EXHIBIT 15

MICROELECTRONIC DEVICES

2nd Edition

Keith Leaver
Imperial College of Science, Technology and Medicine

ICP
Imperial College Press

Published by

Imperial College Press
516 Sherfield Building
Imperial College
London SW7 2AZ

Distributed by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 912805
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Cover design: Courtesy of Mr Peter Cheung

First edition was published by Longman Group UK Limited in 1989.

MICROELECTRONIC DEVICES (2nd Edition)

ISBN 1-86094-013-7
ISBN 1-86094-020-X (pbk)

Printed in the United Kingdom.

Operating regions of the characteristics

The region of Fig. 4.7(a) to the left of the dashed line, in which eqn. (4.4) remains valid, is often known as the TRIODE REGION of the characteristics. In this book we shall refer to it for clarity as the UNSATURATED REGION, to distinguish it clearly from the SATURATION REGION, where the drain current does not increase with V_{DS}. The initial portion of the the unsaturated characteristic, where $V_{DS} \ll V_{GS} - V_T$, is sometimes called the LINEAR region of operation.

A third principal region of the characteristics, in which the transistor is OFF, (i.e. no drain current flows), exists for all gate voltages below threshold. This can be seen in Fig. 4.7(b), which shows how the drain current I_D changes with gate-source voltage V_{GS} when V_{DS} is held constant at a small value. One practical point to note about this diagram is that the threshold of conduction is not as sharp as the above discussion would imply. The threshold voltage is normally defined for a real device by extrapolating the linear part of this characteristic back to the voltage axis. The reason for the curvature near this point is covered in Panel 4.2, which should be read following the discussion of the saturation region of the characteristics in the next section.

4.4 The saturation region of the output characteristic

When $V_{DS} \geq V_{GS} - V_T$ the transistor is said to be SATURATED, meaning that the drain current no longer rises strongly with V_{DS}. At the onset of saturation, the potential relative to the source varies along the channel, as shown in Fig. 4.8(a), reaching the value $(V_{GS} - V_T)$ at the drain. At this point, the voltage between gate and drain, V_{GD}, just equals the threshold voltage V_T. The channel is said to be PINCHED OFF at the drain, and the drain-source voltage V_{DS} at which this occurs is often referred to as the pinch-off voltage. Note, however, that the pinch-off voltage does not have a fixed value, as it is equal to $(V_{GS} - V_T)$.

Any further increase in V_{DS} ensures that the channel becomes fully depleted of carriers at the end nearest the drain. Thus, an extremely short depletion region forms between the channel itself and the drain, as illustrated in Fig. 4.8(b). The electrons in the channel must flow across this depleted region in order to reach the drain. They can do this with the help of even

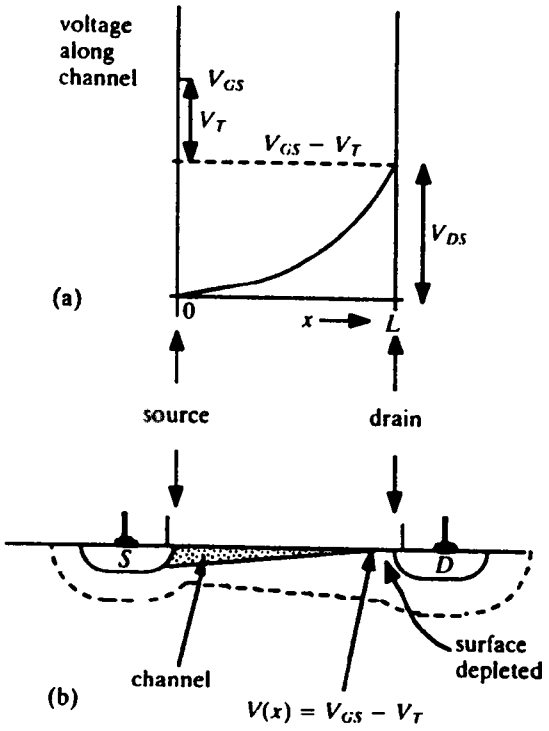

Fig. 4.8 (a) Voltage variation along the channel in saturation (b) a depletion region of length L_D is formed in saturation at the drain end of the channel.

a small voltage difference, because the distance involved is so small that a large electric field strength is easily created. At the point where the conducting channel ends (is 'pinched-off'), the potential is equal to the pinch-off voltage $(V_{GS} - V_T)$. Note that the difference between the voltage V_{DS} and the pinch-off value of $(V_{GS} - V_T)$ appears across the short, high-resistance depletion layer, and is enough to give the few electrons present there a high velocity towards the drain. Indeed, they travel at their maximum speed, the *saturation velocity*.

We expect the current to rise no further, because an increase in V_{DS} no longer raises the velocity of carriers in the conducting channel itself, but merely lengthens the depleted region slightly. The drain current thus saturates, and the characteristics are shown horizontal on the theoretical plots in Fig. 4.7(a). The drain current in saturation is simply found by putting $V_{DS} = (V_{GS} - V_T)$ into eqn. (4.4), thus

$$I_D = \frac{1}{2}\mu_e C_{ox}\frac{W}{L}(V_{GS} - V_T)^2 \quad \text{in saturation} \tag{4.5}$$

However, constancy of the saturation current depends upon maintaining a constant channel length. In practice, the depletion layer at the drain end lengthens as the voltage across it increases. The channel length L decreases slightly in consequence, and the drain current rises a little. This is known as the EARLY effect, after the first person to describe the rather similar behaviour observed in bipolar transistors. It is also referred to as CHANNEL LENGTH MODULATION, and it will be explained at greater length in section 4.6. Figure 4.9 shows some transistor characteristics where the effect is clearly seen.

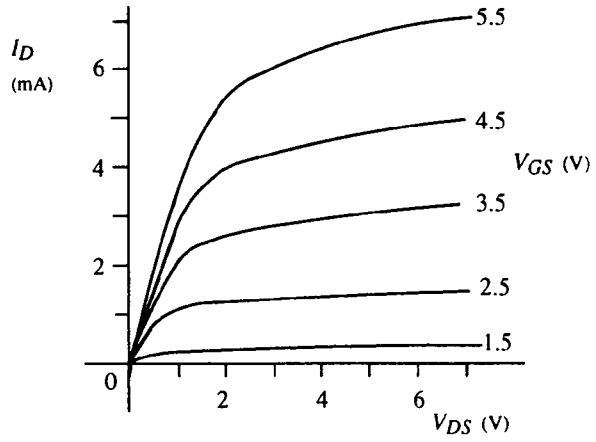

Fig. 4.9 Transistor characteristics showing the effect of channel length modulation in the saturation region.

We can use eqn. (4.5) to find the boundary between the unsaturated and saturation regions on Fig. 4.7(a) by naming the values of I_D and V_{DS} at the onset of saturation as I_{DSat} and V_{DSat} respectively. Then, since $V_{DSat} = V_{GS} - V_T$, eqn. (4.5) gives

$$I_{DSat} = \frac{1}{2}\mu_e C_{ox}\frac{W}{L}V_{DSat}^2$$

This is the equation of a parabola, which is shown dashed in Fig. 4.7(a).

It is worth noting that the term saturation has a quite different meaning in the case of the bipolar transistor, as will be explained in chapter 5.

PANEL 4.2

The pinched-off channel: subthreshold conduction

When $V_{DS} \geq V_{GS} - V_T$, i.e. when the *drain* end of the channel is pinched off, the carrier concentration there is not zero but merely small in magnitude, just as in the depletion region of a *p–n* junction. The gate voltage is unable to maintain *strong inversion* at the drain end of the channel; we say that it supports *weak inversion*. Similarly, weak inversion occurs at the source if we make V_{GS} slightly less than V_T: the transistor is off and, in the normal way, we would then say that the transistor is *below threshold*.

The term 'weak inversion' implies that the channel region *does* contain some free electrons. This is the so-called sub-threshold region of operation, in which the electron concentration in the channel is sufficient only to allow very small currents to flow. In this regime, the following approximate expression for the drain current I_D applies, provided that V_{SB} and V_{DB} are not too close to the pinch-off voltage $V_{GS} - V_T$, which is abbreviated to the symbol V_p:

$$I_D = 2\mu C_{ox} \frac{W}{L}\left(\frac{kT}{e}\right)^2 \left\{\exp\left(\frac{V_p - V_{SB}}{kT/e}\right) - \exp\left(\frac{V_p - V_{DB}}{kT/e}\right)\right\}$$

Here the source and drain voltages are expressed relative to the substrate, or bulk silicon.

More accurate (but still relatively simple) equations for I_D are available for use when V_{SB} and V_{DB} are close to V_p.

The sub-threshold region is important commercially for two reasons:

(i) it determines the 'off' current of a transistor in many digital circuits, and hence the quiescent current drawn by each of the millions of switches used in a complete CMOS integrated circuit,
and

(ii) it can be used instead of the unsaturated region for low-power applications (e.g. in watches, satellites and body implants), and is likely to become more important due to a growing appreciation by more engineers of the circuit design techniques needed for its use.

4.5 The small-signal equivalent circuit

As with any electronic device, an equivalent circuit can be constructed to model the way small changes in input voltage affect both the input current and the output voltages and currents. It is used to represent the behaviour of the transistor when it is employed in a circuit designed to amplify small currents or voltages at the input terminals. The simplest equivalent circuit for a MOSFET is shown in Fig. 4.10(a), alongside the symbol of the transistor which it models. Note the arrow pointing from the *p*-type substrate to the *n*-channel, which identifies the channel type.

The convention followed in chapter 3 for the use of small letters with small subscripts for small-signal voltages and currents is continued here. The circuit of Fig. 4.10(a) is used to model the transistor in the connection shown in Fig. 4.10(b). This is known as the common-source connection, since the source terminal is common to both the input and output.

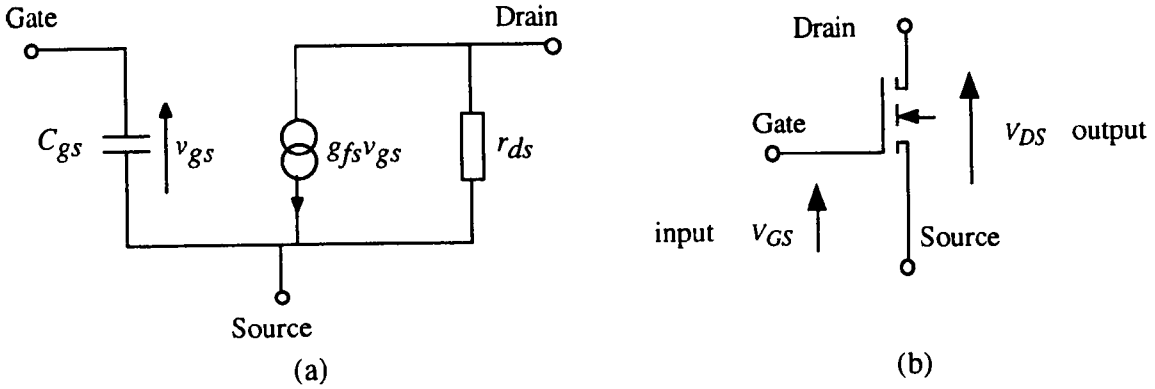

Fig. 4.10 (a) Small-signal equivalent circuit of a MOSFET in the common-source connection (b) Circuit symbol of the *n*-channel MOSFET, shown in the common-source connection.

It is necessary to treat the unsaturated and saturation operating regions of the transistor separately, as in earlier sections. Within each region of operation, the component values depend upon the bias voltages, i.e. upon the dc components of the voltages V_{GS} and V_{DS}.

When used in an amplifier, a MOSFET is usually biased in its saturation region, so we shall concentrate more upon the saturation region in what follows.

Let us first consider how to define each of the component values in Fig. 4.10(a).

The gate-source input capacitance C_{gs}

A MOSFET has almost infinite input resistance, owing to the excellent insulating properties of the gate oxide. Hence, we normally need only to take account of the input capacitance.

The input capacitance can be defined as the small-signal charge q_g on the gate electrode divided by the small-signal gate-source voltage v_{gs}. Thus

$$C_{gs} = \frac{q_g}{v_{gs}} = \frac{\partial Q_G}{\partial V_{GS}}$$

The transconductance g_{fs}

The dependence of the small-signal drain current i_d on small-signal gate-source voltage v_{gs} is modelled by a current generator of strength $g_{fs}v_{gs}$ in Fig. 4.10, where g_{fs} is given by

$$g_{fs} = \frac{i_d}{v_{gs}} = \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}=\text{const}}$$

g_{fs} is referred to as the forward TRANSCONDUCTANCE* of the transistor in the common-source connection. It has the dimensions of a conductance, but is unlike an ordinary conductance in that it connects an *output* current with an *input* voltage. The two subscripts *f* and *s* are used to distinguish this transconductance from others which are sometimes needed, for example the reverse transconductance g_{rs}, used when the drain and source are interchanged, and similar transconductances defined for cases when the terminal which is common to both input and output, is not the source (denoted by the second subscript *s* to the symbol), but the gate or drain.

*The symbol g_m is often used in place of g_{fs}, which is the internationally agreed symbol for this component.

Note from the equation that in evaluating the transconductance, V_{DS} is to be held constant, i.e. $v_{ds} = 0$. Otherwise, some of the small-signal current generated would flow in the resistance r_{ds} in the figure. If we are to measure the strength of the current source in Fig. 4.10 by measuring the external drain current, it must be measured in a short-circuit across the terminals.

The output resistance r_{ds}

The resistance r_{ds} in Fig. 4.10 is commonly known as the OUTPUT RESISTANCE of the transistor, since the drain and source are the output terminals of the amplifier. It models that part of the change in drain-source voltage V_{DS} which is proportional to I_D. Thus the definition of r_{ds} is

$$r_{ds} = \frac{v_{ds}}{i_d} = \left(\frac{\partial V_{DS}}{\partial I_D} \right)_{V_{GS} = \text{const}}$$

We shall use these definitions to find expressions for the component values in the unsaturated region of operation, since this is the easiest case to consider. The saturation region requires a little more development of the ideas introduced in earlier sections.

4.6 Component values in the equivalent circuit: the unsaturated region of operation

The gate-source input capacitance

The two input terminals, the gate and the source, present a very high d.c. input resistance, typically above $10^{12}\ \Omega$ and a small capacitance C_{gs} in the range 0.01–10 pF, which is shown in Fig. 4.10. The input resistance is sufficiently high that for practical purposes it is commonly omitted, as here. Hence at all useful frequencies above d.c., the input impedance is

$$Z_{in} = (j\omega C_{gs})^{-1}$$

Since C_{gs} is small, the input impedance has a very large magnitude, and indeed may be assumed infinite for practical purposes, except at high frequencies, in all regions of operation. We shall discuss what determines the value of C_{gs} later, along with other capacitances which play a role at high frequencies.

Transconductance in the unsaturated region of operation

The definition of g_{fs} above is easily used in conjunction with eqn. (4.4) for I_D to find an expression for the transconductance in the unsaturated region. Differentiating eqn. (4.4) gives the result

$$g_{fs} = \mu_e C_{ox} \left(\frac{W}{L}\right) V_{DS} \quad \text{(unsaturated)}$$

Output resistance in the unsaturated region

By differentiating eqn. (4.4) with respect to V_{DS} it is easy to show that

$$r_{ds} = \mu_e C_{ox} \left(\frac{W}{L}\right) [V_{GS} - V_T - V_{DS}] \quad \text{(unsaturated)}$$

4.7 Component values in the equivalent circuit: the saturation region of operation

Input capacitance

The d.c. input resistance remains effectively infinite.

The input impedance is identical to that given for the unsaturated region, thus:

$$Z_{in} = (j\omega C_{gs})^{-1}$$

although the value of C_{gs} differs somewhat from that in the unsaturated region, as will be discussed later.

Transconductance in the saturation region

To evaluate the transconductance in the MOSFET's saturation region, we differentiate eqn. (4.5) with respect to V_{GS}, and find that

$$g_{fs} = \mu_e C_{ox} \frac{W}{L}(V_{GS} - V_T) \quad \text{(in saturation)} \tag{4.6}$$

Here, V_{GS} denotes the mean gate-source voltage difference (i.e. the d.c. or bias component). Unless the input voltage v_{gs} is small compared to V_{GS}, the transconductance g_{fs} varies with signal amplitude, thereby creating a distortion of the collector current waveform. Note that the designer can vary g_{fs} at will by choosing the ratio W/L.

To estimate a typical value of g_{fs}, we first estimate C_{ox}. The relative permittivity ε_{ox} of SiO_2 is close to 4.0. Thus, assuming an oxide thickness t_{ox} of 0.05 μm, we have

$$C_{ox} = \frac{\varepsilon_{ox}\varepsilon_o}{t_{ox}} = \frac{4 \times 8.8 \times 10^{-12}}{5 \times 10^{-8}} = 7.0 \times 10^{-4} \ \text{pF}/\mu\text{m}^2 .$$

The next point to note is that the mobility μ_e of electrons is about one half of the value used in Chapter 1 appropriate to the doping concentration in the channel, because of extra collisions of the electrons with the surface of the Si at the interface with the SiO_2. The channel thickness is comparable to the mean free path, so that such collisions are relatively frequent. For example, when the doping is low, μ_e is reduced to about 0.07 m^2/Vs. Hence when $W/L = 10$, we have typically

$$\begin{aligned} g_{fs} &= 0.07 \times 7.0 \times 10^{-4} \times 10(V_{GS} - V_T) \quad \text{(in saturation)} \\ &= 0.28 \ \text{mS for every volt by which } V_{GS} \text{ exceeds } V_T . \end{aligned}$$

When compared with the transconductance of a bipolar transistor evaluated in Chapter 5, this is a rather low value when voltages of just a few volts are used.

Output Resistance in the saturation region (The Early effect, or channel length modulation)

The output resistance cannot be calculated directly from eqn. (4.5) for the drain current, since V_{DS} does not appear explicitly there. However, as already explained, the drain current does increase slightly with V_{DS}.

As we have seen, when the gate-drain voltage V_{GD} falls below the value of the threshold voltage V_T, the drain current saturates because the channel at the drain end becomes depleted of carriers. As the drain voltage rises further, with V_{GS} constant, the point in the channel at which pinch-off begins moves slightly away from the drain, as shown in Fig. 4.11.

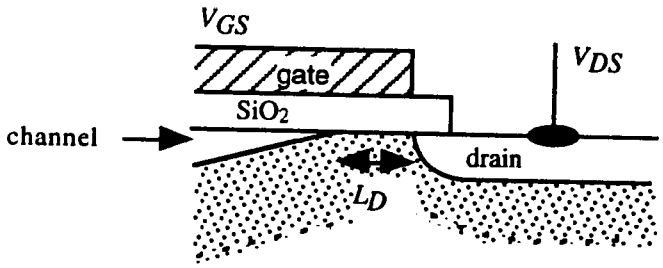

Fig. 4.11 An enlarged cross-sectional view of the channel at the drain end where, in saturation, the channel is pinched-off (i.e. depleted) over a length L_D.

This reduces the effective channel length of the depleted region by the length L_D, so that the drain current may be found by modifying eqn. (4.5) as follows

$$I_D = \frac{1}{2}\frac{\mu C_{ox} W}{(L - L_D)}(V_{GS} - V_T)^2 \qquad (4.7)$$

As L_D rises with increasing V_{DS}, I_D also rises slightly, giving the transistor its finite output resistance r_{ds}. The dependence of L_D upon V_{DS} is not easily calculated, however, and approximations are usually used.

The simplest model of this effect, which is provided within the SPICE computer modelling package, assumes that $L_D = \Lambda V_{DS}$, where Λ is an adjustable parameter chosen by the user, independent of transistor voltages or currents. The corresponding value of r_{ds}, the small-signal output resistance

is then simply found by differentiating eqn. (4.7) to give

$$r_{ds} = \left(\frac{\partial I_D}{\partial V_{DS}} \right)^{-1} = \frac{(L - L_D)}{\Lambda I_D} = \frac{V_A}{I_D}$$

where $V_A = (L - L_D)/\Lambda$ is called the EARLY VOLTAGE. This parameter is approximately constant whenever $L \gg L_D$, i.e. in all except transistors with very short channels. Although there is no simple justification for the assumption that $L_D = \Lambda V_{DS}$, the simple expression for r_{ds} which results does indeed describe the behaviour of real transistors fairly well.

Since the voltage across the depleted channel length L_D is $V_{DS} - V_p$ (where V_p is the pinch-off voltage, equal to $V_{GS} - V_T$), then by applying conventional abrupt p–n junction theory, L_D might be expected to rise in proportion to $(V_{DS} - V_p)^{1/2}$. This is not often the case in practice, since the conventional theory applies only to the doping profile of a planar, abrupt junction, whereas the doping here usually varies with depth as well as distance along the channel. Most versions of SPICE usually provide the option of using this, more complicated, model for the Early effect, in spite of its limitations.

4.8 Small-signal capacitances and transient behaviour

Before discussing the small-signal circuit components which model the high-frequency behaviour of a MOSFET, it is helpful to consider the factors which limit the rate at which the drain current responds to a change in the gate-source voltage. The most fundamental of these is the so-called transit time.

The transit time

The maximum rate at which the drain current I_D can be switched by a change in input voltage is determined by the rate at which the charge in the channel is changed, because it is this charge which carries the current.

Suppose an n-channel enhancement MOSFET is initially its unsaturated (triode) region, with V_{DS} held constant. If V_{GS} is increased instantly, the