Measurement and Modeling of Short-Channel MOS Transistor Gate Capacitances

BING J. SHEU, MEMBER, IEEE, AND PING-KEUNG KO, MEMBER, IEEE

Abstract — A flexible electrometer method for measuring the gate capacitances of small-geometry MOS transistors is described. This technique applies to standard test transistors without requiring any on-chip circuitry. Subfemtofarad accuracy and high resolution (better than 0.1 fF) have been achieved. This technique permits flexibility with regard to choices of dc biases and test devices and provides a good means of monitoring capacitance changes in the device reliability studies. The measured gate capacitances show prominent short-channel effects. A short-channel MOS transistor capacitance model has been developed which takes into account the mobility-degradation effect, velocity-saturation effect, bias-dependent fringing-field effect, as well as source–drain series resistance effect. Good agreement between the measured and modeled results is found. This model can be easily modified for circuit-simulation applications.

I. INTRODUCTION

Precise characterization of gate capacitances of small-geometry MOS transistors is of paramount importance in the design of high-performance MOS VLSI circuits. Recent progress in integrated-circuit technologies has advanced device geometries into the submicrometer range. The near-micrometer and submicrometer transistors show prominent short-channel effects. Knowledge of small-geometry effects is indispensable for the optimization of device structures and circuit performance.

Several MOS capacitance measurement methods were reported in the literature. An open-loop on-chip circuitry approach was pioneered by Iwai et al. [1], [2], while a closed-loop on-chip circuitry approach was proposed by Paulos et al. [3], [4]. The fact that special circuits need to be fabricated near the devices of interest has prevented these methods from being widely adopted. Such methods require a fairly mature fabrication process to support the design of a high-performance operational amplifier. It imposes a severe limitation in the early stages of process development. This paper describes a versatile electrometer method to characterize MOS transistor capacitances. No on-chip circuitry is needed. With this method, subfemtofarad accuracy and high resolution (better than 0.1 fF) have been achieved. The measurement technique permits flexibility with regard to choices of dc biases and test devices.

The measured MOS transistor capacitances show prominent small-geometry effects due to the proximity of various elements of the device. An analytical model has been developed which takes into account the various short-channel effects. This model includes the vertical-field mobility degradation effect, velocity-saturation effect, bias-dependent fringing-field effect, as well as the source–drain series resistance effect.

II. CAPACITANCE MEASUREMENT

A. The Bridge Method

The Wheatstone-bridge method is commonly used in resistance measurement. Balance is achieved by adjusting the resistance in one arm so that no current flows through the detector. For the measurement of capacitance and inductance, an ac bridge is required. Fig. 1 shows the schematic diagram of a general ac bridge which is equivalent to the Wheatstone bridge except that the arms consist of impedance elements rather than resistance elements. The dc voltage source is replaced by an ac signal source, and the detector must be able to detect ac signals. The condition of balance is

$$\frac{Z_1}{Z_2} = \frac{Z_3}{Z_C}. \quad (1)$$

The traditional bridge method is suitable for two-terminal elements.

B. The Four-Port Nulling Method

To measure four-terminal MOS transistor capacitances, the four-port nulling method [5], [6] has been used because an LCR meter or impedance analyzer is sufficient. The four-port nulling method is typically applied to large (100 \times 100 \, \mu m^2 or larger) devices. The test instrument inter-
Fig. 1. The generalized ac bridge circuit.

Fig. 2. Equivalent circuit for interelectrode capacitance measurement using an impedance analyzer [5].

Fig. 3. Open-loop capacitance measurement technique of Iwai et al. [1]. The switch is used to establish the gate bias.

Fig. 4. The ac equivalent circuit for the open-loop method shown in Fig. 3.

Fig. 5. The coulombmeter circuit [4]. The switch is used to establish the gate bias.

faces to the outside world through four measurement ports [7]: low-current, low-potential, high-current, and high-potential ports. Fig. 2 shows the equivalent circuit for the MOS transistor interelectrode capacitance measurement. Both the connections of the high-potential port to the high-current port and the low-potential port to the low-current port should be made as close to the physical device as possible in order to minimize the effect of stray capacitance on the measurement results. The ac test signal is applied to the transistor under test via the high-current port. The second signal source in the low-current port is adjusted to produce a null condition at the low terminal. The magnitude of the ac test signal is monitored at the high-potential port, while the response information is sensed at the low-potential port.

C. An Open-Loop On-Chip Circuitry Method

To characterize capacitances of small-geometry MOS transistors, innovation in the measurement schemes is needed. An open-loop technique using on-chip circuits to measure small parasitic capacitances was first reported by Iwai and Kohyama [8]. This technique has been extended to measure gate capacitances of MOS transistors. Fig. 3 shows one of the three circuit configurations necessary to determine the gate capacitances. The ac equivalent circuit is shown in Fig. 4. By subsequently applying an ac test signal to the drain, source, and substrate terminals, the gate-to-drain $C_{gd}$, gate-to-source $C_{gs}$, and gate-to-bulk $C_{gb}$ capacitances can be solved after some mathematical operations [2].

The open-loop capacitance measurement technique has severe drawbacks which makes it inferior to a closed-loop technique [4]. The input capacitance of the amplifier, which is strongly bias dependent, affects the reference capacitance value. The switch in Fig. 3 is used to establish the dc bias on the floating gate and to calibrate the actual ac grain of the output-buffer circuit. Charge injection from the switch when it turns off perturbs the dc bias of the floating gate. Complicated compensation is required to achieve the desired gate bias.

D. A Closed-Loop Coulombmeter Method

A closed-loop measurement method, as was proposed by Paulos et al. [3], [4], shows some improvements over the open-loop approach. The coulombmeter circuit is used. A simplified schematic diagram of the method is shown in Fig. 5. Drain, source, and bulk biases are directly applied to the transistor terminals, while the gate bias is applied through the noninverting node of the amplifier. A switch is required to periodically establish the dc bias for the gate terminal. Similar to the open-loop technique, it suffers from the charge-injection problem caused by the switch.
TABLE I

A COMPARISON OF SMALL-GEOMETRY MOS TRANSISTOR CAPACITANCE MEASUREMENT METHODS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>transistors, capacitors plus amplifier</td>
<td>transistors, capacitors plus amplifier</td>
<td>transistors only</td>
</tr>
<tr>
<td>test structures</td>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>switch charge injection effect</td>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Cgd, Cgs, Cgb measured independently</td>
<td></td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>sensitive to amplifier input capacitance</td>
<td></td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>require mature process to support amplifier design</td>
<td></td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>applicable process</td>
<td></td>
<td>analog</td>
<td>analog</td>
<td>analog and digital</td>
</tr>
</tbody>
</table>

The closed-loop approach has the following advantages over the open-loop approach. First, any of the three gate capacitances $C_{gd}$, $C_{gs}$, and $C_{gb}$ can be independently measured by applying the ac test signal to the corresponding terminal. Second, the measured result is insensitive to the parasitic capacitance appearing at the summing node of the operational amplifier because of the virtual-ground operation of the amplifier, which keeps the summing-node voltage fixed. Third, one coulombmeter circuit can be physically shared by several test transistors as long as only one transistor is measured at a time. The parasitic capacitances associated with additional test transistors will not degrade the measured accuracy because of the virtual ground operation of the operational amplifier.

Both the open-loop and closed-loop on-chip circuitry approaches are limited by the following factors. A fairly mature fabrication process is needed and a large area is consumed by the circuits as compared to the test transistors. Dedicated masks are needed for the fabrication of these test structures. The performance of the operational amplifier depends highly on each specific fabrication process being used.

III. A FLEXIBLE ELECTROMETER METHOD FOR CAPACITANCE MEASUREMENT

The electrometer method [9] described in this section is superior to the other methods in many respects. It applies to standard transistors. No on-chip circuitry is needed. A comparison of the various small-geometry MOS transistor capacitance measurement methods is summarized in Table I.

A. Measurement System

A schematic diagram of the measurement setup is shown in Fig. 6. The system contains a Hewlett-Packard (HP) 4145 parameter analyzer, under the control of an HP desktop computer. The HP 4145 analyzer supplies programmed dc biases to the test device. The ac test signal is
supplied by an HP signal generator or can be derived from the built-in oscillator of the lock-in amplifier. The coupling transformer and the buffer amplifier convert the primary test signal into a truly floating low-impedance source. In order to achieve high resolution, the voltage level of the ac test signal is adjusted to 50 mV. Configurations for measuring various gate capacitances are selectable with a switching matrix. A high-performance current-to-voltage ($I-V$) converter is built using a high-gain wide-band operational amplifier from Precision Monolithic, Inc. The frequency of the test signal is set at 10 kHz to enhance the measurement resolution. Phase shift due to cable capacitance is insignificant at such low frequencies. The $I-V$ converter connected to the gate terminal converts the gate capacitive current into voltage, which is detected by the lock-in amplifier. An EG&G lock-in amplifier is used. The HP 4145 analyzer monitors the output of the lock-in amplifier and transfers the data to the controller for storage and plotting. The function of the lock-in amplifier in this measurement method is similar to that described in [10] for the measurements of MOS capacitor characteristics. The conversion factor of the $I-V$ converter is chosen as $10^8$ V/A.

**B. Measurement Method**

There are three different gate-related capacitances: the gate-to-source $C_{gs}$, gate-to-drain $C_{gd}$, and gate-to-substrate $C_{gs}$ capacitances. Capacitance $C_{gs}$ is a measure of the change in the gate charge in response to the change in the gate-to-source voltage only. Therefore, the ac test signal is applied to the source terminal, with the gate-to-drain and gate-to-substrate voltages kept constant. Similarly, the ac test signal is applied to the drain and substrate terminals for $C_{gd}$ and $C_{gs}$ capacitance measurements, respectively. The input of the $I-V$ converter is connected to the gate terminal of the test device. The connection cable does not need to be specifically shortened due to the advantage of having the virtual ground operation of the $I-V$ converter. One important feature about this measurement technique is that capacitances $C_{gs}$, $C_{gd}$, and $C_{gb}$ are measured independently. Full-bias-range characterization can be easily achieved by varying the dc biasing voltages $V_{gs}$, $V_{ds}$, and $V_{hs}$.

The measurement system can be calibrated with either high-precision standard capacitors or LCR meters. Fig. 7 shows the external capacitances and parasitic capacitances encountered in the $C_{gs}$ and $C_{gd}$ measurements. External capacitances $C_{gse}$ and $C_{gde}$ are the coupling capacitances between the gate probe and the source and drain probes. Parasitic capacitances $C_{ps}$ and $C_{pd}$ include the overlap capacitances, outer fringing-field capacitances, and metallization capacitances associated with the pads at the source and the drain. The capacitances $C_{gse} + C_{ps}$ and $C_{gde} + C_{pd}$ can be obtained from the measured gate-to-source and gate-to-drain capacitances, respectively, when the device is biased in the strong accumulation region. Such measured capacitances do not contain any contribution from the channel area. Since the external and parasitic capacitances are not strongly bias dependent, the values measured in the accumulation region can also be applied to the strong-inversion region to the first order. The intrinsic capacitances presented in this paper are obtained by subtracting the external and parasitic capacitances from the raw measured data.

The parasitic capacitances scale with device dimensions whereas the external capacitances do not scale. The probe-coupling capacitance poses a fundamental limitation to this measurement technique. Therefore it is essential to keep the external capacitances $C_{gse}$ and $C_{gde}$ as small as possible in order to achieve high measurement accuracy. The coupling capacitance between two ordinary probes is about several picofarads so that accurate measurement of the small-geometry MOS transistor capacitances is impossible. If the probe is teflon shielded, the probe-coupling capacitance is reduced to the subpicofarad range. However, it is still too large to make accurate measurements. By using the coaxial probes, the probe-coupling parasitic capacitance can be reduced down to about 5 fF. Fig. 8 illustrates the resolution of this method with the raw measured data for an NMOS transistor with $W/L = 0.5 \mu m/1.25 \mu m$. High resolution better than 0.1 fF has been achieved.
Fig. 9. Normalized plots of measured $C_{gs}$ and $C_{gd}$ capacitances for a short-channel transistor with $W = 49.2 \mu m$ and $L = 1.0 \mu m$. The gate–source voltage increases in 1-V steps. Capacitance $C_{gd} + C_{gs}$ at $V_{ds} = 0$ V is 84 fF. The external capacitances $C_{gs}$ and $C_{gd}$ are 5 fF in the measurement.

Fig. 10. Normalized plots of measured $C_{gs}$ and $C_{gd}$ capacitances for a long-channel transistor with $W = 99.2 \mu m$ and $L = 99.5 \mu m$. The gate–source voltage increases in 1-V steps. Capacitance $C_{gd} + C_{gs}$ at $V_{ds} = 0$ V is 16.9 pF. The external capacitances $C_{gs}$ and $C_{gd}$ are 5 fF in the measurement.

IV. SHORT-CHANNEL EFFECTS

The devices used in the experiments are n-channel MOS transistors. The gate-oxide thickness is 20.0 nm, substrate doping is $10^{15}$ cm$^{-3}$, and enhancement threshold implant is $8 \cdot 10^{11}$ cm$^{-2}$. Fig. 9 shows the normalized plot of the gate-to-source $C_{gs}$ and gate-to-drain $C_{gd}$ capacitances of a short-channel transistor with $W/L = 49.2 \mu m/1.0 \mu m$. The measured capacitances of a long-channel transistor with $W/L = 99.2 \mu m/99.5 \mu m$ are shown in Fig. 10 for comparison. In these plots, the external and parasitic capacitances have been subtracted. At zero drain–source bias, capacitances $C_{gd}$ and $C_{gs}$ are the same for both long-channel and short-channel devices, as expected. This means that the drain and source have equal amounts of control over the channel charge.

For a long-channel device, capacitance $C_{gd}$ approaches zero and capacitance $C_{gs}$ approaches $2/3 C_{ox}W/L$ as the device operation enters the saturation region. Sharp transitions between the triode and saturation regions can be found. This behavior is well known and modeled [11].

For a short-channel device, several unusual features are observed. First, due to the velocity-saturation effect, the $C_{gd}$ and $C_{gs}$ curves are much smoother than the long-channel device curves. No sharp transition from the triode region to the saturation region can be identified. Second, the $C_{gd}$ curves split in the saturation region for different gate biases. Both velocity saturation and source–drain series resistance contribute to this splitting, with the source–drain resistance being the dominant factor. Third, capacitance $C_{gd}$ saturates to a finite value in the saturation region. This behavior is due to the channel-side fringing fields between the gate and the drain. The magnitude of this fringing capacitance is a strong function of the drain bias, drain-junction depth, and gate-oxide thickness. Hence, it is more important in shorter devices. The measured results for transistors with $W = 49 \mu m$ and effective channel lengths of 4.5 and 0.75 $\mu m$ are shown in Fig. 11. Notice that the limiting saturated values for capacitance $C_{gd}$ are approximately 4 and 14 percent of $C_{ox}W/L$ for these two devices, respectively. Because of the Miller effect, the influence of channel-side fringing capacitance on the performance of high-speed MOS circuits is readily noticeable experimentally.

V. THE SHORT-CHANNEL MODEL

An analytical model for the short-channel MOS transistor capacitances has been developed. The special effects included are:

a) vertical-field mobility-degradation effect;
b) velocity-saturation effect;
c) bias-dependent fringing-field effect; and
d) source–drain series resistance effect.
Fig. 12. Schematic illustration of the three components of the charge stored in the gate when the transistor is biased in the saturation region. $Q_{G1}$: the charge in the source region. $Q_{G2}$: the charge in the drain region. $Q_{G3}$: the charge associated with the channel-side bias-dependent fringing fields.

The analytical capacitance model is based on a hot-electron current model [12] which includes a pseudo two-dimensional analysis of the drain region. In the model, a continuous relationship between carrier velocity and electric field is adopted. A brief description of the drain-current formulation can be found in the Appendix.

When a transistor is biased in the saturation region, the gate charge is made up of three components: $Q_{G1}$, $Q_{G2}$, and $Q_{G3}$. Fig. 12 shows a schematic illustration of the various charge components. The gradual-channel approximation can be applied to the channel near the source terminal. The charge in this portion is designated as $Q_{G1}$. In the drain region, the velocity of mobile carriers saturates. The charge in this portion is designated as $Q_{G2}$. The charge associated with the channel-side fringing-field effect is designated as $Q_{G3}$. Only charge component $Q_{G1}$ has been incorporated into conventional MOS transistor models [5],[13],[14]. The gate-charge derivatives give the gate capacitances

$$C_g = \frac{\partial Q_G}{\partial V_g} \bigg|_{V_{gs},V_{gd}}$$

$$C_{gd} = \frac{\partial Q_G}{\partial V_{gd}} \bigg|_{V_{gs},V_{gd}}$$

and

$$C_{gb} = \frac{\partial Q_G}{\partial V_{gb}} \bigg|_{V_{gs},V_{gd}}$$

The expression for $Q_{G1}$ can be obtained by integrating the distributed charge density over the area of the active gate region:

$$Q_{G1} = \int_0^L Q_n(y) dy = WC_{ox} \int_0^L \left[ V_{gs} - V_T - V(y) \right] dy$$

$$= WC_{ox} \int_0^L \left[ \frac{V_{gs} - V_T - V(y)}{E_y} \right] dV$$

where the integration with respect to the positional increment $dy$ has been converted to an integration with respect to the potential increment $dV$. Let

$$I_{dn} = \frac{I_{ds}}{W\mu_0 C_{ox}} = \frac{(V_{gs} - V_T - V)}{E_y} \frac{I_{dn}}{E_y + \eta E_x}.$$  \hspace{1cm} (4)

Here $E_c$ is the critical field for the velocity-saturation effect, $\eta$ is the vertical-field mobility-degradation coefficient, $\mu_0$ is the low-field mobility, $C_{ox}$ is the thin-oxide capacitance per unit area, and $I_{ds}$ is the drain current. Rearranging (4) we obtain

$$\frac{1}{E_y} = \frac{1}{(1 + \eta E_y)} \left[ \frac{V_{gs} - V_T - V}{I_{dn}} - \frac{1}{E_c} \right].$$  \hspace{1cm} (5)

By substituting (5) into (3) and carrying out the integration, it becomes

$$Q_{G1} = \frac{WC_{ox}}{(1 + \eta E_y)} E_c I_{dn} \int_0^L \left[ (V_{gs} - V_T - V) E_c - I_{dn} \right]$$

$$\cdot (V_{gs} - V_T - V) dV$$

$$= \frac{WC_{ox}}{(1 + \eta E_y)} E_c I_{dn} \left[ \frac{E_c}{3} \left( (V_{gs} - V_T)^3 \right. \right.$$\n
$$\left. - (V_{gs} - V_T - V_{ds})^3 \right]$$

$$- \frac{I_{dn}}{2} \left[ (V_{gs} - V_T)^2 - (V_{gs} - V_T - V_{ds})^2 \right].$$  \hspace{1cm} (6)

If the source–drain series resistance effect is included, the expression becomes

$$Q_{G1} = \frac{WC_{ox}}{(1 + \eta E_y)} E_c I_{dn} \left[ \frac{E_c}{3} \left( (V_{gs} - V_T - I_{ds} R_s)^3 \right. \right.$$\n
$$\left. - (V_{gs} - V_T - V_{ds} + I_{ds} R_s)^3 \right]$$

$$- \frac{I_{dn}}{2} \left( V_{gs} - V_T - I_{ds} R_s \right)^2$$

$$- (V_{gs} - V_T - V_{ds} + I_{ds} R_s)^2 \right].$$  \hspace{1cm} (7)

When the device is biased in the saturation region, $V_{ds}$ in the $Q_{G1}$ expression is replaced by the saturation voltage $V_{dsat}$. Accurate determination of the saturation voltage is crucial in modeling the dc characteristics as well as the capacitance characteristics. It is important to obtain the length of the drain region inside which the carrier velocity saturates. A pseudo two-dimensional analysis is used to find the channel electric field and potential distribution by solving the Poisson equation for the Gaussian box enclosing the drain region [12]. Once the electric field and potential drop inside the drain region is available, the length of the drain region $L_d$ can be easily obtained. Because the carriers travel at the same saturated velocity in the drain region, the multiplication of the charge density with the drain-region length gives

$$Q_{G3} = \frac{I_{dsat}}{v_{sat}} \cdot L_d$$  \hspace{1cm} (8)

where $v_{sat}$ is the carrier saturation velocity. Detailed derivation of the $L_d$ expression can be found in the Appendix.
The bias-dependent fringing-field capacitance associated with \( Q_{G3} \) can be determined by considering the device operation in the drain region. When the device is biased in the strong-inversion region and \( V_{ds} = 0 \), the fringing field is shielded by channel inversion charge. In this case, the drain terminal has a strong control over the channel, and

\[
C_{gdc} = \frac{\partial (Q_{G1} + Q_{G2})}{\partial V_{gd}} \bigg|_{V_{gs}, V_{gb}} = \frac{C_{ox}WL}{2}. \tag{9}
\]

In the saturation region, the drain fringing-field lines can penetrate through the conducting layer and reach the gate terminal. The limiting condition corresponding to nonshielding of the channel-side fringing field is when \( C_{gdc} = 0 \). In this case, the drain has no control over the channel. An analytical expression for the unshielded fringing-field capacitance is given by [15]

\[
C_f = \frac{2 \varepsilon Si}{\pi} \ln \left[ 1 + \frac{X}{T_{ox}} \sin \left( \frac{\pi \varepsilon Si}{2 \varepsilon Si} \right) \right]. \tag{10}
\]

For \( V_{ds} > 0 \), the control by the drain over the channel degrades. Linear interpolation is found to be sufficient for the modeling of the bias-dependent fringing-field effect:

\[
C_{gdr, fringing} = C_f \left( 1 - \frac{2C_{gdc}}{C_{ox}WL} \right). \tag{11}
\]

The two limiting conditions, when \( C_{gdc} = C_{ox}WL/2 \) and when \( C_{gdc} = 0 \), are both satisfied by (11).

VI. RESULTS AND DISCUSSION

Figs. 13 and 14 show the modeled \( CV \) characteristics for a short-channel transistor and a long-channel transistor with the same device parameters as those for the transistors used in Figs. 9 and 10. Good agreement between the modeled results and measured data is found.

When compared to the already published charge and capacitance models for long-channel devices, the salient features of this short-channel transistor model are as follows.

1) The vertical-field and horizontal-field mobility degradation effects are incorporated in the expression for \( Q_{G1} \).

2) The velocity-saturation effect and the channel-length modulation effect have been taken into account in the modeling of the charge component \( Q_{G2} \) through the calculation of the drain-region length \( L_d \).

3) The bias-dependent fringing-field effect is incorporated. The ratio of the charge component \( Q_{G3} \) to the total gate capacitance increases as channel length decreases.

4) The source–drain series resistance effect is included.

VII. CONCLUSION

A flexible electrometer method for measuring the gate capacitances of small-geometry transistors has been described. This direct-on-wafer measurement technique is very suitable for the purposes of device physics study and process control applications and can be easily incorporated into an automated characterization system. It applies to standard test transistors without requiring any on-chip circuitry. Very high accuracy and resolution have been achieved. An analytical model which satisfactorily explained the various features observed in the gate capacitances of short-channel MOS transistors has also been presented. The inclusion of the mobility-degradation effect, velocity-saturation effect, bias-dependent fringing-field effect, and source–drain series resistance effect proves to be adequate for accurate prediction of the gate capacitances.

APPENDIX

THE DRAIN-CURRENT FORMULATION

The linear region drain-current expression including the source–drain series resistance effect, vertical-field mobility degradation effect, and velocity-saturation effect can be formulated as

\[
I_{ds} = \frac{\beta}{1 + \frac{V_{ds}}{\eta E_s}} \left( V_{gs}' - V_T - \frac{V_{ds}'}{2} \right) \frac{V_{ds}'}{E_c L}. \tag{12}
\]
where
\[ \beta = \frac{W}{L} C_{ox} \mu_0 \]
\[ E_x = \frac{V_{gs} - V_T}{T_{ox}}. \quad (13) \]

Here \( W \) and \( L \) are the effective channel width and length, respectively. Since \( V_{gs} = V_g - I_d R_s \) and \( V_{ds} = V_d - 2I_d R_s \), (12) can be rewritten in terms of \( V_{gs} \) and \( V_{ds} \):
\[ I_{ds} = \frac{\beta (V_{gs} - V_T - \frac{V_{ds}}{2})(V_{ds} - 2I_d R_s)}{1 + \frac{V_{ds} - 2I_d R_s}{E_c L} + \frac{\eta(V_{gs} - V_T - I_d R_s)}{T_{ox}}}. \quad (14) \]

With \( \eta(V_{gs} - V_T - I_d R_s)/T_{ox} \ll 1 + V_{ds}/E_c L \) for most cases, (14) can be rewritten in a quadratic equation for \( I_{ds} \):
\[ \frac{2R_s}{E_c L} I_{ds}^2 \left[ 1 + \frac{V_{ds}}{E_c L} + \frac{\eta(V_{gs} - V_T)}{T_{ox}} \right] + 2\beta \left( \frac{V_{gs} - V_T - \frac{V_{ds}}{2}}{2} \right) I_{ds} + \beta \left( \frac{V_{gs} - V_T - \frac{V_{ds}}{2}}{2} \right) V_{ds} = 0. \quad (15) \]

The useful solution to (15) is
\[ I_{ds} = \frac{I_0}{2} \left[ \gamma - \left( \gamma^2 - \frac{4I_0}{I_0} \right)^{1/2} \right] \quad (16) \]
where
\[ \gamma = 1 + \frac{V_{ds}}{E_c L} + \frac{\eta(V_{gs} - V_T)}{T_{ox}} + \frac{2I_d R_s}{V_{ds}} \]

and
\[ I_0 = \frac{E_c L}{2R_s}, \]
\[ I_{do} = \beta \left( \frac{V_{gs} - V_T - \frac{V_{ds}}{2}}{2} \right) V_{ds}. \quad (18) \]

The transition point between the triode region and the saturation region can be determined by considering the drain current being limited by carrier saturation velocity
\[ I_{ds} = v_{sat} WC_{ox} (V_{gs} - V_{td} - V_{ds}) \quad (19) \]

where \( V_{td} \) is the threshold voltage of the channel at the drain, and \( v_{sat} \) is empirically determined to be 0.93 \( v_{sat} \). In this work, the saturation voltage is determined by equating the drain-current expressions (16) and (19). Iteration is needed in finding the solution for the saturation voltage.

When the transistor is biased in the saturation region, the channel-length shortening effect has to be taken into account. A pseudo two-dimensional analysis of the channel electric field in the drain region is employed. By applying Gauss’ Law to the “drain-region” box, the Poisson equation can be expressed as
\[ \frac{dE_y}{dy} + \frac{\epsilon_{ox}}{\epsilon_{si} T_{ox}} (V_{gs} - V_{FB} - 2\phi_f - V(y)) = \frac{qN_{sub} X'_j}{\epsilon_{si} T_{ox}} + \frac{Q_m}{\epsilon_{si} T_{ox}}. \quad (20) \]

Here \( X'_j \) is the effective drain-region depth. Since
\[ V_{gs} - V_{FB} - 2\phi_f - V_{dsat} = \frac{1}{C_{ox}} (qN_{sub} X'_j + Q_m) \quad (21) \]

(20) can be simplified to
\[ X'_j \frac{dE_y}{dy} = \frac{\epsilon_{ox}}{\epsilon_{si} T_{ox}} [V(y) - V_{dsat}]. \quad (22) \]

It then becomes
\[ \frac{dE_y}{dy} = A^2 [V(y) - V_{dsat}]. \quad (23) \]

The boundary conditions at \( y = 0 \) are \( E_y = E_{sat} \) and \( V = V_{dsat} \). Equation (23) can be solved analytically. The solution for channel-length shortening is
\[ L_d = \frac{1}{A} \sinh^{-1} \left[ \frac{A(V_{ds} - V_{dsat})}{E_{sat}} \right]. \quad (25) \]

ACKNOWLEDGMENT

The authors wish to thank Prof. D. O. Pederson and Prof. D. A. Hodges for their continuous encouragement and support. Valuable suggestions from the reviewers are highly acknowledged.

REFERENCES

Bing J. Sheu (S’81–M’85) was born in 1955. He received the B.S.E.E. degree in 1978 from the National Taiwan University, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1983 and 1985, respectively. From 1979 to 1980 he worked as an Engineer in the Chinese Army. In 1981 he was involved in integrated-circuit design for a speech recognition system at Threshold Technology Inc., Cupertino, CA. From 1982 to 1985 he was a Research Assistant in the Electronics Research Laboratory, University of California, Berkeley, working on device modeling, simulation, and circuit design. In 1985 he joined the faculty in Electrical Engineering at the University of Southern California, Los Angeles, as an Assistant Professor. Since 1983 he has served as a consultant to the electronics industry. His research interests include computer-aided design of analog and digital integrated circuits, and transistor modeling with automated parameter extraction.

Dr. Sheu is a member of Phi Tau Phi.

Ping-Keung Ko (S’78–M’82) was born in Hong Kong in 1951. He received the B.Sc. (special honors) degree in physics in 1974 from the University of Hong Kong, and the Ph.D. degree in electrical engineering in 1982 from the University of California at Berkeley. From 1982 to 1983 he was with the silicon-research group at Bell Laboratories, Holmdel, NJ, where he developed a submicrometer high-speed NMOS technology. In January 1984 he joined the Department of Electrical Engineering and Computer Sciences at the University of California at Berkeley where he is at present an Associate Professor. His research interest is in high-speed integrated-circuit technologies and device modeling for circuit simulation.