Digital MOS Integrated Circuits

Edited by
Mohamed I. Elmasry

Digital MOS Integrated Circ

Digital MOS Integrated Circuits

Edited by
Mohamed I. Elmasry
Associate Professor of
Electrical Engineering
University of Waterloo

A volume in the IEEE PRESS Selected Reprint Series,
prepared under the sponsorship of the
IEEE Solid-State Circuits Council.

IEEE
PRESS

5506719

The Institute of Electrical and Electronics Engineers, Inc., New York

PRINTED IN THE UNITED STATES OF AMERICA

Sole Worldwide Distributor (Exclusive of IEEE):

JOHN WILEY & SONS, INC.
605 Third Ave.
New York, NY 10016

Wiley Order Numbers: Clothbound: 0-471-86202-9
Paperbound: 0-471-86203-7

IEEE Order Numbers: Clothbound: PC01503
Paperbound: PP01511

Library of Congress Catalog Card Number 81-6522

Library of Congress Cataloging in Publication Data
Main entry under title:

Digital MOS integrated circuits.

(IEEE Press selected reprint series)
Bibliography: p.
Includes indexes.
1. Integrated circuits—Large scale integration—Addresses, essays, lectures.
2. Integrated circuits—Very large scale integration—Addresses, essays, lectures.
3. Metal oxide semiconductors—Addresses, essays, lectures. I. Elmasry, Mohamed I., 1943-
TK7874.D535 621.381'73 81-6522
ISBN 0-87942-151-7 AACR2
ISBN 0-87942-152-5 (pbk.)

Preface

Over the past two decades, the spectacular development of digital systems in the areas of data and signal processing has been a result of the development of integrated circuit technology. Today, we are constrained only by what it is possible to integrate *economically* on a single silicon chip.

The key to using a technology is in circuits—they are the link between a technology and the system elements. Each new generation of digital, data, or signal processing systems can be linked to an advance in a technology, with its appropriate set of circuit techniques, CAD tools, and system architectures.

In the early 1960's small scale integrated (SSI) bipolar chips with tens of transistors were available. The system's functional complexity was on the order of a flip-flop. About the same time, MOS devices were introduced. They offered advantages over existing digital bipolar logic families, mainly simpler fabrication process and smaller size. In about 1965, medium scaled integrated (MSI) bipolar chips with hundreds of transistors were introduced. The system's functional complexity was on the order of a counter. By 1970, large scale integrated (LSI) bipolar and MOS chips with thousands of transistors were being designed. For the first time, this allowed the integration of a subsystem on a chip—a complexity on the order of a microprocessor. At this point, integrated circuit designers found themselves facing new, interesting, and challenging frontiers by addressing themselves to such problems as system architecture, chip partitioning, and logic simulation. This trend continued and by 1980 very large scale integrated (VLSI) chips, mainly MOS, with tens of thousands of transistors, became the state-of-the-art. Today, complete systems can be integrated on a single silicon chip, and VLSI circuit design has become a most interesting and challenging career.

Will this trend continue? The experience of the semiconductor industry has been that requirements, once identified and stated, are met sooner or later, and it seems that barriers are more apparent than real. At this stage it would appear there are several problems looking for solutions: yield, testing, reliability, CAD tools, standardization, and educating a new generation of VLSI circuit designers. However, these problems, like the problems of the past, will be solved. For example, yields improve as production techniques develop, testing is fundamental to good engineering and rapid advances are being made in understanding the reliability of large systems and the physical factors affecting VLSI chip reliability. Progress is continually being made in developing CAD tools for VLSI chip design and in standardizing many VLSI chip parameters, e.g. a lower operating power supply than the 5V used today. A trend towards educating a new generation of VLSI circuit designers has developed in many universities and is supported by the industry. Textbooks have recently been written to emphasize this trend and this book is an attempt in that direction.

Part I covers the use of MOS devices in digital circuit design. Single-channel MOS and CMOS circuits are discussed. The use of different integrated load structures is analyzed. Some specific circuit techniques, eg. bootstrapping, are described. The dynamic-mode of operating a MOS circuit is discussed.

In Part II, progress in digital MOS VLSI is examined. More specifically, the device-based route to VLSI: scaling of MOSFETs is presented. Some of the state-of-the-art technologies are described. A circuit-based route to VLSI, eg. use of device merging, is also discussed. Finally, VLSI circuit and system design approaches are given.

In Part III, MOS memory cells and their peripherial circuits are examined.

In Part IV, some examples in logic and memory applications of digital MOS are given.

Finally, a selected bibliography with over 500 references, grouped in ten sections, is given.

In conclusion, let me state that a review of predictions made about semiconductor circuits in the past would indicate that those concerned with what was able to be

done were usually more accurate than those concerned with what was not able to be done. Our experience has been that, providing there is a real requirement, the technology, in conjunction with circuit innovations, was developed to meet the need. Thus, if we wish to predict the future capability of digital integrated circuit technology, we must identify the future applications, and then confidently predict that techniques will be invented to meet them.

M. I. Elmasry
Waterloo, Ontario, Canada
July 1980
San Diego, California, U.S.A.
January 1981

Contents

Part I
MOS in Digital Circuit Design

This Part deals with the analysis and the design of MOS digital circuits. It is expected that the reader is familiar with basic MOS physical characteristics, modeling aspects, and general circuit techniques. An excellent treatment of the physical aspects of MOSFETs and semiconductor devices in general is given by Grove [1]. A reference book in this area is Sze [2]. The modeling aspects, are covered by Cobbold [3] and Richman [4]. An excellent treatment of the different aspects of integrated circuits (both bipolar and MOS) is given by Hamilton and Howard [5] and by Muller and Kamins [6]. MOS circuits is the topic of a classic text by Crawford [7], of texts by Penney and Lau [8], Carr and Mize [9], and of the recent system/layout oriented text by Mead and Conway [10].

Dealing with circuit design is different from dealing with circuit analysis. While circuit analysis is an engineering science, circuit design is a science and an art. The artistic aspects of circuit design are difficult to explain because they are not algorithmic. These artistic aspects can also differ from designer to designer with the only reasonable explanation that they are efficient and they work. One of these artistic related aspects is the use of circuit techniques. Most of these techniques are surprisingly universal. For example, the technique of merging different semiconductor devices to achieve VLSI structures is applicable to bipolar circuits (as in integrated injection logic [I^2L]), as well as MOS circuits (as in single device well [SDW] MOSFETs). SDW MOSFETs are explained in Section 2.3. In addition, Section 1.3 deals with a circuit technique which is more suitable to MOS than to bipolar circuits, namely bootstrapping.

Good circuit designers are not born, their skills are developed by gaining knowledge in a relatively *wide* spectrum of engineering sciences. For example, the study of bipolar circuits is recommended for MOS circuit designers. The text by Meyer, Lynn, and Hamilton [11] is a good source of information.

Section 1.1 is a tutorial paper which deals with basic concepts in MOS digital circuit design. NMOS as well as CMOS circuits are discussed. Both dc and transient analysis are presented. Basic building blocks, eg. inverters, logic gates, and flip-flops are explained. Static and dynamic mode of operation are studied.

Section 1.2 contains two papers on depletion type MOSFETs. The paper by Mashuhara *et al.* explains circuit design consideration for using depletion type MOSFETs as loads. Static and transient characteristics are included. Although the technology used to realize the experimental test structures is not the state-of-the-art (1980s), the circuit design approach is still basically the same. In a paper by El-Mansy the analysis and the characterization of depletion-type MOSFETs is presented. A model and its basic elements are given in terms of processing data.

Section 1.3 contains three papers on circuit techniques which are commonly used in MOS digital circuit design. The first two papers deal with a circuit technique which has been widely used and referred to as "bootstrapping." It provides a circuit-based solution to eliminate threshold losses in MOS circuits. The third paper presents a novel MOS circuit configuration which is similar to I^2L bipolar circuit structures, as it uses the same circuit partitioning technique. Thus, it proves the point that circuit techniques are more or less universal.

The last section deals with the comparison of MOS logic gates. The paper by Cook *et al.* compares four-phase dynamic logic circuits, CMOS, single channel static MOS logic circuits with linear (nonsaturated) and depletion-type loads. The parameters used for this comparison are power dissipation, delay, and silicon area. The comparison is done by using the early 1970's MOS technologies; doing a similar comparison using the early 1980's MOS technologies is left as a challenging project to the reader.

The reader is referred to Section 1 of the bibliography which contains references on the use of MOSFETs in digital circuit design.

References

[1] A. S. Grove, "Physics and Technology of Semiconductor Devices," John Wiley, 1976.

[2] S. M. Sze, "Physics of Semiconductor Devices," John Wiley, 1969.

[3] R. S. C. Cobbold, "The Theory and Applications of Field-Effect Transistors," Wiley-Interscience, 1970.

[4] P. Richman, "MOS Field-Effect Transistors and Integrated Circuits," John Wiley, 1973.

[5] D. J. Hamilton and W. G. Howard, "Basic Integrated Circuit Engineering," McGraw-Hill, 1975.

[6] R. S. Muller and T. I. Kamins, "Device Electronics for Integrated Circuits," John Wiley, 1977.

[7] R. H. Crawford, "MOSFET in Circuit Design," McGraw-Hill, 1967.

[8] W. M. Penney and L. Lau, "MOS Integrated Circuits," Van Nostrand Reinhold, 1972.

[9] W. N. Carr and J. P. Mize, "MOS/LSI Design and Application," McGraw-Hill, 1972.

[10] C. Mead and L. Conway, "Introduction to VLSI Systems," Addison-Wesley, 1980.

[11] C. S. Meyer, D. K. Lynn, and D. J. Hamilton, "Analysis and Design of Integrated Circuits," McGraw-Hill, 1968.

Section 1.1
Digital MOS Integrated Circuits

Digital MOS Integrated Circuits: A Tutorial

MOHAMED I. ELMASRY, SENIOR MEMBER IEEE

Abstract—**Basic digital NMOS and CMOS integrated circuits are analyzed. Dc and transient performance is studied using first-order design equations. Static and dynamic circuits are discussed. The effects of device parameters on circuit performance are explained.**

1. Introduction

The dual purpose of this tutorial paper is to review the main characteristics of MOS transistors as they are related to digital circuit design and to analyze basic static and dynamic NMOS digital circuits using first-order dc and transient design equations. The study of digital NMOS circuits is followed by a similar study of digital CMOS circuits, and the most basic digital circuit, the static inverter, is analyzed in detail. This is followed with an explanation of how the design of logic gates and flip-flops, both static and dynamic, is related to the design of the simple static inverter.

2. MOS Device Characteristics

Figure 1 shows a diagrammatic cross-section in a NMOS transistor [1]. It consists of two n^+ regions, introduced in a P substrate by diffusion or ion implantation. In circuit operation, the more positive region is called the drain while the other region is called a source. The surface region between the source and drain is called the channel. The conduction through this channel is controlled by the voltage on the gate, which is either metal or polysilicon. The gate is separated from the channel by a thin layer of a dielectric, usually silicon oxide.

If the voltage of the source terminal is taken as a reference, then V_{DS}, V_{GS}, and V_{BB} are the voltages of the drain, gate, and substrate respectively. V_{BB} is referred to as the back-gate bias. For NMOS digital circuits V_{BB} is an applied negative or zero voltage. The negative V_{BB} is either supplied from off-chip, or is generated on-chip from the available positive power supply V_{DD} (see Appendix A).

An applied positive V_{DS} allows electrons, when present in the channel, to drift from the source to the drain causing I_{DS} to flow from the drain to the source. In depletion-type NMOS devices, electrons are present in the channel even at $V_{GS} = 0$. This is achieved by ion implanting the surface channel with n-type material. An increase in V_{GS} increases I_{DS}. However, if V_{GS} is negative and larger than $|V_T|$, where V_T is the threshold voltage of the depletion-type device, the channel is depleted and I_{DS} is reduced to zero. In enhancement-type NMOS, electrons are only present at the surface if V_{GS} is positive and larger than V_T where V_T is the threshold voltage of the enhancement-type device. An increase in V_{GS} increases I_{DS}.

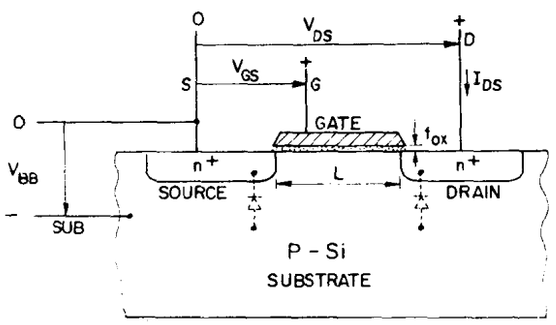


Fig. 1. Diagrammatic cross-section in an NMOS transistor showing parasitic diodes and terminal voltages.

2.1 DC Characteristics

Two important dc characteristics of the NMOS transistor are shown in Fig. 2:

- the drain current I_{DS} vs. the drain voltage V_{DS} for different values of V_{GS} at a given substrate bias V_{BB}; and
- the drain current I_{DS} vs. the gate voltage V_{GS} at a given V_{BB}.

Three regions of operation can be distinguished on the I_{DS} vs. V_{DS} characteristic:

1. The off region, where

$$V_{GS} < V_T$$

$$I_{DS} \simeq 0$$

This region is also referred to as the subthreshold region where I_{DS} increases exponentially with V_{DS} and V_{GS} [1]. The value of I_{DS} in this region is much smaller than its value

Manuscript received April 28, 1981.

The author is with the Department of Electrical Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. This work is supported in part by NSERCC Grants A9334 and PRAI 7705.

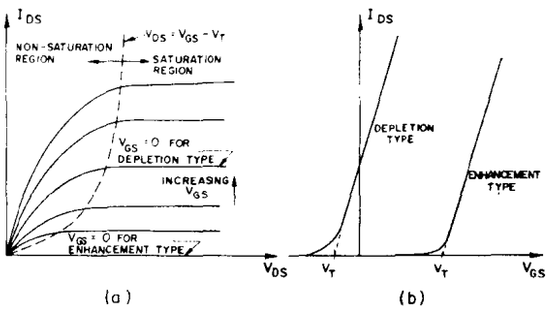


Fig. 2. (a) I_{DS} vs. V_{DS} and (b) I_{DS} vs. V_{GS} for a given V_{BB}. Note that V_T is positive for enhancement-type NMOS devices and negative for depletion-type.

when $V_{GS} > V_T$. Thus, in many NMOS circuits for $V_{GS} < V_T$, the transistor is considered off and $I_{DS} \simeq 0$. However, the small value of I_{DS} in this region could affect the circuit performance as in MOS dynamic memory circuits.

2. The nonsaturation region, where

$$V_{DS} < V_{GS} - V_T$$

$$I_{DS} = \beta[(V_{GS} - V_T)\, V_{DS} - \frac{1}{2} V_{DS}^2] \tag{1}$$

3. The saturation region, where

$$V_{DS} \geq V_{GS} - V_T$$

$$I_{DS} = \frac{\beta}{2} [V_{GS} - V_T]^2 \tag{2}$$

where

$$\beta = \frac{W}{L} \frac{\epsilon_{ox} \mu}{t_{ox}} = \frac{W}{L} K' \tag{3}$$

W = width of the MOS channel

L = length of the MOS channel (in the direction of current flow)

ϵ_{ox} = permittivity of the gate oxide

t_{ox} = thickness of the gate oxide

μ = average surface mobility of carriers (μ_n in the case of electrons in NMOS, and μ_p in the case of holes in PMOS)

V_T = threshold voltage

C_{ox} = gate capacitance per unit area = $\dfrac{\epsilon_{ox}}{t_{ox}}$

2.1.1 *The Conduction Factor K'*

The conduction factor K' ($\equiv \epsilon_{ox}\, \mu / t_{ox}$) is technology dependent and is specified for a given MOS process. Thus it is not a circuit design variable. For a t_{ox} range of 1000-500°A, it has a typical value of 12-25 $\mu A/V^2$ for NMOS devices and 5-10 $\mu A/V^2$ for PMOS. As the value of t_{ox} decreases with advances in technology, K' increases. The difference between K'_n and K'_p results from the fact that $\mu_n \sim 2.5\, \mu_p$.

K' is a function of temperature because of its dependency on μ:

$$\frac{K'}{K'_0} = \left(\frac{T}{T_0}\right)^{-3/2} \tag{4}$$

where K'_0 is the value of K' at room temperature (T_0 = 298°K) and T is the absolute temperature (°K).

2.1.2 *The Geometrical Ratio (W/L)*

This ratio is a circuit design parameter. The minimal value of L (L_{min}) is determined by the MOS fabrication process. L_{min} is determined mainly by the mask channel length, the tolerances on that length and the lateral diffusions of both the source and the drain regions. The minimum value of W is usually in the order of the minimum value of L. Increasing (W/L) will increase the drain current for a given set of operating voltages. However, increasing W increases the gate area and the source and the drain diffusion areas and hence increases the value of the capacitances associated with the gate and with the source-substrate and the drain-substance junctions.

2.1.3 *The Threshold Voltage V_T*

V_T is a function of the MOS processing parameters and the substrate bias V_{BB}. In general, it is also a function of V_{DS}. In order to highlight these different functional dependencies, V_T of an enhancement-type NMOS can be written as:

$$V_T = V_{TO} + \Delta V_T(V_{BB}) - \Delta V_T(V_{DS}) \tag{5}$$

where

$$V_{TO} = \left(\phi_{GS} - \frac{Q_{SS}}{C_{ox}}\right) + \gamma(2\phi_F)^{1/2} + 2\phi_F \tag{5a}$$

$$= V_{FB} + \gamma(2\phi_F)^{1/2} + 2\phi_F$$

$$\Delta V_T(V_{BB}) = \gamma[(|V_{BB}| + 2\phi_F)^{1/2} - (2\phi_F)^{1/2}] \tag{5b}$$

$$\Delta V_T(V_{DS}) = z(V_{DS} + 2|V_{BB}| + 2V_{Bi}) \tag{5c}$$

$$\gamma = \frac{(2\, \epsilon_{ox} q N_B)^{1/2}}{C_{ox}} \tag{5d}$$

$$z = \frac{\eta_0\, (x_j, N_B)}{C_{ox}\, L^n} \tag{5e}$$

ϕ_{GS} = gate voltage necessary to counter balance the gate-to-silicon work function difference.

Q_{SS}/C_{ox} = gate voltage necessary to counter balance the effect of the oxide surface charge Q_{SS}.

$V_{FB} = \phi_{GS} - \dfrac{Q_{SS}}{C_{ox}}$ = gate voltage necessary to cause the flat band condition at the silicon surface, hence the name "flat-band voltage". The flat band condition occurs when the energy bands in the substrate are flat at the surface [1]. In this condition, there is zero electric field at the silicon surface. If the gate voltage is more positive than V_{FB}, for P-type substrates, the silicon surface is in depletion, i.e., there are no mobile carriers at the surface. If the gate voltage is further increased the surface starts to "invert", i.e., electrons are attracted to the surface forming a conductive channel.

$\gamma(2\phi_F)^{1/2}$ = gate voltage necessary to counter balance the effect of the charge created by the exposed dopants at the surface. ϕ_F is the substrate Fermi potential at the surface, and is equal to $\dfrac{KT}{q} \ln \dfrac{N_B}{n_i}$, where $\dfrac{KT}{q}$ is the thermal voltage, n_i is the intrinsic concentration for silicon, $n_i^2 = 1.5 \times$

$10^{33}\ T^3\ e^{-1.2q/KT}\ cm^{-6}$, N_B is the average substrate concentration at the silicon surface. Note that a surface implant at the MOS channel is usually used to influence N_B [2].

$2\phi_F$ = additional gate voltage, by definition necessary to produce a "strong inversion" condition at the silicon surface.

$\Delta V_T(V_{BB})$ = increase in the threshold voltage due to the reverse bias on the back-gate (substrate). If V_{BB} = 0 then $\Delta V_T(V_{BB}) = 0$.

$\Delta V_T(V_{DS})$ = decrease in the threshold voltage due to the short channel effect. For large L, $\Delta V_T(V_{DS})$ tends to zero.

$\eta_0(x_j, N_B)$ = factor which is a function of the source and drain junction depth x_j and N_B [3].

n = factor which ranges between 2.6 and 3.2 for $10^{15}\ cm^{-3} \leqslant N_B \leqslant 10^{16}\ cm^{-3}$, $1.5\ \mu m \leqslant x_j \leqslant 0.41\ \mu m$ [3].

V_{Bi} = source (or drain)-substrate built-in voltage = $\frac{KT}{q} \ln \frac{N_B N}{n_i}$ where N is the average impurity concentration of the source and drain diffusions.

Examining equation (5) reveals the following:

1. V_{TO} is not a function of the operating voltages and is a function of temperature.
2. The sensitivity of the threshold voltage V_T to V_{BB}: $\Delta V_T(V_{BB})$ is determined by γ which is a function of $N_B^{1/2}\ C_{ox}^{-1}$. Increasing N_B or decreasing C_{ox}, i.e., increasing t_{ox}, will increase that sensitivity.
3. The sensitivity of the threshold voltage V_T to V_{DS}: $\Delta V_T(V_{DS})$ is determined by z which tends to zero for long channel devices. z is an empirical factor [3] which is proportional to x_j, $1/N_B$, $1/C_{ox}$ in addition to $1/L$. Decreasing x_j, increasing N_B or decreasing t_{ox} (hence increasing C_{ox}) reduces the sensitivity of V_T to V_{DS} in short channel devices.
4. In short channel and narrow channel devices. γ is a function of both L and W; a decrease in L decreases γ and a decrease in W increases γ. This dependency of γ on L and W affects V_{TO} and $\Delta V_T(V_{BB})$ and could be incorporated empirically in equation (5d). This effect should be considered in the design of digital circuits using MOS devices of small dimensions because a transistor having $W = W_{min}$, $L = L_{min}$ may have a different dc characteristic from one having $W = mW_{min}$ and $L = mL_{min}$, although (W/L) of the two transistors are the same.

Note that temperature affects both K' and V_T in such a way that the effects on I_{DS} could cancel each other. Thus, MOS transistors can be operated so that they exhibit positive, negative, or zero temperature coefficient [1].

2.2 Transient Characteristics

The transient performance of an MOS integrated circuit is a function of the total capacitance at the output node. This capacitance C_{out} is the summation of the parasitic output capacitance C_0 and the input gate capacitance(s) of the loading stage(s) C_{IN}.

The parasitic output capacitance C_0 is the summation of two capacitances:

1. C_J—the junction capacitance of the output diffusion(s). This capacitance varies with the junction voltage V_j:

$$\frac{C_J}{C_{JO}} = \left(1 - \frac{V_j}{V_{Bi}}\right)^n$$

where

C_J = junction capacitance at voltage V

C_{JO} = junction capacitance at zero voltage

V_j = junction voltage (negative for reverse-bias)

V_{Bi} = built-in junction potential

n = factor between 0.5 and 0.3 depending upon junction abruptness

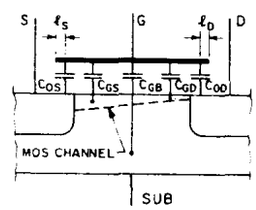
2. C_{INT}—the interconnector capacitance associated with metal, polysilicon, or diffusion interconnection lines. C_{INT} is voltage independent and usually contributes to C_{out} in LSI circuits where complex interconnection patterns exist.

The input capacitance C_{IN} of an MOS transistor consists of the following components [1,4,5,6] as shown in Fig. 3(a).

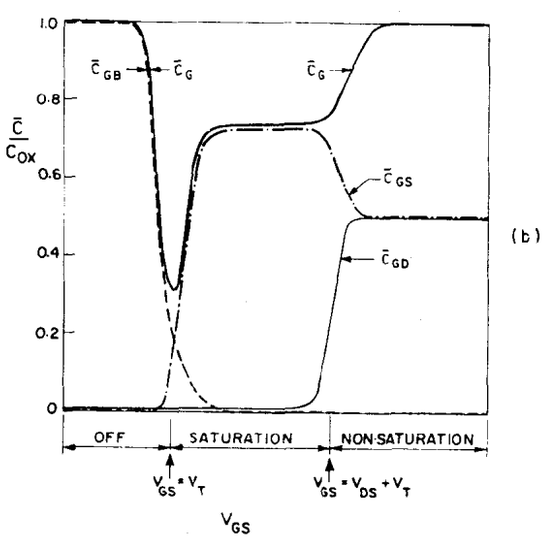
C_{OS}, C_{OD}—the source and the drain overlap capacitances resulting from the overlap of the gate on the source and the drain diffusions: $C_{OS} = C_{ox} \ell_S W$, $C_{OD} = C_{ox} \ell_D W$ where ℓ_S and ℓ_D are the overlap lengths.

C_{GS}, C_{GD}— represent gate to channel capacitances lumped at the source and drain regions of the channel respectively:

$$C_{GS} = C_{ox} WL\ f_S(V),\ C_{GD} = C_{ox} WL\ f_D(V)$$



(a)



(b)

Fig. 3. (a) The different components of C_{IN}; (b) the voltage dependency of the capacitances per unit area: $\bar{C}_{GS}$, $\bar{C}_{GD}$, $\bar{C}_{GB}$, and $\bar{C}_G$ vs. V_{GS}.

C_{GB}—the gate-substrate capacitance:

$$C_{GB} = C_{ox}WL\, f_B(V)$$

where $f_S(V)$, $f_D(V)$ and $f_B(V)$ are voltage dependent functions. Figure 3(b) demonstrates the nature of the voltage dependency of the capacitances per unit areas: $\bar{C}_{GS}$, $\bar{C}_{GD}$, $\bar{C}_{GB}$, and $\bar{C}_G$ of an NMOS transistor vs. V_{GS} where $\bar{C}_G = \bar{C}_{GS} + \bar{C}_{GD} + \bar{C}_{GB}$. When the transistor is off, the only nonzero component is $\bar{C}_{GB}$. This component is due to the series combination of the surface depletion layer and gate oxide capacitances. As the transistor turns on $\bar{C}_{GB}$ reduces to zero because of the shielding effect of the inversion layer. In the nonsaturation region, the source and the drain regions of the MOS channel are inverted and $\bar{C}_{GS} \simeq \bar{C}_{GD} \simeq C_{ox}/2$. In the saturation region, where the drain region of the channel is pinched-off, $\bar{C}_{GD}$ reduces to zero from $C_{ox}/2$ and $\bar{C}_{GS}$ increases from that value to approximately 2/3 C_{ox}. Figure 3(b) shows that $\bar{C}_G$ vs. V_{GS} has a minimum just below $V_{GS} = V_T$.

2.3 Leakage and Breakdown

It is important to consider leakage currents in an MOS chip, especially if the circuit is operating in a dynamic mode and particularly in dynamic memories. The leakage currents are associated with the source and drain p-n junctions. The absolute values of these currents should be minimized, and their variation with temperature should be considered [1].

A potential source of parasitic current in MOS circuits is the current associated with the thick field oxide MOS transistors which have a thick (field) oxide. This current increases exponentially as the gate voltage approaches the threshold voltage of the field-oxide MOSFETs. The threshold voltage of the field-oxide MOSFETs can be increased by increasing the surface doping density under the field oxide.

In MOS circuits, breakdown can occur by different mechanisms. Avalanche breakdown can occur in the reverse-biased drain-substrate junction. Punch-through breakdown can also occur if the depletion region of the drain-substrate junction reaches the source-substrate junction. Either type of breakdown may predominate for a given MOS structure, e.g., punch-through breakdown could be the predominate for short channel MOSFETs. Although both types of breakdowns are related to the *p-n* junction characteristics, they are strongly affected by the presence of a gate oxide and a conducting gate [4].

3. The Static NMOS Inverter

The NMOS inverter, as shown in Fig. 4(a), consists of an enhancement type driver transistor and a load. The load is one of the following: (a) a saturated enhancement-type NMOS device, (b) a nonsaturated enhancement-type NMOS device, (c) a depletion-type NMOS device, or (d) a polysilicon resistor.

Figure 4(b) shows the load lines of the above four loads superimposed on the I_{DS} vs. V_{DS} of the driver. The intersection of the load line with the driver characteristic for $V_{GS} = V_{IN} = V_0$, gives $V_{DS} = V_{OUT} = V_1$ where V_0 is the low voltage level representing logical '0' (see Appendix B), $V_0 < V_{TD}$, V_{TD} is the threshold voltage of the driver and V_1 is the high voltage level representing logical '1'. Similarly, the intersection of the load line with the driver characteristics for $V_{GS} = V_{IN} = V_1$, gives $V_{DS} = V_{OUT} = V_0$.

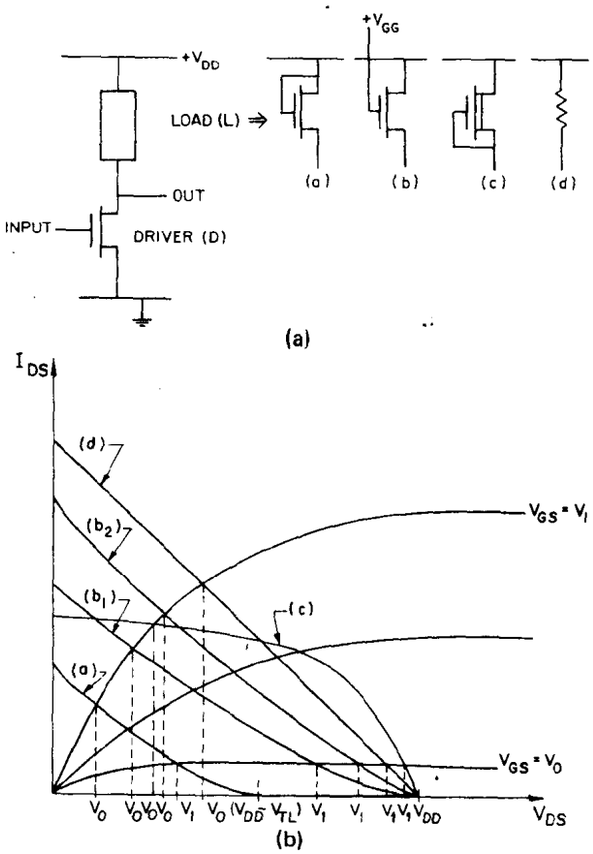


Fig. 4. (a) NMOS inverter with different loads; (b) load lines superimposed on the driver transistor I_{DS} vs. V_{DS} characteristic. V_{GG} for case $(b_2) > V_{GG}$ for case (b_1).

3.1 DC Analysis

3.1.1 *Saturated Enhancement-type Load*

The saturated enhancement-type load was used in the early digital MOS integrated circuits. Figure 5 shows an inverter using NMOS devices. The load device has its gate connected to its drain (i.e., $V_{DS} = V_{GS}$), and operates in the saturation region when it conducts, since $V_{DS} > (V_{GS} - V_{TL})$, where V_{TL} is the threshold voltage of the load device.

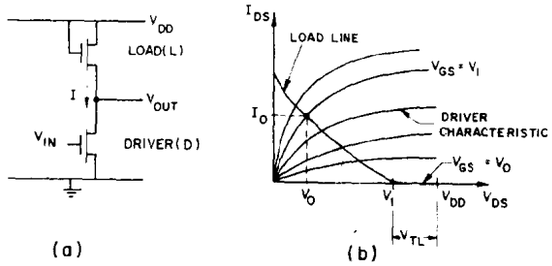


Fig. 5. (a) NMOS inverter with saturated enhancement-type load; (b) load line superimposed on the driver I_{DS} vs. V_{DS} characteristic, showing the operating current I_0 when $V_{OUT} = V_0$.

The dc circuit operation is explained as follows:

(a) when $V_{IN} = V_0$, $V_0 < V_{TD}$, the driver is off, $I \approx 0$ and

$$V_{OUT} = V_1 = V_{DD} - V_{TL} \tag{6}$$

(b) when $V_{IN} = V_1$, $V_1 > V_{TD}$, the driver is operating in the nonsaturation region, the load device is in saturation and $I = I_0$ where:

$$I_0 = \frac{K'\left(\frac{W}{L}\right)_L}{2}[(V_{DD} - V_0) - V_{TL}]^2 \quad (7a)$$

$$= K'\left(\frac{W}{L}\right)_D [(V_1 - V_{TD})V_0 - \frac{1}{2}V_0^2] \quad (7b)$$

If $V_0 \ll V_{DD}$ and $V_0 \ll V_1 - V_{TD}$ then (7) can be simplified:

$$I_0 \simeq \frac{K'\left(\frac{W}{L}\right)_L}{2}(V_{DD} - V_{TL})^2 = \frac{K'\left(\frac{W}{L}\right)_L}{2}V_1^2 \quad (8a)$$

$$\simeq K'\left(\frac{W}{L}\right)_D (V_1 - V_{TD})V_0 \quad (8b)$$

From (8) we obtain:

$$\frac{\left(\frac{W}{L}\right)_D}{\left(\frac{W}{L}\right)_L} \geq \frac{V_1}{2V_0} \quad (9)$$

and

$$V_0 \simeq \frac{I_0}{K'\left(\frac{W}{L}\right)_D (V_1 - V_{TD})} \quad (10)$$

Thus from (9) if V_1/V_0 is taken to be 10, $\left(\frac{W}{L}\right)_D / \left(\frac{W}{L}\right)_L$ should be ≥ 5. Because the ratio of (W/L) of the driver to that of the load must be greater than unity, the circuit is referred to as "a ratioed circuit." To summarize: the voltage levels of the NMOS inverter V_1, V_0, V_t, where V_t is the logic threshold of the inverter (see Appendix B) and the dc power dissipation are given by:

$$V_1 = V_{DD} - V_{TL}$$

$$V_0 \cong \frac{I_0}{K'\left(\frac{W}{L}\right)_D (V_1 - V_{TD})}$$

$$V_t \geq V_{TD} \quad (11)$$

$$P_{DC}\ (@\ V_{OUT} = V_1) \simeq 0$$

$$P_{DC}\ (@\ V_{OUT} = V_0) = I_0 V_{DD}$$

where I_0 is given by (8a).

In (11) I_0 is determined by transient considerations or by the allowable power dissipation. V_{DD} is usually fixed by subsystem design considerations and today it is typically = 5 V.

It is clear that V_1 increases as the power supply voltage V_{DD} increases. The maximum allowable V_{DD} must be less than the junction breakdown voltage of the drain-substrate junction. It should also be less than the voltage at which the parasitic field-oxide MOS transistors start to conduct.

Because the substrate-source junction of the load device is more reverse biased than the driver, V_{TL} is larger than V_{TD} due to the back-gate bias effect. The higher the V_{BB} value of the load, the higher the value of V_{TL}, and the lower the value of V_1. Because $V_1 = V_{DD} - V_{TL}$, the saturated enhancement-type MOS load is said to introduce "threshold losses" in the value of V_1. As we explain in the following sections, depletion-type nonsaturated enhancement type and bootstrapped MOS loads do not introduce threshold losses.

The low logical voltage level, V_0, is a function of the operating current I_0. As I_0 increases, the dc power dissipation increases and the speed of operation also increases as we explain in 3.2. V_0 can be reduced (hence increasing the logic swing and the NM_0 noise margin) by increasing the geometrical ratio $(W/L)_D$ of the driver transistor, hence increasing its size, with respect to $(W/L)_L$.

The threshold voltage V_t of the inverter (see Appendix B) is $\geq V_{TD}$. The higher the ratio $[(W/L)_D/(W/L)_L]$ is, the closer V_t to V_{TD} becomes. If the driver is replaced with a number of stacked transistors connected in a series, as in the case of an MOS NAND gate (see 4), then V_{TD} and hence V_T of the upper input transistors are higher than that of the lower transistors because of the back-gate bias effect. This is a drawback for using NAND gates in NMOS logic design.

3.1.2 *Nonsaturated Enhancement-type Load*

One drawback of using a saturated enhancement-type load is the threshold voltage losses caused by the load, resulting in $V_1 = V_{DD} - V_{TL}$.

This situation can be rectified if the load is operating in the nonsaturation region by connecting its gate to V_{GG} where $V_{GG} > (V_{DD} + V_{TL})$. In this case V_1 can be made close to V_{DD} by increasing V_{GG} as shown in Fig. 4(b). The inverter dc circuit operation is explained as follows:

(a) when $V_{IN} = V_0$, $V_0 < V_{TD}$, the driver is off, the load is nonsaturated and $V_{OUT} = V_1 = V_{DD} - V_{DS}|_{load}$ where $V_{DS}|_{load}$ is the voltage drop across the load device. If the inverter is driving only MOS gates, which is usually the case with high input impedances, then $V_{DS}|_{load} \simeq 0$ and $V_1 \simeq V_{DD}$.

It should be noted that $V_{DS}|_{load}$ can be reduced for a given load current by increasing $(W/L)_L$ or $(V_{GG} - V_{TL})$ at the expense of increasing the gate power dissipation.

(b) when $V_{IN} = V_1$, $V_1 > V_{TD}$, the driver is operating in the nonsaturation region, the load is nonsaturated and $I = I_0$ where

$$I_0 = K'\left(\frac{W}{L}\right)_L [(V_{GG} - V_{TL} - V_0)(V_{DD} - V_0) - \frac{1}{2}(V_{DD} - V_0)^2] \quad (12a)$$

$$= K'\left(\frac{W}{L}\right)_D [(V_1 - V_{TD})V_0 - \frac{1}{2}V_0^2] \quad (12b)$$

If $V_0 \ll V_{DD}$, $V_0 \ll V_1 - V_{TD}$, $V_{GG} - V_{TL} \gg V_{DD}$ then (12) can be simplified:

$$I_0 \simeq K'\left(\frac{W}{L}\right)_L (V_{GG} - V_{TL})V_{DD} \quad (13a)$$

$$\simeq K'\left(\frac{W}{L}\right)_D (V_1 - V_{TD})V_0 \quad (13b)$$