance of maintaining a sharp doping profile for DG FET even though the junction depth is no longer important for DG FET. The gradient is an analytic Gaussian profile with a lateral $\sigma_x = 2, 4, 6$ nm. The silicon channel thickness t_{Si} is 10 nm and the equivalent gate dielectric thickness t_{eq} is 1.5 nm.

adjust the threshold voltage and preserves the drain-field-shielding advantage of the double-gate device structure. However, a thinner gate dielectric also means extra capacitance that does not contribute to channel charge for switching. Since the back-gate FET is very similar to a single-gated SOI FET with an adjustable threshold voltage [49], we focus our discussion here on the DG FET configurations in which both gates are switched.

Double-gate FET electrostatics

The double-gate device structure allows for termination of the drain electric field at the gates and leads to a more scalable FET. To evaluate the scalability of FETs, the concept of the "scale length" for a MOSFET is useful [10, 43, 44, 51]. The electrostatic potential of the MOSFET channel can be approximated by analytically solving the 2D LaPlace equation using the superposition principle (with suitable boundary conditions), and the short-channel behavior can be described by a characteristic "scale length," λ. Table 3.3 of [52] lists the generalized scale length derived by Frank et al. [10, 51] and the simpler, but less accurate, scale length derived by Suzuki et al. [44]. The minimum gate length is jointly determined by the scale length and by the amount of 2D short-channel effects one can tolerate in an application. The 2D short-channel effects can range from increased off-current due to threshold-voltage roll-off, drain-induced barrier lowering (DIBL), and degraded subthreshold slope, to degraded output resistance. Figure 5 of [10] illustrates the trend of these 2D effects as the channel length is decreased with respect to the scale length of the MOSFET. Manufacturing tolerances put a premium on the minimum channel length. With typical tolerances of 20–30% gate-length variation, an L/λ of 1.5 is required.[3] Conventional short-channel-effect theory [23] correlates the junction depth to the short-channel effects. In the case of the DG FET, consideration of junction depth is moot, since the 2D electrostatic behavior is controlled by the thickness of the silicon channel instead of the junction depth. However, the steepness of the source/drain junction is still an important consideration, as in the case of bulk FETs [47]. **Figure 5** illustrates the threshold-voltage roll-off characteristics of the DG FET with lateral junction profile gradients of 2, 4, and 6 nm (Gaussian analytical profile). It is clear that a steep junction gradient commensurate with the channel length is required.

Comparing scale lengths for the DG FET, ultrathin silicon SOI FET, and bulk devices, as well as considering other leakage mechanisms (such as tunneling leakages), leads to the conclusion that the DG FET can be scaled up to 50% further than the bulk FET for some applications [10]. Illustrations of the threshold-voltage roll-off behavior (an example of 2D short-channel effects) comparing DG FET, ultrathin-silicon FET, and ground-plane FET (in which the bottom gate of a DG FET is tied to a fixed bias) can be found in [10, 50] and many other references in the literature and are not repeated here. Similar analyses based on on-current and off-state subthreshold leakage current can be found in [53, 54]. Simply put, the better scalability of DG FET can be used to achieve a shorter channel length using the same gate-oxide thickness, or the same channel length using a thicker gate oxide.

We now turn our discussion to the relationship of the channel inversion charge and the gate voltage. The analytical model of Taur [55] and the numerical modeling of Ieong et al.[4] [56] form the basis for much of this discussion. The gate work function of the two gates can be the same (the SDG, with a symmetric energy-band diagram in the direction normal to the gate electrode) or different (the ADG, with an asymmetric energy-band diagram in the direction normal to the gate electrode), as illustrated in Figure 4. In the subthreshold region, where there is negligible inversion charge, the silicon channel is fully depleted, and the energy bands closely follow the gate bias in a one-to-one relationship. For the SDG, the bands remain flat throughout the subthreshold region, since there is little or no depletion charge. Once inversion charge begins to build up, the mobile charges screen the

[3] As discussed by Frank et al. [10], the "end of scaling" depends on the application at hand, which determines the amount of deleterious 2D short-channel effects one can tolerate. The above $L/\lambda \simeq 1.5$ rule should be considered only a rough guideline.

[4] M. Ieong and H.-S. P. Wong, "Analysis of 25 nm Double-Gate MOSFETs Including Self-Consistent 2-D Quantization Effects," unpublished work, 1999.

gate field and the gate voltage is dropped primarily in the inversion layer. The threshold voltages for the SDG and the ADG with various gate work functions are plotted in **Figure 6** as a function of the silicon channel thickness. The threshold voltage of the SDG has a small channel thickness dependence (≃6 mV/nm), while the threshold voltages of the ADG and BG have a larger dependence on channel thickness (≃28–32 mV/nm) because of the asymmetric band diagram.[4] At the same off-current (same integrated mobile charge in the subthreshold region), the surface potential of the ADG is higher than that of the SDG. In other words, the surface of the ADG is inverted more than the surfaces of the SDG at the same off-current condition. This is because the SDG has a fairly constant charge density throughout the silicon film, while the ADG charge density peaks toward one of the surfaces. This surface-potential difference is carried forward throughout the entire gate-voltage range well into the fully inverted condition in which both surfaces of the ADG are in inversion [55].

For the SDG with a channel thickness greater than 5 nm, there are two distinct charge-density peaks near the two surfaces, and the two inversion layers are basically independent of each other (see the inset of Figure 6) [56]. For a channel thickness less than 5 nm, the two inversion charge peaks begin to merge. For the ADG, the inversion charge forms first at the surface, where the gate work function is lower. Although the ADG has only one predominant channel, the total integrated charge is more than half that for the SDG. The SDG-to-ADG charge ratio is about 2× for a thick silicon channel and approaches 1× for very thin channels. The physics behind this observation is best explained by using the capacitive coupling model of Taur [55]. The ADG gate with a higher gate work function (say, back-gate) induces inversion charge at the opposite surface through the capacitive coupling of the series combination of the back-gate dielectric ($C_{oxb} = \epsilon_{ox}/t_{oxb}$) and the channel capacitance ($C_{Si} = \epsilon_{Si}/t_{Si}$). The total gate capacitance of the ADG is therefore more than the front-gate dielectric capacitance ($C_{oxf} = \epsilon_{ox}/t_{oxf}$) alone. The amount of back-gate coupling obviously depends on the back-gate dielectric thickness and the silicon channel thickness. The thinner the silicon channel with respect to the gate dielectric, the more effective the coupling, and the closer the SDG-to-ADG charge ratio is to unity.

The quantization in the channel introduces several interesting effects, which are discussed here. The threshold-voltage increase due to quantum effects is illustrated in Figure 6 [50] using a simple particle-in-a-box approximation (more accurate in SDG than BG or ADG): $\Delta V_t = -(h^2/4qm^*t_{Si}^2)(\Delta t_{Si}/t_{Si})$ [57], where V_t is the threshold voltage, h is Planck's constant, q is the electronic charge, m^* is the carrier effective mass, and

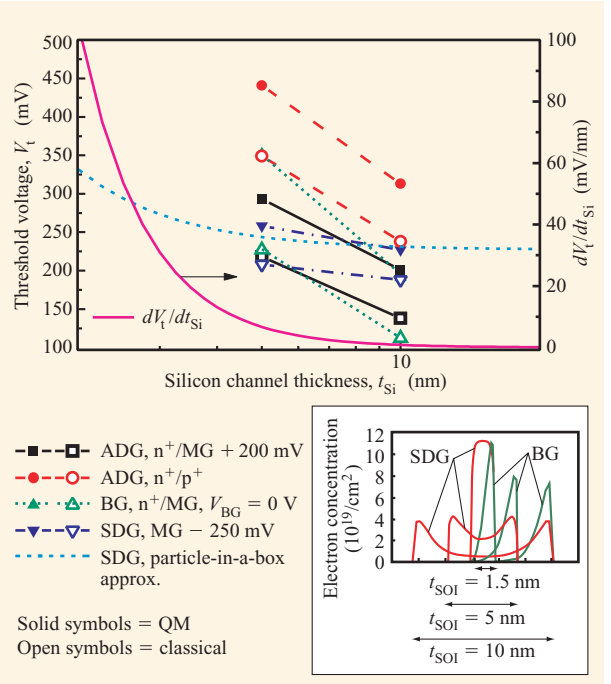


Figure 6

Long-channel threshold voltages computed using a classical and a quantum-mechanical (coupled Schrödinger–Poisson solution) description of the channel for various gate work function combinations. The equivalent gate oxide is 1.5 nm. Adapted with permission from Ieong et al. [56]; © 2000 IEEE. Left axis: Threshold voltage (long channel, with and without QM effects) as a function of the silicon channel thickness. Note that for devices with an asymmetric band diagram (BG, ADG), the dependence of the threshold voltage on silicon channel thickness is larger than SDG because of capacitor divider effect. The estimated shift of the threshold voltage due to quantum confinement of the thin silicon channel according to the particle-in-a-box approximation is added to the classical threshold voltage for t_{Si} = 10 nm and is shown as a dashed curve. It agrees well with the full quantum model. The gate work functions are ADG n^+ and mid-gap plus 200 mV (squares), ADG n^+ and p^+ (circles), BG n^+ and mid-gap back-gate biased at zero volts (up triangles), and SDG both gates at mid-gap minus 250 mV (down triangles). Right axis: The sensitivity of V_t to t_{Si} (dV_t/dt_{Si}) due to quantum effect (particle-in-a-box approximation). The inset shows the electron charge density (computed using the aforementioned quantum-mechanical model) for silicon channel thicknesses of 1.5 nm, 5 nm, and 10 nm at $V_g - V_t$ = 0.8 V. Adapted with permission from Wong et al. [50], © 1998 IEEE; and Ieong et al. [56]; © 2000 IEEE.

t_{Si} is the channel thickness. The quadratic increase of the threshold voltage with decreasing channel thickness (steep rise of dV_t/dt_{Si}) means that channel thickness much below 5 nm will be almost inpractical to manufacture unless an atomically precise method of defining the channel thickness is found. In a more realistic approximation, one solves the one-dimensional Schrödinger equation and

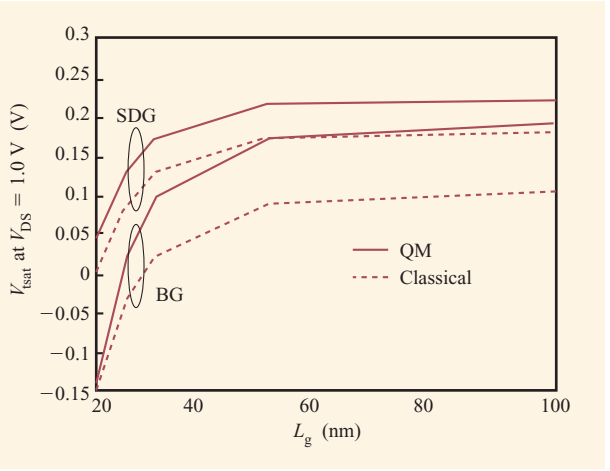


Figure 7

Saturated threshold voltage roll-off as a function of channel length for symmetric double-gate FET (SDG) and ground-plane FET (BG), illustrating the effect of quantization in the channel. Adapted with permission from Ieong et al. [56]; © 2000 IEEE.

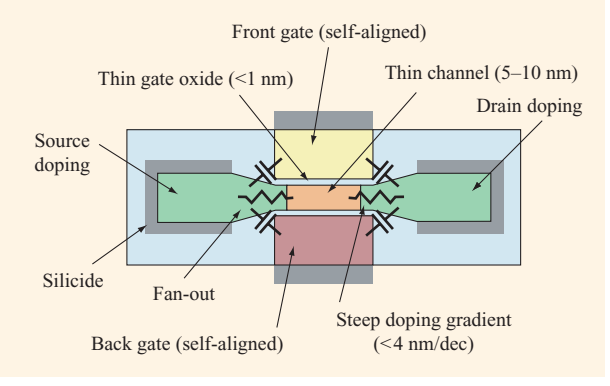


Figure 8

Schematic cross section of an ideal double-gate FET.

incorporates the solution self-consistently in a coupled 2D Poisson and continuity equation solution (with appropriate boundary conditions) [56]. Figure 6 compares the threshold voltages computed from a classical and quantum-mechanical description of the channel for several cases of gate-work-function values. For gates with a symmetric work function, the work function of the gates should be about 250 to 350 mV above/below mid-gap for n/p channels, respectively, for high-performance applications. Another strategy for setting the threshold voltage is to employ a set of asymmetric work-function gates: a) a front gate with a work function close to the conduction band ("n^+ gate") and a back gate with a work function close to the valence band ("p^+ gate"); and b) a front gate with a work function close to the conduction band (valence band) for n-FET (p-FET) and a back gate with a common mid-gap work function (for both n-FET and p-FET). Both of these approaches provide a symmetric threshold voltage for both n-channel and p-channel DG FETs. Obviously, the work-function requirement for the ADG case also depends heavily on the gate dielectric and silicon channel thicknesses. A typical gate dielectric (1 nm) and silicon channel (5–10 nm) thickness required for a sub-50-nm DG FET gives a threshold voltage that is too high if the n^+/p^+ ADG (case a) is used, while case b above has a more appropriate threshold voltage for high-performance applications.

Figure 7[4] [56] shows two noteworthy effects from quantum behavior: 1) threshold-voltage shifts due to quantum effects are larger in DG FETs with asymmetric work-function gates (BG FET and ADG) compared to SDG (see also the values shown in Figure 6); and 2) threshold-voltage roll-off is worse when quantum effects are included. The first observation can easily be understood, since the band diagrams of the BG FET and the ADG are asymmetric (with a high normal electric field), which forces the charge carriers toward one of the surfaces. The second observation is more subtle: As the channel length is shortened, classical (Poisson equation) short-channel effects reduce the normal electric field, thereby reducing the quantization effects ("opening up" the channel potential). This effect is, in fact, more apparent in the BG FET. A similar effect can also be observed in the ADG [56].

Taking into account the short-channel design considerations in [50, 53], the design space for a sub-20-nm DG FET is summarized as follows (**Figure 8**): a channel thickness of 5–10 nm, a gate dielectric of less than 1 nm (equivalent electrical thickness), a source/drain doping profile with less than 4 nm/decade lateral gradient, a highly doped source/drain fan-out structure to reduce series resistance, and a set of self-aligned front and back gates to minimize gate to source/drain overlap capacitances. The "alignment" of the gates refers to both front and back gates aligned with respect to each other as well as to the source/drain doping.

The requirement for a set of self-aligned gates is underscored by the study in [46], which we summarize here. Two scenarios for a non-self-aligned DG FET are 1) misaligned (offset) top and bottom gates of equal size (minimum bottom gate) [45], and 2) an oversized bottom gate to ensure gate overlap of the source/drain [58] [see **Figure 9**, part (b)]. The minimum bottom-gate approach

[4] M. Ieong and H.-S. P. Wong, "Analysis of 25 nm Double-Gate MOSFETs Including Self-Consistent 2-D Quantization Effects," unpublished work, 1999.

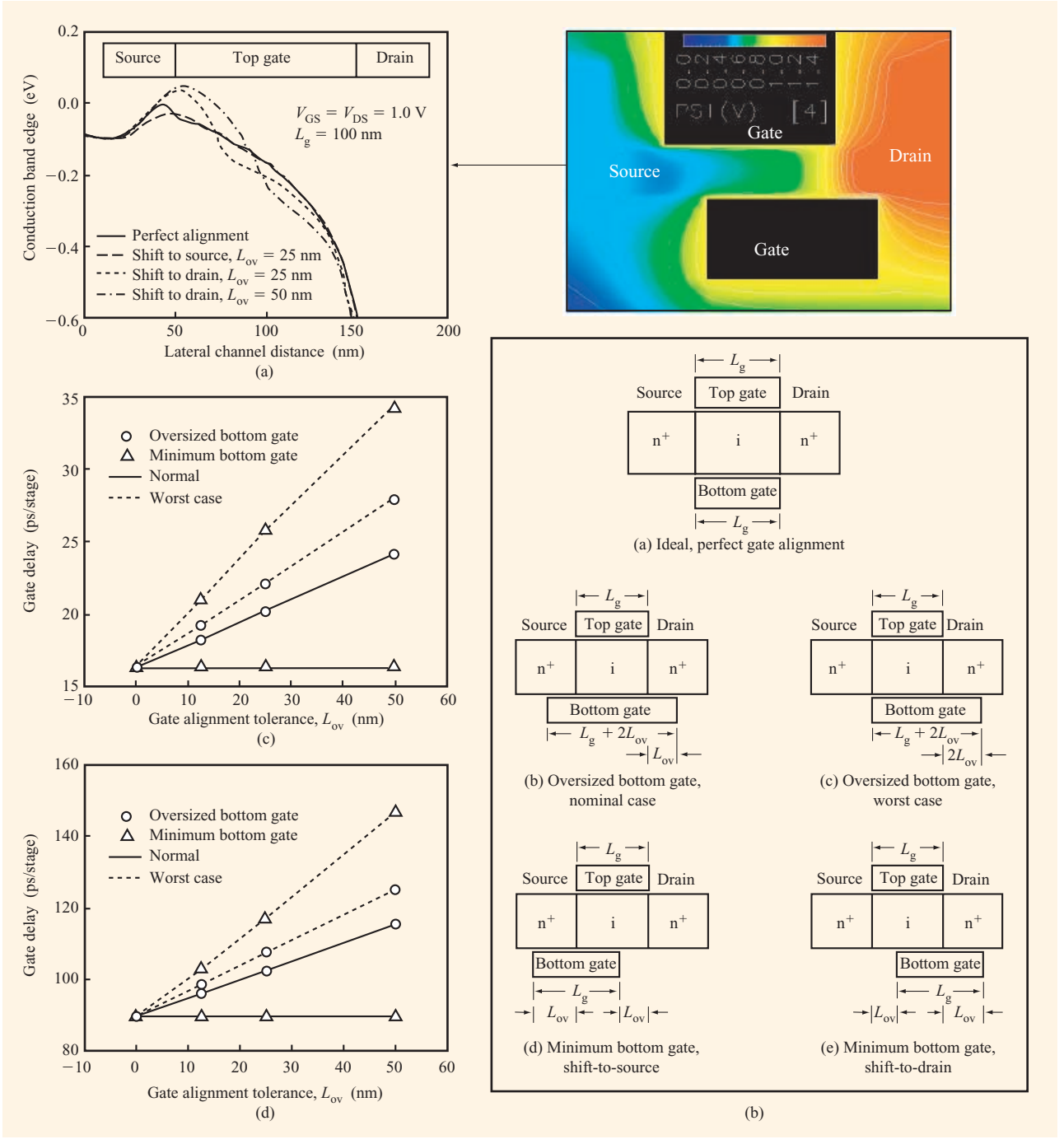


Figure 9

(a) Potential along the back channel from source (left) to drain (right), showing a higher source-side energy barrier in the misaligned minimum-bottom-gate configuration shift-to-drain case which limits the current drive (inset: 2D potential contours of the shift-to-drain case). (b) Illustrations of the possible cases of gate misalignment. MBG = minimum bottom gate, OBG = oversized bottom gate.) (c), (d) Simulated gate delay as a function of the gate alignment tolerance (L_{ov}). Nominal gate length is 100 nm. (c) Unloaded inverter gate delay with fan-in and fan-out of 1. The nominal delay degradation is 24% for OBG, and the worst-case delay degradation is 36% for OBG and 58% for MBG, for an overlay tolerance of 25 nm. (d) Loaded (0.1-pF) three-way NAND gate delay with fan-in and fan-out of 3. The nominal delay degradation is 14% for OBG, and the worst-case delay degradation is 25% for OBG and 30% for MBG, for an overlay tolerance of 25 nm. A 2D device simulator is used to estimate the current–voltage characteristics, which in turn are used in a circuit simulator to estimate the gate delay. Adapted with permission from Wong et al. [46]; © 1994 IEEE.

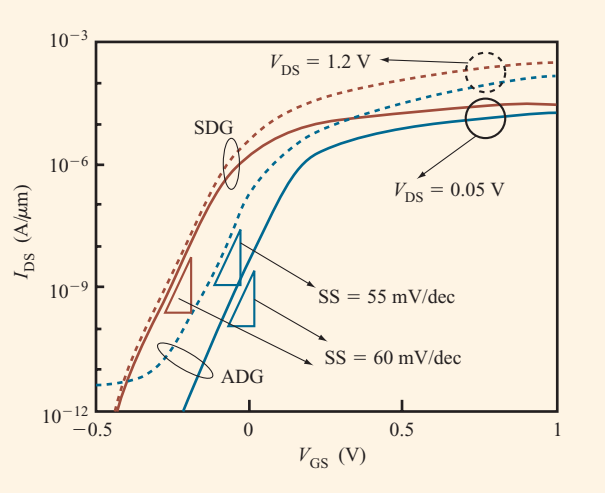


Figure 10

Measured drain current vs. gate voltage characteristics of long-channel ($L = 1\ \mu$m) symmetric (SDG) and asymmetric (ADG) double-gate FET, illustrating the ideal subthreshold slope attainable by double-gate FETs. Device fabrication and parameters are as reported in [61, 155]. Adapted with permission from Ieong et al. [61]; © 2001 IEEE.

minimizes overlap capacitance. The energy barrier for carrier injection from the source for the misaligned gate [offset from the source case; see Figure 9, part (b)] is significantly higher than the nominal, aligned case. The higher source-side energy barrier limits the on-current and degrades the gate delay. The gate-delay degradation amounts to more than a full generation of device scaling performance gain [Figure 9, parts (c)–(d)]. A fully self-aligned fabrication process (both gates aligned with each other and with the source/drain doping) is therefore required for the highest performance benefits. If misalignment is unavoidable, the study in [46] indicates that it is more important to have a large enough bottom gate to ensure adequate gate-to-source overlap in order to attain a high on-current than to minimize parasitic capacitance.

Double-gate FET carrier transport

So far, most of the discussion has been focused on the electrostatics of the FET. We now turn our attention to device design issues including carrier transport. Because of the extremely small depletion capacitance, the gate-to-channel potential coupling is not de-rated by the capacitor divider between the gate-oxide capacitance and the depletion capacitance. The subthreshold slope is therefore 60 mV/decade in the absence of short-channel effects (**Figure 10**). For the same off-current, the threshold voltage can be set about 60 mV lower than bulk FETs, thereby providing more gate overdrive—an important advantage as the power supply is reduced.

Carrier mobility for the DG FET with a thin silicon channel deserves further discussion. Early work on single-gated SOI [59] demonstrated that carrier mobility in SOI follows a "universal mobility" curve similar to that of bulk FETs [60] when the "effective field" is properly taken into account in a single-gated SOI FET. Recent work using a double-gated FET structure has also indicated that the carrier mobility follows the same "universal mobility" behavior for both front-gated and back-gated channels, provided that the gate dielectrics on both sides are of high

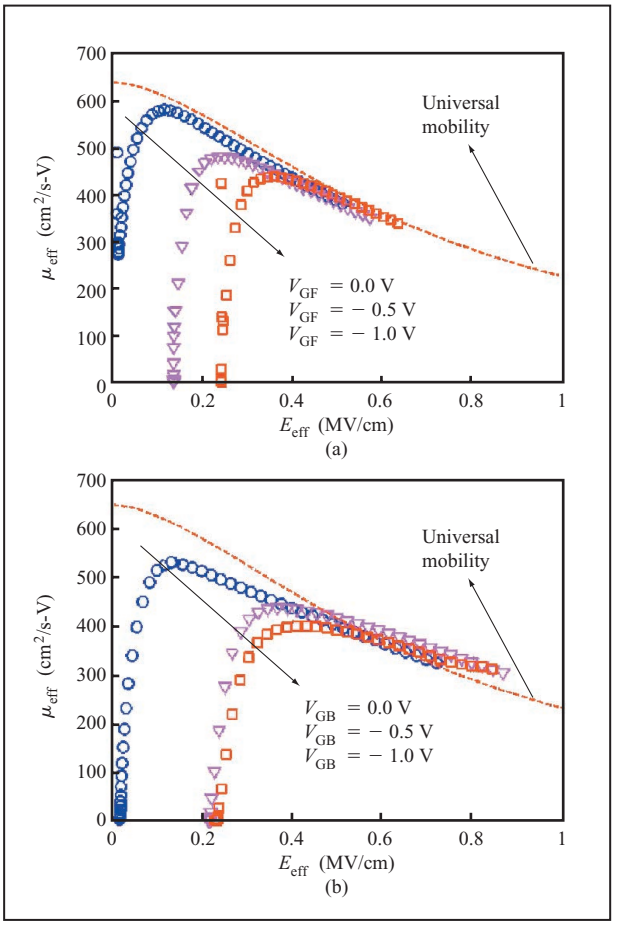


Figure 11

Effective electron mobility vs. effective field for (a) a back-gated channel with various front-gate bias voltages ($V_{GF} = 0.0, -0.5, -1.0$ V), (b) a front-gated channel with various back-gate bias voltages ($V_{GB} = 0.0, -0.5, -1.0$ V). The effective electron mobility follows the conventional "universal mobility" curve even as the surface electric field is varied by varying the back biases. The silicon channel is 44 nm, with a front-gate oxide of 2.4 nm and back-gate oxide of 3 nm. Other device parameters and fabrication details are as reported in [61, 155]. Adapted with permission from Ieong et al. [61]; © 2001 IEEE.

quality and the "effective field" is properly accounted for [61] (see **Figure 11**). However, early reports of carrier mobility measurement for very-thin-silicon-channel FETs (single-gated SOI FET) show a significant degradation of mobility as the silicon channel thickness is decreased below 20 nm [62–64]. Subsequent theoretical calculations suggest a complex behavior of mobility as a function of silicon channel thickness attributed to phonon scattering in the thin confined silicon channel [65–67]. Both findings raise legitimate concern for the DG FET [10], since the thin silicon channel required for control of short-channel effects may result in poor carrier transport. Recent experimental mobility measurements [68] for thin-silicon-channel FETs down to 5 nm as a function of carrier density depict a more complete picture. First, the early experimental results [62–64] may be tempered by a poor back-channel oxide because the thin silicon channel of the single-gated SOI FET was obtained by oxidation thinning from a SIMOX wafer which is known to have a substandard back-channel oxide interface that can degrade carrier transport. In addition, the mobility at low inversion carrier concentration (or, equivalently, low effective electric field) was reported. **Figure 12** summarizes the data to date [62–64, 68]. The electron mobility degradation at channel thicknesses below 20 nm is clearly observable at a low inversion carrier density ($N_{inv} = 10^{12}$ cm^{-2}). However, at the higher inversion carrier density ($N_{inv} = 10^{13}$ cm^{-2}) that is important for nanoscale CMOS, the electron mobility is less sensitive to the silicon channel thickness. While the above results are for single-gated SOI FETs, similar results have been obtained for double-gate operation [69]. The degradation at low inversion carrier density appears to be related to phonon scattering, suggested by the larger degradation seen at lower temperatures [69].

Carrier transport in the DG FET with an undoped channel is superior to that in conventional bulk FETs for two reasons: reduced Coulomb scattering due to fewer ionized dopants in the undoped/low-doped channel, and reduced surface roughness scattering due to a lower surface electric field, as illustrated in **Figure 13**. In bulk FETs, channel doping is employed to set the threshold voltage, and halo or pocket dopings are employed to control the short-channel effects. These ionized depletion charges contribute appreciably to the surface electric field. In a DG FET with an undoped channel, there is no ionized depletion charge; therefore, the surface electric field is contributed entirely from inversion carriers (application of Gauss's law). Even though the carrier mobility follows the "universal mobility" curve [Figure 13(a)], at the same gate overdrive the carrier mobility can be significantly higher [Figure 13(b)] because the effective

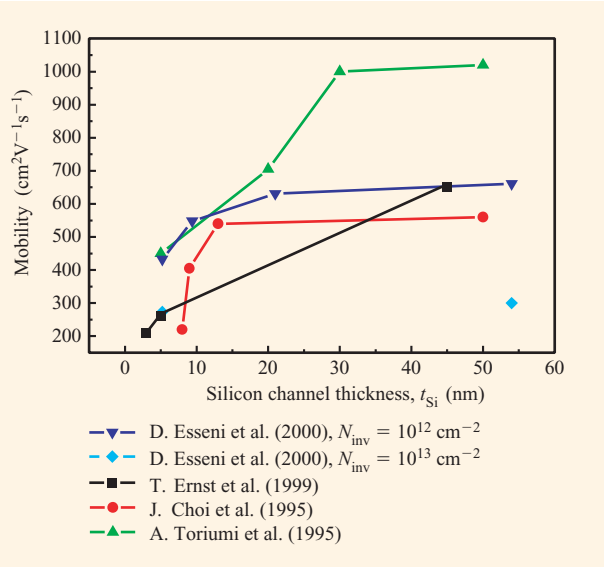


Figure 12

Experimentally measured electron mobilities in thin silicon channels drop substantially below about 10-nm channel thickness. The lines are visual guides and do not suggest trends in the data. The mobility data of Choi et al. [62] and Toriumi et al. [63] are the peak mobility at low effective fields. The electric field corresponding to the mobility data of Ernst et al. [64] was not specified in their paper and is presumed to be the low-field mobility. Data from Esseni et al. [68] indicate that while there is a mobility dependence on silicon channel thickness at low inversion carrier densities, the thickness dependence diminishes at larger carrier densities.

field is lower at the same gate overdrive. Figure 13(a) also shows the mobility and range of effective field for a typical bulk FET. While the bulk FET operates at an effective field above 1 MV/cm, the DG FET with an undoped channel operates at around 0.5 MV/cm, thereby improving the mobility by almost two times. This improved transport potentially provides the DG FET a better CV/I metric because although the capacitance C is doubled in a DG FET, the current I is improved by more than two times because of the better transport.

Device fabrication

Fabrication of the DG FET is difficult. Early experimental work began with non-self-aligned DG FET structures for a first demonstration of the device principles [39, 40, 45, 58, 70], the most advanced being the work of Tanaka et al. [45, 58] where non-self-aligned DG FET circuits were demonstrated. In this section, we focus on the more recent development of the self-aligned DG FET because of its better performance, as described in the section on double-gate FET electrostatics.

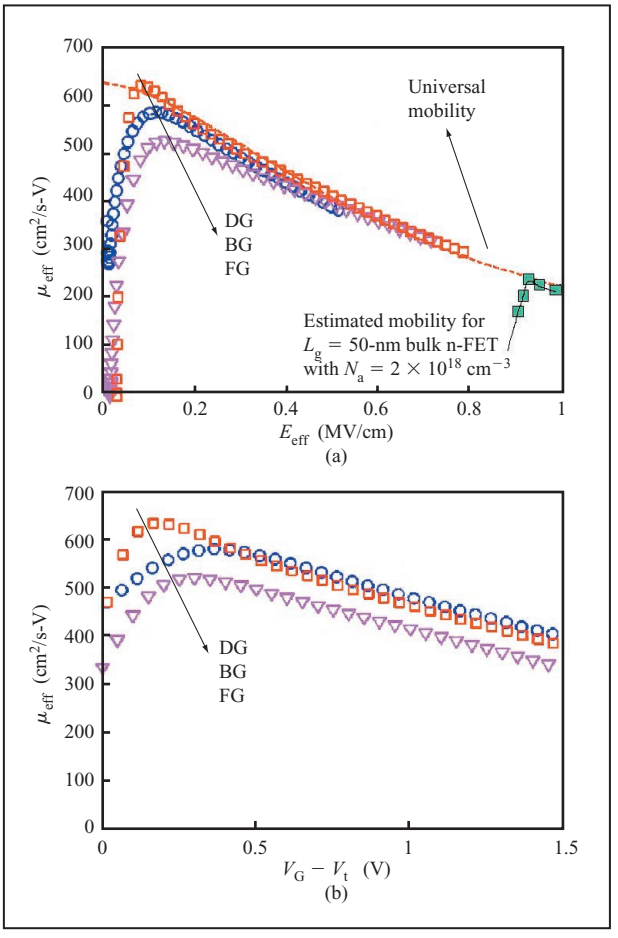
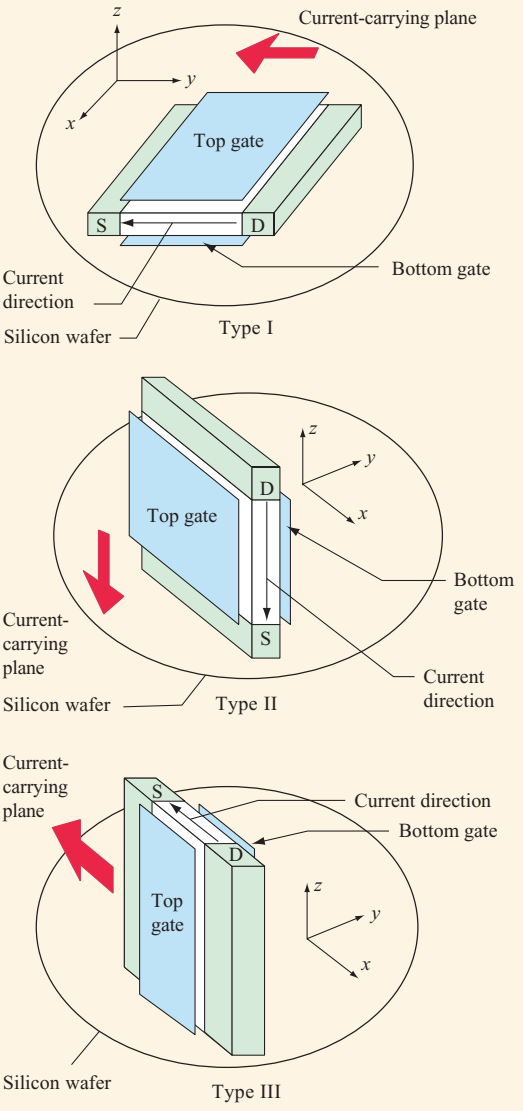


Figure 13

Measured effective electron mobilities for a DG FET (DG) and for a back-gate FET measured using the front gate as the active channel (FG) and the back gate as the active channel (BG) while leaving the other gate at zero bias, as reported in [61]. (a) The effective electron mobilities for DG, FG, and BG FETs follow the "universal mobility" curve. (b) At the same gate overdrive ($V_G - V_t$), the effective mobility for the undoped channel DG FET (and FG and BG as well) is higher than for conventional bulk FETs. This is because the effective field is lower for undoped channel FETs. Adapted with permission from Ieong et al. [61]; © 2001 IEEE.

In general, the DG FET may be fabricated with any of the three orientations depicted in **Figure 14** [9, 48, 71]. The planar structure (Type I) has the advantage of better silicon channel thickness uniformity because film thickness in the plane of the wafer has the best uniformity and controllability. However, the fabrication of a back gate and a thin gate dielectric underneath a single-crystal silicon channel is difficult. In addition, accessing the bottom gate from the top surface for device wiring is not straightforward, and may have a negative impact on device density. The nonplanar structures (Types II and III) allow



Type	Source-to-drain direction	Gate-to-gate direction	Gate length control	Channel thickness control	Top area
I	IP	NPT	Lith/etch	Planar layer	$L \times W$
II	NTP	IP	Planar layer	Lith/etch	$W \times H$
III	IP	IP	Lith/etch	Lith/etch	$L \times H$

L = gate length, W = gate width, H = channel thickness, IP = in plane, NTP = normal to plane

Figure 14

Three possible topologies of the double-gate FET. Adapted with permission from Wong et al. [71]; © 1997 IEEE.

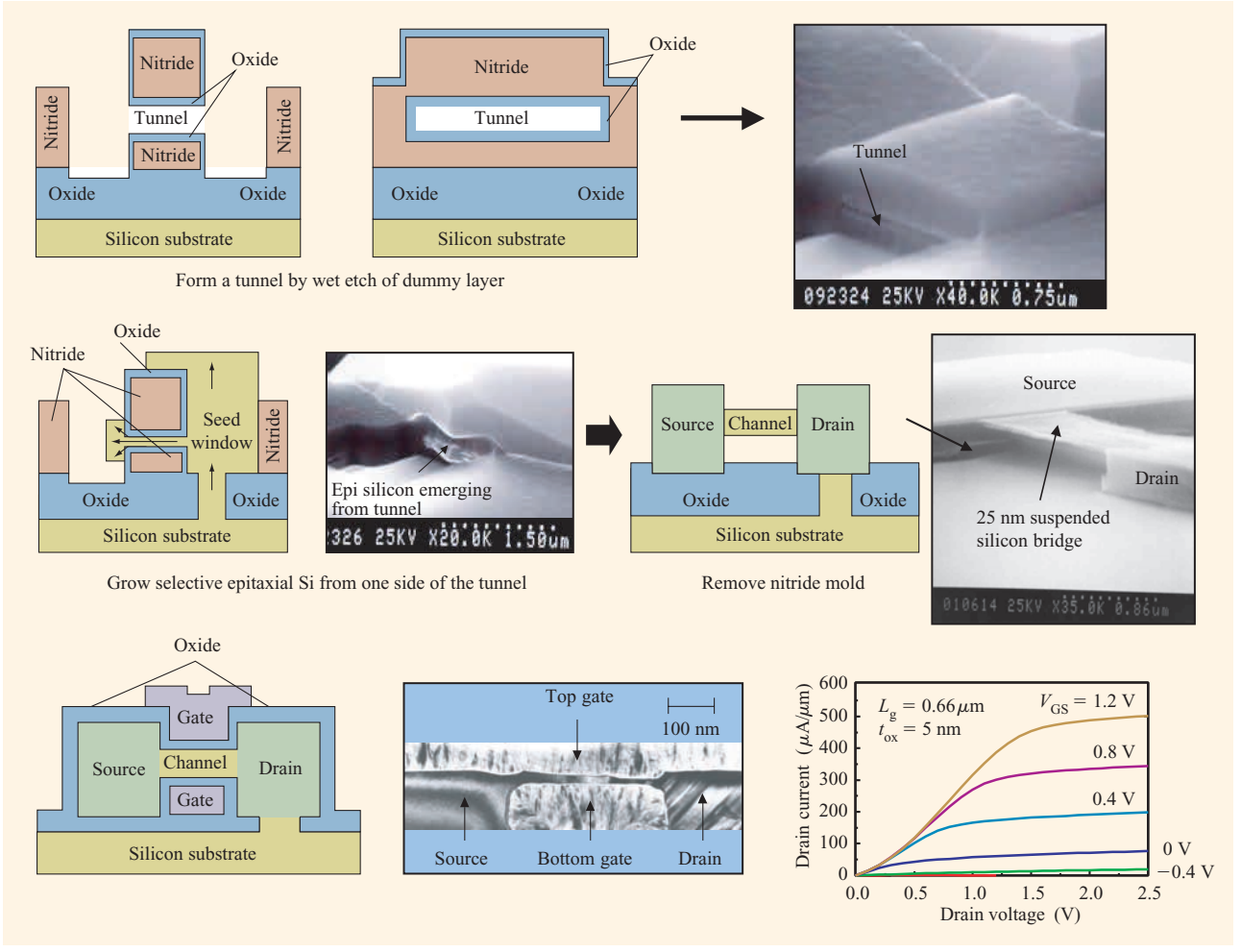


Figure 15

Fabrication sequence and device characteristics of a self-aligned double-gate FET using selective epitaxial growth. Adapted with permission from Wong et al. [71]; © 1997 IEEE.

for easier access and formation of both gates (or a wraparound gate) on crystalline channels with thin gate dielectrics. On the other hand, the channel thickness is defined by lithography and patterning techniques (e.g., reactive ion etching), and may therefore have poorer uniformity than planar films. The device electrostatic design (see the section on double-gate FET electrostatics) requires that the channel thickness be about 1/3 to 1/4 of the channel length. Historically, the smallest dimension patterned on a chip is the gate length. For the nonplanar DG FET structures, the smallest dimension patterned must be considerably smaller than the channel length, which is a major departure from conventional processes. Carrier transport along etched surfaces with different crystallographic orientations [e.g., (110) surface for a notch ⟨110⟩ wafer] may degrade performance, although there is little data to date to support these assumptions. While the topography of Type II and III nonplanar structures may raise some fabrication concerns, it should be noted that this topography problem may not be as severe as it appears, because the height of the vertical structure can be made similar to that found in the gate stacks of planar structures.

The buried gate of the planar, Type I device structure has been demonstrated using two techniques: selective epitaxial growth [71, 72], and wafer bonding and layer transfer [73, 74]. We illustrate these approaches using the two examples below.

Figure 15 shows the fabrication sequence for the first self-aligned DG FET. The fabrication process utilizes a

planar CVD-deposited film in a dummy gate stack as a placeholder for the silicon channel. This dummy gate stack is etched out (forming a tunnel) and then replaced by a single-crystal silicon channel formed by selective epitaxial silicon growth through the tunnel with a seed from one side of the tunnel to prevent grain-boundary formation in the middle of the silicon channel. Excess epitaxial growth is polished away by chemical–mechanical polishing (CMP). Since the selective epitaxial silicon growth is confined to the dielectric tunnel [75–79], the thickness of the silicon channel is determined by the previously deposited CVD film (which has good thickness uniformity). Despite the seemingly difficult task of filling a long, thin tunnel, experimental evidence shows that large aspect ratios of more than 220:1 can be filled [79], well in excess of the requirements for the electrostatic device design for short-channel effects. After the formation of the silicon channel and subsequent CMP planarization to remove excess grown silicon, the source/drain is formed using ion implantation that is self-aligned to the dielectric dummy gate stack. The dummy gate stack is then removed, leaving a silicon channel in the form of a suspended bridge between the source and drain. Finally, the gate dielectric is formed, and the gate material is deposited and patterned. This "double-replacement" process provides a means to form a wraparound gate over the single-crystal silicon channel and self-align the source/drain doping to the channel and the gates. Both the silicon channel and the gate stack are formed by replacing a dummy material/structure with the final material/structure. The device characteristics reported suffer from high series resistance, possibly a result of non-optimized source/drain doping and silicide process. While this process possesses many positive attributes, as outlined above, it does have several shortcomings: 1) lack of an independently adjustable dielectric spacer thickness to separate the gate and the source/drain for parasitic capacitance reduction; 2) difficulty in doping and siliciding the underside of the wraparound gate; 3) the fact that the front and back gates cannot be independently biased; and 4) source/drain silicides that are not self-aligned to the gates. Some of these problems are addressed in the process proposed in [73, 80].

Parts (a)–(e) of **Figure 16** show the fabrication sequence of a triple self-aligned planar double-gate FET using wafer-bonding and layer-transfer techniques [74]. The starting substrate is an unpatterned bonded wafer with the doped polysilicon back gate and thin gate dielectric in place beneath the silicon channel. The bonded interface is 350 nm below the devices [Figure 16(e)]. Locating the bonding interface below the active device region is advantageous for two reasons: 1) any imperfections of the bonding process (stress fields, minor voids, and embedded particles) will not affect the active device; 2) any subsequent fabrication steps that etch to a level below the silicon channel will not expose the bonding interface, thereby avoiding the possibility of delamination of the bonded substrate. The undoped channel was initially thinned to 20–30 nm using oxidation, achieving uniformity to within 1 nm over much of an 8-inch-diameter wafer. The bottom gate and the top gates are separately patterned and accessible on opposite sides of the device width direction [Figure 16(f)]. CMP is employed to planarize the surface every time any topography is generated by patterning (etch). This eliminates the possibility of "stringers" and provides a robust process. For planarization at the front-end process, planarity of nanometer-scale accuracy and uniformity is required. Using custom-designed chemistry and processes, 2–3-nm topography over an 8-inch wafer is achievable. In order to provide access to the bottom gate and provide self-alignment of the bottom gate to the top gate and the source/drain, the source/drain fan-out regions are implemented as doped silicon sidewalls which are subsequently silicided. The sidewall source/drain is used as a self-aligned etch mask to center the length direction of the bottom gate with respect to the top gate and the sidewall source/drain. The undercut bottom gate is passivated with nitride dielectric. Contact to the sidewall source/drain is made by a tungsten plug to the source/drain well, followed by CMP to replanarize the surface. The difficult requirements of this fabrication approach include the following: 1) a precise and controllable back-gate undercut process, 2) a back-gate dielectric that is in place already during the wafer-bonding process and all subsequent process steps, and 3) a carefully controlled sidewall source/drain silicon deposition/etch and subsequent anneal, recrystallization, and dopant profile control via ion implantation of the sidewall and silicidation. Using a dry-etch technique, the back-gate undercut and sidewall source/drain etch can be better controlled. Device scaling requires that the back-gate dielectric be as thin as the front-gate dielectric. The use of a high-k gate dielectric is also a possibility. Both scenarios demand a low-temperature wafer-bonding process (preferably well below 900°C) for the back-gate dielectric to be viable.

Figure 16(g) shows the measured transconductance of a 125-nm-gate-length n-FET fabricated using the aforementioned triple-self-aligned DG FET process. The back gate (BG) is n^+-doped polysilicon, and the front gate (FG) is p^+-doped polysilicon. The double-gate (DG and DG*) mode of operation provides a transconductance that is slightly more than the sum of the FG and BG modes (single-gated, ground-plane operation). An ideal subthreshold slope of 60–62 mV/decade is measured for DG-mode operation. Mobility measurement of the front-gated channel and back-gated channel shows electron and

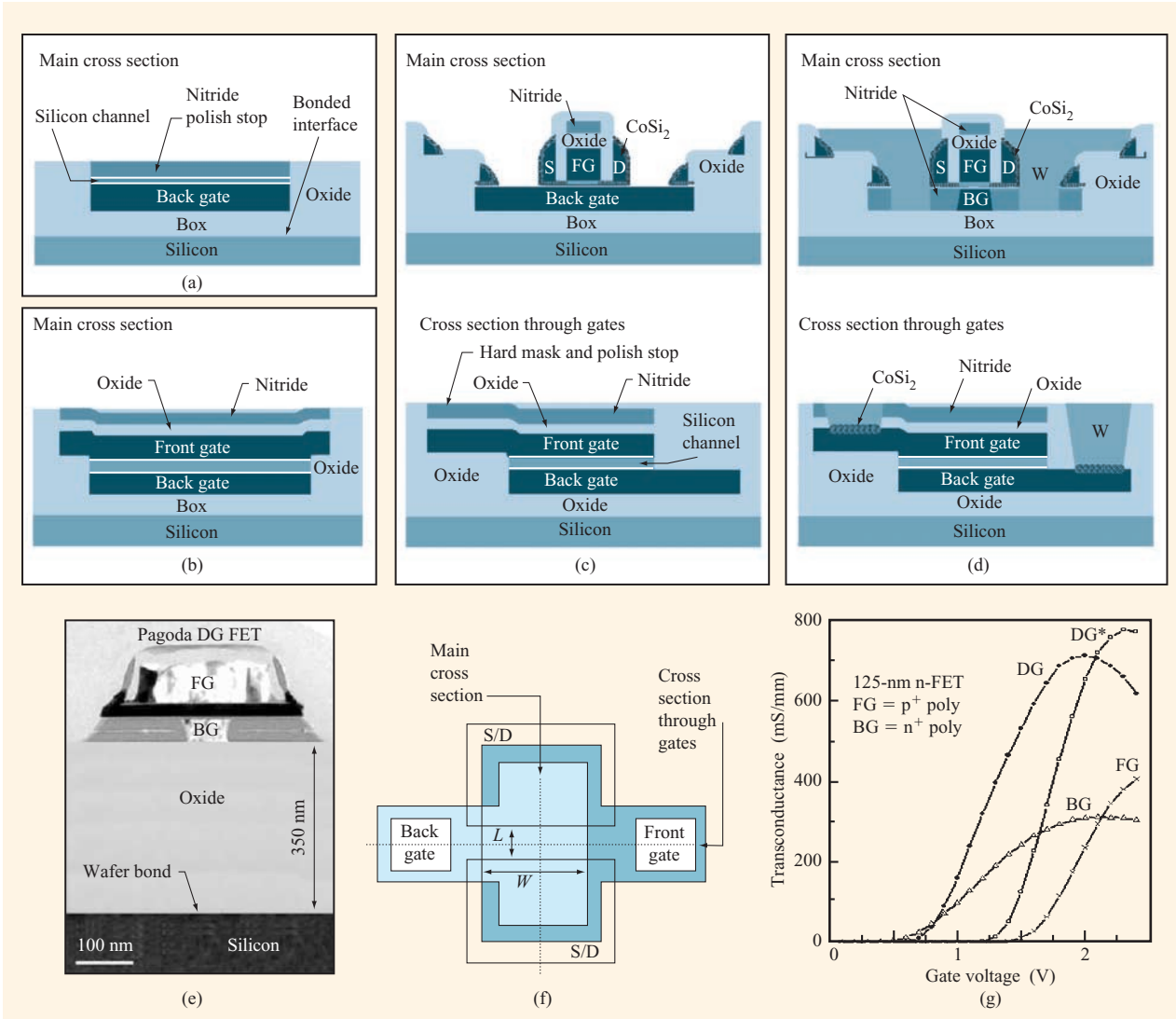


Figure 16

Fabrication sequence and device characteristics of a triple self-aligned double-gate FET using wafer bonding and layer transfer [74]. (a)–(d) Fabrication sequence illustrated by device cross sections. (e) Cross-sectional TEM image of completed device. (f) Top-view layout of a typical FET. (g) Measured transconductance of n-FET biased in the ground-plane mode using the front gate (FG), back gate (BG) as the active channel, as well as double-gate (DG) mode. By using a bias to adjust for the differences of the front- and back-gate work functions, the curve DG* is obtained. The gate dielectrics are 2.4 nm oxide (front gate), and a composite of 2.4 nm nitride on top of 2.4 nm oxide (back gate) with an oxide electrical equivalent of 4 nm. The silicon channel thickness is about 22 nm. Adapted with permission from Guarini et al. [74]; © 2001 IEEE.

hole mobilities following the universal mobility curves, indicating good transport properties for these bonded silicon channels with thin front- and back-gate dielectrics [74]. DG FET circuits are also demonstrated in this work. Ratioed inverters are demonstrated using a DG FET as the load device. The back gate of the load DG FET is employed to control the load current and adjust the input bias of the load device. By using the front gate and the back gate of the same DG FET as two separate inputs of the "inverter" circuit (again with another DG FET as a load), a compact two-input NOR is demonstrated [74]. The tungsten-plugged source/drain well can be shared in a circuit layout to achieve high device density.

The most successful vertical structure as a high-performance device is the Type III structure, also known as the FinFET [81], since the silicon channel protrudes

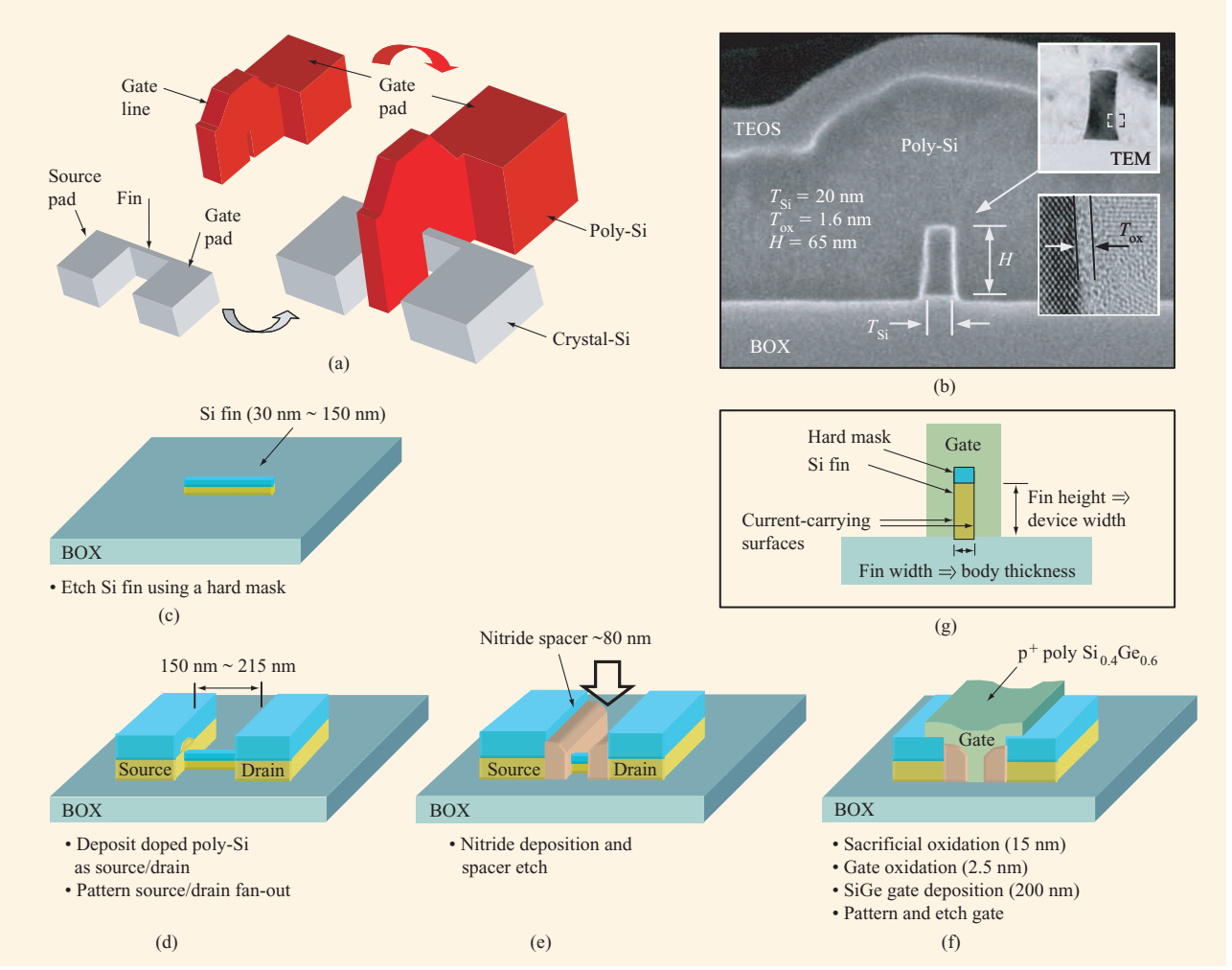


Figure 17

(a) Three-dimensional rendition of a bird's-eye view of the components of a gate-first FinFET double-gate FET. First, the silicon fin and the source/drain pads are etched. Then the gate is deposited and patterned. A gate spacer is formed prior to selective epitaxial growth of the "raised" source/drain (in this case, the epitaxial growth protrudes from the sides of the fin) to provide a source/drain fan-out. (b) Cross-sectional SEM and TEM across the device width, illustrating the fin cross-sectional dimensions and the thin (1.6-nm) gate oxide grown on the sidewall of the fin. Parts (a) and (b) adapted with permission from Kedzierski et al. [85]; © 2001 IEEE. (c)–(f) Fabrication sequence of a gate-last FinFET double-gate FET. (g) Cross section of the silicon fin showing the current-carrying plane. Direction of current flow is into the plane of the diagram. Parts (c)–(g) adapted with permission from Huang et al. [81]; © 1999 IEEE.

from the silicon wafer surface like a fin. **Figure 17** illustrates the fabrication sequence and device structures. The FinFET is essentially a scaled-down version of the DELTA device reported by Hisamoto et al. [82]. Two types of FinFET are reported: 1) a gate-first process in which the source and drain are formed after the formation (patterning) of the gate stack [83–86]; and 2) a gate-last (or replacement-gate) process in which the source and drain are formed before the formation of the gate stack [81, 87].

The fabrication of the FinFET begins with the patterning and etching of a thin fin on an SOI substrate using a hard mask which is retained throughout the fabrication process. The thickness of the fin will be the silicon channel thickness of the DG FET. As discussed earlier, the fin thickness is smaller than the gate length; thus, either electron-beam lithography [81, 84, 86, 87] or optical lithography with extensive linewidth trimming [85] is used to pattern the thin fin. For the gate-first process, the fabrication steps after the fin formation are analogous

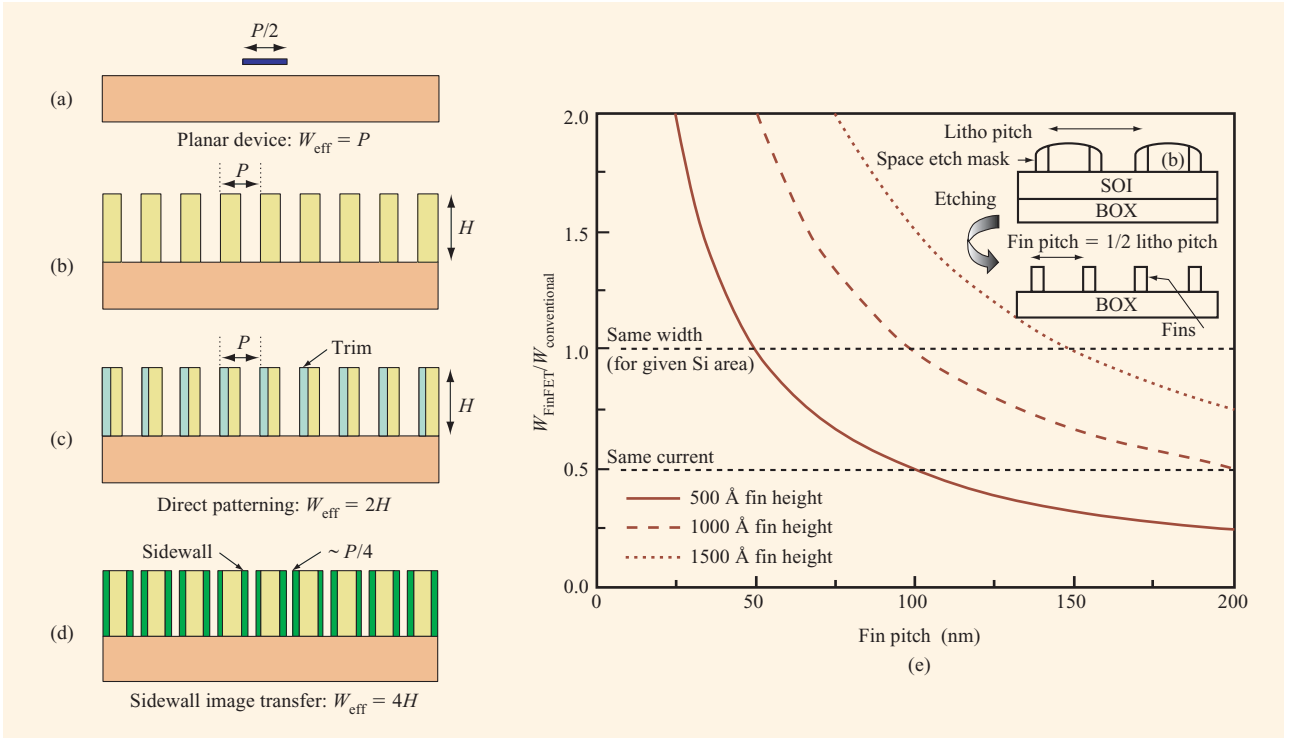


Figure 18

Comparison of the effective device width of DG FET based on device width per unit layout area. (a) Planar DG FET with a device width of P and top layout width of $P/2$. A single-gated FET will have a device width of $P/2$ for the same top layout width of $P/2$. (b) Fins with a lithography pitch of P and fin height H. (c) Fins with a lithography pitch of P, fin height H, and fin thickness reduced by linewidth trimming. The effective device width is $2H$, and the fin pitch is the same as the lithography pitch P. (d) Fins with a lithography pitch of P. Utilizing sidewall image transfer techniques to define fin thickness, the fin pitch is reduced to $P/2$; the effective device width is $4H$. (e) Illustration of the effective device width ratio of FinFET and conventional single-gated FET. To achieve the same (larger) effective device width for the same silicon area, the fin pitch must be the same as (smaller than) the fin height. To achieve the same (larger) drive current, the fin pitch must be the same as (smaller than) half the fin height. Parts (a)–(d) reproduced with permission from [48]; © 1999 IEEE. Part (e) adapted with permission from Tang et al. [54]; © 2001 IEEE.

to the fabrication steps of the conventional bulk FET: After the gate oxide is grown, the gate polysilicon is deposited, patterned, and etched. A sidewall spacer is formed next to the gate. Source/drain and extension implants can be performed before and/or after the gate spacer, using angled implants [85]. A selective epitaxial growth of silicon or germanium from the fin surfaces forms the source/drain fan-out that reduces series resistance [85, 86]. Self-aligned silicide can be formed on the source/drain and gate, as in conventional bulk FETs. For the gate-last process, the source/drain is formed immediately after the fin patterning. Doped polysilicon or polycrystalline SiGe is deposited on the fin, followed by lithographic patterning of the source/drain fan-out pads with a thin slot between the source and drain. This distance between the source and drain determines the gate length. The slot length is further reduced by a dielectric sidewall spacer. Then the gate oxide is grown, and the gate polysilicon (or polySiGe) is deposited and patterned. A potential device density improvement for FinFET using direct etch and sidewall image transfer to generate fins is illustrated in **Figure 18**. The sidewall image transfer technique can reduce the fin pitch by a factor of 2 using the same lithography pitch. The fin pitch must be smaller than the fin height to provide more effective device width than a planar single-gated FET.

The difficulties of fabricating the FinFET DG FET include 1) variability of the fin thickness (and hence silicon channel thickness); 2) highly selective RIE and long over-etches which are required to pattern the gate polysilicon that wraps around the fin (for the gate-first process), and to create the sidewall spacers (for both the gate-first and the gate-last processes); and 3) a difficult integration scheme for the source/drain of complementary

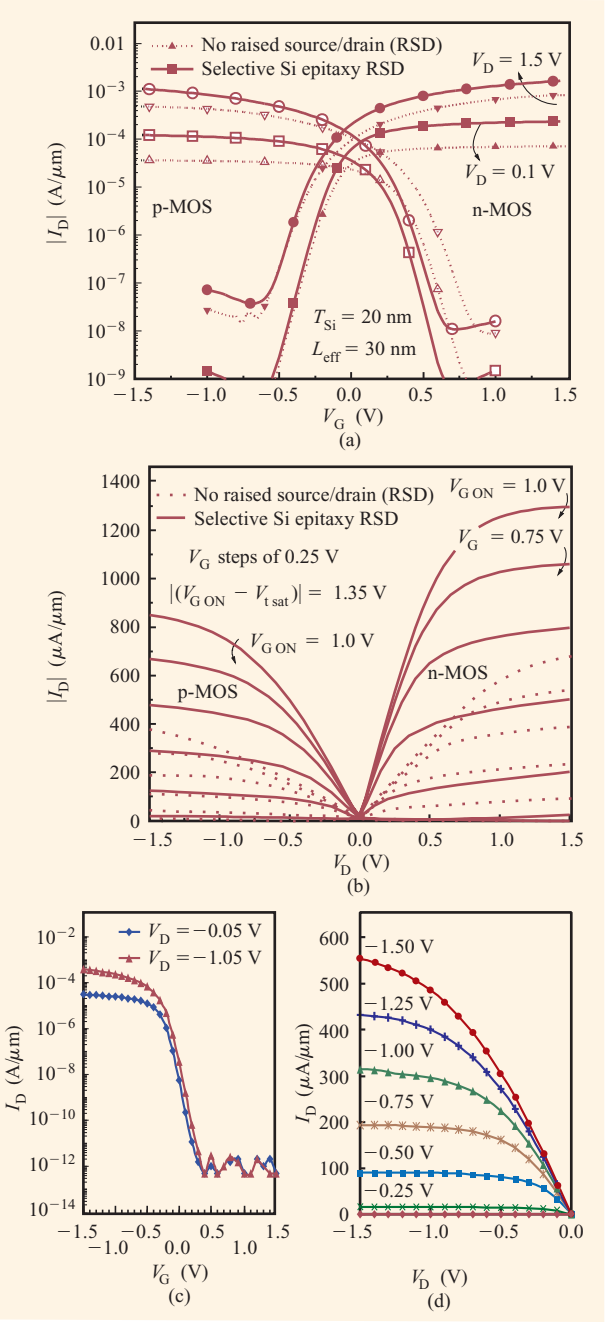


Figure 19

Current–voltage characteristics of FinFET DG FET reported in the literature. (a) Subthreshold and (b) output characteristics of a gate-first FinFET DG FET compared with and without raised source/drain process. The raised source/drain process significantly increases the current drive by reducing the series resistance. Parts (a) and (b) adapted with permission from Kedzierski et al. [85]; © 2001 IEEE. (c) Subthreshold and (d) output characteristics of a gate-last p-FinFET DG FET with 18-nm gate length. Parts (c) and (d) adapted with permission from Huang et al. [81]; © 1999 IEEE. The currents reported are normalized to two times the fin height compared to the device width.

FETs (n-FET and p-FET on the same wafer) for the gate-last process. The apparent difficulty in handling the topography of a "vertical" structure is manifested in the RIE and over-etch issues mentioned above. While tall fins may provide an improvement in density, the fin height must be contained within reasonable limits. The variability of the fin thickness deserves special attention. The variability in fin thickness arises from several sources: linewidth variation of the lithographic process, linewidth variation of the fin etching process, and line-edge roughness of the lithographic process [88], all of which are translated to the final fin dimensions. There is no systematic data yet that correlates fin thickness to processing conditions. The sidewall image transfer technique produces a set of fin surfaces that have correlated roughness; this is a different situation from the direct-etched-fin case, in which the fin surface undulations on either side of the fin are uncorrelated.

Both the gate-first and the gate-last FinFET processes have produced devices with excellent *I–V* characteristics (**Figure 19**). The current drive and the *CV*/*I* of the FinFET rival those of the best conventional bulk devices [13, 14], albeit with a threshold voltage that has to be centered. Despite the unconventional device structure and topology, the minimum gate length achieved is among the shortest [81, 85, 86], including those of conventional FETs [13, 14]. Since the silicon channel thickness is determined by the patterned fin, experimental data comparing the short-channel effect of different silicon thicknesses is readily obtained. The experimental results of Huang et al. [81] and Kedzierski et al. [85] both corroborate the short-channel electrostatic device design and scale length analyses described above. While the carrier-transport properties along the etched fin surfaces have been a concern, the good current drive exhibited by the FinFET DG FET indicates that there is no significant degradation.

Challenges ahead

While there has been tremendous progress both in understanding the device physics of the DG FET and in fabricating it, significant challenges remain to be met before this new device structure is ready for manufacturing. Some of these open issues have no obvious solution at the moment.

First and foremost is the need to set the threshold voltage and to achieve multiple threshold voltages on the same chip. The section on double-gate FET electrostatics outlined the requirements on the work function of the gates to achieve the required threshold voltages for SDG and ADG. At this moment, no material satisfies the requirements for SDG except possibly a midgap-work-function material such as tungsten or $CoSi_2$, with doping of the silicon channel to lower the threshold voltage to the desired level. For ADG, the combination of n^+ (p^+)

polysilicon with tungsten or $CoSi_2$ (materials having a mid-gap work function) for the n-FET (p-FET) gates provides one possible solution, although an integration scheme to achieve self-aligned gates is still lacking. Adjusting the threshold voltage for multiple threshold voltages on the same chip is a different challenge from setting the threshold voltage. For ULSI systems, it is typically necessary to provide a menu of devices with different threshold voltages to allow for the optimization of performance and power consumption. The ability to tune the threshold voltage by about 150 mV is often required. The obvious solution is to dope the channel of the DG FET in analogy with bulk FETs. For devices with geometries in the sub-50-nm gate-length regime, a channel doping of the order of high 10^{18} cm^{-3} is required. This tends to reduce the benefit of the DG FET, since the advantages of an undoped channel (higher mobility, reduced discrete dopant fluctuation effects) will diminish or even vanish. Recent work on gate-material work-function tuning [36] by ion implantation offers interesting opportunities, although much still has to be proven.

Because of the thin silicon channel, the series resistance of the DG FET is of particular concern [50]. Some form of raised source/drain process would be required in order to achieve a source/drain fan-out. The results given in [85] illustrated the efficacy of the solution. Growing selective epitaxial silicon on thin (<15-nm) silicon substrates (or fins) is still difficult, because the thin starting silicon tends to break up during or prior to the epitaxial growth. In addition, as the silicon channel thickness is reduced to less than 10 nm, optimization of the series resistance and parasitic capacitance of the "overlap" region may prove difficult.

Uniformity of the silicon channel is an important concern, especially when the silicon channel is reduced to less than 10 nm to satisfy short-channel control requirements. Silicon channel thickness tolerance may be translated to an equivalent gate-length tolerance by considering a constant scale length [10, 46]. A 15% variation in silicon channel thickness approximately translates into a 10% gate-length variation. This tolerance is added quadratically to the overall tolerances. Maintaining a thin, uniform silicon channel thickness remains a major manufacturing obstacle.

Finally, most double-gate FET devices are structurally different from conventional FETs. Parasitic capacitance (gate-to-source/drain capacitance) critically depends on the device structure. There is no fundamental barrier to achieving a back-gate parasitic capacitance that is as small as the front-gate parasitic capacitance, especially for the FinFET DG FET. While most device results to date have focused on the current drive and static *I–V* characteristics, more work has to be done to characterize and optimize the parasitic capacitance.

Higher mobility

The section on double-gate FET carrier transport has already pointed out the importance of a low-doped channel for carrier transport in DG FET. Another avenue to achieving a higher carrier mobility and saturation velocity is through the choice of material for the FET channel. Fischetti and Laux [89] compared the performance of several semiconductors that have high carrier mobilities and saturation velocities, including Ge, InP, InGaAs, GaAs, and several others. They concluded that materials which provide a significantly higher carrier mobility give only a moderate performance advantage over a lower-mobility material such as silicon. That work succinctly pointed out that under non-equilibrium, high lateral field transport, carrier mobility is not the only determinant of performance (using metrics such as transconductance). The band structure, which determines the density of states (i.e., the inversion capacitance, C_{inv}) [90] and the carrier scattering rates at high carrier energies are just as important as the carrier mobility: a low-carrier-energy, quasi-equilibrium quantity. Furthermore, in the limit in which carriers are transported ballistically, many other factors combine to determine performance [91]. On the other hand, a higher carrier mobility does provide some moderate performance gains, albeit not in proportion to the mobility values [92].

Strained-silicon FET—device concepts

It has been known for some time that carrier mobility in silicon under biaxial tensile strain is enhanced [93–98]. The theory of mobility enhancement for strained silicon is still evolving [96]. The most commonly cited reason for electron mobility enhancement in strained silicon is that under the biaxial tensile strain, the sixfold degeneracy of the conduction band of silicon is lifted, raising the higher-effective-mass fourfold-degenerate ellipsoids and lowering the lower-effective-mass twofold-degenerate ellipsoids. This has the effect of keeping most of the carriers in the lower-energy, lower-effective-mass valleys and reducing the intervalley scattering. In the valence band, the biaxial tensile strain lifts the heavy-hole/light-hole degeneracy at the Γ point, resulting in a smaller in-plane transport effective mass due to band deformation and reduced intervalley scattering.

The use of strained silicon provides a plausible tradeoff between moderate levels of performance enhancement over silicon and ease of fabrication and integration with silicon (as compared to other higher-mobility materials such as Ge, InGaAs, GaAs, and InP). Recent work has provided encouraging experimental evidence that introducing the biaxial tensile strained silicon through a layer of relaxed SiGe may provide adequate performance gains for incorporation into conventional CMOS

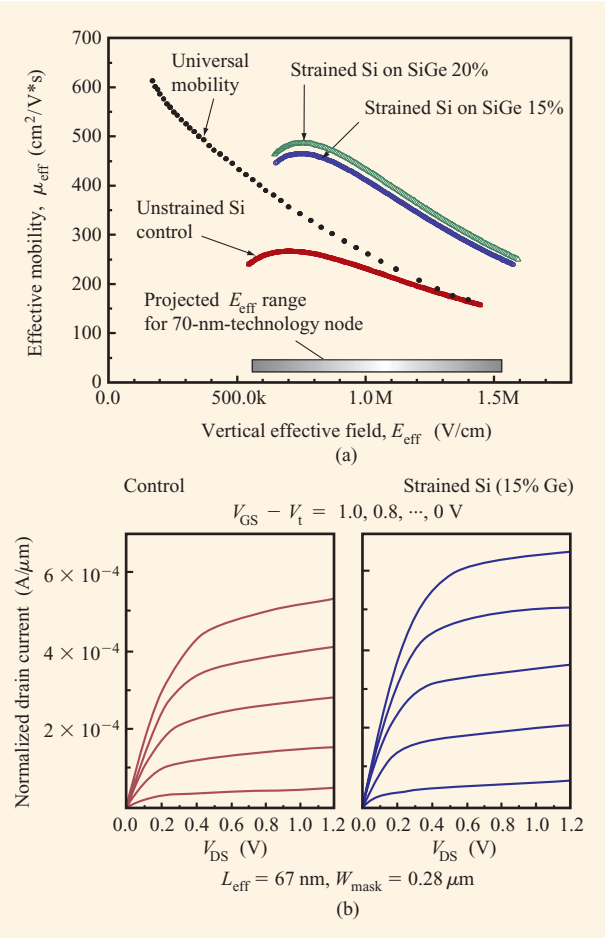


Figure 20

(a) Effective electron mobility vs. effective electric field for a strained silicon on relaxed SiGe FET fabricated using a conventional CMOS process flow. (b) *I–V* characteristics of short-channel FET comparing the current drive of strained-silicon FET with control unstrained-silicon FET. Part (a) reproduced with permission and part (b) adapted with permission from Rim et al. [104]; © 2001 IEEE.

technologies. This section summarizes the recent results, focusing on strained silicon on relaxed SiGe.

Materials and device fabrication

The tensile strain can be introduced in several ways. Local strain caused by the thermal mismatch of silicon and the isolation materials [such as shallow-trench isolation (STI)] has been shown to introduce enough strain to alter device characteristics [97]. A more general approach is to introduce biaxial tensile strain by growing a thin layer of epitaxial silicon on a material with a slightly larger lattice constant, such as relaxed SiGe [95, 98, 99]. The strained silicon must be relatively thin in order to prevent relaxation and strain relief through dislocations. The underlying relaxed SiGe serves as an anchor to constrain the lattice of the strained silicon on top and has no beneficial electrical role otherwise to first order. Compared to a closely related approach of using a pseudomorphic layer of SiGe on silicon, where the carrier transport is in the compressively stressed SiGe layer (see for example [100]), strained silicon on relaxed SiGe has the advantage of having silicon as the top surface. The silicon provides an easy means of forming a good gate dielectric and enables surface channel operation for good short-channel-effect control.

Experimental results for bulk strained silicon on relaxed SiGe FETs and strained silicon on relaxed SiGe on insulator FETs are quite encouraging. For bulk strained silicon on relaxed SiGe, the fabrication begins with the growth of a relaxed SiGe buffer of 1–2 μm by a step-graded approach [101–103]. This step can be performed using rapid thermal chemical vapor deposition (RTCVD) [95, 98] or ultrahigh-vacuum chemical vapor deposition (UHVCVD) [104]. A thin (typically <20 nm) silicon layer is then epitaxially grown on the relaxed SiGe. Because of the larger lattice constant of $Si_{1-x}Ge_x$ ($4 \times x\%$ larger than that of silicon), the top silicon layer is under biaxial tensile strain. Early work on FET fabrication has limited high-temperature processes and ion implantation in the channel to avoid strain relaxation. Recent results indicate that a conventional CMOS process flow can be adopted while still achieving the mobility and current drive enhancement [104]. Process modules such as shallow-trench isolation (STI), channel ion implantation, source/drain extensions and halos, and associated high-temperature activation anneals are shown to have no adverse impact on device characteristics. **Figure 20**, part (a) shows that the effective electron-mobility enhancement over the unstrained-silicon control, as well as the enhancement over the universal mobility curve, persists into an effective electric field (E_{eff}) range (up to 1.5 MV/cm) that is relevant to short-channel devices with high channel/halo dopings (e.g., bulk or PDSOI). A portion of this mobility enhancement is translated into current drive enhancement over the unstrained-silicon control [Figure 20, part (b)]. When the effect of self-heating is accounted for in the measurement, it is expected that the current drive will improve by another 5–10%.

It is also desirable to combine the benefits of enhanced transport in strained-silicon materials with the benefits of silicon-on-insulator (SOI) technologies [e.g., reduced junction capacitances, dynamic floating-body effects (PD SOI), scalability to shorter channel length (ultrathin-body SOI)]. Several approaches have been reported, all of which rely on growing strained silicon on a relaxed SiGe layer on insulator (SGOI). These methods include wafer

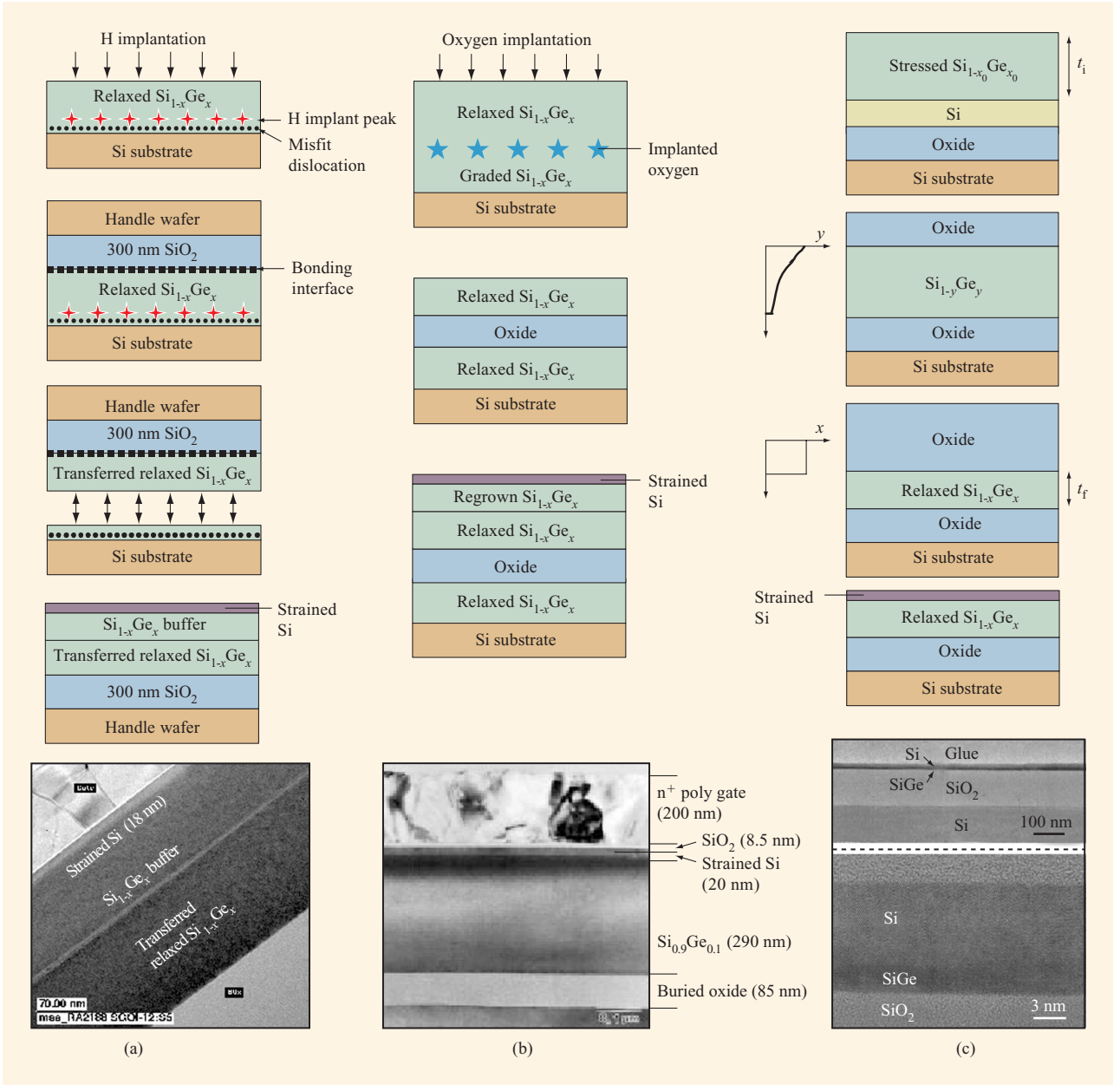


Figure 21

Illustrations of the fabrication sequences of strained-silicon-on-insulator materials: (a) Wafer bonding and layer transfer; (b) SIMOX; (c) oxidation enrichment of SiGe. Part (a) adapted with permission from Huang et al. [106]; © 2001 IEEE. Part (b) adapted with permission from Mizuno et al. [99]; © 2000 IEEE. Part (c) adapted with permission from Tezuka et al. [109].

bonding and layer transfer [105–107], oxygen implantation (SIMOX) [99, 108], and oxidation enrichment of SiGe [109, 110] (see **Figure 21**). For the wafer-bonding and SIMOX approaches, mobility enhancements similar to that obtained in bulk strained-silicon FET have been demonstrated for n-FETs [99, 106, 107] as well as p-FETs [99, 106, 108]. Device design for SOI devices in the sub-50-nm regime calls for a thin SOI layer of less than 50 nm. Therefore, the combined layer thickness of the strained silicon and the relaxed SiGe should be less than 50 nm. In addition, defect density must be low enough for ULSI applications.

The wafer-bonding and layer-transfer technique is the most straightforward and preserves the qualities of the starting relaxed SiGe layer. These include desirable qualities such as a high Ge content and high degree of relaxation of the SiGe buffer layer (e.g., by the step-graded approach). It also retains undesirable qualities such as the defect density of the starting SiGe buffer. A practical limitation of the layer transfer approach is thickness and uniformity control, since the thickness of the SiGe layer on the insulator is dependent on the final polish step. Using the SIMOX approach, the Ge content in the SiGe layer appears to be limited to less than 10% because the melting point of $Si_{1-x}Ge_x$ with $x > 0.1$ is lower than the SIMOX annealing temperature. This limitation can be partially recovered by oxidation enrichment. When the SiGe is oxidized, the Ge is driven from the oxide layer, and the Ge content in the underlying SiGe layer increases (the "snowplow effect"). Thus, the Ge content of the SiGe layer can be "enriched" by this oxidation process. The enrichment process is used in the SIMOX approach to obtain a Ge content greater than 10% in the SGOI film [99]. This process can also be employed to produce a relaxed SiGe layer on insulator (oxide) by oxidizing a film stack consisting of a starting silicon-on-insulator layer with a SiGe layer pseudomorphically grown on top [109]. In this case, thickness uniformity and control will be very similar to that of the starting SOI film.

Electrical results of strained-silicon-on-insulator (SSOI) FETs fabricated from strained silicon grown on SGOI virtual substrates are quite encouraging. Essentially, the mobility enhancements commensurate with the amount of strain introduced as similarly observed in bulk strained-silicon substrates are reproduced in SSOI FETs. These results verify that the concept of combining strained silicon with SOI will retain the benefits of strained silicon.

Challenges ahead

Strained-silicon CMOS is at present one of the most promising technology options for insertion into CMOS technologies. The materials set is friendly to conventional CMOS technologies, and the device performance gains demonstrated so far have been promising. The main issue for the materials development is defect control and understanding. The defect density of the materials must be comparable to those of exiting bulk and SOI materials for the relaxed SiGe materials to be viable in an ULSI application. The nature of the defect is as important as the defect density. Correlating device fail mechanisms to the physical defect will be an important area of work. In particular, for SSOI materials, the nature of the defects can be quite different from that of the bulk relaxed SiGe material. Obviously, there is also a strong dependence of defects on the method of SSOI materials fabrication. It may also be possible to obtain strained-silicon-on-insulator without the underlying SiGe layer (e.g., using compliant substrate concepts [111]), thereby breaking out of the SiGe materials system altogether. In the device area, the fundamental principles of mobility enhancement must be further elucidated. In addition, relating the mobility enhancement to current drive gains for short-channel devices will continue to add to the understanding of the relative role of mobility in describing carrier transport, since strained silicon provides another means of modifying mobility while keeping other device parameters the same. Finally, we note that while it may appear difficult to combine strained silicon with the double-gate FET device structure, in principle nothing precludes this possibility.

3. Nanotechnology

Nanotechnology is a broad term which may refer to such diverse technical disciplines as chemistry, biology, physics, materials, and electrical engineering. Here we focus on the aspect of nanotechnology that pertains to extending and broadening the impact of the microelectronics and semiconductor industry. Therefore, the discussion here is necessarily narrowly focused. Instead of including a comprehensive discussion of this narrowly focused yet still vast subject, we choose to select one example as an illustration of the approach and thought process we adopt to identify and/or adopt nanotechnology for commercial applications. In this example, we examine a potential device technology: the carbon nanotube field-effect transistor (CNFET). This example is chosen because it appears to be one of the nanotechnologies which allows a meaningful comparison with incumbent technologies in that it uses the same circuit and system architecture. It is also one of the most widely studied in the nanoscale science community.

Carbon nanotube FET

Device concepts

Carbon nanotubes are nanoscale high-aspect-ratio cylinders of carbon atoms with exceptional electrical and mechanical properties. The basic properties and preparation of carbon nanotubes are well documented in the literature [112–116]. Many applications for this material have been proposed: electronic switch [117, 118], field emitter for flat-panel displays [119, 120], electrochemical energy storage for batteries and fuel cells [121], interconnect for electronics, probe tips for scanning probe microscopy and lithography [122–124]. Here we focus our discussion on the use of a carbon nanotube as an electronic switching element which performs basic logic functions.

Single-wall carbon nanotubes (SWNT) are two-dimensional graphene sheets rolled into nanometer-

diameter cylinders that can either be 1D metals or semiconductors [125]. With the appropriate chirality [114, 116], SWNT can be semiconducting, with a bandgap inversely proportional to the diameter of the tube: $E_{gap} = 2\gamma_0 a_{C\text{-}C}/d$, where E_{gap} is the bandgap, γ_0 is the carbon-to-carbon tight-binding overlap energy, $a_{C\text{-}C}$ is the nearest-neighbor carbon-to-carbon distance (0.142 nm), and d is the diameter of the nanotube [126]. A typical semiconducting SWNT has a diameter of 1.4 nm and a bandgap of about 0.5–0.65 eV [126]. A straightforward application of this semiconducting property of the carbon nanotube is to form a field-effect transistor (FET) analogous to the MOSFET. Tans et al. [117] and Martel et al. [118] have already demonstrated FET operation using gold electrodes as source and drain contacts and a back-gated structure. Both p- and n-channel carbon nanotube transistors (CNFETs), as well as CNFETs with ambipolar behavior, have been reported [127, 128]. Recently, simple circuits such as inverters [127, 129] and ring oscillators [129] have been successfully fabricated.

Device fabrication

The fabrication of an early CNFET is straightforward: First, the gold electrodes are formed on a silicon wafer with a thermally grown oxide (the gate insulator); then carbon nanotubes are dispersed on the wafer. The wafer substrate serves as the gate electrode of the CNFET. While this process is simple and suitable for a first demonstration of FET action, the source/drain contact resistance clearly limited the current drive of the CNFET [117, 118]. A recent embodiment of the CNFET is shown in the inset of **Figure 22(a)**, where ideas borrowed from conventional microelectronics fabrication are employed to improve the contact resistance. In this process, carbon nanotubes are dispersed onto a heavily doped silicon wafer with a thermally grown oxide (serving as the gate insulator). Then a transition metal (titanium or cobalt) is patterned on the wafer (by lift-off) as the source/drain contact metal. A subsequent anneal at 400°C (Co) or at 820°C (Ti) in an inert ambient forms a low-resistivity cobalt or Ti carbide at the source/drain contact [128, 130]. Figure 22 shows the I–V characteristics of a p-type CNFET with a cobalt source/drain contact [131]. The cobalt contact improves the contact resistance noticeably compared to previous results obtained with the same SWNT material [117, 118] as evidenced from **Figure 22(b)**, where the drain current shows no sign of saturation at large gate voltages. A similar device with a SiO_2 passivation layer reduces the subthreshold slope [131], but does not completely remove the interface traps at the nanotube surface. While this device structure is easy to fabricate and enables advance learning, a device configuration in which the gates are separately controllable (in contrast to a common gate provided by the substrate) is necessary for a technology.

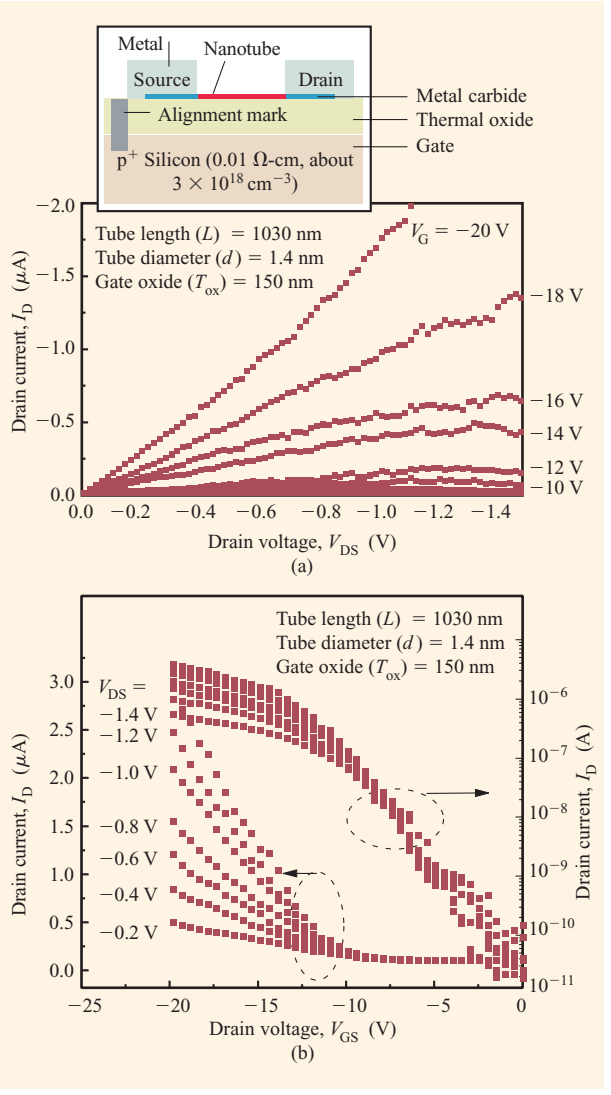


Figure 22

Room-temperature current–voltage characteristics of a carbon nanotube FET (CNFET) with a cobalt source/drain contact. (a) Drain current vs. drain voltage. The inset shows the schematic cross section of a carbon nanotube FET with a metal carbide source and drain contact and a back-gated structure. (b) Drain current vs. gate voltage. Carbon nanotube length = 1030 nm, tube diameter = 1.4 nm, gate-oxide thickness = 150 nm.

Performance assessment

Table 3 summarizes the key device parameters extracted from the CNFET I–V characteristics. The source/drain series resistance is an upper bound estimated from the

Table 3 Selected device parameters of p-type carbon nanotube FETs. The source/drain series resistance is an upper bound estimated from the measured total resistance from source to drain.

Contact scheme	*Cobalt*	*TiC*
Channel length (nm)	1030	800
Nanotube diameter	1.4	1.4
Gate-oxide thickness (nm)	150	150
Source/drain resistance (kΩ)	<25	*
Inverse subthreshold slope (mV/decade)	2500	730
Interface trap capacitance ratio,** C_{it}/C_i	40.9	11.2
Transconductance# (μS)	0.342 (V_D = 1.4 V)	0.280 (V_D = 1.0 V)
"Mobility"# ($cm^2V^{-1}s^{-1}$)	60	68

*This value cannot be accurately estimated for this device.
**Ratio of the capacitance due to interface trap relative to the gate capacitance as estimated from the subthreshold slope assuming that the capacitance due to band bending in the s-SWNT is negligible.
#The transconductance and the mobility values are given without a correction for the contact resistance. After Martel et al. [131].

Table 4 Comparison of device characteristics of carbon nanotube FET and conventional silicon MOSFET.

	p-CNFET[a] *1030 nm*	*100-nm MOSFET*	*25-nm MOSFET*
Transconductance (μS/μm)	122	1000 (n-FET)[b] 460 (p-FET)[b]	1200 (n-FET)[c] 640 (p-FET)[c]
External resistance (Ω-cm per side)	<70	~66 (n-FET)[d] ~143 (p-FET)[d]	~40 (n-FET)[d] ~86 (p-FET)[d]
Gate insulator (nm)	150	2.0	0.8

(a) Device A from Martel et al. [131].
(b) Ghani et al., IEDM 1999 [136].
(c) Chau et al., IEDM 2000 [13]; these values approach the 1500 mS/mm predicted for n-FET for an ideal single-gated bulk MOSFET [47].
(d) ITRS 1999 [1].

measured total resistance from source to drain.[5] In order to assess whether the CNFET is a viable technology, it is useful to compare the device performance of CNFET with that of conventional silicon MOSFETs. **Table 4** compares several key device parameters. It should be noted that several assumptions must be made in order to make a direct comparison. While some of these assumptions may not be valid for various practical or fundamental reasons, Table 4 does serve as a starting point for discussion.

First, it is necessary to make some assumptions about how the CNFET will be used. Here, we assume that a circuit topology similar to today's silicon-based VLSI circuits will be used; that is, the CNFETs are switches that form logic gates interconnected to one another by wires. In this scenario, the CNFET must deliver a current-driving capability similar to or surpassing that of the MOSFET in the future time when the CNFET will be introduced.[6] To provide this large current, it is likely that an array of CNFETs will be needed. If we further assume that the CNFET can be made with equal lines and spaces with dimensions equal to the diameter of the nanotube (1.4 nm), the CNFET characteristics may be compared with those of MOSFETs on a per-unit device-width basis, as in Table 4.

Understanding the carrier transport mechanism of the CNFET is vital to assessing its potential as a device technology. Here, we adopt a phenomenological approach based on a classical description of transport that allows a direct comparison with silicon devices. Although it is clear that the characteristic length for electron–phonon and

[5] It is difficult to extract the series resistance from one device. In addition, the transport physics of the CNFET is not known, which makes any model-based extraction impossible.

[6] This requirement may be relaxed if 1) wiring capacitance per unit length can be reduced drastically, *and* 2) a new circuit topology is adopted in which devices need be connected only locally.

defect scattering in metallic single-wall carbon nanotubes is very long (of the order of microns) and favors ballistic transport [132, 133], this classical approach takes into account the fact that the semiconducting SWNTs show stronger backscattering at room temperature [118, 125, 134].

In this phenomenological model, the channel conductance is given by $G = \mu Q_L/L$, where μ is the carrier "mobility" and Q_L is the charge density per unit length. Note that μ is a phenomenological parameter for these quasi-1D systems [118]. If one draws an analogy with the silicon MOSFET in which the channel charge is related to a threshold voltage V_T, then $Q_L = C_L(V_G - V_T)$, where C_L is the gate-to-channel capacitance per unit length and V_G is the gate voltage [118]. Within this framework, the hole "mobility" of the CNFET is about 60–68 $cm^{-2}V^{-1}s^{-1}$. This hole "mobility" compares very well with surface mobility of holes in MOSFETs, which range from about 160 $cm^{-2}V^{-1}s^{-1}$ at low vertical electric fields (0.1 MV/cm) to about 50 $cm^{-2}V^{-1}s^{-1}$ at high vertical electric fields (1 MV/cm) [60]. The transconductance of the p-type CNFET shown in Figure 22 reaches 122 $\mu S/\mu m$ (assuming equal lines and spaces for a CNFET array) for a 1030-nm-long channel and a gate insulator thickness of 150 nm.[7] Continuing with the classical transport picture in which the CNFET current scales inversely with the channel length, the CNFET is extrapolated to have a transconductance of 1257 $\mu S/\mu m$ at a 100-nm channel length and 5028 $\mu S/\mu m$ at a 25-nm channel length. For silicon MOSFETs, respective transconductances of 1000 $\mu S/\mu m$ and 460 $\mu S/\mu m$ for n-FETs and p-FETs have been achieved at 100-nm gate length [136]. At 25-nm gate length, the transconductances reach 1200 $\mu S/\mu m$ for n-FETs and 640 $\mu S/\mu m$ for p-FETs [13], which approaches the 1500 $\mu S/\mu m$ predicted for n-FETs for an ideal single-gated bulk MOSFET [47]. The transconductance of n-channel double-gate FETs in the 25-nm regime is projected to be of the order of 5000 $\mu S/\mu m$.

The electrostatics of the CNFET is another area that requires detailed analysis, primarily due to the cylindrical geometry, the small size, and the 1D quantized channel (and band structure) of the carbon nanotube. Short-channel effects of the CNFET have not yet been explored. The capacitance for an isolated cylindrical structure (either in a planar gate conductor or coaxial gate conductor configuration [131]) varies inversely and logarithmically with the distance from the center of the conductor [137]. For gate insulator thicknesses that are large compared to the nanotube diameter, the gate capacitance for an isolated CNFET is greater than that for a corresponding parallel-plate capacitor (e.g., in a planar silicon MOSFET) with an equivalent width [131] (due to the fringing field). For example, a 150-nm gate SiO_2 for an isolated CNFET is roughly equivalent to a 3–10-nm gate SiO_2 for a planar MOSFET. However, this advantage due to the fringing field vanishes for gate insulator thicknesses that are comparable to or smaller than the nanotube diameter. Furthermore, for an array of CNFETs, the fringing field diminishes as the packing density of the array increases. At high packing densities, the shielding from neighboring nanotubes renders the electric field pattern similar to the parallel-plate situation. It is therefore important to explore in future work the use of high-*k* gate dielectric materials and to study the electrostatics of alternative gate and channel geometries (such as the coaxial geometry) to maximize the gate-to-channel charge coupling.

In principle, the carbon nanotube has a perfect interface with the gate dielectric because all bonds of the carbon atoms are satisfied in a carbon nanotube. In other words, there are no dangling bonds which form interface states. However, the subthreshold turn-off characteristics shown in Figure 22(b) are very gradual (hundreds of mV/decade). This behavior is probably caused by traps in the gate dielectrics that are close to the interface. Reducing the impurities in the gate dielectric will probably lead to improved subthreshold characteristics (see recent results by Wind et al. [135]).

Challenges ahead

Despite the optimism suggested by the above analysis, the challenges facing CNFET technology are daunting. The first and most important is to understand the device scaling and carrier transport properties of the CNFET (as a function of the gate length, gate dielectric thickness), as well as key device physics such as short-channel effects; gate-dielectric-to-carbon-nanotube interface properties (interface states, charge-trapping instability), CNFET contacts, gate electrode configurations, and their electrostatics. Several papers suggest that carrier transport of the CNFET is dominated by a Schottky tunneling barrier between the source/drain and the channel [128, 138, 139]. This behavior is quite different from that of the silicon MOSFET. These open issues determine the device performance potential of CNFET technology.

The synthesis or growth of carbon nanotubes in predetermined locations is an essential requirement. CNFETs can either be a small fraction of devices in a silicon VLSI chip that provide the highest performance, or they may constitute the entire logic processing system. In either case, one must connect a fairly large number (of the order of millions in the former case and more than 100 million in the latter case) of carbon nanotubes to appropriate contacts. One can imagine a method in which connections are made to nanotubes in predefined locations

[7] Recent experimental results [135] attain a transconductance that is ten times the value reported here, which makes the comparison with silicon MOSFET even more favorable for the CNFET.

(either by *in situ* growth/synthesis or by assembling prefabricated nanotubes at predefined locations), or a method in which nanotubes are synthesized or assembled to connect a preexisting array of interconnections. The growth of carbon nanotubes at predefined locations has been demonstrated to some extent using chemical vapor deposition catalyzed by transition metals such as nickel, cobalt, and iron [140–145], or cobalt silicides [146]. In one example, the growth is catalyzed from nanometer-sized catalyst particles where one nanotube per particle is grown [142]. Using self-assembly techniques, it is possible to assemble a regular array of metal particles on specific sites, thereby providing a means of synthesizing an array of carbon nanotubes. Potential compatibility with silicon devices on-chip is suggested by the work of Mao et al. [146], which showed nanotube growth on $CoSi_2$, and the work of Zhang et al. [130], which showed the formation of metal carbide as well as silicon carbide by thermal solid–solid reaction for forming electrical contacts between a carbon nanotube and a metal or directly with silicon. In addition, Hu et al. [147] demonstrated methods to make a transition from a carbon nanotube to a silicon nanorod using iron-catalyzed growth in ethylene. These results suggest the possibility of integrating conventional semiconductor (silicon) structures with carbon nanotube structures.

A further challenge for CNFET technology in addition to the nanotube synthesis discussed above is the control of the carbon nanotube size, type, and chirality, which uniquely determine its electrical properties. The catalyzed growth described above generally produces bundles of single-wall or multiwall nanotubes with varying size and chirality in a somewhat uncontrolled manner. It has already been shown that carbon nanotubes grown from pre-formed substrates such as zeolites have a tube diameter determined principally by the size of the template (the zeolite) [148] even for an extremely small template diameter (as small as 0.4 nm). It is also suggested that by controlling the size (for example, using self-assembly or similar methods) and surface preparation of the precursor catalyst nanoparticles or nanotemplates, one can control the type of carbon nanotubes grown [149] and thereby control the electrical properties.

Finally, developing a fabrication process and device structure that minimize parasitic resistances and capacitances is important. Self-aligned/self-assembled processes and device structures are useful. Device density improvements can be realized only if the total device area (including the contacts and interconnection regions) is scalable and commensurate with the gate length.

Nanodevices and nanofabrication

Many devices and techniques have been proposed as "future" technologies which may supplant CMOS as the dominant device technology. **Table 5** lists examples of "future" devices proposed in the literature, with a summary of their attributes.[8] While some proposed devices do possess advantages in isolated aspects, none possesses the full suite of capabilities and versatility to replace CMOS. In large-scale systems that consist of more than ten million devices, it is important for the devices (or an aggregate of devices at a functional level) to possess the basic attributes that enable large-scale systems: 1) The output of one device must drive the input of the device in the next stage; 2) the device must restore the degraded incoming signal to a stronger outgoing signal; 3) the device must be immune from the influence of external noise; 4) multiple levels of interconnects must be possible to enable communication of results among a large number of devices; and 5) the device must be manufacturable in large-scale production with a good yield.

The primary challenges facing nanodevice fabrication are 1) making contacts to devices on a nanometer scale, 2) interconnecting the nanodevices massively, and 3) providing a means to input and read out data. These are indeed the challenges facing some of the more widely studied devices such as organic molecular devices, quantum dots, carbon nanotubes, DNA, and various quantum computing implementations. A "top-down" approach such as conventional photolithography is suitable for large systems with diverse length scales. However, the throughput of such methods is limited, since the feature size becomes small and the number of features required increases. A "bottom-up" approach such as self-assembly provides atomic-level control of the feature size. However, it is somewhat unrealistic to expect the self-assembly to possess long-range order and to be able to self-assemble a complete system.[9] A plausible solution is to apply "top-down" techniques to aid in defining areas where "bottom-up" methods will exert atomic-level dimensional control, thereby reducing the accuracy and information required to specify the fabrication process.[10] In this case, the nanofabrication problem is reduced to a problem of integrating the processes together to bridge the different length scales.

4. Discussion

As discussed in the paper by Frank et al. [10], device scaling limits are application-dependent, and the "end of scaling" is brought on by power-consumption constraints.

[8] This list is obviously far from exhaustive.
[9] One might argue that the self-assembling processes exhibited by nature do just that—assemble complete systems. However, this evolution occurs over a period of millions of years, a time horizon quite different from that of this present discussion.
[10] In fact, microelectronics fabrication processes in use today already practice the approach advocated here. For example, using only one photolithographic step to define the gate stack, the entire FET channel and the source/drain areas are fabricated using self-aligned processes to obtain complex double/triple sidewalls, halo and source/drain implants, and silicided contacts.

Table 5 Examples of proposed "future" devices and their salient attributes.

Device	*Possible applications*	*Advantages*	*Disadvantages*	*Remarks*
Single-electron transistors (SET)	Logic element	1. Small size 2. Low power	1. Sensitive to background charge instability. 2. High resistance and low drive current. 3. Cannot drive large capacitive (wiring) loads. 4. Requires geometries $\ll$10 nm for room-temperature operation.	Use of Coulomb blockade in nanocrystal "floating-gate"-type nonvolatile memory demonstrated. May improve retention time.
Quantum dot (quantum cellular automata)	Logic element	Small size	1. Multiple levels of interconnection across long distance difficult. 2. Room-temperature operation difficult. 3. New computation algorithms required. 4. Method of setting the initial state of the system not available. 5. Single defect in line of dots will stop propagation.	Devices demonstrated at low temperatures. QCA architectures extensively investigated.
Resonant tunneling diode (RTD)	1. Logic element 2. Dynamic memory	1. Small size	1. Tunneling process sensitive to small film thickness (tunneling distance) variation, leading to process control difficulties. 2. Requires dc bias, large standby power consumption. 3. Multivalue logic sensitive to noise margin 4. Speed of RTD circuits likely to be determined by the conventional devices required in the circuit.	Small- to medium-scale circuits demonstrated. Most demonstrations on III–V compound semiconductors.
Rapid single-flux quantum (RSFQ) device	Logic element	Very high speed possible	1. Requires liquid helium temperature. 2. Lacks a high-density random-access memory. 3. Requires tight process tolerance.	Very-high-speed (THz) circuits demonstrated.
Two-terminal molecular devices	1. Logic element 2. Memory	1. Small size	1. No inherent device gain. 2. Scaling to large memory size may be difficult without gain. 3. Placement of molecules in a circuit difficult and not yet demonstrated. 4. Temperature stability of organic molecules may be problematic.	Sixteen-bit cross-point memory demonstrated.
Carbon nanotube FET	Logic element	1. Ballistic transport (high speed) 2. Small size	1. Placement of nanotubes in a circuit difficult and not yet demonstrated. 2. Control of electrical properties of carbon nanotube (size, chirality) difficult and not yet achieved.	Device scaling properties not yet explored. Inverter circuit demonstrated.
DNA computing	Logic element	1. High parallelism	1. Imperfect yield. 2. General-purpose computing not possible.	

The standby power increase in recent technology nodes has been driven by increased gate leakage current due to the thin gate dielectric and by increased subthreshold off-current due to reduced threshold voltage to keep the gate overdrive ($V_{DD} - V_t$) large to provide for a large on-current. This trend of standby power increase obviously cannot continue unbounded. Active power can be traded against performance tradeoff (through a reduced power-supply voltage) [8] if the device drive current (at a fixed off-current) improves. The technology options described in Section 2 improve device performance (as outlined in Table 1) while at the same time addressing the power-consumption issue.

The gate leakage current is reduced by $10^3\times$ to $10^6\times$ using high-k gate dielectrics [21]. The use of a metal gate electrode will reduce the T_{inv} by another 0.4–0.5 nm, providing a further 50–100× reduction in gate leakage current. Double-gate FET reduces the subthreshold slope from 80 mV/decade (for bulk FET) to 60 mV/decade, providing at least 10× reduction in subthreshold leakage current. Improving the current drive of the MOSFET relieves the pressure to further reduce the threshold voltage and keep the power-supply voltage high. The improved carrier transport provided by the double-gate FET and strained-silicon FET enables this performance/active-power tradeoff. While there are qualitative evidences that this tradeoff exists [61, 85, 104], it is too early to provide a quantitative estimate.

As device scaling progresses, we expect to see power-management techniques more aggressively deployed to relieve the constraints of standby power consumption. This includes techniques on the device, circuit, and system levels. On the device level, techniques such as active well bias [150, 151] and the back-gate (ground plane) FET [49, 50, 54] can be employed to attain the largest gate overdrive for a fixed off-current. Another example is the use of multiple power supplies and threshold voltages custom-suited for particular circuit blocks or system functions (this is an obvious extension of the SRAM power-management strategy to random logic circuit blocks).

On the system level, embedding the memory subsystem (e.g., DRAM) with the processor on the same chip increases the on-chip communication bandwidth and reduces power dissipation by eliminating the requirement of driving signals off-chip. An expanded hierarchy of on-chip memory will further optimize power consumption. Continued increases in device density enable multiple processors (which may share the same on-chip memory) on the same chip. Three-dimensional (3D) integration of active device layers will reduce the communication distance between circuit blocks in the system, thereby improving execution speed and reducing power devoted to communications (both signal and power distribution). It also opens up the possibility of combining chips optimized for different technologies into a single chip. For high-performance systems in which cooling is an option, lowering the operating temperature below room temperature increases the carrier mobility and sharpens the on–off transition [152]. A significant improvement in device performance can be obtained even without resorting to liquid nitrogen (77 K) temperature [153].

Finally, and perhaps most significantly, we note that a comparison of custom system/circuit design with general-purpose processor design reveals that the energy–delay product can be improved by at least two orders of magnitude through designs custom-optimized for the specific application [9, 154]. In an era in which power consumption dominates system design tradeoffs, we expect a proliferation of special-purpose processing elements designed for minimizing power consumption. This customization will occur at the device, circuit, and system level.

5. Conclusion

From an economic point of view, conventional devices and materials will continue to be employed until they become impractical. Efforts devoted to push conventional approaches as far as they can go [13, 15, 47] continue to be extremely important. In this paper, we review the approaches to circumvent or surmount the barriers to device scaling. These approaches generally fall into two categories: new materials and new device structures. We describe materials innovations for the gate stack and the transistor channel. We also review device structural innovations such as the double-gate FET. While we focus this paper on these unconventional approaches, it is expected that innovations in conventional technologies such as etching, CMP, dopant profile control, and contact formation (silicides) will continue to be advanced and needed even for these "unconventional" approaches. Our application focus is high-performance systems. One may well arrive at a quite different conclusion if considerations are driven by other applications (e.g., low power, analog, sensor systems).

While nanotechnology may be seen by many as a successor to silicon microelectronics technology, it is clear from the results to date that it will be many years before nanotechnology can reach the level of maturity of the present silicon technology. The deployment of nanotechnology will most probably occur first in niche applications that complement conventional silicon technologies. Future social and application changes will further the shift from microelectronics to nanotechnology on a time scale of decades.

In the near term, there are many avenues for system performance improvements stemming from device, circuit, and system optimization that are yet to be

exploited, which may provide orders of magnitude of power–performance tradeoff. However, this may call for system products based on application-specific device, circuit, and system designs, as opposed to today's general-purpose approach. Despite some speculations as to the impending end of progress, there is still plenty of room for continued technological advancement.

Acknowledgments

This paper is distilled from collaborations and discussions over the years with many colleagues too numerous to mention individually, both within IBM as well as in the academic community and the technical community at large. I have learned from many colleagues at IBM with whom I have had the privilege to work: Ph. Avouris, D. Boyd, K. K. Chan, J. O. Chu, D. J. Frank, K. W. Guarini, L. J. Huang, M. Ieong, E. C. Jones, T. Kanarsky, J. Kedzierski, S. Koester, R. Martel, P. M. Mooney, E. J. Nowak, K. Rim, P. M. Solomon, Y. Taur (now at UC San Diego), S. Tiwari (now at Cornell), C. Wann, and J. Welser. Most technical work discussed in this paper was performed in collaboration with these colleagues at various times, with the support of the personnel and facilities at the ASTC and ASTL/S at the East Fishkill and Yorktown Heights sites. I am also indebted to colleagues at IBM with whom I had the opportunity to work closely and share technical information: C. T. Black, D. A. Buchanan, S. L. Cohen, E. Gusev, K. P. Muller, and C. Murray. In addition, discussions with D. A. Antoniadis (MIT), J. Bokor (UC Berkeley), R. H. Dennard, M. Fischetti, C. Hu (UC Berkeley, TSMC), S. Laux, T. H. Ning, D. Radack (DARPA), and T. N. Theis have been particularly important in shaping my thoughts. Many colleagues kindly provided data and information for this paper, cited individually in the references. Critical reading of the manuscript by M. Fischetti, D. Frank, K. Guarini, and K. Rim is much appreciated. Finally, I am particularly grateful to J. Warlaumont and R. Isaac for their support over many years.

References

1. *International Technology Roadmap for Semiconductors*, 1999 Edition, Semiconductor Industry Association (SIA), Austin, Texas: SEMATECH, USA, 2706 Montopolis Drive, Austin, Texas 78741; *http://www.itrs.net/ntrs/publntrs.nsf*.
2. G. Moore, "Progress in Digital Integrated Electronics," *IEDM Tech. Digest*, pp. 11–13 (1975).
3. R. Dennard, F. H. Gaensslen, H. N. Yu, L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE J. Solid State Circuits* **SC-9,** 256 (1974).
4. P. Packan, "Pushing the Limits," *Science* **285,** 2079 (1999).
5. B. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics—I MOS Technology," *Solid State Electron.* **15,** 819 (1972).
6. J. Meindl, "Low Power Microelectronics—Retrospect and Prospect," *Proc. IEEE* **83,** 619 (1995).
7. S. Asai and Y. Wada, "Technology Challenges for Integration Near and Below 0.1 μm," *Proc. IEEE* **85,** 505 (1997).
8. Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S.-H. Lo, G. Sai-Halasz, R. Viswanathan, H.-J. C. Wann, S. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," *Proc. IEEE* **85,** 486 (1997).
9. H.-S. P. Wong, D. Frank, P. M. Solomon, H.-J. Wann, and J. Welser, "Nanoscale CMOS," *Proc. IEEE* **87,** 537 (1999).
10. D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE* **89,** 259–288 (2001).
11. R. W. Keyes, "Fundamental Limits of Silicon Technology," *Proc. IEEE* **89,** 227–239 (2001).
12. J. Plummer and P. Griffin, "Material and Process Limits in Silicon VLSI Technology," *Proc. IEEE* **89,** 240–258 (2001).
13. R. Chau, J. Kavalieros, B. Roberds, R. Schenker, D. Lionberger, D. Barlage, B. Doyle, R. Arghvani, A. Murthy, and G. Dewey, "30 nm Physical Gate Length CMOS Transistors with 1.0 ps n-MOS and 1.7 ps p-MOS Gate Delays," *IEDM Tech. Digest*, pp. 45–48 (2000).
14. R. Chau, "30 nm and 20 nm Physical Gate Length CMOS Transistors," *Proceedings of the IEEE Silicon Nanoelectronics Workshop*, 2001, pp. 2–3.
15. B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, and M. Lin, "15 nm Gate Length Planar CMOS Transistor," *IEDM Tech. Digest*, p. 937 (2001).
16. S.-H. Lo, D. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFETs," *IEEE Electron Device Lett.* **18,** 209–211 (1997).
17. D. A. Buchanan, "Scaling the Gate Dielectric: Materials, Integration, and Reliability," *IBM J. Res. & Dev.* **43,** 245–264 (1999).
18. E. P. Gusev, H.-C. Lu, E. L. Garfunkel, T. Gustafsson, and M. L. Green, "Growth and Characterization of Ultrathin Nitrided Silicon Dioxide Films," *IBM J. Res. & Dev.* **43,** 265–286 (1999).
19. K. Hubbard and D. Schlom, "Thermodynamic Stability of Binary Oxides in Contact with Silicon," *J. Mater. Res.* **11,** 2757 (1996).
20. E. Gusev, E. Cartier, D. Buchanan, M. Gribelyuk, M. Copel, H. Okorn-Schmidt, and C. D'Emic, "Ultra High-*k* Metal Oxides on Silicon: Processing, Characterization, and Integration Issues," *Proceedings of the Conference on Insulating Films on Semiconductors (INFOS)*, 2001.
21. E. Gusev, D. Buchanan, E. Cartier, A. Kumar, D. DiMaria, S. Guha, A. Callegari, S. Zafar, P. Jamison, D. Neumayer, M. Copel, M. Gribelyuk, H. Okorn-Schmidt, C. D'Emic, P. Kozlowski, K. Chan, N. Bojarczuk, L.-A. Rannarsson, P. Ronsheim, K. Rim, R. Fleming, A. Mocuta, and A. Ajmera, "Ultrathin High-*k* Gate Stacks for Advanced CMOS Devices," *IEDM Tech. Digest*, pp. 451–454 (2001).
22. D. Frank and H.-S. P. Wong, "Analysis of the Design Space Available for High-*k* Gate Dielectric in Nanoscale MOSFETs," *Proceedings of the IEEE Silicon Nanoelectronics Workshop*, 2000, pp. 47–48.
23. S. M. Sze, *Physics of Semiconductor Devices*, Wiley, New York, 1981.
24. T. Ning, C. Osburn, and H. Yu, "Emission Probability of Hot Electrons from Silicon into Silicon Dioxide," *J. Appl. Phys.* **48,** 286 (1977).
25. J. Robertson, "Band Offsets of Wide-Band-Gap Oxides and Implications for Future Electronic Devices," *J. Vac. Sci. Technol. B* **18,** 1785–1791 (2000).

26. D. Buchanan, E. Gusev, E. Cartier, H. Okorn-Schmidt, K. Rim, M. Gribelyuk, A. Mocuta, A. Ajmera, M. Copel, S. Guha, N. Bojarczuk, A. Callegari, C. D'Emic, P. Kozlowski, K. Chan, R. J. Fleming, P. Jamison, J. Brown, and R. Arndt, "80 nm Poly-Silicon Gated n-FETs with Ultra-Thin Al_2O_3 Gate Dielectric for ULSI Applications," *IEDM Tech. Digest*, pp. 223–226 (2000).
27. M. Fischetti, D. Neumayer, and E. Cartier, "Effective Electron Mobility in Si Inversion Layers in MOS Systems with a High-κ Insulator: The Role of Remote Phonon Scattering," *J. Appl. Phys.* **90,** 4587 (2001).
28. M. Fischetti, "Long Range Coulomb Interactions in Small Si Devices. Part II: Effective Electron Mobility in Thin Oxide Structures," *J. Appl. Phys.* **89,** 1232–1250 (2001).
29. T.-J. King, J. Pfiester, and K. Saraswat, "A Variable Work-Function Polycrystalline $Si_{1-x}Ge_x$ Gate Material for Submicrometer CMOS Technologies," *IEEE Electron Device Lett.* **12,** 533–535 (1991).
30. J. Hauser, "Gate Dielectrics for Sub-100 nm CMOS," in *IEDM Short Course: Sub-100 nm CMOS*, M. Bohr, Ed., *IEDM Tech. Digest* (1999).
31. Y.-S. Suh, G. Heuss, H. Zhong, S.-N. Hong, and V. Misra, "Electrical Characteristics of $TaSi_xN_y$ Gate Electrodes for Dual Gate Si-CMOS Devices," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, pp. 47–48.
32. D.-G. Park, K.-Y. Lim, H.-J. Cho, T.-H. Cha, J.-J. Kim, J.-K. Ko, I.-S. Yeo, and J.-W. Park, "Novel Damage-Free Direct Metal Gate Process Using Atomic Layer Deposition," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, pp. 65–66.
33. J. Tersoff, "Theory of Semiconductor Heterojunctions: The Role of Quantum Dipoles," *Phys. Rev. B* **30,** 4874 (1984).
34. Y.-C. Yeo, P. Ranade, Q. Lu, R. Lin, T.-J. King, and C. Hu, "Effects of High-κ Dielectrics on the Workfunctions of Metal and Silicon Gates," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, pp. 49–50.
35. A. Chatterjee, R. Chapman, G. Dixit, J. Kuehne, S. Hattangady, H. Yang, G. Brown, R. Aggarwal, U. Erdogan, Q. He, M. Hanratty, D. Rogers, S. Murtaza, S. Fang, R. Kraft, A. Rotondaro, J. Hu, M. Terry, W. Lee, C. Fernando, A. Konecni, G. Wells, D. Frystak, C. Bowen, M. Rodder, and I.-C. Chen, "Sub-100 nm Gate Length Metal Gate NMOS Transistors Fabricated by a Replacement Gate Process," *IEDM Tech. Digest*, pp. 821–824 (1997).
36. Q. Lu, R. Lin, P. Ranade, T.-J. King, and C. Hu, "Metal Gate Workfunction Adjustment for Future CMOS Technology," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, pp. 45–46.
37. I. Polishchuk, P. Ranade, T.-J. King, and C. Hu, "Dual Workfunction Metal Gate CMOS Technology Using Metal Interdiffusion," *IEEE Electron Device Lett.* **22,** 444–446 (2001).
38. T. Sekigawa and Y. Hayashi, "Calculated Threshold-Voltage Characteristics of an XMOS Transistor Having an Additional Bottom Gate," *Solid State Electron.* **27,** 827 (1984).
39. F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-Gate Silicon-on-Insulator Transistor with Volume Inversion: A New Device with Greatly Enhanced Performance," *IEEE Electron Device Lett.* **8,** 410 (1987).
40. J. Colinge, M. Gao, A. Romano-Rodriguez, H. Maes, and C. Claeys, "Silicon-on-Insulator Gate-All-Around Device," *IEDM Tech. Digest*, p. 595 (1990).
41. C. Fiegna, H. Iwai, T. Wada, T. Saito, E. Sangiorgi, and B. Ricco, "A New Scaling Methodology for the 0.1–0.025 μm MOSFET," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 1992, p. 33.
42. D. Frank, S. Laux, and M. Fischetti, "Monte Carlo Simulation of a 30nm Dual-Gate MOSFET: How Far Can Si Go?," *IEDM Tech. Digest*, p. 553 (1992).
43. R. Yan, A. Ourmazd, and K. Lee, "Scaling the Si MOSFET: From Bulk to SOI to Bulk," *IEEE Trans. Electron Devices* **39,** 1704 (1992).
44. K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling Theory for Double-Gate SOI MOSFET's," *IEEE Trans. Electron Devices* **40,** 2326 (1993).
45. T. Tanaka, K. Suzuki, H. Horie, and T. Sugii, "Ultrafast Low-Power Operation of $p^+ - n^+$ Double-Gate SOI MOSFETs," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 1994, p. 11.
46. H.-S. Wong, D. Frank, Y. Taur, and J. Stork, "Design and Performance Considerations for Sub-0.1μm Double-Gate SOI MOSFET's," *IEDM Tech. Digest*, p. 747 (1994).
47. Y. Taur, C. Wann, and D. J. Frank, "25 nm CMOS Design Considerations," *IEDM Tech. Digest*, pp. 789–792 (1998).
48. H.-S. P. Wong, "Novel Device Options for Sub-100 nm CMOS," in *IEDM Short Course: Sub-100 nm CMOS*, M. Bohr, Ed., presented at the IEEE International Electron Devices Meeting, 1999.
49. I. Yang, C. Vieri, A. Chandrakasan, and D. Antoniadis, "Back-Gated CMOS on SOIAS for Dynamic Threshold Voltage Control," *IEEE Trans. Electron Devices* **44,** 822 (1997).
50. H.-S. Wong, D. Frank, and P. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25 nm Channel Length Generation," *IEDM Tech. Digest*, p. 407 (1998).
51. D. Frank, Y. Taur, and H.-S. P. Wong, "Generalized Scale Length for Two-Dimensional Effects in MOSFET's," *IEEE Electron Device Lett.* **19,** 385 (1998).
52. C. Y. Chang and S. M. Sze, Eds., *ULSI Devices*, Wiley, New York, 2000, Ch. 3.
53. L. Chang, S. Tang, T.-J. King, J. Bokor, and C. Hu, "Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs," *IEDM Tech. Digest*, pp. 719–722 (2000).
54. S. Tang, L. Chang, N. Lindert, Y.-K. Choi, W.-C. Lee, X. Huang, V. Subramanian, J. Bokor, T.-J. King, and C. Hu, "FinFET—A Quasi-Planar Double-Gate MOSFET," *ISSCC Tech. Digest*, pp. 118–119 (2001).
55. Y. Taur, "Analytic Solutions of Charge and Capacitance in Symmetric and Asymmetric Double-Gate MOSFETs," *IEEE Trans. Electron Devices* **48,** 2861 (2001).
56. M. Ieong, H.-S. P. Wong, Y. Taur, P. Oldiges, and D. Frank, "DC and AC Performance Analysis of 25 nm Symmetric/Asymmetric Double-Gate, Back-Gate and Bulk CMOS," *Proceedings of the Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2000, pp. 147–150.
57. C. Kittel, *Introduction to Solid State Physics*, Wiley, New York, 1956, Ch. 11, p. 283.
58. T. Tanaka, H. Horie, S. Ando, and S. Hijiya, "Analysis of p+ Poly Si Double-Gate Thin-Film SOI MOSFETs," *IEDM Tech. Digest*, p. 683 (1991).
59. M. Sherony, L. Su, J. Chung, and D. Antoniadis, "SOI MOSFET Effective Channel Mobility," *IEEE Trans. Electron Devices* **41,** 276–278 (1994).
60. S. Takagi, I. Iwase, and A. Toriumi, "On the Universality of Inversion-Layer Mobility in n- and p-Channel MOSFET's," *IEDM Tech. Digest*, p. 398 (1988).
61. M. Ieong, E. Jones, T. Kanarsky, Z. Ren, O. Dokumaci, R. Roy, L. Shi, T. Furukawa, Y. Taur, R. Miller, and H.-S. P. Wong, "Experimental Evaluation of

Carrier Transport and Device Design for Planar Symmetric/Asymmetric Double-Gate/Ground-Plane CMOSFETs," *IEDM Tech. Digest*, pp. 441–444 (2001).
62. J.-H. Choi, Y.-J. Park, and H.-S. Min, "Electron Mobility Behavior in Extremely Thin SOI MOSFETs," *IEEE Electron Device Lett.* **16,** 527 (1995).
63. A. Toriumi, J. Koga, H. Satake, and A. Ohata, "Performance and Reliability Concerns of Ultra-Thin SOI and Ultra-Thin Gate Oxide MOSFETs," *IEDM Tech. Digest*, p. 847 (1995).
64. T. Ernst, D. Munteanu, S. Cristoloveanu, T. Ouisse, S. Horiguchi, Y. Ono, Y. Takahashi, and K. Murase, "Investigation of SOI MOSFETs with Ultimate Thickness," *Microelectron. Eng.* **48,** 339–342 (1999).
65. M. Shoji and S. Horiguchi, "Electronic Structures and Phonon-Limited Electron Mobility of Double-Gate Silicon-on-Insulator Si Inversion Layers," *J. Appl. Phys.* **85,** 2722 (1999).
66. F. Gamiz, J. A. Lopez-Villanueva, J. Roldan, J. Carceller, and P. Cartujo, "Monte Carlo Simulation of Electron Transport Properties in Extremely Thin SOI MOSFETs," *IEEE Trans. Electron Devices*, p. 1122 (1998).
67. F. Gamiz, J. Roldan, P. Cartujo-Cassinello, J. Carceller, J. Lopez-Villanueva, and S. Rodriguez, "Electron Mobility in Extremely Thin Single-Gate Silicon-on-Insulator Inversion Layers," *J. Appl. Phys*, p. 6269 (1999).
68. D. Esseni, M. Mastrapasqua, G. Celler, F. Baumann, C. Fiegna, L. Selmi, and E. Sangiorgi, "Low Field Mobility of Ultra-Thin SOI n- and p-MOSFETs: Measurements and Implications on the Performance of Ultra-Short MOSFETs," *IEDM Tech. Digest*, pp. 671–674 (2000).
69. D. Esseni, M. Mastrapasqua, C. Fiegna, G. Celler, L. Selmi, and E. Sangiorgi, "Low Field Electron Mobility in Double-Gate Ultra-Thin SOI MOSFETs," *IEDM Tech. Digest*, p. 445 (2001).
70. E. Leobandung, J. Gu, L. Guo, and S. Chou, "Wire Channel and Wrap Around Gate Metal Oxide Semiconductor Field Effect Transistors with a Significant Reduction of Short Channel Effects," *J. Vac. Sci. Technol. B* **15,** 2791–2794 (1997).
71. H.-S. Wong, K. Chan, and Y. Taur, "Self-Aligned (Top and Bottom) Double-Gate MOSFET with a 25 nm Thick Silicon Channel," *IEDM Tech. Digest*, p. 427 (1997).
72. G. Neudeck, T.-C. Su, and J. P. Denton, "Novel Silicon Epitaxy for Advanced MOSFET Devices," *IEDM Tech. Digest*, pp. 169–172 (2000).
73. J.-H. Lee, G. Tarashi, A. Wei, T. Langdo, E. A. Fitzgerald, and D. Antoniadis, "Super Self-Aligned Double-Gate (SSDG) MOSFETs Utilizing Oxidation Rate Difference and Selective Epitaxy," *IEDM Tech. Digest*, p. 71 (1999).
74. K. W. Guarini, P. M. Solomon, Y. Zhang, K. K. Chan, E. C. Jones, G. M. Cohen, A. Krasnoperova, M. Ronay, O. Dokumaci, J. J. Bucchignano, C. C. Jr., C. Lavoie, V. Ku, D. C. Boyd, K. S. Petrarca, I. V. Babich, J. Treichler, P. M. Kozlowski, J. S. Newbury, C. P. D Emic, R. M. Sicina, and H.-S. P. Wong, "Triple-Self-Aligned, Planar Double-Gate MOSFETs: Devices and Circuits," *IEDM Tech. Digest*, pp. 425–428 (2001).
75. S. Venkatesan, C. Subramanian, G. Neudeck, and J. Denton, "Thin-Film Silicon-on-Insulator (SOI) Device Applications of Selective Epitaxial Growth," *Proceedings of the International SOI Conference*, 1993, p. 76.
76. J. Siekkinen, G. Neudeck, J. Glenn, and S. Venkatesan, "A Novel High-Speed Silicon Bipolar Transistor Utilizing SEG and CLSEG," *IEEE Trans. Electron Devices*, p. 862 (1994).
77. A. Ogura and Y. Fujimoto, "Extremely Thin and Defect-Free Si-on-Insulator Fabrication by Tunnel Epitaxy," *Appl. Phys. Lett.* **57,** 2806 (1990).
78. A. Ogura, A. Furuya, and R. Koh, "50-nm-Thick Silicon-on-Insulator Fabrication by Advanced Epitaxial Lateral Overgrowth: Tunnel Epitaxy," *J. Electrochem. Soc.* **140,** 1125 (1993).
79. H.-S. Wong, K. Chan, Y. Lee, P. Roper, and Y. Taur, "Ultra-Thin, Highly Uniform Thin Film SOI MOSFET with Low Series Resistance Fabricated Using Pattern-Constrained Epitaxy (PACE)," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 1996, p. 94.
80. G. Cohen and H.-S. P. Wong, "Self-Aligned Double-Gate MOSFET by Selective Epitaxy and Silicon Wafer Bonding Techniques," Technical Report, 1999; patent pending.
81. X. Huang, W.-C. Lee, C. Ku, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," *IEDM Tech. Digest*, p. 67 (1999).
82. D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "A Fully Depleted Lean-Channel Transistor (DELTA)—A Novel Vertical Ultra Thin SOI MOSFET," *IEDM Tech. Digest*, p. 833 (1989).
83. D. Fried, A. Johnson, E. Nowak, J. Rankin, and C. Willets, "A Sub-40 nm Body Thickness N-Type FinFET," *Proceedings of the Device Research Conference*, 2001, p. 24.
84. N. Lindert, L. Chang, Y.-K. Choi, E. Anderson, W.-C. Lee, T.-J. King, J. Bokor, and C. Hu, "Sub-60-nm Quasi-Planar FinFETs Fabricated Using a Simplified Process," *IEEE Electron Device Lett.* **22,** 487–489 (2001).
85. J. Kedzierski, D. M. Fried, E. J. Nowak, T. Kanarsky, J. H. Rankin, H. Hanafi, W. Natzle, D. Boyd, Y. Zhang, R. A. Roy, J. Newbury, C. Yu, Q. Yang, P. Saunders, C. P. Willets, A. Johnson, S. P. Cole, H. E. Young, N. Carpenter, D. Rakowski, B. A. Rainey, P. E. Cottrell, M. Ieong, and H.-S. P. Wong, "High-Performance Symmetric-Gate and CMOS-Compatible V_t Asymmetric-Gate FinFET Devices," *IEDM Tech. Digest*, pp. 437–440 (2001).
86. Y.-K. Choi, N. Lindert, P. Xuan, S. Tang, D. Ha, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET Process Technology for Nanoscale CMOS," *IEDM Tech. Digest*, p. 421 (2001).
87. D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, "A Folded-Channel MOSFET for Deep-Sub-Tenth Micron Era," *IEDM Tech. Digest*, p. 1032 (1998).
88. P. Oldiges, Q. Lin, K. Petrillo, M. Sanchez, M. Ieong, and M. Hargrove, "Modeling Line Edge Roughness Effects in Sub 100 Nanometer Gate Length Devices," *Proceedings of the Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2000, pp. 131–134.
89. M. Fischetti and S. Laux, "Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structures—Part II: Submicrometer MOSFETs," *IEEE Trans. Electron Devices* **38,** 650–660 (1991).
90. S. Takagi, M. Takaytanagi-Takagi, and A. Toriumi, "Accurate Characterization of Electron and Hole Inversion-Layer Capacitance and Its Impact on Low Voltage Operation of Scaled MOSFETs," *IEDM Tech. Digest*, pp. 619–622 (1998).
91. P. Solomon and S. Laux, "The Ballistic FET: Design, Capacitance and Speed Limit," *IEDM Tech. Digest*, p. 95 (2001).
92. M. Lundstrom, "On the Mobility Versus Drain Current Relation for a Nanoscale MOSFET," *IEEE Electron Device Lett.* **22,** 293–295 (2001).
93. T. Vogelsang and H. R. Hofmann, "Electron Transport in Strained Silicon Layers on $Si_{1-x}Ge_x$ Substrates," *Appl. Phys. Lett.* **63,** 186 (1993).

94. D. Nayak, J. Woo, J. Park, K. Wang, and K. MacWilliams, "High-Mobility p-Channel Metal-Oxide-Semiconductor Field-Effect Transistors on Strained Si," *Appl. Phys. Lett.* **62,** 2853–2855 (1993).
95. J. Welser, J. Hoyt, S. Takagi, and J. Gibbons, "Strain Dependence of the Performance Enhancement in Strained-Si n-MOSFETs," *IEDM Tech. Digest*, pp. 373–376 (1994).
96. M. Fischetti and S. Laux, "Band Structure, Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiGe Alloys," *J. Appl. Phys.* **80,** 2234 (1996).
97. S. Tiwari, M. Fischetti, P. Mooney, and J. Welser, "Hole Mobility Improvement in Silicon-on-Insulator and Bulk Silicon Transistors Using Local Strain," *IEDM Tech. Digest*, pp. 939–941 (1997).
98. K. Rim, J. Hoyt, and J. Gibbons, "Transconductance Enhancement in Deep Submicron Strained-Si n-MOSFETs," *IEDM Tech. Digest*, p. 707 (1998).
99. T. Mizuno, N. Sugiyama, H. Satake, and S. Takagi, "Advanced SOI-MOSFETs with Strained-Si Channel for High Speed CMOS—Electron/Hole Mobility Enhancement," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2000, p. 210.
100. K. Ismail, "Si/SiGe CMOS: Can It Extend the Lifetime of Si," *ISSCC Tech. Digest*, pp. 116–117 (1997).
101. P. Mooney, "Strain Relaxation and Dislocations in SiGe/Si Structures," *Mater. Sci. Eng.* **R17,** 105–146 (1996).
102. E. Fitzgerald, Y. Xie, D. Monroe, P. Silverman, J. Kuo, A. Kortan, F. Theil, and B. Weir, "Relaxed Ge_xSi_{1-x} Structures for III/V Integration with Si and High Mobility Two-Dimensional Electron Gases in Si," *J. Vac. Sci. Technol. B* **10,** 1087 (1992).
103. K. Ismail, F. Nelson, J. Chu, and B. Meyerson, "Electron Transport Properties of Si/SiGe Heterostructures: Measurements and Device Implications," *Appl. Phys. Lett.* **63,** 660–662 (1993).
104. K. Rim, S. Koester, M. Hargrove, J. Chu, P. M. Mooney, J. Ott, T. Kanarsky, P. Ronsheim, M. Ieong, A. Grill, and H.-S. P. Wong, "Strained Si NMOSFETs for High Performance CMOS Technology," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, p. 59.
105. L.-J. Huang, J. Chu, C. Canaperi, C. D'Emic, R. Anderson, S. Koester, and H.-S. P. Wong, "SiGe-on-Insulator Prepared by Wafer Bonding and Layer Transfer for High-Performance Field-Effect Transistors," *Appl. Phys. Lett.* **78,** 1267 (2001).
106. L.-J. Huang, J. Chu, S. A. Goma, C. D'Emic, S. J. Koester, D. F. Canaperi, P. M. Mooney, S. A. Cordes, J. L. Speidell, R. M. Anderson, and H.-S. P. Wong, "Carrier Mobility Enhancement in Strained Si-on-Insulator Fabricated by Wafer Bonding," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, p. 57.
107. Z.-Y. Cheng, M. Currie, C. Leitz, G. Taraschi, E. Fitzgerald, J. Hoyt, and D. Antoniadis, "Electron Mobility Enhancement in Strained-Si n-MOSFETs Fabricated on SiGe-on-Insulator (SGOI) Substrates," *IEEE Electron Device Lett.* **22,** 321–323 (2001).
108. T. Mizuno, N. Sugiyama, A. Kurobe, and S. Takagi, "Advanced SOI p-MOSFETs with Strained-Si Channel on SiGe-on-Insulator Substrate Fabricated by SIMOX Technology," *IEEE Trans. Electron Devices* **48,** 1612–1618 (2001).
109. T. Tezuka, N. Sugiyama, and S. Takagi, "Fabrication of Strained Si on an Ultrathin SiGe-on-Insulator Substrate with a High-Ge Fraction," *Appl. Phys. Lett.* **79,** 1798–1800 (2001).
110. T. Mizuno, N. Sugiyama, T. Tezuka, and S. Takagi, "Novel Fabrication Technique for Relaxed SiGe-on-Insulator Substrates Without Thick SiGe Buffer Structures," *Proceedings of the International Conference on Solid State Devices and Materials (SSDM)*, 2001, pp. 242–243.
111. A. R. Powell, S. S. Iyer, and F. K. LeGoues, "New Approach to the Growth of Low Dislocation Relaxed SiGe Material," *Appl. Phys. Lett.* **64,** 1856–1858 (1994).
112. S. Iijima, "Helical Microtubules of Graphitic Carbon," *Nature* **354,** 56 (1991).
113. S. Iijima and T. Ichihashi, "Single-Shell Carbon Nanotubes of 1-nm Diameter," *Nature* **363,** 603 (1993).
114. M. Dresselhaus, G. Dresselhaus, and P. Eklund, *Science of Fullerenes and Carbon Nanotubes*, Academic Press, New York, 1996.
115. P. Avouris, T. Hertel, R. Martel, T. Schmidt, H. Shea, and R. Walkup, "Carbon Nanotubes: Nanomechanics, Manipulation, and Electronic Devices," *Appl. Surf. Sci.* **141,** 201–209 (1999).
116. M. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds., *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Springer-Verlag, New York, 2001.
117. S. J. Tans, A. R. M. Verschueren, and C. Dekker, "Room-Temperature Transistor Based on a Single Carbon Nanotube," *Nature* **393,** 49–51 (1998).
118. R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single- and Multi-Wall Carbon Nanotube Field-Effect Transistors," *Appl. Phys. Lett.* **73,** 2447–2449 (1998).
119. Z. Ren, Z. Huang, J. Xu, J. Wang, P. Bush, M. P. Siegal, and P. Provencio, "Synthesis of Large Arrays of Well-Aligned Carbon Nanotubes on Glass," *Science* **282,** 1105–1107 (1998).
120. Q. Wang, A. Setlur, J. Lauerhaas, J. Dai, E. Seelig, and R. Chang, "A Nanotube Based Field-Emission Flat Panel Display," *Appl. Phys. Lett.* **72,** 2912–2913 (1998).
121. G. Che, B. B. Lakshmi, E. Fisher, and C. R. Martin, "Carbon Nanotube Membranes for Electrochemical Energy Storage and Production," *Nature* **393,** 346–349 (1998).
122. H. Dai, J. Hafner, A. Rinzler, D. Colbert, and R. E. Smalley, "Nanotubes as Nanoprobes in Scanning Probe Microscopy," *Nature* **384,** 147–150 (1996).
123. S. S. Wong, E. Joselevich, A. Woolley, C. L. Cheung, and C. M. Leiber, "Covalently Functionalized Nanotubes as Nanometre-Sized Probes in Chemistry and Biology," *Nature* **394,** 52–55 (1998).
124. H. Dai, N. Franklin, and J. Han, "Exploiting the Properties of Carbon Nanotubes for Nanolithography," *Appl. Phys. Lett.* **73,** 1508–1510 (1998).
125. P. L. McEuen, M. Bockrath, D. H. Cobden, Y.-G. Yoon, and S. Louie, "Disorder, Pseudospins, and Backscattering in Carbon Nanotubes," *Phys. Rev. Lett.* **83,** 5098–5101 (1999).
126. J. Wildoer, L. Venema, A. Rinzler, R. Smalley, and C. Dekker, "Electronic Structure of Atomically Resolved Carbon Nanotubes," *Nature* **391,** 59–64 (1998).
127. V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Carbon Nanotube Inter- and Intramolecular Logic Gates," *Nano Lett.* **1,** 453–456 (2001).
128. R. Martel, V. Derycke, C. Lavoie, J. Appenzeller, K. K. Chan, J. Tersoff, and P. Avouris, "Ambipolar Electrical Transport in Semiconducting Single-Wall Carbon Nanotubes," *Phys. Rev. Lett.* **87,** 256805 (2001).
129. A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker, "Logic Circuits with Carbon Nanotube Transistors," *Science*, October 4, 2001. Science Online, *10.1126/science.1065824*.
130. Y. Zhang, T. Ichihashi, E. Landree, F. Nihey, and S. Iijima, "Heterostructures of Single-Walled Carbon Nanotubes and Carbide Nanorods," *Science* **285,** 1719–1722 (1999).
131. R. Martel, H.-S. P. Wong, K. Chan, and P. Avouris, "Carbon Nanotube Field Effect Transistors for Logic Applications," *IEDM Tech. Digest* (2001).

132. J. Appenzeller R. Martel, P. Avouris, H. Stahl, and B. Lengeler, "Optimized Contact Configuration for the Study of Transport Phenomena in Ropes of Single-Wall Carbon Nanotubes," *Appl. Phys. Lett.* **78,** 3313 (2001).
133. C. White and T. Todorov, "Carbon Nanotubes as Long Ballistic Conductors," *Nature* **393,** 240–242 (1998).
134. A. Bachtold, M. S. Fuhrer, S. Plyasunov, M. Forero, E. H. Anderson, A. Zettl, and P. L. McEuen, "Scanned Probe Microscopy of Electronic Transport in Carbon Nanotubes," *Phys. Rev. Lett.* **84,** 6082–6085 (2000).
135. S. Wind, J. Appenzeller, R. Martel, V. Derycke, and P. Avouris, "Vertical Scaling of Single-Wall Carbon Nanotube CMOS Field Effect Transistors Using Top Gate Electrodes," *Appl. Phys. Lett.,* 2002, in press.
136. T. Ghani, S. Ahmed, P. Aminzadeh, J. Bielefeld, P. Charvat, C. Chu, M. Harper, P. Jacob, C. Jan, J. Kavalieros, C. Kenyon, R. Nagisetty, P. Packan, J. Sebastian, M. Taylor, J. Tsai, S. Tyagi, S. Yang, and M. Bohr, "100 nm Gate Length High Performance Low Power CMOS Transistor Structure," *IEDM Tech. Digest*, pp. 415–418 (1999).
137. S. Ramo, J. Whinnery, and T. van Duzer, *Fields and Waves in Communication Electronics*, Wiley, New York, 1965, p. 188.
138. C. Zhou, J. Kong, and H. Dai, "Electrical Measurements of Individual Semiconducting Single-Walled Carbon Nanotubes of Various Diameters," *Appl. Phys. Lett.* **76,** 1597–1599 (2000).
139. K. Liu, M. Burghard, S. Roth, and P. Bernier, "Conductance Spikes in Single-Walled Carbon Nanotube Field-Effect Transistor," *Appl. Phys. Lett.* **75,** 2494–2496 (1999).
140. H. Dai, A. Rinzler, P. Nikolaev, A. Thess, D. Colbert, and R. Smalley, "Single-Wall Nanotubes Produced by Metal-Catalyzed Disproportionation of Carbon Monoxide," *Chem. Phys. Lett.* **260,** 471–475 (1996).
141. M. Terrones, N. Grobert, J. Olivares, J. Zhang, H. Terrones, K. Kordatos, W. Hsu, J. Hare, P. Townsend, K. Prassides, A. Cheetham, H. Kroto, and D. Walton, "Controlled Production of Aligned-Nanotube Bundles," *Nature* **388,** 52–55 (1996).
142. Z. Ren, Z. Huang, D. Wang, J. Wen, J. Xu, J. Wang, L. Calvet, J. Chen, J. Klemic, and M. Reed, "Growth of a Single Freestanding Multiwall Carbon Nanotube on Each Nanonickel Dot," *Appl. Phys. Lett.* **75,** 1086–1088 (1999).
143. L. Sun, J. Mao, Z. Pan, B. Chang, W. Zhou, G. Wang, L. Qian, and S. Xie, "Growth of Straight Nanotubes with a Cobalt-Nickel Catalyst by Chemical Vapor Deposition," *Appl. Phys. Lett.* **74,** 644–646 (1999).
144. S. Fan, M. Chapline, N. Franklin, T. Tombler, A. M. Cassell, and H. Dai, "Self-Oriented Regular Arrays of Carbon Nanotubes and Their Field Emission Properties," *Science* **283,** 512–514 (1999).
145. J. Li, C. Papadopoulos, J. Xu, and M. Moskovits, "Highly-Ordered Carbon Nanotube Arrays for Electronics Applications," *Appl. Phys. Lett.* **73,** 367–369 (1999).
146. J. Mao, L. Sun, L. Qian, Z. Pan, B. Chang, W. Zhou, G. Wang, and S. Xie, "Growth of Carbon Nanotubes on Cobalt Disilicide Precipitates by Chemical Vapor Deposition," *Appl. Phys. Lett.* **72,** 3297–3299 (1998).
147. J. Hu, M. Ouyang, P. Yang, and C. Lieber, "Controlled Growth and Electrical Properties of Heterojunctions of Carbon Nanotubes and Silicon Nanowires," *Nature* **399,** 48–51 (1999).
148. N. Wang, Z. Tang, G. Li, and J. Chen, "Single-Walled 4A Carbon Nanotube Arrays," *Nature* **408,** 50–51 (2000).
149. R. Schlittler, J. Seo, J. Gimzewski, C. Durkan, M. Saifullah, and M. Welland, "Single Crystals of Single-Walled Carbon Nanotubes Formed by Self-Assembly," *Science* **292,** 1136–1139 (2001).
150. C. Wann, J. Harrington, R. Mih, S. Biesemans, K. Han, R. Dennard, O. Prigge, C. Lin, R. Mahnkopf, and B. Chen, "CMOS with Active Well Bias for Low-Power and RF/Analog Applications," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2000, pp. 158–159.
151. S.-F. Huang, C. Wann, Y.-S. Huang, C.-U. Lin, T. Schafbauer, S.-M. Cheng, Y.-C. Cheng, D. Vietzke, M. Eller, C. Lin, Q. Ye, N. Rovedo, S. Biesemans, P. Nguyen, R. Dennard, and B. Chen, "Scalability and Biasing Strategy for CMOS with Active Well Bias," *Symposium on VLSI Technology*, *Digest of Technical Papers*, 2001, pp. 107–108.
152. F. H. Gaensslen, V. L. Rideout, E. Walker, and J. Walker, "Very Small MOSFET's for Low-Temperature Operation," *IEEE Trans. Electron Devices* **24,** 218–229 (1977).
153. S. Wind, L. Shi, K.-L. Lee, R. Roy, Y. Zhang, E. Sikorski, P. Kozlowski, C. D'Emic, J. J. Bucchignano, H.-J. Wann, R. Viswanathan, J. Cai, and Y. Taur, "Very High Performance 50 nm CMOS at Low Temperature," *IEDM Tech. Digest*, p. 928 (1999).
154. D. Frank, "Design Considerations for CMOS Near the Limits of Silicon," *Proceedings of the Workshop on Ultimate Integration of Silicon*, January 20–21, 2000.
155. E. Jones, M. Ieong, O. Dokumaci, R. Roy, L. Shi, T. Furukawa, R. Miller, and H.-S. P. Wong, "High Performance of Planar Double Gate MOSFETs with Thin Backgate Dielectrics," *Proceedings of the Device Research Conference*, 2001, pp. 28–29.

Received November 6, 2001; accepted for publication March 6, 2002

H.-S. Philip Wong *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (hspwong@us.ibm.com).* Dr. Wong received the Ph.D. degree in electrical engineering from Lehigh University, Pennsylvania, in 1988, joining the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, that same year as a Research Staff Member. He is now Senior Manager of the Exploratory Devices and Integration Technology Department. His department is responsible for defining and executing IBM's exploratory devices and technology roadmap for silicon technology. While he manages a wide range of technical activities from e-beam lithography to quantum device modeling, Dr. Wong maintains an active personal research career that centers on solid-state devices, device physics and fabrication technology, applications of microelectronic systems, and solid-state image sensors. He is a Fellow of the IEEE and serves the IEEE Electron Devices Society (EDS) as an elected AdCom member. Since 1998 he has been a member of the IEDM Committee and the ISSCC Program Committee. He is a Distinguished Lecturer of the IEEE EDS. Dr. Wong has taught several short courses at the IEDM, ISSCC, SOI, Symposium on VLSI Technology, and SPIE conferences.