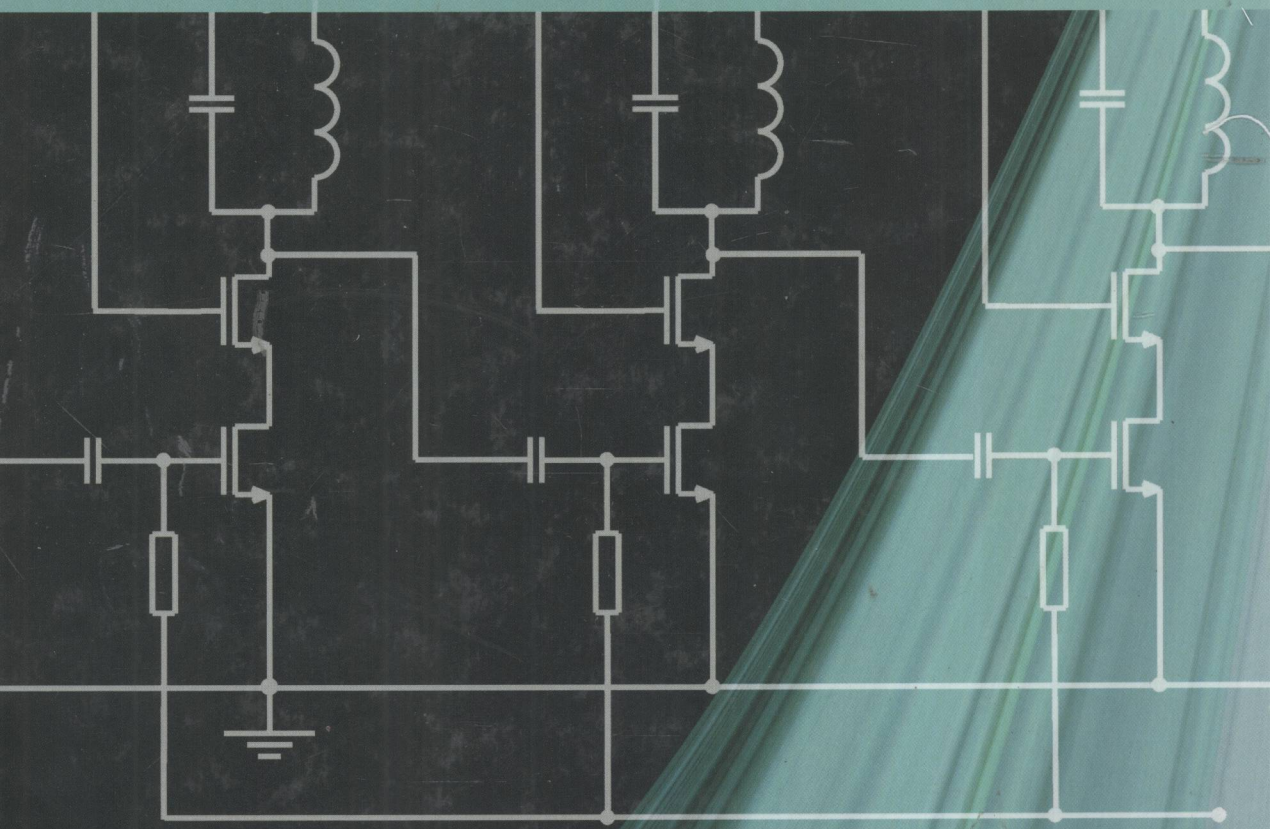
Fundamentals of High-Frequency CMOS Analog Integrated Circuits



Duran Leblebici
Yusuf Leblebici

CAMBRIDGE

Fundamentals of High-Frequency CMOS Analog Integrated Circuits

Duran Leblebici
Istanbul Technical University (ITU)

Yusuf Leblebici
Swiss Federal Institute of Technology in Lausanne (EPFL)

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521513401

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Leblebici, Duran.
Fundamentals of high-frequency CMOS analog integrated circuits / Duran Leblebici, Yusuf Leblebici.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-51340-1 (hardback)
1. Metal oxide semiconductors, Complementary. 2. Linear integrated circuits.
3. Analog electronic systems. 4. Very high speed integrated circuits. 5. Oscillators, Electric.
I. Leblebici, Yusuf. II. Title.
TK7871.99.M44L4335 2009
621.3815–dc22 2009007320

ISBN 978-0-521-51340-1 hardback

Fundamentals of High-Frequency CMOS Analog Integrated Circuits

With a design-centric approach, this textbook bridges the gap between fundamental analog electronic circuits textbooks and more advanced RF IC design texts. The structure and operation of the building blocks of high-frequency ICs are introduced in a systematic manner, with an emphasis on transistor-level operation, the influence of device characteristics and parasitic effects, and input–output behavior in the time and frequency domains.

Key features include:

- solved design examples to guide the reader through the decision process that accompanies each design task, with an emphasis on key trade-offs;
- coverage of the major issues that must be taken into account when combining analog and digital circuit building blocks;
- key criteria and parameters that are used to describe system-level performance;
- simple circuit models to enable a robust understanding of high-frequency design fundamentals;
- SPICE simulations that are used to check results and fine-tune the design.

This textbook is ideal for senior undergraduate and graduate courses in RF CMOS circuits, RF circuit design, and high-frequency analog circuit design. Analog integrated circuit designers and RF circuit designers in industry who need help making design choices will also find this a practical and valuable reference.

DURAN LEBLEBICI is Professor Emeritus of Electrical and Electronics Engineering at Istanbul Technical University (ITU). He has been teaching a range of undergraduate and graduate courses, from device electronics and fabrication technologies to integrated electronic circuits and RF IC design, for more than 40 years. He is the author of three textbooks in the field of electronics. He also established the first microelectronics laboratory and the first VLSI design house at ITU. He has received the Distinguished Service Award of the Turkish Scientific and Technological Research Council (TUBITAK) in 1992, in recognition of his services to microelectronics education.

YUSUF LEBLEBICI is Director and Chair Professor of the Microelectronic Systems Laboratory at the Swiss Federal Institute of Technology in Lausanne (EPFL). He has previously worked as a faculty member at the University of Illinois at Urbana-Champaign, at Istanbul Technical University, and at Worcester Polytechnic Institute (WPI), where he established and directed the VLSI Design Laboratory, and also served as a project director at the New England Center for Analog and Mixed-Signal IC Design. He is a co-author of more than 150 scientific articles and three textbooks.

To
Yıldız Leblebici – wife, mother, and colleague.

Preface

In the first half of the twentieth century the radio was the main activity area of the electronics industry and, correspondingly, RF circuits occupied a considerable part in the electronic engineering curriculum and books published in this period. Properties of resonance circuits and electronic circuits using them, single-tuned amplifiers as the input stages of receivers, double-tuned circuits as IF amplifiers, RF sinusoidal oscillators and high-power class-C amplifiers have been investigated in depth. It must be kept in mind that the upper limit of the radio frequencies of those days was several tens of MHz, the inductors used in tuned circuits were air-core or ferrite-core coils with inductance values in the micro-henries to milli-henries range, having considerably high quality factors, ranging from 100 to 1000, and the tuning capacitors were practically lossless.

The knowledge developed for the vacuum-tube circuits easily adapted to the transistors with some modifications related to the differences of the input and output resistances of the devices. In the meantime, the upper limit of the frequency increased to about 100 MHz for FM radio and to hundreds of MHz for UHF-TV. The values of the inductors used in these circuits correspondingly decreased to hundreds to tens of nano-henries. But these inductors were still wound, high-Q discrete components.

In the second half of the twentieth century, the emergence of integrated circuits drastically increased the reach of electronic engineering. Digital electronics on one side and analog electronics using the potentials of the operational amplifiers on the other side forced the curricula and the textbooks to skip certain old and "already known" subjects (among them resonance circuits and tuned amplifiers), to open room to these new subjects. The development of (inductorless) active filters that replaced the conventional passive LC filters extensively used in telecommunication systems even decreased the importance of inductors.

Rapid development of CMOS technology and the steady decrease of the dimensions of the devices according to "Moore's Law" led to an increase in the complexity of ICs (from now on, the VLSI circuits) and the operating frequencies as well. In the digital realm this helped to improve the performances of digital computers and digital telecommunication systems. In the analog realm the operation frequencies of the circuits increased to the GHz region and correspondingly the inductance values decreased to below 20 nH, so that now it was possible to realize them as on-chip components. Hence the freedom from external bulky discrete inductors opened a new

horizon towards small and light-weight mobile systems; the mobile telephone, GPS systems, Bluetooth, etc.

But there is a problem related to this development; the quality factors of on-chip inductors are very low, usually around 10–100, and tuning capacitors are not lossless any more. Most of the earlier theory (and design practice) that was developed with very high-Q discrete components does not easily translate to such integrated high-frequency circuits and on-chip components with less than ideal characteristics. Instead of relying on the comprehensive theory and analytical design of high-frequency analog circuits that was in place many decades earlier, most of the new-generation designers were tempted to adopt rather ad-hoc design strategies not grounded in sound theory, and to explain away the inevitable inconsistencies as "secondary effects". To make matters even more complicated, systematic treatment of subjects such as high-frequency circuit behavior and resonance/tuned circuits has been missing from the electrical engineering curricula for several decades, and analog designers entering the field of RF/high-frequency design had to re-learn these subjects.

One of the objectives of this book is to fill this gap; to introduce the fundamental aspects of high-frequency circuit operation, to systematically discuss the behavior of key components (in particular, submicron MOSFETs and on-chip passive components), summarize the behavior of the series and parallel resonance circuits in detail and investigate the effects of the losses of on-chip inductors and capacitors that are usually not taken into account in formulas derived for high-Q resonance circuits.

Since all these circuits are being developed and used usually in the GHz range, sometimes close to the physical limits of the devices, it becomes necessary to recapitulate the behavior of MOS transistors together with their important parasitics and the frequency related secondary effects.[1] In Chapter 1 the basic current–voltage relations of MOS transistors are derived, taking into account the parabolic (not linear) shape of the inversion charge profile that leads to a different approach to understand the channel shortening effect and the gate-source capacitance of the transistor. The velocity saturation effect and the behavior of a MOS transistor under a velocity saturation regime are also investigated as an important issue for small geometry devices. Although it is not used extensively for HF applications, the sub-threshold regime is also investigated in brief.

In addition to the intrinsic behavior of MOS transistors the parasitics – that are inevitable and have severe effects on the overall behavior especially at high frequencies – have been discussed, mostly in connection with the BSIM3 parameters. The properties, limits and parasitics of the passive on-chip components, namely resistors, fixed value and variable capacitors (varactors) and inductors are also summarized in this chapter from a realistic and design-oriented point of view.

The subject of Chapter 2 is the DC properties of the basic analog MOS circuits that will be investigated in the following chapters. The interactive use of analytic

[1] This part of the book must not be considered as an alternative to the existing sound and comprehensive models such as EKV and others, but rather as an attempt to explain the behavior of MOS transistors, based on the basic laws of electrostatics and circuit theory that all electronics students already know.

expressions – that provide interpretable knowledge about the basic behavior of the circuit – and SPICE simulations – that give the designer a possibility to "experiment", to fine-tune and optimize the circuit, all secondary and parasitic effects included – are exemplified throughout the chapter. It is believed that to have the ability of using together the analytical expressions and the power of SPICE is a "must" for an analog designer.

In Chapter 3 the frequency-dependent behaviors of the basic circuits are given, not limited to the frequency characteristics of the gain but including the input and output impedances. The important properties of them, usually not dealt with in books, are investigated and their effects on the performance of wide-band circuits are underlined. The basics of the techniques used to enhance the gain; the additive approach (distributed amplifiers) and the cascading strategies to reach to the wide-band amplifiers (not only voltage amplifiers, but also current amplifiers, transadmittance amplifiers and transimpedance amplifiers) are systematically investigated.

In Chapter 4 first the resonance circuits are recapitulated with this approach and the behavioral differences of high-Q and low-Q resonance circuits are underlined. Afterwards, tuned amplifiers are systematically investigated taking into account the low-Q effects, not only for single-tuned amplifiers but also for double-tuned and staggered tuned amplifiers that are not covered in many new (and even older) books in detail. The LNA, that is one of the most important classes of tuned amplifiers, is also investigated in this chapter together with the noise behavior of MOS transistors, that is developed with a different approach.

LC sinusoidal oscillators are discussed in Chapter 5 with the negative resistance approach and the classical positive feedback approach as well, with emphasis on the effects of the low-Q components. The problems related to the frequency stability of LC oscillators are discussed and the phase noise in LC oscillators is investigated with a different approach.

The last chapter is devoted to a summary of the higher-level system view of HF analog circuits, especially in the context that virtually all such high-frequency circuits are eventually integrated with considerable digital circuitry for interface, post-processing and calibration purposes – and that such integration is increasingly done on the same silicon substrate. The traditional system-level view of high-frequency components and circuits is strongly influenced by conventional (all-analog) modulation and transmission systems modeling, which is based almost exclusively on the frequency domain. The behavior of all digital systems, on the other hand, is preferably described in the time domain. While the translation between these two domains is (in theory) quite straightforward, the designers must develop a sense of how some of their choices in the analog realm eventually influence the behavior of the digital part, and vice versa. Data converters (analog-to-digital and digital-to-analog converters) naturally play an important role in this translation between domains, and Chapter 6 attempts to summarize the key criteria and parameters that are used to describe system-level performance.

The target audience of this book includes advanced undergraduate and graduate-level students who choose analog/mixed-signal microelectronics as their area of

specialization, as well as practicing design engineers. The required background that is needed to follow the material is consistent with the typical physics, math and circuits background that is acquired by the third (junior) year of a regular electrical and computer engineering (ECE) curriculum.

Solved design examples are provided to guide the reader through the decision process that accompanies each design task, emphasizing key trade-offs and eventual approximations.

A number of individuals have contributed with their time and their efforts to the creation of this textbook. In particular, both authors would like to thank Mrs Yildiz Leblebici who read the entire manuscript, carefully checked the analytical derivations throughout all chapters, and provided valuable insight as an experienced electronics teacher. The authors also acknowledge the generous support of Mr Giovanni Chiappano from austriamicrosystems A.G.[2] for offering the use of transistor parameters in numerous examples.

The idea of this book was originally launched with the enthusiastic encouragement of Dr Philip Meyler of Cambridge University Press, who saw the need for a design-oriented text in this field and patiently followed through its early development. We are deeply grateful to Dr Julie Lancashire, our publisher, for her guidance, support and encouragement over the years, leading up to the final stages of production. The editorial staff of Cambridge University Press has been wonderfully supportive throughout this project. We would especially like to thank Ms Sarah Matthews and Mr Christopher Miller for their valuable assistance.

Last but not least, the authors would also like to thank all reviewers who read all or parts of the manuscript and provided very valuable comments.

[2] Device characteristics and technology parameters provided by austriamicrosystems A.G. will be labeled with the acronym "AMS" throught the text.

Contents

1 Components of analog CMOS ICs

1.1 MOS transistors

The basic structure of an n-channel metal–oxide–semiconductor (NMOS) transistor built on a p-type substrate is shown in Fig. 1.1. The MOS transistor consists of two disjoint p-n junctions (source and drain), bridged by a MOS capacitor composed of the thin gate oxide and the poly-silicon gate electrode. If a positive voltage with sufficiently large magnitude is applied to the gate electrode, the resulting vertical electric field between the substrate and the gate attracts the negatively charged electrons to the surface. Once the electron concentration on the surface exceeds the majority hole concentration of the p-type substrate, the surface (the channel) is said to be inverted, i.e., a conducting channel is formed between the source and the drain. The carriers, i.e., the electrons in an NMOS transistor, enter the channel region underneath the gate through the source contact, leave the channel region through the drain contact, and their movement in the channel region is subject to the control of the gate voltage.

To ensure that both p-n junctions are continuously reverse biased, the substrate potential is kept lower than the source and drain terminal potentials. Note that the device structure is symmetrical with respect to the drain and source regions; the different roles of these two regions are defined only in conjunction with the applied terminal voltages and the direction of the drain current. In an n-channel MOS (NMOS) transistor, the source is defined as the n+ region which has a lower potential than the other n+ region, the drain. This means that the current flow direction is from the drain to the source. By convention, all terminal voltages of the device are defined with respect to the source potential.

The value of the gate-to-source voltage (V_{GS}) necessary to cause surface inversion (to create the conducting channel) is called the threshold voltage V_T. This quantity depends on various device and process parameters such as the work function difference between the gate and the substrate, the substrate (surface) Fermi potential, the depletion region charge concentration, the interface charge concentration, the gate oxide thickness and oxide (dielectric) permittivity, as well as the concentration of the channel implantation that is used to adjust the threshold voltage level.

If the applied gate-to-source voltage exceeds the threshold voltage of the MOS transistor, a sufficiently high concentration of electrons is achieved in the channel region, leading to surface inversion. Thus, an n-type conducting channel is formed between the source and the drain, which is capable of carrying the drain (channel) current. If a small positive voltage is applied to the drain, a current proportional to this voltage will start to flow from the drain to the source through the conducting channel. The effective

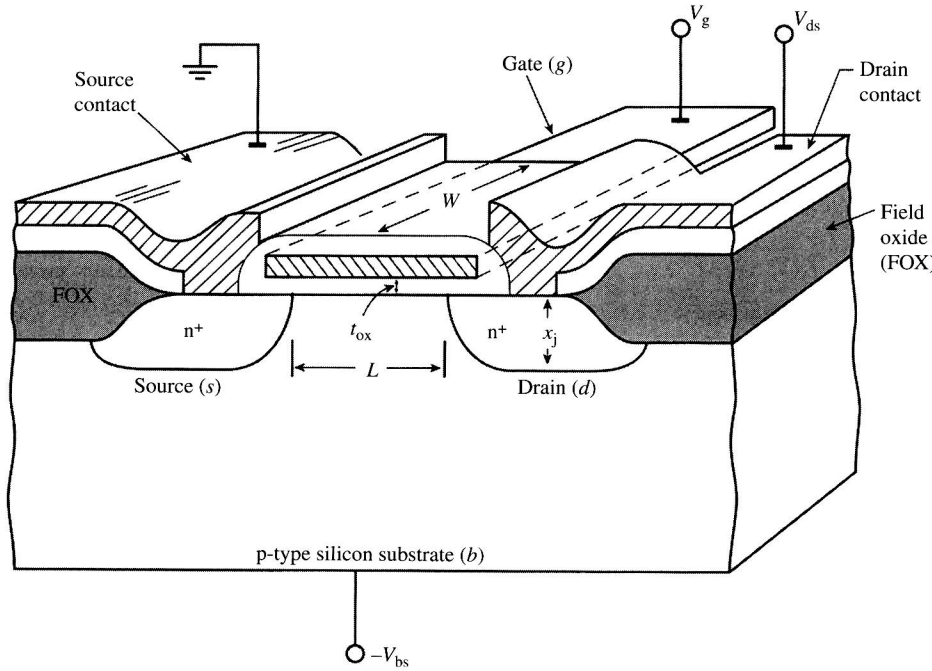


Figure 1.1 Simplified cross-section view of an n-channel MOS (NMOS) transistor (after Taur and Ning [13]).

resistivity of the continuous inversion layer between the source and the drain depends on the gate voltage. This operating mode is called the linear (or triode) mode, where the channel region acts as a voltage-controlled resistor. During this operating mode, the electron velocity in the channel is usually much lower than the drift velocity limit.

As the applied drain voltage is increased, the inversion layer charge and the channel depth at the drain end start to decrease. Eventually, when the drain voltage reaches a limit value called the saturation voltage ($V_{D(sat)}$), the inversion charge at the drain is reduced – theoretically – to zero, and the velocity of electrons – theoretically – reaches very high values, as discussed in the following sections. This event is named as the "pinch-off" of the channel. Beyond the pinch-off point, i.e., for drain voltage values larger than the saturation voltage, electrons travel in a very shallow pinched-off channel with a very high velocity, which is called the "saturation velocity". This operating regime is known as the saturation mode.

If the transistor is formed on an n-type substrate, using two p+ regions as source and drain, this structure is called a p-channel MOS (PMOS) transistor. In a PMOS transistor, the fundamental mechanisms of surface inversion and channel conduction are exactly the same as in NMOS transistors, although the majority of carriers consist of holes, not electrons. Thus, the gate-to-source voltage applied to the gate electrode to achieve surface inversion must be negative. Also, it should be taken into account that the hole mobility is considerably smaller than the electron mobility at room

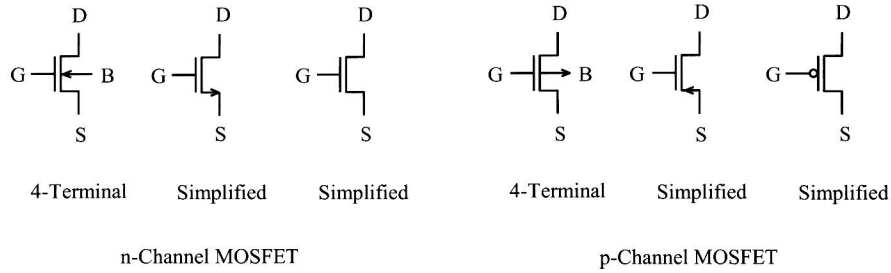


Figure 1.2 Commonly used circuit symbols for NMOS and PMOS transistors.

temperature, which leads to a smaller effective channel conductance for the PMOS transistor with the same channel dimensions. Nevertheless, the complementary nature of NMOS/PMOS biasing and operating conditions offers very useful circuit implementation possibilities, which underlines the importance and the wide-spread use of complementary MOS (CMOS) circuits in a very large range of applications.

Commonly used circuit symbols for n-channel and p-channel MOS transistors are shown in Fig. 1.2. While the four-terminal representation shows all external terminals of the device, the three-terminal symbol is usually preferred for simplicity. Unless noted otherwise, the substrate terminals are always assumed to be connected to the lowest potential for NMOS devices, and to the highest potential for PMOS devices.

1.1.1 Current–voltage relations of MOS transistors

The basic (so-called Level-1) current–voltage relations of a MOS transistor are given in most basic electronics textbooks. Since these relations contain a small number of parameters, they are convenient for hand calculations. The parameters of these expressions are:

- the mobility of electrons (or holes), μ;
- the gate capacitance per unit area, C_{ox};
- the threshold voltage of the transistor, V_T;
- the gate-length modulation coefficient, λ;
- and the aspect ratio of the transistor, (W/L).

In the following, these relations are derived with a different approach, to remind the reader of the fundamentals, and also to clarify the understanding of device behavior. In addition, the derivation presented here is based on a realistic profile of the channel-region inversion charge (as calculated from the fundamental electric field expressions), as opposed to the classical gradual channel approach which assumes linear charge profiles in the channel. This model represents the transistor under moderate to strong inversion conditions with reasonable accuracy for hand calculations, provided that the channel length is not too short and the transistor is not in the velocity saturation region.

For short-channel MOS transistors in which the carrier velocities reach saturation, i.e. approach a limit velocity, this model is no longer valid. Since a great majority of transistors realized in analog MOS integrated circuits today have channel lengths in

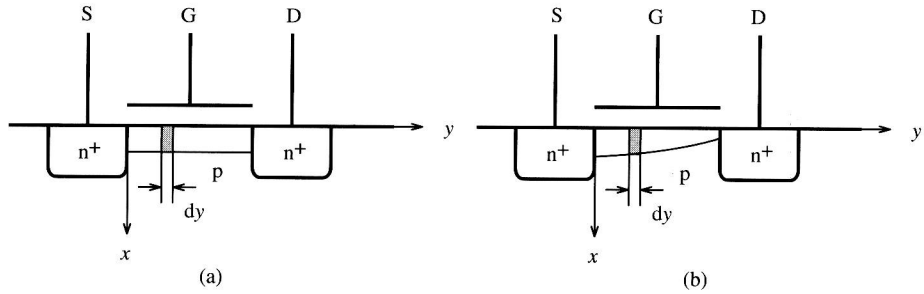


Figure 1.3 Inversion channel profiles of an NMOS transistor for (a) $V_{GS} > V_T$, $V_{DS} = 0$, (b) $V_{GS} > V_T$, $V_{DS} > 0$.

the sub-half-micron range, they may easily enter the velocity saturation region. Therefore it is necessary to derive rules to check whether a transistor is operating in the velocity saturation region or not, and to obtain expressions that are valid for velocity saturated transistors.

1.1.1.1 The basic current–voltage relations without velocity saturation

The cross-section of an NMOS transistor having an inversion layer along the channel owing to a gate–source voltage greater than the threshold voltage and zero source–drain voltage is shown in Fig. 1.3(a). Since there is no surface inversion for gate voltages smaller than the threshold voltage V_T, the value of the inversion charge density is

$$Q_i = -C_{ox}(V_{GS} - V_T)\,[\text{coulomb/cm}^2]$$

where $(V_{GS} - V_T)$ is the "effective gate voltage" for this case. The amount of the inversion charge of a transistor having a channel length L and a channel width W is

$$\bar{Q}_i = -C_{ox}WL(V_{GS} - V_T)$$

The minus signs in front of these expressions denote that this is a negative charge, since the carriers in the inversion layer of an NMOS transistor are electrons.

When we apply a positive drain voltage with respect to the source, a drain current (I_D) flows in the $-y$ direction and is constant along the channel (Fig. 1.3(b)). However, owing to the voltage drop on the channel resistance, the voltage along the channel is not constant. Although the effective gate voltage – inducing the inversion charge – at the source end of the channel is $(V_{GS} - V_T)$, it decreases along the channel and becomes equal to $(V_{GD} - V_T) = (V_{GS} - V_{DS} - V_T)$ at the drain end. If the channel voltage with respect to the source is denoted by $V_c(y)$, the effective gate voltage as a function of y can be written as

$$V_{eff}(y) = (V_{GS} - V_T) - V_c(y) \tag{1.1}$$

and the amount of the inversion charge in an infinitesimal channel segment dy is

$$d\bar{Q}_i(y) = -C_{ox}W[(V_{GS} - V_T) - V_c(y)]dy \tag{1.2}$$

Since the drain current is constant along the channel, for any y position the current can be expressed as

$$I_D = \frac{d\bar{Q}_i(y)}{dt} = \frac{d\bar{Q}_i(y)}{dy/v(y)} \tag{1.3}$$

where $v(y)$ is the velocity of electrons at position y, and can be expressed in terms of the electron mobility and the electric field strength at y:

$$v(y) = \mu_n E(y) = -\mu_n \frac{dV_c(y)}{dy} \tag{1.4a}$$

Using (1.2), (1.3) and (1.4a),

$$I_D = \mu_n C_{ox} W[(V_{GS} - V_T) - V_c(y)] \frac{dV_c(y)}{dy}$$

which gives the electric field strength as

$$E(y) = \frac{dV_c(y)}{dy} = \frac{I_D}{\mu_n C_{ox} W[(V_{GS} - V_T) - V_c(y)]} \tag{1.4b}$$

and

$$\frac{I_D}{\mu_n C_{ox} W} dy = [(V_{GS} - V_T) - V_c(y)] dV_c(y)$$

After integration from the source end ($y = 0$) to y we obtain

$$\frac{I_D}{\mu_n C_{ox} W} y = (V_{GS} - V_T) V_c(y) - \frac{1}{2} V_c^2(y) \tag{1.5}$$

From (1.5), the channel voltage $V_c(y)$ corresponding to a certain gate voltage and drain current can be deduced as

$$V_c(y) = (V_{GS} - V_T) \mp \sqrt{(V_{GS} - V_T)^2 - \frac{2I_D}{\mu_n C_{ox} W} y} \tag{1.6}$$

To fulfill the obvious physical condition $V_c(0) = 0$, the sign before the square-root term has to be taken as minus. The effective gate voltage is found from (1.1) and (1.6):

$$\begin{aligned} V_{eff}(y) &= (V_{GS} - V_T) - V_c(y) \\ &= (V_{GS} - V_T) \sqrt{1 - \frac{2I_D}{\mu_n C_{ox} W (V_{GS} - V_T)^2} y} \end{aligned} \tag{1.7}$$

It is useful to interpret (1.7) for certain cases.

(a) For $y = 0$ (at the source end of the channel) the effective channel voltage is $V_{eff}(0) = (V_{GS} - V_T)$, as expected.
(b) If $V_c(L) = V_{DS} = (V_{GS} - V_T)$, the effective channel voltage at $y = L$ is equal to zero and the channel is pinched-off at the drain end of the channel. For this case