Nicolas Sklavos · Michael Hübner
Diana Goehringer · Paris Kitsos *Editors*

System-Level Design Methodologies for Telecommunication

Springer

Nicolas Sklavos • Michael Hübner
Diana Goehringer • Paris Kitsos
Editors

System-Level Design Methodologies for Telecommunication

Springer

Editors
Nicolas Sklavos
Technological Ed Institute of
Patras
Hellas
Greece

Diana Goehringer
Ruhr-Universität Bochum
Bochum
Germany

Michael Hübner
Ruhr-Universität Bochum
Bochum
Germany

Paris Kitsos
Technological Educational Inst of Patras,
Informatics & MM
KNOSSOSnet Research Group
Pyrgos
Greece

ISBN 978-3-319-00662-8 ISBN 978-3-319-00663-5 (eBook)
DOI 10.1007/978-3-319-00663-5
Springer Dordrecht Heidelberg London NewYork

Library of Congress Control Number: 2013947735

Contents

Contributors

D. Amaxilatis Computer Technology Institute and Press "Diofantus", Patras, Greece

George Asimakopoulos Department of Telecommunication Systems and Networks (TESYD), Technological Educational Institute of Messolonghi, Hellas, Greece
e-mail: asim@teimes.gr

Labros Bisdounis Electrical Engineering Department, Technological Educational Institute of Patras, 1, M. Alexandrou Street, 263 34 Patras, Greece
e-mail: bisdounis@teipat.gr

David Fuschelberger Department of Telecommunication Systems and Networks (TESYD), Technological Educational Institute of Messolonghi, Messolonghi, Greece
e-mail: d.fuschelberger@gmail.com

Apostolos P. Fournaris Electrical and Computer Engineering Department, University of Patras, Patras, Greece
e-mail: apofour@ieee.org

Angelina Gkioni Department of Telecommunication Systems and Networks (TESYD), Technological Educational Institute of Messolonghi, Hellas, Greece
e-mail: angegkio@gmail.com

L. Gortzis University of Patras, 26504 Rio, Greece

V. Georgitzikis Computer Technology Institute and Press "Diofantus", Patras, Greece
e-mail: georgitzik@ceid.upatras.gr

Georgios Keramidas Electrical and Computer Engineering Department, University of Patras, Patras, Greece
e-mail: keramidas@ece.upatras.gr

Spiros Louvros Department of Telecommunication Systems and Networks (TESYD), Technological Educational Institute of Messolonghi, Hellas, Greece
e-mail: splouvros@gmail.com

Fotis Plessas Department of Telecommunication Systems and Networks, Technological Education Institute of Messolonghi, Messolonghi, Greece
e-mail: fotis.plessas@gmail.com

Aggeliki Pragiati Hellenic Telecommunications & Post Commission (EETT), Athens, Greece
e-mail: apragiati@eett.gr

Ch. Pylarinou University of Patras, 26504 Rio, Greece

Vassilios Triantafyllou Department of Telecommunication Systems and Networks (TESYD), Technological Educational Institute of Messolonghi, Hellas, Greece
e-mail: triantaf@teimes.gr

Nikolaos Terzopoulos Department of Computing & Communication Technologies, Faculty of Technology, Design & Environment, Oxford Brookes University,1 Wheatley, Oxford, UK

S. Zimeras University of the Aegean, Karlovassi, 83200 Samos, Greece
e-mail: zimste@aegean.gr

Chapter 5
Modeling the Operation of CMOS Primitive Circuits and MOSFET Devices

Labros Bisdounis

Abstract Estimation of complementary metal-oxide semiconductor (CMOS) circuits' behavior, in terms of analysis and computation of their dynamic characteristics (such as propagation delay, transition time, and energy dissipation) is today a standard part of digital circuit design. Since these characteristics are critical design parameters in CMOS digital circuits, much effort has to be devoted for the extraction of accurate, analytical expressions for primitive circuits. Using transistor-level simulators with continuous-time modeling of the devices such as SPICE can be very expensive in terms of storage and computation time. Hence, much of the research has addressed the development of analytical timing and energy dissipation models, without the necessity of expensive numerical iterations. This chapter mainly regards the methodology for the derivation of closed-form, accurate expressions for the aforementioned parameters. The operational conditions of primitive CMOS structures are determined and the differential equations describing their operation are solved analytically by using appropriate approximations in order to simplify the modeling procedure, without significant influence in the accuracy. As a case study, the CMOS inverter is used. Following a detailed analysis of the inverter operation, accurate expressions for its output response are derived for the different operation regions, and based on this analysis, analytical expressions for the calculation of the timing and energy parameters can be produced. The derived models account for the influ ences of input voltage transition time, device sizes, parasitic capacitances, output load, as well as small-geometry device effects. The inverter model can be extended to multi-input CMOS gates by using reduction techniques of series-connected and parallel-connected transistors. Since the accuracy of the used MOSFET device I–V model determines the accuracy of primitive circuits' timing and energy models to a large extent, accurate and compact device models that take into account the influences of predominant effects in modern nanometer device technologies should be adopted.

L. Bisdounis (✉)
Electrical Engineering Department, Technological Educational Institute of Patras,
1, M. Alexandrou Street, 263 34 Patras, Greece
e-mail: bisdounis@teipat.gr

N. Sklavos et al. (eds.), *System-Level Design Methodologies for Telecommunication,*
DOI 10.1007/978-3-319-00663-5_5,

5.1 Introduction

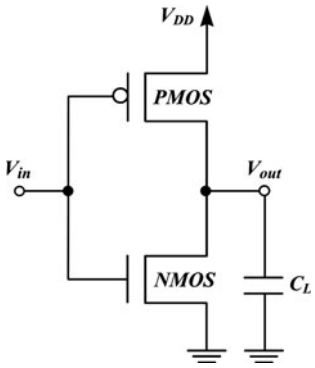
Switching speed is traditionally a critical performance parameter in circuits and systems [1]. In addition, energy dissipation is a very critical parameter that has to be taken into account during circuit design, to avoid problems related to heating in high-performance applications and those related to energy savings in battery-oriented portable applications [2].

Thus, computer-aided design tools should include efficient methods for fast and accurate computation of such parameters, in terms of accurate, analytical expressions for timing and energy models of basic circuits [3–7]. From the practical use point of view, improved analytical timing and energy dissipation models can be applied to various types of simulation (i.e., fast-timing simulation, switch-level and gate-level timing, and mixed-mode simulation), subcircuits' timing/energy analysis and characterization, as well as to transistor sizing. Examples of application of analytical models in fast-timing and gate-level simulation are the ILLIADS [8, 9] and HALOTIS [10] tools, respectively. The main advantage of the analytical modeling approach is that it does not require presimulation as in tabular methods or empirical equation techniques, in order to improve the simulation speed. On the other hand, since analytical models for circuit primitives (like the CMOS inverter) preserve the nonlinearity of the devices, they incorporate the impact of modern technologies' effects resulting in high accuracy, without the need for computational expensive iterations and numerical methods used in SPICE-like circuit simulators.

The methodology for analytical modeling of CMOS primitive circuits' operation is constituted from the following main steps:

- Understanding of the circuit operation.
- Creation of a circuit model including the parasitics to be considered.
- Derivation of the differential equation that describes the circuit operation, i.e., application of the Kirchhoff's current law at the output node of the primitive circuit.
- Definition of the operating regions of the primitive CMOS circuit's transistors in every time interval of the circuit operation.
- Adoption of an appropriate current model for the transistors in the main operating regions (cutoff, linear, saturation). The adopted transistor model should match the I–V characteristics of the transistors by accounting first-order and main second-order effects [11, 12], according to the used technology process. The transistor current model should combine simplicity to provide the ability for derivation of explicit expressions for design parameters, and accuracy to account for the influence of main device physical mechanisms.
- Definition of the circuit operational conditions, which determines the bounds of different operating regions.
- Derivation of the differential equation analytical solution in each region of operation, in order to determine the output voltage waveform of the circuit. The output voltage waveform is determined as an integration of the device currents in contrast to the average current method where the current is assumed equal to the average of its values at the limits of the time interval of interest.

Fig. 5.1 The CMOS inverter



- Since the target is to combine simplicity and accuracy, the methodology should include creative and reasonable assumptions in case the differential equation cannot be solved analytically. Circuit simulations can be used to define candidate regions for approximations.
- Use of smart techniques (such as Taylor series expansions) to solve boundary equations in order to avoid numerical approaches.
- Adoption of efficient techniques for the reduction of complex circuits (i.e., multi-input CMOS gates) to primitive circuits (i.e., CMOS inverter) [6, 13, 14].

5.2 Analytical Modeling of CMOS Primitive Circuits

In order to apply the aforementioned methodology to a case study primitive circuit, the CMOS inverter (Fig. 5.1) and its energy dissipation are considered. The importance of modeling the timing response and the energy dissipation of CMOS inverters comes from the fact that the clock distribution networks and busses in digital integrated circuits are based on inverters or inverter-like circuits, which have to be carefully modeled since these circuits account for a great fraction of circuits' delay and energy dissipation [15]. Another equally important reason to obtain accurate models for the inverter delay and energy dissipation is that several methods for reducing multi-input CMOS gates to equivalent inverters have been proposed [6, 13, 14].

The energy dissipation of a CMOS inverter is constituted from the dynamic energy dissipation due to charge and discharge of the output load during the inverter switching, the leakage energy dissipation, and the short-circuit energy dissipation due to the direct current path from power supply to ground during the inverter switching [1]. During the rising transition of the output node, half of the supplied energy is stored in the load capacitance (and will be lost to the ground during the falling output transition), and the rest is lost in the form of heat at the PMOS device. The supplied energy and consequently the dynamic part of energy dissipation are independent from the output waveform of the inverter [1]. The leakage energy dissipation is caused by the reverse-bias diode leakage current at the device diffusion areas, the subthreshold current through the turned-off transistor channel, and the tunneling current through

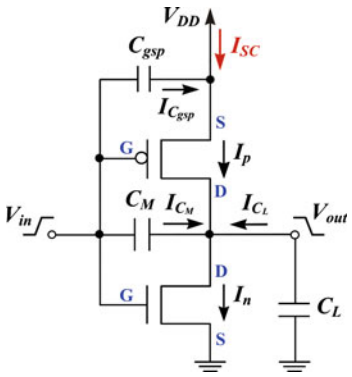


Fig. 5.2 CMOS inverter model used for the computation of the short-circuit energy dissipation

the very thin gate oxide. This type of energy dissipation is mainly dependent on the technology process parameters and it is significan for nanometer transistors [1, 2].

The short-circuit energy dissipation depends on the transition time of the inverter input voltage transition time, the output load, the supply voltage, and the internal design characteristics of the inverter [6, 16]. Hence, the computation of this type of energy dissipation is complex, and accurate analysis of the output voltage waveform is required by following the steps of the aforementioned methodology.

For the computation of the inverter short-circuit energy dissipation, analytical expressions of the output waveform in the required operating regions, are derived. These expressions take into account the current through both transistors and the influence of the gate-drain coupling capacitance, and they are valid for a wide range of input transition times, output loads, supply voltages, and device sizes. In addition, the influence of the short-circuiting transistor's gate-source capacitance is included when computing the short-circuit energy dissipation of the inverter. The circuit model that includes the parasitics to be considered is shown in Fig. 5.2.

The derivations presented further are for a rising input ramp: $V_{\text{in}} = V_{\text{DD}} \cdot (t/\tau)$ for $0 \le t \le \tau$, $V_{\text{in}} = 0$ for $t \le 0$, and $V_{\text{in}} = V_{\text{DD}}$ for $t \ge \tau$, where τ is the input rise time. The analysis for a falling input is symmetrical. The differential equation describing the discharge of the load capacitance C_{L} for the CMOS inverter (Fig. 5.2), taking into account the current through the gate-drain coupling capacitance, (C_{M}) is written as [3]:

$$C_{\text{L}} \frac{\text{d}V_{out}}{\text{d}t} = C_{\text{M}} \left(\frac{\text{d}V_{\text{in}}}{\text{d}t} - \frac{\text{d}V_{\text{out}}}{\text{d}t} \right) + I_{\text{p}} - I_{\text{n}} \tag{5.1}$$

After normalizing voltages with respect to the supply voltage (V_{DD}), and using the variable $x = t/\tau$, the differential Eq. (5.1) is written as

$$\frac{\text{d}u_{\text{out}}}{\text{d}x} = c_{\text{m}} + \frac{(I_{\text{p}} - I_{\text{n}})\tau}{(C_{\text{L}} + C_{\text{M}})V_{\text{DD}}}, \tag{5.2}$$

where

$$c_{\text{m}} = \frac{C_{\text{M}}}{C_{\text{L}} + C_{\text{M}}} \tag{5.3}$$

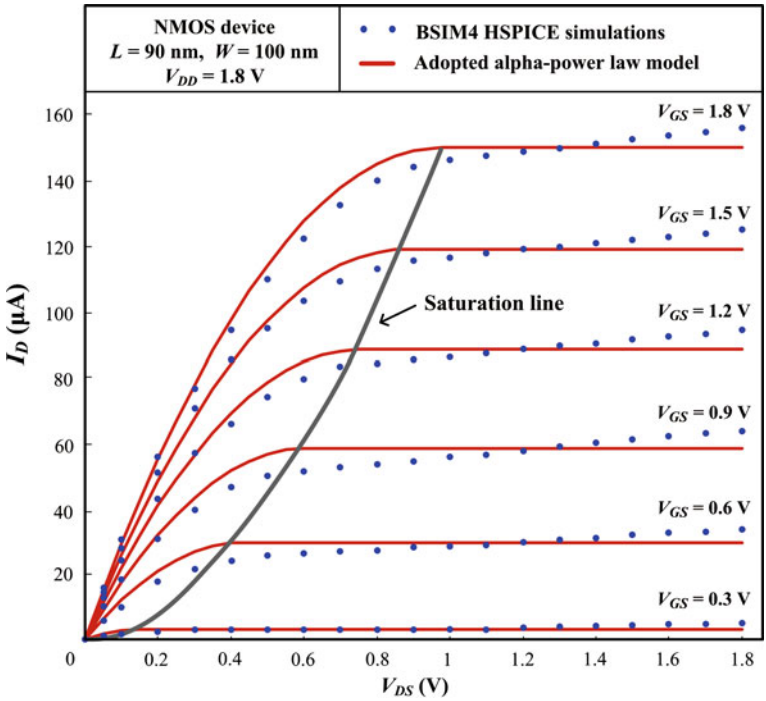


Fig. 5.3 The 90 nm NMOS device I–V characteristics

For the device drain current, the following accurate and simple version of the alpha-power law MOSFET model [17, 18] is used.

In triode or linear region ($V_{DS} \leq V'_{DO}$),

$$I_D = B(V_{GS} - V_T)^{\alpha} \left(2 - \frac{V_{DS}}{V'_{DO}}\right) \frac{V_{DS}}{V'_{DO}} \tag{5.4}$$

In saturation region ($V_{DS} > V'_{DO}$),

$$I_D = B(V_{GS} - V_T)^{\alpha}, \tag{5.5}$$

$$V'_{DO} = K(V_{GS} - V_T)^{\alpha/2} \tag{5.6}$$

The parameters of the adopted MOSFET model are determined by the used technology process or extracted by suitable fitting of device characteristics produced by simulation using BSIM4 predictive models [19, 20]. α is the velocity saturation index, V'_{DO} is the drain-source saturation voltage, B is the transconductance parameter extracted for $V_{GS} = V_{DD}$, and $V_{DS} = 3/4 \cdot V_{DD}$ (NMOS device), $V_{DS} = 4/5 \cdot V_{DD}$ (PMOS device), V_T is the threshold voltage of the device, and K is a constant determined by the drain-source saturation voltage at $V_{GS} = V_{DD}$. Such drain current equations model with sufficient accuracy the behavior of near − 100 nm devices (Fig. 5.3).

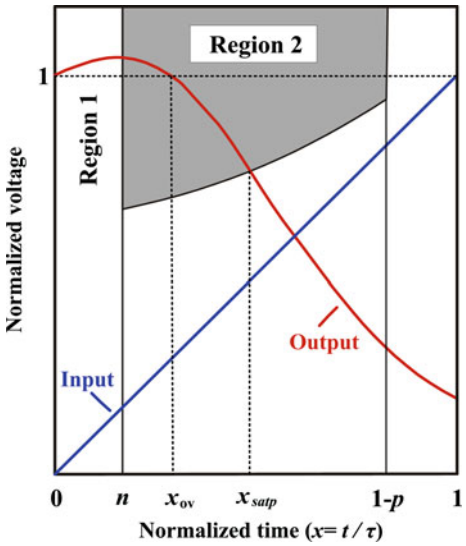


Fig. 5.4 Initial operation regions of the primitive circuit

The PMOS device current in the triode region (i.e., for $1 - u_{out} < u'_{dop}$) is given by the following equation:

$$I_p = k_{lp1}\,(1 - x - p)^{\alpha_p/2}(1 - u_{out}) - k_{lp2}(1 - u_{out})^2, \tag{5.7}$$

where p is the normalized threshold voltage of the device and k_{lp1}, k_{lp2} are constants dependent on the supply voltage and on the parameters K, B, and α of the device.

The NMOS device current in the saturation region (i.e., for $u_{out} \geq u'_{don}$) is given by

$$I_n = k_{sn}(x - n)^{\alpha_n} \tag{5.8}$$

where n is the normalized threshold voltage of the device and k_{sn} is a constant dependent on the supply voltage and on the parameters B and α of the device.

For the evaluation of the short-circuit energy dissipation, analytical expressions of the output waveforms in the two first inverter operating regions (Fig. 5.4) are required. For $0 \leq x \leq n$, the NMOS device is off and the PMOS device is in the triode region whereas for $n < x \leq x_{satp}$, the NMOS device is saturated and the PMOS device remains in the triode region. Part of the charge from the input, injected through the gate-to-drain coupling capacitance, causes an overshoot at the early part of the output voltage waveform. During the overshoot, there is no current from power supply to ground because the output voltage is greater than the supply voltage. In region 1, the differential Eq. (5.2) cannot be solved analytically. The main influence on the output voltage waveform in this region is due to the charge from the input, injected through the gate-to-drain coupling capacitance, whereas the influence of the PMOS transistor current is quite smaller, so if one applies careful approximations in the expression of the PMOS device current, the accuracy of the resulting output

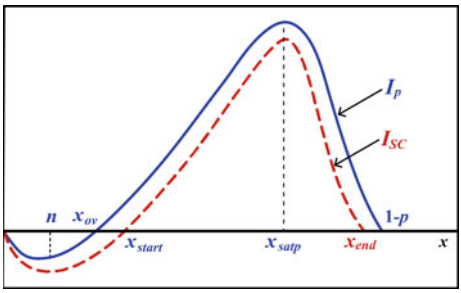


Fig. 5.5 PMOS device current and short-circuit current waveforms

voltage expression will not be affected. Therefore, an average value of x ($x = n/2$) is used in the first term of PMOS current (Eq. (5.7)), and an approximated expression for u_{out} ($u_{\text{out}} = 1 + c_{\text{m}} \cdot x$) is used in the quadratic term of the PMOS current. This approximated expression takes into account only the charge through the gate-to-drain coupling capacitance, but it is used only in the quadratic term of the PMOS transistor current [i.e., $(1 - u_{\text{out}})^2$]. The approximation is reasonable since in this operating region, the main influence on the output voltage waveform is due to the charge injected through the gate-to-drain coupling capacitance. Moreover, the charge contributed by the quadratic term of the PMOS transistor current is very small due to the small values of the PMOS device normalized drain-source voltage (i.e., $1 - u_{\text{out}}$) in this operating region. After the adoption of both approximations, u_{out} is given as

$$u_{\text{out}} = 1 + \frac{c_{\text{m}}}{\text{C}^3 \text{A}_{\text{lp1}}^3}[2\,\text{A}_{\text{lp2}} c_{\text{m}}(e^{-x\,\text{C}\,\text{A}_{\text{lp1}}} - 1) + 2\text{C}\,\text{A}_{\text{lp1}}\,\text{A}_{\text{lp2}}\,c_{\text{m}}\,x +$$

$$\text{C}^2\,\text{A}_{\text{lp1}}^2\,(1 - e^{-x\,\text{C}\,\text{A}_{\text{lp1}}} - \text{A}_{\text{lp2}}\,c_{\text{m}}\,x^2)], \tag{5.9}$$

where A_{lp1}, A_{lp2}, and C are constants dependent on the supply voltage, the normalized threshold voltages n and p, the load and gate-drain capacitances, the input transition time, and the parameters K, B, and α of the PMOS device.

In order to obtain an expression of the output voltage waveform for $n < x \leq x_{\text{satp}}$, the PMOS current (Fig. 5.5) is approximated by a linear function of x:

$$I_{\text{p}} = I_{\text{pn}} + S_1(x - n), \tag{5.10}$$

where I_{pn} is the PMOS device current for $x = n$. The current slope S_1 is computed by equating the exact PMOS current in the triode region with the approximated one at $x = (1 - p)/2$.

By using the linear approximation of the PMOS current, the differential Eq. (5.2) in region 2 is solved and the output voltage waveform is described by

$$u_{\text{out}} = u_{\text{n}} + (x - n)c_{\text{m}} + \text{d}I_{\text{pn}}(x - n) + \frac{\text{d}S_1(x - n)^2}{2} - \frac{\text{A}_{\text{sn}}(x - n)^{\alpha_{\text{n}}+1}}{\alpha_{\text{n}} + 1}, \tag{5.11}$$

where A_{sn} and d are constants dependent on the supply voltage, the load and gate-drain capacitances, the input transition time, and the parameters B and α of the NMOS device, and u_n is the normalized output voltage at $x = n$.

The short-circuit energy dissipation for a rising input is the energy of the current (I_{SC}), which is provided from the power supply (Fig. 5.2). The current through the PMOS device includes two non-short-circuit current components: the current flowing through C_{gsp} and the current flowing from the output to the supply node during the overshoot of the output signal. The short-circuit energy dissipation during the falling output transition is defined as

$$E_{SC} = V_{DD} \int_{x_{start}}^{x_{end}} I_{SC}\,\tau \mathrm{d}x = V_{DD} \left(\int_{x_{start}}^{x_{satp}} I_{SC}\,\tau \mathrm{d}x + \int_{x_{satp}}^{x_{end}} I_{SC}\,\tau \mathrm{d}x \right), \tag{5.12}$$

where $I_{SC} = I_p - I_{C_{gsp}}$ and $I_{C_{gsp}} = C_{gsp}(V_{DD}/\tau)$.

In the firs integral of (5.12), the following linear approximation of the PMOS device current is used:

$$I_p = S_2(x - x_{ov}). \tag{5.13}$$

The current slope S_2 is computed by equating the exact PMOS current in the triode region with that of (5.12), at the middle of the interval $[x_{ov}, x_{satp}]$. In the second integral of (5.12), the exact PMOS saturation current expression is used. After that, the inverter short-circuit energy dissipation is computed as

$$E_{SC} = \frac{V_{DD}}{2}(x_{satp} - x_{start}) \left[(x_{satp} + x_{start} - 2x_{ov})S_2 - \frac{2C_{gsp}V_{DD}}{\tau} \right] +$$

$$\frac{V_{DD}k_{sp}\tau}{\alpha_p + 1}[(1 - p - x_{satp})^{\alpha_p+1} - (1 - p - x_{end})^{\alpha_p+1}] - C_{gsp}V_{DD}^2(x_{end} - x_{satp}). \tag{5.14}$$

The computation of the boundary values x_{start} and x_{end} is achieved by setting the short-circuit current (I_{SC}) to zero, when the PMOS device operates in the linear and saturation region, respectively:

$$S_2(x_{start} - x_{ov}) - C_{gsp}(V_{DD}/\tau) = 0, \tag{5.15}$$

$$k_{sp}(1 - x_{end} - p)^{\alpha_p} - C_{gsp}(V_{DD}/\tau) = 0. \tag{5.16}$$

For the computation of the normalized time point x_{satp}, suitable second-order Taylor series expansions of u_{out} and u'_{dop} should be used. In Fig. 5.6a, the inverter short-circuit energy per transition is plotted as a function of the input transition time and for two different values of output load. In addition, in Fig. 5.6b, the inverter short-circuit energy per transition is plotted as a function of the supply voltage. The channel

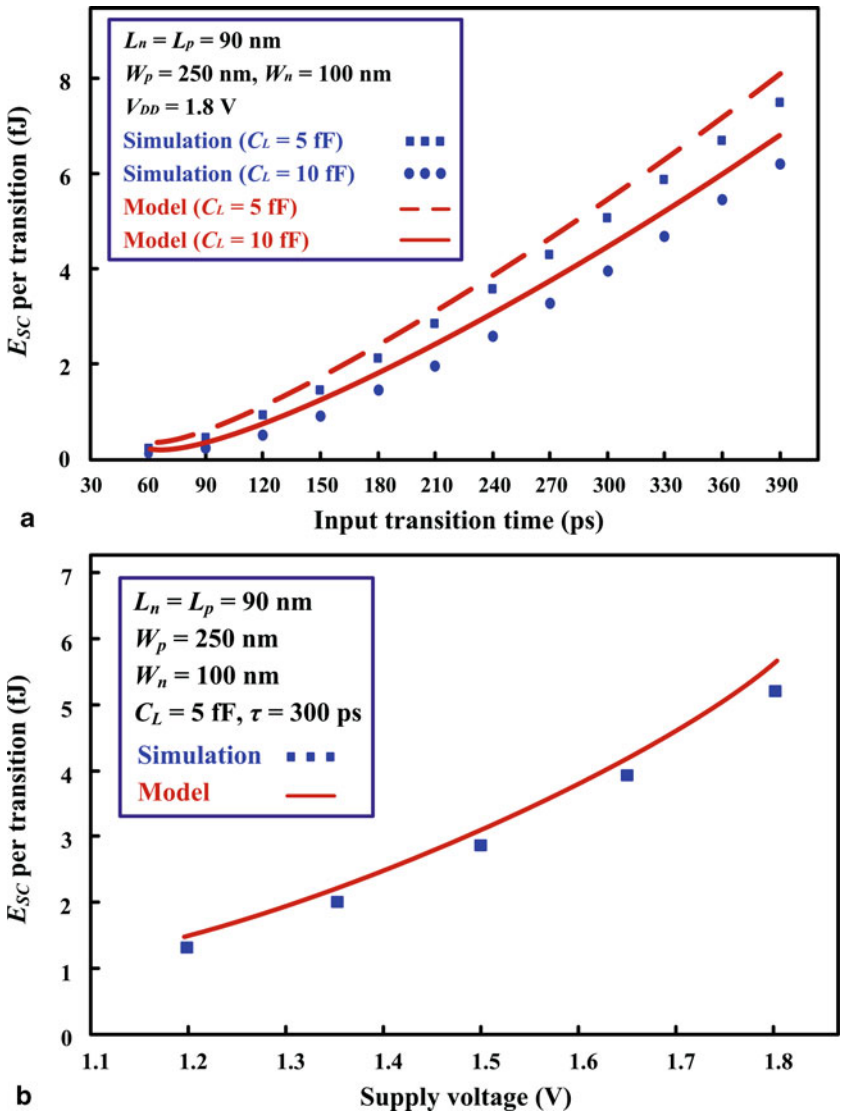


Fig. 5.6 **a** Inverter short-circuit energy dissipation per input transition for two different values of output load, **b** Inverter short-circuit energy dissipation per supply voltage

length of the used devices was 90 nm, and the results show very good agreement with BSIM4 [21] and HSPICE [22] simulations.

As described in the case of short-circuit energy dissipation modeling, the combination of saturation, triode, and cutoff devices' operation modes define different operating regions of the CMOS inverter. For the separation of the inverter operation into regions, the input voltage slope is also considered. Such definition of operating regions can be extended to the overall inverter operation range as shown in Fig. 5.7 [3].

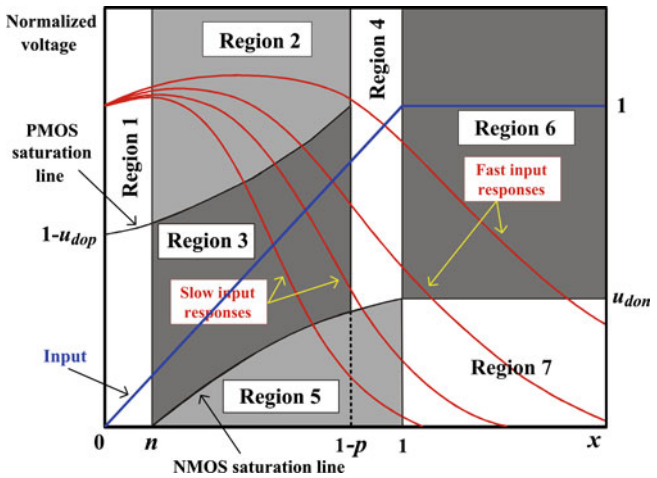


Fig. 5.7 Operating regions of the primitive circuit [3]

The differential equation can be defined and solved for all operating regions by using appropriate approximations when needed [3]. Then, the analytical output waveform expressions can be used for the derivation of explicit formulae for the computation of the primitive circuit's dynamic characteristics, such as propagation delay and output voltage transition time.

The inverter model can be extended to more complex CMOS primitive circuits such as multi-input CMOS gates [6, 13, 14]. Such extension is performed by reducing each gate to an equivalent inverter. This procedure requires the modeling of the series-connected and parallel-connected transistors (i.e., their reduction to single equivalent transistors), and the reduction of overlapping inputs to a single effective input signal. For a successful reduction of a gate, one should take into account the transition time of the gate inputs, the number of switching inputs of the gate, the position of the switching inputs, the body effect, the output load, and the internal node capacitances.

5.3 Compact Modeling of MOSFET Devices

The accuracy and computational efficiency of primitive circuit analysis models is directly affected by the accuracy and the simplicity of the MOSFET model used. Such a model must meet two equally important requirements: (1) accurate transistor drain current prediction and (2) simplicity of the device model to obtain explicit expressions for design parameters (i.e., transient response and energy dissipation).

In most existing current device models, the effects that determine the device behavior are accounted for through physical and empirical parameters. With the growing complexity of physical mechanisms in nanometer devices, device models become very complex and employ a large number of parameters to provide the highest

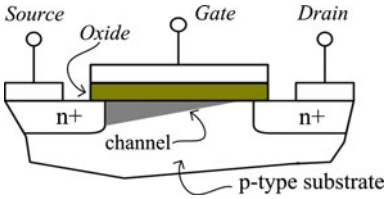


Fig. 5.8 The MOSFET transistor

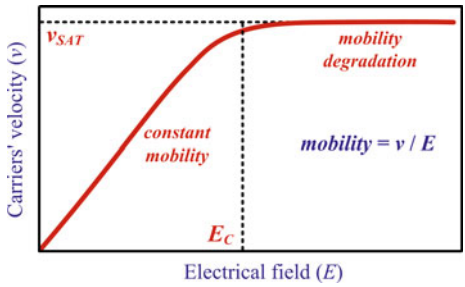


Fig. 5.9 Electric fiel to carrier velocity dependence

accuracy [11, 12]. Although these complex but accurate models can be handled by circuit simulators, they do not satisfy the requirement of computational efficiency. Hence, compact device models are needed, as simple as possible, to take into account the influence of essential physical mechanisms in nanometer devices by using few parameters extracted through measurements or simulations [12].

In order to strengthen the modeling methodology of CMOS circuit primitives, an accurate and compact I–V model for nanometer MOSFETs is required. The I–V equations of such model may use empirical parameters to match measured or simulated device characteristics, and should take into account the influence of predominant effects in nanometer devices, such as [11, 12, 23]: mobility degradation and velocity saturation, channel-length modulation, drain-induced barrier lowering (DIBL), body effect, narrow-channel width effect, and source-drain parasitic resistance.

5.3.1 Modeling of Small Dimension Effects in MOSFET devices

In MOSFET devices (Fig. 5.8), for small electric fields the carriers' mobility is constant and independent of the applied electric field As shown in Fig. 5.9, when the horizontal electrical field (moving the channel carriers) reaches a critical value, the carriers' velocity tends to saturate owing to scattering effect (i.e., electrons moving in semiconductor material collide with silicon atoms). This effect is more pronounced for reduced channel length that implies higher horizontal electric field for equivalent drain-source voltages [23]. The vertical electric field originating from the gate voltage further inhibits channel carrier mobility. This field pushes carriers toward the gate oxide and the carriers' mobility is reduced due to carrier collisions with the oxide–channel interface.

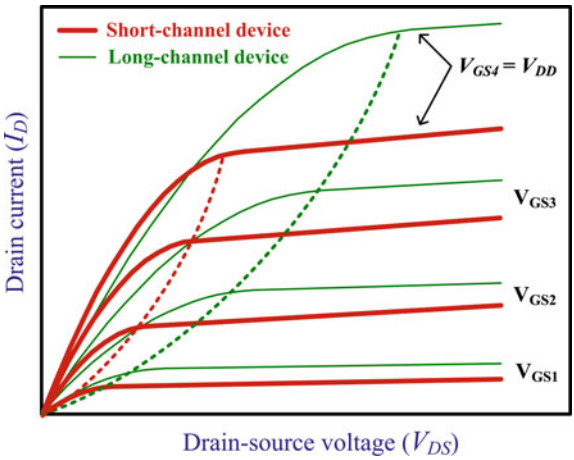


Fig. 5.10 MOSFET characteristics in the absence and presence of velocity saturation

The influence of mobility degradation and velocity saturation effects on the MOSFET device output characteristic curves is illustrated in Fig. 5.10. In short-channel devices, the saturation occurs at smaller drain-source voltages, and the spacing of the I–V curves in saturation is not according to square law, but becomes nearly proportional to gate-source voltage increment. As described in [17] as well as by (5.5) and (5.6), the modeling of mobility degradation and velocity saturation effects is achieved by employing the velocity saturation index (α) to describe the power laws featuring the drain current and the drain-source saturation voltage.

Channel length modulation (CLM) refers to the shortening of the length of the inverted channel region with increase in drain bias. When the device operates in saturation and the drain voltage increases, the uninverted region at the vicinity of the drain (pinch-off region) expands toward the source, shortening the length of the channel region (Fig. 5.8). Due to the fact that resistance is proportional to length, shortening the channel decreases its resistance, causing an increase in current with increase in drain bias for a device operating in saturation [1, 11]. The effect is more pronounced when the source-to-drain separation is short (i.e., in deep-submicrometer and nanometer devices) [23].

In a MOSFET device, a potential barrier exists between the source and the channel, which is controlled by the gate voltage. When the gate voltage is increased, the barrier between the source and the channel is decreased, increasing the carriers' injection from the source to the channel over the lowered barrier. In very short-channel devices, as drain voltage increases, more depletion is performed by the drain bias, and the electric field at the drain penetrates to the source region causing an additional decrease of the barrier at source (Fig. 5.11). This is referred to as DIBL effect [11]. As a result, the device can conduct significan drain current due to an increase in carriers injected from the source. DIBL affects the drain current versus drain bias curve, causing the current to increase with drain bias in the saturation region of operation (i.e., at high drain-to-source voltages). This current increase is additional to that caused by the CLM effect. The dependence between the drain current and the drain-source voltage

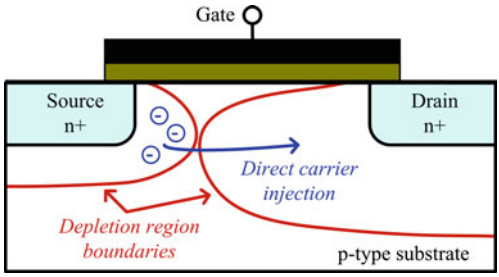


Fig. 5.11 Drain-induced barrier lowering (DIBL) effect in a MOSFET device

in the saturation region, which is due to CLM and DIBL effects, could be modeled through the inclusion of two additional empirical fitting arameters (A, D):

$$I_{\mathrm{D}} = B(V_{\mathrm{GS}} - V_{\mathrm{T}})^{\alpha}[\mathrm{A} + \mathrm{D}(V_{\mathrm{DS}} - V_{\mathrm{DO}})], \tag{5.17}$$

where V_{DO} is the drain-source saturation voltage of the device at $V_{\mathrm{GS}} = V_{\mathrm{DD}}$. Given the practical target of the model, that is, the analytical modeling of primitive circuits' operations, this linear dependence maintains the simplicity, while providing the required accuracy by including the influenc of both effects.

When a positive source-bulk voltage (V_{SB}) is applied, the bulk is at a negative potential with respect to the source, and this increases the depletion between the source and the bulk. The minority electrons attracted from the p-type bulk have to overcome this increase in depletion, and therefore the gate voltage required to form and maintain an inversion layer or channel (i.e., threshold voltage) becomes higher. This is referred to as body effect [11, 23]. For the determination of the device threshold voltage when V_{SB} is positive, a linear approximation (5.18) of the BSIM4 model [21] expression (5.19), describing the body effect, is used

$$V_{\mathrm{TH}} = V_{\mathrm{TO}} + \gamma V_{\mathrm{SB}}, \tag{5.18}$$

$$V_{\mathrm{TH}} = V_{\mathrm{TO}} + \mathrm{K}_1(\sqrt{\varphi_{\mathrm{s}} + V_{\mathrm{SB}}} - \sqrt{\varphi_{\mathrm{s}}}) + \mathrm{K}_2 V_{\mathrm{SB}}. \tag{5.19}$$

where V_{TO} is the threshold voltage for $V_{\mathrm{SB}} = 0$, φ_{s} is the inversion surface potential, K_1, K_2 are the BSIM4 body effect coefficients and γ is the simplified body effect coefficient

In MOSFET devices fabricated by using the local oxidation of silicon (LOCOS) process (Fig. 5.12) [23], the depletion region is not limited to just the area below the thin oxide, since the polysilicon gate overlaps the field oxide on both sides of the channel region, along the width direction of the device. For large device widths, the part of the depletion region on the sides is a small percentage of the total depletion region. As the device width is scaled down, the depletion charge under the gate is reduced but the fringing charge remains relatively unchanged, constituting a significant proportion. The gate is responsible for depleting a larger region and hence, higher gate voltage is required, resulting in increased threshold voltage. In effect, this results in lower driving capability per width unit of narrow-width devices in comparison with the driving capability per width unit of wide-width devices. The

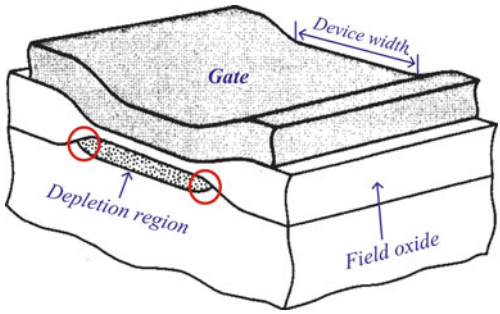


Fig. 5.12 Cross-section along the width of a MOSFET device in which the polysilicon gate overlaps the field oxide of the device [23]

prediction of the drain current for varying device widths (i.e., the inclusion of narrow-channel width effects) is obtained by computing the transconductance parameter B of the device as a quadratic function of the device channel width (W) [24]:

$$B = \beta_1 + \beta_2 W + \beta_3 W^2, \tag{5.20}$$

where the coefficient β_i are determined by fittin the quadratic plot to the B versus W plot, once for a given nanometer technology.

It has to be mentioned that for devices fabricated with the shallow-trench isolation process [23], the fringing field from the gate regions beyond the channel edges support depletion charges in the channel, and in contrast to the LOCOS process, the fringing field makes the depletion region deeper, thus increasing the surface potential, and helping the start of the inversion layer. In effect, this results in higher driving capability per width unit of narrow-width devices in comparison with that of wide-width devices. However, the transconductance parameter of the device can be predicted by using the aforementioned technique that is based on (5.20).

In long-channel devices, the source-drain parasitic resistance is negligible compared with the channel resistance. However, in very short-channel devices, it can be an appreciable fraction of the channel resistance and can therefore cause significan current degradation [11]. The most severe current degradation occurs in the triode region, i.e., for low values of drain-source voltage. This is because the channel resistance is low (the slope of the drain current versus drain-source voltage curve is high) under such bias conditions. An expression-based explanation can be easily derived by Eqs. (5.21)–(5.23).

$$I_{\mathrm{D}} = \frac{V_{\mathrm{DS}}}{R_{\mathrm{ch}} + R_{\mathrm{sd}}}, \tag{5.21}$$

$$R_{\mathrm{ch}} = \frac{V_{\mathrm{DS}}}{I_{\mathrm{D-without}\ R_{\mathrm{sd}}}}, \tag{5.22}$$

$$I_{\mathrm{D}} = \frac{I_{\mathrm{D-without}\ R_{\mathrm{sd}}}}{1 + (R_{\mathrm{sd}} I_{\mathrm{D-without}\ R_{\mathrm{sd}}})/V_{\mathrm{DS}}}. \tag{5.23}$$

Since the drain current dependence of the drain-source voltage is small in saturation, the current in this region is least affected by the parasitic resistance. The

source-drain parasitic resistance effect can be taken into account by adopting lower transconductance parameter in triode region than that of the saturation region. The transconductance parameters in both operating regions are derived by using proper fitting points on the device I–V output characteristics, as described in Sect. 5.3.2.

5.3.2 *Putting it All Together: A Compact MOSFET Current Model*

After the discussion of the predominant effects that impact the transistor behavior when technology scales down, compact model equations are introduced to describe the drain current of nanometer devices. Such equations take into account effects such as: mobility degradation and velocity saturation, channel-length modulation, DIBL, body effect, narrow-channel width effect, and source-drain parasitic resistance.

As mentioned earlier in this chapter, the objective of the derived device current equations is to strengthen the modeling methodology of CMOS circuit primitives. The adopted approach is based on the derivation of compact expressions that fit transistor curves over both bias ranges. The expressions use empirical parameters to specifically match measured device characteristics [25]. Empirically based models afford the possibility of developing analytical timing and energy dissipation models for CMOS primitive circuits, without the necessity of expensive numerical iterations [7, 25].

The combination of the advantages of existing compact models [14, 17, 18] with the modeling of the effects mentioned in Sect. 5.3.1, leads to the following strong inversion drain current MOSFET model:

For $V_{DS} \leq V'_{DO}$ (triode region),

$$I_D = B_{tri}(V_{GS} - V_T)^{\alpha} \left(2 - \frac{V_{DS}}{V'_{DO}}\right) \frac{V_{DS}}{V'_{DO}}. \tag{5.24}$$

For $V_{DS} > V'_{DO}$ (saturation region),

$$I_D = B_{sat}(V_{GS} - V_T)^{\alpha}[A + D(V_{DS} - V_{DO})]. \tag{5.25}$$

where

$$V'_{DO} = V_{DO}\left(\frac{V_{GS} - V_T}{V_{DD} - V_T}\right)^{\frac{\alpha}{2}}, \tag{5.26}$$

$$V_T = V_{TO} + \gamma V_{SB}, \quad B_{tri} = \frac{I'_{DO}}{(V_{DD} - V_T)^{\alpha}}, \quad B_{sat} = \frac{I_{DO}}{(V_{DD} - V_T)^{\alpha}},$$

$$A = \frac{I'_{DO}}{I_{DO}}, \text{ and } D = \frac{1 - A}{V_{DD} - V_{DO}}.$$

V_{TO} stands for the zero back-gate bias threshold voltage, γ is the coefficien accounts for the body effect, V_{DO} is the drain-source saturation voltage at $V_{GS} = V_{DD}$, and

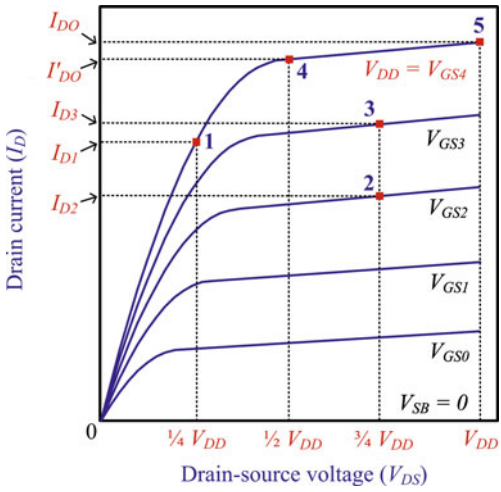


Fig. 5.13 Selected fittin points for extracting parameters of the NMOS device (the fitting points for the PMOS device are selected similarly according to the mentioned details) [25]

B_{tri} and B_{sat} are transconductance parameters for triode and saturation region, respectively. I_{DO} is the drain current at $V_{\text{GS}} = V_{\text{DS}} = V_{\text{DD}}$, whereas I'_{DO} is the drain current at $V_{\text{GS}} = V_{\text{DD}}$, $V_{\text{DS}} = 1/2 \cdot V_{\text{DD}}$ (for the NMOS device), and $V_{\text{DS}} = 2/3 \cdot V_{\text{DD}}$ (for the PMOS device).

The model parameters are extracted by using appropriately selected fitting points on the device I–V curves, as illustrated in Fig. 5.13 [25]. I_{DO} and I'_{DO} are easily obtained from the output MOSFET characteristics (points 5 and 4, respectively). α is extracted from the following equation, which is derived from (5.25) by using the fittin points 2 and 3 [17]:

$$\alpha = \frac{\ln\left(\frac{I_{\text{D3}}}{I_{\text{D2}}}\right)}{\ln\left(\frac{V_{\text{GS3}} - V_{\text{TO}}}{V_{\text{GS2}} - V_{\text{TO}}}\right)}. \tag{5.27}$$

V_{DO} is computed by combining (5.24) and (5.26) and using the fitting point 1:

$$V_{\text{DO}} = \frac{I'_{\text{DO}} V_{\text{DD}} + V_{\text{DD}} \sqrt{I'_{\text{DO}}(I'_{\text{DO}} - I_{\text{D1}})}}{4 I_{\text{D1}}}. \tag{5.28}$$

By simulating the device, we obtain I_{DO} and I'_{DO} and consequently B_{tri} and B_{sat} for the minimum and the maximum used device width, as well as for few intermediate width values, in order to fit B_{tri} versus W and B_{sat} versus W plots to quadratic plots, such as:

$$B_{\text{tri}} = \beta_{\text{t1}} + \beta_{\text{t2}} W + \beta_{\text{t3}} W^2, \tag{5.29}$$

$$B_{\text{sat}} = \beta_{\text{s1}} + \beta_{\text{s2}} W + \beta_{\text{s3}} W^2. \tag{5.30}$$

Given the applicative target of the presented MOSFET current model that is the analysis and modeling of primitive CMOS circuits, an accurate characterization

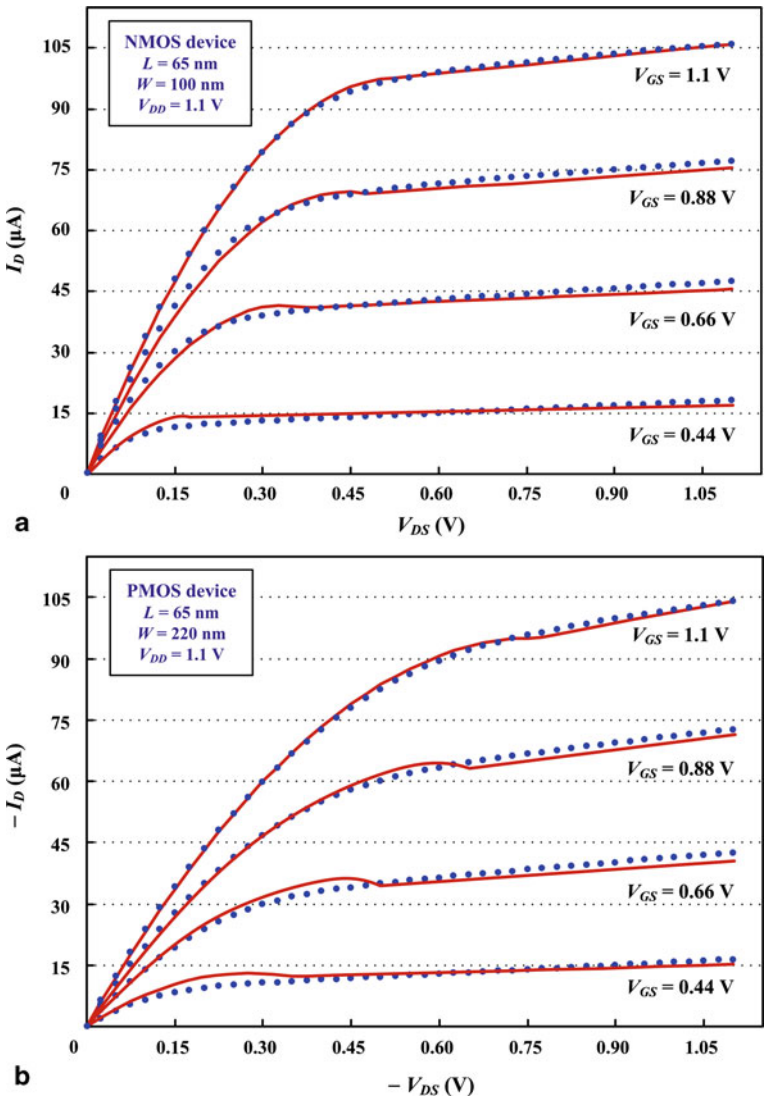


Fig. 5.14 I–V plots for 65 nm CMOS technology, **a** NMOS device, **b** PMOS device (*continuous lines* correspond to model results and *dots* correspond to BSIM4 HSPICE simulations)

of the current behavior in the subthreshold operating region is not essential. The presented MOSFET current model can lead to accurate estimation of the dynamic characteristics of CMOS primitive circuits, even considering a current abruptly going to zero for $V_{GS} = V_T$, and not modeling second-order effects (such as DIBL) within the threshold voltage expression.

The derived MOSFET current model has been validated by using a 65 nm CMOS technology. Figure 5.14 presents the NMOS and PMOS output characteristic curves for device widths of 100 nm and 220 nm, respectively. In Fig. 5.15, the

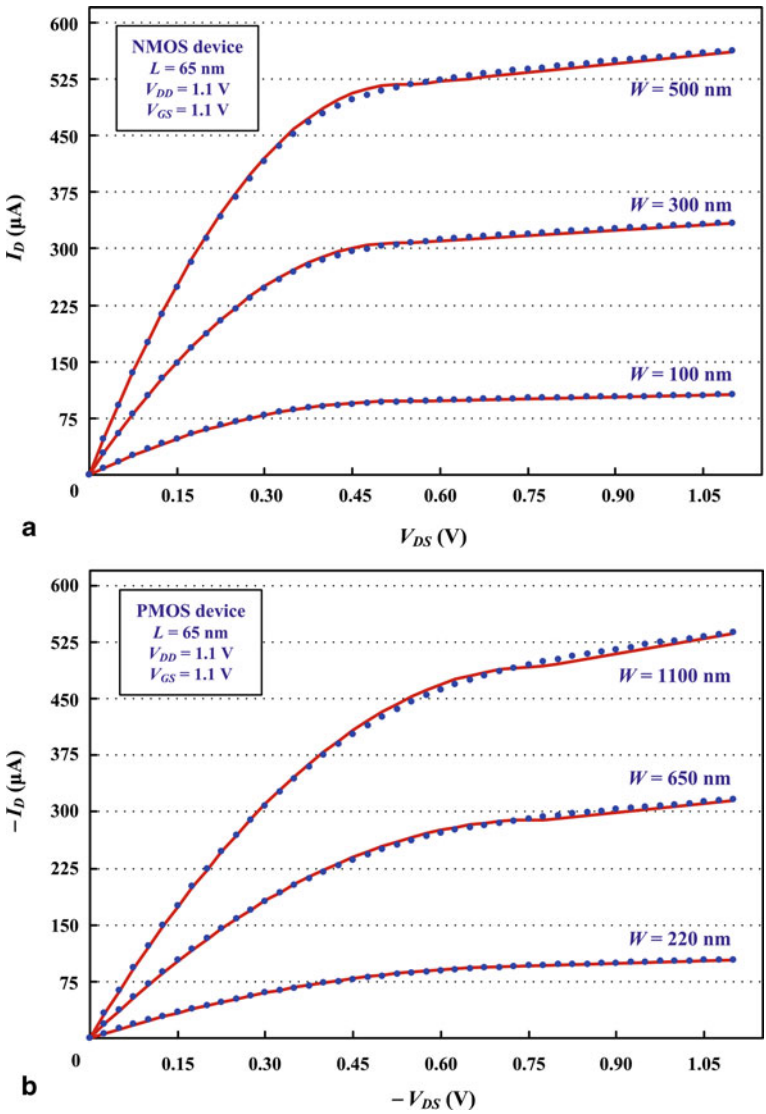


Fig. 5.15 I–V plots for different device channel widths, **a** NMOS device, **b** PMOS device (*continuous lines* correspond to model results and *dots* correspond to BSIM4 HSPICE simulations)

model is validated for several device widths. Continuous lines correspond to the model results, whereas dots correspond to the BSIM4 HSPICE simulations. The experimental results derived by the model show very good agreement with the simulation data extracted by using predictive technology models [19, 20].

The exhibited discontinuity at the boundary between triode and saturation operating regions is due to the fact that a basic concern of the device current model is

to avoid complex dependence on the drain-source voltage. Such discontinuity could cause problems in circuit simulators that require continuity of the functions, but should not cause problems in case of use for the analytical computation of dynamic characteristics in primitive circuits.

5.4 Conclusion

In this chapter, a methodology for the modeling of dynamic characteristics (such as transient response and energy dissipation) of CMOS primitive circuits has been presented. As a case study, the CMOS inverter has been used. Following a detailed analysis of its operation, accurate formulae for its output response are derived for the different operating regions, and based on this analysis, analytical expressions for the calculation of the timing and energy parameters can be produced. The derived models account for the influences of circuit design and operational parameters, as well as of device parameters. After an analysis of predominant effects in modern nanometer device technologies, a compact device current model has been presented that exhibits simplicity and is accurate enough to strengthen the modeling methodology of CMOS circuit primitives.

References

1. Weste, N. H. E., Harris, D. M. (2011). *CMOS VLSI design: A circuits and systems perspective*. Boston: Pearson Education.
2. Rabaey, J. (2009). *Low power design essentials*. New York: Springer.
3. Bisdounis, L., Nikolaidis, S., Koufopavlou, O. (1998). Analytical transient response and propagation delay evaluation of the CMOS inverter for short-channel devices. *IEEE Journal of Solid-State Circuits, 33*(2), 302–306.
4. Rossello, J. L., Segura, J. (2004). An analytical charge-based compact delay model for submicrometer CMOS inverters. *IEEE Transactions on Circuits and Systems I, 51*(7), 1301–1311.
5. Kabbani, A., Al-Khalili, D., Al-Khalili, A. J. (2003). Technology-portable analytical model for DSM CMOS inverter transition time estimation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 22*(9), 1177–1187.
6. Bisdounis, L., Koufopavlou, O. (2000). Short-circuit energy dissipation modeling for submicrometer CMOS gates. *IEEE Transactions on Circuits and Systems I, 47*(9), 1350–1361.
7. Consoli, E., Giustolisi, G., Palumbo, G. (2012). An accurate ultra-compact I-V model for nanometer MOS transistors with applications on digital circuits. *IEEE Transactions on Circuits and Systems I, 59*(1), 159–169.
8. Shih, Y. H., Leblebici, Y., Kang, S. M. (1993). ILLIADS: A fast timing and reliability simulator for digital MOS circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 12*(9), 1387–1402.
9. Kutuk, H., Goknar, I. C., Kang, S. M. (1999). Interconnect simulation in a fast timing simulator ILLIADS-I. *IEEE Transactions on Circuits and Systems I, 46*(1), 178–189.
10. Bellido, M. J., Juan, J., Valencia, M. (2006). *Logic-timing simulation and the degradation delay model*. London: Imperial College Press.

11. Taur, Y., Ning, T. (2009). *Fundamentals of modern VLSI devices*. Cambridge: Cambridge University Press.
12. Bhattacharyya, A. B. (2009). *Compact MOSFET models for VLSI design*. Singapore: Wiley.
13. Chatzigeorgiou, A., Nikolaidis, S., Tsoukalas, I. (1999). A modeling technique for CMOS gates. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 18*(5), 557–575.
14. Sakurai, T., Newton, A. R. (1991). Delay analysis of series-connected MOSFET circuits. *IEEE Journal of Solid-State Circuits, 26*(2), 122–131.
15. Liu, D., Svensson, C. (1994). Power consumption estimation in CMOS VLSI chips. *IEEE Journal of Solid-State Circuits, 29*(6), 663–670.
16. Bisdounis, L. (2010). Short-circuit energy dissipation model for sub –100 nm CMOS buffers. In Proceedings of the 17th IEEE international conference on electronics, circuits and systems, Athens, 12–15 December 2010, pp. 615–618.
17. Sakurai, T., Newton, A. R. (1990). Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas. *IEEE Journal of Solid-State Circuits, 25*(2), 584–594.
18. Nose, K., Sakurai, T. (2000). Analysis and future trend for short-circuit power. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 19*(9), 1023–1030.
19. Zhao, W., Cao, Y. (2007). Predictive technology model for nano-CMOS design exploration. *ACM Journal of Emerging Technologies in Computing Systems, 3*(1), 2–17.
20. Arizona State University (2006). Predictive technology model. http://ptm.asu.edu/cgi-bin/test/nanocmos.cgi.
21. Liu, W. (2001). *MOSFET models for SPICE simulation including BSIM 3v3 and BSIM4*. New York: Wiley.
22. Synopsys Inc. (2007). *HSPICE simulation and analysis user guide*. Mountain View: Synopsys Inc.
23. Tsividis, Y. (2003). *Operation and modeling of the MOS transistor*. New York: Oxford University Press.
24. Chandra, N., Kumar, A., Bhattacharyya, A. B. (2009). Extended Sakurai-Newton MOSFET model for ultra-deep-sub-micrometer CMOS digital design. In Proceedings of the 22nd international conference on VLSI design, New Delhi, 5–9 January 2009, pp. 247–252.
25. Bisdounis, L. (2012). An accurate and compact MOSFET I-V model for nanometer CMOS circuit analysis. In Proceedings of the 2nd pan-hellenic conference on electronics and telecommunications, Thessaloniki, 16–18 March 2012.