CMOS Circuits Basics

2

Contents

Although a CMOS chip is a complex object comprising logic, memory, analog, and I/O functions, significant insight can be gained from the simulated and measured behaviors of circuit elements and small circuit blocks. The basic components and building blocks of digital logic circuits and their electrical properties are described. Circuit simulations are set up with BSIM models for plotting *I–V* and *C–V* characteristics of MOSFETs and extracting their key parameters. A methodology to characterize logic gates typically found in a standard cell library is introduced using an inverter as an example. Lookup tables for computing signal delays in combinational logic circuits with different input signal waveforms and load capacitances are generated, highlighting their interdependencies. Delay chains and ring oscillator configurations used for model validation in silicon hardware are described and simulated to extract delay parameters of logic gates. The foundations laid here including Monte Carlo analysis for determining parameter spreads are used throughout the book.

M. Bhushan, M.B. Ketchen, *CMOS Test and Evaluation*,
DOI 10.1007/978-1-4939-1349-7_2

The spreads in propagation delays and power levels of circuit blocks reflect the ranges of cycle time and operating power of a digital CMOS chip as a whole. The behavior of relatively simple circuit blocks can also be easily related to their constituent MOSFETs, interconnects, and parasitics. Small circuit blocks, which may be characterized in detail, therefore provide direct links to both the product chip and the underlying silicon technology elements as indicated in Fig. 2.1. With the physical insight acquired through such an approach, a common platform for communication among silicon technology, circuit design, design tools, test and characterization teams emerge.

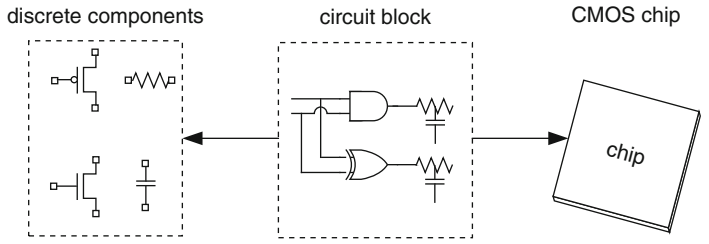


Fig. 2.1 Linking circuit blocks to silicon technology and CMOS product

Computer modeling and simulation play an important role in both design and test of microelectronic products. While this has been true all along, the compute resources now available generally far exceed the requirements of most such tasks. Indeed the compute power present in a personal computer purchased for a few hundred dollars in 2014 far exceeds that on the original Apollo lunar lander in late 1960s! The real challenge facing the design and test community today is the effective use of compute power to get the job done in a systematic, efficient, and accurate manner. Sophisticated quantum mechanical models may be helpful and even essential in understanding some of the basic physical behavior of highly scaled devices as we approach the nano-regime, yet this complexity can be simplified and encapsulated in a compact model, along with parasitic resistances, capacitances, and inductances, for efficient and accurate representation of device behavior over a practical range of use.

Typically CMOS chip designs are carried out on a workstation equipped with custom vendor supported tools. Licensing fees for use of these tools can be significant. Recognizing that not everyone has access to such tools, we have opted to use LTspice IV released by Linear Technology as the simulation tool [1]. LTspice is widely used by circuit designers and can be downloaded for free to run on a personal computer. All the examples and problems described in this book may be adapted to other versions of SPICE simulation tools [2, 3] as well.

SPICE simulators require compact models for MOSFETs, diodes, interconnects, and other parasitic components. Berkeley short-channel IGFET models (BSIM) for MOSFETs are presently the industry standard [4]. We have used BSIM predictive technology models (PTM) released by Arizona State University [5]. These models use a set of simplified equations to describe critical electrostatic behavior and carrier transport rather than the full set used in more complex BSIM models [6, 7]. Published data from early technology development as well as from previous technology generations are used for building more realistic models in advance of full technology development. The PTM models take into account limits of scaling due to manufacturability and fabrication cost and some new features introduced in successive technology nodes. These models do not represent any particular silicon foundry.

The 45-nm technology PTM models for high-performance (HP) and low-power (LP) devices are used in circuit simulation examples. While generally realistic in their representation of 45-nm technology these models do occasionally exhibit unusual behavior. As an example the temperature dependence of MOSFET drive current in the saturation mode is much stronger than normally observed. Most circuit simulation examples and problems are carried out at 25 °C to avoid operating conditions where the models are weak.

In this chapter circuit simulation techniques for MOSFET and logic gate characterization are introduced. The foundations laid here will be used in other chapters. For completeness, a brief overview of circuit components, chip design methodology, and test are included. Circuit components and their basic properties are described in Sect. 2.1. Circuit simulations and SPICE commands for characterization and model-to-hardware correlation of MOSFETs and logic gates are covered in Sect. 2.2.

In-depth treatment of CMOS circuits can be found in many excellent textbooks [8–11]. Other useful text book references cover CMOS devices [12, 13] and silicon fabrication technology [14, 15].

2.1 Circuit Components and Building Blocks

A schematic cross section for a CMOS planar process with four metal interconnecting layers is shown in Fig. 2.2. MOSFETs and diodes, the two active elements in CMOS circuits, are delineated in the silicon substrate. Metal and dielectric isolation layers are deposited on top of the active devices and patterned for making wire interconnections. Connections to the chip package are made through solder balls at the top of the metal wire stack by flip-chip bonding, or by wire bonding to I/O pads in the top metal layer.

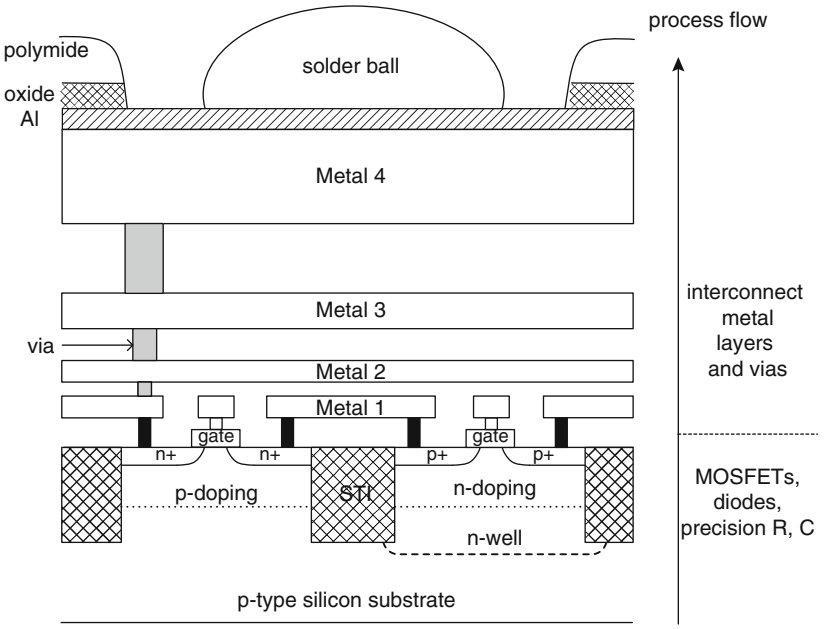


Fig. 2.2 Schematic cross section of part of a CMOS circuit showing an n-FET, a p-FET, four metal layers with inter-level vias and a solder ball for connecting to the chip package

An integrated CMOS fabrication process is very complex with many flavors of active devices differing in their electrical properties, and up to 15 or more interconnecting metal layers. The translation from circuit schematic to data input required for fabrication is through physical mapping to layers comprising two-dimensional geometric shapes and alignment of each layer with respect to other layers. Each layer is assigned a key and a color or shading to distinguish it from other layers in the drawing, and is defined with opaque and transparent areas on a photomask. This photomask is in turn used to transfer the layer shapes to a silicon wafer coated with a photosensitive material (photoresist). The wafer is exposed to ultraviolet light through the photomask and the exposed photoresist chemically processed to develop the pattern. Subsequent processing such as dopant implantation, material deposition, reactive-ion etching to remove material from unwanted areas, chemical mechanical polishing to obtain a planar surface for forming metal interconnects, and thermal and other treatments accomplish the three-dimensional physical realization of each layer with its desired properties.

Electrical and other material properties of the layers, circuit components and devices defined by the layers, and parasitic elements associated with the circuit components are described in compact models released by the silicon foundry. The models include the nominal values of the parameters, and the range of expected variations in key parameters of circuit elements introduced during manufacturing. CMOS chips are generally designed to operate with circuit properties varying within their published range.

Physical layer dimensions and compact models are two of the key inputs to circuit simulation tools. As an example, the circuit symbol of a resistor shown in Fig. 2.3a is an electrical representation of a metal wire. The physical layout of the metal wire, drawn as a rectangle with a layer key designated to a specific metal layer (e.g., metal layer M3) is shown in Fig. 2.3b. It has a drawn or design length l, width w, and is placed at a lateral distance s from a neighboring wire on the same layer. Properties of the metal wire, its actual dimensions when printed on silicon (l, w_{eff}), its temperature dependence defined by a temperature coefficient of resistance (TCR), and parameter tolerances ($\pm 3\sigma$) are included in the compact models. These parameters are used in model equations of the type shown in Fig. 2.3c.

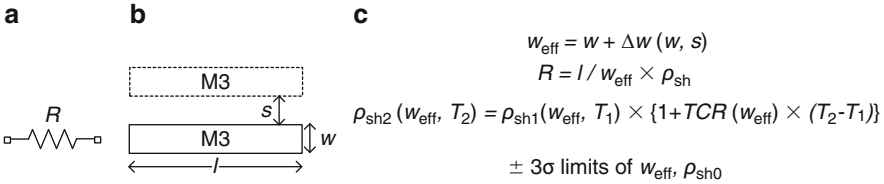
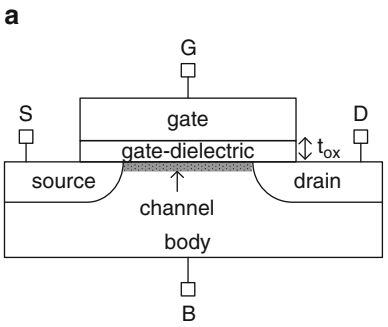


Fig. 2.3 (**a**) Circuit symbol of a resistor and (**b**) top view of the physical layout of a metal wire resistor of length l, width w, and nearest neighbor spacing s drawn in metal layer M3. (**c**) Model equations describing effective width w_{eff}, wire resistance R, sheet resistance ρ_{sh}, and parameter tolerances

In circuit simulations and in analyzing CMOS electrical test data, it is important to understand the relationships among the physical dimensions of circuit components and their electrical properties. At the circuit simulation stage, the details of the silicon fabrication process need not be considered. However, some knowledge of silicon processing is needed when analyzing process-induced variability and its impact on chip functionality and yield. A basic description of silicon process steps and sources of variations in the manufacturing process is covered in Chap. 6. Detailed description of CMOS processing can be found in any one of several books on this topic [14, 15].

2.1.1 MOSFETs

A cross section schematic view of a MOSFET with its source (S), drain (D), gate (G), and body (B) terminals is shown in Fig. 2.4a. The gate electrode is separated from the body by a thin insulating layer of silicon oxide or an alternative dielectric material of thickness t_{ox}. When a voltage of appropriate polarity with respect to source is applied to the G terminal (positive for n-FET and negative for p-FET), minority carriers in the body are pulled towards the surface and a conducting channel is formed. With the source and drain regions contacting the channel, carrier transport occurs across the channel in the presence of an electric field.



b

	n-FET	p-FET
body	p	n
source	n+	p+
drain	n+	p+
gate	p+ polySi / metal	n+ polySi / metal

Fig. 2.4 (**a**) Schematic cross section of a MOSFET and (**b**) table listing the doping types of the body, source, drain, and gate in n-FETs and p-FETs

In complementary MOS (CMOS) technology, an n-type MOSFET (n-FET) is formed in a p-type body and a p-type MOSFET (p-FET) is formed in an n-type body. Conduction in the channel is primarily by electrons in an n-FET and by holes in a p-FET. The source and drain regions are heavily doped (n+ for n-FET and p+ for p-FET) and make low-resistance contact to the channel. The gate material is doped polysilicon, p+ for n-FET and n+ for p-FET. A highly conductive silicide film covers the gate, source, and drain regions to reduce the parasitic resistances. In advanced technologies with a high-K dielectric gate insulator, the gate electrode is a metal stack with tailored work functions for n-FETs and p-FETs. The doping types of different MOSFET regions are included in the table in Fig. 2.4b.

A schematic cross section with an n-FET and a p-FET in a single p-type silicon substrate is shown in Fig. 2.5. An n-type doped region (n-well) is created for the body of the p-FET and the two MOSFET types are isolated by shallow trench oxide (STI) regions. A polysilicon or metal layer (PS) forms the gate electrode. Connection to the substrate or body (B) is made through a heavily doped silicon layer (p+ for n-FET and n+ for p-FET). All of the MOSFET terminals are contacted by the first metal layer (M1) through H0 vias in a dielectric isolation layer (not shown).

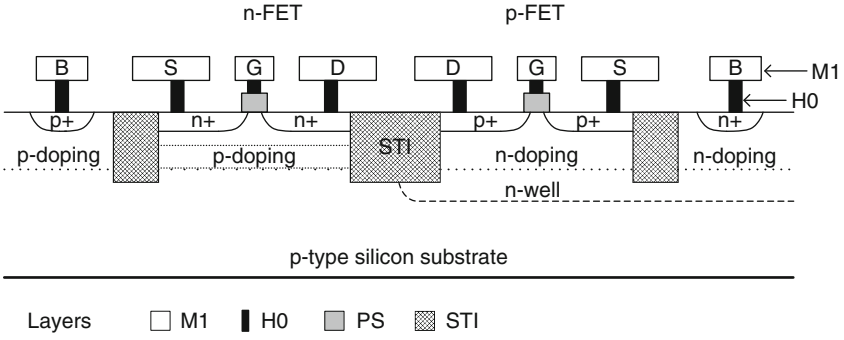
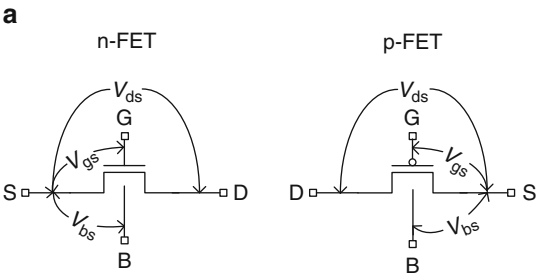


Fig. 2.5 Schematic cross sections of an n-FET and a p-FET with body contacts

Circuit symbols for the n-FET and p-FET with all four terminals are shown in Fig. 2.6a. Bias voltages are measured with respect to the S terminal: drain-to-source voltage V_{ds}, gate-to-source voltage V_{gs} and body-to-source voltage V_{bs}. The S terminal is typically held at GND for the n-FET and at V_{DD} for the p-FET. The B terminal is generally tied to the source terminal ($V_{bs} = 0$). The MOSFET schematic can then be drawn with three terminals (D, G, S). In twin well-bulk silicon CMOS technology, the B terminal may be biased independently. The nominal voltages in the conducting state of the MOSFETs are listed in Fig. 2.6b.



b

	n-FET	p-FET
source	GND	V_{DD}
drain	V_{DD}	GND
gate*	V_{DD}	GND

* in full saturation

Fig. 2.6 (**a**) Circuit symbols for n-FET and p-FET indicating voltage bias parameters and (**b**) nominal voltage bias values in conducting state

Physical layer mappings of an n-FET and a p-FET are shown in Fig. 2.7. Layer DF defines the source and drain regions and the gate region is defined by the PS layer adjacent to DF. Metal M1 wires connect to PS and DF layers through H0 vias in the dielectric isolation layer. Key parameters are drawn channel length L_p widths W_n (n-FET), and W_p (p-FET), and length of source and drain regions L_{ds}.

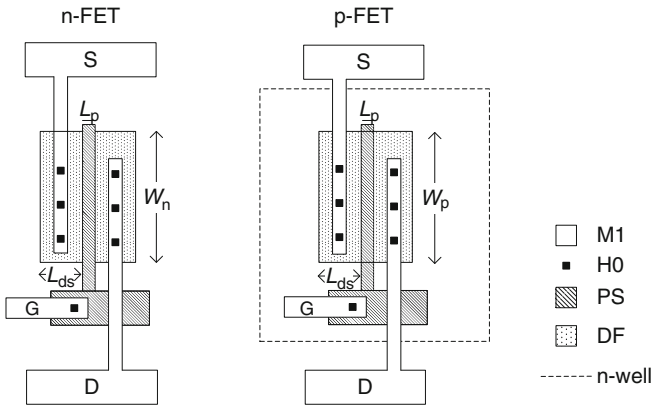


Fig. 2.7 Physical layout of an n-FET and a p-FET. Source and drain diffusion areas are implemented in the DF layer and PS is the gate layer

MOSFET (n-FET) drain-to-source current I_{ds} is measured by holding V_{gs} constant and sweeping V_{ds} from GND to the power supply voltage V_{DD}, or by holding V_{ds} constant and sweeping V_{gs}. Normalized I_{ds}–V_{ds} and I_{ds}–V_{gs} curves obtained in this manner are shown in Fig. 2.8a and b respectively. The I_{ds}–V_{ds} characteristics are separated into linear and saturation regions. In the linear region, I_{ds} varies linearly with V_{gs} and V_{ds}. In the saturation region, in an ideal MOSFET, I_{ds} is nearly independent of V_{ds} but increases linearly with (V_{gs}–V_t), where V_t is the threshold voltage at which the channel begins to conduct.

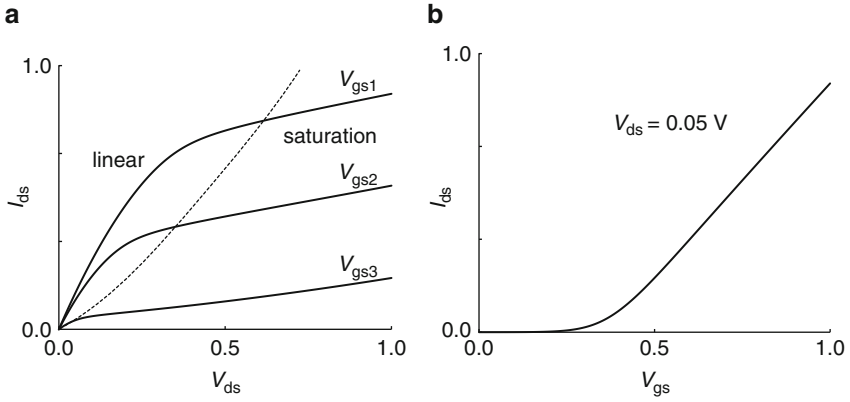


Fig. 2.8 An n-FET (**a**) I_{ds}–V_{ds} characteristics and (**b**) I_{ds}–V_{gs} characteristics

The range over which I_{ds} varies as V_{gs} is increased spans several decades. This can be viewed by plotting I_{ds}–V_{gs} characteristics on a log-linear scale as shown in Fig. 2.9. Ideally a MOSFET draws current only in the "on" state but there is also a small current component in the off-state ($V_{gs} = 0$). The current in the off-state I_{off} is measured at $V_{gs} = 0$. As V_{ds} is increased, I_{off} increases and V_t is lowered. At a low V_{ds} (typically 0.05 V), a MOSFET is in the linear region of its operating characteristics and $V_t = V_{tlin}$. When V_{ds} is set equal to V_{DD}, the MOSFET is in the saturation region and $V_t = V_{tsat}$.

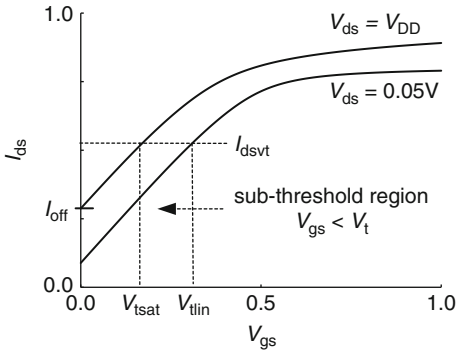


Fig. 2.9 Log(I_{ds})–V_{gs} characteristics of an n-FET with $V_{ds} = V_{DD}$ (saturation region) and $V_{ds} = 0.05$ V (linear region)

A convenient way to measure the threshold voltage V_t in both linear and saturation regions is illustrated in Fig. 2.9. Here V_t is defined as the value of V_{gs} required to reach a fixed I_{ds} ($=I_{dsvt}$). The value of I_{dsvt} is selected to be in the range of $\sim 100 \times I_{off}$. In another method, V_{tlin} is obtained by extrapolating the linear section of the I_{ds}–V_{gs} characteristic to $I_{ds} = 0$, and V_{tsat} is obtained by extrapolating the linear section of the $\sqrt{I_{ds}}$–V_{gs} characteristic to $I_{ds} = 0$.

Note that the definition of V_t is not precise and depends on the value of V_{ds} and the method of extraction. It does serve as a very useful single parameter for comparing MOSFET properties in subthreshold, linear and saturation regions even though its absolute value may vary with the choice of extraction.

MOSFET I–V characteristics are highly nonlinear. Long-channel MOSFET behavior in the linear and saturation regions is described by Eqs. 2.1 and 2.2 below:

Linear (non-saturation) region:

$$I_{ds} = \beta\left\{\left(V_{gs} - V_t\right)V_{ds} - \frac{V_{ds}^2}{2}\right\}, \quad 0 < V_{ds} < \left(V_{gs} - V_t\right). \tag{2.1}$$

Saturation region:

$$I_{ds} = \frac{\beta}{2}\left\{\left(V_{gs} - V_t\right)^2\right\}, \quad 0 < \left(V_{gs} - V_t\right) < V_{ds}. \tag{2.2}$$

The gain factor β in Eqs. 2.1 and 2.2 is given by

$$\beta = \frac{\mu_{eff}\varepsilon\varepsilon_0}{t_{ox}}\left\{\frac{W}{L_p}\right\}, \tag{2.3}$$

where μ_{eff} is the effective carrier mobility, t_{ox} is the thickness of the gate-dielectric, ε is the gate-dielectric constant, and ε_0 ($=8.854 \times 10^{-12}$ F/m) is the permittivity of free space.

In the subthreshold region shown in Fig. 2.9, the slope of V_{gs} vs. $\log(I_{ds})$ is the subthreshold slope (SS) defined as

$$SS = \frac{dV_{gs}}{d(\log_{10} I_{ds})}, \tag{2.4}$$

and expressed in mV/decade. It indicates the increase in V_{gs} in mV corresponding to a $10\times$ increase in I_{ds}. Although I_{off} is a strong function of V_{ds}, SS is nearly independent of V_{ds}.

As MOSFETs are scaled to smaller dimensions, phenomena such as short-channel effect (SCE), drain-induced barrier lowering (DIBL), gate-induced drain leakage (GIDL), and strain-enhanced mobility effects become increasingly significant. The basic expressions in Eqs. 2.1–2.4 are modified and many additional terms added to incorporate these effects.

In any CMOS technology node, several different types of MOSFET pairs are offered. These MOSFET types differ in nominal values of L_p, V_t, t_{ox}, and mobility μ_{eff} to meet circuit and reliability requirements of different circuit topologies and applications. Additional physical layers are added to the layouts shown in Fig. 2.7 for silicon processing. In advanced technologies, a nominal value of L_p is specified for logic circuits and a wider range of L_p or a discrete set of L_p values is allowed for analog circuit applications. MOSFET width dimensions W_n and W_p vary in the designs and typically a minimum allowed width is specified.

A MOS capacitor is formed by the gate and the body of the MOSFET. It has an intrinsic capacitance arising from the inversion and depletion layers in the channel region. In addition there are parasitic capacitances associated with overlap regions between the gate and source/drain, and with the p/n junctions at the source/drain boundaries as shown in Fig. 2.10a.

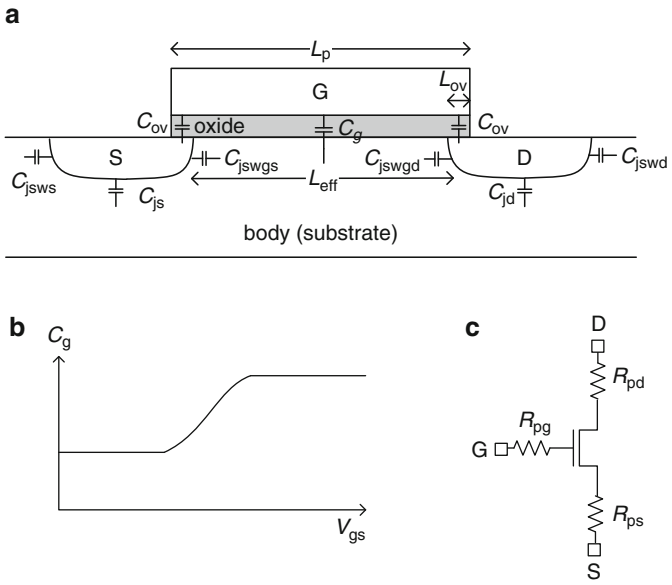


Fig. 2.10 An n-FET: (**a**) gate, source, and drain capacitances, (**b**) gate capacitance C_g as a function of V_{gs} and (**c**) parasitic series resistances

The intrinsic MOSFET gate capacitance is a function of V_{gs}, increasing with the formation of the inversion layer. The gate-dielectric capacitance C_{ox} per unit area varies inversely with the oxide (gate-dielectric) thickness t_{ox}. In the subthreshold region of operation, the capacitances associated with the gate-dielectric and the depletion layers in the channel are in series and the gate capacitance per unit width C_g is given by

$$C_g = L_p\left(\frac{1}{C_{ox}} + \frac{1}{C_d}\right)^{-1}, \tag{2.5}$$

where C_d is the depletion layer capacitance per unit area. In the linear region, the inversion layer screens the body of the MOSFET from the gate electrode. The MOS capacitance is then just the gate-dielectric capacitance C_{ox}. In the saturation region with a significant V_{ds}, the inversion charge layer density decreases in the drain region and C_g is reduced [12].

The parasitic capacitances associated with the gate-to-source and gate-to-drain overlap regions C_{ov}, have a direct overlap component and fringing components at the outer and inner boundaries. These capacitances are in parallel with the gate-to-channel capacitance and add to the intrinsic gate capacitance. The overlap capacitance on the drain side is a weak function of gate-to-drain voltage. The contribution of C_{ov} to the total gate capacitance C_g increases as the effective channel length L_{eff}, is decreased. Its contribution is also affected during switching of a CMOS circuit when the gate and source/drain voltages vary in opposite sense with time (Miller effect). The variation in C_g with V_{gs} is shown in Fig. 2.10b.

The p/n junction capacitance of the source and drain diffusion regions have an area component, and perimeter components for the STI and gate bounded perimeters. The total capacitance of source diffusion region, C_{ds} is given by

$$C_{ds} = L_{ds}WC_{js} + (2L_{ds} + W)C_{jsws} + WC_{jswgs}, \tag{2.6}$$

where C_{js} is the capacitance per unit area and C_{jsws} and C_{jswgs} are the capacitances per unit perimeter length on the STI and gate sides respectively. A similar expression holds for the capacitance of the drain diffusion region C_{dd}, with corresponding parameters C_{jd}, C_{jswd}, and C_{jswgd}.

In addition, there are parasitic series resistances associated with the source, drain, and gate contacts which include H0 vias and M1 and higher metal layers and spreading resistance of the diffusion regions (Fig. 2.10c). These resistances are extracted from the physical layout using parasitic component models provided by the silicon foundry and included in the circuit netlist. Of these, the resistance in series with the S node, R_{ps}, is the most significant as it reduces V_{gs} by the *IR* drop across it. Since the onset of conduction is brought about by an exponential increase in I_{ds} with V_{gs}, a small change in V_{gs} has a significant impact.

Equations describing MOSFET *I–V* and *C–V* characteristics contain many terms to incorporate device physical dependencies on various process and operating parameters. BSIM compact models used in circuit simulations are generated by fitting the measured *I–V* characteristics of representative MOSFETs in a technology to the MOSFET equations. There are typically hundreds of fitting parameters required to obtain a suitable fit over the desired range of MOSFET dimensions, voltages, and temperatures. However, the basic shape of the curves can be summarized in terms of a few point values as described in Sect. 2.2.2. These parameters are correlated with circuit switching speeds and power drawn in the AC and standby states.

2.1.2 Interconnects

Metal wires are used for (1) interconnecting terminals of MOSFETs and other devices, (2) interconnecting circuit blocks, (3) distributing clock signals across chip, (4) providing a power supply grid and (5) connecting to chip I/Os. Metal wire resistances and capacitances are dependent on the physical properties of metal and dielectric layer composites, layer dimensions, and near neighbor interactions.

The resistance R of a wire of length l, width w, and thickness d shown in Fig. 2.11a is given by

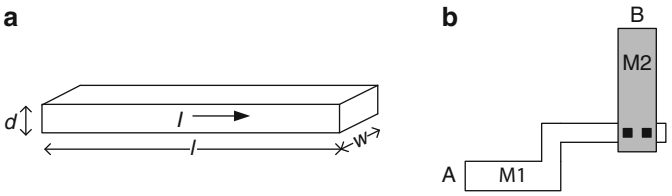


Fig. 2.11 (**a**) An isolated metal wire of length l, width w, and thickness d, and (**b**) top view of an interconnect path from point A to B through metals M1 and M2 and two inter-level vias

$$R = \rho\frac{l}{wd} = \rho_{\text{sh}}\frac{l}{w} = \rho_{\text{sh}}n_{\text{sq}}, \tag{2.7}$$

where ρ is the resistivity of the metal, ρ_{sh} is the sheet resistance, and n_{sq} is the number of squares. The quantity n_{sq} represents the number of squares ($l = w$) that fill up the resistor geometry irrespective of the magnitude of l and w. Thus a section of wire with $l = w = 10$ μm has the same resistance as a section with $l = w = 1$ μm. Resistances of vias connecting metal wires in one layer to those in metal layers above and/or below are added to get the resistance of a wire path in multiple layers. The values of ρ_{sh} of metal layers and resistances of vias of fixed dimensions are available from the silicon foundry for estimating interconnect resistances.

Although it is straightforward to compute the resistance of a wire of rectangular geometry, many wires have complex shapes with width variations and bends as illustrated in Fig. 2.11b. Wire models are provided by the silicon foundry and used in conjunction with physical extraction tools to determine interconnect resistances.

Wire resistance is a function of temperature. For metal wires the resistance varies linearly with temperature over practical temperature ranges

$$R_2 = R_1\{1 + \text{TCR}(T_2 - T_1)\}, \tag{2.8}$$

where R_1 and R_2 are resistances at temperatures T_1 and T_2, and TCR is the temperature coefficient of resistance.

In addition to parasitic resistances, metal interconnects also add parasitic capacitances to circuits. A capacitor formed by two parallel plates of length l and width w and separated by a dielectric of thickness h is shown in Fig. 2.12a.

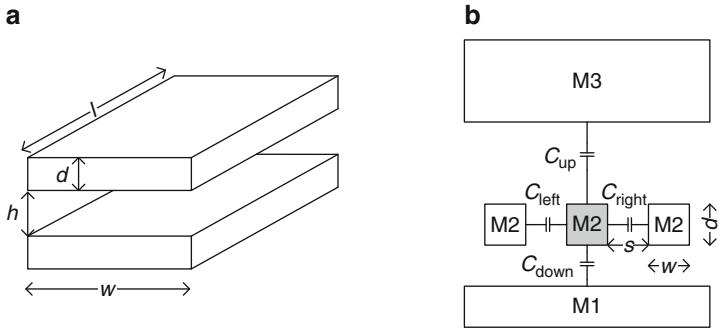


Fig. 2.12 (**a**) A parallel plate capacitor with plates of length l, width w, and separation h, and (**b**) M2 signal wire (shown in *gray tone*) and its capacitance components C_{left}, C_{right}, C_{up}, and C_{down} with its nearest neighbors

In the simplest case of a parallel plate capacitor with plate dimensions larger than the separation h, the capacitance C, is given by

$$C = \frac{\varepsilon\varepsilon_0 lw}{h} = \frac{\varepsilon\varepsilon_0 A}{h} \quad \text{for} \quad h \ll l, w. \tag{2.9}$$

In general, the wire capacitance has components to the nearest neighbors on the same layer and to the wires in layers above and below. These capacitance components are depicted as C_{left}, C_{right}, C_{up}, and C_{down} in Fig. 2.12b. The total switching capacitance of a wire during a transition is determined by the relative node voltages of the surrounding wires. If all neighboring nodes are at a fixed potential, and the signal switches between V_{DD} and GND, the total switching capacitance per unit length is given by

$$C_{\text{w}} = C_{\text{up}} + C_{\text{down}} + C_{\text{left}} + C_{\text{right}}. \tag{2.10}$$

The capacitance between wires is doubled when switching simultaneously in opposite sense and is zero when switching in the same sense.

Interconnect resistances and capacitances add to signal propagation delay in a circuit. A lumped element equivalent circuit shown in Fig. 2.13a may be used in circuit simulations. Here R_{w} and C_{w} are the resistance and capacitance per unit length of the wire. For very long wires, inductance L comes into play as well and in that case a distributed transmission line *RLC* model shown in Fig. 2.13b is used. The inductance per unit length is given by

$$L_{\text{w}} = g\mu_0, \tag{2.11}$$

where $\mu_0 = 1.26 \times 10^{-6}$ H/m (=1.26 pH/μm) is the permeability of free space and g is a geometric factor with a value in the range of 0.5–1.5.

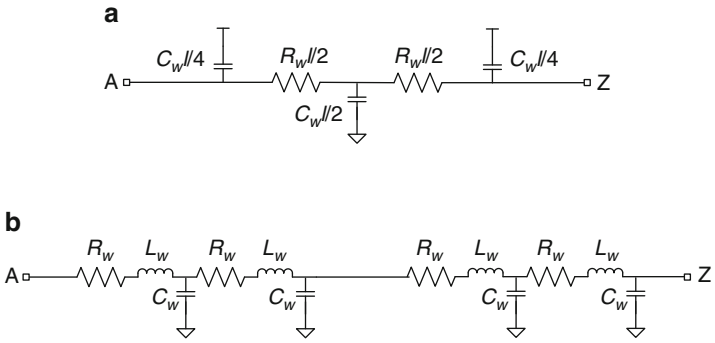


Fig. 2.13 (**a**) Lumped *RC* model of a wire of length *l*, and (**b**) distributed transmission line *RLC* model of a long wire showing four sections with length $l = 1$ in each section

2.1.3 Passive *R* and *C* Components

Silicon DF or polysilicon PS layers with higher sheet resistances than metal interconnect layers are utilized for defining precision resistors. Such resistors are used in analog and I/O circuits. The temperature dependence of these resistors is nonlinear and quadratic and higher order terms may be required in an equation analogous to Eq. 2.8.

On-chip decoupling capacitors (DECAPs) are used to help maintain a stable power supply voltage during switching of AC circuits. These are placed in unused silicon area on the chip in close proximity to regions of high switching activity. The schematic cross section of a metal-dielectric-silicon DECAP is shown in Fig. 2.14a. This configuration gives a high capacitance per unit area (~10 fF/μm^2). Resistance of the diffusion and PS layers appears in series with the capacitor. The equivalent circuit in Fig. 2.14b is used to compute the *RC* time constant and the maximum frequency for effective operation of the DECAP.

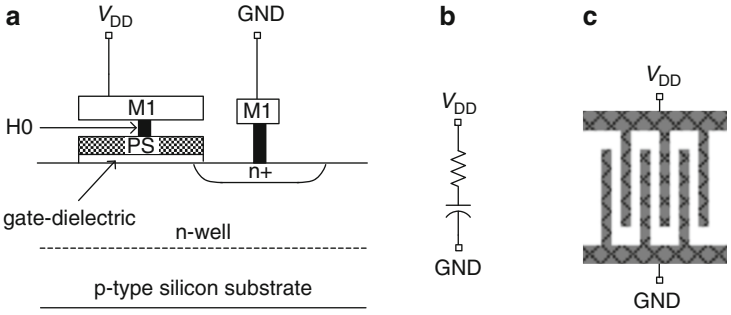


Fig. 2.14 (**a**) Schematic cross section of a metal-dielectric-silicon DECAP and (**b**) its equivalent circuit, and (**c**) interdigitated metal DECAP

Interdigitated metal capacitors are also offered for use as DECAPs in many CMOS technologies. The metal fingers in several metal layers are stacked to

increase the capacitance per unit area. This type of capacitor may be used when the metal wiring is less dense and large open areas in the metal layers are available. Typical capacitance per unit area is a factor of 10 lower than for MOS capacitors.

2.1.4 Logic Gates

Static CMOS circuits comprise combinational logic gates with one or more inputs and one output. The simplest logic gate is an inverter with a single input. The circuit schematic of an inverter and its symbol are shown in Fig. 2.15a and b respectively. An inverter performs the function of inverting the input signal voltage at its output, i.e., if the input voltage is at V_{DD} ("1" or high), its output voltage is at GND ("0" or low), and conversely if the input voltage is at "0" its output voltage is at "1". The logic truth table describing this function is shown in Fig. 2.15c.

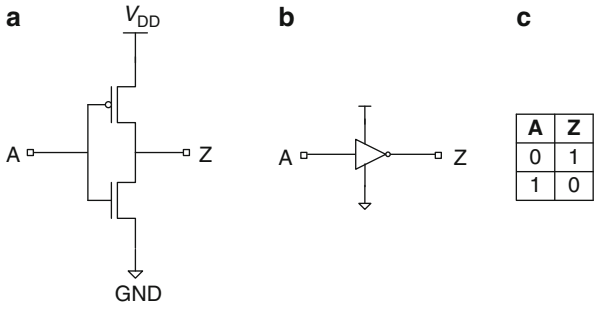


Fig. 2.15 An inverter: (**a**) circuit schematic, (**b**) symbol and (**c**) logic truth table

Input and output voltage waveforms of an inverter driving a fixed capacitive load C_L are shown in Fig. 2.16. The input signal is rising and the output voltage level is pulled down from a "1" to a "0" as the signal passes through the inverter. This is called a pull-down (PD) transition. Conversely, with input signal falling and output rising, the transition is a pull-up (PU) transition. The input signal has a rise time for the signal to increase from $0.1 \times V_{DD}$ to $0.9 \times V_{DD}$, of τ_r. The signal propagation delay τ through the inverter is measured between the $V_{DD}/2$ points of the input and output waveforms.

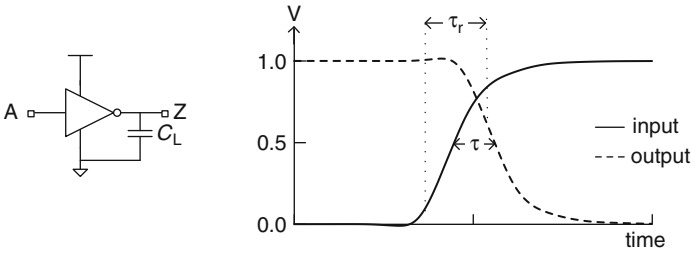


Fig. 2.16 Voltage waveforms at the input and output of an inverter driving a fixed load capacitance C_L

Even though the function of an inverter is a simple logical inversion, this structure's characteristics, such as delay and power consumption, are representative of other static CMOS logic gates described in this section and relate directly to CMOS chip cycle time and power.

Two other basic logic functions are AND and OR. Circuit symbol, logical function, and truth tables for 2-input AND and OR gates are shown in Fig. 2.17. The output Z of an AND gate is a logical "1" only if both inputs are at "1" and expressed as $Z = A \cdot B$, where A and B represent the input levels. The output of an OR gate is a logical "1" if either or both inputs are at "1" and expressed as $Z = A + B$.

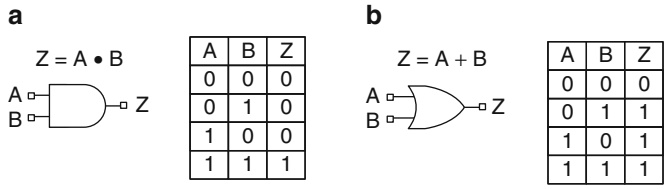


Fig. 2.17 Circuit symbol and truth table of (**a**) AND gate and (**b**) OR gate

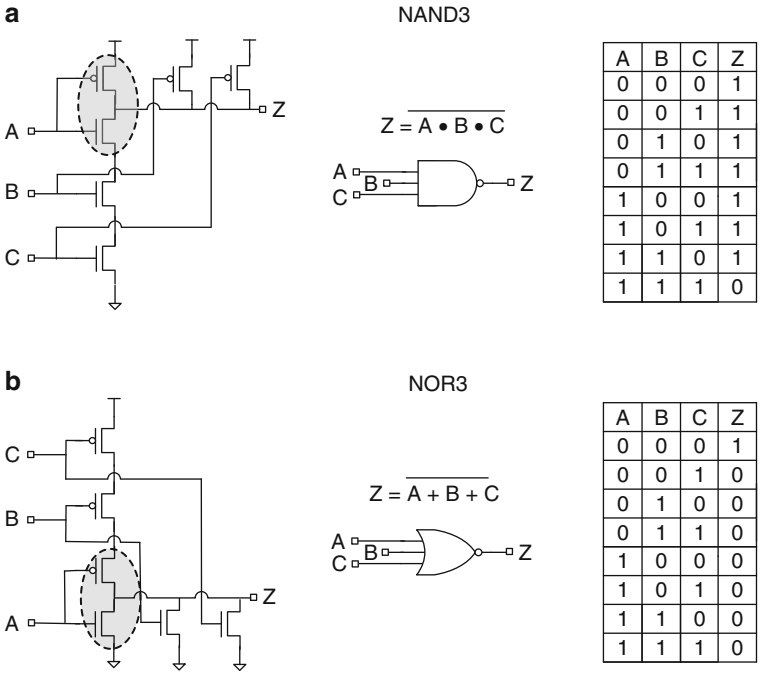
Static CMOS circuits use inversion logic in which the outputs of the logic gates are inverted. Hence the basic functions become NAND (AND + inverter) and NOR (NOR + inverter).

NAND and NOR logic gates may have two, three or more inputs, and a single output. Circuit schematics of a 3-input NAND gate (NAND3) and a 3-input NOR gate (NOR3) are shown in Fig. 2.18. The output signal voltage is a function of the states of all the inputs as listed in the truth tables. A gray ellipse highlights the inverter section of the schematic. A NAND gate can be described as an inverter with one or more n-FETs in series with the inverter n-FET and p-FETs in parallel with the inverter p-FET. Similarly a NOR gate can be described as an inverter with series p-FETs and parallel n-FETs.

Other commonly used logic gates are XOR (exclusive OR), XNOR (exclusive NOR), AOI (AND-OR-Invert), and OAI (OR-AND-Invert). Their symbols and logical functions are shown in Fig. 2.19. The output of an XOR2 or exclusive OR gate is a "0" when both of its inputs are at the same logic level, either "1" or "0". Its output is a "1" when only one of its two inputs is at "1". XNOR2 inverts the output of XOR2.

Another set of circuits that is useful for evaluating silicon technology performance comprises inverters driving a single ended n-FET (n-passgate or NPG) or p-FET (p-passgate or PPG) or a transmission gate (TG). These circuit configurations are shown in Fig. 2.20. The passgates act as switches to pass or to block signals.

A significant fraction of the signal propagation delay through a logic gate can be attributed to the internal resistances and capacitances of the MOSFETs and the gate capacitance of the following logic gates. Interconnecting wires add to the *RC* delay, their contribution increasing with interconnect length. Short interconnects are



A	B	C	Z
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

A	B	C	Z
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	0

Fig. 2.18 Circuit schematic, symbol, and truth table of (**a**) 3-input NAND gate (NAND3) and (**b**) 3-input NOR gate (NOR3). *Gray ellipses* indicate inverter schematic

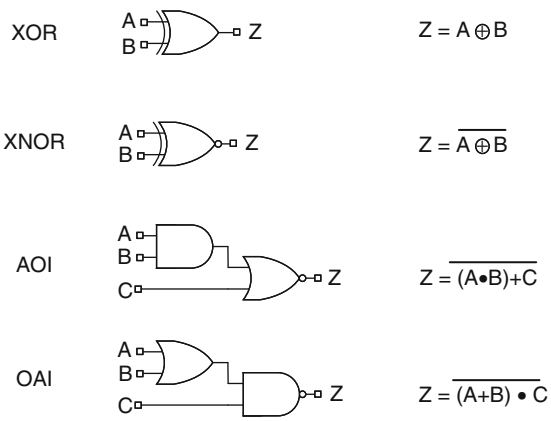


Fig. 2.19 Circuit symbols and logic functions of XOR2, XNOR2, and three input AOI and OAI gates

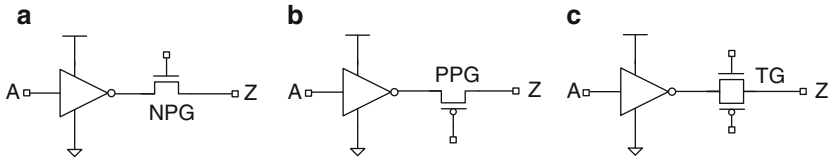


Fig. 2.20 Circuit schematic of an inverter driving (**a**) an n-FET passgate (NPG), (**b**) a p-FET passgate (PPG) and (**c**) a transmission gate (TG)

primarily capacitive whereas long interconnects are modeled as distributed *RC* networks as shown in Fig. 2.13a.

Interconnect delay responds differently to changes in voltage and temperature than do MOSFET delays. It is instructive to simulate the behavior of a logic gate driving both interconnect and MOSFET gate loads. Two circuit configurations for studying logic gate response to interconnect loads are shown in Fig. 2.21. The ratio of interconnect to gate load may be varied to characterize their relative contributions to propagation delay.

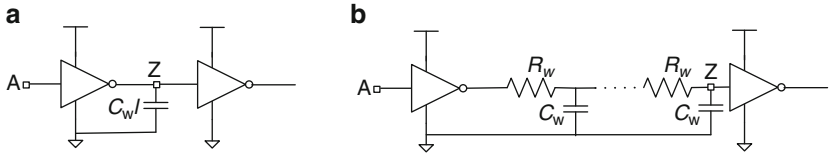


Fig. 2.21 Inverter driving (**a**) interconnect capacitive load and (**b**) interconnect distributed *RC* load

2.2 SPICE Simulations

In this section, methodologies for setting up circuit simulations of MOSFETs and logic gates are described. The simulations are geared towards evaluating technology models, circuit performance, and power. The small circuit blocks investigated here can be simulated on a personal computer using simulation tools available in the public domain or with licensed software and tools for designing large CMOS chips on work stations. Emphasis is placed on the role of the basic underlying physics at the device level and on illustrating how this propagates to a physical understanding of circuit behavior.

The essential elements of a circuit simulation setup are shown in Fig. 2.22. A schematic editor may be used for a pictorial representation of a circuit. A netlist describes the circuit elements and their interconnections as a text description of a circuit schematic. Although not essential, the netlist can be generated automatically with a netlist generator coupled to a schematic editor. This automates the tedious task of generating a netlist manually and facilitates hierarchical views of complex circuits which are then easier to debug.

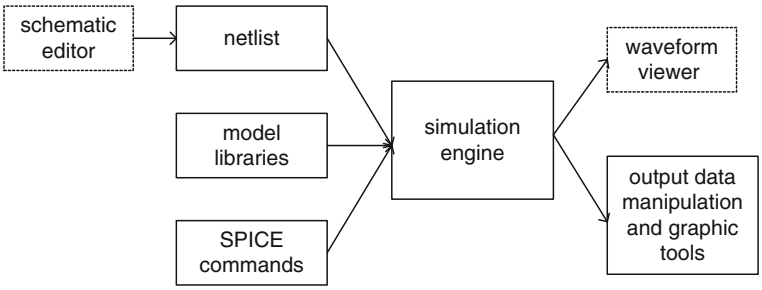


Fig. 2.22 Block diagram showing elements of a circuit simulation setup

Model libraries contain parameters and equations describing the behavior of each circuit element. SPICE commands describe the settings under which circuits are simulated and assert output requests. The simulation engine processes these inputs and generates the requested outputs. A waveform viewer is a convenient tool for viewing the voltages and currents at different nodes. Raw data output may be exported to a data analysis tool or a spreadsheet for further manipulation and for generating customized graphical displays.

The sources of MOSFET models and tools for simulation and analysis used in this book are listed in Table 2.1.

Table 2.1 Models and tools used for circuit simulations and analysis

Components and tools	Source	Model/version
MOSFET models	BSIM PTM (Arizona State University)	45 nm (HP & LP), 32 nm (HP & LP), 22 nm (HP & LP)
Circuit component library	LTspice	IV
Schematic editor	LTspice	IV
Netlister	LTspice	IV
Simulation engine	LTspice	IV
Waveform viewer	LTspice	IV
Data export and graphics	Microsoft Office Excel	2003 or later

BSIM models for MOSFETs issued by U.C. Berkley are charted by a consortium of semiconductor companies and simulation tool vendors. Silicon foundries release BSIM models for each technology offering. These models are used for chip design and for manufacturing process control. There may be several model releases by a foundry for the same technology node reflecting technology modifications that impact circuit design. The models are proprietary and typically available only to customers of the foundry.

The predictive technology models (PTM) are distributed by Arizona State University. The model cards can be downloaded from their website [5]. These

models are based on CMOS scaling rules in advance of full technology maturity for early circuit design analysis. They do not represent any particular silicon foundry offering but are for the most part a reasonable representation of a technology node. Hence, the circuit simulation data presented here are for illustration purposes only.

The schematic editor and waveform viewing capabilities of LTspice IV are very useful for generating circuit netlists and for debug. LTspice can be run on Microsoft Windows or OS X operating systems and contains model libraries for standard circuit elements such as resistors, capacitors, and voltage sources. The examples given here are generic although some SPICE commands may be specific to LTspice. With minor adaptation, the commands can be used with other versions of SPICE such as HSPICE and Ngspice [2, 3].

Each circuit component in a schematic is identified with a unique symbol. A circuit component may have multiple instances. Each instance in a circuit is given a unique instance name. Some properties of an instance may be assigned in the schematic as attributes. The attributes may be values or variables. The circuit element names and their equivalent LTspice names, symbols, and key attributes are listed in Fig. 2.23, with further elaboration in Fig. 2.24 and Table 2.3. Note that in the three terminal symbols for nmos and pmos the body is tied to the source terminal.

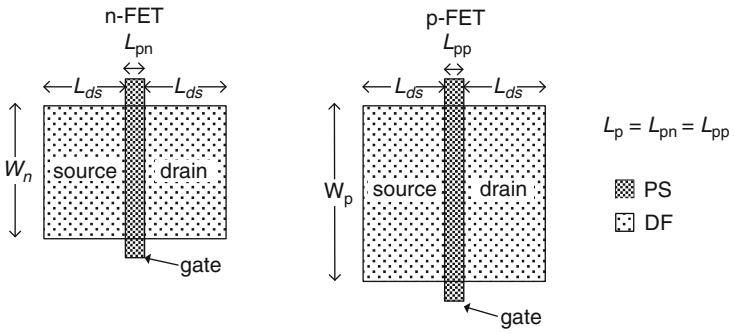


Fig. 2.24 MOSFET physical layouts showing the dimensions of gate, source, and drain regions

We begin with simulating MOSFET characteristics and follow this with circuit simulations of small circuit blocks. Three different circuit configurations are used for characterizing such circuit blocks. These are:

- Individual circuit blocks with variable input waveforms and output load
- Delay chains comprising series-connected circuit blocks
- Ring oscillators comprising series-connected circuit blocks configured in a closed loop

Individual circuit blocks are simulated in a manner similar to that used for characterization of logic gates in a standard cell library. Delay chain and ring

circuit element	LTspice component name	symbol	attributes
n-FET	nmos		*w l lds ad as pd ps nf delvto*
p-FET	pmos		*w l lds ad as pd ps nf delvto*
resistor	res		*R*
capacitor	cap		*C*
voltage source	bv, voltage		*V*
current source	bi, current		*I*

Fig. 2.23 Symbols and attributes used in LTspice simulations for different circuit elements

oscillator configurations are commonly used in test structures placed on silicon to monitor the CMOS process and for model-to-hardware correlation. Such applications of delay chains and ring oscillators are covered in Chaps. 4–10.

2.2.1 PTM (BSIM)

The PTM releases are available from the 180 nm technology node through the 7 nm node. Models in version V2.1 have been released in 16, 22, 32, and 45 nm for high performance (HP) and low power (LP) applications. Models for future technologies (7, 10, 14, 16, and 20 nm nodes) at the time of publication of this book (2014) use multi-gate transistors. Model releases for these advanced technologies are for high performance (HP) and low standby power (LSTP).

The 45 nm PTM for high performance (HP) and low power (LP) MOSFETs were released in September of 2008 and November of 2008 respectively [5]. The BSIM4 model cards for these MOSFETs are included in Appendix B. The model cards have been modified to run in LTspice. The model level has been changed from 54 to 14. A prefix "m" has been added to the MOSFET model names: nmos to **m**nmos and pmos to **m**pmos to meet LTspice requirements.

The nominal operating voltage V_{DD} and channel length L_p for the 45 nm PTM models are listed in Table 2.2. It is assumed that these parameters are normally distributed with a mean corresponding to the nominal value and a standard deviation σ specified in the BSIM models (Sect. 9.1.2). Systematic process variations in L_p and V_t are expressed as standard deviations of their distributions, σL_p and σV_t respectively. Typically circuit properties are examined over $\pm 3\sigma L_p$ and $\pm 3\sigma V_t$ ranges. Other sources of variations are covered in Chap. 6.

Table 2.2 Nominal parameters for PTM models for the 45 nm technology node

MOSFET	PTM	V_{DD} (V)	L_p (μm)	σL_p (μm)	σV_t (V)
n-FET	HP	1.0	0.045	0.015	0.020
	LP	1.1	0.045	0.015	0.020
p-FET	HP	1.0	0.045	0.015	0.020
	LP	1.1	0.045	0.015	0.020

Parasitic resistances and capacitances which are normally extracted from physical layouts are generally not included in the examples in this book. This simplifies the netlists, and the results are independent of physical layout styles. Full parasitic extraction is essential for precise evaluation of circuit behavior and for model-to-hardware correlation.

2.2.2 MOSFET Characteristics

MOSFET parameters related to a physical layout are specified as attributes to the instance symbol. Representative layout of an n-FET and a p-FET along with their key dimensions are shown in Fig. 2.24. Channel length, L_{pn} for n-FET and L_{pp} for p-FET, is the gate dimension in the direction of current flow. Generally a common value of gate length L_p ($=L_{pn}=L_{pp}$) is used for both n-FET and p-FET. The channel widths and corresponding widths of source and drain regions are W_n for n-FET and W_p for p-FET. The n-FET and p-FET widths are allowed to vary independently. In this symmetric layout, the lengths of source and drain regions are both equal to L_{ds}. The source and drain area (*as* and *ad*) and perimeter (*ps* and *pd*) are defined in terms of W_n, W_p, and L_{ds}. These relationships are listed in Table 2.3. The LTspice parameters are in normal italic font (lower case) whereas the parameters in the text, figures, and equations use the normal convention of subscripts, superscripts, and upper and lower cases.

Table 2.3 MOSFET parameters in LTspice, and corresponding symbols used in the text

Parameter	Description	n-FET	p-FET
w	Channel width	W_n	W_p
l	Channel length	L_{pn}	L_{pp}
lds	Diffusion area length	L_{ds}	L_{ds}
as	Source diffusion area	$L_{ds} \times W_n$	$L_{ds} \times W_p$
ad	Drain diffusion area	$L_{ds} \times W_n$	$L_{ds} \times W_p$
ps	Source diffusion perimeter	$2 \times (L_{ds} + W_n)$	$2 \times (L_{ds} + W_p)$
pd	Drain diffusion perimeter	$2 \times (L_{ds} + W_n)$	$2 \times (L_{ds} + W_p)$

It is sometimes convenient to assign variables to the MOSFET attributes. Values of the variables are then assigned in LTspice commands. In a hierarchical schematic view, values of the variables can be assigned at the top level of the hierarchy as

described in Sect. 2.2.4. The variables assigned to MOSFET attributes are listed in Table 2.4. The variables for n-FET have a prefix 'n' and for p-FET have a prefix 'p'.

Table 2.4 Variables used in n-FET (prefix 'n') and p-FET (prefix 'p') attributes

Parameter	Description	n-FET	p-FET
w	Channel width	*nw*	*pw*
l	Channel length	*nl*	*pl*
lds	Diffusion area length	*lds*	*lds*
nf	Number of fingers	*nnf*	*pnf*
delvto	ΔV_t	*ndelvto*	*pdelvto*

The circuit schematic of an n-FET for LTspice simulation of I_{ds}–V_{ds} and I_{ds}–V_{gs} characteristics is shown in Fig. 2.25a, and for a p-FET in Fig. 2.25b. Note that the MOSFET node voltages are measured with respect to source, which is at GND for n-FET and at V_{DD} for p-FET. The current and voltages, I_{ds}, V_{ds}, and V_{gs}, are positive for n-FET and negative for p-FET in their on-states.

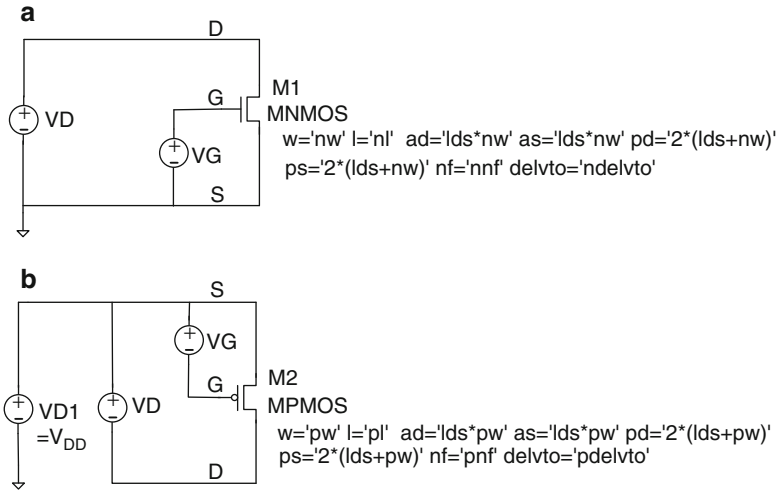


Fig. 2.25 Circuit schematic for simulating I_{ds}–V_{ds} characteristics in LTspice for (**a**) an n-FET and (**b**) a p-FET

We start with a complete LTspice input file for simulating the I_{ds}–V_{ds} characteristics of an n-FET. The text is separated into sections each beginning with a comment line in bold and preceded by *. First the LTspice, BSIM library paths, and model cards are specified. Simulator options are set to aid convergence and improve error tolerance. These values override the default values in the simulator. The netlist gives connectivity of the two power supplies (VD and VG) with drain, gate, source, and body terminals of the NMOS instance M1, in that order.

The variables used in the attributes of n-FET instance M1 are assigned values characteristic of the 45 nm technology node. The nominal value of the channel length $L_p = nl$ is 0.045 μm. The n-FET width $W_n = nw$ is set to 1.0 μm, convenient for obtaining the current drive per unit width. The temperature is set at room temperature, 25 °C, overriding the default temperature of 27 °C.

The simulation is run by sweeping the power supply voltage VD from 0.0 V to V_{DD} in steps of 0.01 V. Five such sweeps are made with VG power supply voltages at fractional values of V_{DD}, i.e., V_{gs}/V_{DD} of 0.0 V, 0.25, 0.50, 0.75, and 1.0. Alternatively five instances of identical n-FETs may be included in the netlist, each with a unique value of VG. For plotting I_{ds}–V_{gs} characteristics of an n-FET in the linear and saturation regions, the power supply VG in Fig. 2.25a is swept from 0.0 V to V_{DD} and VD is stepped from 0.05 V to V_{DD}.

An LTspice deck for simulating n-FET I_{ds}–V_{ds} characteristics is listed below:

*** n-FET I_{ds}-V_{ds} characteristics using 45 nm PTM HP models**
*** measure n-FET I_{on} in μA/μm at 1.0V, 25°C**
*** library paths and models cards**
.lib C:\Program Files\LTC\LTspiceIV\lib\cmp\standard.mos
.inc C:\Users\model_library
.model NMOS NMOS
***simulator options**
.option gmin=1e-14 abstol=0.1pA reltol=1e-5 noopiter
*** netlist: voltage sources VG, VD and MOSFET M1 (drain. gate, source, body)**
VG g 0
VD d 0
*M1 d g 0 0 MNMOS w='nw' l='nl' ad='lds*nw' as='lds*nw' pd='2*(lds+nw)' ps='2*(lds+nw)' nf='nnf' delvto='ndelvto'*
*** variables**
.param nw=1u nl=0.045u lds=0.12u nnf=1 ndelvto=0.0
.temp 25
*** DC sweep**
.dc VD 0 1.0 0.01
.step VG 0 1.0 0.25
.end

LTspice features output currents flowing through the MOSFET instance. These are labeled Id(m1), Ig(m1), Is(m1), and Ib(m1) as the drain, gate, source, and body currents in instance M1. The currents as a function of the voltage being swept (VD or VG) may be viewed in the waveform viewer. For customized graphics, the data points may be exported to a spreadsheet (e.g., Microsoft Office Excel) or another software tool with similar capabilities.

In the case of a p-FET, D is at GND potential and all voltages are measured with respect to the S terminal which is at a higher potential. The current flow is via holes from source to drain and the sign of the conventional current is negative, opposite

that in an n-FET. It is convenient to multiply the p-FET I_{ds} and V_{ds} by -1 so that the n-FET and p-FET characteristics may be overlaid for comparison.

The I_{ds}–V_{ds} characteristics of an n-FET at V_{gs}/V_{DD} values of 0.5, 0.75, and 1.0 are shown in Fig. 2.26a. The V_{ds} range is from 0 to V_{DD}. In Fig. 2.26b, the I_{ds}–V_{gs} characteristics are plotted at V_{ds} values of 0.05 and 1.0 V.

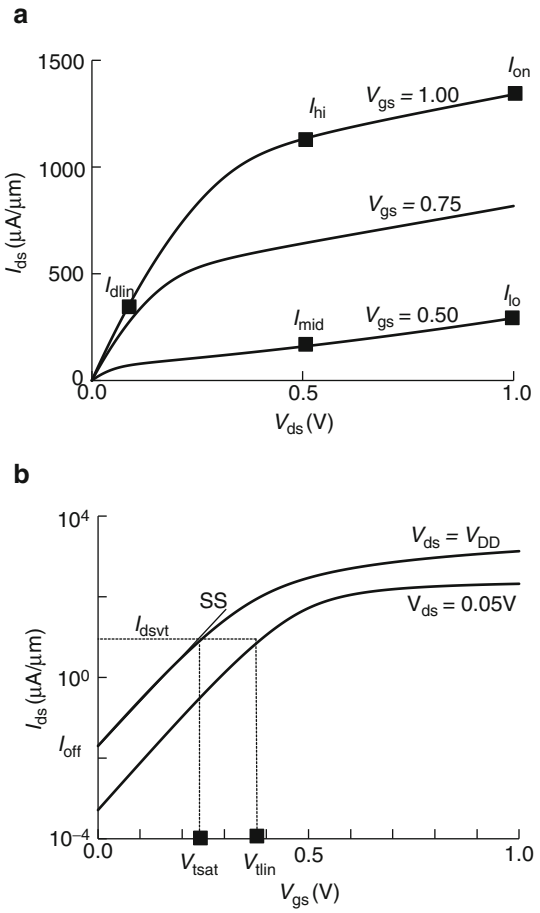


Fig. 2.26 n-FET: (**a**) I_{ds}–V_{ds} characteristics and (**b**) I_{ds}–V_{gs} characteristics. 45 nm PTM HP models at 25 °C

The I_{ds}–V_{ds} and I_{ds}–V_{gs} characteristics are summarized in terms of a few locations on the curves. The location of the parameters on I_{ds}–V_{ds} and I_{ds}–V_{gs} plots are shown in Fig. 2.26. The definitions of measurement locations for I_{on}, I_{hi}, I_{lo}, I_{mid}, and I_{off} are given in Table 2.5. In thin gate-dielectric MOSFETs, the leakage current through the gate-dielectric I_{gl} can become significant. It is dependent on both V_{gs} and V_{ds} and is maximum when $V_{gs} = 1.0$ V and $V_{ds} = 0.0$ V.

Table 2.5 Source, drain, and gate voltages for measuring n-FET parameters with the body tied to source

MOSFET parameter	Source	Drain	Gate
I_{off}	GND	V_{DD}	0
I_{dlin}	GND	0.05–0.10 V	V_{DD}
I_{mid}	GND	$V_{DD}/2$	$V_{DD}/2$
I_{lo}	GND	V_{DD}	$V_{DD}/2$
I_{hi}	GND	$V_{DD}/2$	V_{DD}
I_{on}	GND	V_{DD}	V_{DD}
I_{gl}	GND	GND	V_{DD}
V_{tlin}	GND	0.05 V	
V_{tsat}	GND	V_{DD}	

LTspice commands for measuring I_{on} in μA/μm are shown below. Normalized values of other I_{ds} parameters (I_{hi}, I_{lo}, I_{mid}, I_{off}, and I_{gl}) with appropriate V_{ds} and V_{gs} bias voltages are obtained in a similar fashion.

*** measure I_{on}**
.measure dc Ion find id(M1) at=1.0 **set VG=1.0*
.measure Ion_uA_um param = 'Ion/nw'

The threshold voltage V_t is obtained from I_{ds}–V_{gs} characteristics. Here V_t is defined at a fixed current, I_{dsvt} (positive for n-FET, negative for p-FET). For the 45-nm high-performance (HP) technology node the I_{dsvt} values are selected to be

$$I_{dsvt} = 300\ \text{nA} \times \frac{W_n}{L_p}; \quad \text{n-FET}, \tag{2.12}$$

and

$$I_{dsvt} = 100\ \text{nA} \times \frac{W_p}{L_p}; \quad \text{p-FET}. \tag{2.13}$$

The subthreshold slope (SS) is calculated as the slope of $\log_{10}(I_{ds})$ vs. V_{gs} line between the values of $V_{gs} = 0$ and $V_{gs} = V_t$.

The netlist for power supplies and LTspice commands to measure I_{off} in nA/μm, V_t and SS for a fixed value of VD are listed below:

*** measure n-FET I_{off}, V_t and SS**
.dc VG 0 1.0 0.01
.step VD 0.05 1.0 0.95
*.measure dc Ioff_n find id(M1)*1000000000 at=0*
.measure dc Vtn find V(G) when id(M1)=(300e-9(nw/nl))*
.measure SS param=Vtn/(log10(300(nw/nl))-log10(Ioff_n))*1000*

**** measure p-FET V_t***
*.measure dc Vt_g find V(G) when id(M2)=(-1*100e-9*(pw/pl))*
.measure Vtp param = 'V(Vdd)-Vt_g'

Key derived parameters for MOSFETs are listed in Table 2.6. The effective current, I_{eff} is a measure of the current drive of the MOSFET during switching. It correlates well with the signal propagation delay of logic gates as discussed in Sect. 2.2.3. In short-channel MOSFETs, V_t is lowered with increase in V_{ds}. This effect is known as drain-induced barrier lowering (DIBL), the difference between V_{tlin} and V_{tsat} is a measure of its magnitude.

Table 2.6 Definitions of calculated electrical parameters of a MOSFET

Parameter	Definition	Comments
I_{eff}	$(I_{hi}+I_{lo})/2$	Correlates to circuit delay
SS	$dV_{gs}/d(\log_{10} I_{ds})$	mV/decade, $0 < V_{gs} < V_t$
DIBL	$(V_{tlin} - V_{sat})$	Varies with L_p (SCE)
g_m	dI_{ds}/dV_{gs} at constant V_{ds}	Saturation region
g_{ds}	dI_{ds}/dV_{ds} at constant V_{gs}	Saturation region
γ	$1 + dV_t/dV_{bs}$	$V_{bs} \approx 0$

The output conductance g_{ds} is the change in output current I_{ds} in the saturation region with output voltage V_{ds}. The transconductance g_m is the change in I_{ds} in response to a change in input voltage V_{gs} and is a measure of the gain of a MOSFET. Both g_{ds} and g_m are important parameters in the design of analog circuits.

The subthreshold characteristics of a MOSFET can be modulated by independently controlling its body voltage, V_{bs}. With the body at a negative voltage with respect to the source in an n-FET, there is an increase in V_t and hence a decrease in I_{off} as $|V_{bs}|$ is increased. The body-effect coefficient γ gives a measure of the sensitivity of V_t to V_{bs} near $V_{bs} = 0$. This effect can be utilized in reducing off-state power as discussed in Chap. 4.

The I_{ds}–V_{ds} and I_{ds}–V_{gs} plots for an n-FET and p-FET are shown in Fig. 2.27 (all p-FET currents and voltages are shown as positive). The n-FET and p-FET characteristics are overlaid for ease of visual comparison. From the I_{ds}–V_{ds} plots it is apparent that the n-FET current is higher than the p-FET, and this difference varies within the I_{ds}–V_{ds} space.

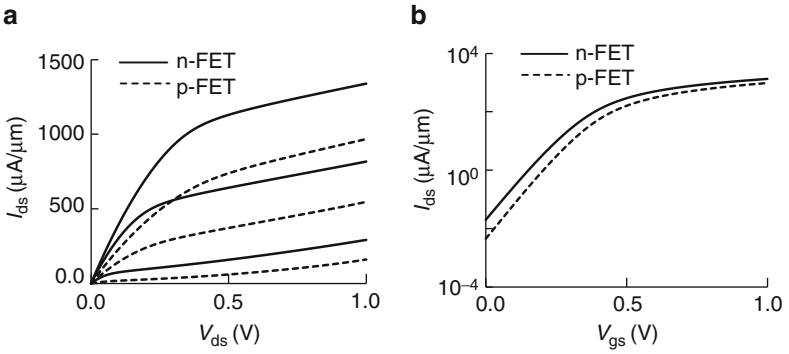


Fig. 2.27 n-FET and p-FET characteristics: (**a**) I_{ds}–V_{ds} for $V_{gs} = 0.5$, 0.75 and 1.0 V and (**b**) I_{ds}–V_{gs} plots for $V_{ds} = 1.0$ V. 45 nm PTM HP models at 25 °C

Visual inspection of the I_{ds}–V_{ds} curves gives a qualitative assessment of the relative strengths of the n-FET and the p-FET. For a quantitative assessment, the I_{ds} values specified in Tables 2.5 and 2.6 are recorded. These values for HP and LP models are shown in Table 2.7. The HP model has >3× higher current in the saturation region than LP resulting in faster switching speeds in CMOS circuits. This comes at a cost of higher I_{off} and higher power in the off-state by a factor of >200.

Table 2.7 Properties of nominal p-FETs and n-FETs in 45 nm PTM models for HP @ 1.0 V and LP @ 1.1 V, 25 °C

	Model	V_{DD} (V)	I_{on} (μA/μm)	I_{eff} (μA/μm)	I_{dlin} (nA/μm)	I_{mid} (μA/μm)	I_{off} (nA/μm)
n-FET	HP	1.0	1,339	711	394	159	20
p-FET	HP	1.0	968	450	229	60	4.5
n-FET	LP	1.1	525	239	186	7.6	0.024
p-FET	LP	1.1	310	136	92	4.9	0.020

The relative strengths or current drive capabilities of n-FET and p-FET in a technology are critical in sizing the MOSFET widths in logic gates. Typically the I_{on} ratio is used as a measure of n/p strength. However, it is apparent from Fig. 2.27a that relative current drives of n-FET and p-FET are functions of V_{ds} and V_{gs}. By taking the ratios of the n-FET and p-FET I_{ds} parameters listed in Table 2.7, the relative strengths of the two MOSFET types I_{ds}(n/p) are evaluated in different regions of the I_{ds}–V_{ds}–V_{gs} space. These ratios are listed in Table 2.8. The same data are presented in a bar chart in Fig. 2.28.

Table 2.8 Ratios of n-FET and p-FET key I_{ds} values in 45 nm PTM models @ 1.0 V, 25 °C

Model	V_{DD} (V)	I_{on} (n/p)	I_{eff} (n/p)	I_{dlin}(n/p)	I_{mid} (n/p)	I_{off} (n/p)
HP	1.0	1.38	1.58	1.72	2.65	4.44
LP	1.1	1.69	1.75	2.02	1.55	1.20

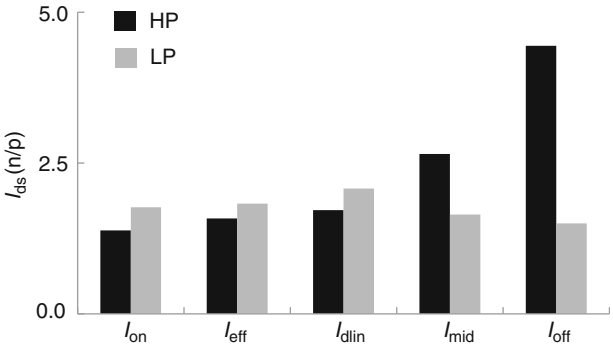


Fig. 2.28 Bar chart showing n/p ratios in Table 2.8

Let us consider the variation in two of the most significant MOSFET parameters, L_p and V_t. The parameter values may be changed in LTspice simulations by using an adder to the nominal value. An adder parameter for V_t is provided in the BSIM model. This parameter, *delvto*, has a nominal value of 0. For varying L_p, adder parameters *ndl* for n-FET and *pdl* for p-FET are introduced. The LTspice commands for these variables are listed below:

```
* specifying parameter values for ΔLp and ΔVt
.param nl = '(0.045u+ndl)'
.param ndl = 0.015u          * increase n-FET Lp by 0.015 μm
.param ndelvto = 0.020       * increase n-FET Vt by 20 mV
.param pdelvto = - 0.020     * increase p-FET Vt by 20 mV
```

Note that to increase $|V_t|$, *ndelvto* is positive and *pdelvto* is negative.

The variations of I_{on} and I_{off} for an n-FET as a function of L_p are shown in Fig. 2.29. The range of L_p shown corresponds to a variation of $\pm 3\sigma L_p$ as listed in Table 2.2 and includes 99.7 % of the population. Over this range, the I_{on} swing with respect to nominal is from +22 % (L_p at its -3σ value) to −14 % (L_p at its $+3\sigma$ value). At these same extreme values of L_p, I_{off} varies from 27× to 0.2× of its nominal value.

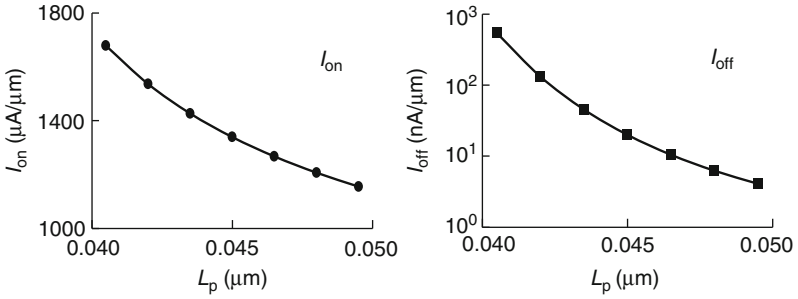


Fig. 2.29 n-FET current vs. channel length, L_p: (**a**) I_{on} and (**b**) I_{off}. 45 nm PTM HP models @ 1.0 V, 25 °C

The variations of I_{on} and I_{off} for an n-FET as a function of a shift in V_{tn} from nominal ($ndelvto = \Delta V_{tn}$) over the $\pm 3\sigma V_{tn}$ range are shown in Fig. 2.30. I_{on} varies linearly with V_{tn}, with a slope of 2.41 μA/μm per mV shift in V_{tn}. I_{off} vs. $\Delta|V_{tn}|$ is plotted on a log-linear scale in Fig. 2.30b. The value of SS obtained from the slope of the line fit to $\log_{10} I_{off}$ vs. $\Delta|V_{tn}|$ is 82.5 mV/decade.

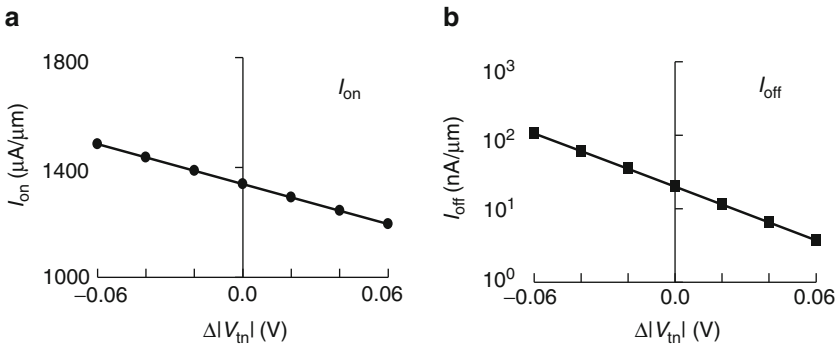


Fig. 2.30 n-FET current vs. $\Delta|V_{tn}|$: (**a**) I_{on} and (**b**) I_{off}. 45 nm PTM HP models @ 1.0 V, 25 °C

MOSFET parameters influencing the DC characteristics such as L_p and V_t vary nearly independently within their specified distributions. There is some interplay between these parameters such as V_t roll-off at short channels ($L_p <$ nominal). To obtain the full range of variations, Monte Carlo simulations are carried out by independently varying all of the MOSFET parameters for which distributions are available. This is described in Sect. 2.2.7.

Another important subject of MOSFET characterization and simulation is capacitance. The C_g–V_{gs} characteristics of an n-FET may be simulated using the circuit schematic shown in Fig. 2.31a. The S and D terminals of the n-FET are connected together and a DC voltage bias VGB is applied to the G terminal. A small signal voltage source VM connected in series with VGB swings the voltage at terminal G from (VGB − VM/2) to (VGB + VM/2) as shown in Fig. 2.31b. The current flowing through VM is integrated to give the charge transferred in raising the voltage across the capacitor by VM. The capacitance, C_g as a function of VGB $= V_{gs}$ is calculated using the equation

$$C_g = \frac{Q_c}{\text{VM}} = \frac{\int I(\text{VM})dt}{\text{VM}}. \tag{2.14}$$

Here Q_c is the total charge transferred during the VM transition. The pulse rise time is selected to be sufficiently short so that the charging current is at least a few orders of magnitude larger than the gate-to-body leakage current.

The netlist and LTspice commands for simulating the circuit shown in Fig. 2.31a are listed below.

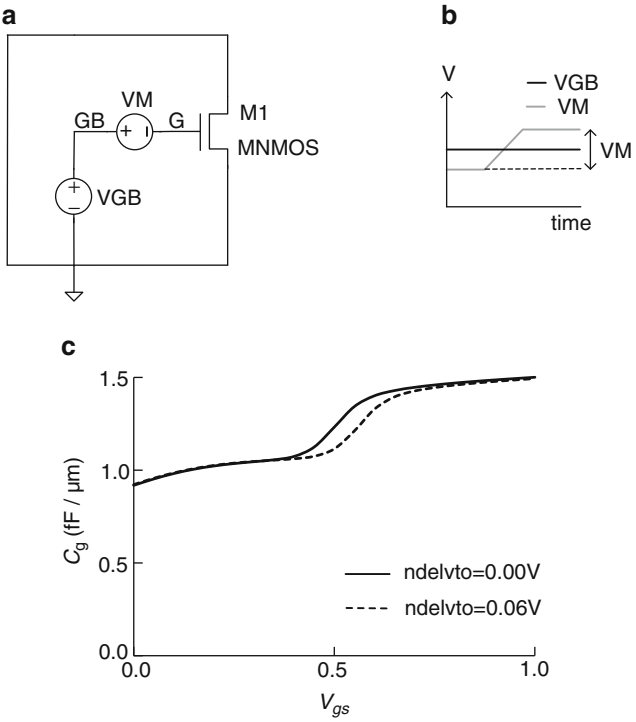


Fig. 2.31 (**a**) Circuit schematic for simulating C_g–V_{gs} characteristics of an n-FET, (**b**) voltage levels for VGB and VM, and (**c**) C_g vs. V_{gs} for an n-FET for nominal V_{tn} and with $\Delta|V_{tn}| = 0.06$ V. 45 nm PTM HP models at 25 °C

*** netlist and commands for measuring n-FET C_g in fF/µm at V_{gs} = pvdd**

```
.VGB GB 0 =pvdd
.VM  GB G pulse(-0.005 0.005 0 100e-12 100e-12 5e-9 10e-9 1)
.step param pvdd 0 1.0 0.05
.measure tran i_charge_rise integ i(VM) from=0 to 100E-12
.measure cg param= 'i_charge_rise/0.01/nw*1e15'
.tran 1E-9
```

Simulation results for a 45 nm HP n-FET with $W_n = 1.0$ µm, and $L_p = 0.045$ µm are shown in Fig. 2.31c. The plots compare C_g vs. V_{gs} for a nominal V_t n-FET with an n-FET whose V_t is raised by 0.06 V. The C_g values in the off- and on-states for the two cases are the same but the transition region for the higher V_t n-FET is shifted to the right as expected.

The circuit configuration described in Fig. 2.31a is similar to the arrangement used for measuring C_g of MOSFETs in silicon. These measurements are used for model building and for monitoring the process in the manufacturing line.

A circuit for directly obtaining a plot of C_g vs. V_{gs} in an LTspice waveform viewer is shown in Fig. 2.32. The voltage of the VG power supply is increased linearly from 0 to V_{DD} with VD = 0. The charging current is measured through the zero-voltage source VM. As C_g is charged to 1.0 V, the instantaneous current through VM is directly proportional to the instantaneous value of C_g, while V_{gs} is linearly proportional to time. By integrating the charging current through VM and dividing by the final value of VGB, the average switching capacitance is obtained.

In logic gates, when an n-FET is in the on-state, its V_{ds} = "0" and when it is in the off-state, its V_{ds} = "1". The circuit schematic in Fig. 2.32a can also be used for measuring n-FET capacitance under different switching conditions. The rising waveform for V_{gs} and the simultaneously falling waveform for V_{ds} are shown in Fig. 2.32b. This imitates the case of an n-FET in an inverter as its output switches from a "1" to a "0" (PD transition). In Fig. 2.32c, C_g–V_{gs} characteristics during an inverter PD transition and with $V_{ds} = 0$ are compared. Clearly the capacitance during switching, with V_{gs} and V_{ds} transitioning in opposite sense, is higher due to the Miller effect.

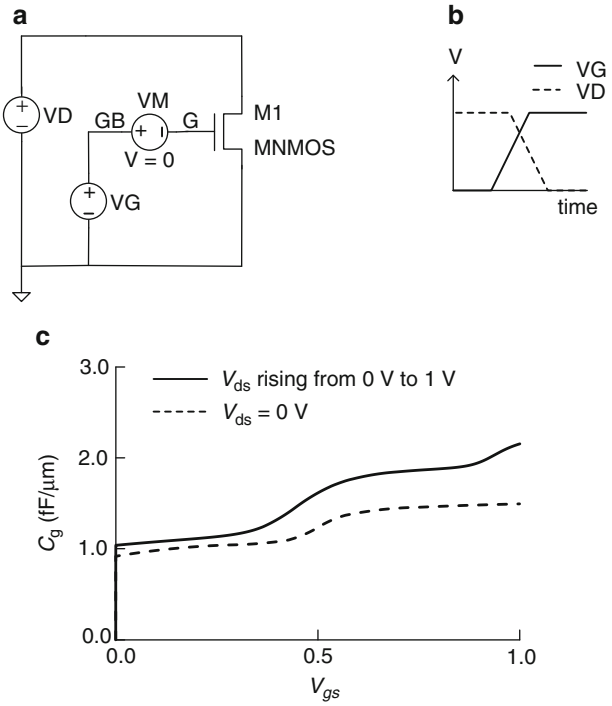


Fig. 2.32 (**a**) Circuit schematic for simulating C_g–V_{gs} characteristics of an n-FET with V_{gs} rising and V_{ds} falling, (**b**) voltage levels for VG and VD, and (**c**) C_g vs. V_{gs} for an n-FET during a PD transition and with $V_{ds} = 0$. 45 nm PTM HP models at 25 °C

The C_g values of an n-FET and a p-FET from different simulation conditions are summarized in Table 2.9. The average capacitance during a switching transition of an inverter is nearly equal to the on-state capacitance if V_{gs} and V_{ds} are changing in the opposite sense and ~20 % lower if $V_{ds} = 0$. Average gate capacitances for different V_{gs}, V_{ds}, and V_{bs} waveforms may be extracted in a similar fashion.

Table 2.9 C_g for n-FET and p-FET at static points and averaged over different transitions. 45 nm PTM HP models @ 25 °C

Method	$\|V_{gs}\|$ (V)	$\|V_{ds}\|$ (V)	n-FET C_{gn}(fF/μm)	p-FET C_{gp} (fF/μm)
Static	0.0	0.0	0.99	0.91
Static	1.0	0.0	1.49	1.53
Average	0.0–1.0	0.0	1.24	1.21
Average	0.0–1.0	1.0–0.0	1.53	1.51

A circuit schematic for obtaining g_m and g_{ds} of an n-FET from simulations is shown in Fig. 2.33. For measuring g_m, I_{ds} is measured with a DC bias applied to the gate and modulated by VMG at a fixed V_{ds} (=VDB). For measuring g_{ds}, I_{ds} is measured with a DC bias applied to the drain and modulated by VMD at a fixed V_{gs} (=VGB). LTspice commands for determining g_m are listed below:

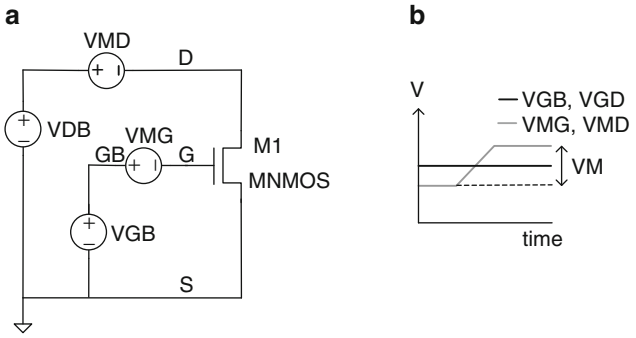


Fig. 2.33 (**a**) Circuit schematic, and (**b**) voltage waveforms for simulating n-FET transconductance g_m, and output conductance, g_{ds}

*** voltage sources and gm measurement**

.VMG GB G PULSE(-0.01 0.01 100e-12 20e-12 20e-12 5e-9 10e-9 1)
.measure tran ids1 avg id(M1) from= 0E-12 to =50E-12
.measure tran ids2 avg id(M1) from =150E-12 to =200E-12
.measure del_ids param=(ids2 -ids1)
*.measure gm param = del_ids/0.02*1e6*

MOSFETs for analog circuits may be engineered with longer channel lengths than high-performance MOSFETs and are described by different BSIM models.

2.2.3 Standard Cell Library Book Characteristics

In CMOS circuit design, individual logic gates are characterized by measuring the propagation delays through the gate and the output waveform rise and fall times while varying the input signal waveform shape and output capacitance load C_L. This type of characterization can be easily carried out in simulation and is convenient for building a set of lookup tables for gate delays which are then used to compute cycle time for logic paths comprising multiple gates. However, the data obtained from simulation at the gate level cannot be easily validated in hardware. The difficulty of accurately measuring the input and output waveform shapes and propagation delays of the order of a few picoseconds on silicon hardware prohibits direct experimental verification of simulation results on individual gates.

A standard cell library comprises physical layouts, circuit schematics and symbols for commonly used logic gate families and small circuit blocks. There are multiple books of each logic gate type with different strengths (MOSFET widths, width ratios, and V_t values). Netlists of logic books with all of the parasitic resistances and capacitances extracted from physical layouts are included.

Electrical characteristics of each logic book are determined from circuit simulations as shown in Fig. 2.34. The library book X1, represented by a box in Fig. 2.34a, may be a logic gate LG, or a circuit block, and may have an inverted or non-inverted output signal. The book drives a fixed capacitive load of value C_L. A zero-voltage source VI is connected in series with the input node A to measure the current drawn for charging the input capacitance of the logic book. The value of the input capacitance C_{in}, is determined by integrating the current through the voltage source VI $\{\int I(\mathrm{VI})dt\}$ during a transition.

The voltage sources for the power supply and for launching a pulse at input node A are shown in Fig. 2.34b. The voltage waveform of the input pulse is depicted in Fig. 2.34c. After time T_d, the signal rises to V_{DD} or "1" with a transition time from GND to V_{DD} of T_r. After time T_{on}, the signal falls to GND or "0" with a transition time of T_f. The waveform is generated using a PULSE function in SPICE and the signal rises and falls linearly with time. Alternatively other waveform shapes described by exponential or PWL functions may be used. The standard definition of signal rise and fall times τ_r and τ_f is the time for the voltage level to rise or fall between $0.1 \times V_{DD}$ and $0.9 \times V_{DD}$. In case of the linear ramp shown here, $\tau_{ri} = 0.8 \times T_r$ and $\tau_{fi} = 0.8 \times T_f$.

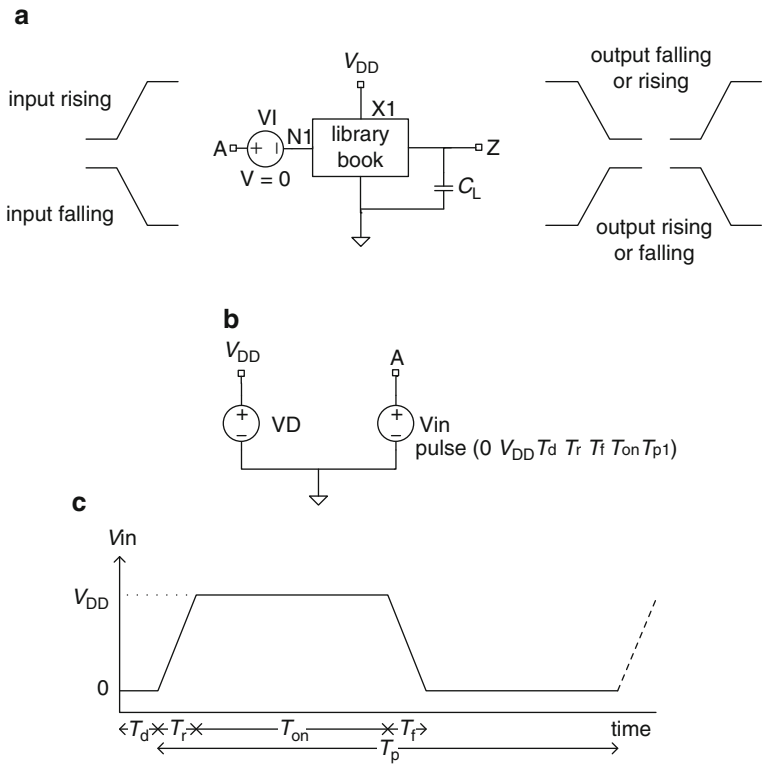


Fig. 2.34 Circuit simulation setup for library book characterization: (**a**) circuit schematic, (**b**) voltage sources and (**c**) input signal waveform for one cycle with T_d = time delay, T_r = rise transition time, T_{on} = on time, T_f = fall transition time, and T_p = pulse period

Circuit simulations are carried out by varying the input rise time τ_{ri}, fall time τ_{fi}, and output load C_L. The propagation delays of the signal through the logic book for pull-down (PD) and pull-up (PU) transitions, τ_{pd} and τ_{pu}, and the corresponding output signal fall time τ_{fo}, and rise time τ_{ro} are measured. The parameter names and corresponding LTspice names are listed in Table 2.10.

An inverter logic book circuit schematic is shown in Fig. 2.35a. The attributes of the n-FET and the p-FET of the inverter are displayed in the schematic. As in the case of MOSFET characterization in Sect. 2.2.2, variables are assigned to the MOSFET properties. The physical layout of an inverter with one PS finger ($nnf = pnf = 1$) is shown in Fig. 2.35b. This design with $W_n = 0.4$ µm and $W_p = 0.6$ µm is used as a "standard" inverter design for characterization using 45 nm PTM HP models. The sum of these widths, $(W_n + W_p) = 1.0$ µm is selected for the convenience of directly expressing capacitance in units of fF/µm. The ratio, $W_p/W_n = 1.5$ is selected to render nearly equal delays for PD and PU transitions. The length of the source and drain regions L_{ds} is 0.12 µm corresponding to a contacted PS pitch of 0.165 µm. The source and drain capacitance values are

Table 2.10 Circuit and LTspice parameter names and descriptions

	Parameter	LTspice parameter	Description
Inputs	τ_{fi}	*tfi*	Input signal fall time
	τ_{ri}	*tri*	Input signal rise time
	C_L	*cl*	Load capacitance
Outputs	τ_{pd}	*tpd*	PD delay
	τ_{pu}	*tpu*	PU delay
	τ_p	*delay_stage*	Average delay $= (\tau_{pu} + \tau_{pu})/2$
	τ_{fo}	*tfo*	Output signal fall time
	τ_{ro}	*tro*	Output signal rise time
	C_{in}	integ *i*(*vi*)/*pvdd*	Input capacitance
	IDDQ	*i*(*vd*) at time $< T_d$	Current in the off-state

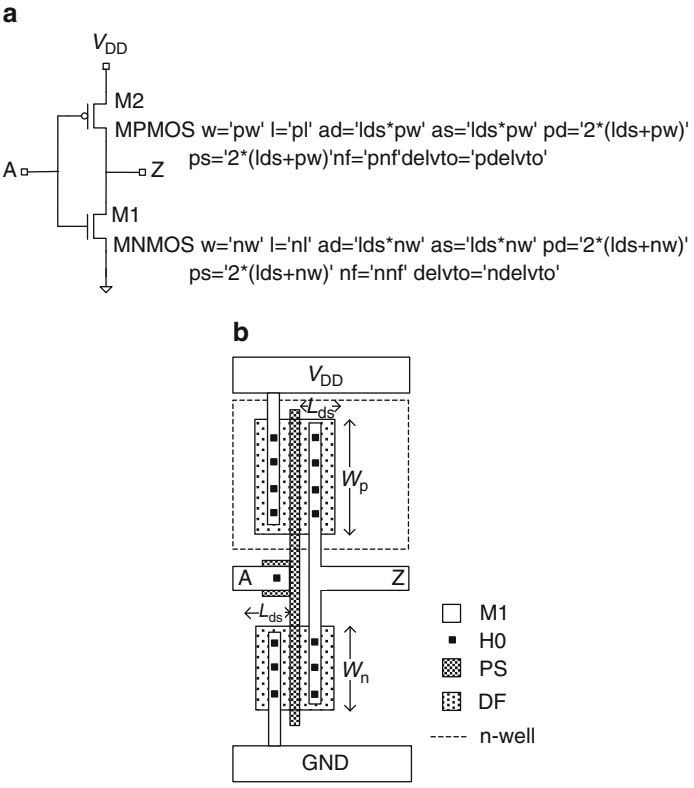


Fig. 2.35 (**a**) Circuit schematic of an inverter with MOSFET attributes, (**b**) physical layout of an inverter with one PS finger

computed during simulations from the BSIM model equations for junction areas and perimeters listed in Table 2.3. Other parasitic resistances and capacitances associated with interconnect wiring are not included in the simulation results described here.

The inverter is defined as a subcircuit (subckt) "inv" and instantiated as X1 in the circuit schematic in Fig. 2.34a. The netlist of "inv" subckt in LTspice is listed below:

```
* netlist for inverter subcircuit
.subckt  inv A VDD 0 Z
M1 Z A 0 0 MNMOS w='nw' l='nl' ad='lds*nw' as='lds*nw' pd='2*(lds+nw)'
ps='2*(lds+nw)' nf='nnf' delvto='ndelvto'
M2 Z A VDD VDD MPMOS w='pw' l='pl' ad='lds*pw' as='lds*pw' pd='2*(lds+pw)'
ps='2*(lds+pw)' nf='pnf' delvto='pdelvto'
```

A full netlist with representative parameter values and commands follows:

```
*inverter characterization
*netlist
XX1 N1 VDD 0 Z inv
C1   Z 0 =cl
VI A N1 0
VD VDD 0 =pvdd
Vin A 0 PULSE(0 =pvdd 100e-12 =tr =tf 400e-12 2e-9 1)
*parameter values
.param pvdd=1.0
.temp 25
.param nw=0.4u pw=0.6u lds=0.12u nnf=1 pnf=1
.param nl='(0.045u+ndl)'  pl='(0.045u+pdl)'  ndl=0.0u  pdl=0.0u
.param ndelvto=0  pdelvto= 0
.param tr=20E-12 tf=20E-12
.tran 1E-9
*CL parameter sweeps
.step param cl 0f 15f 3f
*measure τpd, τri, τfo
.measure tran tpd + trig v(A) val=0.5*pvdd rise=1 + targ v(Z) val=0.5*pvdd fall=1
.measure tran tri + trig v(A) val=0.1*pvdd rise=1+ targ v(A) val=0.9*pvdd rise=1
.measure tran tfo + trig v(Z) val=0.1*pvdd fall=1+ targ v(Z) val=0.9*pvdd fall=1
*compute input capacitance Cin in fF for signal rising (PD transition)
.measure tran in_charge_rise integ i(VI) from=100e-12 to=150e-12
.measure Cin_rise  param = 'in_charge_rise/pvdd*1e15'
*measure leakage current (IDDQ) in off-state
.measure tran IDDQ avg i(VD)*-1 from=10e-12 to=50E-12
```

Characteristics of the standard inverter PD transitions with input rise times τ_{ri} of 6, 16, and 32 ps, and C_L values of 3 and 15 fF are obtained from circuit simulations

using 45 nm PTM HP models at 1.0 V and 25 °C. The findings are recorded in Table 2.11. From a closer inspection of the data in Table 2.11 it is apparent that τ_{pd} increases with τ_{ri} and C_L, while C_{in} remains nearly constant. IDDQ is measured with input A at "0", prior to any transition. Lookup tables of this type for PD and PU transitions over a range of τ_{ri} and C_L are generated using design automation software tools for a wide variety of gates.

Table 2.11 Standard inverter parameters for PD transitions with variable inputs τ_{ri} and C_L. 45 nm PTM HP models @ 1.0 V, 25 °C

τ_{ri} (ps)	C_L (fF)	τ_{pd} (ps)	τ_{fo} (ps)	C_{in} (fF)	IDDQ, A at "0" (nA)
8	3.0	10.3	15.9	1.50	9.35
8	15.0	19.5	34.4	1.51	9.35
32	3.0	15.3	21.5	1.50	9.35
32	15.0	24.9	37.6	1.51	9.35
56	3.0	18.1	28.2	1.50	9.35
56	15.0	30.2	43.4	1.50	9.35

The simulation sets described above are repeated at different values of V_{DD}, temperature, and MOSFET parameters L_p and V_t. In order to limit the number of simulation runs and the size of lookup tables, a few fixed simulation corners are defined as nominal (NOM), worst-case (WC), and best-case (BC) for delays.

Typical scenarios for defining NOM, WC, and BC simulation corners are shown in Table 2.12. Here it is assumed that delay decreases with V_{DD} and increases with temperature, L_p and V_t. Although generally true, the delay trend with temperature may reverse in some technologies. The V_{DD} and temperature values in these corners are based on product specifications and environmental operating conditions. MOSFET parameter values are based on the parameter distributions in the models as illustrated in Table 2.2. Circuit design methodology varies from product to product and simulation corners may be defined at $\pm 2\sigma$, $\pm 3\sigma$ or other suitable parameter values. Additional corners may be defined as needed.

Table 2.12 Three commonly used simulation corners to cover product application conditions and process variations

Simulation corner	V_{DD}	Temperature	L_p	V_{tn}, V_{tp}
Nominal (NOM)	Nominal	Nominal	Nominal	Nominal
Worst-case (WC)	Minimum	Maximum	Maximum	High
Best-case (BC)	Maximum	Minimum	Minimum	Low

The total delay of a logic path comprising different logic gate library books is calculated using the delay of each gate in the lookup tables. C_L is replaced by C_{in} of the following gate, and the output rise/fall time of the preceding gate becomes the input rise/fall time of the following gate. A logic path may be represented as a black box characterized by an input capacitance and output rise/fall times, similar in

many respects to Fig. 2.34a. The path may, however, have multiple inputs and outputs. This methodology continues up the circuit hierarchy.

Although lookup tables are extremely useful in timing tools for circuit design, it is difficult to obtain physical insight into the behavior of logic gates and circuits from lengthy tables. In addition, a software bug in generating such tables can introduce timing errors which may require costly fixes if detected later in hardware test. It is therefore important to understand the dependencies of circuit delays and capacitances on the properties of MOSFETs and parasitic elements under different environmental conditions. This knowledge can then be applied to validate timing and power EDA tools with test cases designed to catch nonphysical circuit behavior.

An equivalent *RC* model of a logic gate is helpful in gaining physical insight and in relating gate delay and power characteristics to the constituent MOSFET and parasitic elements. We will demonstrate this with an inverter. This methodology is extended to other logic gates in Chap. 5 for more extensive model-to-hardware correlation of CMOS chip circuitry.

The circuit schematic and an equivalent *RC* circuit for an inverter are shown in Fig. 2.36a and b respectively. The inverter changes state as its internal switch is toggled between two possible positions in response to changes in voltage levels at input A. With input A at "0", output Z is at "1", connected to V_{DD} through R_{swp}.

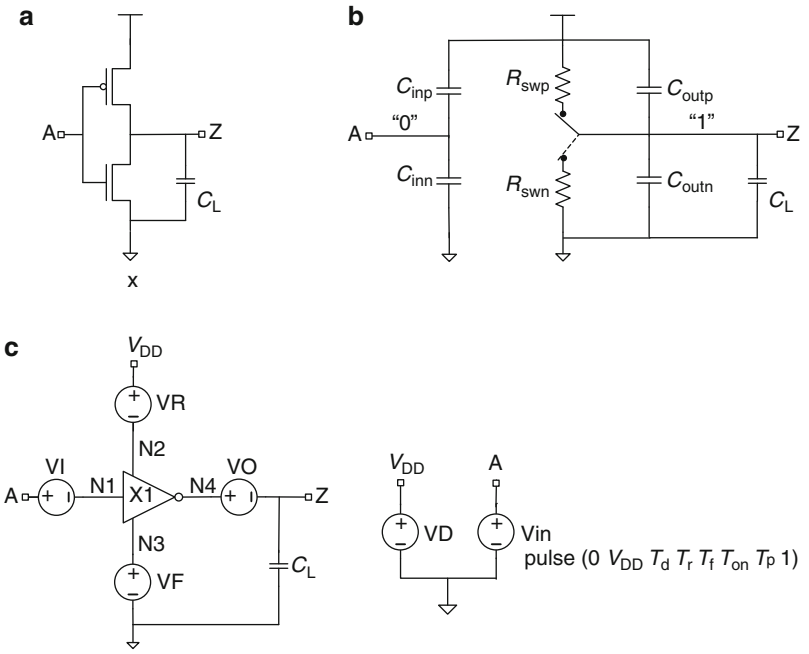


Fig. 2.36 (**a**) Circuit schematic of an inverter driving a capacitive load C_L, (**b**) equivalent *RC* model, and (**c**) circuit simulation setup for inverter characterization

When input A transitions from a "0" to a "1", the switch toggles to its lower position, and output Z is pulled down to a "0" with an *RC* time delay determined by R_{swn} and the total capacitance being discharged. When input A subsequently transitions from a "1" to a "0", the switch closes through R_{swp} connecting Z to V_{DD} and the capacitance is charged. Input capacitances C_{inn} and C_{inp} are charged and discharged through the power supply Vin.

The capacitances C_{in} and C_{out} are the sum of their n-FET and p-FET components:

$$C_{in} = C_{inn} + C_{inp}, \tag{2.15}$$

$$C_{out} = C_{outn} + C_{outp}. \tag{2.16}$$

The PD and PU delays are given by

$$\tau_{pd} = R_{swn}(C_{out} + C_L), \tag{2.17}$$

$$\tau_{pu} = R_{swp}(C_{out} + C_L). \tag{2.18}$$

The average delay τ_p is given by

$$\tau_p = \frac{\tau_{pu} + \tau_{pd}}{2} = \frac{1}{2}(R_{swn} + R_{swp})(C_{out} + C_L), \tag{2.19}$$

or

$$\tau_p = R_{sw}(C_{out} + C_L), \tag{2.20}$$

where R_{sw} is the average switching resistance for PD and PU transitions.

The effective resistances of the n-FET and the p-FET during switching are R_{swn} and R_{swp} and are related to their current drive strengths. Input capacitance C_{in} of the inverter is the sum of gate, overlap and other parasitic capacitances of the n-FET and p-FET denoted by C_{inn} and C_{inp} respectively. On the output side, internal capacitances C_{outn} and C_{outp} of the inverter add to the load. The main contributors to C_{outn} and C_{outp} are the junction capacitances on the drain side of the MOSFETs with additional contributions from gate and overlap capacitances [12].

Keeping the *RC* model in mind, simulations of an inverter are carried out to determine key parameters that characterize its properties and performance. A circuit schematic to determine C_{in}, C_{out}, R_{swn}, R_{swp}, and their relationships to gate delay and to the properties of the constituent MOSFETs is shown in Fig. 2.36c. Zero-voltage sources VR, VF, and VO are included in addition to VI introduced in the schematic in Fig. 2.34a. The charging and discharging currents through these sources are integrated over time to obtain the corresponding capacitances. The charging current for total output capacitance $(C_L + C_{out})$ flows through VR during a PU transition and the discharging current for $(C_L + C_{out})$ during a PD transition flows through VF. A small additional current flowing between V_{DD} and GND when

both n-FET and p-FET are momentarily "on" during switching adds to the extracted capacitance. This is discussed in more detail in Sect. 4.3.2.

C_{in} is determined by integrating the charging current through zero-voltage source VI. Similarly, C_L can also be measured from the charging current through VO. As C_L is an input parameter in the simulation runs, its measured value may be compared with the input value to validate the simulation setup and SPICE commands. R_{swn} and R_{swp} are calculated from measured τ_{pd}, τ_{pu}, and capacitances using Eqs. 2.17 and 2.18. The average delay τ_p and average switching resistance R_{sw} are also computed from the measured data.

Circuit simulations are carried out while varying input fall and rise times, τ_{fi} and τ_{ri}, capacitive load C_L, total n-FET and p-FET width ($W_n + W_p$), and width ratio W_p/W_n. All simulation runs are made at the nominal V_{DD} of 1.0 V for 45 nm PTM HP models. Instead of viewing the simulation results in a table, the data are displayed graphically in Figs. 2.37, 2.38, 2.39, 2.40, 2.41, 2.42, and 2.43. Linear equations describing a least square fit to the data and the regression coefficient R^2 (Sect. 9.2) indicating the goodness of fit are displayed on the plots wherever applicable.

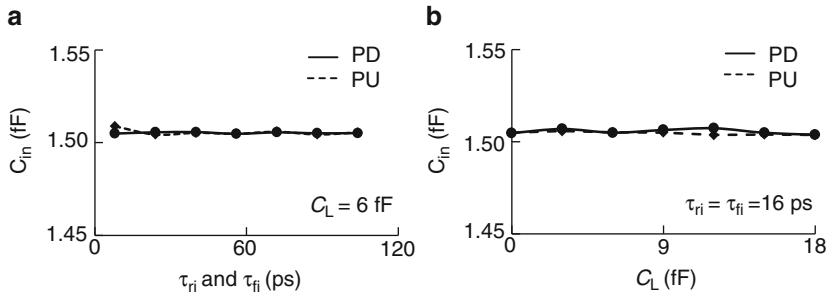


Fig. 2.37 Standard inverter input capacitance C_{in} vs. (**a**) input τ_{ri} and τ_{fi} with $C_L = 6$ fF and (**b**) output load C_L with $\tau_{ri} = \tau_{fi} = 16$ ps. 45 nm PTM HP models @ 1.0 V, 25 °C

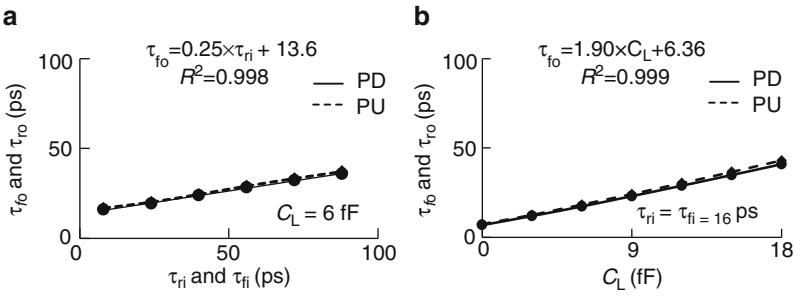


Fig. 2.38 Standard inverter output τ_{fo} and τ_{ro} vs. (**a**) input τ_{ri} and τ_{fi} with $C_L = 6$ fF and (**b**) output load C_L with $\tau_{ri} = \tau_{fi} = 16$ ps. 45 nm PTM HP models @ 1.0 V, 25 °C

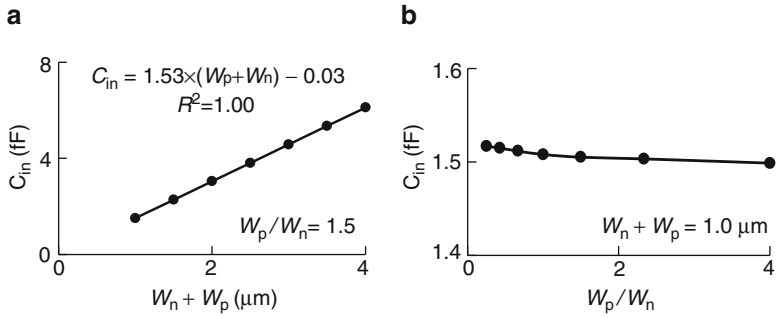


Fig. 2.39 Inverter C_{in} as a function of: (**a**) $(W_n + W_p)$ with $W_p/W_n = 1.5$ and (**b**) W_p/W_n with $(W_n + W_p) = 1.0$ µm. 45 nm PTM HP models @ 1.0 V, 25 °C

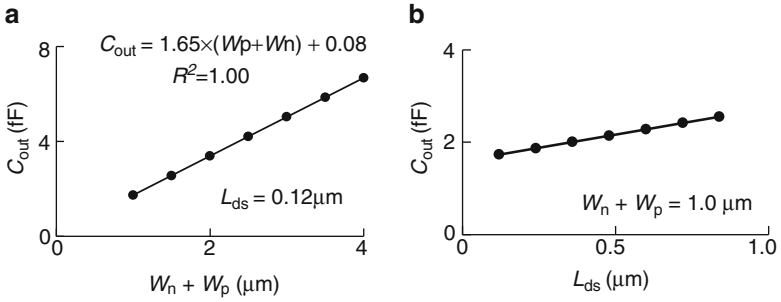


Fig. 2.40 Inverter C_{out} as a function of (**a**) $(W_n + W_p)$, with $W_p/W_n = 1.5$ and $L_{ds} = 0.12$ µm and (**b**) L_{ds} with $(W_n + W_p) = 1.0$ µm. 45 nm PTM HP models @ 1.0 V, 25 °C

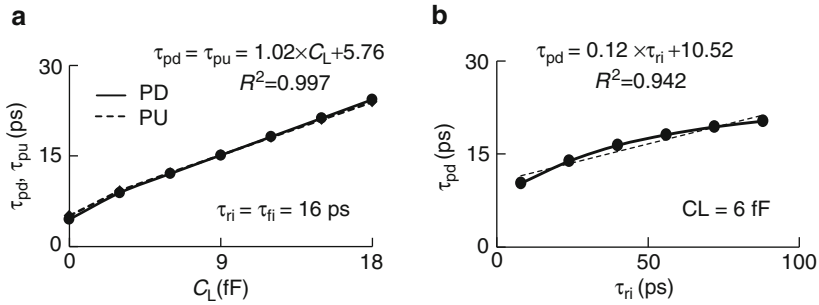


Fig. 2.41 Standard inverter: (**a**) signal propagation delays, τ_{pd} and τ_{pu} vs. C_L and (**b**) τ_{pd} vs. τ_{ri} for $C_L = 6$ fF. 45 nm PTM HP models @ 1.0 V, 25 °C

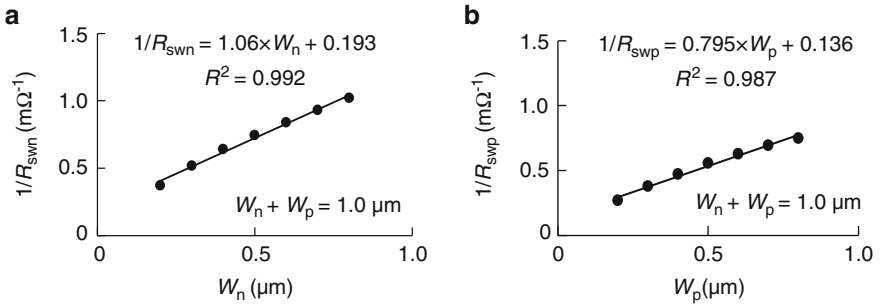


Fig. 2.42 Inverter (FO = 4): (**a**) $1/R_{swn}$ as a function of W_n and (**b**) $1/R_{swp}$ as a function of W_p. 45 nm PTM HP models @ 1.0 V, 25 °C

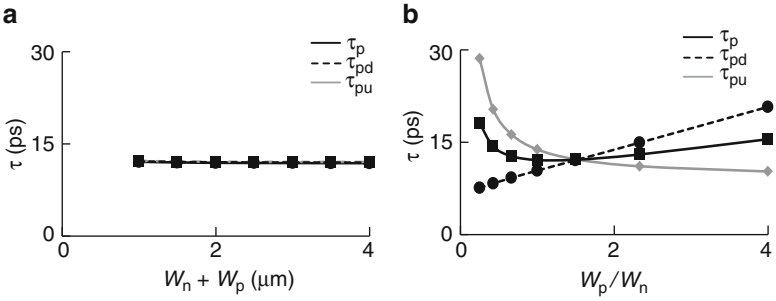


Fig. 2.43 τ_p, τ_{pd}, and τ_{pu} of inverter (FO = 4) with $\tau_{fi} = \tau_{ri} = 16$ ps: (**a**) vs. $(W_n + W_p)$ and (**b**) vs. W_p/W_n. 45 nm PTM HP models @ 1.0 V, 25 °C

In Fig. 2.37a, C_{in} is plotted as a function of τ_{fi} and τ_{ri} with C_L fixed at 6 fF. In Fig. 2.37b, C_{in} is plotted as a function of C_L with τ_{fi} and τ_{ri} fixed at 16 ps. From the plots we can see that C_{in} is nearly constant (~1.51 fF) over a wide range of τ_{fi} and τ_{ri}, and C_L. Since C_{in} is the total gate capacitance of the inverter MOSFETs, an approximate value of C_{in} can be obtained from the equation below:

$$C_{in} = C_{gn} \times W_n + C_{gp} \times W_p. \tag{2.21}$$

With $W_n = 0.4$ µm and $W_p = 0.6$ µm in this inverter design and using n-FET and p-FET C_g values listed in Table 2.9 for the inverter PD case (V_{gs} rising, V_{ds} falling), $C_{in} = (1.53 \times 0.4 + 1.51 \times 0.6) = 1.52$ fF, close to the value of 1.51 fF in Fig. 2.37a. Further simplification is made by assuming $C_{gp} = C_{gn}$ as a first approximation, and C_{in} expressed as

$$C_{in} = c_{in}\left(W_n + W_p\right), \tag{2.22}$$

where c_{in} is the normalized input capacitance per unit width. A normalized output capacitance per unit width c_{out} is defined in a similar fashion.

Knowing C_{in}, C_L is calibrated in units of fanout (FO), where

$$FO = \frac{C_L}{C_{in}}. \tag{2.23}$$

Approximating C_{in} as 1.5 fF, a load capacitance $C_L = 6$ fF is equivalent to FO = 4. In CMOS technology evaluation and in circuit design an inverter FO = 3 or FO = 4 is used as a reference. Following this general practice, in many of the examples C_L is set at 6 fF for a standard inverter (FO = 4) design.

In, Fig. 2.38a output parameters τ_{fo} and τ_{ro} are plotted as a function of τ_{ri} and τ_{fi} with C_L fixed at 6 fF. In Fig. 2.38b, τ_{fo} and τ_{ro} are plotted as a function of C_L with τ_{ri} and τ_{fi} fixed at 16 ps. The range of C_L covers FO = 0 (unloaded) to FO = 12. Simulated data points are fit to linear equations to estimate the output parameters from known input parameters.

Linear superposition of these equations allows approximate estimations of τ_{fo} and τ_{ro} at any combination of τ_{fi}, τ_{ri}, and C_L. As an example, with $\tau_{ri} = 30$ ps and $C_L = 12$ fF, a τ_{fo} estimation of 32.5 ps $\{=0.25 \times 30 + 13.6 + 1.9 \times (12 - 6)\}$ matches the simulated value of 32 ps within the accuracy required in circuit designs. Typically, for noise and delay considerations, maximum allowed values of τ_{fo} and τ_{ro} on a CMOS chip are limited to one-third of the clock cycle time. For a clock cycle time of 250 ps (frequency = 4 GHz), τ_{fo} and τ_{ro} are limited to 83 ps. Hence, with $\tau_{ri} = 83$ ps, this standard inverter can drive a load of ~33 fF (FO = 21) for $\tau_{fo} = 83$ ps. The simulated value of τ_{fo} is 77 ps, smaller than estimated because of small deviations from linearity.

Next, the inverter *RC* parameters are related to inverter design dimensions. All the simulations are done with $\tau_{fi} = \tau_{ri} = 16$ ps and $C_L = 6$ fF (FO = 4) except when these input parameters are the variables being studied.

In Fig. 2.39a, C_{in} is plotted as a function of $(W_n + W_p)$ while maintaining $W_p/W_n = 1.5$ and still using a one finger inverter design, ignoring any width bias in the model. C_{in} increases with $(W_n + W_p)$ as expected from Eq. 2.22. A linear fit of the simulated data gives $C_{in} \sim 1.53 \times (W_n + W_p)$ fF, and $c_{in} = 1.53$ fF/μm.

In Fig. 2.39b, C_{in} is plotted as a function of W_p/W_n while keeping $(W_n + W_p) =$ 1.0 μm. There is a small decrease in C_{in} as W_p/W_n increases. From Table 2.9, C_{gn} (=1.53 fF/μm) is slightly larger than C_{gp} (=1.51 fF/μm). Hence, from Eq. 2.21, C_{in} is expected to decrease with increase in W_p/W_n.

The inverter output capacitance C_{out} is determined from the difference between the charging currents through voltage sources VF (output falling) or VR (output rising), and VO. Its dependence on source and drain diffusion region areas is seen by plotting C_{out} as a function of $(W_n + W_p)$ and L_{ds} in Fig. 2.40. Linear fits of the data provide expressions to compute C_{out} for any value of $(W_n + W_p)$ or L_{ds}. As in the case of C_{in}, for a fixed L_{ds}, C_{out} is also proportional to $(W_n + W_p)$.

In Fig. 2.41a, PD and PU delays τ_{pd} and τ_{pu} are plotted as a function of C_L. The delays increase linearly with C_L at the rate of 1.02 ps/fF which corresponds to ~1.54 ps/FO for $\tau_{ri} = \tau_{fi} = 16$ ps. The slope of this plot in delay/FO defines the logical effort of a gate. The concept of logical effort for determining gate delays has

been described in engineering textbooks [8, 11]. Here we have extended the delay model to explore the dependencies of gate delays on MOSFET properties and input signal waveform shapes.

In Fig. 2.41b, τ_{pd} is plotted as a function of input rise time τ_{ri}. The delay increases with increase in τ_{ri}, and the behavior can be approximated as a linear dependence. From the linear fit over the range simulated, the estimated inverter delay τ_p (FO = 4, C_L = 6 fF), with $\tau_{ri} = \tau_{fi} = 16$ ps, is 11.88 ps. The results obtained from simulations for these precise input values are $\tau_p = 12.14$ ps and $\tau_{fo} = 17.4$ ps.

Effective switching resistances R_{swn} and R_{swp} are calculated from the measured delays and capacitances using Eqs. 2.17 and 2.18. R_{swn} and R_{swp} decrease with increase in the corresponding current drives and in turn MOSFET widths. In Fig. 2.42a, $1/R_{swn}$ is plotted as a function of W_n, and in Fig. 2.42b $1/R_{swp}$ is plotted as a function of W_p. The dependence of $1/R_{swn}$ and $1/R_{swp}$ on corresponding MOSFET widths can also be described by linear equations displayed on the plots. Thus $R_{swn} \times W_n$ and $R_{swp} \times W_p$ are nearly constant for practical MOSFET finger widths.

Rewriting Eq. 2.19 and expressing C_L in terms of FO gives

$$\tau_p = \frac{1}{2}\left(R_{swn} + R_{swp}\right)(\text{FO} \times C_{in} + C_{out}). \tag{2.24}$$

It is convenient to introduce normalized parameters r_{swn} and r_{swp},

$$r_{swn} = R_{swn} \times W_n, \tag{2.25}$$

$$r_{swp} = R_{swp} \times W_p. \tag{2.26}$$

When MOSFET widths are selected to give $\tau_{pd} \approx \tau_{pu}$, $R_{swn} = R_{swp}$, and

$$\tau_p \approx r_{sw} \times (\text{FO} \times c_{in} + c_{out}), \tag{2.27}$$

where r_{sw} $\{=R_{sw} \times (W_n + W_p)\}$ in Ω-μm is the average switching resistance normalized to total width. Following Eqs. 2.22 and 2.27, the average delay τ_p is nearly independent of $(W_n + W_p)$.

In Fig. 2.43a, τ_{pd}, τ_{pu}, and τ_p are plotted as functions of $(W_n + W_p)$, with $W_p/W_n = 1.5$, and the load C_L adjusted for FO = 4 (=4 × $c_{in}[W_n + W_p]$). The delays are nearly independent of $(W_n + W_p)$ as predicted by Eq. 2.27. In Fig. 2.43b, τ_{pd}, τ_{pu}, and τ_p are plotted as a function of W_p/W_n, with $(W_n + W_p) = 1.0$ μm. In this case the r_{sw} value in Eq. 2.27 varies with W_n and W_p excursions from that needed to maintain $\tau_{pd} \approx \tau_{pu}$. The PD delay increases with W_p/W_n (W_n decreasing, R_{swn} increasing), and PU delay increases with decreasing W_p/W_n (W_p decreasing, R_{swp} increasing). The average delay τ_p, increases as W_p/W_n is allowed to move away from its ideal value of 1.5. The variation of τ_p with W_p/W_n around this value is fairly small.

The R_{sw} of a logic gate is the average of its PD and PU resistances during switching and inversely proportional to its average current drive capability. In a PD transition R_{swn} is related to the I_{on} or I_{eff} of the n-FET and in a PU transition

R_{swp} is related to the I_{on} or I_{eff} of the p-FET. In Fig. 2.44, $1/R_{swn}$ vs. I_{effn} and $1/R_{swp}$ vs. I_{effp} for an inverter (FO = 4) are plotted. Variations in R_{swn} and R_{swp} are obtained by varying W_p/W_n at constant $(W_n + W_p)$. I_{effn} and I_{effp} for the corresponding widths are obtained from I_{ds}–V_{ds} characterization of MOSFETs as described in Sect. 2.2.2.

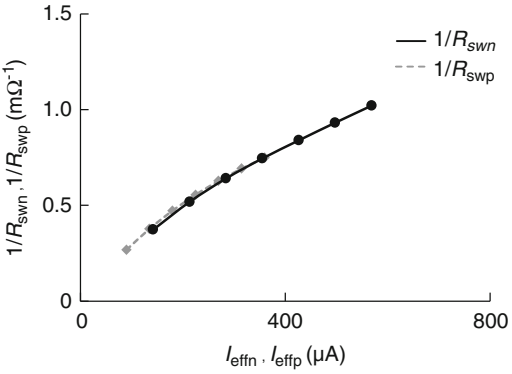


Fig. 2.44 Inverter (FO = 4) $1/R_{swn}$ vs. I_{effn}, and $1/R_{swp}$ vs. I_{effp} obtained by varying W_p/W_n with $(W_n + W_p) = 1.0$ μm. 45 nm PTM HP models @ 1.0 V, 25 °C

It is instructive to track I_{ds} of MOSFETs during a switching transition. Let us consider an n-FET in an inverter undergoing a PD transition. Initially, the inverter input voltage or V_{gs} of the n-FET is "0", and the inverter output voltage or V_{ds} of the n-FET is "1". The n-FET is in the off-state, with $I_{ds} = I_{off}$. As the input voltage transitions from "0" to "1", n-FET V_{gs} and I_{ds} increase while V_{ds} is decreasing. As the n-FET passes through the saturation region, its I_{ds} begins to decrease with V_{ds} ultimately going to zero as the transition is completed. The instantaneous resistance of the n-FET, at time t, $V_{ds}(t)/I_{ds}(t)$, averaged over the switching time is equivalent to the R_{swn} of the inverter.

By overlaying the $I_{ds}(t)$–$V_{ds}(t)$ trajectory on the DC I_{ds}–V_{ds} characteristics of a MOSFET, $I_{ds}(t)$ values can be compared to the MOSFET parameters listed in Table 2.7. Such plots for an n-FET and a p-FET during PD and PU transitions are shown in Fig. 2.45. I_{ds} values in the plots are normalized to the n-FET and p-FET widths for direct visual comparison. The $I_{ds}(t)$ trajectory passes near the I_{lo} and I_{hi} locations for $V_{ds} > V_{DD}/2$.

From Fig. 2.45 it is apparent that the average current during switching is better approximated as $I_{eff} = (I_{hi} + I_{lo})/2$ than as I_{on}. As I_{eff} is inversely proportional to R_{sw}, by selecting $W_p/W_n = I_{effn}/I_{effp}$, $R_{swp} \approx R_{swn}$, and $\tau_{pu} \approx \tau_{pd}$. As an example, the I_{eff} (n/p) ratio from Table 2.8 is 1.58, slightly larger than the W_p/W_n value of 1.50 in our standard inverter.

The increase in τ_{pd} with τ_{ri} is explained by comparing I_{ds} trajectories of the n-FET in the inverter for $\tau_{ri} = 16$ and 80 ps in Fig. 2.46a. As τ_{ri} increases, V_{gs} increases more gradually with time and the current drive of the n-FET is lowered. As a result both τ_{pd} and τ_{fo} increase. Similar plots for n-FET trajectories for

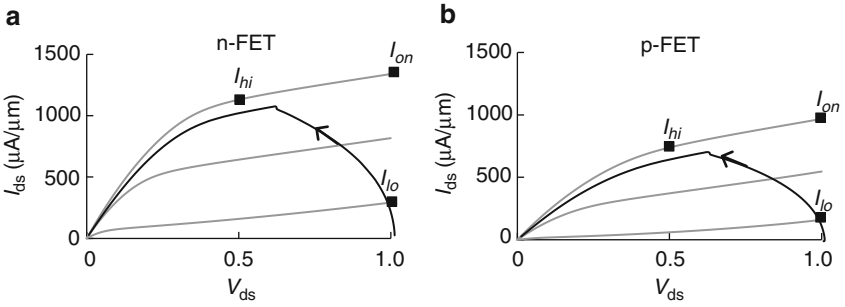


Fig. 2.45 Normalized I_{ds}–V_{ds} characteristics and inverter (FO = 4) trajectory during switching for (**a**) an n-FET, and (**b**) a p-FET. 45 nm PTM HP models @ 1.0 V, 25 °C

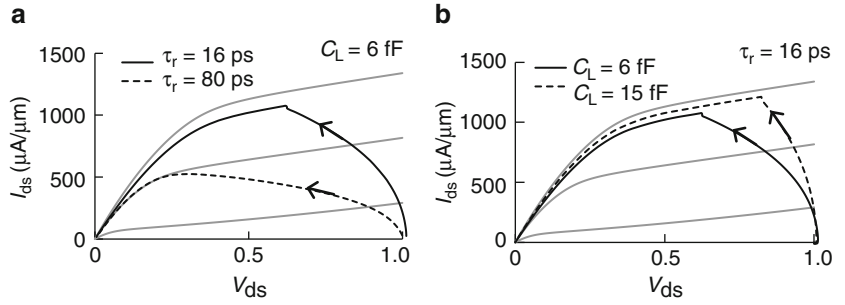


Fig. 2.46 (**a**) I_{ds}–V_{ds} characteristics and the inverter (FO = 4) trajectory during switching for an n-FET (**a**) with $\tau_{ri} = 16$ and 80 ps, and (**b**) with $C_L = 6$ and 15 fF. 45 nm PTM HP models @ 1.0 V, 25 °C

$C_L = 6$ fF (FO = 4) and 15 fF (FO = 10) with $I_{ds}(t)$ increasing with C_L are shown in Fig. 2.46b.

With the circuit simulation results described above we can see that the highly nonlinear behavior of MOSFETs is transformed into a set of circuit parameters that are either independent of other parameters or have a near linear dependence on them. These observations are stated below:

- C_{in} is independent of input τ_{ri}, τ_{fi} and output load C_L
- C_{in} is proportional to $(W_n + W_p)$
- C_{out} is proportional to $(W_n + W_p)$ and L_{ds}
- τ_p, τ_{pu}, and τ_{pd} vary linearly with C_L and FO
- τ_{pu} varies linearly with τ_{fi}, and τ_{pd} varies linearly with τ_{ri}
- τ_{pu} and τ_{pd} are nearly equal when $W_p/W_n = I_{effn}/I_{effp}$
- $1R_{swn}$ in a PD transition varies linearly with W_n and I_{effn}
- $1/R_{swp}$ in a PU transition varies linearly with W_p and I_{effp}

Delay parameters of our standard inverter, with $W_n = 0.4$ µm and $W_p = 0.6$ µm and driving a fixed load C_L of 6 fF (FO = 4), obtained from circuit simulations using 45 nm PTM HP models are shown in Table 2.13.

Table 2.13 Simulation results for the standard inverter book. 45 nm PTM HP models @ 1.0 V, 25 °C

Parameter	Nominal value	Variation with τ_{ri}	Variation with C_L	Variation with $(W_n + W_p)$, $W_p/W_n = 1.5$
c_{in}	1.50 fF/µm	None	None	Very weak
c_{out}	1.73 fF/µm	None	None	Very weak
r_{sw}	1,574 Ω µm	Weak	Weak	Weak
τ_{pd}	12.1 ps	~Linear	~Linear	Weak
τ_{pu}	12.2 ps	~Linear	~Linear	Weak
τ_{fo}	17.4 ps	~Linear	~Linear	Weak
τ_{ro}	18.2 ps	~Linear	~Linear	Weak

Nominal values are for standard inverter $W_n = 0.4$ µm and $W_p = 0.6$ µm with $C_L = 6$ fF (FO = 4) and $\tau_{ri} = \tau_{fi} = 16$ ps. Dependencies of delay parameters with τ_{ri}, C_L, and $(W_n + W_p)$ are indicated

The current drawn by the inverter in its quiescent state, IDDQ, is dependent on MOSFET leakage currents and the input node voltage. Measurements of standby and active power are covered in Sects. 2.2.4 and 2.2.5, and in more detail in Chap. 4.

2.2.4 Delay Chains

The signal propagation delay through a circuit can be increased to a few hundred picoseconds or more by connecting a number of logic gates in series. Delays in this range can be measured on silicon for model-to-hardware correlation with an accuracy of a few %. If all the logic gates in the chain are nominally identical, the average delay τ_p through a single gate is obtained by dividing the total delay across the chain by the number of gates.

The circuit schematic in Fig. 2.47 shows four inverters with different values of $(W_n + W_p)$ and W_p/W_n connected in series. An input signal with a rise time of τ_{ri1} at A1 is propagated through inverter X1 with a load capacitance equal to input capacitance C_{in2} of inverter X2. Output signal fall time τ_{fo1} and the propagation delay through X1, τ_{pd1}, are determined by the characteristics of inverter X1, τ_{ri1} and C_{in2}. The falling signal at A2 propagates through the second inverter X2. The output rise time τ_{ro3} and delay τ_{pu2} through X2 are determined by the characteristics of inverter X2, τ_{fi2} (=τ_{fo1}) and C_{in3}. The process of alternate stages going through PD and PU transitions continues until the signal reaches the end of the chain. The delay across the chain τ, is given by

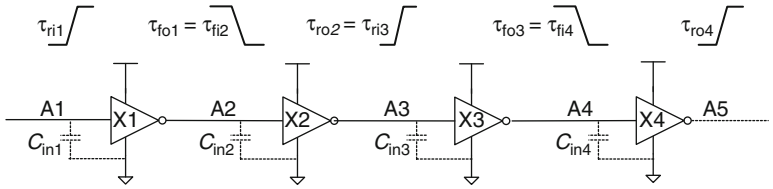


Fig. 2.47 Four inverters of different designs in a delay chain. Input and output waveforms at each node and input capacitances are indicated

$$\tau = \tau_{\mathrm{pd1}} + \tau_{\mathrm{pu2}} + \tau_{\mathrm{pd3}} + \tau_{\mathrm{pu4}} + \cdots. \tag{2.28}$$

The value of τ can be calculated from lookup tables similar to Table 2.11 for each inverter as explained in the previous section. Alternatively, a circuit simulation can be carried out treating the entire chain as a single unit. However, measurement of τ in the hardware does not provide sufficient information to characterize each inverter in the chain individually.

If all the inverters in a long delay chain are identical, the capacitive loads on each inverter are also identical. The input signal waveform shape and MOSFET widths of the inverter may be carefully adjusted such that the rise times and fall times at the inputs of alternating stages are nearly the same ($\tau_{\mathrm{ri1}} = \tau_{\mathrm{fi2}} = \tau_{\mathrm{ri3}} = \cdots$). In such an arrangement $\tau_{\mathrm{pd1}} = \tau_{\mathrm{pu2}} = \tau_{\mathrm{pd3}} = \cdots$. The average delay τ_{p} ($=\tau_{\mathrm{pd}} = \tau_{\mathrm{pu}}$) across one inverter is obtained by dividing τ by the number of inverter stages in the chain. This approach may be used for any logic gate with inverting or non-inverting characteristics.

The load on a logic gate stage (LG), in addition to the C_{in} of the subsequent stage in a chain, may be realized in several different ways. A fixed capacitive load C_{L} at the output of each stage can represent a desired FO following the relationship in Eq. 2.23. For a correct representation of a logic gate driving other logic gates, the fixed capacitor is replaced with additional logic gates operating as capacitive loads.

The circuit schematic corresponding to true logic gate FO is compacted with the use of a current multiplier circuit shown in Fig. 2.48. The charging current during switching through zero-voltage source VX, I(VX), is measured. The value of a current source connected between the output of the logic gate and GND terminal is set as $(\mathrm{XL} - 1) \times I(\mathrm{VX})$. This circuit behaves as if the total gate load being driven is XL times the capacitive load of the single logic gate at the output node. The parameter XL (=FO) is a user-defined variable.

The three schemes described above for adding a capacitive load at the output of a logic gate are shown in Fig. 2.49. For a pure C_{L} load in Fig. 2.49a, the value of C_{in} for an LG has to be determined for each gate type and each simulation condition (V_{DD}, temperature, L_{p}, V_{t}, etc.) to get equivalent FO. Schematics of one LG driving two additional identical LGs for FO = 3 is shown in Fig. 2.49b. In this representation, a schematic is created for a specific FO value and C_{L} of equivalent FO load added to the load LGs. The schematic with an XL load in Fig. 2.49c is compact and its FO value is set as a variable so that the same schematic may be used for simulating LGs with different FOs.

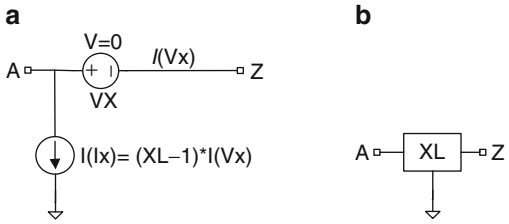


Fig. 2.48 A capacitance load multiplier (**a**) circuit schematic and (**b**) symbol

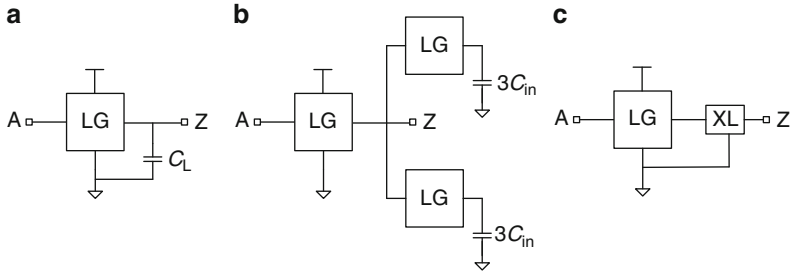


Fig. 2.49 Circuit schematics of logic gates (LGs) driving an additional capacitive load: (**a**) fixed capacitive load C_L, (**b**) logic gate loads for FO = 3 and (**c**) variable XL load

The schematic of a delay chain may be further compacted by creating a hierarchy. The schematic of one stage in the delay chain comprising a logic gate LG, load current multiplier XL, and a fixed capacitive load C_L to represent interconnect capacitance is shown in Fig. 2.50a. This subcircuit (subckt) is represented by a symbol LGXL in Fig. 2.50b. The sources to provide the power supply voltage V_{DD} and the voltage signal pulse at node A are shown in Fig. 2.50c. A delay chain with ten identical LGXL stages is shown in Fig. 2.50d. This hierarchical schematic is less cluttered and provides the option of defining MOSFET parameters, XL and C_L in LTspice commands at the top level. The schematic may be used as a template in which different types of logic gates are instantiated.

If the logic gate LG has more than one input, the additional inputs are tied to a "1" or "0" voltage level such that the output of the LG has the desired level (inverting or non-inverting). The power supply nodes in each stage are vdd and vss and may be connected to V_{DD} and GND respectively or biased at different voltage levels.

The circuit schematic for stage 7 in Fig. 2.50d is expanded to show the wiring of this stage. A zero-voltage source V7 is placed at the output of the LG in stage 7 to measure the charging and discharging current during a transition for computing the load capacitance. Another zero-voltage source V7S is placed in the GND connection of this stage to measure the total discharging current during a PD transition. The charging current during a PU transition can be measured by inserting a third zero-voltage source (not shown) in the vdd connection to stage 7.

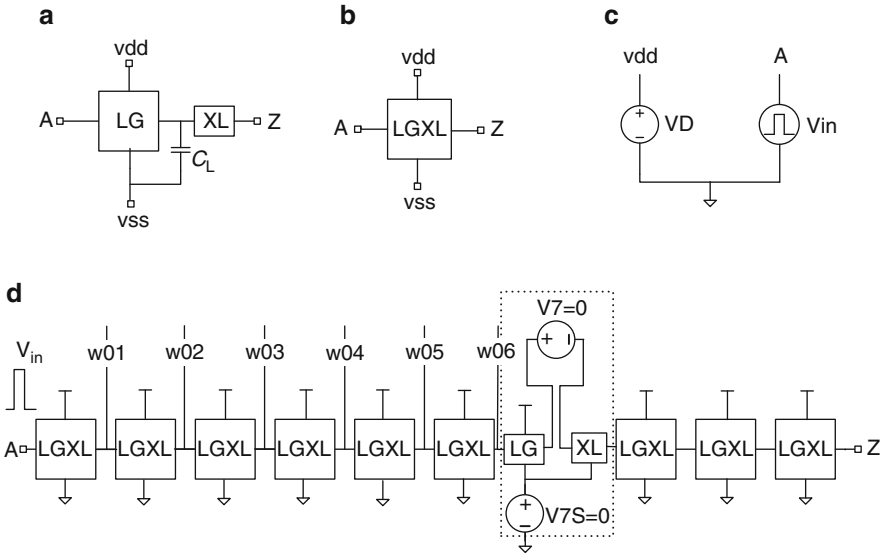


Fig. 2.50 (**a**) Circuit schematic of a logic gate with load multiplier XL and interconnect capacitance C_L, (**b**) symbol of the stage LGXL in (**a**), (**c**) DC power supply and signal source to enable signal propagation through the delay chain, and (**d**) a chain of 10 LGXL stages with zero-voltage sources added in stage 7

A single pulse is launched at the input A of the delay chain comprising identical inverters. The rise and fall times may be different as the signal traverses the delay chain. A typical scenario with standard inverter stages is shown in Fig. 2.51 with $\tau_{ri} = 80$ ps at input A. At the output of the first stage, node w01, $\tau_{fo1} = 36$ ps, at the output of the second stage, node w02, $\tau_{ro2} = 24$ ps, at the output of the third stage, node w03, $\tau_{fo3} = 21$ ps, and at w04, $\tau_{ro4} = 20$ ps. The signal waveform shape stabilizes after the first three stages and $\tau_r \approx \tau_f$. If there is no load on the last stage in the chain at node Z, $\tau_{roz} \approx \tau_{foz} \approx 8.5$ ps. This has a small impact on the delay of the previous stage in the chain.

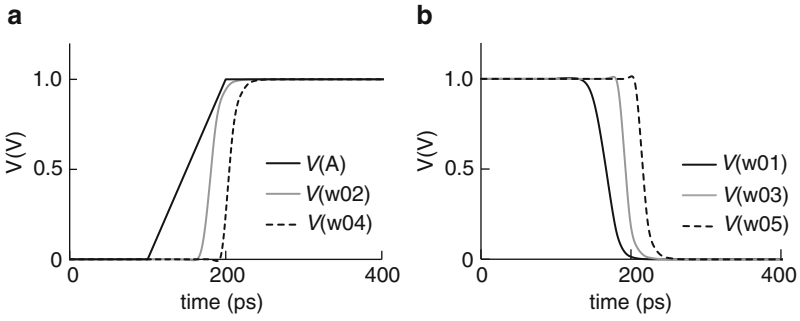


Fig. 2.51 Signal waveforms in an inverter FO = 4 delay chain at nodes shown in Fig. 2.50: (**a**) rising edges at A, w02 and w04 and (**b**) falling edges at w01, w03 and w05. 45 nm PTM HP models @ 1.0 V, 25 °C

In the middle of the chain, the output waveforms for all PD transitions are identical. Similarly, output waveforms of all PU transitions are identical. This holds for any value of W_p/W_n in the inverter, an inverter design with $\tau_{pd} = \tau_{pu}$ being a special case. The characteristics of a logic gate in the middle of a chain are equivalent to those of the single logic gate described in Sect. 2.2.3 for the special case of τ_{ri}, τ_{fi}, and C_L (=XL × C_{in}) being matched.

Circuit simulations are carried out to measure signal propagation delays through stages 5 and 6 to avoid any end-effects. The waveforms at nodes w04, w05, and w06 are shown in Fig. 2.52. The delay across a stage is measured at signal levels of $V_{DD}/2$ at the input and output of a stage. The signal rise and fall times are measured between signal levels of 10 and 90 % of V_{DD}. PD and PU delays are measured across stage 5 (nodes w04 and w05). The delay between nodes w04 and w06 gives

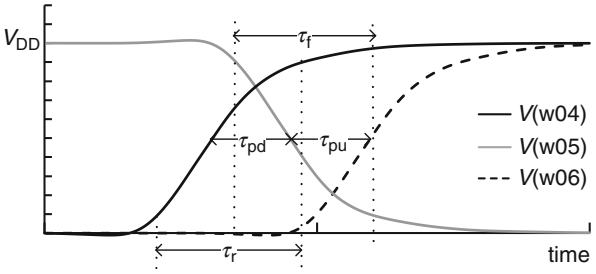


Fig. 2.52 Voltage waveforms at nodes w04, w05, and w06 and measured parameters τ_{pd}, τ_{pu}, τ_r, and τ_f. 45 nm PTM HP models @ 1.0 V, 25 °C

$$2\tau_p = (\tau_{pu} + \tau_{pd}). \tag{2.29}$$

Using Eqs. 2.20 and 2.24, average delay τ_p, with any FO (=XL) and additional capacitive load C_L is expressed as

$$\tau_p = R_{sw}(\text{FO} \times C_{in} + C_{out} + C_L). \tag{2.30}$$

The parameter FO × C_{in} is determined by integrating the charging current through zero-voltage source V7 in stage 7. The total switching capacitance, C_{sw} (= FO × C_{in} + C_{out} + C_L), is measured by integrating the discharging current for a PD transition through the zero-voltage source V7S. With $C_L = 0$, C_{in}, C_{out}, and R_{sw} can be determined for the PD transition. Similarly, by measuring the charging current through V_{DD} connection to stage 7, these parameters for a PU transition may be obtained. Interconnect wire load of length l is modeled as a capacitance ($C_L = C_w l$) for short wires and as an RC network shown in Fig. 2.13 for long wires (Sect. 10.3.2).

The leakage current in the quiescent state, IDDQ, is measured prior to the arrival of the first pulse edge. An average IDDQ per stage (IDDQ/stage) is obtained by dividing the measured IDDQ by the number of stages (=10 in this example). This is an average current of inverters with inputs a "0" and at "1". Subthreshold leakage

current contributions to IDDQ are from an n-FET when the inverter input is a "0" and a p-FET when its input is a "1". Gate-dielectric leakage contributions may become significant in some CMOS technologies and add to IDDQ (Sect. 4.2.1)

A netlist of the delay chain circuit shown in Fig. 2.50 can be generated from the schematic. LTspice commands for measuring τ_p, C_{in}, and IDDQ are listed below:

***delay chain voltage sources and measurements**

```
.Vin A 0 PULSE(0 =pvdd 100e-12 20e-12 20e-12 400e-12 1000e-12 1)
.param XL =4
.measure tran tpd+trig v(w04) val=0.5*pvdd rise=1+ targ v(w05) val=0.5*pvdd fall=1
.measure tran tpu+trig v(w04) val=0.5*pvdd fall=1+ targ v(w05) val=0.5*pvdd rise=1
.measure tran tprr+trig v(w04) val=0.5*pvdd rise=1+targ v(w06) val=0.5*pvdd rise=1
.measure tran i_charge_fall integ I(v7) from=100e-12 to=300e-12
.measure tran i_charge_rise integ i(V7) from=600e-12 to 800e-12
.measure cin param ='(i_charge_rise - i_charge_fall)/(2*pvdd*xl)*1e15'
.measure tp param='tprr/2*1e12'
.measure tran IDDQ avg I(VD)*-1 from=10e-12 to=50E-12
.measure IDDQ_stage  param=' IDDQ/10 '
```

Circuit simulations are carried out for the standard inverter with $W_n = 0.4$ µm, $W_p = 0.6$ µm, $L_p = 0.045$ µm, $L_{ds} = 0.12$ µm at nominal V_t values using 45 nm PTM HP models. The MOSFET widths are selected to give equal PD and PU delays at $V_{DD} = 1.0$ V. With $C_L = 0$, parameter XL is used to define FO.

The simulated delay per stage τ_p and the load capacitance as a function of FO are shown in Fig. 2.53. Note that the load increases linearly with FO (=XL) and the increase in delay with FO is 2.13 ps/FO, a value higher than 1.54 ps/FO for the inverter book obtained with $\tau_{ri} = \tau_{fi}$ held constant at 16 ps The load capacitance measured using the zero-voltage source V7, also varies linearly with FO, giving 1.50 fF/FO. Its value is consistent with the simulation results on a single inverter in Sect. 2.2.3. This calibration can be used to compute propagation delay and C_{in} of an inverter for any FO. The value of C_{out} is independent of FO.

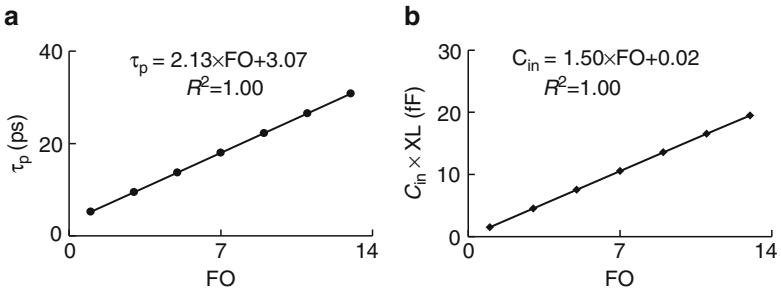


Fig. 2.53 Simulated values of a standard inverter as a function of FO (=XL): (**a**) τ_p and (**b**) load capacitance (=C_{in} × XL). 45 nm PTM HP models @ 1.0 V, 25 °C

From the known value of C_{out} (=1.83 fF for our standard inverter), R_{sw} is determined using Eq. 2.30. R_{sw} is plotted vs. FO in Fig. 2.54a. As FO increases, R_{sw} becomes nearly constant. The normalized I_{ds}–V_{ds} trajectories of the n-FET during a PD transition in the inverter with FO = 1 and FO = 13, superimposed on n-FET I_{ds}–V_{ds} DC characteristics are shown in Fig. 2.54b. These trajectories are slightly different than those shown in Fig. 2.46b because of differences in input and output waveforms for the transitions and the C–V characteristics of the load capacitance.

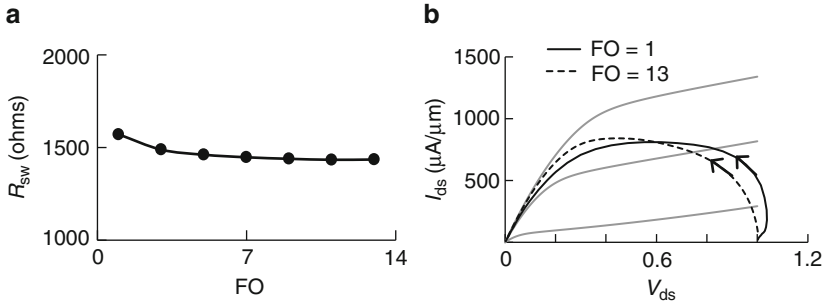


Fig. 2.54 (**a**) R_{sw} vs. FO, and (**b**) inverter n-FET I_{ds}–V_{ds} trajectory during switching. 45 nm HP models @ 1.0 V, 25 °C

Variations in MOSFET parameters impact signal waveforms, propagation delay, and power during switching and in the quiescent state. The two parameters with the largest impact are channel length L_p and threshold voltage V_t. Other parameters such as overlap capacitances of the MOSFETs, gate-dielectric thickness, diffusion area capacitance, and source-drain series resistance also introduce variability in circuit delays.

In Fig. 2.55a, average inverter (FO = 4) delay τ_p is plotted as a function of L_p for the $\pm 3\sigma L_p$ range of 0.0405 to 0.0495 μm. In Fig. 2.55b, τ_p is plotted as a function of the shift in V_t from the nominal $\Delta|V_t|$ of both n-FET and p-FET over the $\pm 3\sigma V_t$

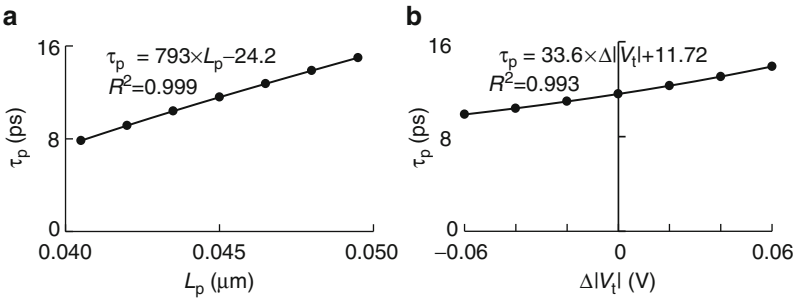


Fig. 2.55 Standard inverter FO = 4 average delay τ_p vs. (**a**) L_p and (**b**) $\Delta|V_t|$. 45 nm PTM HP models @ 1.0 V, 25 °C

range. Data are fit to get a linear relationship between τ_p and L_p and τ_p and $\Delta|V_t|$. The equations displayed in Fig. 2.55a, b are used for hand calculations of τ_p for any value of L_p or $\Delta|V_t|$ under the simulation conditions of $V_{DD} = 1.0$ V, 25 °C.

The τ_p value increases from 7.8 ps for minimum L_p to 15.0 ps for maximum L_p, a 1.92× increase in delay over the full range L_p. This is comparable to 1.88× increase in $(W_n \times I_{effn} + W_p \times I_{effp})$ over the same L_p range. Similarly, τ_p increases from 9.8 to 13.9 ps, a 1.41× increase over the full $\Delta|V_t|$ range compared with 1.49× increase in $(W_n \times I_{effn} + W_p \times I_{effp})$. Hence, the spread in $(W_n \times I_{effn} + W_p \times I_{effp})$ gives a rough estimate of the spread in τ_p.

The variations in C_{in} and IDDQ/stage as a function of L_p are shown in Fig. 2.56a, b respectively. Over the $\pm 3\sigma$ L_p range, IDDQ varies by >40× whereas C_{in} remains within ±4 %. Change in IDDQ affects the off-state leakage power, an important parameter in low-power electronic applications and in high-performance applications where the off-state power is a significant fraction of total power.

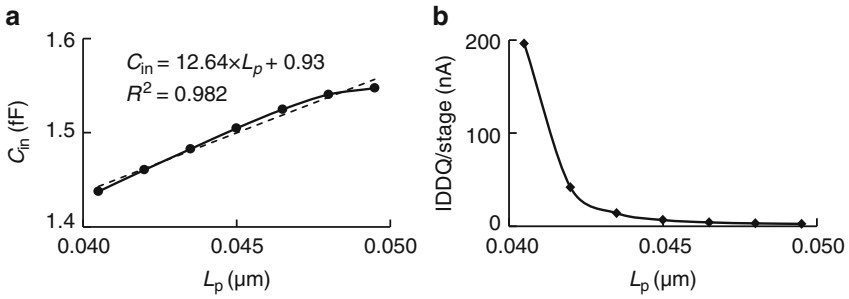


Fig. 2.56 Standard inverter (FO = 4): (**a**) C_{in} vs. L_p and (**b**) IDDQ/stage vs. L_p. 45 nm PTM HP models @ 1.0 V, 25 °C

2.2.5 Ring Oscillators

Although a delay chain is very useful for obtaining model-based logic gate delays, it is a nontrivial task to measure the signal propagation delays in a chain embedded in a CMOS chip with a high degree of accuracy. As described in Chap. 5, delay chain measurements require clock signals, latches, and control circuitry to capture voltage signals and facilitate relative timing measurements.

An alternate method to measure logic gate delays in silicon is to use ring oscillators. A ring oscillator (RO) is formed by connecting the output of a delay chain to its input, thus forming a closed loop. The signal edge launched at any node of the RO travels around the loop continuously. As a result the voltage at any node of the ring oscillates with a period proportional to the signal propagation delay through the circuit stages.

The circuit schematic of a ring oscillator with 51 inverting stages is shown in Fig. 2.57a. It comprises five instances of the 10-stage delay chain block, LGXL10, shown in Fig. 2.50d. A 2-input NAND gate is included in the loop to enable and disable the oscillations with the aid of an external input signal at node EBL.

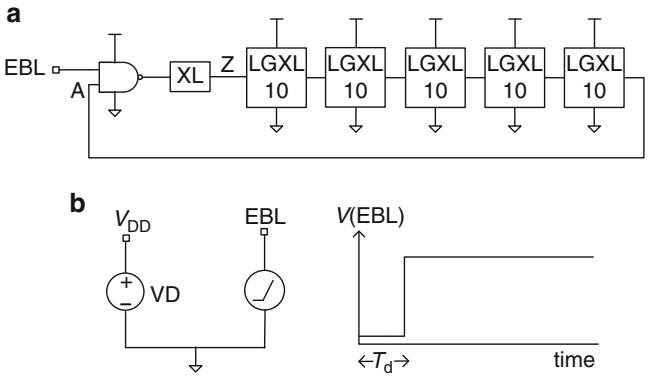


Fig. 2.57 (**a**) Circuit schematic of a ring oscillator with 51 inverting stages: each LGXL10 is a chain of 10 inverting logic gate stages with FO = XL, (**b**) voltage sources to power and enable oscillations

The signal input at node EBL is shown in Fig. 2.57b. At time $<T_d$, the input level at node EBL is a "0" and the output node of the NAND2, Z is a "1" as is node A. If the gates in the chain follow inverting logic, the voltage level at the output of each gate alternates between "0" and "1". At time $=T_d$, the voltage at EBL rises to a "1" and node Z switches to a "0". This signal edge falling from a "1" to a "0" at Z travels through the 50 stages and arrives at node A. With the EBL input of the NAND2 at "1", node Z switches to a "1". Again the second signal edge rising from a "0" to a '1" at Z travels through the loop and node Z falls back to "0" again. This cycle is repeated as long as EBL is held at "1". The voltage at any node oscillates as shown in Fig. 2.58a. For sustainable oscillations, the number of inverting stages must be odd.

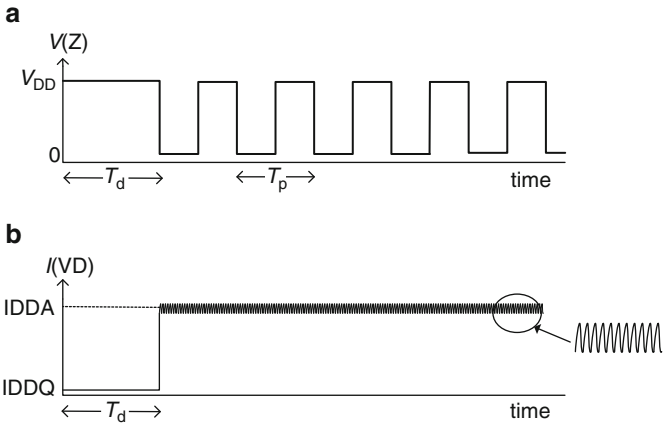


Fig. 2.58 Ring oscillator (**a**) node voltage oscillation, and (**b**) current drawn by the power supply VD in quiescent and active modes. The IDDA ripple frequency corresponds to the stage delay

If the RO comprises $(2\alpha+1)$ inverting stages, the period of oscillation T_p is given by

$$T_p = 2\tau_p(2\alpha+1) = \frac{1}{f}, \tag{2.31}$$

where τ_p is the average of PU and PD delays of a stage, f is the frequency of oscillation and α is an integer. Measurements of f in the MHz range for an RO embedded in a CMOS chip can be easily made with an off-the-shelf frequency counter. Typically RO frequencies for $\alpha < 50$ are in the GHz range. A frequency divider circuit at the output of the RO is used to lower the frequency for a direct measurement. Alternatively, a counter circuit using the on-chip clock as a reference is employed for measuring RO frequencies, and the data streamed out for analysis.

The time variation of current drawn by the RO power supply VD, is shown in Fig. 2.58b. At time $<T_d$, with the RO in its quiescent state, this current is the sum of the leakage currents of all the stages in the RO. Since the input nodes of successive stages are alternating between "0" and "1", the leakage current contributions of successive stages alternate in value. The measured quiescent current IDDQ may be divided by $(2\alpha+1)$ to obtain the average IDDQ/stage. In the oscillating state, with EBL held at "1", current is drawn by each stage as it switches. The magnitude of the current varies on the time scale of the switching transitions (~10 to 100 ps). The current measured externally with a longer integration time is essentially constant. Measurement of currents drawn by an RO in the oscillating state, IDDA, and in the quiescent state, IDDQ, along with frequency of oscillation are used to determine the power drawn during switching and the switching capacitances and resistances of the stages in the RO.

As the signal edge traverses the loop, there is only one stage switching at any instant and all other stages are in the quiescent state. The total current drawn is the sum of switching current I_{sw} of one stage and the IDDQ of the remaining stages. For an RO with $(2\alpha+1)$ stages

$$\text{IDDA} = I_{sw} + \frac{\text{IDDQ}}{\text{stage}} \times 2\alpha. \tag{2.32}$$

The power drawn in the quiescent state is given by

$$P_{off} = \frac{\text{IDDQ}}{\text{stage}} \times (2\alpha+1) \times V_{DD}. \tag{2.33}$$

The AC component of power (ignoring short-circuit power discussed in Section 4.3.2) is

$$P_{sw} = I_{sw} \times V_{DD} = \left(\text{IDDA} - \frac{\text{IDDQ}}{\text{stage}} \times 2\alpha\right) \times V_{DD}. \tag{2.34}$$

From Eq. 2.34 it is apparent that if IDDQ $\ll I_{sw}$, P_{sw} is essentially independent of the number of stages whereas P_{off} increases with the number of stages. Power is an important consideration when designing stand-alone ROs for model-to-hardware correlation (Sect. 5.2.3) and for technology performance evaluation (Sect. 10.4).

The power consumed during switching can also be expressed in terms of the switching capacitance, C_{sw}

$$P_{sw} = \frac{1}{2} C_{sw} V_{DD}^2 \{2(2\alpha + 1)\} f = \frac{1}{2\tau_p} C_{sw} V_{DD}^2, \tag{2.35}$$

where C_{sw} is the switching capacitance in each transition and there are $2(2\alpha + 1)f$ transitions per second.

For $\alpha \gg 1$, the factor $2\alpha/(2\alpha + 1) \approx 1$, and using Eqs. 2.31, 2.34, and 2.35, C_{sw} is expressed in terms of measured IDDA and IDDQ as

$$C_{sw} = \frac{(\text{IDDA} - \text{IDDQ})}{\{2 \times (2\alpha + 1) V_{DD} f\}} = \frac{(\text{IDDA} - \text{IDDQ})}{V_{DD}} \times 2\tau_p. \tag{2.36}$$

The average switching resistance R_{sw} can thus be estimated from the measured parameters IDDA, IDDQ, and τ_p as

$$R_{sw} = \frac{V_{DD}}{2 \times (\text{IDDA} - \text{IDDQ})}. \tag{2.37}$$

The RO circuit parameters and their relationships to electrical measurements in silicon hardware are summarized in Table 2.14.

Table 2.14 RO circuit parameter names, units, and description

Parameter	Unit	Expression	Description
T_p	ps		Period of oscillation
IDDQ	mA	Avg I(VD) at time $< T_d$	Current in quiescent state
IDDA	mA	Avg I(VD) at time $> T_d$	Current in active state
f	MHz	$1/T_p$	Frequency of oscillation
τ_p	ps	$T_p/\{2(2\alpha + 1)\}$	Average delay/stage
P_{off}	mW	IDDQ $\times V_{DD}$	Off-state power
P	mW	IDDA $\times V_{DD}$	Active power
P_{sw}	mW	(IDDA – IDDQ) $\times V_{DD}$	Switching power
C_{sw}	fF	$2\tau_p \times$ (IDDA – IDDQ)/V_{DD}	Switching capacitance/stage
R_{sw}	Ω	$V_{DD}/\{2 \times$ (IDDA – IDDQ)$\}$	Switching resistance of a stage

LTSpice commands for measuring period T_p and signal rise and fall times at any node are listed below. The period of oscillation of the RO is measured in the tenth cycle to allow sufficient time for the oscillations to stabilize. In PD-SOI technology, MOSFET body voltages change with time during switching, and sufficient time must be allowed for the body potentials to equilibrate. Techniques for achieving steady-state conditions in SOI are described elsewhere [17].

***commands for RO measurements**

*.measure tran tp + trig v(Z) val=0.5*pvdd rise=10+ targ v(Z) val=0.5 rise=11*
*.measure tran tr + trig v(Z)=0.1*pvdd rise=10+ targ v(Z)=0.9*pvdd rise=10*
*.measure tran tf + trig v(Z)=0.1*pvdd fall=10+ targ v(Z)=0.9*pvdd fall=10*

In estimating τ_p, C_{sw} and IDDQ per stage, it is assumed that the insertion of a NAND2, or another scheme to enable the oscillations, is equivalent to adding one more identical stage in the loop. Generally, the delay through the NAND2 is different than the delay of a standard stage, and this introduces an error in computing average delay parameters per stage.

There are four ways to reduce the error in computing average delay parameters per stage:

1. Increase the number of stages in the RO to reduce the effect of the NAND2 imbalance
2. Match the NAND2 delay with the LGXL block delay by adjusting XL at the output of the NAND2
3. Replace the NAND2 with a different enable scheme that matches LGXL delay
4. Use circuit simulations to estimate the error correction for the NAND2 for each RO design

Increasing the number of stages in the RO has the advantage that the same circuit schematic template may be used for different LGXL topologies and simulation conditions. However, the simulation time increases with the number of stages.

Measurement of IDDQ and IDDA in silicon requires ROs with independent power supplies. This can be done for RO test structures placed in the scribe-line. In the case of ROs embedded in a CMOS chip and tied to the chip power grid, differential measurement techniques are used to characterize the delay parameters C_{sw} and R_{sw} per stage [16]. These and other considerations in the design and measurement of RO-based test structures for model-to-hardware correlation and for monitoring CMOS technology are covered in Chaps. 5 and 10.

2.2.6 Comparison of Logic Gate Characterization Methods

Each of the three circuit templates for characterization of logic circuit blocks as described in Sect. 2.2.3 (single block), Sect. 2.2.4 (delay chain), and Sect. 2.2.5 (ring oscillator) have unique features. A single circuit block has complete flexibility

in its input and output settings in circuit simulations, but the characterization results cannot be directly validated in silicon. A delay chain is useful for both simulations and measurements in silicon. Ring oscillator characteristics are relatively easy to measure in silicon compared with delay chains. Circuit simulations for ROs, however, take considerably longer than for open-ended delay chains. The simulator may take a long time to find a DC operating point for a closed loop, and timing tools generally cannot handle this circuit configuration.

The simulation results obtained from these three approaches may differ slightly. A comparison of simulation results for a single inverter book, an inverter delay chain, and a ring oscillator comprising 50 inverter stages and a NAND2 is shown in Table 2.15. The inverter schematic and design dimensions used for all three methods are identical. The NAND2 gate in the ring oscillator is counted as equivalent to one inverter stage, thereby resulting in small differences in τ_p from the delay chain and RO. Related differences are seen in derived parameters C_{sw} and R_{sw}.

Table 2.15 Simulated parameters for inverter FO = 4, $W_p = 0.6$ µm, $W_n = 0.4$ µm. 45 nm PTM HP models @ 1.0 V, 25 °C

Parameter	Single gate[a]	Delay chain	Ring oscillator
τ_p (ps)	12.13	11.59	11.68
C_{in} (fF)	1.50	1.50	1.51
IDDQ (nA) (FO = 1)	5.95[b]	6.15	6.18
C_{sw} (fF)	7.70	7.87	7.88
R_{sw} (Ω)	1,574	1,471	1,481
$\tau_i \approx \tau_f$ (ps)	~16 to 18	~20	~20

[a] $C_L = 6$ fF (FO = 4 equivalent)
[b] Average IDDQ with inputs at "1" and at "0"

In a single inverter book, the input signal waveform is user defined as a linear ramp (pulse function in LTspice) whereas the input waveform shapes in the delay chain and RO are as shown in Fig. 2.52. Also, the inverter book has a fixed load capacitance whereas in the delay chain and RO the load capacitance is a function of time-varying voltage on the gate terminal during switching (MOSFET gate load). IDDQ/stage values for the delay chain and RO include gate-dielectric leakage current whereas the single inverter does not.

The simulator is able to provide outputs with a high degree of accuracy. However, the method used for simulating circuit characteristics must be clearly specified when comparing different BSIM models. Measurements made on silicon with different test structure designs and test environments add another level of complexity to accurate model-to-hardware correlation.

2.2.7 Monte Carlo Analysis

In Figs. 2.55 and 2.56 the variations of circuit parameters with L_p or $\Delta|V_t|$ are demonstrated by assigning identical values of L_p or $\Delta|V_t|$ to all the n-FETs and p-FETs in the inverter chain. In silicon processing, across chip, across wafer and lot-to-lot variations described in Chap. 6 cause L_p and V_t of n-FETs and p-FETs to vary independently over their $\pm 3\sigma$ ranges. In addition, random dopant fluctuations result in V_t variations in identical MOSFETs on the chip even when placed in close proximity. This random variation in V_t is given by σV_{tr} where

$$\sigma V_{tr} = \frac{A_{vt}}{\sqrt{WL_p}}. \tag{2.38}$$

Here A_{vt} is a technology-dependent constant that is assigned a value of 0.004 V-μm when using 45 nm PTM HP models. The value of σV_{tr} increases as the MOSFET area is reduced.

Circuit simulations for taking into account, systematic process variations and random V_t variations are carried out with Monte Carlo analysis. The simulations are run repeatedly, and in each run randomly picked values are assigned to the variables. The results contain all possible outcomes over the variable ranges, similar to the data for silicon hardware manufactured over a period of time.

Monte Carlo simulations are run by assigning a nominal value to a parameter of interest and its range of variation. If the parameter is normally distributed (Gaussian distribution) then a mean (nominal) and a σ value are assigned. The probability that a specific parameter value is picked follows the probability in the normal distribution. The probability of a variable having a value within $\pm 1\sigma$ is 68.3 % whereas the probability of having a value beyond $\pm 3\sigma$ is 0.3 % (Sect. 9.1.2).

LTspice commands for running Monte Carlo simulations for n-FET V_t adder variable *ndelvto* are listed below.

***commands for Monte Carlo simulations with variables ΔV_{tn} (ndelvto)**

```
.param ndelvto={normal(ndelvtonom, ndelvtosigma)}
.function normal(nom, sigma) if (run==1, nom, nom+(gauss(sigma)))
.param ndelvtonom=0
.param ndelvtosigma=0.02
.step param run 1 500 1                    * number of cases=500
```

Here *ndelvto* is a Gaussian (normal) function with a mean of *ndelvtonom* and a standard deviation of *ndelvtosigma*. A value of 0.0 V is assigned to *ndelvtonom* and *ndelvtosigma* = 0.02 V. The 'run' parameter specifies the number of cases to simulate. In the first simulation run, *ndelvto* is assigned the nominal value of 0.0 V. The results of the first run may be compared with a single standard nominal run to validate the simulation setup. In subsequent runs, the value of *ndelvtosigma* is randomly picked from its normal distribution with $\sigma = 0.02$ V.

In LTspice the assignment of variables in a Monte Carlo simulation is dependent on the circuit topology and its hierarchy. Each instance in the schematic is treated as one unit. Constituent circuit components of a unit with the same variable names are assigned identical parameter values. Parameter values assigned to components within each unit are independent of the values assigned within any other unit. The outcome of this feature is illustrated with two different hierarchical arrangements in Fig. 2.59.

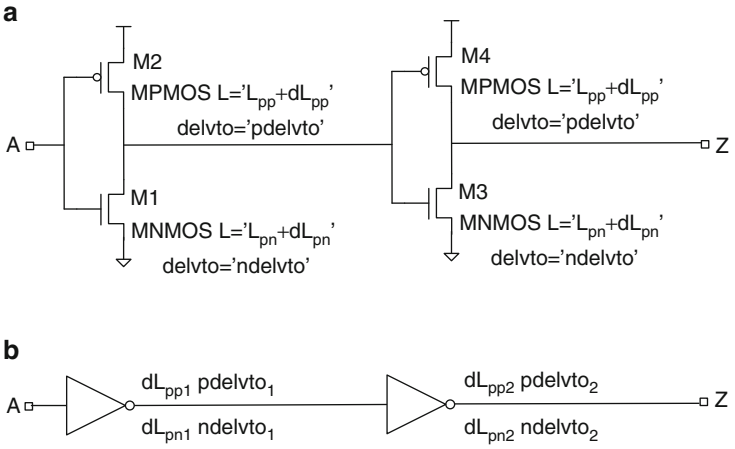


Fig. 2.59 Circuit schematics with different hierarchical arrangements for running Monte Carlo simulations in LTspice: (**a**) all n-FETs (and p-FETs) with identical parameter values in each run, and (**b**) different randomly assigned parameter values for n-FET (and p-FET) in each inverter in each run

In the circuit schematic in Fig. 2.59a, there are two inverters with two n-FETs and two p-FETs. When running Monte Carlo simulations, the randomly picked values for the n-FET V_t adder *ndelvto*, as an example, are identical for both n-FETs. Similarly, both p-FETs are assigned identical V_t values using *pdelvto*. This type of circuit schematic is used for simulating systematic process variations (Sect. 6.4.1). The nominal value *ndelvtonom* is set = 0, and $ndelvtosigma = \sigma V_{ts}$ defines the standard deviation for systematic V_t variations in both n-FETs.

The circuit schematic in Fig. 2.59b shows two inverter symbols at the top level schematic. In Monte Carlo simulations for *ndelvto*, the values of *ndelvto* (*ndelvto*1 and *ndelvto*2) are picked at random for each n-FET in each inverter. This type of schematic is used for simulating a systematic offset in the nominal value (e.g., ΔV_{ts}) along with $ndelvtosigma = \sigma V_{tr}$. In this case, all n-FETs in the circuit schematic will have V_t shifted by ΔV_{ts}. In addition, the V_t of each n-FET will be modulated by the random σV_{tr} assignment in a simulation run. For random variation alone, *ndelvtonom* is set = 0.

The methodology described above applies only to LTspice simulators. Other simulators have different ways of handling random variable assignments when running Monte Carlo simulations. It is prudent to output the assigned values of variables for the circuit components of interest with 100 or more Monte Carlo runs and verify the desired parameter distributions.

In Fig. 2.60 the probability distribution of τ_p for 500 cases for the circuit in Fig. 2.50d, with systematic L_p ($L_{pp} = L_{pn}$), V_{tn} and V_{tp} variations in their respective $\pm 3\sigma$ ranges, is shown. The mean of the distribution is 11.49 ps and $\sigma\tau_p = 1.32$ ps. The maximum and minimum values of τ_p are 15.32 and 7.57 ps respectively. The distribution may get wider when variations in other parameters such as C_{ov} and drain-to-source resistance R_{ds} are included.

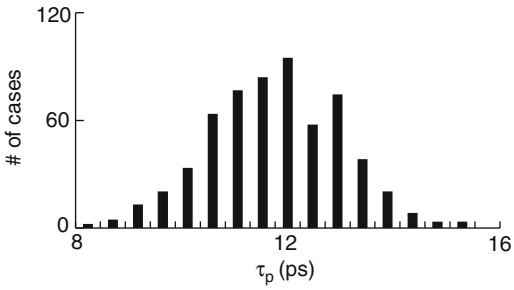


Fig. 2.60 Histograms of average delay τ_p of a standard inverter (FO = 4) for systematic L_p, V_{tn}, and V_{tp} variations. Monte Carlo simulations with 500 cases. 45 nm PTM HP models @ 1.0 V, 25 °C

The number of simulation runs may be reduced dramatically by performing a worst-case analysis. In this case, only values at the extremes of the parameter distributions, (nominal $+3\sigma$) and (nominal -3σ), are assigned. The LTspice command for the worst-case function is

.function wc(nom,tol) if (run == 1, nom, if(flat(1)z> 0,nom + tol, nom − tol))

The SPICE function *flat*(1) returns a random number between −1 and +1 and *tol* may be set to the 3σ value of the parameter. The parameter value is (*nom* + *tol*) if *flat*(1) is between −1 and 0, and (*nom-tol*) if *flat*(1) is between 0 and +1.

Monte Carlo simulations serve well to get an insight into the effect of process and other environmental variations on circuit behavior. If the range and shape of parameter distributions are realistic, the simulation results provide a fair representation of measured electrical test data in the hardware.

2.3 Summary and Exercises

The basic properties of CMOS circuit components and logic gates are reviewed. Circuit simulation setups using LTspice and PTM models, as described in this chapter, are used in examples and exercises throughout the book. Key MOSFET I_{ds}

and capacitance parameters are defined and extracted from simulated *I–V* and *C–V* characteristics. Parameter values for n-FETs and p-FETs are compared to assess their relative strengths.

Logic gate characterization is exemplified with a standard inverter design for generating a lookup table of signal propagation delays by varying capacitive load C_L and signal input signal rise and fall times (τ_{ri}, τ_{fi}). The dependencies of signal PD and PU delays, input capacitance of the inverter, and output signal rise and fall times on τ_{ri}, τ_{fi}, and C_L are graphically illustrated and analyzed. An equivalent *RC* model of the inverter is described in terms of average switching resistances and capacitances.

Special features and circuit simulation setups of commonly used test structures comprising delay chains and ring oscillators for measuring logic gate delays and power in silicon are described. Examples are provided to show the variation of average delay with MOSFET parameter variations. A methodology for running Monte Carlo simulations with systematic and random silicon process variations is introduced.

Exercises are designed for step-by-step evaluation of simulation setups, device models, and circuit power and performance, and for cross-checking and validating the results. Exercises 2.1–2.3 deal with aspects of MOSFET characterization, and Exercises 2.4–2.6 with library book characterization. Exercises 2.7–2.10 are designed to set up delay chain and RO circuit simulations as well as Monte Carlo simulations. These circuit simulation configurations are used in many of the examples and exercises in subsequent chapters.

Note: It would be helpful to create a reference table similar to Table A.4 in Appendix A for the available models at the nominal corner

2.1. (a) Simulate n-FET and p-FET DC I_{ds}–V_{ds} characteristics, and generate a table similar to Table 2.7. Do the I_{ds} parameter values match the published values if available? If not, are these within a reasonable range for the corresponding technology node?

(b) Simulated n-FET I_{on} is found to be lower by 1 %, and I_{off} higher by 20 %, than published values. What would you look for in the simulation setup to correct these errors?

(c) If physical layout/parasitic extraction models are available, simulate n-FET and p-FET I_{ds}–V_{ds} characteristics with a netlist including parasitics and calculate the % change due to parasitics in each of the I_{ds} parameters. Which parameters are affected most and why? If parasitic extraction models are not available, add resistances ($=0.02 * V_{DD}/I_{on}$) in series with S, D, and G terminals for this exercise.

2.2. (a) Simulate n-FET *C–V* characteristics and generate a table similar to Table 2.9.

(b) Simulate *C–V* characteristics and determine average C_g with (1) D held at V_{DD}, S at GND, and (2) D and S held at V_{DD}. How do you explain the difference in C_g values?

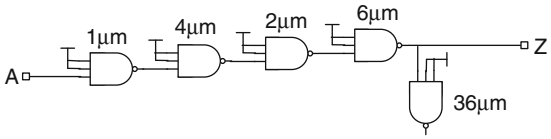
(c) Measure capacitance using the method described in Fig. 2.32 with VD = 0 and VG ramped from GND to V_{DD} and then back to GND. Compare this with the C_g values in the table and explain the difference?

2.3. The mean I_{on} values for isolated n-FETs provided by a silicon foundry are ~5 % higher than the measured values on a test chip designed by the product team. The physical layouts of individual n-FETs in the test chip are identical to the foundry designs except that the test structure is configured to measure 10 n-FETs in parallel on the test chip.
(a) Propose at least four possible causes of this discrepancy?
(b) Recommend additional information/measurements to assist in debug.

2.4. Using the methodology described for library book characterization, the τ_{pd} and τ_{pu} values for a standard inverter for different $\tau_{ri} = \tau_{fi}$ and $C_L = 0.1$ and 1.5 fF are listed in the table below:
(a) Note that τ_{pd} decreases at long rise times whereas τ_{pu} continues to increase. Explain why.
(b) What design or simulation parameters can be changed so that τ_{pd} will continue to increase with τ_{ri}. How does this affect τ_{pu} behavior with τ_{ri}?
(c) Using the data in the table, estimate approximate delay across three inverter books connected in series, each with $C_L = 0.1$ fF, and $\tau_{ri} = \tau_{fi} = 8$ ps?

$\tau_{ri} = \tau_{fi}$ (ps)	τ_{pd} (ps) $C_L = 0.1$ fF	τ_{pu} (ps) $C_L = 0.1$ fF	τ_{pd} (ps) $C_L = 1.5$ fF	τ_{pu} (ps) $C_L = 1.5$ fF
8	4.1	4.3	7.2	7.2
48	5.7	8.3	12.1	14.3
88	5.5	10.5	13.6	18.2
128	4.9	12.4	14.1	21.2
168	4.1	14.0	14.2	23.7
208	3.1	15.5	14.0	25.9
248	2.1	16.9	13.6	28.0

2.5. (a) Design and characterize a NAND3B gate with $(W_n + W_p) = 1.0$ μm and $\tau_{pd} \approx \tau_{pu}$. Obtain plots similar to Figs. 2.37, 2.38, 2.39, 2.40, and 2.41 in the nominal corner.
(b) Set SPICE commands for the following input signal waveforms: (1) linear ramp (pulse function), (2) sine-squared function, (3) output signal of an inverter (FO = 3).
(c) Adjust the waveform parameters to get the same measured τ_r and τ_f.
(d) Measure τ_{pd} and τ_{pu} across the NAND3B using the three different waveform shapes and compare the results.
(e) Using the fitting parameters from the plots from (a), calculate the signal propagation delay from A to Z for the circuit shown below. Compare the calculated delay with simulated delay. The $(W_n + W_p)$ values for each gate in the chain are as indicated, with constant W_p/W_n as sized in (a).



2.6. (a) Estimate C_L/FO for an inverter, NAND2T, NAND3T, and NOR2T. Set $(W_n + W_p) = 1.0$ μm, and W_p/W_n sized to get $\tau_{pd} \approx \tau_{pu}$ for each gate with $\tau_{ri} = \tau_{fi} = 16$ ps.
(b) Set W_p and W_n of non-switching MOSFETs to 0.5× of the value determined in (a). What is the new C_L/FO for these gates? Explain the results.

2.7. (a) Using a delay chain for simulations, design a standard inverter for the model library such that $(W_n + W_p) = 1.0$ μm and $\tau_{pd} \approx \tau_{pu}$ at nominal V_{DD} and 25 °C (nominal corner). How does W_p/W_n compare with I_{onn}/I_{onp} and I_{effn}/I_{effp}?
(b) Measure τ_{pd} and τ_{pu} for this standard inverter at (1) $0.8 \times V_{DD}$, 25 °C, (2) $1.2 \times V_{DD}$, 25 °C, (3) $1.0 \times V_{DD}$, −25 °C and (4) $1.0 \times V_{DD}$, 75 °C. Compare τ_{pd}, τ_{pu}, and τ_{pd}/τ_{pu} in these four simulation corners. Plot τ_{pd} and τ_{pu} as a function of V_{DD} and temperature including the data at $1.0 \times V_{DD}$, 25 °C. Can you use the plots to predict τ_{pd} and τ_{pu} at other V_{DD} and temperature values?
(c) Set *ndelvto* = +0.06 V and *pdelvto* = −0.06 V. Is τ_{pd}/τ_{pu} expected to increase or decrease at nominal V_{DD}? Simulate and compare τ_{pd}/τ_{pu} with the nominal V_t inverter.

2.8. (a) Set up a template for an RO circuit with 50 LGXL stages and a NAND2 enable gate.
(b) Determine τ_p of the NAND3B logic gate used in Exercise 2.5(a) in the RO configuration and compare with previous results.
(c) An RO circuit fails to oscillate in simulation. What are the possible sources of error in (1) RO circuit, and (2) simulation setup?

2.9. (a) Set up ROs with 4, 10, 24, 50, and 100 standard inverter (FO = 3) stages and a NAND2 gate (FO = 1, $W_n + W_p$ = 1 μm) enable gate. Determine τ_p and P_{ac} in the nominal corner for each design and compare.
(b) Compare τ_p values of the ROs with the τ_p value from the delay chain simulation in Exercise 2.7.
(c) It is desirable to minimize RO area on silicon. How many stages would you recommend for the RO to reduce error in τ_p to <0.5 %?

2.10. (a) Run Monte Carlo simulations for 500 cases to measure τ_{pd} and τ_{pu} with $\pm 3\sigma$ systematic variations (Gaussian distribution) in L_p ($L_{pn} = L_{pp}$), V_{tn}, and V_{tp} for a NAND2T logic gate book designed to have $\tau_{pd} \approx \tau_{pu}$ in the nominal corner.
(b) Plot histograms of L_p, V_{tn}, and V_{tp} measure the range. How do these compare with their corresponding 6σ ranges?
(c) Plot histograms of τ_{pd} and τ_{pu} compare their ranges.

References

1. Linear technology. Design simulation and device models. http://www.linear.com/designtools/software/. Accessed 21 Jul 2014
2. Synopsys HSPICE. http://www.synopsys.com/Tools/Verification/AMSVerification/CircuitSimulation/HSPICE/Pages/default.aspx. Accessed 21 Jul 2014
3. NGSPICE Mixed mode-mixed signal simulator. http://ngspice.sourceforge.net. Accessed 21 Jul 2014
4. BSIM group. http://www-device.eecs.berkeley.edu/bsim/. Accessed 21 Jul 2014
5. Predictive technology models website. http://ptm.asu.edu/latest.html. Accessed 21 Jul 2014
6. Cao Y, Sato T, Sylvester D, Orshansky M, Hu C (2000) New paradigm of predictive MOSFET and interconnect modeling for early circuit design. In: Proceedings of the IEEE 2000 custom integrated circuits conference, pp 201–204
7. Zhao W, Cao Y (2006) New generation of predictive technology model for sub-45 nm early design exploration. IEEE Trans ED-53:2816–2823
8. Weste NH, Harris D (2010) CMOS VLSI design: a circuit and systems perspective, 4th edn. Addison-Wesley, Boston
9. Baker RJ (2010) CMOS circuit design, layout and simulation, 3rd edn. Wiley, Hoboken
10. Rabaey JM, Chandrakasan A, Nikolic B (2003) Digital integrated circuits, 2nd edn. Prentice Hall, Upper Saddle River
11. Sutherland I, Sproull B, Harris D (1999) Logical effort: designing fast CMOS circuits. Morgan Kaufmann, New York
12. Taur Y, Ning TH (2009) Fundamentals of modern VLSI devices, 2nd edn. Cambridge University Press, New York
13. Sze SM (2006) Semiconductor devices: physics and technology, 3rd edn. Wiley, Hoboken
14. Campbell SA (2001) The science and engineering of microelectronic fabrication, 2nd edn. Oxford University Press, Oxford
15. Jaeger RC (2001) Introduction to microelectronic fabrication, vol 5, 2nd edn, Modular series on solid state devices. Prentice Hall, Upper Saddle River
16. Bhushan M, Ketchen MB (2011) Microelectronic test structures for CMOS technology. Springer, Berlin
17. Joshi RV, Kroell K, Chuang CT (2004) A novel technique for steady state analysis for VLSI circuits in partially depleted SOI. In: Proceedings of the 17th international conference on VLSI design, pp 832–836

http://www.springer.com/978-1-4939-1348-0

CMOS Test and Evaluation
A Physical Perspective
Bhushan, M.; Ketchen, M.B.
2015, XIII, 424 p. 338 illus., Hardcover
ISBN: 978-1-4939-1348-0