5

Compact Models for Small Geometry MOSFETs

5.1 Introduction

In this chapter the compact models for small geometry MOSFET (metal-oxide-semiconductor field-effect transistors) devices are presented. The continuous scaling of MOSFET devices toward decananometer regime has resulted in higher device density and faster circuit speed along with higher power dissipation [1–4]. Many new physical phenomena became significant with the device dimension rapidly approaching its physical limit. These include small geometry effects [5–8], channel length modulation (CLM) [9], drain-induced barrier lowering (DIBL) [10], velocity saturation [11], mobility degradation due to high vertical electric field [12], impact ionization [13], band-to-band tunneling [14], velocity overshoot [15], self-heating [16], inversion-layer quantization [17–19], polysilicon depletion [20], and process variability [21,22]. Thus, accurate MOSFET models that include the observed new physical phenomena are crucial to design and optimization of advanced very-large-scale-integrated (VLSI) circuits using nanoscale complementary metal-oxide-semiconductor (CMOS) technologies. In this chapter, we will use regional modeling approach to develop compact MOSFET models to accurately simulate different physical and small geometry effects in advanced VLSI circuits. First of all, we will derive different analytical expressions to model the deviation of long channel V_{th} model derived in Chapter 4 due to geometry and different physical effects and present an accurate V_{th} model for circuit CAD. Then we derive drain current model for short channel MOSFET devices considering high-field effects causing mobility degradation and velocity saturation.

5.2 Threshold Voltage Model

MOSFET threshold voltage model developed in Chapter 4 assumes uniformly doped substrate and neglects geometry effects on device performance. The expression for V_{th} for long channel MOSFETs with uniformly doped substrate is given by Equation 4.12 and can be generalized as

$$V_{th} = V_{fb} + \phi_s + \gamma\sqrt{\phi_s - V_{bs}} \tag{5.1}$$

where:

V_{fb}, ϕ_s, γ, and V_{bs} are the flat band voltage, surface potential, body effect coefficient, and back gate or body bias, respectively

Note that in Equation 5.1, $\phi_s = 2\phi_B$ in strong inversion as shown in Equation 4.12. In Equation 5.1, the body effect coefficient is defined as

$$\gamma = \frac{\sqrt{2qK_{si}\varepsilon_0 N_b}}{C_{ox}} \tag{5.2}$$

where:

q, K_{si}, ε_0, N_b are the electronic charge, permittivity of silicon, permittivity of free space, and substrate concentration, respectively

If we define $V_{TH0} = V_{th}$ @ $V_{bs} = 0$, then we can show

$$V_{th} = V_{TH0} + \gamma\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right) \tag{5.3}$$

Equation 5.3 models V_{th} for large geometry MOSFET devices of uniformly doped substrate with doping concentration, N_b. In Sections 5.2.1 and 5.2.2, we will derive analytical expressions to consider nonuniform substrate doping and different physical and geometrical effects in modeling V_{th} for advanced MOSFET devices.

5.2.1 Effect of Nonuniform Channel Doping on Threshold Voltage

In nanoscale MOSFET devices, the channel doping concentration N_b varies both vertically and laterally [23–26]. In advanced CMOS technology, the channel doping concentration is vertically nonuniform due to threshold voltage adjust implant dopants and laterally nonuniform due to halo doping implant around the source-drain (S/D) extension (SDE) regions as shown in Figure 5.1a and b.

In a conventional CMOS technology, the type of impurity for V_{th} adjust doping is the same as the channel doping. Thus, the V_{th} adjust implant in the channel increases the channel doping concentration near the surface, that is, provides high–low doping profile [19]. In some advanced technology, the threshold voltage adjust implant creates low–high implant or super-steep-retrograde channel doping profile [19]. The nonuniform vertical channel doping causes a strong dependence of the depletion charge, Q_b, on the applied body bias, V_{bs}, as shown in Figure 5.2a [22]. On the other hand, the nonuniform lateral channel doping causes strong dependence of V_{th} on the channel length (L) as shown in Figure 5.2b [25,26].

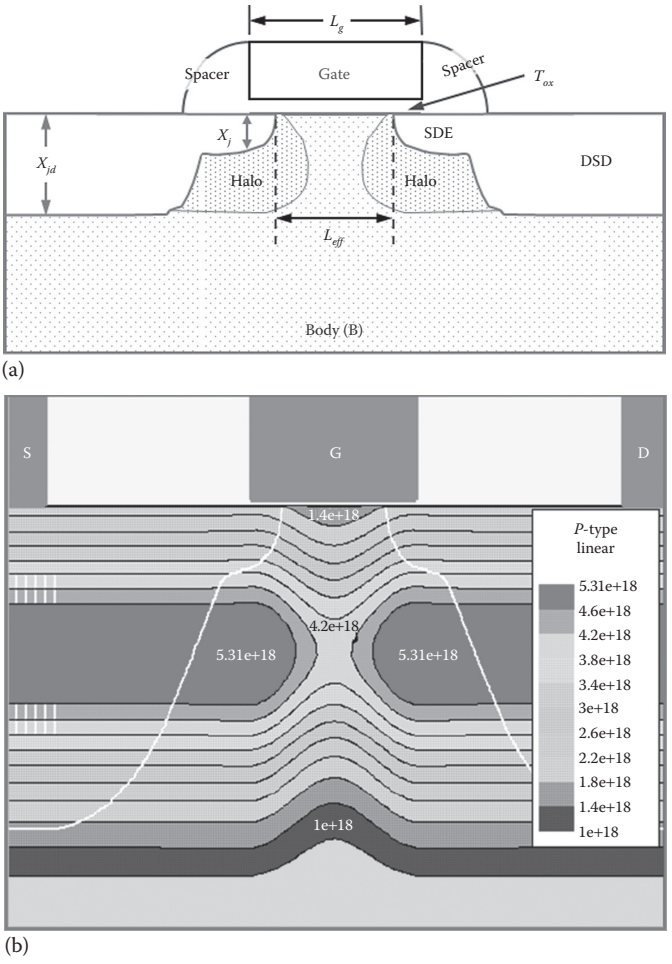

FIGURE 5.1
2D cross-section of a MOSFET device: (a) threshold adjust and halo implant causing nonuniform channel doping profile and (b) simulated 2D-doping contours of a typical double-halo *n*MOSFET device with laterally and vertically nonuniform *p*-type channel doping generated using device CAD MEDICI; 2D cross-section shows S, G, and D are the source, gate, and drain terminals, respectively, and the outline of SDE and deep source-drain (DSD) junctions. (Data from S. Saha, *Proc. SPIE Conf.*, 5042, 172–179, 2003.)

5.2.1.1 Threshold Voltage Modeling for Nonuniform Vertical Channel Doping Profile

Due to nonuniform channel doping, the body effect coefficient γ depends on the body bias, V_{bs}. For the simplicity of mathematical formulation, let us approximate the nonuniform vertical channel doping profile by a *high–low* step function as shown in Figure 5.3 with uniform concentration N_{CH} from

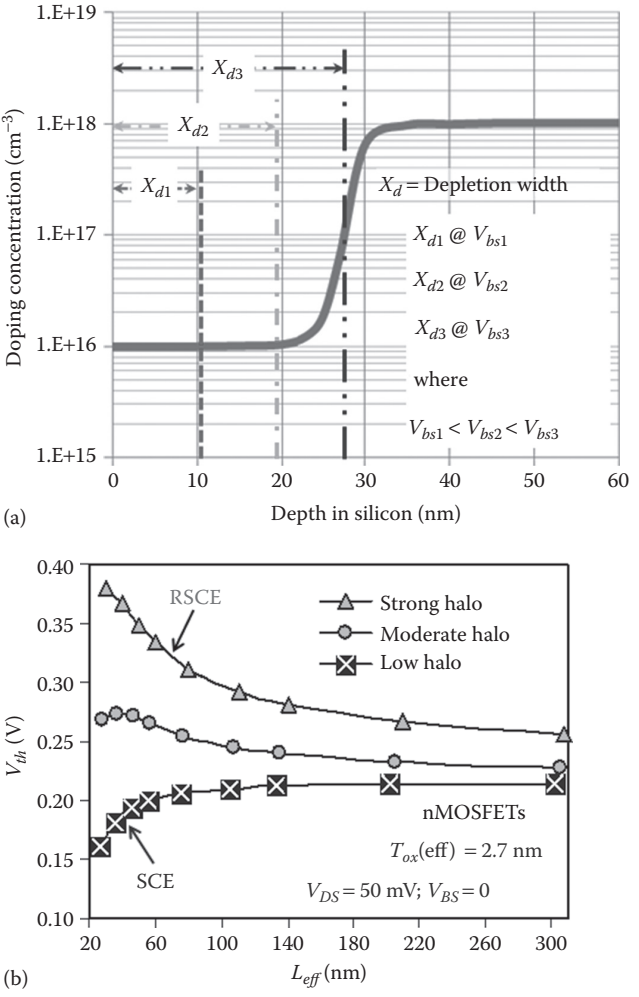

FIGURE 5.2
Effect of nonuniform channel doping profile on MOSFET devices: (a) body bias, V_{bs}, dependence of channel depletion widths (X_{d1}, X_{d2}, and X_{d3}) and bulk charge, Q_b, due to nonuniform vertical channel doping profile and (b) channel length dependence of V_{th} due to nonuniform lateral channel doping profile. (Data from S. Saha, *Proc. SPIE Conf.*, 3881, 195–204, 1999.)

the Si/SiO_2 interface to a depth X_T and N_{SUB} from X_T to the bottom of the silicon substrate.

With reference to Figure 5.3, let us assume that V_{bx} is the body bias required to fully deplete the region X_T. Then, for the applied body bias V_{bs}, we can show from Equation 5.3

$$V_{th} = V_{TH0} + \gamma_1\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right); \quad |V_{bs}| < |V_{bx}| \tag{5.4}$$

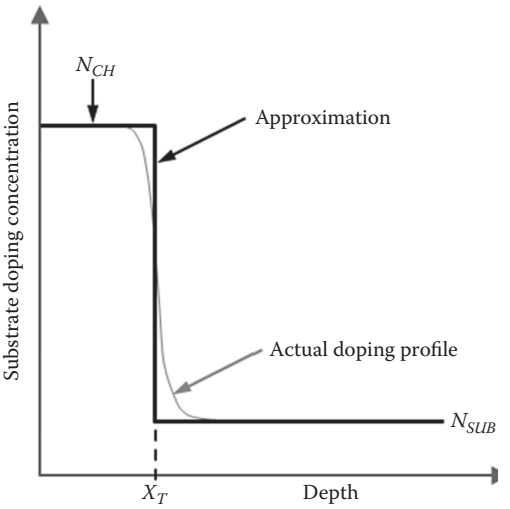

FIGURE 5.3
A typical nonuniform vertical channel doping profile of a MOSFET due to threshold voltage adjust implant approximated to a high–low step profile; N_{CH} and N_{SUB} are the channel doping concentrations at the surface and deep into the substrate, respectively; X_T is the transition depth of doping concentration from the high level to low level.

$$V_{th} = V_{TH0} + \gamma_1\left(\sqrt{\phi_s - V_{bx}} - \sqrt{\phi_s}\right) + \gamma_2\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s - V_{bx}}\right); \quad |V_{bs}| > |V_{bx}| \quad (5.5)$$

It is to be noted that V_{bs} and $V_{bx} < 0$ for *n*-channel MOSFETs (nMOSFETs) and >0 for *p*-channel MOSFETs (pMOSFETs). In Equations 5.4 and 5.5, the body effect coefficients γ_1 and γ_2 are given by

$$\gamma_1 = \frac{\sqrt{2qK_{si}\varepsilon_0 N_{CH}}}{C_{ox}} \quad \text{and} \quad \gamma_2 = \frac{\sqrt{2qK_{si}\varepsilon_0 N_{SUB}}}{C_{ox}} \quad (5.6)$$

Equations 5.4 and 5.5 are complex because these require knowledge of the shape of channel doping profile and the exact voltages to deplete different regions of the profile. Therefore, a unified expression for V_{th} is used to model the nonuniform vertical channel doping profile given by [27–29]

$$V_{th} = V_{TH0} + K_1\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right) - K_2 V_{bs} \quad (5.7)$$

where K_1 and K_2 are the parameters to model the vertically nonuniform channel doping profile and determined by fitting Equation 5.7 to the measured $I_{ds} - V_{gs}$ data for large geometry devices (e.g., W/L = 10 μm/10 μm) at low $V_{ds} \approx 50$ mV. The relation between K_1 and K_2 and γ_1 and γ_2 can be determined by solving Equations 5.5 and 5.7 at an intermediate bias $V_{bm} > V_{bx}$. Since Equations 5.5 and 5.7 represent the same V_{th} versus V_{bs} characteristics of a device, at a particular body bias, $V_{bs} = V_{bm}$, we must have the conditions [29]

$$V_{th}\left(\text{Equation 5.5}\right)\Big|_{V_{bs}=V_{bm}} = V_{th}\left(\text{Equation 5.7}\right)\Big|_{V_{bs}=V_{bm}}$$

$$\frac{d}{dV_{bs}}V_{th}\left(\text{Equation 5.5}\right)\Big|_{V_{bs}=V_{bm}} = \frac{d}{dV_{bs}}V_{th}\left(\text{Equation 5.7}\right)\Big|_{V_{bs}=V_{bm}}$$

Using the above conditions, we can show

$$\gamma_1\left(\sqrt{\phi_s - V_{bx}} - \sqrt{\phi_s}\right) + \gamma_2\left(\sqrt{\phi_s - V_{bm}} - \sqrt{\phi_s - V_{bx}}\right) = K_1\left(\sqrt{\phi_s - V_{bm}} - \sqrt{\phi_s}\right) - K_2 V_{bm} \tag{5.8}$$

$$-\frac{\gamma_2}{2\sqrt{\phi_s - V_{bm}}} = -\frac{K_1}{2\sqrt{\phi_s - V_{bm}}} - K_2 \tag{5.9}$$

Solving Equations 5.8 and 5.9 simultaneously, we can show that

$$\begin{aligned} K_1 &= \gamma_2 - \frac{2(\gamma_1 - \gamma_2)\left(\sqrt{\phi_s - V_{bx}} - \sqrt{\phi_s}\right)\left(\sqrt{\phi_s - V_{bm}}\right)}{2\sqrt{\phi_s}\left(\sqrt{\phi_s - V_{bm}} - \sqrt{\phi_s}\right) + V_{bm}}; \\ K_2 &= \frac{(\gamma_1 - \gamma_2)\left(\sqrt{\phi_s - V_{bx}} - \sqrt{\phi_s}\right)}{2\sqrt{\phi_s}\left(\sqrt{\phi_s - V_{bm}} - \sqrt{\phi_s}\right) + V_{bm}} \end{aligned} \tag{5.10}$$

If K_1 and K_2 are not given, they can be computed from Equation 5.10 using the channel doping concentration [27–29]. From Equation 5.6, we note that for a conventional high–low channel doping profile, $\gamma_1 > \gamma_2$; then, from Equation 5.10, the value of $K_2 > 0$. And, therefore, the devices with conventional high–low channel doping profile are sensitive to strong body bias. On the other hand, in the devices with low–high channel doping profile, $\gamma_1 < \gamma_2$; therefore, from Equation 5.10, $K_2 < 0$ and the devices are less sensitive to strong body bias.

5.2.1.2 Threshold Voltage Modeling for Nonuniform Lateral Channel Doping Profile

In advanced CMOS technologies, localized high doping concentration regions of the same doping type as the channel are used near the source and drain ends of the channel [3,4,23–26] to suppress SCEs [5,6]. This localized concentration of additional channel doping near the source and drain ends of the channel is referred to as the *halo* or *pocket* doping as shown in Figure 5.1a and is a critical technology parameter for optimizing the performance of MOSFET devices. Due to halo doping, the channel doping profile becomes laterally nonuniform with higher concentration at the source and drain ends and low concentration near the channel as shown in Figure 5.1b. This nonuniform lateral channel doping profile can also be modeled by two step functions as shown in Figure 5.4.

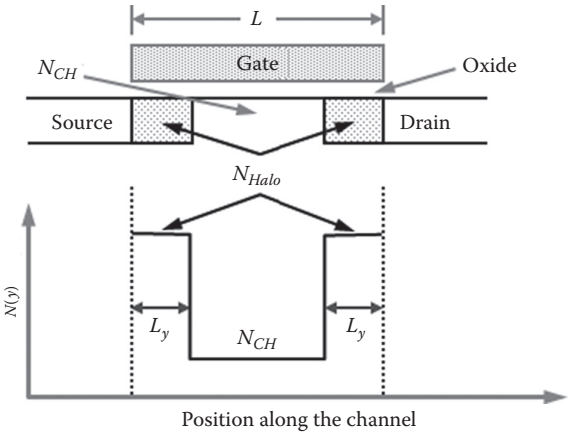

FIGURE 5.4
Nonuniform lateral channel doping profile due to halo/pocket implant approximated by two step profiles at the source and drain ends overlapping the gate; N_{Halo} and N_{CH} are the halo and channel doping concentrations of the same type of dopants, respectively; L_y is the width of the halo doping concentration inside the channel region.

In Figure 5.4, L is the channel length, L_y is the length of each halo region, N_{Halo} is the uniform concentration in each halo region L_y, and N_{CH} is the uniform concentration in the region $L - 2L_y$. If N_{eff} is the average channel doping concentration, then the channel charge per unit area is given by

$$Q = qN_{eff}L = \left[qN_{CH}L + q\left(N_{Halo} - N_{CH}\right)\left(2L_y\right)\right] \tag{5.11}$$

After simplifying Equation 5.11, we can show that the effective channel concentration due to lateral channel doping profile is given by

$$\begin{aligned} N_{eff} &= N_{CH}\left[1 + 2L_y\left(\frac{N_{Halo} - N_{CH}}{N_{CH}}\right)\frac{1}{L}\right] \\ &= N_{CH}\left[1 + \frac{L_{PE0}}{L}\right] \end{aligned} \tag{5.12}$$

where L_{PE0} is defined as

$$L_{PE0} = 2L_y\frac{N_{Halo} - N_{CH}}{N_{CH}} \tag{5.13}$$

L_{PE0} is a model parameter and is obtained by optimizing I–V data at $V_{bs} = 0$ for short channel and wide MOSFET devices. Similarly, we can show that the effective channel doping concentration with applied V_{bs} is

$$N_{eff}(V_{bs}) = N_{CH}\left[1 + \frac{L_{PEB}}{L}\right] \tag{5.14}$$

where:

L_{PEB} is a model parameter obtained by optimizing *I–V* data for different V_{bs} for short channel and wide MOSFET devices

Now, by introducing the nonuniform lateral channel doping, the expression for V_{th} at $V_{bs} = 0$ is given by

$$\begin{aligned} V'_{TH0} &= V_{fb} + \phi_s + \frac{\sqrt{2qK_{si}\varepsilon_0 N_{CH}\left(1+\frac{L_{PE0}}{L}\right)}}{C_{ox}}\sqrt{\phi_s} \\ &= V_{fb} + \phi_s + \gamma\left(\sqrt{1+\frac{L_{PE0}}{L}}\right)\sqrt{\phi_s} \\ &= V_{fb} + \phi_s + \gamma\sqrt{\phi_s} + \gamma\left(\sqrt{1+\frac{L_{PE0}}{L}}-1\right)\sqrt{\phi_s} \\ &= V_{TH0} + \gamma\left(\sqrt{1+\frac{L_{PE0}}{L}}-1\right)\sqrt{\phi_s} \end{aligned} \tag{5.15}$$

where we have added and subtracted $\gamma\sqrt{\phi_s}$ and used Equation 5.1 at $V_{bs} = 0$ along with $\gamma = \sqrt{2qK_{si}\varepsilon_0 N_{CH}}/C_{ox}$ to obtain Equation 5.15 for threshold voltage at $V_{bs} = 0$ due to the halo doping. Similarly, the expression for V_{th} at V_{bs} with halo doping can be shown as

$$V_{th} = V'_{TH0} + \gamma\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right).\left(\sqrt{1+\frac{L_{PEB}}{L}}\right) \tag{5.16}$$

Then combining Equations 5.15 and 5.16, we get the general expression for threshold voltage for modeling nonuniform lateral channel doping profile of MOSFETs as

$$V_{th} = V_{TH0} + \gamma\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right).\left(\sqrt{1+\frac{L_{PEB}}{L}}\right) + \gamma\left(\sqrt{1+\frac{L_{PE0}}{L}}-1\right)\sqrt{\phi_s} \tag{5.17}$$

Now, considering the vertically nonuniform channel doping profile and K_1 and K_2 parameters, we can write the combined expression for substrate doping effect on V_{th} as

$$V_{th} = V_{TH0} + K_1\left(\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}\right).\left(\sqrt{1+\frac{L_{PEB}}{L}}\right) + K_1\left(\sqrt{1+\frac{L_{PE0}}{L}}-1\right)\sqrt{\phi_s} - K_2 V_{bs} \tag{5.18}$$

Thus, the effect of nonuniform substrate doping introduces the set of compact model parameters $\{K_1, K_2, L_{PE0}, L_{PEB}\}$. In Equation 5.18,

- K_1 and K_2 model the effect of nonuniform vertical channel doping profile on V_{th};
- L_{PE0} and L_{PEB} model the nonuniform lateral channel doping profile on V_{th} at $V_{bs} = 0$ and $|V_{bs}| > 0$, respectively;
- At $V_{bs} = 0$, as L decreases, Equation 5.18 shows that V_{th} increases, showing reverse short channel effect due to halo-doping profile as shown in Figure 5.2b.

In long channel devices, halo/pocket implant causes a significant *drain-induced threshold voltage shift* (DITS) [30,31]. The applied V_{ds} reduces the drain barrier in the long channel MOSFET devices with halo implant. For large V_{ds}, the shift ΔV_{th}(DITS) due to DITS is given by [30]

$$\Delta V_{th}(\text{DITS}) \cong nv_{kT}.\ln\left(\frac{L}{L+\text{DVTP0}.\left(1+e^{-\text{DVTP1}.V_{ds}}\right)}\right) \tag{5.19}$$

where:
*DVTP*0 and *DVTP*1 are fitting parameters
nv_{kT} depends on subthreshold slope as discussed in Chapter 4

5.2.2 Small Geometry Effect on Threshold Voltage Model

The MOSFET threshold voltage is sensitive to both the channel length (Figure 5.2b) and the channel width [32]. Experimental data show that V_{th} decreases with the decrease of channel length called SCE whereas it increases with the decrease of channel width referred to as the narrow width effect (NWE). Therefore, it is critical to determine the shift in the long channel threshold voltage due to SCE and NWE and develop an expression for threshold voltage that accurately models the nanoscale device technology for circuit CAD. In Sections 5.2.2.1 and 5.2.2.2, we will develop mathematical expressions of the shift in the threshold voltage due to small geometry effects and present a threshold voltage expression to model all geometries in an advanced technology.

5.2.2.1 Threshold Voltage Model for Short Channel MOSFET Devices

For short channel devices, SCE or the decrease in V_{th} with the decreases in L is caused by the *bulk-charge sharing* between the gate and S/D *pn*-junctions as shown in Figure 5.5. Figure 5.5 shows that a significant amount of the bulk charge Q_b near the source and drain ends is controlled by reversed bias S/D *pn*-junctions. As a result, gate-induced Q_b decreases as channel length decreases (i.e., less V_{gs} is used to induce the same amount of Q_b). Since $Q_s = Q_b + Q_i$, for the same V_{gs}, Q_i increases as the devices are scaled down. Thus, less gate voltage is required to turn on the device, causing V_{th} decrease

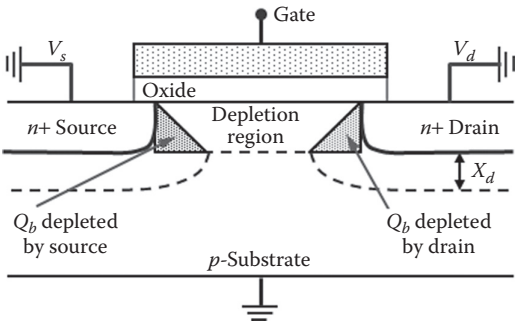

FIGURE 5.5
Short channel effect in MOSFETs caused by bulk-charge sharing by the gate and S/D *pn*-junctions; a significant part of channel depletion is caused by S/D regions; the source and drain each contributes an amount of channel charge Q_b to the total channel charge in silicon; X_d is the width of the S/D depletion region at zero bias condition.

as channel length decreases. The physics of SCE can be understood by a simple mathematical model based on charge sharing [32,33]. However, this model is not suitable for circuit CAD. Therefore, compact models are developed to calculate the shift in V_{th} due to SCE for circuit CAD.

Again, for a particular $V_{gs} > V_{th}$, as the drain voltage increases, the depletion region near the drain end of the channel gradually increases and extends toward the source end of the channel. As a result, the potential barrier to the inversion charge near the source end is reduced so that more carriers are injected from the source to the channel as V_{ds} increases. Thus, for a particular value of V_{gs} more inversion charges are injected as L decreases. This is referred to as the DIBL, causing V_{th} fall with the increase in V_{ds} as shown in Figure 5.6.

In order to model SCE, we solve Poisson's equation in the y direction along the channel. It can be shown that the shift in V_{th} due to SCE and DIBL is given by [34]

$$\Delta V_{th}(\text{SCE, DIBL}) = -\theta_{th}(L_{eff})\left[2(V_{bi} - \phi_s) + V_{ds}\right] \tag{5.20}$$

where

$$\theta_{th}(L_{eff}) = \frac{1}{2\left[\cosh\left(L_{eff}/2l_t\right) - 1\right]} \tag{5.21}$$

V_{bi} is the built-in potential of the S/D *pn*-junctions and is given by (Equation 2.84)

$$V_{bi} = v_{kT} \ln\left(\frac{N_{CH}N_{SD}}{n_i^2}\right) \tag{5.22}$$

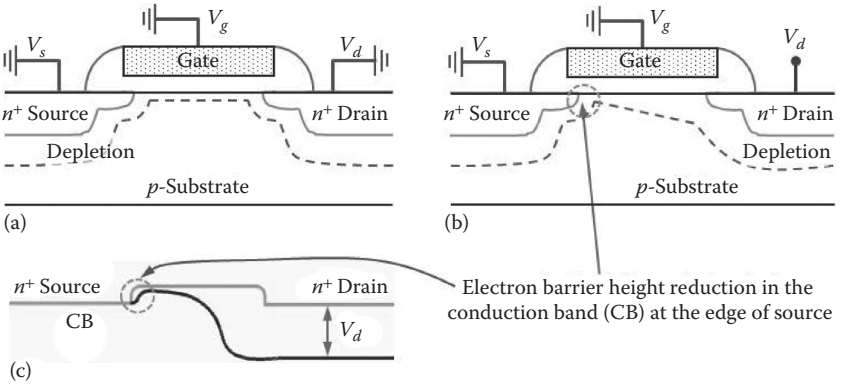

FIGURE 5.6
Short channel effect in MOSFETs due to drain voltage V_{ds}–DIBL in an n-channel device: (a) $V_{gs} = 0$ and $V_{ds} = 0$, (b) $V_{gs} = 0$ and V_{ds} = supply voltage, V_{dd}, and (c) plot of conduction bands along the length of the device under zero bias (top curve) and at drain bias conditions (bottom curve).

where N_{CH} and N_{SD} are the effective channel and S/D doping concentrations, respectively, and l_t represents the characteristic length given by

$$l_t = \sqrt{\frac{K_{si}T_{ox}W_d}{K_{ox}\eta}} \tag{5.23}$$

With depletion width $W_d = \sqrt{2K_{si}\varepsilon_0\left(\phi_s - V_{bs}\right)/qN_{CH}}$ and η (ETA) is a fitting parameter so that W_d/η = average width of the depletion region along the length of the channel.

Equation 5.20 shows that ΔV_{th} depends linearly on V_{ds} showing that V_{th} decreases as V_{ds} increases due to DIBL. In order to improve modeling flexibility for different technologies, different model parameters are introduced to get

$$\theta_{th}(\text{SCE}) = \frac{0.5\text{DVT0}}{\cosh\left(\text{DVT1}.L_{eff}/l_t\right) - 1} \tag{5.24}$$

$$\Delta V_{th}(\text{SCE}) = -\theta_{th}(\text{SCE})\left(V_{bi} - \phi_s\right) \tag{5.25}$$

$$l_t = \sqrt{\frac{K_{si}T_{ox}W_d}{K_{ox}}}\left(1 + \text{DVT2}.V_{bs}\right) \tag{5.26}$$

Similarly, the shift in threshold voltage due to DIBL is described by

$$\theta_{th}(\text{DIBL}) = \frac{0.5}{\cosh\left(\text{DSUB}.L_{eff}/l_{t0}\right) - 1} \tag{5.27}$$

$$\Delta V_{th}(\text{DIBL}) = -\theta_{th}(\text{DIBL}) \cdot (\text{ETA0} + \text{ETAB} \cdot V_{bs}) \cdot V_{ds} \tag{5.28}$$

where l_{t0} is the characteristics length without bias and is given by

$$l_{t0} = \sqrt{\frac{K_{si} T_{ox} W_{d0}}{K_{ox} \eta}} \tag{5.29}$$

with zero bias depletion width is given by (Equation 2.97), $W_{d0} = \sqrt{2K_{si}\varepsilon_0\phi_s / qN_{CH}}$.

In summary, the parameters used in BSIM4 (Berkeley Short Channel IGFET Model, version 4) compact MOS models for *SCE* and *DIBL* modeling are *DVT*0, *DVT*1, *DVT*2, *DSUB*, *ETA*0, and *ETAB*. Where, *DVT*2 and *ETAB* account for substrate bias effect on *SCE* and *DIBL*, respectively.

5.2.2.2 Threshold Voltage Modeling for Narrow Channel MOSFET Devices

In addition to channel length effect on V_{th}, narrow channel widths also affect V_{th}. These effects can be understood physically with reference to local oxidation of silicon (LOCOS) isolation process in CMOS technology as shown in Figure 5.7. LOCOS isolation process has been used prior

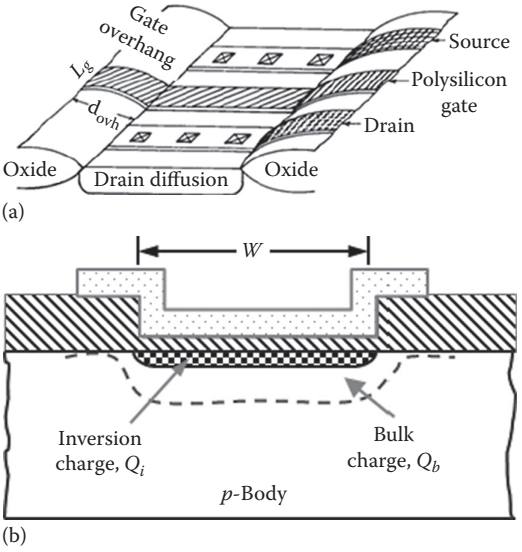

FIGURE 5.7
Narrow channel effect in MOSFETs: (a) gate overlap over isolation oxide and (b) additional bulk charge, Q_b controlled by gate bias due to gate overlap.

to shallow trench isolation (STI) techniques used in advanced CMOS technology.

Figure 5.7b shows two-dimensional (2D) cross-section of a MOSFET device along the channel width direction from the layout shown in Figure 5.7a. As shown in Figure 5.7b, the depletion layer does not abruptly change from deep to shallow at the edge of gate oxide. Therefore, there is a transition region and some spreading of field lines outside *W*. Thus, the gate charge Q_g supports some charge outside *W*. Since $Q_g = Q_b + Q_i$, for the same gate bias, Q_b is higher for narrow devices (i.e., gate is required to induce more Q_b out of the same Q_g), resulting in lower Q_i and consequently higher V_{th}.

For STI devices, the fringing field from the gate regions beyond the channel edges support channel depletion charges. This fringing field causes deeper depletion resulting in higher band bending and, therefore, an increase in the surface potential ϕ_s near the STI channel edge. The higher ϕ_s induces channel inversion near the STI at a lower V_{gs} than the rest of the channel. Thus, it takes lower effective V_{gs} to reach maximum channel depletion and the formation of inversion layer. Since the percentage contribution of the fringing field increases as the channel width *W* decreases, V_{th} tends to decrease with decreasing *W* in MOSFETs using STI technology (in contrast to LOCOS isolation technology), resulting in *inverse NWE*.

The physics of NWE can be understood by charge-sharing model similar to SCE [32]. However, these models are not suitable for compact modeling of billions of transistors in a VLSI circuit. Besides, NWE depends on the isolation technology. Therefore, universally accurate physical model is not available. For compact modeling, an empirical model can be developed, based on the observation of NWE from the experimental data. We know that V_{th} is directly proportional to gate oxide thickness T_{ox} and surface potential ϕ_s and experimentally it is found that V_{th} is inversely proportional to the channel width, *W*; therefore, for a long channel device, the shift in V_{th} due to NWE can be expressed as

$$\Delta V_{thW} \propto \frac{T_{ox}}{W_{eff}}\phi_s$$

or (5.30)

$$\Delta V_{thW} = K_3 \frac{T_{ox}}{W_{eff}}\phi_s$$

where:

K_3 is the constant of proportionality and is a *W*-dependent model parameter extracted from the measurement data

W_{eff} is the effective channel width

In order to model V_{bs} dependence of NWE, a model parameter K_{3B} can be used. Thus, Equation 5.30 can be expressed to include the effect of body bias as

$$\Delta V_{thW} = \left(K_3 + K_{3B}V_{bs}\right)\frac{T_{ox}}{W'_{eff} + W_0}\phi_s \tag{5.31}$$

Thus, three fitting parameters K_3, K_{3B}, and W_0 are required to model NWE. Here W'_{eff} is the effective channel width with the additional fitting parameter W_0 for accurate fitting of the measured data. Equation 5.31 models NWE in MOSFETs; however it does not model SCE of the narrow devices. In order to model SCE in narrow channel devices, we use Equations 5.24 and 5.25 to obtain the shift in V_{th} for narrow- and short channel devices as

$$\Delta V_{thWL} = \frac{0.5\text{DVT0}W}{\cosh\left(\text{DVT1}W.L_{eff}W_{eff}/l_{tw}\right)-1}\left(V_{bs} - \phi_s\right) \tag{5.32}$$

where:

$l_{tw} = \sqrt{K_{si}T_{ox}W_d/K_{ox}}\left(1+\text{DVT2}W.V_{bs}\right)$ is the characteristic length for short and narrow devices

Equation 5.32 models V_{th} shift in the short- and narrow channel devices whereas Equation 5.31 models that in narrow- and long channel devices. The final expression for V_{th} including nonuniform substrate concentration and small geometry effects is given by

$$\begin{aligned} V_{th} &= V_{th}(n\text{on} - \text{uniform substarate}) + \Delta V_{th}(\text{NWE}) + \Delta V_{th}(\text{NWE},\text{SCE}) \\ &\quad + \Delta V_{th}(\text{SCE}) + \Delta V_{th}(\text{DIBL}) + \Delta V_{th}(\text{DITS}) \end{aligned} \tag{5.33}$$

Thus, combining Equations 5.18, 5.19, 5.25, 5.28, 5.31, and 5.32, we can show the expression for V_{th} as used in BSIM4 model for circuit CAD.

$$\begin{aligned} V_{th} &= V_{TH0} + K_{1ox}\left(\sqrt{\phi_s - V_{bseff}} - \sqrt{\phi_s}\right).\left(\sqrt{1+\frac{L_{PEB}}{L_{eff}}}\right) \\ &\quad + K_{1ox}\left(\sqrt{1+\frac{L_{PE0}}{L_{eff}}} - 1\right)\sqrt{\phi_s} - K_{2ox}V_{bseff} \\ &\quad + \left(K3 + K3B.V_{bseff}\right)\frac{T_{ox}}{W'_{eff} + W_0}\phi_s \\ &\quad - \frac{1}{2}\left[\frac{\text{DVT0}W}{\cosh\left[\text{DVT1}W\left(L_{eff}.W_{eff}/l_{tw}\right)\right]-1} + \frac{\text{DVT0}}{\cosh\left[\text{DVT1}\left(L_{eff}/l_t\right)\right]-1}\right]\left(V_{bseff} - \phi_s\right) \\ &\quad - \frac{1}{2}\frac{\left(\text{EAT0} + \text{EATB.V}_{bseff}\right)}{\cosh\left(\text{DSUB}\left(L_{eff}/l_{t0}\right)\right)-1}.V_{ds} - nv_{kT}.\ln\left[\frac{L_{eff}}{L_{eff} + \text{DVTP0}\left(1+e^{-\text{DVTP1}.V_{ds}}\right)}\right] \end{aligned} \tag{5.34}$$

where we have used the effective channel length (L_{eff}) and effective channel width (W_{eff}) in Equation 5.34. Equation 5.34 shows the overall V_{th} expression for MOSFET devices to accurately model geometry and substrate doping dependence on device performance. In real CAD implementation the following modifications are made [28].

1. Electrical oxide thickness, TOXE dependence is introduced in model parameters $K1$ and $K2$ to improve scalability of V_{th} model over TOXE as

$$K_{1ox} = K1.\frac{\text{TOXE}}{\text{TOXM}}$$

and (5.35)

$$K_{2ox} = K2.\frac{\text{TOXE}}{\text{TOXM}}$$

   where:
   TOXM is the model parameter required to fit device characteristics
2. In order to set a lower bound for the body bias during circuit simulations to prevent occurrence of unreasonable values during iterations in CAD environment, V_{bs} is implemented as [28]

$$V_{bseff} = V_{bc} + \frac{1}{2}\cdot\left[\left(V_{bs} - V_{bc} - \delta_1\right) + \sqrt{\left(V_{bs} - V_{bc} - \delta_1\right)^2 - 4\delta_1 \cdot V_{bc}}\right] \quad (5.36)$$

   where $\delta_1 = 1$ mV and V_{bc} is the maximum allowable V_{bs} and found from $dV_{th}/dV_{bs} = 0$ to be

$$V_{bc} = 0.9\left(\phi_s - \frac{K1^2}{4K2^2}\right) \quad (5.37)$$

   For positive V_{bs}, there is need to set an upper bound for the body bias as [28]

$$V_{bseff} = 0.95\phi_s - \frac{1}{2}\cdot\left[\left(0.95\phi_s - V'_{bseff} - \delta_1\right) + \sqrt{\left(0.95\phi_s - V'_{bseff} - \delta_1\right)^2 - 4\delta_1 \cdot 0.95\phi_s}\right] \quad (5.38)$$

Effective Channel Length and Width: The effective channel length (L_{eff}) and width (W_{eff}) used in Equation 5.34 are given by

$$L_{eff} = L_{drawn} - 2\Delta L \quad (5.39)$$

$$W_{eff} = W_{drawn} - 2\Delta W \quad (5.40)$$

where ΔL and ΔW are model parameters that include S/D overlap under the gate and poly overlap along the width direction, respectively, and are given by

$$\Delta L = L_{INT} + \frac{L_L}{L^{L_{LN}}} + \frac{L_W}{L^{L_{WN}}} + \frac{L_{WL}}{L^{L_{LN}} L^{L_{WN}}} \tag{5.41}$$

$$\Delta W = W_{INT} + \mathrm{DWG}.V_{gsteff} + \mathrm{DWB}\left(\sqrt{\phi_s - V_{bseff}} - \sqrt{\phi_s}\right) + \frac{W_L}{L^{W_{LN}}} + \frac{W_W}{W^{W_{WN}}} + \frac{W_{WL}}{L^{W_{LN}} W^{W_{WN}}} \tag{5.42}$$

where L_{INT}, W_{INT}, DWG, and DWB are extracted from the measured data. Other parameters in Equations 5.41 and 5.42 are fitting parameters to improve the modeling accuracy (and rarely used). In Equation 5.42, V_{gsteff} is the effective value of $(V_{gs} - V_{th})$ used to ensure the channel charge continuity at the weak and strong inversion regions in the regional model. V_{gsteff} is obtained by equating channel charge at the weak inversion and at strong inversion at the transition point.

5.3 Drain Current Model

The total current density (J) in a MOSFET is the sum total of the *electron* and *hole* current densities J_n and J_p, respectively. And, the total J_n and J_p are the sum of the drift component of the respective carriers due to electric field E and diffusion component of the respective carriers due to the concentration gradient along the channel as discussed in Chapter 4 (Section 4.4) and is given by

$$\begin{aligned} J_n &= qn\mu_n E + qD_n \nabla n \\ J_p &= qp\mu_p E - qD_p \nabla p \end{aligned} \tag{5.43}$$

where:

q is the electronic charge
n and p are the electron and hole concentrations, respectively
∇n and ∇p are the electron and hole concentration gradient, respectively
μ_n and μ_p are the electron and hole surface mobility, respectively

The accuracy of MOSFET drain current model depends on the accuracy of inversion layer mobility model. Therefore, in the following section, we will derive the surface mobility model used in circuit CAD for small geometry MOSFETs.

5.3.1 Surface Mobility Model

In Chapter 4, we have assumed a constant surface mobility, μ_s for modeling MOSFET drain current, I_{ds}. This assumption is not valid under high electric field operation of the devices. As the vertical electric field E_x and lateral electric field E_y increase with increasing gate voltage V_{gs} and drain voltage

V_{ds}, respectively, the inversion carriers suffer increased scattering. Therefore, μ_s strongly depends on E_x and E_y. Let us consider the effect of E_x only on the surface mobility, that is, $V_{ds} \sim 0$. For the simplicity of I_{ds} modeling, let us define an *effective mobility* as the average mobility of carriers given by

$$\mu_{eff} = \frac{\int_0^{X_{inv}} \mu_s(x,y) \cdot n(x,y) dx}{\int_0^{X_{inv}} n(x,y) dx} \tag{5.44}$$

Using the definition of mobility from Equation 5.44 in Equation 4.64, we can write

$$I_{ds} = \frac{W}{L} \mu_{eff} \int_0^{V_{ds}} Q_i dV \tag{5.45}$$

In reality, μ_{eff} is highly reduced by large vertical electric field due to the high applied V_{gs}. The vertical electric field E_x pulls the inversion layer electrons in nMOSFETs toward the surface causing higher surface scattering as well as Coulomb scattering due to the interaction of electrons with oxide charges (Q_f, N_{it}) discussed in Chapter 2. Since the electric field varies vertically through the inversion layer, the average field in the inversion layer is given by

$$E_{eff} = \frac{E_{x1} + E_{x2}}{2} \tag{5.46}$$

where:

E_{x1} is the vertical electric field at the Si/SiO_2 interface

E_{x2} is the vertical electric field at the channel/depletion layer interface as shown in Figure 5.8

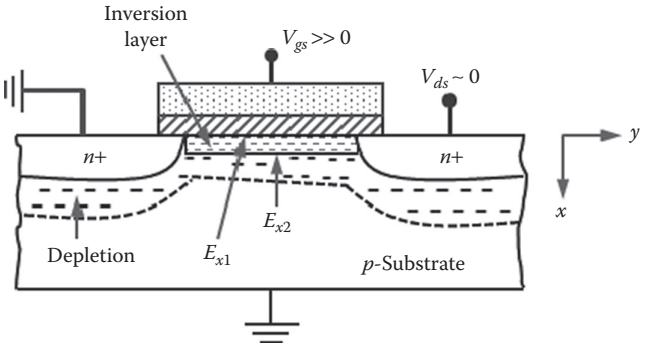

FIGURE 5.8
Effective vertical electric field on MOSFET inversion carriers due to the large applied gate bias V_{gs}: E_{x1} is the vertical electric field at the Si/SiO_2 interface and E_{x2} is the vertical electric field at the channel/depletion layer interface.

Now, from Gauss's law we can show that

$$E_{x1} - E_{x2} = \frac{Q_i}{K_{si}\varepsilon_0}$$

and (5.47)

$$E_{x2} = \frac{Q_b}{K_{si}\varepsilon_0}$$

where:

Q_i and Q_b are the inversion and bulk-charge densities, respectively, due to the applied V_{gs}

Substituting for E_{x1} and E_{x2} from Equation 5.47 into Equation 5.46, we can show

$$E_{eff} = \frac{1}{K_{si}\varepsilon_0}\left(\frac{1}{2}Q_i + Q_b\right) \tag{5.48}$$

In order to represent both electrons and holes, the general expression for the effective electric field is expressed as

$$E_{eff} = \frac{1}{K_{si}\varepsilon_0}\left(\eta Q_i + Q_b\right) \tag{5.49}$$

where:

the constant $\eta = 1/2$ for electrons and $\eta = 1/3$ for holes [35–37]

The measured μ_{eff} versus E_{eff} plots show a *universal behavior* independent of doping concentration at high effective vertical electrical fields and dependence on the substrate doping concentration and interface charge at low effective vertical electric fields as shown in Figure 5.9a.

The experimentally observed universal mobility behavior is due to the relative contributions of different scattering mechanisms [38,39] set by the strength of vertical electrical fields as shown in Figure 5.9b. As shown in Figure 5.9b, μ_{eff} is determined by Coulomb scattering of the ionized impurities and oxide charges, *phonon* scattering due to thermal vibration, and *surface roughness* scattering at the Si/SiO_2 interface. At high vertical electric fields, *surface roughness scattering* dominates as the carrier confinement is close to the interface, resulting in a decrease of μ_{eff} with the increase of E_{eff} as observed in Figure 5.9a.

The deviation from the *universal behavior* observed in Figure 5.9a, particularly in the heavily doped substrates at low effective electric fields, is due to the ionized impurity scattering, Coulomb scattering, and phonon scattering. At low effective vertical electric fields, Q_i is low and $<<Q_b$. As a result, the ionized impurity scattering and Coulomb scattering by ionized impurities and oxide charges become dominant scattering mechanisms in

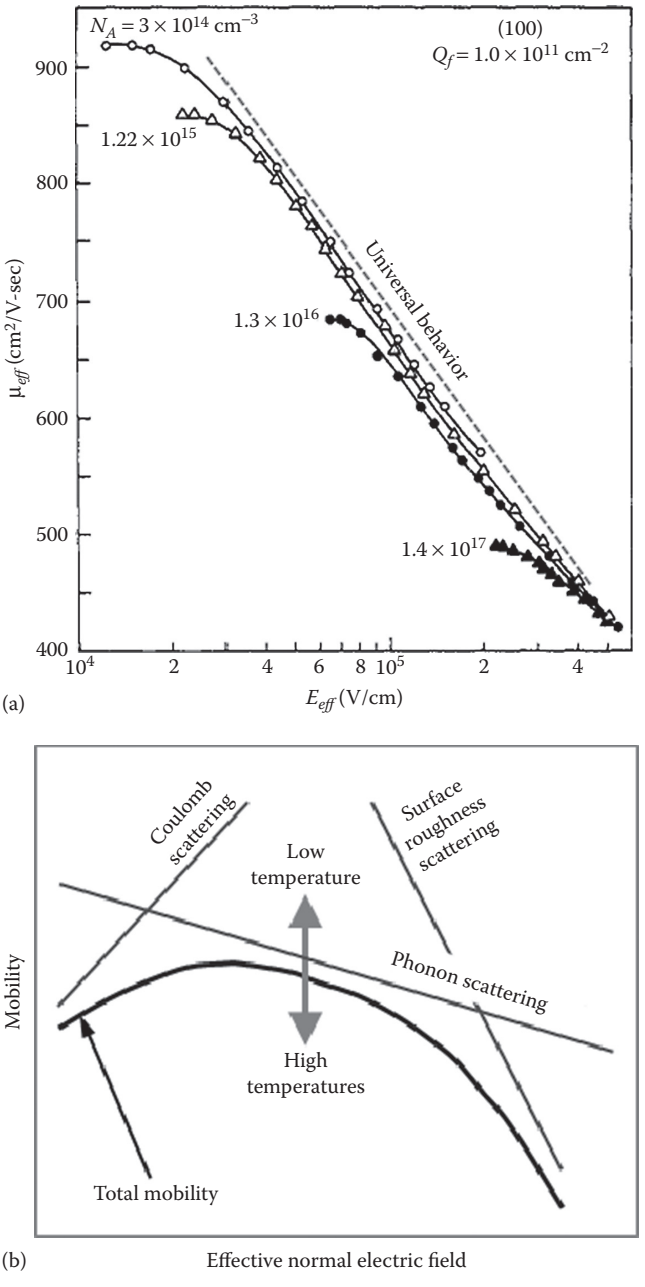

FIGURE 5.9
Low field mobility of inversion carriers in MOSFETs: (a) universal mobility behavior of inversion layer electrons in nMOSFET devices (Data from S.C. Sun and J.D. Plummer, *IEEE Trans. Electron Dev.*, 27, 1497–1508, 1980.) and (b) physical mechanisms showing the dependence of the inversion layer mobility on the effective vertical electric field.

the depletion region of a MOSFET device and μ_{eff} becomes strong function of channel doping concentration as observed experimentally. As the effective electric field increases, the phonon scattering due to lattice vibration becomes important. Thus, phonon scattering is weakly dependent on vertical electric fields and has the strongest temperature dependence on μ_{eff} as shown in Figure 5.9b.

The previous physical analysis describes the behavior of μ_{eff} versus E_{eff}. However, we need to develop an effective mobility model that can be used in drain current calculation to account for the vertical field effects on device performance. In order to develop μ_{eff} model for circuit CAD, we substitute the expressions for Q_b and Q_i for a MOSFET in Equation 5.48. For MOSFETs with threshold voltage V_{th} at strong inversion, the inversion charge is given by

$$Q_i = -C_{ox}\left(V_{gs} - V_{th}\right) \tag{5.50}$$

Again, we know,

$$V_{th} = V_{fb} + 2\phi_B + \frac{Q_s}{C_{ox}} \cong V_{fb} + 2\phi_B - \frac{Q_b}{C_{ox}} \tag{5.51}$$

where we have assumed that $Q_s \cong Q_b$. Therefore, from Equation 5.51 we get

$$Q_b = -C_{ox}\left(V_{th} - V_{fb} - 2\phi_B\right) \tag{5.52}$$

Now, substituting the expressions for Q_i and Q_b from Equations 5.50 and 5.51, respectively, in Equation 5.48, we get

$$\begin{aligned} E_{eff} &= \frac{C_{ox}}{K_{si}\varepsilon_0}\left[\frac{V_{gs} - V_{th}}{2} + V_{th} - V_{fb} - 2\phi_B\right] \\ &= \frac{K_{ox}\varepsilon_0}{T_{ox}} \cdot \frac{1}{2K_{si}\varepsilon_0}\left[V_{gs} + V_{th} - \left(2V_{fb} + 4\phi_B\right)\right] \cong \frac{1}{6T_{ox}}\left[V_{gs} + V_{th} - \left(2V_{fb} + 4\phi_B\right)\right] \end{aligned} \tag{5.53}$$

In the above expression, we have used $K_{ox}/K_{si} \cong 1/3$. Typically $\left(V_{gs} + V_{th}\right) >> \left(2V_{fb} + 4\phi_B\right)$; therefore, after simplification of Equation 5.53 we get

$$E_{eff} \cong \frac{V_{gs} + V_{th}}{6T_{ox}} \tag{5.54}$$

Now, we know that the unified formulation of effective mobility is given by the empirical relation [34,40,41]

$$\mu_{eff} = \frac{\mu_0}{\left[1 + \left(E_{eff}/E_0\right)\right]^{\nu}} \tag{5.55}$$

where:

μ_0 is concentration-dependent surface mobility
E_0 is the critical electric field
ν is a constant

Since the parameter $\nu << 1$, we can use Taylor's series expansion of the denominator and neglect the higher order terms to obtain

$$\left(1+\frac{E_{eff}}{E_0}\right)^{\nu} = 1+\nu\frac{E_{eff}}{E_0}+\frac{\nu(\nu-1)}{2!}\left(\frac{E_{eff}}{E_0}\right)^2+\cdots \tag{5.56}$$

Now substituting for E_{eff} from Equation 5.54 to the right-hand side of Equation 5.56, we get

$$\begin{aligned}\left(1+\frac{E_{eff}}{E_0}\right)^{\nu} &\cong 1+\frac{\nu}{E_0}\left(\frac{V_{gs}+V_{th}}{6T_{ox}}\right)+\frac{\nu(\nu-1)}{2E_0^2}\left(\frac{V_{gs}+V_{th}}{6T_{ox}}\right)^2+\cdots \\ &= 1+\frac{\nu}{6E_0}\left(\frac{V_{gs}+V_{th}}{T_{ox}}\right)+\frac{\nu(\nu-1)}{72E_0^2}\left(\frac{V_{gs}+V_{th}}{T_{ox}}\right)^2+\cdots \\ &= 1+U_a\left(\frac{V_{gs}+V_{th}}{T_{ox}}\right)+U_b\left(\frac{V_{gs}+V_{th}}{T_{ox}}\right)^2\end{aligned} \tag{5.57}$$

where we have defined $U_a \equiv \nu/6E_0$ and $U_b \equiv \nu(\nu-1)/72E_0^2$ as the model parameters to be extracted from the measured I_{ds} *versus* V_{gs} characteristics of MOSFET devices at low drain bias, V_{ds}. Therefore, combining Equations 5.55 and 5.57, the simplified low lateral field mobility model for MOSFET inversion carriers can be shown as [27,28]

$$\mu_{eff} = \frac{\mu_0}{1+U_a\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]+U_b\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]^2} \tag{5.58}$$

In order to improve the modeling accuracy at high body bias, a term U_cV_{bs} is introduced in the denominator of Equation 5.58 so that

$$\mu_{eff} = \frac{U_0}{1+\left(U_a+U_cV_{bs}\right)\cdot\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]+U_b\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]^2} \tag{5.59}$$

where:

$U_0 \equiv \mu_0$

The alternative expression to include the body bias dependence on μ_{eff} is

$$\mu_{eff} = \frac{U_0}{1+\left\{U_a\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]+U_b\left[\left(V_{gs}+V_{th}\right)/T_{ox}\right]^2\right\}\left(1+U_cV_{bs}\right)} \tag{5.60}$$

The mobility Equations 5.58 through 5.60 have been derived assuming strong inversion condition. In the strong inversion regime, the inversion carrier mobility is a function of gate bias. In the subthreshold region the accuracy of the mobility is not critical since Q_{inv} varies with V_{gs} and cannot be modeled accurately. Therefore, in subthreshold regime, the mobility is usually modeled as a constant concentration dependent mobility.

To ensure the continuity of the mobility model, BSIM mobility model is modified based on the V_{gsteff} expression to obtain the basic empirical models as [28]

$$\mu_{eff} = \frac{U_0}{1+\left(U_A + U_C V_{bseff}\right).\left[\left(V_{gsteff}+2V_{th}\right)/T_{OX}\right]+U_B\left[\left(V_{gsteff}+2V_{th}\right)/T_{OX}\right]^2} \tag{5.61}$$

or

$$\mu_{eff} = \frac{U_0}{1+\left[U_A \cdot\left[\left(V_{gsteff}+2V_{th}\right)/T_{OX}\right]+U_B\left[\left(V_{gsteff}+2V_{th}\right)/T_{OX}\right]^2\right]\left(1+U_C V_{bseff}\right)} \tag{5.62}$$

where:

V_{bseff} is the effective value of body bias to set the upper limit of computation as defined earlier

The BSIM4 model parameter set for the basic mobility model is {U_0, U_A, U_B, U_C} and is extracted from the $I_{ds} - V_{gs}$ characteristics at low V_{ds} with body bias. Different options of Equation 5.59 have been implemented in BSIM4 model and readers are encouraged to look at the users' manual to use the appropriate model and extract the appropriate model parameters for circuit CAD [28]. It can be observed from the earlier defined mobility models that μ_{eff} approaches a constant value of U_0 for $V_{gs} < V_{th}$ as used in the subthreshold regime.

The expression for V_{gsteff} is obtained by equating the channel charge of weak and strong inversions at the transition point for model continuity in the entire range of device operation and can be shown as [28]

$$V_{gsteff} = \frac{nv_{kT}\ln\left\{1+\exp\left[m^*\left(V_{gs}-V_{th}\right)/nv_{kT}\right]\right\}}{m^*+nC_{ox}\sqrt{2\phi_s/qK_{si}\varepsilon_0 N_{CH}}\exp\left\{-\left[\left(1-m^*\right)\left(V_{gs}-V_{th}-V_{off}\right)/2nv_{kT}\right]\right\}} \tag{5.63}$$

It should be pointed out that all of the mobility models given earlier account for only the influence of the vertical electrical field due to V_{gs} at low lateral electric field and often referred to as the *low-field mobility* model. The influence of the lateral electric field due to the applied V_{ds} on device performance is modeled in drain current by considering the velocity saturation in MOSFET devices under high lateral electric field.

5.3.2 Subthreshold Region Drain Current Model

The subthreshold current model is the same as derived for the long channel devices in Chapter 4 with minor change for improving the accuracy of data fitting and is given by [27,28]

$$I_{ds} = I_{s0} e^{\left(V_{gs} - V_{th} - V_{OFF}\right)/nv_{kT}} \left[1 - e^{-\left(V_{ds}/v_{kT}\right)}\right]; \quad V_{gs} < V_{th} \tag{5.64}$$

where V_{OFF} is the model parameter to account for the difference between V_{th} in the strong inversion and the subthreshold region and I_{s0} is given by (see Equation 4.118)

$$I_{s0} = \mu_s \left(W_{eff}/L_{eff}\right) C_d v_{kT}^2 \tag{5.65}$$

In Chapter 4 (Equation 4.127), we have shown that the subthreshold slope is given by

$$S = 2.3 n v_{kT} \tag{5.66}$$

where the ideality factor is given by

$$n = 1 + \frac{C_d}{C_{ox}} + \frac{C_{IT}}{C_{ox}} \tag{5.67}$$

In BSIM [27,28] compact models, a parameter called *NFACTOR* is introduced to ensure accurate calculation of C_d and is extracted from the measured data. Again, in short channel devices the surface potential in the channel is determined by both V_{gs} and V_{ds} through the coupling of C_{ox} and C_{dsc} as shown in Figure 5.10. The coupling capacitance $C_{dsc}(L)$ is an exponential function of L. Therefore, in BSIM4 the parameter n is modeled as

$$\begin{aligned} n = 1 &+ NFACTOR \frac{C_d}{C_{ox}} + \frac{C_{IT}}{C_{ox}} \\ &+ \frac{\left(C_{DSC} + C_{DSCD}.V_{ds} + C_{DSCB}.V_{bseff}\right)\left(0.5/\cosh(DVT1.L_{eff}/l_t) - 1\right)}{C_{ox}} \end{aligned} \tag{5.68}$$

where:

C_{DSC}, C_{DSCD}, and C_{DSCB} are the model parameters that describe the coupling between the channel and the drain

C_{DSCD} and C_{DSCB} represent the drain bias and body bias dependence of channel/drain coupling, respectively

5.3.3 Linear Region Drain Current Model

The high lateral electric field along the channel due to the applied V_{ds} significantly effects device performance. As we observe from Figure 5.11 that for electrons in silicon, the drift velocity v_d saturates near $E \sim 10^4$ V cm^{-1}.

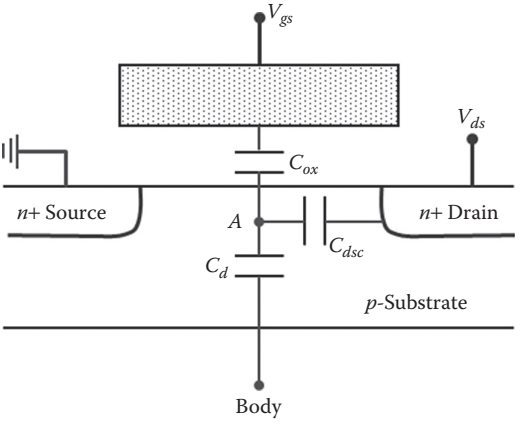

FIGURE 5.10
MOSFET device showing gate capacitance C_{ox}, bulk capacitance C_d, and source drain to channel coupling capacitances C_{dsc}; all the capacitances have an effect on the channel potential and subthreshold conduction.

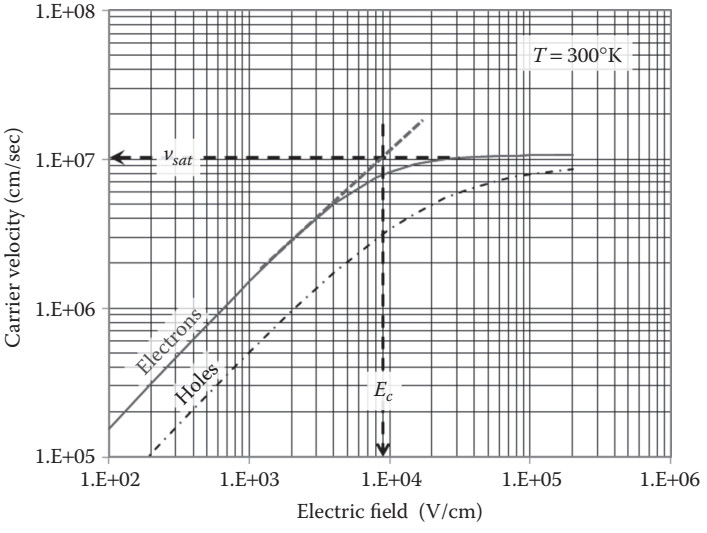

FIGURE 5.11
Drift velocity versus electric field showing carrier velocity saturation in silicon at an electric field near 1×10^4 V cm^{-1}.

As a result, the relation $v_d = \mu E$ does not hold at high electric field. Since average electric field for short channel devices $> 10^4$ V cm^{-1}, small geometry MOSFET devices will operate at $v_d = v_{sat} \cong 1 \times 10^7$ cm sec^{-1}, that is, the saturation velocity of electrons.

We discussed earlier that the mobility is not a constant at high electric field; therefore, we must account for the high lateral electric field effects in

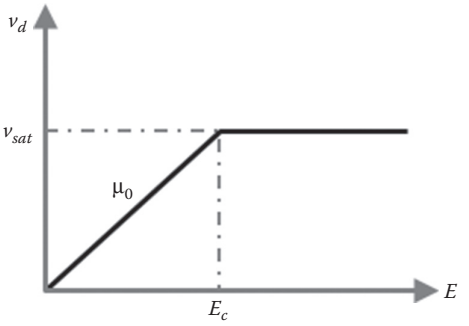

FIGURE 5.12
Drift velocity, v_d versus lateral electrical field, E; piecewise linear mobility behavior of inversion layer electrons due to high E along the channel of MOSFETs; v_{sat}, μ_0, and E_c are the saturation velocity of inversion carriers, concentration-dependent mobility of inversion carriers, and critical electric field at which carrier velocity saturates, respectively.

the expression for I_{ds} derived from simple theory (Chapter 4). Thus, at high electric field along the channel, MOSFET devices operate at a drift velocity, $v_d = v_{sat}$ [11]. Then with reference to Figure 5.11, we assume a v_d versus E piecewise linear model for I–V modeling as shown in Figure 5.12. Thus, at a particular lateral electric field, E_y, we can write [11]

$$v_d = \begin{cases} \dfrac{\mu_{eff} E_y}{1+\left(E_y/E_c\right)}, & \left(E_y < E_c\right) \\ v_{sat}, & \left(E_y > E_c\right) \end{cases} \tag{5.69}$$

As shown in Figure 5.12, we assume that v_d saturates abruptly at a critical lateral electric field E_c along the channel.

In Figure 5.12, E_c is the field at which carriers are velocity saturated, that is, at $E_y = E_c$, $v_d = v_{sat}$. Then from Equation 5.69 we can show [11]

$$v_{sat} = \frac{\mu_{eff} E_c}{2} \quad \text{or} \quad E_c = \frac{2v_{sat}}{\mu_{eff}} \tag{5.70}$$

We will use Equation 5.69 to derive linear region drain current expression to account for the high lateral field along the channel due to V_{ds}.

Now, we know that the current density at any point y along the channel in the y direction of an nMOSFET is given by $J_n(y) = nqv(y) = Q_i v(y)$, where n, q, and $v(y)$ are the inversion carrier density, electronic charge, and drift velocity of inversion layer electrons, respectively; $Q_i = nq$ is the inversion carrier charge per unit area. Using the expression for Q_i from Chapter 4,

Equation 4.95, we can write the general expression for drain current in the linear regime as

$$I_{ds} = I(y) = W_{eff} C_{ox} \left[V_{gs} - V_{th} - A_{bulk} V(y) \right] v(y) \tag{5.71}$$

where:
$V(y)$ = potential difference between the drain and channel at y
$v(y)$ is the carrier velocity at any point y in the channel
A_{bulk} is the body effect coefficient, α (Equation 4.96)

Then substituting Equation 5.69 in Equation 5.71, we get for $E_y < E_c$

$$I_{ds} = W_{eff} C_{ox} \left[V_{gs} - V_{th} - A_{bulk} V(y) \right] \frac{\mu_{eff} E(y)}{1 + \left[E(y)/E_c \right]} \tag{5.72}$$

After simplification, we can show from (5.72),

$$E(y) = \frac{I_{ds}}{W_{eff} \mu_{eff} C_{ox} \left[V_{gs} - V_{th} - A_{bulk} V(y) \right] - \left(I_{ds}/E_c \right)} = \frac{dV(y)}{dy}$$

or (5.73)

$$I_{ds} dy = \left(W_{eff} \mu_{eff} C_{ox} \left[V_{gs} - V_{th} - A_{bulk} V(y) \right] - \frac{I_{ds}}{E_c} \right) dV(y)$$

Integrating Equation 5.73 from ($y = 0$, $V(y) = 0$) to ($y = L_{eff}$, $V(y) = V_{ds}$) and after simplification, we get the linear region ($V_{ds} < V_{dsat}$) current as

$$I_{ds} = \frac{W_{eff}}{L_{eff} \left[1 + \left(V_{ds}/L_{eff} E_c \right) \right]} \mu_{eff} C_{ox} \left(V_{gs} - V_{th} - \frac{1}{2} A_{bulk} V_{ds} \right) V_{ds} \tag{5.74}$$

From Equation 5.74 note that the effect of high lateral electric field is the apparent increase in L_{eff} for higher V_{ds}, thus decreasing the linear current. Also, note that Equation 5.74 is valid when parasitic S/D series resistance, $R_{ds} = 0$. For $R_{ds} > 0$, the drain current is modified as [28]

$$I_{ds} = \frac{I_{ds0}}{1 + \left(R_{ds} I_{ds0} / V_{ds} \right)} \tag{5.75}$$

where:
I_{ds0} is the drain current at $R_{ds} = 0$ and is given by Equation 5.74

5.3.4 Saturation Region Drain Current Model

Let us assume that V_{dsat} is the drain saturation voltage at which the inversion carriers attain saturation velocity v_{sat}, that is, at $E_y = E_c$. Using the condition,

$v(y) = v_{sat}$ at $E_y = E_c$, in Equation 5.71, we get the saturation region ($V_{ds} \geq V_{dsat}$) drain current as

$$I_{ds} = W_{eff} C_{ox} \left(V_{gs} - V_{th} - A_{bulk} V_{dsat} \right) v_{sat} \tag{5.76}$$

Using the expression for v_{sat} from Equation 5.70, we get from Equation 5.76 an alternate expression for drain current in the saturation region of MOSFETs as

$$I_{ds} = \frac{1}{2} W_{eff} \mu_{eff} C_{ox} \left(V_{gs} - V_{th} - A_{bulk} V_{dsat} \right) E_c \tag{5.77}$$

Again, Equations 5.76 and 5.77 are valid when $R_{ds} = 0$ and must be modified for $R_{ds} > 0$.

In order to derive the expression for saturation drain voltage V_{dsat}, we recognize that I_{ds} given by Equations 5.74 and 5.77 must be continuous at $V_{ds} = V_{dsat}$; therefore, equating Equation 5.74 to Equation 5.77, we get

$$\frac{W_{eff}}{L_{eff}\left[1 + \left(V_{dsat}/L_{eff}E_c\right)\right]} \mu_{eff} C_{ox} \left(V_{gs} - V_{th} - \frac{1}{2} A_{bulk} V_{dsat} \right)$$

$$= \frac{1}{2} W_{eff} \mu_{eff} C_{ox} \left(V_{gs} - V_{th} - A_{bulk} V_{dsat} \right) E_c \tag{5.78}$$

or

$$\frac{2}{E_c L_{eff} + V_{dsat}} \left(V_{gs} - V_{th} - \frac{1}{2} A_{bulk} V_{dsat} \right) V_{dsat} = \left(V_{gs} - V_{th} - A_{bulk} V_{dsat} \right)$$

After simplification of Equation 5.78, we can show

$$V_{dsat} = \frac{E_c L_{eff} \left(V_{gs} - V_{th} \right)}{A_{bulk} E_c L_{eff} + \left(V_{gs} - V_{th} \right)} \tag{5.79}$$

For $R_{ds} > 0$, V_{dsat} is higher than that given by Equation 5.79 and can be calculated from Equations 5.75 and 5.77 with two model parameters, $A1$ and $A2$, to account for the nonsaturating effect of I–V characteristics [28,41].

The I_{dsat} model described in Equations 5.76 and 5.77 must be corrected for output resistance, R_{out}, due to (1) CLM, (2) DIBL, and (3) substrate current–induced body effect (SCBE).

5.3.5 Bulk-Charge Effect

When V_{ds} is large and/or when the channel length is long, the depletion region thickness of the channel is not uniform along the channel length. This will cause V_{th} to vary along the channel. This effect is called the

bulk-charge effect as discussed in Chapter 4, defining the parameter called, α (Equation 4.96). In BSIM4, the parameter A_{bulk} is used to model the bulk-charge effect including both short channel effects and narrow channel effects and is given by

$$A_{bulk} = \left\{1 + F\_doping \cdot \left[\frac{A0 \cdot L_{eff}}{L_{eff} + 2\sqrt{XJ \cdot X_{dep}}} \cdot \left[1 - AGS \cdot V_{gsteff}\left(\frac{L_{eff}}{L_{eff} + 2\sqrt{XJ \cdot X_{dep}}}\right)^2\right] + \frac{B0}{W'_{eff} + B1}\right]\right\} \cdot \frac{1}{1 + KETA \cdot V_{bseff}} \quad (5.80)$$

where, *F_doping* models nonuniform doping profiles and is given by

$$F\_doping = \frac{\sqrt{1 + (LPEB/L_{eff})}K_{1ox}}{2\sqrt{\phi_s - V_{bseff}}} + K_{1ox} - K3B\frac{TOXE}{W'_{eff} + W_0} \quad (5.81)$$

where:

K_{1ox} and K_{2ox} are defined in Equation 5.35

In Equation 5.80, the model parameters introduced to characterize A_{bulk} are A0, AGS, B0, B1, and KETA. These parameters are extracted from the measured *I–V* data. It is found that the value of A_{bulk} increases with the increase in *L* and approaches 1 for shorter devices. This is due to the fact that for short channel devices, the depletion width is almost uniform from source to drain, whereas for long channel devices the depletion near the drain end is much wider than that near the source end of the channel.

5.3.6 Output Resistance

I_{ds}–V_{ds} plot along with the output resistance (R_{out}), which is reciprocal of its first-order derivative, is shown in Figure 5.13 [27,28]. As shown in Figure 5.13, the behavior of R_{out} is characterized by four separate regions based on different physical mechanisms. These regions are (1) triode or linear, (2) CLM, (3) DIBL, and (4) SCBE. Three mechanisms CLM, DIBL, and SCBE affect R_{out} in the saturation region; however, each of them dominates in one of the three distinct regions as shown in Figure 5.13.

We know that I_{ds} depends on both V_{gs} and V_{ds}, and from Figure 5.13, we find that I_{ds} is weakly dependent on V_{ds} in the saturation region (CLM and DIBL).

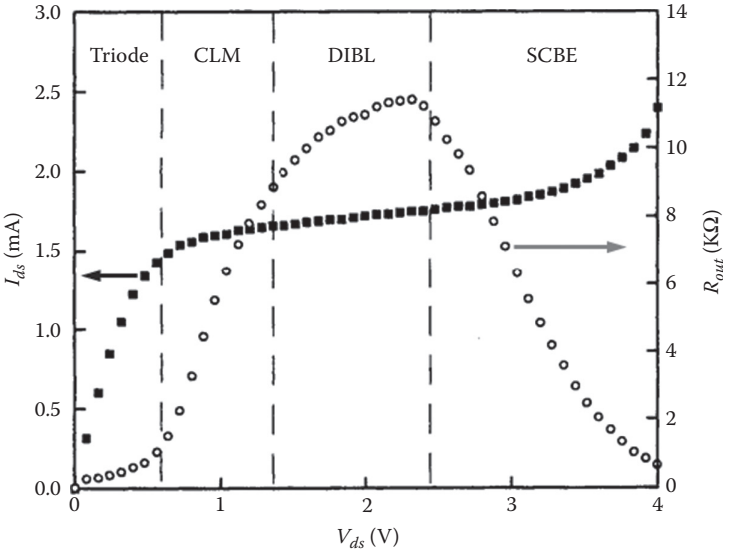

FIGURE 5.13
MOSFET output characteristics: drain current, I_{ds}, and output resistance, R_{out}, of an nMOSFET device divided into different operating regions based on different physical mechanisms. (Reproduced with permission from J.H. Huang et al., *International Electron Devices Meeting 1992*, Technical Digest, pp. 569–572, IEEE, 1992. Copyright 1992 IEEE.)

Since the saturation region I_{ds} depends weakly on V_{ds}, we can use Taylor series expansion of I_{ds} @ $V_{ds} = V_{dsat}$ and neglect the higher order terms to get

$$\begin{aligned} I_{ds}\left(V_{gs}, V_{ds}\right) &= I_{ds}\left(V_{gs}, V_{dsat}\right) + \frac{dI_{ds}\left(V_{gs}, V_{ds}\right)}{dV_{ds}}\left(V_{ds} - V_{dsat}\right) \\ &= I_{dsat}\left(1 + \frac{1}{I_{dsat}}\frac{dI_{ds}\left(V_{gs}, V_{ds}\right)}{dV_{ds}}\left(V_{ds} - V_{dsat}\right)\right) \\ &= I_{dsat}\left(1 + \frac{V_{ds} - V_{dsat}}{V_A}\right) \end{aligned} \tag{5.82}$$

where I_{dsat} and V_A are given by

$$\begin{aligned} I_{dsat} &= I_{ds}\left(V_{gs}, V_{dsat}\right) \\ V_A &= I_{dsat}\left(\frac{dI_{ds}}{dV_{ds}}\right)^{-1} \end{aligned} \tag{5.83}$$

In Equation 5.83, the expression for I_{dsat} is given by Equations 5.76 and 5.77. In Equation 5.82, V_A is called the *early voltage* (following the original term

used in describing bipolar junction transistor output resistance) and is introduced for the analysis of the output resistance of MOSFET devices in the saturation region. In order to model V_A, we have to consider the contributions of CLM, DIBL, and SCBE components on output resistance as described next:

The early voltage due to CLM is given by

$$V_{ACLM} = I_{dsat}\left(\frac{dI_{ds}}{dL}.\frac{dL}{dV_{ds}}\right)^{-1} = C_{clm}.\left(V_{ds} - V_{dsat}\right) \tag{5.84}$$

The early voltage due to DIBL is given by

$$V_{ADIBL} = I_{dsat}\left(\frac{dI_{ds}}{dV_{th}}.\frac{dV_{th}}{dV_{ds}}\right)^{-1} \tag{5.85}$$

The early voltage due to SCBE is caused by the reduction of V_{th} due to the substrate current induced forward biasing of the source/channel *pn*-junction (as discussed in Section 5.4). Therefore, the early voltage due to SCBE can be defined as

$$V_{ASCBE} = I_{dsat}\left(\frac{dI_{ds}}{dV_{ds}}\right)^{-1} = \frac{L_{eff}}{PSCBE2}\exp\left(\frac{PSCBE1.l_c}{V_{ds} - V_{dsat}}\right) \tag{5.86}$$

where:

*PSCBE*1 and *PSCBE*2 are model parameters extracted from $I_{ds}-V_{ds}$ plots in the saturation region

l_c is the characteristic length of the impact ionization region at the drain-end of MOSFETs

In addition, for long channel devices with halo implant we have to consider the component of early voltage, V_{ADITS} due to *DITS*.

5.3.7 Unified Drain Current Equation

In the regional modeling approach, separate model expressions for each region of MOSFET device operation such as subthreshold and strong inversion as well as the linear and saturation regions are developed. Although these expressions can accurately describe device behavior within their own respective region of operation, problems are likely to occur in the transition region between two well-described regions. In order to address this persistent problem, a unified model should be synthesized to preserve the region-specific accuracy and to ensure continuity of current and conductance and their derivatives in all transition regions.

In order to ensure this continuity, a unified current expression based on continuous channel charge and mobility is used in BSIM4 model. Thus, a single *I*–*V* equation is obtained and is given by [28]

$$I_{ds} = \frac{I_{ds0}}{1+\left(R_{ds}I_{ds0}/V_{dseff}\right)}\left[1+\frac{1}{C_{lm}}\ln\left(\frac{V_A}{V_{Asat}}\right)\right].\left(1+\frac{V_{ds}-V_{dseff}}{V_{ADIBL}}\right). \quad (5.87)$$
$$\left(1+\frac{V_{ds}-V_{dseff}}{V_{ASCBE}}\right).\left(1+\frac{V_{ds}-V_{dseff}}{V_{ADITS}}\right)$$

where V_{Asat} = early voltage @ $V_{ds} = V_{dsat}$; $V_A = V_{Asat} + V_{ACLM}$; and I_{ds0} is given by

$$I_{ds0} = \frac{W_{eff}}{L_{eff}\left[1+\left(V_{dseff}/E_cL_{eff}\right)\right]}C_{ox}\mu_{eff}V_{gsteff}V_{dseff}\left(1-\frac{V_{dseff}}{2V_{bs}}\right) \quad (5.88)$$

Also, an effective drain voltage, V_{dseff}, is a function that guarantees continuity of I_{ds} and its derivatives at V_{dsat} with a user defined parameter δ (DELTA) and is given by

$$V_{dseff} = V_{dsat} - \frac{1}{2}\left[V_{dsat} - V_{ds} - \delta + \sqrt{\left(V_{dsat}-V_{ds}-\delta\right)^2 - 4\delta.V_{dsat}}\right] \quad (5.89)$$

V_{dseff} along with the optimized value of δ ensures continuity of *I–V* plot and its derivatives from linear to saturation regimes. It is shown that the unified Equation 5.87 addresses the continuity from the subthreshold to linear region also by the introduction of the parameter V_{gsteff} given in Equation 5.63.

5.3.8 S/D Parasitic Series Resistance

The S/D parasitic series resistance, R_{ds}, of advanced MOSFET devices is modeled as

$$R_{ds} = \frac{R_{DSW} + \left[1 + P_{RWG}V_{gsteff} + P_{RWB}\left(\sqrt{\phi_s - V_{bseff}} - \sqrt{\phi_s}\right)\right]}{\left(10^6 W'_{eff}\right)^{WR}} \quad (5.90)$$

where:

R_{DSW}, P_{RWG}, P_{RWB}, and WR are model parameters
P_{RWG} and P_{RWB} are gate- and body bias–dependent parameters
WR is empirical fitting parameters to improve the accuracy of the model

5.3.9 Polysilicon Gate Depletion

When a gate voltage is applied to a heavily doped polysilicon gate, for example, nMOSFETs with *n*+ polysilicon (poly-Si) gate, a thin depletion layer in the poly-Si can be formed at the interface between the poly-Si and the gate oxide. This depletion layer is very thin because of the high doping concentration in the poly-Si gate. However, its effect cannot be ignored for devices with gate oxides thinner than 10 nm [28].

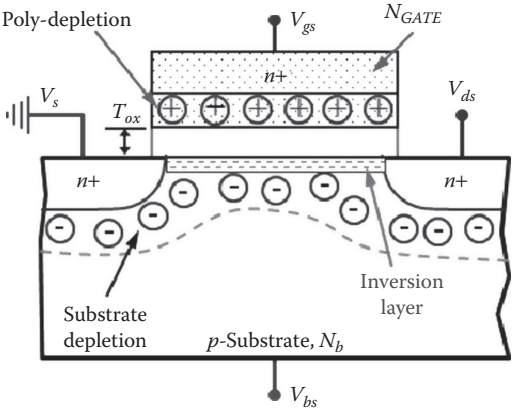

FIGURE 5.14
Charge distribution in an nMOSFET device due to polysilicon gate depletion effect as the device operates in the strong inversion region.

Figure 5.14 shows an nMOSFET device with the depletion region in the n+ poly-Si gate. The doping concentration in the poly-Si gate is N_{GATE} and the doping concentration in the substrate is N_{SUB}. The gate oxide thickness is T_{ox}. If we assume that the doping concentration in the gate is infinite, then no depletion region will exist in the gate, and there would be no one sheet of positive charge at the interface between the poly-Si gate and gate oxide. In reality, the doping concentration is finite. The positive charge near the interface of the poly-Si gate and the gate oxide is distributed over a finite depletion region with thickness X_p. The depletion width in the substrate is X_d. In the presence of the depletion region, the voltage drop across the gate oxide and the substrate will be reduced, because part of the gate voltage will be dropped across the depletion region in the gate. That means the effective gate voltage will be reduced.

Let us assume that the potential drop in the depletion layer X_p in the polysilicon gate is ϕ_p; following the procedure discussed in Section 3.4.2.1 [Equation 3.62], we can show

$$\phi_p = \frac{qN_{GATE}}{2K_{si}\varepsilon_0} X_p^2 \tag{5.91}$$

where:

N_{GATE} is the effective concentration in the poly-depletion region

If E_p is the electric field at the poly-Si/SiO_2 interface, then the depletion charge in the poly is given by (Equation 3.64)

$$Q_{GATE} = \sqrt{2qK_{si}\varepsilon_0 N_{GATE}\phi_p} \tag{5.92}$$

Again, from Gauss's law we get $K_{ox}\varepsilon_0 E_{ox} = Q_{GATE}$; therefore, from Equation 5.92 we get

$$E_{ox} = \frac{1}{K_{ox}\varepsilon_0}\sqrt{2qK_{si}\varepsilon_0 N_{GATE}\phi_p} \tag{5.93}$$

Now, the applied gate voltage with additional voltage drop in the poly-depletion region is given by

$$V_{gs} = V_{fb} + \phi_s + \phi_p + V_{ox} \tag{5.94}$$

Since $V_{ox} = E_{ox}T_{ox}$, we can simplify Equation 5.94 using Equation 5.93 as

$$V_{gs} = V_{fb} + \phi_s + \phi_p + \frac{T_{ox}}{K_{ox}\varepsilon_0}\sqrt{2qK_{si}\varepsilon_0 N_{GATE}\phi_p} \tag{5.95}$$

After simplification we can show from Equation 5.95

$$a\left(V_{gs} - V_{fb} - \phi_s - \phi_p\right)^2 - \phi_p = 0 \tag{5.96}$$

where we defined

$$a = \frac{K_{ox}^2\varepsilon_0^2}{2qK_{si}\varepsilon_0 N_{GATE}T_{ox}^2} \tag{5.97}$$

Now let us define that the effective gate voltage due to additional voltage drop in the poly is given by $V_{gseff} = (V_{gs} - \phi_p)$; then rearranging Equation 5.96 we get

$$a\left[\left(V_{gs} - \phi_p\right) - \left(V_{fb} + \phi_s\right)\right]^2 + \left(V_{gs} - \phi_p\right) - V_{gs} = 0$$

or (5.98)

$$a\left[V_{gseff} - \left(V_{fb} + \phi_s\right)\right]^2 + V_{gseff} - V_{gs} = 0$$

After simplification of Equation 5.98, we can show

$$aV_{gseff}^2 - \left[2a\left(V_{fb} + \phi_s\right) - 1\right]V_{gseff} + \left[a\left(V_{fb} + \phi_s\right)^2 - V_{gs}\right] = 0 \tag{5.99}$$

Now, we solve the quadratic Equation 5.99 on V_{gseff} due to poly gate depletion. Solving V_{gseff} we get

$$V_{gseff} = \left(V_{fb} + \phi_s\right) - \frac{1}{2a} \pm \frac{1}{2a}\sqrt{\left(2a\left(V_{fb} + \phi_s\right) - 1\right)^2 - 4a^2\left(V_{fb} + \phi_s\right)^2 + 4aV_{gs}} \tag{5.100}$$

Since $(V_{gs} - \phi_p) > 0$, we consider the positive sign of Equation 5.100, to get

$$
\begin{aligned}
V_{gseff} &= \left(V_{fb}+\phi_s\right)-\frac{1}{2a}+\frac{1}{2a}\sqrt{\left(2a\left(V_{fb}+\phi_s\right)-1\right)^2-4a^2\left(V_{fb}+\phi_s\right)^2+4aV_{gs}} \\
&= V_{fb}+\phi_s-\frac{1}{2a} \\
&\quad+\frac{1}{2a}\sqrt{4a^2\left(V_{fb}+\phi_s\right)^2-4a\left(V_{fb}+\phi_s\right)+1-4a^2\left(V_{fb}+\phi_s\right)^2+4aV_{gs}} \\
&= V_{fb}+\phi_s-\frac{1}{2a}+\frac{1}{2a}\sqrt{1-4a\left(V_{fb}+\phi_s\right)+4aV_{gs}} \\
&= V_{fb}+\phi_s+\frac{1}{2a}\left(\sqrt{1+4a\left(V_{gs}-V_{fb}-\phi_s\right)}-1\right)
\end{aligned}
\tag{5.101}
$$

Now, substituting the expression for a from Equation 5.97 in Equation 5.101, we can show

$$
V_{gseff} = V_{fb}+\phi_s+\frac{qK_{si}\varepsilon_0 N_{GATE}T_{ox}^2}{K_{ox}^2\varepsilon_0^2}\left(\sqrt{1+\frac{2K_{ox}^2\varepsilon_0^2\left(V_{gs}-V_{fb}-\phi_s\right)}{qK_{si}\varepsilon_0 N_{GATE}T_{ox}^2}}-1\right) \tag{5.102}
$$

For metal gate $K_{si} = 0$; therefore, Equation 5.91 shows that there are no gate depletion and $V_{gs} = V_{gseff}$.

Due to polysilicon gate depletion, the effective gate voltage can be reduced by about 10%. We can estimate the drain current reduction in the linear region as a function of V_{gs}. Assume that V_{ds} is very small (e.g., 50 mV). The linear drain current is proportional to $C_{ox}(V_{gs} - V_{th})$. The ratio of the linear drain current with and without polysilicon gate depletion is equal to

$$
\frac{I_{ds}\left(V_{gseff}\right)}{I_{ds}\left(V_{gs}\right)} \cong \frac{V_{gseff}-V_{th}}{V_{gs}-V_{th}} \tag{5.103}
$$

Since $V_{gs} > V_{gseff}$, Equation 5.103 shows that $I_{ds}(V_{gseff})$ is reduced due to polysilicon depletion effect. A significant capacitance reduction has been observed in MOSFETs with oxide thickness less than 5 nm. Thus, the polysilicon depletion effect has to be accounted for in modeling the capacitance characteristics of devices with very thin oxide thickness.

5.3.10 Temperature Dependence

The temperature dependence of the major BSIM model parameters are briefly described next with reference to the reference temperature T_{NOM}.

At any temperature T, the temperature dependence of threshold voltage is modeled by

$$V_{th}(T) = V_{th}\left(T_{NOM}\right) + \left(KT_1 + \frac{KT_{1L}}{L_{eff}} + KT_2 V_{bseff}\right)\left(\frac{T}{T_{NOM}} - 1\right) \tag{5.104}$$

where:

KT_1, KT_{1L}, and KT_2 are the model parameters to characterize the temperature dependence of threshold voltage for different channel lengths and body biases

The temperature dependence of carrier mobility is given by

$$\begin{aligned} U_0(T) &= U_0\left(\frac{T}{T_{NOM}}\right)^{UTE} \\ U_A(T) &= U_A + U_{A1}\left[\frac{T}{T_{NOM}} - 1\right] \\ U_B(T) &= U_B + U_{B1}\left[\frac{T}{T_{NOM}} - 1\right] \\ U_C(T) &= U_C + U_{C1}\left[\frac{T}{T_{NOM}} - 1\right] \end{aligned} \tag{5.105}$$

where:

UTE is the parameter to model the temperature dependence of concentration dependent mobility

U_{A1}, U_{B1}, and U_{C1} are the parameters to model the temperature dependence of mobility parameters U_A, U_B, and U_C, respectively, as discussed in Section 5.3.1

The temperature dependence of the saturation velocity is defined by model parameter *AT* as

$$v_{sat}(T) = v_{sat} - AT\left(\frac{T}{T_{NOM}} - 1\right) \tag{5.106}$$

The temperature dependence of S/D series resistance is modeled by a parameter *PRT* such that

$$R_{DSW}(T) = R_{DSW} - PRT\left(\frac{T}{T_{NOM}} - 1\right) \tag{5.107}$$

The temperature coefficients are optimized to fit the measurement data obtained at the target range of operating temperatures.

5.4 Substrate Current Model

The channel electrons traveling through high electric field near the drain end of the channel can become highly energetic. These high energetic electrons are called *hot electrons* and can cause impact ionization generating electrons and holes [42–44]. The holes go into the substrate creating substrate current I_{sub} as shown in Figure 5.15. Some of the electrons have enough energy to overcome the Si/SiO_2 energy barrier generating gate current I_g as shown in Figure 5.15. And, some are collected to the drain, contributing to the drain current. The maximum electric field E_m near the drain has the greatest control of hot carrier effects.

Figure 5.16 shows the detailed mechanism of hot carrier effects on nMOSFET device performance.

Figure 5.16 shows the effect of high drain bias $V_{ds} > V_{dsat}$ on nMOSFET devices at strong inversion, $V_{gs} > V_{th}$. As shown in Figure 5.16, the inversion layer electrons traveling under high electric field cause the following:

1. High energetic electrons traveling along the channel acquire energy from the electric field and become hot;

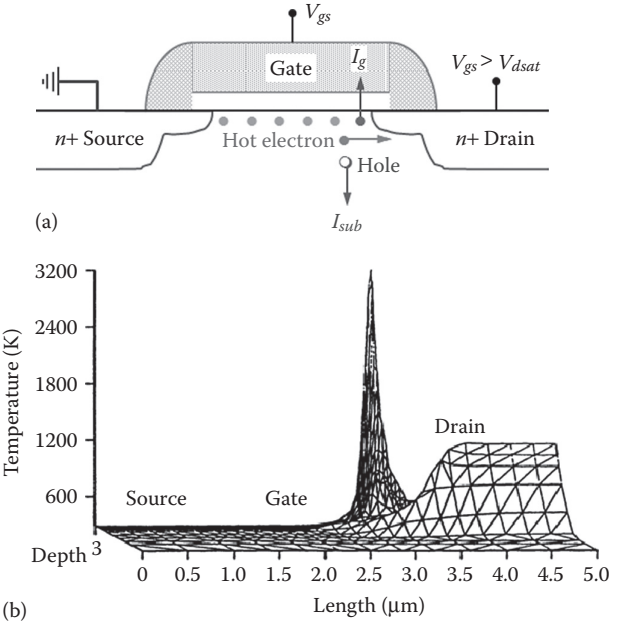

FIGURE 5.15
Hot carrier effect in MOSFETs: (a) channel hot electrons in an nMOSFET device contributing to the drain current and generating gate current and (b) electron temperature near the drain end of the channel of the nMOSFET.

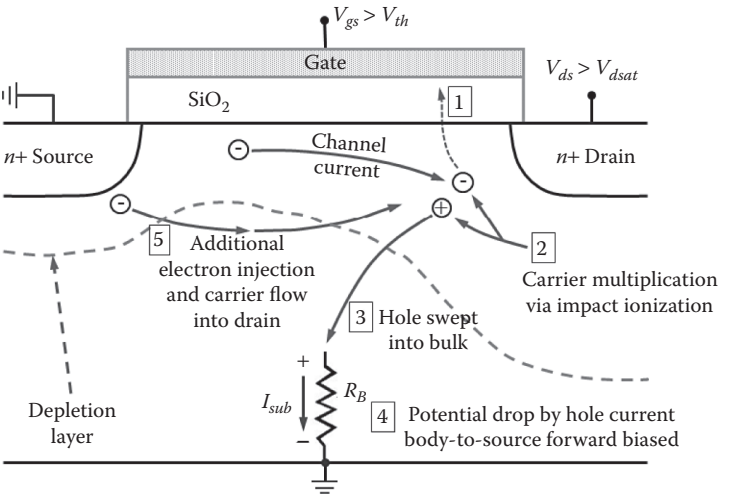

FIGURE 5.16
Cross section of an nMOSFET device in saturation showing hot carrier effects: different physical mechanisms include (1) electron injection into the oxide generating gate current, (2) carrier multiplication by impact ionization, (3) hole flow in the bulk, (4) substrate current flow due to holes, and (5) secondary impact ionization generating additional drain current; the substrate current flow causes a potential drop on the substrate due to the finite substrate resistance R_B, thus forward biasing the source-body *pn*-junction.

2. These hot electrons cause carrier multiplication due to impact ionization by collision with the silicon atoms and breaking covalent bonds, thus creating electrons and holes;
3. Holes are swept into the substrate due to the favorable electric field producing substrate current, I_{sub};
4. I_{sub} flowing through the bulk causes a potential drop in the body, which forward biases the source channel *pn*-junction, thus reducing the source channel potential barrier, $\phi_{bi}(s)$, and enabling more carrier injection from the source to channel;
5. Additional carrier injection due to reduced $\phi_{bi}(s)$ causes more carrier flow in the drain, thus increasing I_{ds} referred to the SCBE discussed earlier.

From the above discussions, we find that the substrate current in an nMOSFET device is due to the holes that are generated by impact ionization of channel hot electrons as they travel from the source to drain. The total drain current, I_{ds}, including the substrate current due to impact ionization is given by

$$I_{ds} = I_{dsat} + I_{sub} \tag{5.108}$$

where:
I_{dsat} is the saturation drain current

If M is the avalanche multiplication factor due to impact ionization, then I_{sub} can be expressed as

$$I_{sub} = (M-1)I_{dsat} \tag{5.109}$$

where M is given by

$$M = \frac{1}{1-\int \alpha_n dy}$$

or (5.110)

$$M - 1 = M\int \alpha_n dy$$

where:
α_n is the electron impact ionization coefficient per unit length and is a strong function of the channel electric field E

In order to derive a generalized expression for I_{sub}, we replace I_{dsat} by I_{ds}. Then from Equations 5.109 and 5.110, we can show

$$I_{sub} = I_{ds}M\int \alpha_n dy \tag{5.111}$$

Since I_{sub} resulting from the channel hot electrons impact ionization process is 3–5 orders of magnitude smaller than the drain current I_{ds}, it can be considered as a low-level avalanche current. For low-level multiplication $M \approx 1$, and therefore, Equation 5.111 becomes

$$I_{sub} = I_{ds}\int_0^{l_i} \alpha_n dy \tag{5.112}$$

where y is the distance along the channel with $y = 0$ representing the start of the impact ionization region, and l_i is the length of the drain section where impact ionization takes place as shown in Figure 5.17. Several forms for α_n have been proposed but most commonly used form is

$$\alpha_n = A_i \exp\left[-\frac{B_i}{E}\right] \tag{5.113}$$

where:
A_i and B_i are called the impact ionization coefficients

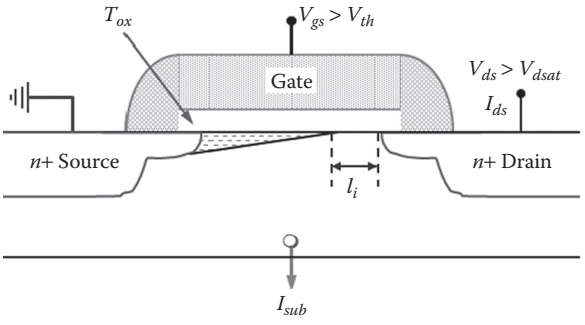

FIGURE 5.17
Hot carrier current effect in nMOSFETs showing the impact ionization region, l_i, at the drain end of the device.

Most of the reported data on α_n have been measured in bulk silicon and the constants A_i and B_i show a wide range of values [42–44]. Slotboom et al. [44] have measured α_n at the surface and in the bulk silicon and reported the values for the constants, which are provided in Table 5.1.

Due to the exponential dependence of α_n on electric field as shown in Equation 5.113, it is easy to see that the impact ionization will dominate at the position of the maximum electric field. In a MOSFET, the maximum electric field E_m is present at the drain end as shown in Figure 5.18a. The sharp maximum E_m shown in Figure 5.18a can be reduced by device

TABLE 5.1
Surface and Bulk Impact Ionization Coefficients in Silicon

α_n	A_i (cm^{-1})	B_i (V cm^{-1})
Surface	2.45×10^6	1.92×10^6
Bulk	7.03×10^5	1.23×10^6

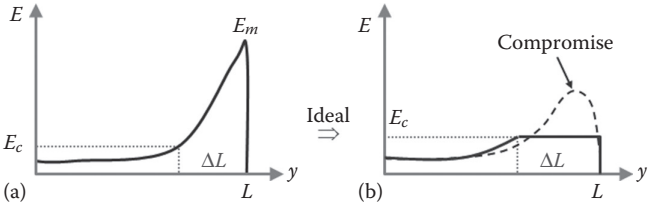

FIGURE 5.18
Hot carrier effect in nMOSFETs: (a) maximum electric field, E_m, at the drain end of the channel and (b) smoother E_m to reduce the effect of substrate current on device performance.

optimization as shown in Figure 5.18b. Therefore, we expect the impact ionization integral in Equation 5.112 to be dominated by the maximum electric field E_m at the drain end of the channel. Substituting Equation 5.113 in Equation 5.112 we get

$$I_{sub} = I_{ds} A_i \int_0^{l_i} \exp\left(-\frac{B_i}{E(y)}\right) dy \tag{5.114}$$

In order to solve Equation 5.114, we first calculate the electric field in the channel. Based on a pseudo-2D analysis [45], it can be shown that the channel electric field E can be expressed as

$$E(y) = -\frac{dV}{dy} = \sqrt{\frac{\left(V(y) - V_{dsat}\right)^2}{l_i^2} + E_c^2} \tag{5.115}$$

where E_c represents the channel electric field at which the carriers reach velocity saturation (at $y = 0$ and $E = E_c$) and the corresponding voltage at that point is the saturation voltage V_{dsat}. E_c is about 2×10^4 V cm^{-1} for electrons. The parameter l_i can be treated as an effective impact ionization length and is given by

$$l_i^2 = \frac{\varepsilon_{si}}{\varepsilon_{ox}} T_{ox} X_j \tag{5.116}$$

where:

T_{ox} is the gate oxide thickness
X_j is the S/D junction depth

Although Equations 5.115 and 5.116 were derived for conventional S/D junctions, they are still valid for lightly-doped drain (LDD) as well as SDE MOSFET structures. For LDD and SDE devices, X_j is the junction depth of the LDD or SDE region. The maximum field E_m, which occurs at the drain end, can easily be obtained replacing $V(y)$ by V_{ds} in Equation 5.115. Again, since $E_c^2 << \left(V_{ds} - V_{dsat}\right)^2 / l_i^2$ in Equation 5.115, neglecting E_c results in the following approximate expression for E_m, we get

$$E_m \cong \frac{\left(V_{ds} - V_{dsat}\right)}{l_i} \tag{5.117}$$

Now, we replace dy in Equation 5.114 by $\left(dy/dE\right)dE = -E^2\left(dy/dE\right)d\left(1/E\right)$ to get

$$I_{sub} = -I_{ds} A_i \int_{Ec}^{E_m} \exp\left(-\frac{B_i}{E(y)}\right) E^2 \frac{dy}{dE} d\left(\frac{1}{E}\right) \tag{5.118}$$

From Pseudo-2D analysis of the velocity saturation region, we can show

$$E(y) = E_c \cosh\left(\frac{y}{l_i}\right) = E_c \frac{\exp(y/l_i) - \exp(-y/l_i)}{2} \cong E_c \frac{1}{2}\exp\left(\frac{y}{l_i}\right) \tag{5.119}$$

Since l_i is very small and y/l_i is a very large number, $\exp(-y/l_i)$ is negligibly small; then differentiating Equation 5.119 we get

$$\frac{dE}{dy} = E_c \frac{1}{2}\exp\left(\frac{y}{l_i}\right).\left(\frac{1}{l_i}\right) = \frac{E}{l_i} \tag{5.120}$$

Therefore,

$$-E^2\left(\frac{dy}{dE}\right) = -E^2\left(\frac{l_i}{E}\right) = l_i E \tag{5.121}$$

Substituting Equation 5.121 in Equation 5.118, we get

$$I_{sub} = -I_{ds} A_i \int_{Ec}^{E_m} l_i E \exp\left(-\frac{B_i}{E(y)}\right) d\left(\frac{1}{E}\right) \tag{5.122}$$

Since the exponential term in Equation 5.122 has a pronounced peak at $E = E_m$, we evaluate it at $E = E_m$ and let it be constant over the region so that we can remove it from the integral. After this simplification, Equation 5.122 can be solved for I_{sub} as

$$I_{sub} = -I_{ds} A_i l_i E_m \int_{Ec}^{E_m} \exp\left(-\frac{B_i}{E(y)}\right) d\left(\frac{1}{E}\right) \tag{5.123}$$

After integration and simplification, we can show assuming $E_c << E_m$,

$$I_{sub} \cong I_{ds} \frac{A_i}{B_i} l_i E_m \exp\left(-\frac{B_i}{E_m}\right) \tag{5.124}$$

Substituting for E_m from Equation 5.117 and Equation 5.124 can be expressed in terms of drain voltage as

$$I_{sub} \cong I_{ds} \frac{A_i}{B_i}\left(V_{ds} - V_{dsat}\right)\exp\left(-\frac{l_i B_i}{V_{ds} - V_{dsat}}\right) \tag{5.125}$$

Equation 5.125 is used for substrate current modeling. Note that Equation 5.125 is independent of device geometry. In order to model channel length dependence of I_{sub}, the ratio A_i/B_i can be replaced by $(\alpha_0 + \alpha_1/L_{eff})$ to express

$$I_{sub} \cong \left(\alpha_0 + \frac{\alpha_1}{L_{eff}}\right)\left(V_{ds} - V_{dsat}\right)\exp\left(-\frac{\beta}{V_{ds} - V_{dsat}}\right).I_{dsa} \tag{5.126}$$

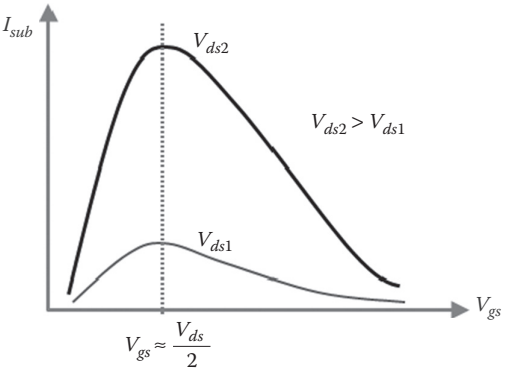

FIGURE 5.19
Impact ionization induced substrate current I_{sub} versus gate voltage V_{gs} characteristics of nMOSFET devices for two different values of V_{ds}; typically, for any value of V_{ds}, the value of I_{sub} attains a maximum value at $V_{gs} \approx V_{ds}/2$.

where:
$\beta = l_i B_i$
I_{dsa} is the drain current without the impact ionization

Thus, the basic parameter set for modeling I_{sub} is $\{\alpha_0, \alpha_1, \beta\}$ which is obtained by optimizing the measurement data for MOSFET devices.

Figure 5.19 shows a typical I_{sub} versus V_{gs} plot for two values of V_{ds}. It is found that for a given value of V_{ds}, initially I_{sub} increases with increasing V_{gs} due to an increase in the drain current (i.e., increase in the inversion charge density from weak to strong inversion regime as V_{gs} increases from 0 to strong inversion). Further increase in V_{gs} eventually results in a decrease in I_{sub} due the reduction in the effective width of the pinch-off region, resulting in an increase in V_{dsat}, which in turn reduces the electric field along the channel. Thus, as V_{gs} increases, I_{sub} increases first, reaches its peak value at a certain V_{gs}, and then decreases resulting in a bell-shaped curve with its maximum occurring at a gate voltage, $V_{gs} \approx 0.5V_{ds}$. However, in nanoscale devices, the lateral electric field along the direction of current flow is extremely high and due to local carrier heating, the entire channel length is velocity saturated. Therefore, for any nanoscale MOSFETs, the impact ionization occurs at a lower value of $V_{gs} > V_{th}$ and the value of V_{gs} at I_{sub}(peak) is almost independent of V_{ds} [43].

In order to extract the impact ionization parameters A_i, B_i, and l_i, the general Equation 5.125 can be expressed as [46,47]

$$\ln(Y) = mX + c \tag{5.127}$$

where

$$Y = \frac{I_{sub}}{I_{ds}(V_{ds} - V_{dsat})}$$
$$X = \frac{1}{(V_{ds} - V_{dsat})} \tag{5.128}$$
$$m = -l_i B_i$$

and

$$c = \ln\left(\frac{A_i}{B_i}\right)$$

Equation 5.127 represents a straight line with the slope, m, and intercept, c, given by Equation 5.128. Thus, $\ln\left[I_{sub}/I_{ds}\left(V_{ds} - V_{dsat}\right)\right]$ versus $1/\left(V_{ds} - V_{dsat}\right)$ plot is a straight line with a slope $m = -l_i B_i$ and the intercept $c = \ln(A_i/B_i)$. From such plots for MOSFETs with different processing parameters, the value of l_i can be determined [47] as shown in Figure 5.20.

As discussed in Section 5.3.6, substrate current I_{sub} flowing into the substrate increases drain current significantly, resulting in lower output resistance as shown in Figure 5.13. This is due to the fact that I_{sub} flowing to the substrate causes an IR drop in the substrate, resulting in a body bias; this body bias forward biases the source/body pn-junction thus lowering the source to chain potential barrier for carriers. As a result, more carriers are injected from the source to the inversion channel, causing a significant increase in I_{ds}, which is referred to as the SCBE. The SCBE causes V_{th} drop and manifold increase in I_{sub} and consequently, I_{dS} as shown in Figure 5.13

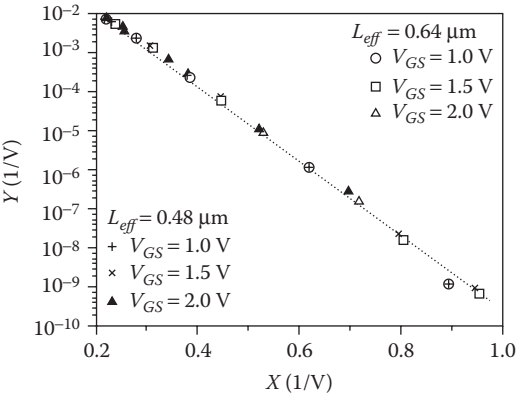

FIGURE 5.20
Plot of $Y = I_{sub}/\left[I_{ds}\left(V_{ds} - V_{dsat}\right)\right]$ versus $X = \left(V_{ds} - V_{dsat}\right)^{-1}$ with different V_{gs} for LDD type nMOSFETs of different channel length; all data are obtained under $V_{bs} = 0$ for $W = 40$ µm devices and $T_{ox} = 150$ A. (Data from S. Saha, *Solid-State Electron.*, 37, 1786–1788, 1994.)

5.4.1 Gate-Induced Drain Leakage Body Current Model

When $V_{gs} < 0$ (or $V_{gs} = 0$) and high V_{ds} is applied to the device as shown in Figure 5.21, the electric field is very high in the drain region. This high electric field causes a large band bending, which results in *band-to-band tunneling* (BTBT). As a result a significant amount of drain leakage current is observed.

The drain leakage current due to BTBT is related to the generation of carriers in the drain overlap region under the gate as shown in Figure 5.21. From the basic device physics, we know that a positive gate bias tends to invert the *p*-type channel. Similarly, a negative gate bias tends to invert the *n*-type drain junction in the overlap region. The inversion of the drain does not easily take place, since the drain is doped more heavily than the channel. Nevertheless, when V_{gd} is fairly negative, the applied drain bias at least causes the overlap region to be depleted of carriers. As the minority carriers, generated either by BTBT or trap-assisted tunneling, arrive at the surface to attempt to form the inversion layer, they immediately get swept laterally to the substrate. The current that flows as a result of the carriers being swept from the overlap region constitutes the *gate-induced drain leakage* (GIDL) current, I_{gidl}. In the framework of this explanation, we see that GIDL is not an SCE. The leakage current tends to be significant in LDD devices where the overlapped region is lightly doped. GIDL is, generally, less a severe in nanometer-scale devices whose drain extension forms a heavily doped junction.

Similar current is also observed at the source end of the device. The components of body current observed are gate-induced drain leakage and gate-induced source leakage (GISL). The general expressions to model GIDL and GISL are given by

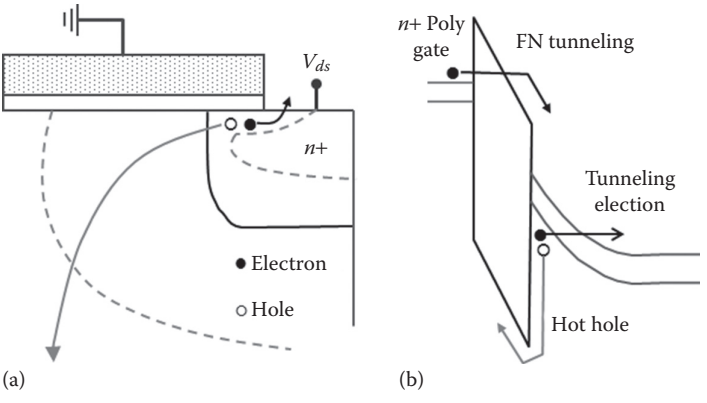

FIGURE 5.21
GIDL current in an nMOSFET device: (a) gated diode, at the drain MOSFET only, showing electron–hole pair generation and transport and (b) Fowler-Nordheim (FN) tunneling due to high lateral electric field by applied drain voltage.

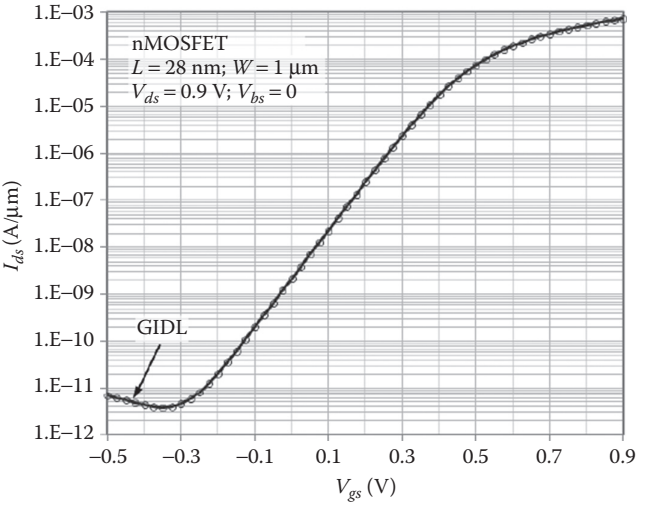

FIGURE 5.22
GIDL in MOSFETs: I_{ds} versus V_{gs} characteristics of an nMOSFET device showing the effect of GIDL on a 28 nm nMOSFET performance for $V_{gs} < 0$.

$$I_{gidl} = NF.AGIDL.W_{eff}\left(\frac{V_{ds} - V_{gseff} - EGIDL}{3TOXE}\right)\exp\left(\frac{-3TOXE.BGIDL}{V_{ds} - V_{gseff} - EGIDL}\right)\frac{V_{DB}^3}{CGIDL + V_{DB}^3}$$

and (5.129)

$$I_{gisl} = NF.AGISL.W_{eff}\left(\frac{-V_{ds} - V_{gseff} - EGISL}{3TOXE}\right)\exp\left(\frac{-3TOXE.BGISL}{-V_{ds} - V_{gseff} - EGISL}\right)\frac{V_{SB}^3}{CGISL + V_{SB}^3}$$

The model parameters (AGIDL, AGISL), (BGIDL, BGISL), (CGIDL, CGISL), and (EGIDL, EGISL) are obtained from the measured $I_{ds} - V_{gs}$ data obtained for $-V_{gs}$ to $+V_{gs}$ at $V_{ds} = V_{dd}$ (supply voltage); NF is the number of fingers used in the layout for MOSFETs. GIDL must be accounted if the standby current of a circuit is an important specification. Figure 5.22 shows GIDL effect in a 28 nm channel length nMOSFET device.

5.4.2 Gate Current Model

As the oxide becomes progressively thinner in each generation of IC technology, the magnitude of the direct tunneling currents through the oxide becomes more significant. In direct tunneling, the carriers from the inversion layer of silicon surface can tunnel directly through the energy gap of the SiO_2 layer instead of tunneling into the conduction band of the SiO_2 layer.

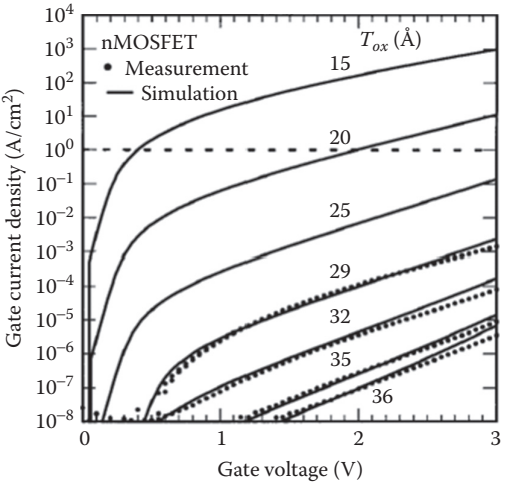

FIGURE 5.23
Measured and simulated tunneling currents in thin oxide polysilicon gate MOSFET devices. The horizontal broken line indicates a tunneling current level of 1 A cm^{-2}. (Data from S.-H. Lo et al., *IEEE Electron Device Lett.*, 18, 209–211, 1997.)

The direct tunneling current can be very large for advanced CMOS technologies with oxide thickness of about 1 nm. Figure 5.23 shows the plots of measured and simulated tunneling current versus gate voltage in polysilicon-gate MOSFETs with different gate oxide thicknesses [48]. Figure 5.23 shows that I_{gate} is extremely high for thinner $T_{ox} < 2$ nm due to direct tunneling gate leakage current. Therefore, it is critical to model gate current of advanced MOSFETs for circuit design.

There are five tunneling components of gate current, I_g, as shown in Figure 5.24. They are

1. I_{gd} = gate-to-drain current between the gate and the heavily doped drain junction
2. I_{gcd} = gate-to-channel current and to the drain
3. I_{gs} = gate-to-source current between the gate and the heavily doped source diffusion
4. I_{gcs} = gate-to-channel current and to the source
5. I_{gb} = gate-to-substrate tunneling current (accumulation and inversion)

The detailed analysis of these tunneling currents unavoidably involves quantum mechanical analysis [48–56]; however, the analytical expressions for compact gate current modeling are described in BSIM4 [28].

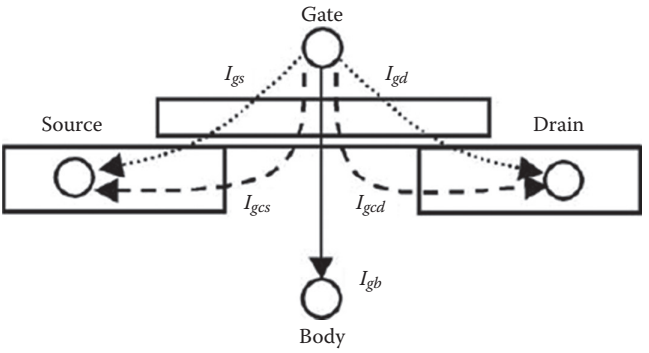

FIGURE 5.24
Gate current model: different components of gate tunneling current in nanometer-scale MOSFETs.

5.5 Summary

This chapter presents compact MOSFET models for small geometry devices. In order to develop accurate small geometry compact model, the different structural and physical effects are modeled in device threshold voltage. First of all, the nonuniform substrate doping is modeled in device threshold voltage. Then the model for small geometry effects such as short channel effect, reverse short channel effect, narrow width and reverse-NWEs are included in the threshold voltage model. In this chapter, the accurate mobility model is derived to account for the effect of high gate bias on device performance. With accurate mobility model, the regional drain current models for the linear and saturation regions are developed to model high lateral electric field and velocity saturation due to high drain bias. Finally, the compact models for hot carrier–induced substrate current for MOSFETs devices are presented.

Exercises

5.1 Consider an nMOSFET device with channel doping concentration $N_a = 1 \times 10^{18}$ cm^{-3} and $T_{ox} = 3$ nm. Assume $Q_f = 0$, $V_{sb} = 0$, and $n+$ degenerately doped poly gate.

a. Calculate the value of long channel threshold voltage V_{th0}.

b. Derive the expressions for $K1$ and $K2$ in terms of the channel and substrate body effect coefficients and an intermediate substrate bias.

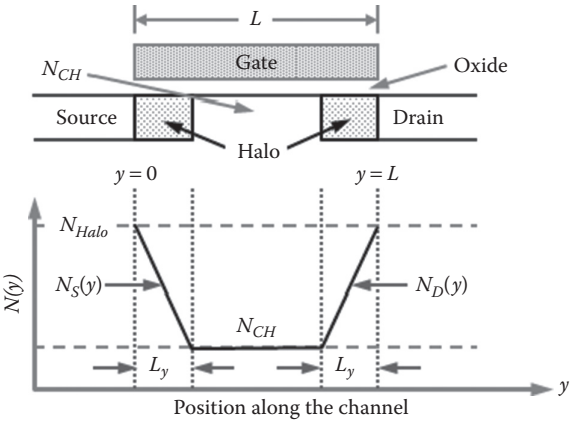

FIGURE E5.2.1
Triangular halo/pocket doping profiles for MOSFET device structure.

Discuss the impact of the model parameters *K*1 and *K*2 on V_{th} of an MOS transistor.

5.2 In this problem you will use the triangular halo doping profiles shown in Figure E5.2.1 to model the halo doping distribution near the source and drain ends of a MOSFET channel. Given: L = channel length, L_y = halo spread inside L at the source and drain ends, N_{Halo} = maximum halo concentration, and N_{CH} = channel doping concentration:

a. Show that the halo doping profile $N_S(y)$ at any point y near the source end of the channel is given by

$$N_S(y) = N_{CH}\left(\frac{y}{L_y}\right) + N_{Halo}\left[1 - \left(\frac{y}{L_y}\right)\right]$$

b. Show that the halo doping profile $N_D(y)$ at any point y near the drain end of the channel is given by

$$N_D(y) = N_{CH}\left[\left(\frac{L}{L_y}\right) - \left(\frac{y}{L_y}\right)\right] + N_{Halo}\left[1 - \left\{\left(\frac{L}{L_y}\right) - \left(\frac{y}{L_y}\right)\right\}\right]$$

5.3 In order to develop V_{th}-model for nonuniform lateral channel doping profile, we used piecewise box-shaped step functions for N_{Halo} to represent a constant channel doping concentration near the source and drain ends of the channel while N_{CH} to represent a constant channel concentration, where $N_{Halo} > N_{CH}$. In reality, the halo doping profile near the source and drain ends can be more accurately modeled by a triangular-shaped function. Use triangular profiles [Figure E5.2.1] to

represent the halo doping, model V_{th} for nonuniform lateral channel doping. Given L = channel length, and L_y = halo spread inside L at the source/drain ends:

a. Derive an expression for the average channel doping concentration to account for the halo doping in the channel. Clearly define all parameters and explain any assumptions you make.

b. Show the expressions for model parameters from your work in part (a).

c. How would you extract the model parameters obtained in part (b)?

d. Compare the model parameters in part (b) with that derived using box-shaped profiles given by Equation 5.12. Explain.

5.4 In order to derive an effective inversion carrier mobility model, it is shown that the effective channel electrical field, $E_{eff} = [0.5Q_{inv} + Q_b]/\varepsilon_{si}$, where Q_{inv} and Q_b are the inversion charge and bulk (depletion) charge under the gate, respectively, and ε_{si} is the dielectric constant of silicon. The dependence of surface mobility μ_s on process parameters such as T_{ox} and N_{sub} and the terminal voltages are lumped in E_{eff}. Assume $V_{gs} > V_{th}$ and small V_{ds}:

a. Show that $E_{eff} \cong (V_{gs} + V_{th})/6T_{ox}$.

b. If the effective mobility is modeled by: $\mu_{eff} = \mu_0/[1 + E_{eff}/E_0)]^{\nu}$, where $\mu_s = \mu_0$ @ $V_{gs} = 0$ and E_0 and ν are parameters determined from the measured data. Use the expression for E_{eff} in part (a) to show that:

$$\mu_{eff} = \frac{\mu_0}{1 + U_a\left[\left(V_{gs} + V_{th}\right)/T_{ox}\right] + U_b\left[\left(V_{gs} + V_{th}\right)/T_{ox}\right]^2}$$

where:

U_a and U_b are the model parameters that are determined experimentally from *I–V* data of MOSFET devices

Clearly state any assumptions you make.

5.5 An nMOSFET device is designed with a gate oxide thickness of 5 nm and a uniformly doped substrate with $N_a = 5 \times 10^{17}$ cm^{-3}. Assuming that the "ON" state of this device is characterized by $\phi_s = 2\phi_B$ and the "OFF" state by $\phi_s = \phi_B$, estimate the ratio of ON to OFF currents flowing in the device.

5.6 Complete the mathematical steps to show that the MOSFET drain current expression in the linear region is given by Equation 5.74.

5.7 Complete the mathematical steps to show that the general expression for MOSFET saturation drain voltage is given by Equation 5.79.