CHAPTER

# Leakage currents

# 5

## CHAPTER OUTLINE



During the past four decades, while IC manufacturers were continually reducing the physical size of planar silicon MOSFETs in order to improve their speed and power efficiency and to lower the fabrication cost per transistor, an undesirable effect was growing in parallel. Because of short-channel effects, the leakage current and, consequently, the leakage (static) power dissipation were increasing. Today, conventional planar, bulk MOSFET scaling has reached a point where almost half of the power that is dissipated in a chip is due to the static leakage power. The conventional bulk MOSFET scaling is coming to an end not because of fabrication difficulties, rather because of the fact that further scaling would not decrease power dissipation and may in fact increase it. As described in Chapter 1, the FinFET architecture could greatly reduce the short-channel effects, convincing the industry to alter the architecture of the traditional MOSFET from planar to FinFET.

If we plot the drain current $I_d$ of a typical MOSFET (on a logarithmic scale) as a function of its gate voltage $V_{gs}$ for a nonzero drain voltage $V_{ds} \neq 0$, the value of the intersection with the drain current axis gives the off-state leakage current (see Figure 5.1). Typically, the transistor's off-state leakage current for $V_{ds}$ equal to the supply voltage $V_{dd}$ is defined as the off-state current, $I_{off}$. Ideally, the transistor is

FinFET Modeling for IC Simulation and Design. http://dx.doi.org/10.1016/B978-0-12-420031-9.00005-1

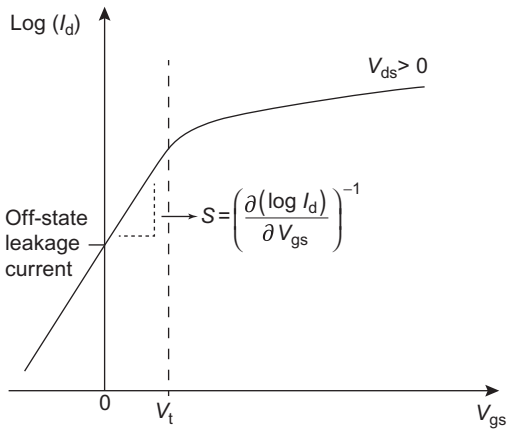


**FIGURE 5.1**

The transistor's drain current (on a logarithmic scale) as a function of its gate voltage. The off-state leakage current, threshold voltage, and subthreshold swing are marked.

supposed to be completely off at this bias point, but as explained later in this chapter, there always exist leakage currents due to different mechanisms. These sources of leakage current include the weak-inversion current between the drain terminal and source terminal, the substrate and drain junction leakage currents (both forward[1] and reverse diode currents), the gate-induced drain leakage (GIDL) current between the drain terminal and the substrate terminal, and a portion of the gate oxide tunneling current.

From those, the junction leakage and gate oxide tunneling currents extend to the transistor's on state and add to impact ionization leakage, which becomes noticeable in the on state. In addition, there might be leakage currents between terminal pairs other than those involving the drain; for instance, gate-induced source leakage (GISL) between the source and substrate terminals. BSIM-CMG is equipped with models that can simulate the leakage currents of all the terminals.[2]

In Section 5.1, the weak-inversion current is reviewed, with a focus on the terminology used in the field. In Section 5.2, an approach similar to the one originally used in BSIM4 is described to develop a GIDL/GISL model for BSIM-CMG. Section 5.3 discusses the gate oxide tunneling mechanisms and formulations. Finally, Section 5.4 explains the impact ionization model. The junction leakage component is reviewed in detail in Chapter 9.

[1]Forward leakage can occur under intentional forward well bias and under voltage spikes that may even lead to latch-up.

[2]Note that for a FinFET on a silicon-on-insulator (SOI) substrate (BULKMOD = 0), the substrate leakage current will flow out of the source; that is, the holes will be injected into the source and appear as an additional drain-source leakage.

## 5.1 WEAK-INVERSION CURRENT

Assuming a room-temperature n-channel MOSFET with $V_{gs} < V_t$, there are always some electrons in the source diffusion region that have enough energy to pass over the source-channel barrier and reach the drain side (see Figure 5.2). These electrons will create a nonzero $I_d$ for $V_{ds} > 0$. This is known as the weak-inversion or subthreshold current, and it is the dominant leakage mechanism in modern devices. Since the number of these carriers is exponentially increased by an applied gate voltage below the threshold voltage, the weak-inversion current is represented by a straight line with a finite slope in a semilog plot as shown in Figure 5.1. The reverse of the slope of this line is known as the subthreshold swing $S$, and it has units of millivolts of the gate voltage per decade of the drain current. The ideal value of $S$ is approximately 60 mV per decade. This value is a fundamental limit and represents the fact that $V_g$ needs to be increased by at least 60 mV in order to achieve a factor of 10 increase in current over the potential barrier. To beat this limit, the carriers must additionally tunnel through the barrier. Tunneling may be afforded by a microelectromechanical system switch [1] or a tunneling transistor based on gate-induced band-to-band-tunneling [2].

Why is the subthreshold leakage a big concern in the recent CMOS technology nodes? As can be visualized through Figure 5.1, reducing $V_t$, which is equivalent to shifting the entire curve to the left, or increasing $S$, which makes the slope shallower, will increase the off-state leakage current exponentially, and hence will increase the static power dissipation. Short-channel MOSFETs inherently have smaller threshold voltages owing to two-dimensional electrostatics which originates from the proximity of the source and drain regions and their charge sharing with the gate (the effect known as $V_t$ roll-off). The effect is enhanced at higher drain voltages in short-channel

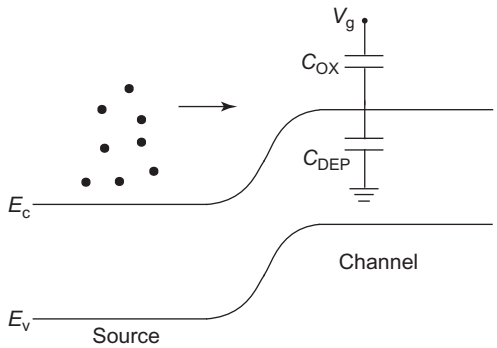


**FIGURE 5.2**

The potential barrier at the source/channel determines the subthreshold current. The current can be increased by 10 times for a 60 mV decrease in the potential barrier or equivalently for a 60 × (1 + $C_{DEP}/C_{OX}$ ) mV increase in $V_g$. The thin body of the FinFET becomes fully depleted for a small value of applied $V_g$; this makes $C_{DEP}$ equal zero and $S$ roughly 60 mV for the FinFET.

devices because the drain region is close enough that the drain voltage can affect and lower the source-channel barrier height at the channel-dielectric interface (the effect known as drain-induced barrier lowering, DIBL). Typical values of DIBL for high-performance, 32 nm node, planar bulk MOSFETs are around 100 mV/V; this means there is a 100 mV shift in $V_t$ for 1 V of applied drain voltage. In very short channel devices, $S$ is also affected and becomes larger (the effect known as subthreshold swing degradation). This is because the drain is so close that it can lower the source-channel barrier height for paths a few nanometers below the surface, resulting in subsurface leakage. Typical values of $S$ for high-performance, 32 nm node, planar bulk MOSFETs are in the range of 70-100 mV per decade.

By providing a tighter electrostatic control around the channel, the FinFET has demonstrated a great ability in controlling short-channel effects and suppression of the off-state leakage current. The values of DIBL and $S$ for a high-performance, 22 nm node FinFET are approximately 50 mV/V and approximately 70 mV per decade, respectively, leading to low values of $I_{off}$ in the range of 20-100 nA/μm [3]. Still, determining the subthreshold behavior in scaled FinFETs through a careful modeling of effects which reduce $V_t$ or degrade $S$ (and hence worsen the off-state leakage current) is of great importance for IC designs. This is especially critical for low-power, mobile circuit applications. As discussed in Chapter 3, the BSIM-CMG model employs a current equation in its core model which is valid for a long-channel FinFET from weak inversion (subthreshold) to strong inversion. For the implementation of $V_t$ roll-off, DIBL, and $S$ degradation models, please refer to the real device models described in Chapter 4.

## 5.2 GATE-INDUCED SOURCE AND DRAIN LEAKAGES

Figure 5.3 illustrates the cross-section of an n-channel, double-gate FinFET and its energy-band diagram for the gate-drain overlap region when a low gate voltage and a high drain voltage are applied. If the band bending at the oxide interface is greater than or equal to the energy band gap $E_g$ of the drain material, band-to-band tunneling will take place. The electrons in the valence band of the n-type drain will tunnel through the thinned band gap into the conduction band, and they will be collected at the drain contact to be a part of the drain current, whereas the reaming holes will be collected at the substrate contact (the source contact in the case of a FinFET on an SOI substrate) and will contribute to the substrate (source) leakage. This phenomenon, which was first elucidated and modeled by researchers at the University of California, Berkeley [4], discerns a potential major contributor to the off-state leakage current (see Figure 5.4) and is called the gate-induced drain leakage (GIDL) current. Depending on the voltages applied, there might also exist a gate-induced source leakage (GISL) current.

But what are the prerequisites for the GIDL current to flow? First, there must be band bending greater than $E_g$ so that the valence band energy states overlap the conduction band energy states as shown in Figure 5.3. In that case the semiconductor

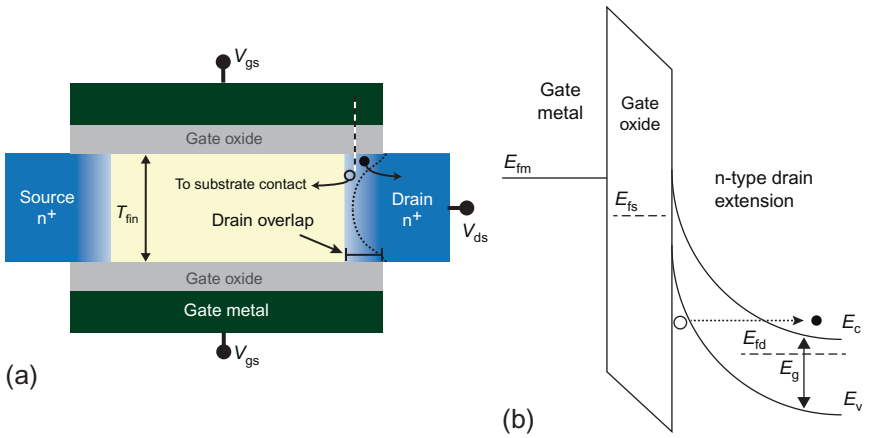


**FIGURE 5.3**

(a) Cross-section of the fin of a FinFET, and (b) illustration of the energy-band diagram along the dashed line in (a). The substrate contact is normal to and below the page.

surface in the gate-drain overlap region is in *deep depletion*, with the band bending being much larger than $2\varphi_B$.[3]

The surface potential can exceed $2\varphi_B$ because there is no inversion hole layer at the surface. There is no hole layer at the surface because any hole there would drift and diffuse to the body/substrate because of the built-in junction potential plus any substrate-drain reverse bias. However, with a forward-biased substrate-drain junction, the holes may remain at the interface and form an inversion layer and cause the band bending to be pinned at roughly $2\varphi_B$, a value smaller than $E_g$, thereby suppressing the GIDL current. In the case of a FinFET on an SOI substrate, holes build up in the floating body and raise the body potential until the body-source junction is slightly forward biased, enabling the GIDL-generated holes to be injected into the $n^+$ source. Second, the electric field needs to be large; that is, the tunneling barrier needs to be narrow. Compared with a planar MOSFET, both of these conditions are more difficult to meet in a FinFET because the potential at both sides of the thin fin is raised or lowered by the same $V_g$. Therefore, lightly doped and very thin fin FinFETs can have negligible GIDL. Try to convince yourself of this by looking at Figure 5.3 and remember the Poisson equation.

In addition, defects or traps in tunneling lead to trap-assisted band-to-band tunneling by providing stepping stones along the tunneling path; therefore, GIDL current is larger in the presence of defects created by ion implantation. Use of solid-source diffusion instead of implantation for drain creation or use of a laser for activation of dopants and annealing has been shown to reduce GIDL [5, Chapter 3].

[3] $\varphi_B$ is the difference between the Fermi potential and the intrinsic potential in the drain.

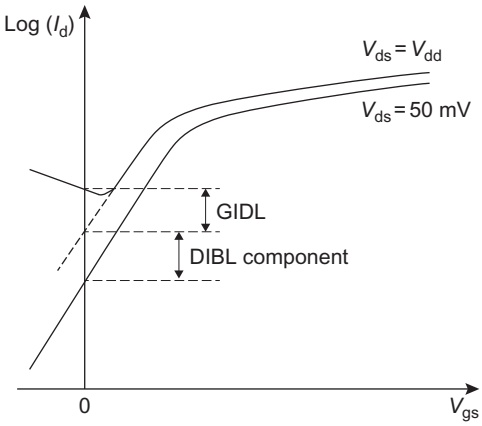


**FIGURE 5.4**

Contributions of DIBL and GIDL to the transistor's off-state leakage current. The position of the dip caused by GIDL will vary around $V_{gs} = 0$ depending on $V_{dd}$, the channel material, doping, and trap density.

### 5.2.1 GIDL/GISL CURRENT FORMULATION IN BSIM-CMG

The band-to-band tunneling current density from the Wentzel-Kramers-Brillouin (WKB) approximation is given by

$$J = A \times E_s \times e^{-B/E_s}, \tag{5.1}$$

where $A$ is a preexponential constant related to the density of states of both the emitting side and the receiving side, $B$ is a physical exponential parameter which depends on $E_g$ and the carrier's effective mass in the tunneling direction (approximately 20 MV/cm for silicon), and $E_s$ is the surface electric field in the drain. With use of Gauss's law at the onset of GIDL, when the band bending in the drain is equal to $E_g$, $E_s$ is given by

$$E_s = \frac{V_{ds} - V_{gs} + V_{fbsd} - E_g}{\epsilon_{ratio} \times EOT}, \tag{5.2}$$

where $V_{fbsd}$ is the flat-band voltage between the gate and the drain, $\epsilon_{ratio}$ is the ratio of the dielectric constant of the substrate material EPSRSUB over that of silicon dioxide, and EOT is the equivalent oxide thickness. Equations (5.1) and (5.2) lead to the following equation for the GIDL current in BSIM-CMG:

$$I_{gidl0} = NFIN_{total} \times W_{eff} \times AGIDL \times \left( \frac{V_{dg} + V_{fbsd} - EGIDL}{\epsilon_{ratio} \times EOT} \right)^{PGIDL}$$
$$\times e^{\left(-(\epsilon_{ratio} \times EOT \times BGIDL)/(V_{dg} + V_{fbsd} - EGIDL)\right)} \tag{5.3}$$

where the constants $A$ and $B$ in Equation (5.1) and $E_g$ in Equation (5.2) have been replaced by the model parameters AGIDL and BGIDL, and EGIDL, respectively, and PGIDL has been introduced for more flexibility in fitting the measured data.

The GISL current is calculated in the same manner:

$$I_{\text{gisl0}} = \text{NFIN}_{\text{total}} \times W_{\text{eff}} \times \text{AGISL} \times \left( \frac{V_{\text{sg}} + V_{\text{fbsd}} - \text{EGISL}}{\epsilon_{\text{ratio}} \times \text{EOT}} \right)^{\text{PGISL}} \times e^{\left(-(\epsilon_{\text{ratio}} \times \text{EOT} \times \text{BGISL})/(V_{\text{sg}} + V_{\text{fbsd}} - \text{EGISL})\right)} \tag{5.4}$$

In addition to the $V_{\text{dg}}$ dependence present in Equation (5.3), in bulk FinFETs (BULKMOD $\neq 0$), the GIDL current is also affected by the substrate bias for small values of $V_{\text{de}}$ (the drain to substrate voltage) as the deep depletion condition in the drain surface starts to fail. The total GIDL current is obtained by multiplying $I_{\text{gidl0}}$ from Equation (5.3) by an empirical factor for modeling the low $V_{\text{de}}$ effect as follows:

$$I_{\text{gidl}} = \begin{cases} I_{\text{gidl0}} \times \dfrac{{V_{\text{de}}}^3}{\text{CGIDL} + {V_{\text{de}}}^3} & V_{\text{de}} \geq 0 \\ 0 & V_{\text{de}} < 0 \end{cases} \tag{5.5}$$

In Equation (5.5), CGIDL is a non-negative fitting parameter. A similar equation holds for the GISL current:

$$I_{\text{gisl}} = \begin{cases} I_{\text{gisl0}} \times \dfrac{{V_{\text{se}}}^3}{\text{CGISL} + {V_{\text{se}}}^3} & V_{\text{se}} > 0 \\ 0 & V_{\text{se}} \leq 0 \end{cases} \tag{5.6}$$

For a FinFET on an SOI substrate (BULKMOD = 0), $I_{\text{gidl0}}$ and $I_{\text{gisl0}}$ are multiplied by $V_{\text{ds}}$ and $V_{\text{sd}}$, respectively. These terms are negligible in comparison with the exponential terms proceeding them, but will guarantee that no GIDL or GISL current is flowing when the drain and source are at the same voltage.

## 5.3 GATE OXIDE TUNNELING

For decades, to help the gate to keep its supremacy against the drain in controlling the source-to-channel barrier, the gate silicon dioxide (silicon oxynitride) thickness was scaled down in proportion to $L_g$. In the early years of this century, the tunneling through the scaled silicon oxynitride started to dominate the transistor's off-state leakage current, making it intolerable. A thicker dielectric layer with a higher dielectric constant ($\kappa$) was required. A thick, high-$\kappa$ gate oxide could retain the control of the gate over the channel with orders of magnitude reduction in dielectric leakage current compared with $SiO_2$ of the same EOT. Furthermore, a metal gate eliminates the polysilicon gate depletion effect which was effectively increasing the gate dielectric thickness and thus reducing the gate control of the channel. High-$\kappa$ oxides, in general, were also found to form a better interface with metal gates

than the traditional polysilicon gate. This led to the introduction of high-$\kappa$ metal-gate technology in the 45 nm node [6], which was extended to the successive nodes. However, the gate tunneling leakage through the gate oxide remains a significant and increasing concern as each new technology generation requires a smaller EOT.

### 5.3.1 GATE OXIDE TUNNELING FORMULATION IN BSIM-CMG

The gate oxide tunneling in the BSIM-CMG model inherits a similar formulation to that of BSIM4. Although the formulation has been derived for a polysilicon-silicon oxide gate stack, it turns out to be accurate enough to be used for high-$\kappa$ metal-gate technology thanks to its flexibility. As illustrated in Figure 5.5, the gate tunneling current is composed of several mechanisms: the gate-to-body leakage current $I_{gb}$, the leakage currents through gate-to-source and gate-to-drain overlaps $I_{gs}$ and $I_{gd}$, and the gate-to-inverted channel tunneling current $I_{gc}$. Part of $I_{gc}$ is collected by the source ($I_{gcs}$), while the rest goes to the drain ($I_{gcd}$). $I_{gb}$, $I_{gs}$, $I_{gd}$, and $I_{gc}$ are determined from the MOS capacitor, dielectric leakage model described below. Then, $I_{gc}$ is extended to nonzero $V_{ds}$ and partitioned into $I_{gcs}$ and $I_{gcd}$.

On the basis of the early work of Lee and Hu [7], the dielectric tunneling leakage current density of a MOS capacitor can be modeled as

$$J_g = A \times \left(\frac{\text{TOXREF}}{\text{TOXG}}\right)^{\text{NTOX}} \times \frac{V_{ge} \times V_{aux}}{\text{TOXG}^2} \times e^{-B\times(\alpha-\beta.|V_{ox}|)\times(1+\gamma.|V_{ox}|)\times\text{TOXG}}, \qquad (5.7)$$

where $A = q^2/(8\pi h\varphi_b)$, $B = \left(8\pi\sqrt{2qm_{ox}\varphi_b^{3/2}}\right)/3h$, $\varphi_b$ is the tunneling barrier height, $m_{ox}$ is the effective carrier mass in the oxide, TOXG is the oxide thickness (different from the physical oxide thickness TOXP to introduce more flexibility), TOXREF is the reference oxide thickness at which all the parameters are extracted, NTOX is a fitting parameter that defaults to 1, $V_{aux}$ is an auxiliary function which represents the density of tunneling carriers as well as available energy states to tunnel

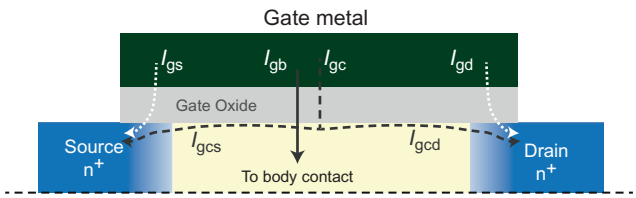


**FIGURE 5.5**

Half cross-section of the fin of the FinFET shown in Figure 5.2. The components of the tunneling current are shown. $I_{gs}$ and $I_{gd}$ are tunneling currents in the gate-to-source/drain overlap regions; $I_{gb}$ flows between the gate and the body; $I_{gc}$ is the gate-to-channel tunneling current and it is partitioned into $I_{gcs}$ and $I_{gcd}$, which flow out of the source and drain, respectively.

into, and $\alpha$, $\beta$, and $\gamma$ are fitting parameters. Depending on the mode of operation (accumulation or depletion/inversion) and the gate tunneling component of interest, the values of $m_{ox}$, $\varphi_b$, and $V_{aux}$ will be different, as explained below.

### 5.3.2 GATE-TO-BODY TUNNELING CURRENT IN DEPLETION/INVERSION

Figure 5.6 schematically demonstrates the dominant leakage mechanism between the gate and the body in depletion/inversion, represented by $I_{gbinv}$. In both p-type MOS (PMOS) and n-type MOS (NMOS), the electrons tunnel from the valence band of the body into the gate material. For this case, the values of $A$, $B$, and $V_{aux}$ for a Si-$SiO_2$ interface (silicon as the body and silicon oxide as the gate oxide) are calculated to be

$$A = 3.75956 \times 10^{-7} \left(\frac{\mathrm{A}}{\mathrm{V}^2}\right), \tag{5.8}$$

$$B = 9.82222 \times 10^{11} \left(\frac{\mathrm{g}}{\mathrm{Fs}^2}\right)^{0.5}, \tag{5.9}$$

and

$$V_{aux,igbinv} = \mathrm{NIGBINV} \times \frac{kT}{q} \times \ln\left(1 + \mathrm{e}^{(V_{ox} - \mathrm{EIGBINV})/(\mathrm{NIGBINV} \times kT/q)}\right). \tag{5.10}$$

In Equation (5.10), NIGBINV and EIGBINV are model parameters.

The total gate tunneling current $I_{gbinv}$ is then given by

$$I_{gbinv} = \mathrm{NFIN}_{total} \times W_{eff} \times L_{eff} \times A \times \left(\frac{\mathrm{TOXREF}}{\mathrm{TOXG}}\right)^{\mathrm{NTOX}} \times \frac{V_{ge} \times V_{aux,igbinv}}{\mathrm{TOXG}^2} \times \mathrm{e}^{-B \times (\mathrm{AIGBINV}(T) - \mathrm{BIGBINV}.q_{ia}) \times (1 + \mathrm{CIGBINV}.q_{ia}) \times \mathrm{TOXG}}, \tag{5.11}$$

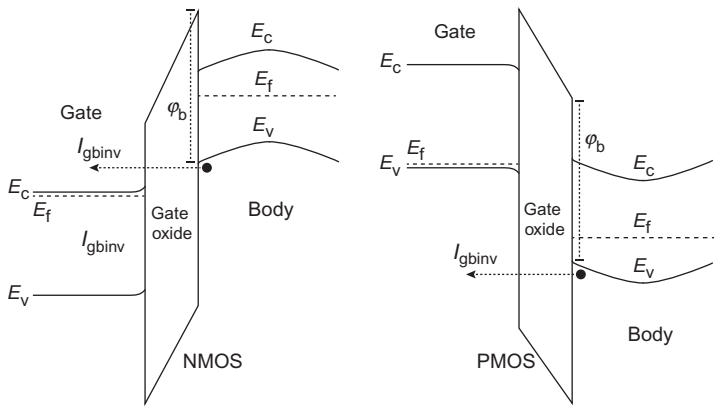


**FIGURE 5.6**

In both NMOS and PMOS, tunneling of valance-band electrons from the body into the gate is the principal cause of the gate-to-body tunneling current in inversion, $I_{gbinv}$.

where $\alpha$, $\beta$, $\gamma$, and $V_{ox}$ in Equation (5.7) have been replaced by the model parameters AIGBINV($T$), BIGBINV, CIGBINV, and the average charge in the channel, $q_{ia}$, respectively.[4] The last approximation is valid since we assume that the fin is fully depleted and the body charge $q_{ba}$ is a fixed value which can be incorporated into other model parameters.

### 5.3.3 GATE-TO-BODY TUNNELING CURRENT IN ACCUMULATION

In accumulation, the dominant leakage current between the gate and the body, $I_{gbacc}$, is the tunneling of conduction-band electrons. In NMOS, the electrons tunnel from the conduction band of the gate material into the conduction band of the body, and in PMOS they tunnel in the reverse direction (see Figure 5.7). For this case, the values of $A$, $B$, and $V_{aux}$ for a polysilicon-silicon oxide-silicon structure are calculated to be

$$A = 4.97232 \times 10^{-7} \left(\mathrm{A/V^2}\right), \tag{5.12}$$

$$B = 7.45669 \times 10^{11} \left(\frac{\mathrm{g}}{\mathrm{Fs^2}}\right)^{0.5}, \tag{5.13}$$

and

$$V_{aux,igbinv} = \mathrm{NIGBACC} \times \frac{kT}{q} \times \ln\left(1 + \mathrm{e}^{(V_{fb}-V_{ge})/(\mathrm{NIGBACC}\times kT/q)}\right). \tag{5.14}$$

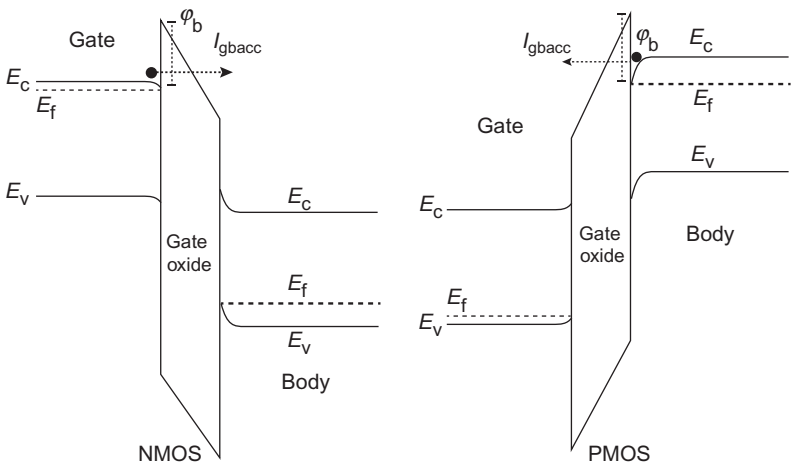


**FIGURE 5.7**

In accumulation, conduction-band electrons tunnel from the gate into the body in NMOS and from the body into the gate in PMOS.

[4]Note that all the charges in the BSIM-CMG model are normalized with respect to $C_{ox}$. That is why one can substitute a voltage with a charge.

In Equation (5.14), NIGBACC is a model parameter.

The total gate tunneling current $I_{\text{gbacc}}$ is then given by

$$I_{\text{gbacc}} = \text{NFIN}_{\text{total}} \times W_{\text{eff}} \times L_{\text{eff}} \times A \times \left(\frac{\text{TOXREF}}{\text{TOXG}}\right)^{\text{NTOX}} \times \frac{V_{ge} \times V_{\text{aux,igbinv}}}{\text{TOXG}^2}$$
$$\times \, \text{e}^{-B\times(\text{AIGBACC}(T)-\text{BIGBACC}.q_{\text{acc}})\times(1+\text{CIGBACC}.q_{\text{acc}})\times\text{TOXG}}, \tag{5.15}$$

where $\alpha$, $\beta$, $\gamma$, and $V_{\text{ox}}$ in Equation (5.7) have been replaced by the model parameters AIGBACC($T$), BIGBACC, CIGBACC, and $q_{\text{acc}}$, respectively.

For BULKMOD $\neq 0$, $I_{\text{gb}}$ (i.e., $I_{\text{gbinv}} + I_{\text{gbacc}}$) simply flows from the gate into the substrate. For BULKMOD = 0, $I_{\text{gb}}$ mostly flows into the source because the potential barrier for holes is typically lower at the source side. To ensure continuity when $V_{\text{ds}}$ switches sign, $I_{\text{gb}}$ is partitioned into a source component $I_{\text{gbs}}$ and a drain component $I_{\text{gbd}}$ using the following partitioning scheme:

$$I_{\text{gbs}} = \left(I_{\text{gbinv}} + I_{\text{gbacc}}\right) \times W_{\text{f}} \tag{5.16}$$

$$I_{\text{gbd}} = \left(I_{\text{gbinv}} + I_{\text{gbacc}}\right) \times W_{\text{r}}, \tag{5.17}$$

where

$$W_{\text{f}} = \frac{1}{2} + \frac{1}{2} \times \tanh\left(\frac{0.6 \times q \times V_{\text{ds}}}{kT}\right) \tag{5.18}$$

and

$$W_{\text{r}} = \frac{1}{2} - \frac{1}{2} \times \tanh\left(\frac{0.6 \times q \times V_{\text{ds}}}{kT}\right). \tag{5.19}$$

### 5.3.4 GATE-TO-CHANNEL TUNNELING CURRENT IN INVERSION

As shown in Figure 5.8, in inversion, the electrons (holes in PMOS) tunnel from the inversion channel into the conduction band of the gate (valance band for PMOS). This results in different values of $A$ and $B$ for NMOS and PMOS:

$$A = \begin{cases} 4.97232 \times 10^{-7}\,(\text{A/V}^2) & \text{for NMOS,} \\ 3.42536 \times 10^{-7}\,(\text{A/V}^2) & \text{for PMOS} \end{cases} \tag{5.20}$$

and

$$B = \begin{cases} 7.45669 \times 10^{11}\left(\frac{\text{g}}{\text{Fs}^2}\right)^{0.5} & \text{for NMOS,} \\ 1.16645 \times 10^{12}\left(\frac{\text{g}}{\text{Fs}^2}\right)^{0.5} & \text{for PMOS.} \end{cases} \tag{5.21}$$

The auxiliary function $V_{\text{aux,igc}}$ can be shown to be

$$V_{\text{aux,igc}} = V_{\text{ox}}/V_{\text{ge}} \times \left(V_{\text{ge}} - 0.5{\cdot}V_{\text{dsx}} + 0.5{\cdot}V_{\text{es}} + 0.5{\cdot}V_{\text{ed}}\right). \tag{5.22}$$

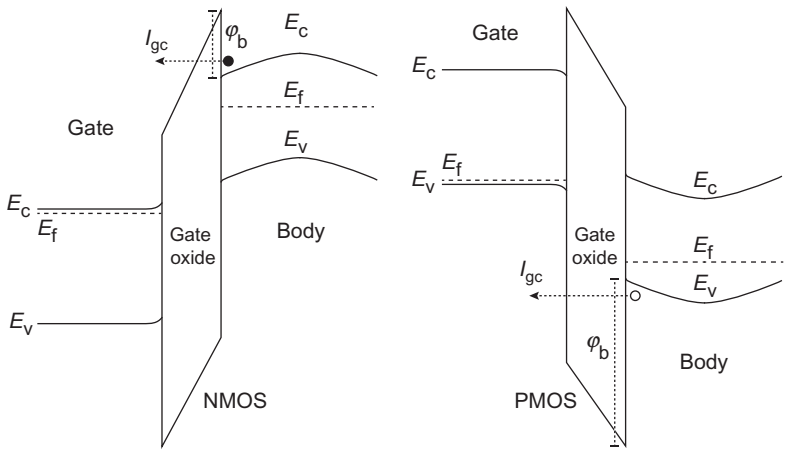


**FIGURE 5.8**

In inversion for NMOS, conduction-band electrons tunnel from the channel into the gate, whereas in PMOS, the valance-band holes tunnel from the channel into the gate.

The total gate-to-channel tunneling component at zero $V_{ds}$ can be written as

$$I_{gc0} = \text{NFIN}_{total} \times W_{eff} \times L_{eff} \times A \times \left(\frac{\text{TOXREF}}{\text{TOXG}}\right)^{\text{NTOX}} \times \frac{V_{ge} \times V_{aux,igcv}}{\text{TOXG}^2} \times e^{-B\times(\text{AIGC}(T)-\text{BIGBC}.q_{ia})\times(1+\text{CIGC}.q_{ia})\times\text{TOXG}}. \quad (5.23)$$

To consider the drain bias effect, a current continuity equation is solved analytically along the channel which extends $I_{gc0}$ to nonzero $V_{ds}$ and splits it into two components, $I_{gcs}$ and $I_{gcd}$. For a detailed discussion on the derivation of this physical current partitioning factor, refer to [8]. The expressions for $I_{gcs}$ and $I_{gcd}$ are as follows:

$$I_{gcs} = I_{gc0}\frac{\text{PIGCD} \times |V_{dseff}| + e^{(-\text{PIGCD}.V_{dseff})} - 1}{\text{PIGCD}^2 \times V_{dseff}^2}, \quad (5.24)$$

$$I_{gcd} = I_{gc0}\frac{(\text{PIGCD} \times |V_{dseff}| + 1) \times e^{(-\text{PIGCD}.V_{dseff})}}{\text{PIGCD}^2 \times V_{dseff}^2}, \quad (5.25)$$

where PIGCD is a fitting parameter added for flexibility with a default value of unity.

### 5.3.5 GATE-TO-SOURCE/DRAIN TUNNELING CURRENT

The $n^+$ ($p^+$) gate to $n^+$ ($p^+$) source and drain currents $I_{gs}$ and $I_{gd}$ are principally caused by the tunneling of the conduction-band electrons in NMOS and valance-band holes in PMOS as shown in Figure 5.9. In NMOS, the electrons tunnel from the conduction band of the body into the gate. In PMOS, the holes tunnel from the valance band of the body into the gate.

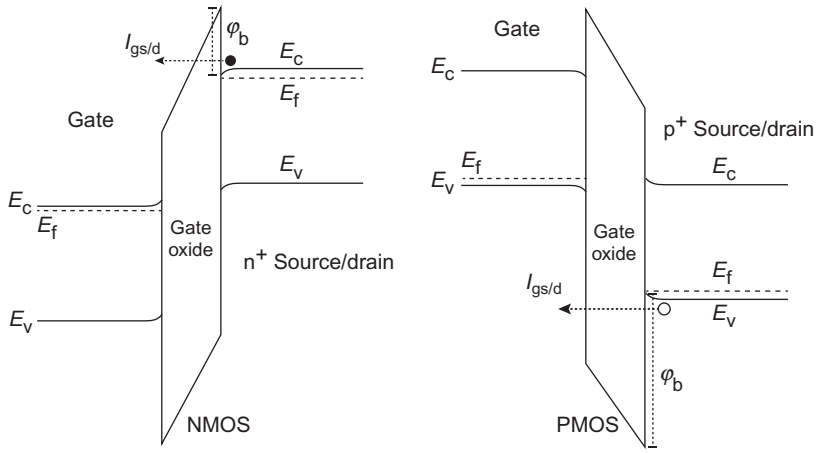


**FIGURE 5.9**

In NMOS the tunneling of the conduction-band electrons and in PMOS the tunneling of the valance-band holes make the source/drain-gate overlap leakage. The figure shows the band diagrams for an inverted channel. In accumulation, the direction of the tunneling is reversed.

For this case, the parameters $A$ and $B$ are naturally equal to those given by Equation (5.20) and Equation (5.21), respectively. If the gate material is a metal, $V_{aux}$ is also simplified to be equal to $|V_{gs}|$ and $|V_{gd}|$ for $I_{gs}$ and $I_{gd}$, respectively.

The total gate-to-source extension tunneling component is

$$I_{gs} = \mathrm{NFIN}_{total} \times W_{eff} \times \mathrm{DLCIGS} \times A \times \left(\frac{\mathrm{TOXREF}}{\mathrm{TOXG} \times \mathrm{POXEDGE}}\right)^{\mathrm{NTOX}} \times \frac{V_{gs} \times |V_{gs}|}{(\mathrm{TOXG} \times \mathrm{POXEDGE})^2} \times e^{-B \times (\mathrm{AIGS}(T) - \mathrm{BIGS}.|V_{gs}|) \times (1 + \mathrm{CIGS}.|V_{gs}|) \times \mathrm{TOXG} \times \mathrm{POXEDGE}}. \quad (5.26)$$

In Equation (5.26), DLCIGS is the length of the gate-source overlap region and POXEDGE is a factor for the gate oxide thickness in the source/drain extension regions.

Similarly, for the total gate-to-drain extension tunneling component, we have

$$I_{gd} = \mathrm{NFIN}_{total} \times W_{eff} \times \mathrm{DLCIGD} \times A \times \left(\frac{\mathrm{TOXREF}}{\mathrm{TOXG} \times \mathrm{POXEDGE}}\right)^{\mathrm{NTOX}} \times \frac{V_{gd} \times |V_{gd}|}{(\mathrm{TOXG} \times \mathrm{POXEDGE})^2} \times e^{-B \times (\mathrm{AIGD}(T) - \mathrm{BIGD} \times |V_{gd}|) \times (1 + \mathrm{CIGD} \times |V_{gd}|) \times \mathrm{TOXG} \times \mathrm{POXEDGE}}, \quad (5.27)$$

where DLCIGD is the length of the gate-drain overlap region.

## 5.4 IMPACT IONIZATION

In the transistor on state, because of the high electric field near the drain end of the channel, carriers in this region can gain enough kinetic energy to ionize the lattice atoms when they collide. This collision frees an electron from the valance-band and leaves a hole behind. The generated hole will drift to the substrate and it will increase the substrate leakage. The released high-energy electron (hot carrier) is collected by the drain and it will be a part of the drain current. Also, there is chance that the generated hot electron travels along the gate field and penetrates into the gate oxide. Hot carrier injection into the gate oxide over time can damage the oxide and cause reliability problems.

The local impact ionization current $I_{ii}(y)$ can be written as a function of the channel current (increase in the number of carriers will increase the chance of collisions) and the strength of the local electric field (the stronger the electric field, the higher the kinetic energy of the carriers) as follows:

$$I_{ii}(y) = I_{ds}.A_i e^{-B_i/E_l(y)} \tag{5.28}$$

where $A_i$ and $B_i$ are two material constants and represent how often the impact ionization events take place and the critical field to trigger the events, respectively, and $E_l(y)$ is the longitudinal electric field along the transport direction. By integrating Equation (5.28) along the length of the channel where velocity saturation happens, we can write the total impact ionization current as

$$I_{ii} = I_{ds}.A_i \int_{y=0}^{y=\acute{l}} e^{-B_i/E_l(y)} dy, \tag{5.29}$$

where $y = 0$ is the starting point of the velocity saturation region and $\acute{l}$ is the length of this region. The integration in Equation (5.29) can be performed (see [9, Chapter 4] for details) to give the following equation for the BSIM-CMG impact ionization model:

$$I_{ii} = \frac{A_i}{B_i}.I_{ds}.(V_{ds} - V_{dsat}).e^{(-B_i.\lambda)/(V_{ds}-V_{dsat})}, \tag{5.30}$$

where $V_{dsat}$ is the saturation voltage and $\lambda$ is the characteristic length (see Chapter 4). The first impact ionization model (IIMOD = 1) implements Equation (5.30) as

$$I_{ii} = \left(\text{ALPHA1} + \frac{\text{ALPHA0}}{L_{eff}}\right).I_{ds}.(V_{ds} - V_{dseff}).e^{-\text{BETA0}/(V_{ds}-V_{dseff})}. \tag{5.31}$$

In Equation (5.31), $V_{dseff}$ is the effective drain voltage resulting from the smooth transition of $V_{ds}$ to $V_{dsat}$ (see Chapter 4), ALPHA1 and BETA0 are a fitting parameter, and the term ALPHA0/$L_{eff}$ has been introduced to improve the length dependence of $I_{ii}$ over a wide range of channel lengths.

There are approximations involved in deriving Equations (5.30) and (5.31), including a linear dependence of $E_l(y)$ on $(V_{ds} - V_{dsat})$. The BSIM-CMG's second

impact ionization model can be activated (IIMOD = 2) and used if a more flexible model is needed:

$$I_{\text{ii}} = \left(\text{ALPHA1} + \frac{\text{ALPHA0}}{L_{\text{eff}}}\right).I_{\text{ds}}.\exp\left(\frac{V_{\text{diff}}}{\text{BETAII2} + \text{BETAII1}.V_{\text{diff}} + \text{BETAII0}.V_{\text{diff}}^2}\right), \tag{5.32}$$

$$V_{\text{diff}} = V_{\text{ds}} - V_{\text{dsatii}}, \tag{5.33}$$

$$V_{\text{dsatii}} = V_{\text{gsStep}}\left(1 - \frac{LII}{L_{\text{eff}}}\right), \tag{5.34}$$

$$V_{\text{gsStep}} = \left(\frac{\text{ESATII}.L_{\text{eff}}}{1 + \text{ESATII}.L_{\text{eff}}}\right)\left(\frac{1}{1 + \text{SII1}.V_{\text{gsfbeff}}} + \text{SII2}\right)\left(\frac{\text{SII0}.V_{\text{gsfbeff}}}{1 + \text{SIID}.V_{\text{ds}}}\right). \tag{5.35}$$

Here, BETAII0, BETAII1, BETAII2, and SIID are parameters for $V_{\text{ds}}$-dependence, LII is a channel-length-dependent parameter, SII0, SII1, and SII2 are $V_{\text{gs}}$-dependent fitting parameters, and ESATII is the channel saturation field with the default value of $1 \times 10^7$ V/m.

## REFERENCES


[1] H. Kam, V. Pott, R. Nathanael, J. Jeon, E. Alon, T.-J.K. Liu, Design and reliability of a micro-relay technology for zero-standby-power digital logic applications, Electron Devices Meeting (IEDM), 2009 IEEE International, 7–9 December 2009, pp. 1–4.

[2] K. Jeon, W.Y. Loh, P. Patel, et al., Si tunnel transistors with a novel silicided source and 46 mV/dec swing, in: VLSI Symp. Tech. Dig., June 2010, pp. 121–122.

[3] C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, P. Bai, A 22 nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications, Electron Devices Meeting (IEDM), 2012 IEEE International, 10–13 December 2012, pp. 3.1.1–3.1.4.

[4] T.-Y. Chan, J. Chen, P.-K. Ko, C. Hu, The impact of gate-induced drain leakage current on MOSFET scaling, Electron Devices Meeting, 1987 International, vol. 33, 1987, pp. 718–721.

[5] C.C. Hu, Modern Semiconductor Devices for Integrated Circuits, Prentice Hall, Upper Saddle River, 2009.

[6] C. Auth, M. Buehler, A. Cappellani, C.-H. Choi, G. Ding, W. Han, S. Joshi, B. McIntyre, M. Prince, P. Ranade, J. Sandford, C. Thomas, 45 nm high-k + metal-gate strain-enhanced transistors, Intel Technol. J. 12 (2) (2008) 77–86.

[7] W.-C. Lee, C. Hu, Modeling gate and substrate currents due to conduction- and valence-band electron and hole tunneling [CMOS technology], 2000 Symposium on VLSI Technology. Digest of Technical Papers, 13–15 June 2000, pp. 198–199.

[8] K.M. Cao, W.-C. Lee, W. Liu, X. Jin, P. Su, S.K.H. Fung, J.X. An, B. Yu, C. Hu, BSIM4 gate leakage model including source-drain partition, Electron Devices Meeting, 2000. IEDM'00. Technical Digest International, 10–13 December 2000, pp. 815–818.
[9] W. Liu, C.C. Hu, BSIM4 and MOSFET Modeling for IC Simulation, World Scientific, Singapore, 2011.