Introduction to MOSFETs

Lecture notes by:

Mark Lundstrom
Purdue University
March 2015

The four lectures included here are taken from a set of lecture notes on nanotransistors that will be published by World Scientific Publishing Company.

Lecture 2

The Transistor as a Black Box

2.1 Introduction

The goal for these lectures is to relate the internal physics of a transistor to its terminal characteristics; i.e. to the currents that flow through the external leads in response to the voltages applied to those leads. This lecture will define the external characteristics that subsequent lectures will explain in terms of the underlying physics. We'll treat a transistor as an engineer's "black box," as shown in Fig. 2.1. A large current flows through terminals 1 and 2, and this current is controlled by the voltage on (or, for some transistors) the current injected into terminal 3. Often there is a fourth terminal too. There are many kinds of transistors [1], but all transistors have three or four external leads like the generic one sketched in Fig. 2.1. The names given to the various terminals depends on the type of transistor. The IV characteristics describe the current into each lead in terms of the voltages applied to all of the leads.

Before we describe the IV characteristics, we'll begin with a quick look at the most common transistor – the field-effect transistor (FET). In these lectures, our focus is on a specific type of FET, the silicon Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET). A different type

of FET, the High Electron Mobility Transistor (HEMT), find use in radio frequency (RF) applications. Bipolar junction transistors (BJTs) and heterojunction bipolar transistor (HBTs) are also used for RF applications. Most of the transistors manufactured and used today are one of these four types of transistors. Although our focus in on the Si MOSFET, the basic principles apply to these other transistors as well.

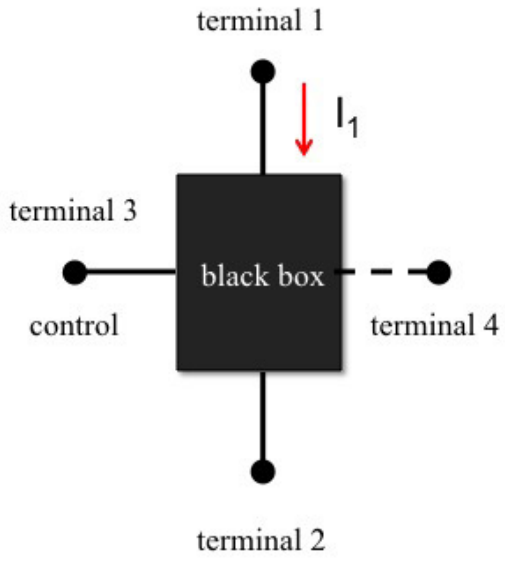

Fig. 2.1 Illustration of a transistor as a black box. The currents that flow in the four leads of the device are controlled by the voltages applied to the four terminals. The relation of the currents to the voltages is determined by the internal device physics of the transistor. These lectures will develop simple, analytical expressions for the current vs. voltage characteristics and relate them to the underlying device physics.

2.2 Physical structure of the MOSFET

Figure 2.2 (same as Fig. 1.1) shows a scanning electron micrograph (SEM) cross section of a Si MOSFET circa 2000. The drain and source terminals (terminals 1 and 2 in Fig. 2.1 are clearly visible, as are the gate electrode (terminal 3 in Fig. 2.1) and the Si body contact (terminal 4 in Fig. 2.1).

Note that the gate electrode is separated from the Si substrate by a thin, insulating layer that is less than 2 nm thick. In present-day MOSFETs, the gap between the source and drain (the channel) is only about 20 nm long.

Also shown in Fig. 2.2 is the schematic symbol used to represent MOSFETs in circuit diagrams. The dashed line represents the channel between the source and drain. It is dashed to indicate that this is an *enhancement mode* MOSFET, one that is only "on" with a channel present when the magnitude of the gate voltage exceeds a critical value known as the *threshold voltage*.

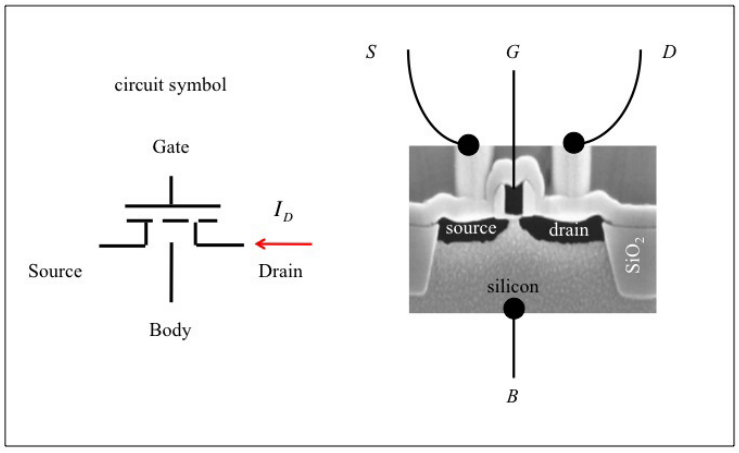

Fig. 2.2 The n-channel silicon MOSFET. Left: The circuit schematic of an enhancement mode MOSFET showing the source, drain, gate, and body contacts. The dashed line represents the channel, which is present when a large enough gate voltage is applied. Right: An SEM cross-section of a silicon MOSFET circa 2000. The source, drain, gate, silicon body, and gate insulator are all visible. (This figure is the same as Fig. 1.1.)

Figure 2.3 compares the cross-sectional and top-views of an n-channel, silicon MOSFET. On the left is a "cartoon" illustration of the cross-section, similar to the SEM in Fig. 2.2. An n-channel MOSFET is built on a p-type Si substrate. The source and drain regions are heavily doped n-type regions; the transistor operates by controlling conduction across the channel that separates the source and drain. On the right right side of Fig. 2.3 is a top view of the same transistor. The large rectangle is the transistor itself. The black squares on the two ends of this rectangle are contacts to the source and drain regions, and the black rectangle in the middle is the gate electrode. Below the gate is the gate oxide, and under it, the p-type silicon channel.

The channel length, L, is a critical parameter; it sets the overall "footprint" (size) of the transistor, and determines the ultimate speed of the transistor (the shorter L is, the faster the ultimate speed of the transistor). The width, W, determines the magnitude of the current that flows. For a given technology, transistors are designed to be well-behaved for channel lengths greater than or equal to some minimum channel length. Circuit designers specify the lengths and widths of transistors to achieve the desired circuit performance. For the past several decades, the minimum channel length (and, therefore, the minimum size of a transistor) has steadily shrunk, which has allowed more and more transistors to be placed on an integrated circuit "chip" [2, 3].

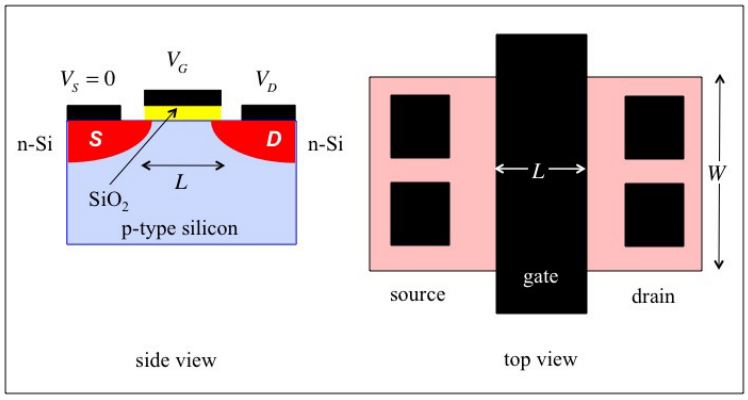

Fig. 2.3 Comparison of the cross-sectional, side view (left) and top view (right) of an n-channel, silicon MOSFET.

In the n-channel MOSFET shown in Fig. 2.3, conduction is by electrons in the conduction band. As shown in Fig. 2.4, it is also possible to make the complementary device in which conduction is by electrons in the valence band (which can be visualized in terms of "holes" in the valence band). A p-channel MOSFET is built on an n-type substrate. The source and drain regions are heavily doped p-type; the transistor operates by controlling conduction across the n-type channel that separates the source and drain. Note that $V_{DS} < 0$ for the p-channel device and that $V_{GS} < 0$ to turn the device on. Also note that the drain current flows out of the drain, rather than into the drain as for the n-channel device. Modern electronics is largely built with CMOS (or complementary MOS) technology for which every n-channel device is paired with a p-channel device.

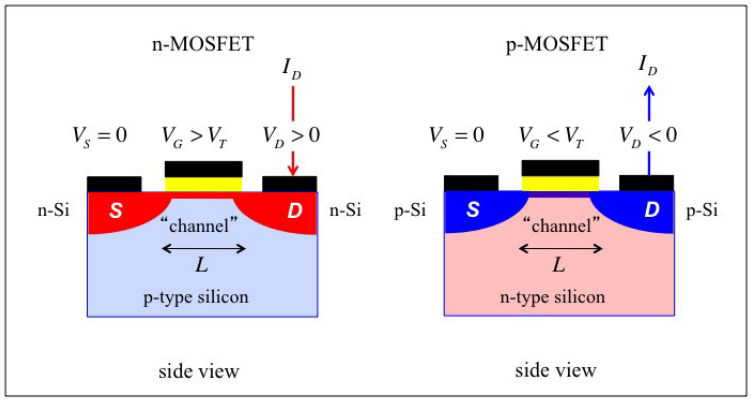

Fig. 2.4 Comparison of an n-channel MOSFET (left) and a p-channel MOSFET (right). Note that $V_{DS}, V_{GS} > 0$ for the n-channel device and $V_{DS}, V_{GS} < 0$ for the p-channel device. The drain current flows in the drain of an n-channel MOSFET and out the drain of a p-channel MOSFET.

For circuit applications, transistors are usually configured to accept an input voltage and to operate at a certain output voltage. The input voltage is measured across the two input terminals and the output voltage across the two output terminals. The input current is the current that flows into one of the two input terminals and out the other, and the output current is the current that flows into one of the two output terminals and out the other. (By convention, the "circuit convention," the current is considered to be positive if it flows into a terminal, so the drain current of an n-channel MOSFET is positive, and the drain current of a p-channel MOSFET is negative.) Since we only have three terminals (the body contact is special - it tunes the operating characteristics of the MOSFET), one of terminals must be connected in common to both the input and the output. Possibilities are *common source*, *common drain*, and *common gate* configurations.

Figure 2.5 shows an n-channel MOSFET connected in the common source configuration. In this case, the DC output current is the drain to source current, I_{DS}, and the DC output voltage is the drain to source voltage, V_{DS}. The DC input voltage is the gate to source voltage, V_{GS}. For MOSFETs, the DC gate current is typically very small and can usually be neglected.

Our goal in this lecture is to understand the general features of transistor IV characteristics and to introduce some of the terminology used. Two types of IV characteristics of are interest; the first are the *output charac-*

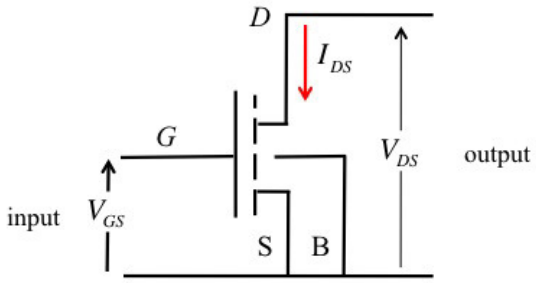

Fig. 2.5 An n-channel MOSFET configured in the common source mode. The input voltage is V_{GS}, and the output voltage, V_{DS} . The output current is I_{DS}, and the gate current is typically negligibly small, so the DC input current is assumed to be zero.

teristics, a plot of the output current, I_{DS}, vs. the output voltage, V_{DS}, for a constant input voltage, V_{GS}. The second IV characteristic of interest is the *transfer characteristic*, a plot of the output current, I_{DS}, as a function of the input voltage, V_{GS} for a fixed output voltage, V_{DS}. In the remainder of this lecture, we treat the transistor as a black box, as in Fig. 2.1, and simply describe the IV characteristics and define some terminology. Subsequent lectures will relate these IV characteristics to the underlying physics of the device.

2.3 IV Characteristics

Figure 2.6 shows the IV characteristics of a simple device, a resistor. For an ideal resistor, the current is proportional to the voltage according to $I = V/R$, where R is the resistance in Ohms. Figure 2.7 shows the IV characteristics of a current source. For an ideal current source, the current is independent of voltage, but real current sources show some dependence of the current on the voltage across the terminals. Accordingly, a real current source can be represented as an ideal current source in parallel with an ideal resistor, as shown in Fig. 2.7. The output characteristics of a MOSFET look like a resistor for small V_{DS} and like a current source for large V_{DS}.

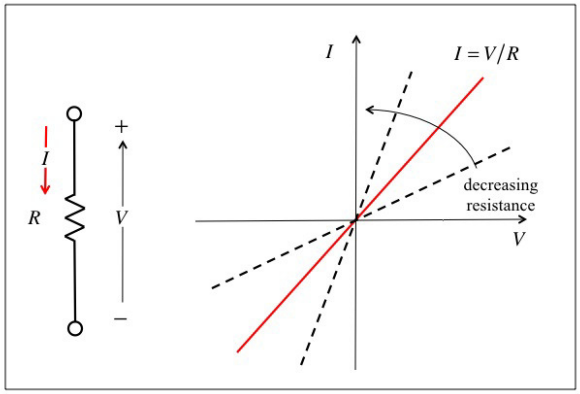

Fig. 2.6 The IV characteristics of an ideal resistor.

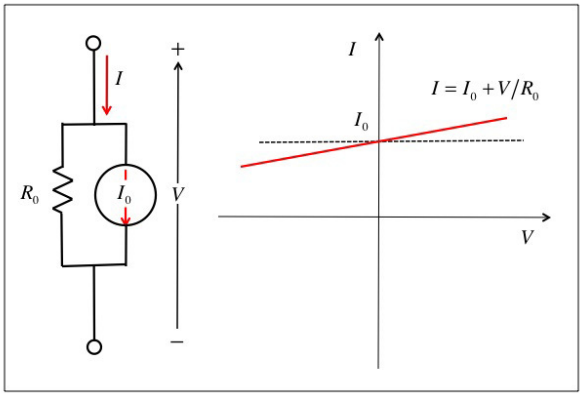

Fig. 2.7 The IV characteristics of a current source. The dashed line is an ideal current source, for which the current is independent of the voltage across the terminals. Real current sources show some dependence of the current on the voltage, which can be represented by a ideal current source in parallel with a resistor, R_0, as shown on the left.

The output characteristics of an n-channel MOSFET are shown in Fig. 2.8 (same as Fig. 1.2). Each line in the family of characteristics corresponds to a different input voltage, V_{GS}. For V_{DS} less than some critical value (called V_{DSAT}), the current is proportional to the voltage. In this small V_{DS} (*linear* or *ohmic*) region, a MOSFET operates like a resistor with the resistance being determined by the input voltage, V_{GS}. For $V_{DS} > V_{DSAT}$, (the *saturation* or *beyond pinch-off* region), the MOSFET operates as a

current source with the value of the current being determined by V_{GS}. The current increases a little with increasing V_{DS}, which shows that the current source has a finite output resistance, r_d. A third region of operation is the *subthreshold* region, which occurs for V_{GS} less than a critical voltage, V_T, the threshold voltage. For $V_{GS} < V_T$, the drain current is very small and not visible when plotted on a linear scale as in Fig. 2.8.

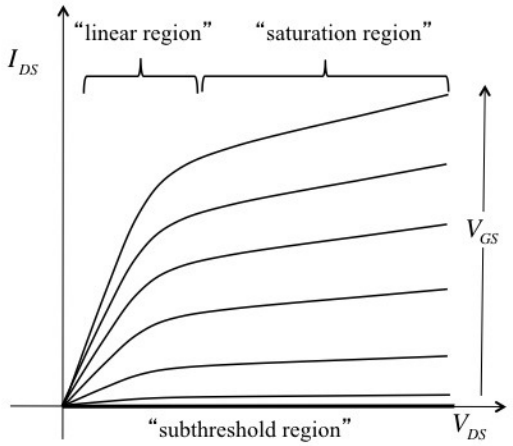

Fig. 2.8 The common source output IV characteristics of an n-channel MOSFET. The vertical axis is the current that flows between the drain and source, I_{DS}, and the horizontal axis is the voltage between the drain and source, V_{DS}. Each line corresponds to a different gate voltage, V_{GS}. The two regions of operation, linear (or ohmic) and saturation (or beyond pinch-off) are also labeled. (This figure is the same as Fig. 1.2.)

Figure 2.9 compares the output and transfer characteristics for an n-channel MOSFET. The output characteristics are shown on the left. Consider fixing V_{DS} to a small value and sweeping V_{GS}. This gives the line labeled V_{DS1} in the transfer characteristics on the right. If we fix V_{DS} to a large value and sweep V_{GS}, then we get the line labeled V_{DS2} in the transfer characteristic. The transfer characteristics also show that for $V_{DS} < V_T$, the current is very small. A plot of $\log_{10}(I_{DS})$ vs. V_{GS} is used to resolve the current in this subthreshold region (see Fig. 2.12).

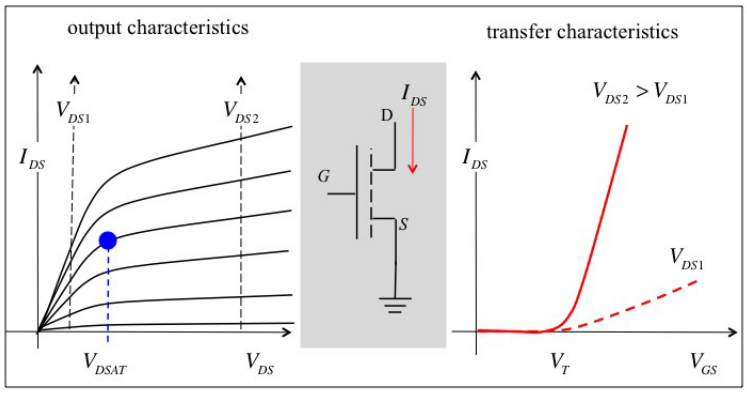

Fig. 2.9 A comparison of the common source output characteristics of an n-channel MOSFET (left) with the common source transfer characteristics of the same device (right). The line labeled V_{DS1} in the transfer characteristics is the low V_{DS} line indicated on the output characteristic on the left, and the line labeled V_{DS2} in the transfer characteristic corresponds to the high V_{DS} line indicated on the output characteristic.

2.4 MOSFET device metrics

The performance of a MOSFET can be summarized by a few device metrics as listed below.

on-current, I_{ON}, in μA/μm
on resistance R_{ON}, in $\Omega - \mu$m
output resistance, r_d, in $\Omega - \mu$m
transconductance, g_m, in μS/μm.
off-current, I_{OFF}, in μA/μm
subthreshold swing, S, in mV/decade
drain-induced barrier lowering, $DIBL$, in mV/V
threshold voltage, V_T(lin) and V_T(sat) in V
drain saturation voltage, V_{DSAT}, in V

The units listed above are those that are commonly used, which are not necessarily MKS units. For example, the transconductance is not typically quoted in Siemens per meter (S/m), but in micro-Siemens per micrometer, μS/μm or milli-Siemens per millimeter, mS/mm.

As shown in Fig. 2.10, several of the device metrics can be determined from the common source output characteristics. The *on-current* is the maximum drain current, which occurs at $I_{DS}(V_{GS} = V_{DS} = V_{DD})$, where V_{DD} is the power supply voltage. Note that the drain to source current, I_{DS},

is typically measured in μA/μm, because the drain current scales linearly with width, W. The *on-resistance* is the minimum channel resistance, which is one over dI_{DS}/dV_{DS} in the linear region for $V_{GS} = V_{DD}$. The units are $\Omega - \mu$m. The *output resistance* is one over dI_{DS}/dV_{DS} in the saturation region; typically quoted at $V_{GS} = V_{DD}$. The units are $\Omega - \mu$m. The *transconductance* is dI_{DS}/dV_{GS} at a fixed drain voltage. It is typically quoted at $V_{DS} = V_{DD}$ and is measured in μS/μm. To get the actual drain current and transconductance, we multiply by the width of the transistor in micrometers. To get the actual on-resistance and output resistance, we divide by the width of the transistor in micrometers.

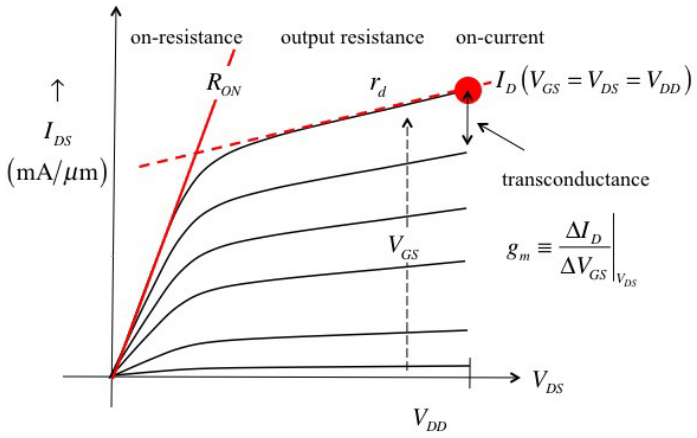

Fig. 2.10 The common source output characteristics of an n-channel MOSFET with four device metrics indicated.

As shown in Fig. 2.11, additional device metrics can be determined from the common source transfer characteristics with the current plotted on a linear scale. The two different IV characteristics are for low V_{DS} (linear region of operation) and for high V_{DS} (saturation region). The on-current noted in Fig. 2.10 is also shown in Fig. 2.11. If we find the point of maximum slope on the IV characteristics, plot a line tangent to the curve at that point, and read off the x-axis intercept, we find the threshold voltage. Note that there are two threshold voltages, one obtained from the linear region plot, V_T(lin) and another from the saturation region plot, V_T(sat) and that V_T(sat) $< V_T$(lin). Note that the off to on transition is gradual and V_T is approximately the point at which this transition is complete.

Finally, the *off-current*, $I_{DS}(V_{GS} = 0, V_{DS} = V_{DD}$, is also indicated in Fig. 2.11, but it is too small to read from this plot.

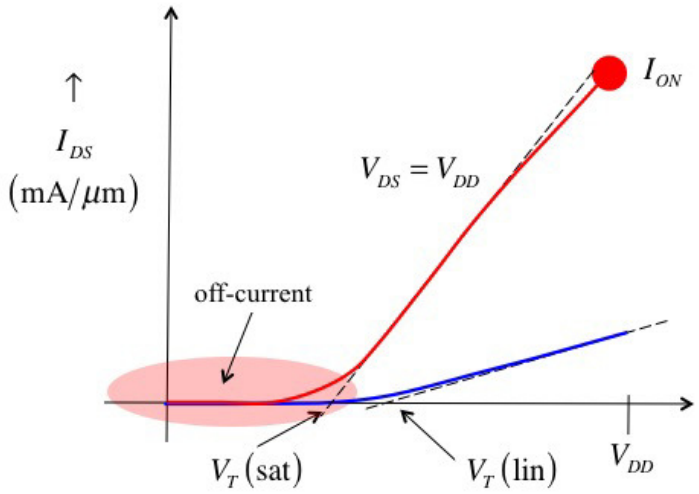

Fig. 2.11 The common source transfer characteristics of an n-channel MOSFET with three device metrics indicated, V_T(lin) and V_T(sat), and the on-current. The drain current, I_{DS}, is plotted on a linear scale in this plot.

To revolve the subthreshold characteristics, we should plot the drain current on a logarithmic scale, as shown in Fig. 2.12. Both the off-current, $I_{OFF} = I_{DS}(V_{GS} = 0, V_{DS} = V_{DD})$, and the on-current, $I_{ON} = I_{DS}(V_{GS} = V_{DS} = V_{DD})$, are identified in this figure. The subthreshold current in a well-behaved MOSFET increases exponentially with V_{GS}. The *subthreshold swing*, SS, is given by

$$SS = \frac{k_B T}{q} \left[\frac{d(\log_{10} I_{DS})}{dV_{GS}} \right] \tag{2.1}$$

and is typically quoted in millivolts per decade. In words, the subthreshold swing is the change in gate voltage (typically quoted in millivolts) needed to change the drain current by a factor of 10. The smaller the SS, the lower is the gate voltage needed to switch the transistor from off to on. As we'll discuss in Sec. 2, the physics of subthreshold conduction dictate that $SS \geq 60$ mV/decade. In a well-behaved MOSFET, the subthreshold

swing is the same for the low and high V_{DS} transfer characteristics. An increase of SS with increasing V_{DS} is often observed and is attributed to the influence of two-dimensional electrostatics (which will also be discussed in Sec. 2).

Finally, we note that the subthreshold IV characteristics are shifted to the left for increasing drain voltages. This shift is attributed to an effect known as *drain-induced barrier lowering* ($DIBL$) and is defined as the horizontal shift in the low and high V_{DS} subthreshold characteristics divided by the difference in drain voltages (typically $V_{DD} - 0.05$ V). DIBL is closely related to the two threshold voltages shown in Fig. 2.11. An ideal MOSFET has zero DIBL and a threshold voltage that does not change with drain voltage, i.e., $V_T(\text{lin}) = V_T(\text{sat})$.

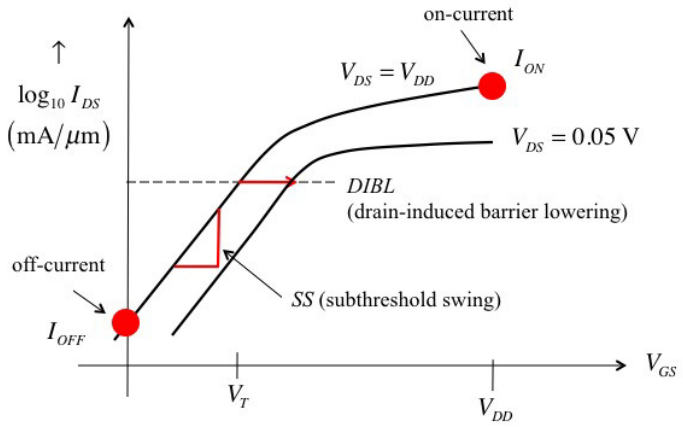

Fig. 2.12 The common source transfer characteristics of an n-channel MOSFET with two additional device metrics indicated, SS and $DIBL$ The drain current, I_{DS}, is plotted on a logarithmic scale in this plot.

As mentioned earlier, it is important to note that threshold voltage is not a precisely defined quantity. It is approximately the gate voltage at which significant drain current begins to flow, and there are different ways to specify this voltage. For example, it may be determined from the x-intercept of a plot of I_{DS} vs. V_{GS} as indicated in Fig. 2.11. Alternatively, one could specify a small drain current (e.g. perhaps 10^{-7}A/μm as in the horizontal dashed line in Fig. 2.12) and simply define V_T as the gate voltage needed to achieve this current. When a threshold voltage is quoted, one

should, therefore, be sure to understand exactly how V_T was defined.

Finally, a word about notation. In Figs. 2.2 and 2.4, we define the current flowing into the drain of an n-MOSFET as I_D. Ideally, the same current flows out the channel and $I_D = I_S$. In practice, there may be some leakage currents (i.e. some of the drain current may flow to the substrate), so that $I_D > I_S$. We'll not be concerned with these leakage currents in these notes and will assume that $I_D = I_S = I_{DS}$, where I_{DS} is the current that flows from the drain to the source.

2.5 Summary

In this lecture we described the shape of the IV characteristics of a MOSFET and defined several metrics that are commonly used to characterize the performance of MOSFETs. We briefly discussed the physical structure of a MOSFET, but did not discuss what goes on inside the black box to produce the IV characteristics we described. Subsequent lectures will focus on the physics that leads to these IV characteristics and on developing simple expressions for several of the key device metrics.

2.6 References

There are many type of transistors, for an incomplete list, see:.

[1] Kwok K. Ng "A survey of semiconductor devices," *IEEE Trans, Electron Devices*, **43**, pp. 1760-1766, 1996.

For an introduction to Moore's Law and its implications for electronics, see:

[2] "Moore's law," `http://en.wikipedia.org/wiki/Moore's_law`, July 19, 2013.

[3] M. Lundstrom,"Moore's Law Forever?" an Applied Physics Perspective, *Science*, **299**, pp. 210-211, January 10, 2003.

Lecture 3

The MOSFET: A barrier-controlled device

3.1 Introduction

Most transistors operate by controlling the height of an energy barrier with an applied voltage. This includes so-called field-effect transistors (FET's), such as MOSFETs, JFETs (junction FET's), HEMTs (high electron mobility transistors, which are also FET's) as well as BJT's (bipolar junction transistors) and HBT's (heterojunction bipolar transistors) [1, 2]. The operating principles of these transistors are most easily understood in terms of energy band diagrams, which provide a qualitative way to understand MOS electrostatics. The energy band view of a MOSFET is the subject of this lecture.

3.2 Equilibrium energy band diagram

As sketched in Fig. 3.1, the MOSFET is inherently a two-dimensional (or even three-dimensional) device. For a complete understanding of the device, we must understand multi-dimensional energy band diagrams, but most of the essential principles can be conveyed with 1D energy band diagrams. Accordingly, we aim to understand the energy vs. position plot along the surface of the MOSFET as indicated in Fig 5.1.

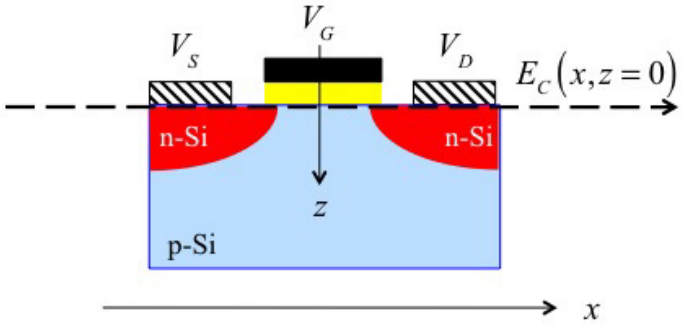

Fig. 3.1 Sketch of a MOSFET cross-section showing the line along the Si surface for which we will sketch the energy vs. position, $E_c\,(x, z = 0)$, from the source, across the channel, to the drain. The y-axis is out of the page, in the direction of the width of the transistor, W.

The source and drain regions of the n-channel MOSFET are heavily doped n-type, and the channel is p-type. In a uniformly doped bulk semiconductor, the bands are independent of position with the Fermi level near E_c for n-type semiconductors and near E_v for p-type. The upper part of Fig. 3.2 shows separate n-type, p-type, and n-type regions. We conceptually put these threes regions together to draw the energy band diagram. In equilibrium, we begin with the principle that the Fermi level (electrochemical potential) is constant. Far to the left, deep in the source, E_c must be near E_F, and far to the right, deep in the drain, the same thing must occur. In the channel, E_v must be near E_F. In order to line up the Fermi levels in the three regions, the source and drain energy bands must drop in energy until E_F is constant (or, equivalently, the channel must rise in

energy). The alignment of the Fermi levels occurs because electrons flow from higher Fermi level to lower Fermi level (from the source and drain regions to the channel), which sets up a charge imbalance and produces an electrostatic potential difference between the two regions. The source and drain regions acquire a positive potential (the so-called *built-in potential*), which lowers the bands according to

$$\begin{aligned} E_c(x) &= E_{co} - q\psi(x) \\ E_v(x) &= E_{vo} - q\psi(x) \end{aligned}, \tag{3.1}$$

where the subscript "o" indicates the value in the absence of an electrostatic potential, ψ.

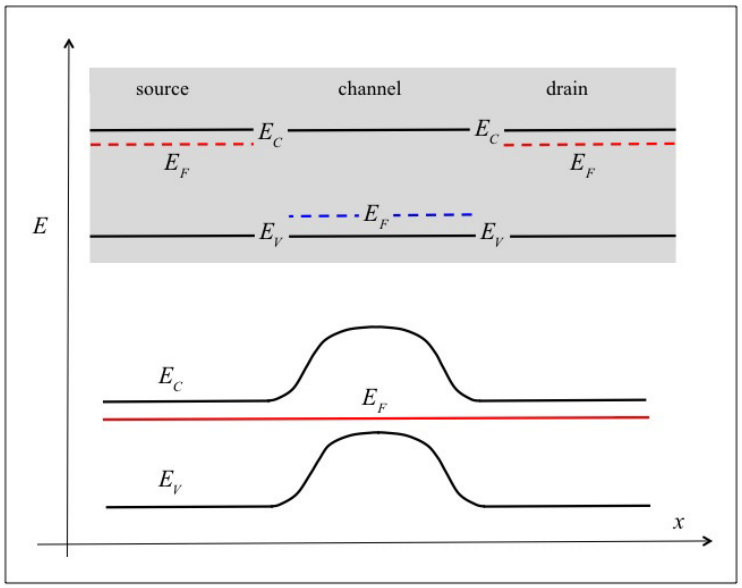

Fig. 3.2 Sketch of the equilibrium energy band diagram along the top surface of a MOSFET. Top: separate n-type, p-type, and n-type regions representing the source, channel, and drain regions. Bottom: The resulting equilibrium energy band diagram when all three regions are connected and $V_S = V_G = V_D = 0$.

Because the device of Fig. 3.2 is in equilibrium, no current flows. Note that there is a potential energy barrier that separates electrons in the source from electrons in the drain. This barrier will play an important role in our understanding of how transistors work. The next step is to understand how the energy bands change when voltages are applied to the gate and drain terminals.

3.3 Application of a gate voltage

Figure 3.3 shows what happens when a positive voltage is applied to the gate. In this figure, we show only the conduction band, because we are discussing an n-channel MOSFET for which the current is carried by electrons in the conduction band. Also shown in Fig. 3.3 are the metal source and drain contacts. (We assume ideal contacts, for which the Fermi levels in the metal align with Fermi level in the semiconductor in equilibrium.) Since $V_S = V_D = 0$, the Fermi levels in the source, device, and drain all align; the device is in equilibrium, and no current flows.

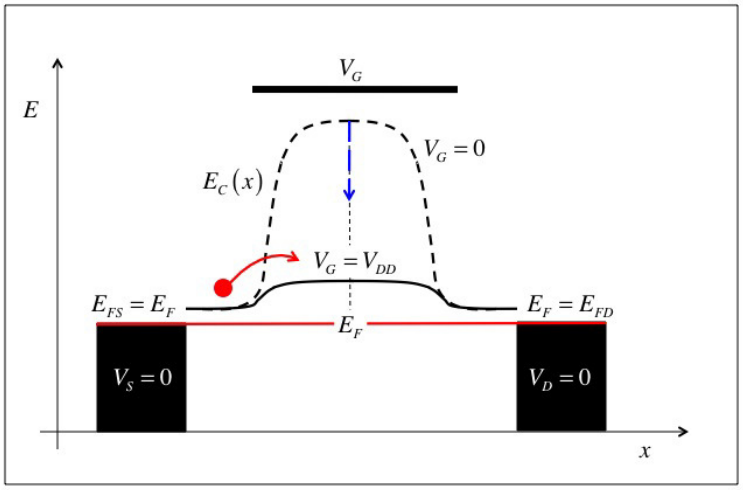

Fig. 3.3 Sketch of the equilibrium electron potential energy vs. position for an n-channel MOSFET for low gate voltage (dashed line) and for high gate voltage (solid line). The voltages on the source, drain, and gate electrodes are zero. The Fermi levels in the source and source contact, in the channel, and in the drain and drain contact are all equal, $E_{FS} = E_F = E_{FD}$ because the device is in equilibrium with no voltage applied to the source and drain contacts. The application of a gate voltage does not disturb equilibrium because the gate electrode is insulated from the silicon by the gate oxide insulator.

Also shown in Fig. 3.3 is what happens when a positive gate voltage is applied. The gate electrode is separated from the silicon channel by an insulating layer of SiO_2, but the positive potential applied to the gate influences the potential in the semiconductor. A positive gate voltage increases the electrostatic potential in the channel, which lowers the conduction band according to eqn. (5.1).

It is important to note that the application of a gate voltage does not affect the Fermi level in the underlying silicon. A positive gate voltage lowers the Fermi level in the gate electrode, but the gate electrode is isolated from the underlying silicon by the gate oxide. The Fermi level in the device can only change if the source or drain voltages change, because the source and drain Fermi levels are connected to the Fermi level in the device.

We conclude that the application of a gate voltage simply raises or lowers the potential energy barrier between the source and drain. The device remains in equilibrium, and no current flows. The fact that the device is in equilibrium even with a gate voltage applied simplifies the analysis of MOS electrostatics, which we will discuss in the next few lectures.

3.4 Application of a drain voltage

Figure 3.4 shows what happens when a large drain voltage is applied. The source is grounded, so the quasi-Fermi level (electrochemical potential) in the source does not change from equilibrium, but the positive drain voltage lowers the quasi-Fermi level in the drain. Lowering the Fermi level lowers E_c too, because $E_F - E_c$ determines the electron density. Electrostatics will attempt to keep the drain neutral, so $n \approx n_0 \approx N_D$, where N_D is the doping density in the drain. The resulting energy band diagrams under low and high gate voltages are shown in Fig. 5.4. Note that we have only shown the quasi-Fermi levels in the source and drain, but $F_n(x)$ varies smoothly across the device. In general, numerical simulations are needed to resolve $F_n(x)$, but it is clear that there will be a slope to $F_n(x)$, so current will flow. The device is no longer in equilibrium when the electrochemical potential varies with position.

Consider first the case of a large drain voltage and small gate voltage as shown in the dashed line of Fig. 3.4. In a well-designed transistor, the height of energy barrier between the source and the channel is controlled only (or mostly) by the voltage applied to the gate. If the gate voltage is low, the energy barrier is high, and very few electrons have enough energy to surmount the barrier and flow to the drain. The transistor is in the *off-state* corresponding to the $I_{DS} \approx 0$ part of the IV characteristic of Fig. 2.10. Current flows, but only the small leakage off-current, I_{OFF} (Fig. 2.12).

When a large gate voltage is applied in addition to the large drain voltage (shown as the solid line in Fig. 3.4), the gate voltage increases the

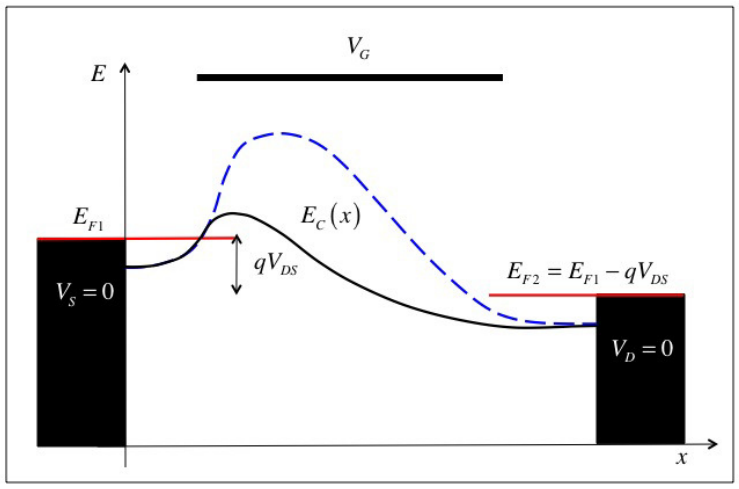

Fig. 3.4 Sketch of $E_c(x)$ vs. x along the channel of an n-channel MOSFET. Dashed line: Large drain voltage and small gate voltage. Solid line: Large drain voltage and large gate voltage.

electrostatic potential in the channel and lowers the height of the barrier. If the barrier is low enough, a significant fraction of the electrons in the source can hop over the energy barrier and flow to the drain. The transistor is in the *on-state* with the maximum current being the on-current, I_{ON}, at $V_{GS} = V_{DS} = V_{DD}$ of Fig. 2.10.

3.5 Transistor Operation

Figure 3.4 illustrates the basic operating principle of most transistors – controlling current by modulating the height of an energy barrier with an applied voltage. We have described the physics of the off-state and on-state of the IV characteristic of Fig. 2.10, but the entire characteristic can be understood with energy band diagrams. Figure 3.5 shows numerical simulations of the conduction band vs. gate voltage in the linear region of operation. Note that under high gate voltage, $E_c(x)$ varies linearly with position in the channel, which corresponds to a constant electric field, as expected in the linear region of operation where the device acts as a gate voltage controlled resistor.

Figure 3.6 shows simulations of the conduction band vs. gate voltage in the saturated region of operation. As the gate voltage pushes the potential

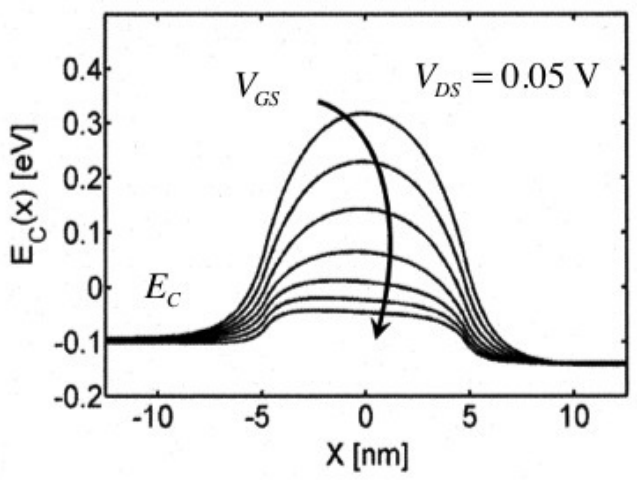

Fig. 3.5 Simulations of $E_c(x)$ vs. x for a short channel transistor. A small drain voltage is applied, so the device operates in the linear region. Each line corresponds to a different gate voltage, with the gate voltage increasing from the top down. The simulations were performed with the nanoMOS simulation program (Z. Ren, et al., nanoMOS 2.5: A Two -Dimensional Simulator for Quantum Transport in Double-Gate MOSFETs, *IEEE Trans. Electron. Dev.*, **50**, pp. 1914-1925, 2003).

energy barrier down, electrons in the source hop over the barrier and then flow down hill to the drain. This figure also illustrates why the drain current saturates with increasing drain voltage. It is the barrier between the source and channel that limits the current. Electrons that make it over the barrier flow down hill and out the drain. Increasing the drain voltage (assuming that it does not lower the source to channel barrier) should not increase the current. Note also that even under very high gate voltage, a small barrier remains. Without this barrier and its modulation by the gate voltage, we would not have a transistor.

3.6 IV Charcteristic

The mathematical form of the IV characteristic of a transistor can also be understood with the help of energy band diagrams and a simple, thermionic

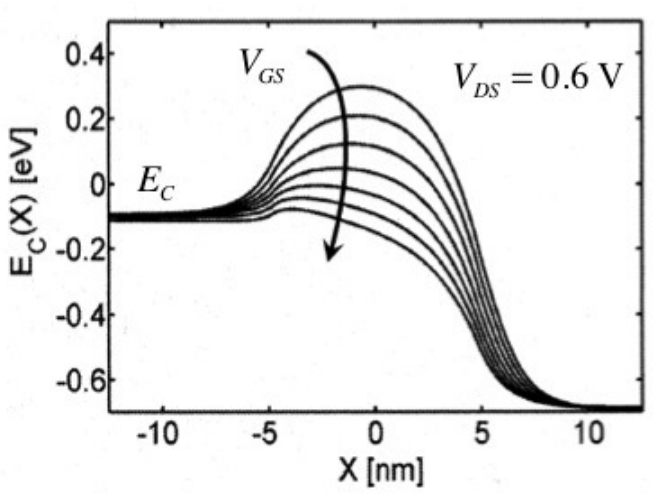

Fig. 3.6 Simulations of $E_c(x)$ vs. x for a short channel transistor. A large drain voltage is applied, so the device operates in the saturation region. Each line corresponds to a different gate voltage, with the gate voltage increasing from the top down. The simulations were performed with the nanoMOS simulation program (Z. Ren, et al., nanoMOS 2.5: A Two -Dimensional Simulator for Quantum Transport in Double-Gate MOSFETs, *IEEE Trans. Electron. Dev.*, **50**, pp. 1914-1925, 2003).

emission model. Consider first the common source characteristic of Fig. 2.10. The net drain current is the current from the left to right (from the source, over the barrier, and out the drain) minus the current form the right to left (from the drain, over the barrier, and out the source):

$$I_{DS} = I_{LR} - I_{RL} \,. \tag{3.2}$$

The probability that an electron can surmount the energy barrier and flow from the source to the drain is $\exp(-E_{SB}/k_BT)$, where E_{SB} is the barrier height from the source to the top of the barrier, so the current from the left to the right is

$$I_{LR} \propto e^{-E_{SB}/k_BT} \,. \tag{3.3}$$

The probability that an electron can surmount the barrier and flow from the drain to the source is $\exp(-E_{DB}/k_BT)$, where E_{DB} is the barrier height from the drain to the top of the barrier. The current from the right to left

is, therefore,

$$I_{RL} \propto e^{-E_{DB}/k_BT} \,. \tag{3.4}$$

Because the drain voltage pulls the conduction band in the drain down, $E_{DB} > E_{SB}$. When there is no DIBL, $E_{DB} = E_{SB} + qV_{DS}$, so $I_{RL}/I_{LR} = \exp(-qV_{DS}/k_BT)$, and we can write the net drain current as

$$I_{DS} = I_{LR} - I_{RL} = I_{LR}\left(1 - e^{-qV_{DS}/k_BT}\right) \,. \tag{3.5}$$

At the top of the barrier, there are two streams of electrons, one moving to the right and one to the left. They have the same kinetic energy, so their velocities, υ_T, are the same. Current is charge times velocity. For a MOSFET, the charge flows in a two-dimensional channel, so it is the charge per area in C/cm^2 that is important. The left to right current is $I_{LR} = WQ_n^+(x=0)\upsilon_T$, where $Q_n^+(x=0)$ is the charge in C/cm^2 at the top of the barrier due to electrons with positive velocities, and W is the width of the MOSFET. Similarly, $I_{RL} = WQ_n^-(x=0)\upsilon_T$. We find the total charge by adding the charge in the two streams,

$$\begin{aligned} Q_n(x=0) &= \frac{I_{LR} + I_{RL}}{W\upsilon_T} \\ &= \frac{I_{LR}}{W\upsilon_T}\left(1 + I_{RL}/I_{LR}\right) \\ &= \frac{I_{LR}}{W\upsilon_T}\left(1 + e^{-qV_{DS}/k_BT}\right) \end{aligned} \,. \tag{3.6}$$

Finally, if we solve eqn. (3.6) for I_{LR} and insert the result in eqn. (3.5), we find the IV characteristic of a ballistic MOSFET as

$$I_{DS} = W|Q_n(x=0)|\upsilon_T \frac{\left(1 - e^{-qV_{DS}/k_BT}\right)}{\left(1 + e^{-qV_{DS}/k_BT}\right)} \,. \tag{3.7}$$

In Lecture 13, we will derive eqn. (3.7) more formally, learn some of its limitations, and define the velocity, υ_T. The general form of the ballistic IV characteristic is, however, easy to understand in terms of thermionic emission in a barrier controlled device.

Now let's examine the general result, eqn. (3.7) under low and high drain bias. For small drain bias, a Taylor series expansion of the exponentials gives

$$I_{DS} = W|Q_n(x=0)|\frac{\upsilon_T}{2k_BT/q}V_{DS} = G_{CH}V_{DS} = V_{DS}/R_{CH} \,, \tag{3.8}$$

where G_{CH} (R_{CH}) is the channel conductance (resistance). Equation (3.8) is a ballistic treatment of the the linear region of the IV characteristic in Fig. 2.10.

Consider next the high V_{DS}, saturated region of the common source characteristic of Fig. 2.10. In this case, $I_{RL} \ll I_{LR}$ and the drain current saturates at $I_{DS} = I_{LR}$. In the limit, $V_{DS} \gg k_B T/q$, eqn. (3.7) becomes

$$I_{DS} = |Q_n(x=0)|\upsilon_T \,. \tag{3.9}$$

The high V_{DS} current is seen to be independent of V_{DS}, but we will see later that DIBL causes $Q_n(x=0)$ to increase with drain voltage, so I_{DS} does not completely saturate.

Having explained the common source IV characteristic, we now turn to the transfer characteristic of Fig. 2.12. The transfer characteristic is a plot of I_{DS} vs. V_{GS} for a fixed V_{DS}. Let's assume that we fix the drain voltage at a high value, so the current is given by eqn. (3.9) and the question is: "How does $Q_n(x=0)$ vary with gate voltage?"

For high drain voltage, $I_{RL} = 0$, so eqn. (3.6) gives

$$|Q_n(x=0)| = \frac{I_{LR}}{W\upsilon_T} \,. \tag{3.10}$$

The current, I_{LR} is due to thermionic emission over the source to channel barrier. Application of a gate voltage lowers this barrier, so we can write:

$$I_{LR} \propto e^{-E_{SB}/k_B T} = e^{-\left(E^0_{SB} - qV_{GS}/m\right)/k_B T} \,, \tag{3.11}$$

where E^0_{SB} is the barrier height from the source to the top of the barrier at $V_{GS} = 0$, and $1/m$ is the fraction of the gate voltage that gets to the semiconductor surface (some of the gate voltage is dropped across the gate oxide). From eqns. (3.11) and (3.10), we find

$$Q_n(V_{GS}) = Q_n(V_{GS}=0)\; e^{qV_{GS}/mk_B T} \,. \tag{3.12}$$

From eqns. (3.12) and (3.9), we see that the current increases exponentially with gate voltage,

$$I_{DS} = W|Q_n(V_{GS}=0)|\upsilon_T \; e^{qV_{GS}/mk_B T} \,. \tag{3.13}$$

In fact, it is easy to show that to increase the current by a factor of ten (a decade), the gate voltage must increase by $2.3mk_BT \geq 0.060$ V at room temperature. This 60 mV per decade is characteristic of thermionic emission over a barrier.

According to eqn. (3.13), the drain current is independent of drain voltage; in practice, there is a small increase in drain current with increasing drain voltage because the drain voltage "helps" the gate pull down the source to channel barrier. This is the physical explanation for DIBL – it

is due to the two-dimensional electrostatics that we will discuss in Lecture 10.

Equation (3.13) describes the exponential increase of I_{DS} with V_G observed in Fig. 2.12 below threshold, but above threshold, the drain current of a MOSFET does not increase exponentially with gate voltage; it increases approximately linearly with gate voltage. Above threshold, eqn. (3.3) still applies, it is just that the decrease in the height of the potential barrier is no longer propositional to V_{GS} above threshold; there is a lot of charge in the semiconductor, which screens the charge on the gate, and make it difficult for the gate voltage to push the barrier down. The parameter, m, becomes very large. The same considerations apply to the charge as well. Below threshold, eqn. (3.12) shows that the charge varies exponentially with gate voltage, but above threshold, we will find that it varies linearly with gate voltage.

When we discuss MOS electrostatics in Lectures 8 and 9, we will show that above threshold, the charge increases linearly with gate voltage as in eqn. (3.14) below.

$$\begin{aligned} Q_n(V_{GS}, V_{DS}) &= -C_{ox}\left(V_{GS} - V_T\right) \\ V_T &= V_{T0} - \delta V_{DS} \end{aligned}, \tag{3.14}$$

where Q_n is the mobile electron charge, $C_{ox} = \kappa_{ox}\epsilon_0/t_{ox}$, where t_{ox} is the oxide thickness, is the gate capacitance per unit area. Also in eqn. (3.14), V_T is the threshold voltage, and δ is the drain-induced barrier lowering (DIBL) parameter. (We'll see later that the appropriate capacitance to use is a little less than C_{ox}.)

This discussion shows that the IV characteristics of a ballistic MOSFET can be easily understood in terms of thermionic emission over a gate controlled barrier. When we return to this problem in Lecture 13, we will learn a more formal and more comprehensive way to treat ballistic MOSFETs, but the underlying physical principles will be the same.

3.7 Discussion

Transistor physics boils down to electrostatics and transport. The energy band diagram is a qualitative illustration of transistor electrostatics. In practice, most of transistor design is about engineering the device so that the energy barrier is appropriately manipulated by the applied voltages. The design challenges have increased as transistors have gotten smaller

and smaller, and we understand transistor electrostatics better now, but the basic principles are the same as they were in the 1960's.

Figure 3.7 illustrates the key principles of a well-designed short channel MOSFET. The top of the barrier is a critical point; it marks the beginning of the channel and is also called the *virtual source*. In a well-designed MOSFET, the height of the source to channel energy barrier is strongly controlled by the gate voltage and only weakly dependent on the drain voltage.

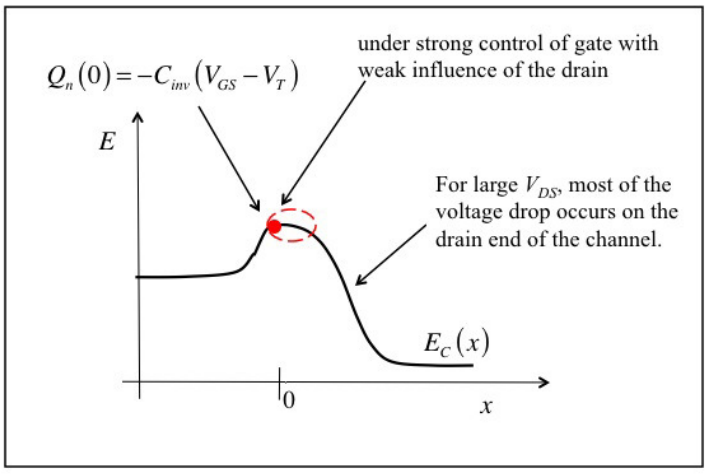

Fig. 3.7 Sketch of a well-designed short channel MOSFET under high gate and drain bias. In a well-designed short channel MOSFET, the charge at the top of the barrier is very close to the value it would have in a long channel device, for which which the lateral electric field could be neglected. In a well-designed MOSFET, there is a low lateral electric field near the beginning of the channel and under high V_{DS}, the drain voltage has only a small influence on the region near the top of the barrier.

Under low V_{DS} and high V_{GS}, the potential drops approximately linearly in the channel, so the electric field is approximately constant. Under high drain and gate bias, the electric field is high and varies non-linearly with position. Near the beginning the the channel (near the top of the barrier) the electric field is low, but near the drain, the electric field is very large. In the saturation region, increases in drain voltage increase the potential drop in the high field part of the channel but leave the region near the top of the barrier relatively unaffected (if DIBL is small). Since the region near the top of the barrier controls the current, the drain current is relatively

insensitive to the drain voltage in the saturation region.

Note from Fig. 3.7 that electrons that surmount the barrier and flow to the drain gain lot of kinetic energy. Some energy will be lost by electron-phonon scattering, but in a nanoscale transistor, there is not enough time for electrons to shed their kinetic energy as they flow to the drain. Accordingly, the velocity is very high in the part of the channel where the lateral potential drop (electric field) is high. Because current is the product of charge times velocity, the electron charge density will be very low in the region where the velocity is high. The part of the channel where the lateral potential drop is large and the electron density low is known in classical MOS theory as the *pinch-off* region. In a short channel device, the pinch-off region can be a substantial part of the channel, but for an electrostatically well-designed MOSFET, there must always be a small region near the source where the potential is largely under the control of the gate, and the lateral potential drop is small.

Figure 3.8 is a sketch of a long channel transistor under high gate a drain bias. Compared to the short channel transistor sketched in Fig. 3.7, we see that the low-field region under the control of the gate is a very large part of the channel, but there is still a short, pinch-off region near the drain. The occurrence of the pinch-off region under high drain bias is what causes the current to saturate. In the saturation or *beyond pinch-off* region, the current is mostly determined by transport across the low-field part of the channel, which is near the source, but most of the potential drop across the channel occurs in the high-field portion of the channel, which is near the drain. Once electrons enter the pinch-off region, they are quickly swept out to the drain.

In a well-designed MOSFET, the region near the top of the barrier is under the strong control of the gate voltage and only weakly affected by the drain voltage. The goal in transistor design is to achieve this performance as channel length scaling brings the drain closer and closer to the source. Once electrons hop over the source to channel barrier, they can flow to the drain. While the electrostatic design principles of MOSFETs have not changed much for the past five decades (but they have gotten much more challenging), the nature of electron transport in transistors has changed considerably as transistors have become smaller and smaller. A proper treatment of transport in nanoscale transistors is essential to understanding and designing these devices and and will be our focus beginning in Lecture 14.

We have discussed 1D energy bands for a MOSFET by sketching $E_c\,(x)$

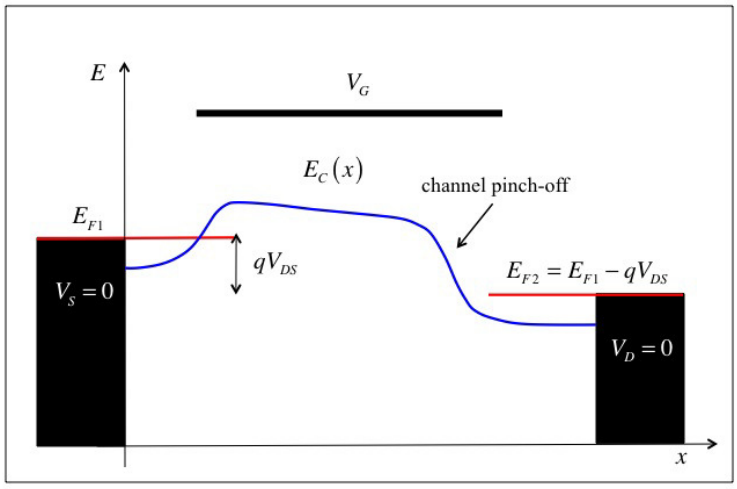

Fig. 3.8 Sketch of a long channel MOSFET under high gate and drain bias. In this case, the low lateral electric occupies a substance part of the channel the pinch-off region is short. Additional increases in V_{DS}, lengthen the pinch-off region a bit, but in a long channel transistor, if occupies a small portion of the channel.

for $z = 0$, the surface of the silicon. Figure 3.9 shows these energy band diagrams in two dimensions. Figure 3.9a is a sketch of the device. Figure 3.9b shows a device in equilibrium with $V_S = V_D = 0$ and the gate voltage adjusted so that the bands are flat in the direction normal to the channel. Figure 3.9c shows the device with a large gate voltage applied, but with V_S and V_D still at zero volts. Note that E_c along the surface of the device is just like the solid line in Fig. 3.3. Figure 5.9d shows the energy band diagram with large gate and drain voltages applied. In this case, E_c along the surface is just like the solid line in Fig. 3.4.

Finally, we note that the energy band diagrams that we have sketched are similar to the energy band diagrams for a bipolar transistor [1, 2]. In fact, the two devices both operate by controlling current by manipulating the height of an energy barrier [3]. The source of the MOSFET is analogous to the emitter of the BJT, the channel to the base of the BJT, and the drain to the collector of a BJT. This close similarity will prove useful in understanding the operation of short channel MOSFETs.

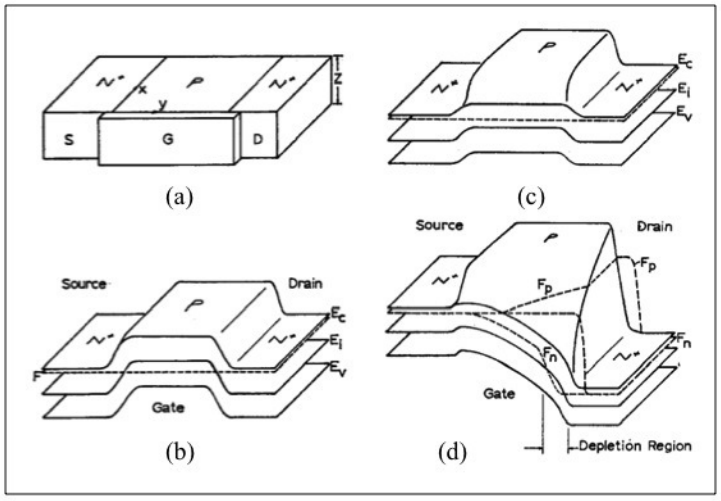

Fig. 3.9 Two dimensional energy band diagram for an n-channel MOSFET. (a) the device structure, (b) the equilibrium energy band diagram, (c) an equilibrium energy band diagram with a large gate voltage applied, and (d) the energy band diagram with large gate and drain voltages applied. (From Fig. 1 in H.C Pao and C.T. Sah, "Effects of Diffusion Current on the Charateristis of Metal-Oxide (Insulator)-Semiconductor Transistors," *Solid-State Electron.* **9**, pp. 927-937, 1966.)

3.8 Summary

The MOSFET operates by controlling current through the manipulation of an energy barrier with a gate voltage. Understanding this gives a clear, physical understanding of how long and short channel MOSFETs operate. The control of current by a energy barrier is what gives a transistor its characteristic shape.

We can write the drain current as

$$I_{DS} = W|Q_n\left(V_{GS}, V_{DS}\right)|\left\langle \upsilon \right\rangle \ . \tag{3.15}$$

This equation simply says that the drain current is proportional to the amount of charge in the channel and how fast that charge is moving. (The sign of Q_n is negative and because the current is defined to be positive when it flows into the drain, the absolute value is taken.) The charge, Q_n, flows into the channel to balance the charge on the gate electrode. While the shape of the IV characteristic is determined by MOS electrostatics, the magnitude of the current depends on how fast that charge flows.

3.9 References

Most of the important kinds of transistors are discussed in these texts:

[1] Robert F. Pierret *Semiconductor Device Fundamentals, 2nd Ed.*, , Addison-Wesley Publishing Co, 1996.

[2] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, 2nd Ed., Oxford Univ. Press, New York, 2013.

Johnson describes the close relation of bipolar and field-effect trnsistors.

[3] E.O. Johnson, "The IGFET: A Bipolar Transistor in Disguise," *RCA Review*, **34**, pp. 80-94, 1973.

Lecture 4

MOSFET IV: Traditional Approach

4.1 Introduction

The traditional approach to MOSFET theory was developed in the 1960's [1 - 4] and although they have evolved considerably, the basic features of the models used today are very similar to those first developed more than 50 years ago. My goal in this lecture is to briefly review the traditional theory of the MOSFET as it is presented in most textbooks (e.g. [5, 6]). Only the essential ideas of the traditional approach will be discussed. For example, we shall be content to compute the linear region current, and the saturated region current and not the entire IV characteristic. Only the above threshold IV characteristics will be discussed, not the subthreshold characteristics. Those interested in a full exposition of traditional MOSFET theory should consult standard texts such as [7, 8]. Later in these lecture notes, we will develop a much different approach to MOSFET theory – one better suited to the physics of nanoscale transistors, but we will also show, that it can be directly related to the traditional approach reviewed in this lecture.

4.2 Current, charge, and velocity

Figure 4.1 is a "cartoon" sketch of a MOSFET for which the drain to source current can be written as in eqn. (1.1),

$$I_{DS} = W|Q_n(x)|\langle v(x)\rangle \ , \tag{4.1}$$

where W is the width of the transistor in the y-direction, Q_n is the mobile sheet charge in the $x-y$ plane ($\mathrm{C/m^2}$), and $\langle v\rangle$ is the average velocity at which the charge flows. We assume that the device is uniform in the y-direction and that current flows in the x-direction from the source to the drain. The quantity, Q_n, is called the *inversion layer charge* because it is an electron charge in a p-type material. The electron charge and velocity vary with position along the channel, but the current is constant if there is no electron recombination or generation. Accordingly, we can evaluate the current at the location along the channel where it is the most convenient to do so.

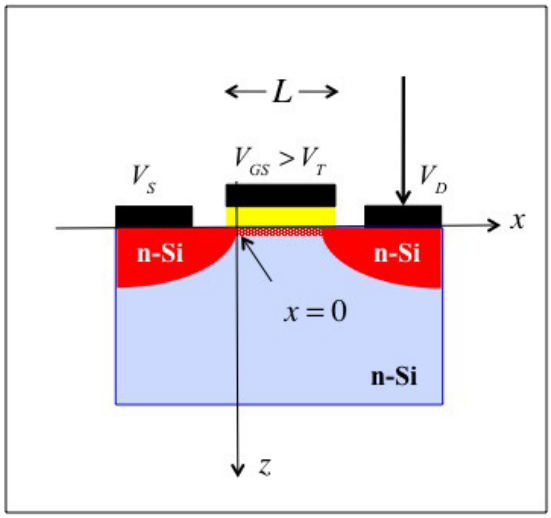

Fig. 4.1 Sketch of a simple, n-channel, enhancement mode MOSFET. The z-direction is normal to the channel, and the y-axis is out of the page. The beginning of the channel is located at $x = 0$. An inversion charge is present in the channel because $V_{GS} > V_T$; it is uniform between $x = 0$ and $x = L$ as shown here, if $V_S = V_D = 0$.

Consider the MOSFET of Fig. 4.1 with $V_S = V_D = 0$, but with $V_G > 0$. The MOSFET is in equilibrium and no current flows. In this case, the inversion layer charge is independent of x. As we will discuss in Sec. 2, there very little charge when the gate voltage is less than a critical value, the threshold voltage, V_T. For $V_{GS} > V_T$, the charge is negative and proportional to $V_{GS} - V_T$,

$$Q_n(V_{GS}) = -C_{ox}\left(V_{GS} - V_T\right) , \tag{4.2}$$

where C_{ox} is the gate oxide capacitance per unit area,

$$C_{ox} = \frac{\kappa_{ox}\epsilon_0}{t_{ox}} \quad \mathrm{F/m^2} , \tag{4.3}$$

with the numerator being the dielectric constant of the oxide and the denominator the thickness of the oxide. (As we'll discuss in Lecture 8, the gate capacitance is actually somewhat less than C_{ox} when the oxide is thin.) For $V_{GS} \leq V_T$, the charge is assumed to be negligibly small.

When $V_D > V_S$, the inversion layer charge density varies with position along the channel, and so does the average velocity of electrons. As we shall see when we discuss MOS electrostatics, in a well-designed transistor, Q_n at the beginning of the channel is given by eqn. (4.2). Accordingly, we will evaluate I_{DS} at $x = 0$, where we know the charge, and we only need to deduce the average velocity, $\langle \upsilon\,(x = 0)\rangle$.

4.3 Linear region

In the small V_{DS}, or linear region of the output characteristics (Fig. 2.8), a MOSFET acts as a voltage controlled resistor. Above threshold, the electric field in the channel is constant, and we can write the average velocity as

$$\langle \upsilon \rangle = -\mu_n \mathcal{E} = -\mu_n V_{DS}/L . \tag{4.4}$$

Using Eqns. (4.2) and (4.4) in (4.1), we find

$$\boxed{I_{DS} = \frac{W}{L}\mu_n C_{ox}\left(V_{GS} - V_T\right)V_{DS} ,} \tag{4.5}$$

which is the classic expression for the small V_{DS} drain current of a MOSFET. Note that we have labeled the mobility as μ_n, but in traditional

MOS theory, this mobility is called the *effective mobility*, μ_{eff}. The effective mobility is the depth-averaged mobility in the inversion layer. It is smaller than the electron mobility in the bulk, because *surface roughness scattering* at the oxide-silicon interface lowers the mobility.

4.4 Saturated region: Velocity saturation

In the large V_{DS}, or saturated region of the output characteristics (Fig. 2.8), a MOSFET acts as a voltage controlled current source. For a relatively small drain to source voltage of about 1 V, the electric electric field in the channel of a modern short channel (≈ 20 nm) MOSFET is very high – well above the ≈ 10 kV/cm needed to saturate the velocity in bulk Si (recall Fig. 4.5). If the electric field is large across the entire channel for $V_{DS} > V_{DSAT}$, then the velocity is constant across the channel with a value of υ_{sat}, and we can write the average velocity as

$$\langle \upsilon(x) \rangle = \upsilon_{sat} \approx 10^7 \quad \text{cm/s}\,. \tag{4.6}$$

Using eqns. (4.2) and (4.6) in (4.1), we find

$$\boxed{I_{DS} = WC_{ox}\upsilon_{sat}\left(V_{GS} - V_T\right)\,,} \tag{4.7}$$

which is the classic expression for the *velocity saturated* drain current of a MOSFET. Note that in practice, the current does not completely saturate, but increases slowly with drain voltage. In a well-designed Si MOSFET, the output conductance is primarily due to DIBL as described by eqn. (3.11).

Finally, we should note that it is now understood that in a short channel MOSFET, the maximum velocity in the channel does not saturate – even when the electric field is very high. Nevertheless, the traditional approach to MOSFET theory, still presented in most textbooks, assumes that the electron velocity saturates when the electric field in the channel is large.

4.5 Saturated region: Classical pinch-off

Consider next a long channel MOSFET under high drain bias. In this case, the electric field is moderate, and the velocity is not expected to saturate. Nevertheless, we still find that the drain current saturates, so it must be for a different reason. This was the situation in early MOSFET's for which

the channel length was about 10 micrometers (10,000 nanometers), and the explanation for drain current saturation was *pinch-off* near the drain.

Under high drain bias, the potential in the channel varies significantly from V_S at the source to V_D at the drain end (See Ex. 4.2). Since it is the difference between the gate voltage and the Si channel that matters, eqn. (4.2) must be extended as

$$Q_n(V_{GS}, x) = -C_{ox}\big(V_{GS} - V_T - V(x)\big)\,, \tag{4.8}$$

where $V(x)$ is the potential along the channel. According to eqn. (4.8), when $V_D = V_{GS} - V_T$, at the drain end, we find $Q_n(V_{GS}, L) = 0$. We say that the channel is pinched off at the drain. Of course, if $Q_n = 0$, then eqn. (4.1) states that $I_{DS} = 0$, but a large drain current is observed to flow. This occurs because in the pinched off region, carriers move very fast in the high electric field, so Q_n is finite, although very small. The current saturates for drain voltages above above $V_{GS} - V_T$ because the additional voltage is dropped across the small, pinched off part of the channel. The voltage drop across the conductive part of the channel remains at about $V_{GS} - V_T$. We are now ready to compute the saturated drain current due to pinch-off.

Figure 4.2 is an illustration of a long channel MOSFET under high gate bias and for a drain bias greater than $V_{GS} - V_T$. Over most of the channel, there is a strong inversion layer, and $\upsilon(x) = \mu_n \mathcal{E}(x)$. When carriers enter the pinched-off region, the large electric field quickly sweeps the carriers across and to the drain. (The energy band view of pinch-off was presented in Fig. 3.8.)

In the part of the channel where the inversion charge density is large, we can write the average velocity as

$$\langle \upsilon(x) \rangle = -\mu_n \mathcal{E}(x)\,. \tag{4.9}$$

The voltage at the beginning of the channel is $V(0) = V_S = 0$, and the voltage at the end of the channel where it is pinched off off is $V_{GS} - V_T$. The electric field at the beginning of the channel is (see Ex. 4.2)

$$\mathcal{E}(0) = \frac{V_{GS} - V_T}{2L'}\,, \tag{4.10}$$

where the factor of two comes from a proper treatment of the nonlinear electric field in the channel and L' is the length of the part of the channel that is not pinched off. Using eqn. (4.10) in (4.9), we find

$$\langle v(0) \rangle = -\mu_n \mathcal{E}(0) = -\mu_n \frac{V_{GS} - V_T}{2L'} \,. \tag{4.11}$$

Finally, using eqns. (4.2) and (4.11) in (4.1), we find

$$\boxed{I_{DS} = \frac{W}{2L} \mu_n C_{ox} \left(V_{GS} - V_T \right)^2 \,,} \tag{4.12}$$

the so-called *square law IV characteristic* of a long channel MOSFET. In practice, the current does not completely saturate, but increases slowly with drain voltage as the pinched-off region slowly moves towards the source, which effectively decreases the length of the conductive part of the channel, L'.

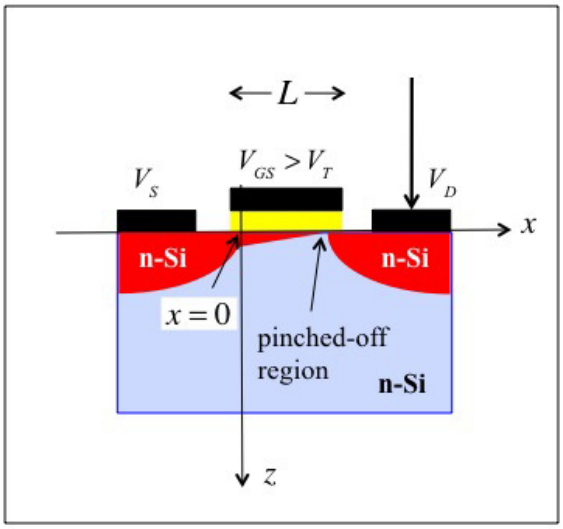

Fig. 4.2 Sketch of a long channel MOSFET showing the pinched-off region. Note that the thickness of the channel on this figure is used to illustrate the magnitude of the charge density (more charge near the source end of the channel than near the drain end). The channel is physically thin in the z-direction near the source end, where the gate to channel potential is large and physically thicker near the drain end, where the gate to channel potential is smaller The length of the part of the channel where Q_n is substantial is $L' < L$.

Exercise 4.1: Complete, square law IV characteristic

Equations (4.5) and (4.12) describe the linear and saturation region currents as given by the traditional square law theory of the MOSFET. In this exercise, we'll compute the complete IV characteristic from the linear region to the saturation region. We begin with eqn. (4.1) for the drain current and use eqn. (4.4) for the velocity to write

$$I_{DS} = W|Q_n(x)|\langle v(x)\rangle = W|Q_n(x)|\mu_n \frac{dV}{dx}\,. \tag{4.13}$$

Next, we use eqn. (4.8) for the charge to write,

$$I_{DS} = W\mu_n C_{ox}\big(V_{GS} - V_T - V(x)\big)\frac{dV}{dx}\,, \tag{4.14}$$

then separate variables and integrate across the channel to find,

$$I_{DS}\int_0^L dx = W\mu_n C_{ox}\int_{V_S}^{V_D}(V_{GS} - V_T - V)\,dV\,, \tag{4.15}$$

where we have assumed that I_{DS} is constant (no recombination-generation in the channel) and that μ_n is constant as well. Integration gives us the IV characteristic of the MOSFET,

$$I_{DS} = \frac{W}{L}\mu_n C_{ox}\big[(V_{GS} - V_T)\,V_{DS} - V_{DS}^2/2\big]\,. \tag{4.16}$$

Equation (4.16) gives the drain current for $V_{GS} > V_T$ and for $V_{DS} \leq (V_{GS} - V_T)$. The charge in eqn. (4.8) goes to zero at $V_{DS} = V_{GS} - V_T$, which defines the beginning of the pinch-off region. The current beyond pinch-off is found by evaluating eqn. (4.16) for $V_{DS} = V_{GS} - V_T$ and is

$$I_{DS} = \frac{W}{2L'}\mu_n C_{ox}\,(V_{GS} - V_T)^2 \tag{4.17}$$

and only changes for increasing V_{DS} because of channel length shortening due to pinch-off (i.e. $L' < L$).

Equations (4.16) and (4.17) give the square law IV characteristics of the MOSFET – not just the linear and saturated regions, but the entire IV characteristics.

Exercise 4.2: Electric field vs. position in the channel

In the development of the traditional model, we asserted that the electric field in the channel was V_{DS}/L under low drain bias and $(V_{GS}-V_T)/2L'$ under high drain bias in a long channel MOSFET. In this exercise, we will compute the electric field in the channel and show that these assumptions are correct.

Beginning with eqn. (4.14), we can use (4.16) for I_{DS} to find

$$\frac{1}{L'}lbigl[(V_{GS}-V_T)\,V_{DS}-V_{DS}^2/2\big]=\big(V_G-V_T-V(x)\big)\frac{dV}{dx}\,, \tag{4.18}$$

then we separate variables and integrate from the source at $x=0, V_S=0$ to an arbitrary location, x, in the channel where $V=V(x)$. The result is

$$\big[(V_{GS}-V_T)\,V_{DS}-V_{DS}^2/2\big]\frac{x}{L'}=(V_{GS}-V_T)\,V(x)-V^2(x)/2\,, \tag{4.19}$$

which is a quadratic equation for $V(x)$ that can be solved to find

$$V(x)=(V_{GS}-V_T)\left[1-\sqrt{1-\frac{2(V_{GS}-V_T)V_{DS}-V_{DS}^2/2}{(V_{GS}-V_T)^2}\left(\frac{x}{L'}\right)}\right]. \tag{4.20}$$

Equation (4.20) can be differentiated to find the electric field. Let's examine the electric field for two cases. First, assume small V_{DS}, the linear region of operation, where eqn. (4.20) becomes

$$V(x)=(V_{GS}-V_T)\left[1-\sqrt{1-\frac{2V_{DS}}{(V_{GS}-V_T)}\left(\frac{x}{L'}\right)}\right], \tag{4.21}$$

and the square root can be expanded for small argument ($\sqrt{(1-x)}\approx 1-x/2$) to find

$$V(x)=V_{DS}\frac{x}{L} \tag{4.22}$$

(Note that $L'=L$ for small V_{DS}.) Finally, differentiating eqn. (4.22), we find that the electric field for small V_{DS} is

$$-\frac{dV(x)}{dx}=\mathcal{E}=-\frac{V_{DS}}{L}\,, \tag{4.23}$$

which is the expected result.

Next, let's evaluate the electric field under pinched-off conditions, $V_{DS} = V_{GS} - V_T$. Equation (4.20) becomes

$$V(x) = (V_{GS} - V_T)\left[1 - \sqrt{1 - x/L'}\,\right], \tag{4.24}$$

and the electric field is

$$\mathcal{E}(x) = -\frac{dV}{dx} = -\frac{(V_{GS} - V_T)}{2L'}\left[\frac{1}{\sqrt{1 - x/L'}}\right]. \tag{4.25}$$

At $x = 0$, eqn. (4.25) gives the result, eqn. (4.10), which we simply asserted earlier. At $x = L'$, where the channel is pinched-off, we find $\mathcal{E}(L') \to \infty$. This result should be expected because in our model, $Q_n = 0$ at the pinch-off point, so it takes an infinite electric field to carrier a finite current.

4.6 Discussion

i) velocity saturation and drain current saturation

Equations (4.5), (4.7), and (4.12) describe the linear and saturation region IV characteristics of MOSFETs according to traditional MOS theory. We have presented two different treatments of the saturated region current; in the first, drain current saturation was due to velocity saturation in a high channel field, and in the second, it was due to the development of a pinched-off region near the drain end of the channel. When the average electric field in the channel is much larger than the critical field for velocity saturation ($\approx$ 10 kV/cm) then we expect to use the velocity saturation model. We should use the velocity saturation model when

$$\frac{V_{GS} - V_T}{L} \gg \mathcal{E}_{cr} \approx 10\ \text{kV/cm}\,. \tag{4.26}$$

Putting in typical numbers of $V_{GS} = V_{DD} = 1$ V and $V_T = 0.2$ V, we find that the velocity saturation model should be used when $L \lesssim 1\,\mu$m. Indeed, velocity saturation models first began to be widely-used in the 1980's when channel lengths reached one micrometer [9].

Figure 4.3 shows the common source output characteristics of an n-channel Si MOSFET with a channel length of about 60 nm. It is clear from the results that $I_{DS} \propto (V_{GS} - V_T)$ under high drain bias, so that

the velocity saturation model of eqn. (4.7) seems to describe this device. Indeed, the observation of a saturation current that varies linearly with gate voltage is taken as the "signature" of velocity saturation.

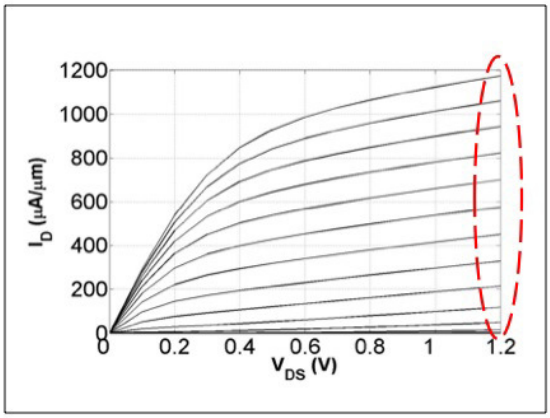

Fig. 4.3 Common source output characteristics of an n-channel. Si MOSFET with a gate length of $L \approx 60$ nm. The top curve is for $V_{GS} = 1.2$ V and the step is 0.1 V. Note that for large V_{DS}, the drain current increases linearly with gate voltage. This behavior is considered to be the signature of velocity saturation in the channel. The device is described in C. Jeong, D. A. Antoniadis and M.S. Lundstrom, "On Backscattering and Mobility in Nanoscale Silicon MOSFETs, *IEEE Trans. Electron Dev.*, **56**, pp. 2762-2769, 2009.

For the MOSFET of Fig. 4.3, $V_T \approx 0.4$ V. For the maximum gate voltage, the pinch-off model would give a drain saturation voltage of $V_{DSAT} = V_{GS} - V_T \approx 0.8$ V, which is clearly too high for this device and tells us that the drain current is not saturating due to classical pinch-off. References [7] and [8] discuss the calculation of V_{DSAT} in the presence of velocity saturation.

Although velocity saturation models seem to accurately describe short channel MOSFETs, there is a mystery. Detailed computer simulations of carrier transport in nanoscale MOSFETs show that the velocity **does not saturate** in the high electric field portion of a short channel MOSFET. There is simply not enough time for carriers to scatter enough to saturate the velocity; they traverse the channel and exit the drain too quickly. Nevertheless, the IV characteristic of Fig. 4.3 tell us that the velocity in the channel saturates. Understanding this is a mystery that we will unravel as

we explore the nanoscale MOSFET.

ii) device metrics

Equations (4.5) and (4.7) describe the IV characteristic of modern short channel MOSFETs and can be used to relate some of the device metrics listed in Sec. 2.4 to the underlying physics. Using these equations, we find:

$$\boxed{\begin{aligned} &I_{ON} = WC_{ox}v_{sat}\left(V_{DD} - V_T\right) \qquad\qquad V_T = V_{T0} - \delta V_{DS} \\ &R_{ON} = \left(\left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{GS}=V_{DD},V_{DS}\approx 0}\right)^{-1} = \left(\frac{W}{L}\mu_n C_{ox}\left(V_{GS} - V_T\right)\right)^{-1} \\ &g_m^{sat} = \left.\frac{\partial I_{DS}}{\partial V_{GS}}\right|_{V_{GS}=V_{DS}=V_{DD}} = WC_{ox}v_{sat} \\ &r_d = \left(\left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{GS}=V_{DD},V_{DS}>V_{DSAT}}\right)^{-1} = \frac{1}{g_m^{sat}\delta} \\ &|A_v| = g_m^{sat} r_d = \frac{1}{\delta} \end{aligned}} \tag{4.27}$$

The parameter, $|A_v|$ is the *self-gain*, an important figure of merit for analog applications.

Finally, we should discuss energy band diagrams. While energy bands did not appear explicitly in our discussion, they are present implicitly. The beginning of the channel, $x = 0$, is the top of the energy barrier in Figs. 3.5 and 3.6 (or close to the top of the barrier [10]). As we'll discuss later, in a well-designed MOSFET, the charge at the top of the barrier is given by eqn. (4.2). This charge comes from electrons in the source that surmount the energy barrier. The location at the beginning of the channel where eqn. (4.2) applies is also known as the *virtual source*.

The energy band view is especially helpful in understanding pinch-off. From the illustration in Fig. 4.2, it can be confusing as to how carriers can leave the end of the channel and flow across the pinched-off region. Energy bands make it clear. As was shown in Fig. 3.6, the pinched-off region is the high electric field region near the drain, where the slope of $E_c(x)$ is the steepest. But electrons that enter this region from the channel simply flow downhill and out the drain. There is nothing to stop them when the enter the pinched-off region.

4.7 Summary

Our goal in this lecture has been to discuss the traditional approach to MOSFET *IV* theory. In practice, there are several complications to consider, such as the role of the depleted change in Eqn. (4.8), current for an arbitrary drain voltage, etc. The reader is referred to references [5-8] for more discussion, but the essential features the traditional approach are easy to grasp, and will give is a point of comparison for the much different picture of the nanoscale MOSFET that will be developed in subsequent lectures.

Returning again to eqn. (4.1), we see that the drain current is proportional to the product of charge and velocity. The charge is controlled by MOS electrostatics (i.e. by manipulating the energy barrier between the source and the channel). The traditional approach to MOS electrostatics is still largely applicable, with some modifications due to quantum confinement (recall Sec. 3.8). The lectures in Sec. 2 will review the critically important electrostatics of the MOSFET.

4.8 References

The mathematical modeling of transistors began in the 1960's. Some of the early papers MOSFET IV characteristics are listed below.

[1] S.R. Hofstein and F.P. Heiman, "The Silicon Insulated-Gate Field- Effect Transistor, *Proc. IEEE*, **51**, pp. 1190-1202, 1963.

[2] C.T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors," *IEEE Trans. Electron Devices*, **11**, pp. 324-345, 1964.

[3] H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE J. Solid State Circuits*, **SC-3**, 1968.

[4] B.J. Sheu, D.L. Scharfetter, P.-K. Ko, and M.-C. Jeng, "BSIM: Berkeley Short-Channel IGFET Model for MOS Transistors," *IEEE J. Solid-State Circuits*, **SC-22**, pp. 558-566, 1987.

The traditional theory of the MOSFET reviewed in this chapter is the ap-

proach use in textbooks such as the two listed below.

[5] Robert F. Pierret *Semiconductor Device Fundamentals, 2^{nd} Ed.*, , Addison-Wesley Publishing Co, 1996.

[6] Ben Streetman and Sanjay Banerjee, *Solid State Electronic Devices, 6^{th} Ed.*, Prentice Hall, 2005.

For authoritative treatments of classical MOSFET theory, see:

[7] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd Ed., Oxford Univ. Press, New York, 2011.

[8] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, 2nd Ed., Oxford Univ. Press, New York, 2013.

As channel lengths shrunk to the micrometer scale, velocity saturation became important. The following paper from that era discusses the impact on MOSFETs and MOSFET circuits.

[9] C.G. Sodini, P.-K. Ko, and J.L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Dev.*, **31**, pp. 1386 - 1393, 1984.

The virtual source or beginning of the channel is not always exactly at the top of the energy barrier, as discussed by Liu.

[9] Y. Liu, M. Luisier, A. Majumdar, D. Antoniadis, and M.S. Lundstrom, "On the Interpretation of Ballistic Injection Velocity in Deeply Scaled MOSFETs," *IEEE Trans. Electron Dev.*, **59**, pp. 994-1001, 2012.

Lecture 5

MOSFET IV: The virtual source model

5.1 Introduction

In Lecture 4, we developed expressions for the linear and saturation region drain currents as:

$$\begin{aligned} I_{DLIN} &= \frac{W}{L}\mu_n C_{ox}\left(V_{GS}-V_T\right)V_{DS} \\ I_{DSAT} &= WC_{ox}\upsilon_{sat}\left(V_{GS}-V_T\right) \end{aligned} . \tag{5.1}$$

These equations assume $V_{GS} > V_T$, so they cannot describe the subthreshold characteristics. As shown in Fig. 5.1, these equations provide a rough description of I_{DS} vs. V_{DS}, especially if we include DIBL as in eqn. (3.11), so that the finite output conductance is included. If we define the drain saturation voltage as the voltage where $I_{DLIN} = I_{DSAT}$, we find

$$V_{DSAT} = \frac{\upsilon_{sat}L}{\mu_n} . \tag{5.2}$$

For $V_{DS} \ll V_{DSAT}$, $I_{DS} = I_{DLIN}$, and for $V_{DS} \gg V_{DSAT}$, $I_{DS} = I_{DSAT}$.

Traditional MOSFET theory develops expressions for I_{DS} vs. V_{DS} that smoothly transition from the linear to saturation regions as V_{DS} increases from zero to V_{DD} [1, 2]. The goal in this lecture is to develop a simple, semi-empirical expression that describes the complete $I_{DS}(V_{DS})$ characteristic from the linear to saturated region. The approach is similar to the so-called *virtual source MOSFET model* that has been developed and successfully used to describe a wide variety of nanoscale MOSFETs [3]. We'll take a different approach to developing a virtual source model and begin with the traditional approach, and then use the VS model as a framework for subsequent discussions. As we extend and interpret the VS model in subsequent lectures, we'll develop a simple, physical model that provides an accurate quantitative descriptions of modern transistors.

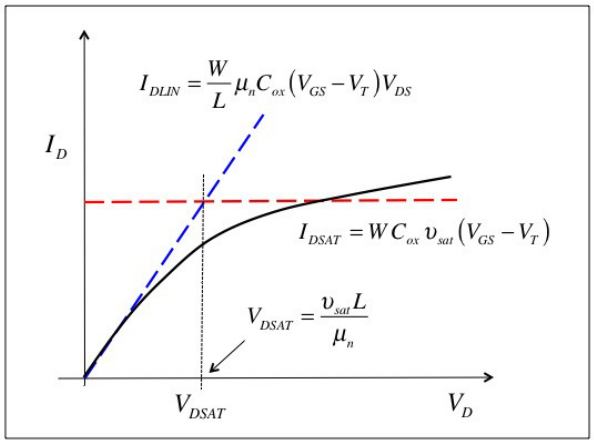

Fig. 5.1 Sketch of a common source output characteristic of an n-channel MOSFET at a fixed gate voltage (solid line). Also sketched as dashed lines are the linear and saturation region characteristics as described by eqns. (5.1).

5.2 Channel velocity vs. drain voltage

The drain current is proportional to the product of charge at the beginning of the channel times the average carrier velocity at the beginning of the channel. From eqn. (4.1) at the beginning of the channel, we have

$$I_{DS}/W = |Q_n(x=0)|\upsilon(x=0) \ . \tag{5.3}$$

Equation (5.1) for the linear current can we re-written in this form as

$$\begin{aligned} I_{DLIN}/W &= |Q_n(V_{GS})|\, \upsilon(V_{DS}) \\ Q_n(V_{GS}) &= -C_{ox}\,(V_{GS} - V_T) \\ \upsilon(V_{DS}) &= \left(\mu_n \frac{V_{DS}}{L}\right) \end{aligned} \quad . \tag{5.4}$$

Similarly, eqn. (5.1) for the saturation current can be re-written as

$$\begin{aligned} I_{DSAT}/W &= |Q_n(V_{GS})|\, \upsilon(V_{DS}) \\ Q_n(V_{GS}) &= -C_{ox}\,(V_{GS} - V_T) \\ \upsilon(V_{DS}) &= \upsilon_{sat} \end{aligned} \quad . \tag{5.5}$$

If we can find a way for the average velocity to go smoothly from its value at low V_{DS} to υ_{sat} at high V_{DS}, then we will have a model that covers the complete range of drain voltages.

The VS model takes an empirical approach and writes the average velocity at the beginning of the channel as [3]

$$\begin{aligned} \upsilon(V_{DS}) &= F_{SAT}(V_{DS})\upsilon_{sat} \\ F_{SAT}(V_{DS}) &= \frac{V_{DS}/V_{DSAT}}{\left[1 + (V_{DS}/V_{DSAT})^{\beta}\right]^{1/\beta}} \end{aligned} , \tag{5.6}$$

where V_{DSAT} is given by eqn. (5.2) and β is an empirical parameter chosen to fit the measured IV characteristic.

The form of the drain current saturation function, F_{SAT}, is motivated by the observation that the lower of the two velocities in eqns. (5.4) and (5.5) should be the one that limits the current. We might, therefore, expect

$$\frac{1}{\upsilon(V_{DS})} = \frac{1}{(\mu_n V_{DS}/L)} + \frac{1}{\upsilon_{sat}} , \tag{5.7}$$

which can be re-written as

$$\upsilon(V_{DS}) = \frac{V_{DS}/V_{DSAT}}{[1 + (V_{DS}/V_{DSAT})]}\upsilon_{sat} . \tag{5.8}$$

Equation (758) is similar to Eqn. (5.6), except that (5.6) introduces the

empirical parameter, β, which is adjusted to better fit data. Typical values of β for n- and p-channel Si MOSFETs are between 1.4 and 1.8 [3].

Equations (5.3), (4.2), and (5.6) give us a description of the above-threshold MOSFET for any drain voltage from the linear to the saturated regions.

5.3 Level 0 VS model

Our simple model for the above threshold MOSFET is summarized as follows:

$$\boxed{\begin{aligned} &I_{DS}/W = |Q_n(0)|\, \upsilon(0) \\ \\ &Q_n(V_{GS}) = 0 \quad V_{GS} \leq V_T \\ &Q_n(V_{GS}) = -C_{ox}\,(V_{GS} - V_T) \quad V_{GS} > V_T \\ &V_T = V_{T0} - \delta V_{DS} \\ \\ &\langle \upsilon(V_{DS}) \rangle = F_{SAT}(V_{DS}) \upsilon_{sat} \\ &F_{SAT}(V_{DS}) = \frac{V_{DS}/V_{DSAT}}{\left[1 + (V_{DS}/V_{DSAT})^{\beta}\right]^{1/\beta}} \\ \\ &V_{DSAT} = \frac{\upsilon_{sat} L}{\mu_n} \end{aligned}} \quad , \tag{5.9}$$

With this simple model, we can compute reasonable MOSFET IV characteristics, and the model can be extended step by step to make it more and more realistic. There are only six device-specific input parameters to this model: C_{ox}, V_T, μ_n, υ_{sat}, L, and β. The level 0 model does not describe the subthreshold characteristics, but after discussing MOS electrostatics in the next few lectures, we will be able to include the subthreshold region. Series resistance is important in any real device, and can be readily included as discussed next.

5.4 Series resistance

As illustrated on the left of Fig. 5.2, we have developed expressions for the IV characteristic of an intrinsic MOSFET — one with no series resistance between the intrinsic source and drain and the metal contacts to which the voltages are applied. In practice, these series resistors are always there and must be accounted for.

The figure on the right in Fig. 5.2 shows how the voltages applied to the terminals of the device are related to the voltages on on the intrinsic contacts. Here, V'_D, V'_S, and V'_G refer to the voltages on the terminals and V_D, V_S, and V_G refer to the voltages on the intrinsic terminals. (No resistance is shown in the gate lead, because we are considering D.C. operation now.) Since the D.C. gate current is zero, a resistance in the gate has no effect. Gate resistance is, however, an important factor in the R.F. operation of transistors.)

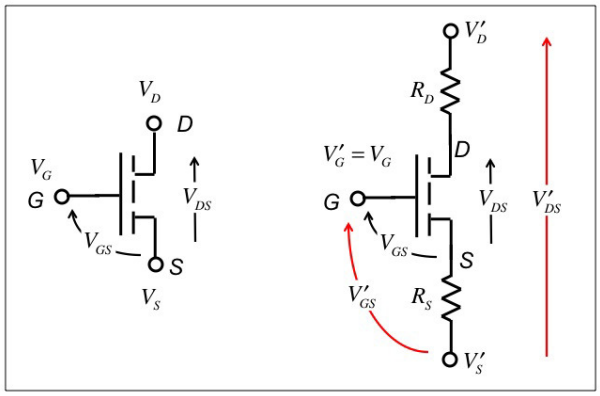

Fig. 5.2 Series resistance in a MOSFET. Left: the intrinsic device. Right: The actual, extrinsic device showing how the voltages applied to the external contacts are related to the voltages on the internal contacts.

From Fig. 5.2, we relate the internal (unprimed) voltages to the external (primed) voltages by

$$V_G = V'_G$$

$$V_D = V'_D - I_{DS}\left(V_G, V_S, V_D\right) R_D \,, \tag{5.10}$$

$$V_S = V'_S + I_{DS}\left(V_G, V_S, V_D\right) R_S$$

Since we know the IV characteristic of the intrinsic device, $I_{DS}\left(V_G, V_S, V_D\right)$, Equations (5.10) are two equations in two unknowns – the internal voltages, V_D and V_S. Given applied voltages on the gate, source, and drain, V'_G, V'_S, V'_D, we can solve these equations for the internal voltages, V_S and V_D, and then determine the current, $I_{DS}\left(V'_G, V'_S, V'_D\right)$.

Figure 5.3 illustrates the effect of series resistance on the IV characteristic. In the linear region, we can write the current of an intrinsic device as

$$I_{DLIN} = \frac{W}{L}\mu_n C_{ox}\left(V_{GS} - V_T\right) V_{DS} = V_{DS}/R_{ch} \,. \tag{5.11}$$

When source and drain series resistors are present, the linear region current becomes

$$I_{DLIN} = V_{DS}/R_{tot} \,, \tag{5.12}$$

where

$$R_{tot} = R_{ch} + R_S + R_D = R_{ch} + R_{DS} \,. \tag{5.13}$$

(It is common to label the sum of R_S and R_D as R_{SD}). So the effect of series resistance in the linear region is to simply lower the slope of the IV characteristic as shown in Fig. 5.3.

Figure 5.3 also shows that series resistance decreases the value of the saturation region current. In an ideal MOSFET with no output conductance, the drain series resistance has no effect in the saturation region where $V_D > V_{DSAT}$, but the source resistance reduces the intrinsic V_{GS}, so eqn. (5.1) becomes

$$I_{DSAT} = W C_{ox} \upsilon_{sat}\left(V_{GS} - I_{DSAT} R_S - V_T\right) \,. \tag{5.14}$$

Series resistance lowers the internal gate to source voltage of a MOSFET, and therefore lowers the saturation current. The maximum voltage applied

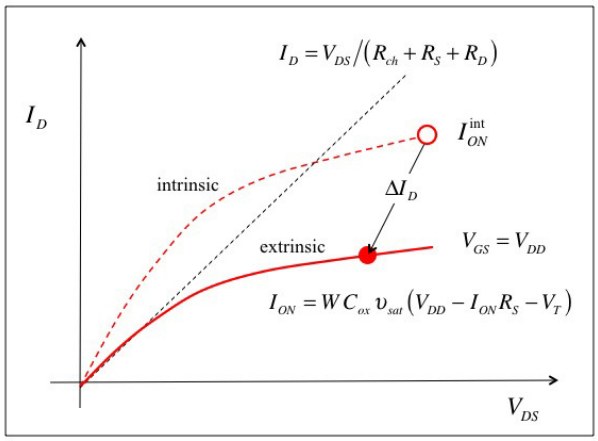

Fig. 5.3 Illustration of the effect of series resistance on the IV characteristics of a MOSFET. The dashed curve is an intrinsic MOSFET for which $R_S = R_D = 0$. As indicated by the solid line, series resistance increases the channel resistance and lowers the on-current.

between the gate and source is the power supply voltage, V_{DD}. Series resistance will have a small effect if $I_{DSAT}R_S \ll V_{DD}$. For high performance, we require

$$R_S \ll \frac{V_{DD}}{I_{DSAT}}\,. \tag{5.15}$$

Modern Si MOSFETs deliver about 1 mA/μm of on-current at $V_{DD} = 1$ V. Accordingly, R_S must be much less than 1000 $\Omega - \mu$m; series resistances of about 100 $\Omega - \mu$m are needed. Although we will primarily be concerned with understanding the physics of the intrinsic MOSFET, we should be aware of the significance of series resistance when analyzing measured data. As channel lengths continue to scale down, keeping the series resistance to a manageable level is increasingly difficult.

Exercise 5.1: Analysis of Experimental Data

Use eqn. (5.14) and the *IV* characteristic of Fig. 4.3, to deduce the "saturation velocity" for the on-current. Note that we'll regard the saturation velocity as an empirical parameter used to fit the data of Fig. 4.3 and will compare it to the high-field saturation velocity for electrons in bulk Si.

Assume the following parameters:

$$\begin{aligned} I_{ON} &= 1180 \quad \mu\text{A}/\mu\text{m} \\ C_{ox} &= 1.55 \times 10^{-6} \quad \text{F/cm}^2 \\ R_{DS} &= 220 \quad \Omega \\ V_T &= 0.25 \quad \text{V} \\ V_{DD} &= 1.2 \quad \text{V} \\ W &= 1 \quad \mu\text{m} \end{aligned} \quad ,$$

Solving eqn. (5.14) for v_{sat}, we find

$$v_{sat} \equiv v_{inj} = \frac{I_{DSAT}}{WC_{ox}\left(V_{GS} - V_T\right)} \, .$$

$$V_{GS} = V_{DD} - I_{DSAT}R_{SD}/2 \, .$$

Putting in numbers, we find

$$v_{sat} = 0.92 \times 10^7 \quad \text{cm/s} \, .$$

It is interesting to note that the velocity we deduce is close to the high-field, butlk saturation velocity of Si (1×10^7 cm/s), but the physics of velocity saturation in a nanoscale MOSFET is actually quite different from the physics of velocity saturation in bulk Si under high electric fields. Accordingly, from now on, we will give v_{sat} a different name, the *injection velocity*, v_{inj}.

5.5 Discussion

One might have expected the traditional model that we have developed to be applicable only to long channel MOSFETs because it is based on assumptions such as diffusive transport in the linear region and high-field velocity saturation in the saturated region. Surprisingly, we find that it accurately describes the IV characteristics of MOSFETs with channel lengths less than 100 nm as shown in Fig. 5.4. To achieve such fits, we view two of the physical parameters in our VS model as empirical parameters that are fit to measured data, and we find that with relatively small adjustments in these parameters, excellent fits to most transistors can be achieved. The fact that this simple traditional model describes modern transistors so well, tells us that it captures something essential about the physics of MOSFETs.

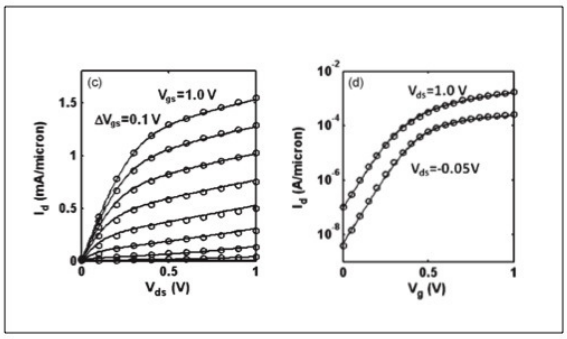

Fig. 5.4 Measured and fitted VS model data for 32 nm n-MOSFET technology. Left: Common source output characteristic. Right: Transfer characteristic. The VS model used for these fits is an extension of the model described by Eqns. (7.9) that uses an improved description of MOS electrostatics to treat the subthreshold as well as above threshold conduction. (From [3].)

5.6 Summary

In this lecture we recast traditional MOSFET theory in the form of a simple virtual source model. Application of this simple model to modern transistors shows that it describes them remarkably well. This is a consequence of the fact that it describes the essential features of the barrier controlled model of a transistor (i.e. MOS electrostatics). The weakest part of the model is the transport model, which is based on the use of a mobility and saturated velocity. Because of the simplified transport model, we need to regard the mobility and saturation velocity in the model as fitting parameters that can be adjusted to fit experimental data.

In the next few lectures (Sec. 2 of the series of lectures), we will review MOS electrostatics and learn how to describe subthreshold as well as above-threshold conduction. The result will be an improved VS model, but mobility and saturation velocity will still be viewed as fitting parameters. Beginning in Sec. 3, we'll discuss transport and learn how to formulate the VS model so that transport is described physically.

5.7 References

For a thorough treatment of classical MOSFET theory, see:

[1] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, 3rd Ed., Oxford Univ. Press, New York, 2011.

[2] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, 2nd Ed., Oxford Univ. Press, New York, 2013.

The MIT Virtual Source Model, which provides a framework for these lectures, is described in:

[3] A. Khakifirooz, O.M. Nayfeh, and D.A. Antoniadis, "A Simple Semiempirical Short-Channel MOSFET CurrentVoltage Model Continuous Across All Regions of Operation and Employing Only Physical Parameters," *IEEE Trans. Electron. Dev.,* **56**, pp. 1674-1680, 2009.