Chapter 9

The "Long" Metal-Oxide-Semiconductor Field-Effect Transistor

version 2.4 (updated: July 31, 2012), ©Prentice Hall

Outline

The Metal-Oxide-Semiconductor Field-Effect Transistor or MOSFET is arguably the most important semiconductor device ever invented. Integrated circuits based on MOSFETs are ubiquitous in our modern society. There are many that we readily see, such as microprocessors in computers or digital-signal processors in wireless phones, but there are many more that we do not see, such as chips embedded in home appliances, industrial tools, toys, cars, etc. As a result of the explosion of logic ICs, MOSFET technology has become a commodity: it is widely standardized and it is readily available to IC designers around the world. This has fueled, in turn, the widespread use of MOSFETs in many other applications that go beyond logic, such as analog and communications systems. Current examples are wireless LANs, cell phones, and optical fiber transceiver modules, among many others.

There have been many contributors to the conception and development of the MOSFET. Ross postulated in 1955 that an inversion layer could be induced electrostatically by an electrode placed close to the surface of a semiconductor. The first reduction to practice of the MOSFET had to wait until SiO_2 with sufficient quality was developed. The first MOSFET was demonstrated in 1959 by Dawon Kahng and Martin Atalla at Bell Labs. The integrated MOSFET and MOSFET IC's quickly followed. While these represented huge advances in microelectronics, a truly revolutionary event was the invention of CMOS in 1963 by Frank Wanlass at Fairchild. CMOS refers to Complementary MOS, or the pairing of an n-channel MOSFET (based on electrons) and a p-channel MOSFET (based on holes) to form a logic gate. The distinctive feature of CMOS, in contrast with other logic families, is that it performs logic while consuming no DC power. This unique characteristic, perhaps more than anything else, underpins the microelectronics revolution and has led to ultra-large scale integrated circuits. More details about the evolution of MOSFET design and key milestones along the way are given in Sec. 10.5.6.

A schematic cross section of a modern integrated *n-channel* MOSFET is shown in Fig. 9.1. At its heart, the MOSFET consists of a metal-oxide-semiconductor structure with an n^+-region on each side. The metal of the MOS structure is referred to as *gate*. The two n^+-regions are referred to as *source* and *drain*. The p-type well surrounding the device is referred to as *body*. A MOSFET is a four-terminal device. The gate, source, drain, and body regions are all contacted separately. In modern technology, the gate is made out of n^+-polySi and it is typically silicided on top to reduce its resistance. The surface of the source and drain, as well as the body contact, are all silicided for the same reason. To facilitate electrical contact, a p^+-region is implanted at the surface of the body contact. The device is isolated through shallow trenches that cut below the heavily-doped regions of the source, drain, and body. In the planar view shown at the bottom of Fig. 9.1 two key dimensions are defined. *Gate length* and *gate width* refer to the length and width of the intrinsic portion of the device defined by the overlap of the polySi gate and the area enclosed by the shallow trench isolation. These are critical dimensions with a major impact on most electrical figures of merit of the MOSFET. [1]

The complementary p-channel MOSFET looks similar with p and n regions interchanged. Its operation is similar to that of the n-channel MOSFET but the signs of voltages and currents are reversed. Since it is based on hole transport, the p-channel MOSFET has lower performance.

The heart of the MOSFET is the intrinsic region, also referred to as *channel* region. This

[1] Notice the odd choice of notation: the width of the gate tends to be larger than its length. This is for historical reasons.

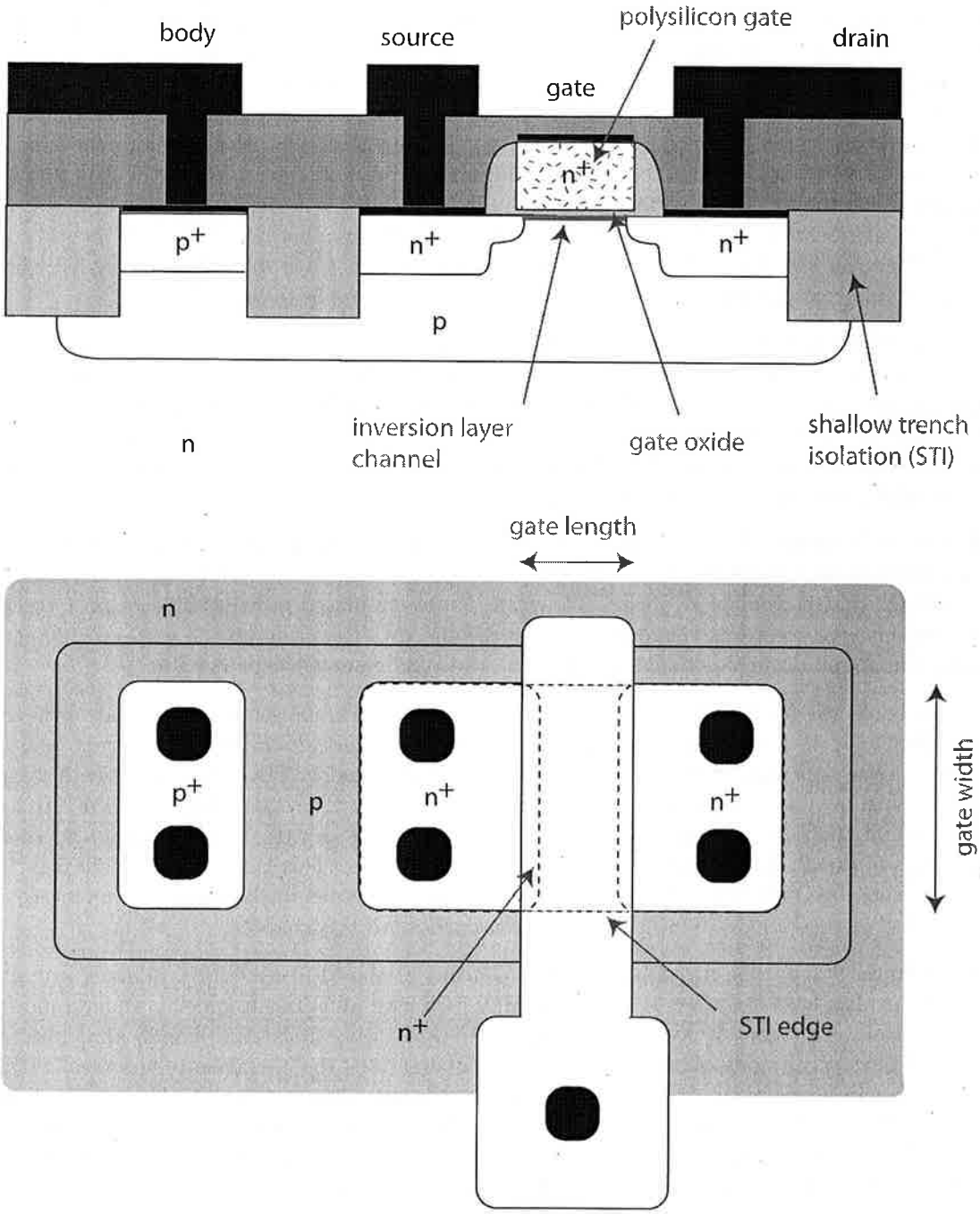


Figure 9.1: Simplified schematic diagram of a modern MOSFET: cross section (top), layout at the wafer surface (bottom).

is the MOS portion of the device directly underneath the gate between source and drain. The rest is often referred to as the extrinsic portions of the MOSFET. When a gate voltage in excess of the threshold voltage is applied to the gate of the MOS structure, an inversion layer forms in the semiconductor right below the oxide/semiconductor interface. By virtue of the existence of a finite overlap of the gate above the n^+ source and drain regions (this is very important to insure low parasitic resistance), the inversion layer creates a conducting channel that connects the source and the drain. If a positive voltage is applied to the drain with respect to the source, electrons will flow from the source to the drain through the inversion layer. When the gate voltage is brought below the threshold voltage, there is no inversion layer in the MOS structure and the conducting path between source and drain is broken. In this manner, the MOSFET behaves as a switch. Electron flow from source to drain is controlled by the gate voltage; it is enabled when the gate voltage is above threshold and it is forbidden in the opposite case.

The MOSFET can also behave as an amplifier. With the gate voltage above threshold and an inversion layer connecting source and drain, the higher the gate voltage, the more electrons are induced in the inversion layer, and the higher the current that flows between source and drain. As we will study in this chapter, under the right conditions, the current between the source and drain can be made independent of the drain voltage. This "isolation between input and output" is an essential property of a good amplifying device.

A word on notation. In most cases, a MOSFET features a rather symmetrical design. The source and drain are interchangeable. What makes the source and the drain be identified as such is the voltage that is applied to them. The drain is always biased positive with respect to the source. In this way, electrons flow out of the source and into the drain. If the voltage difference between source and drain was to change its sign, we would change the names too.

In this book, the MOSFET is presented in two chapters. The present chapter deals with the "long" MOSFET, while the next one is concerned with the "short" MOSFET. The terms "short" and "long" refer to length of the gate of the transistor (marked in Fig. 9.1). As we will learn, these are somehow relative terms to other key dimensions of the device. Modern MOSFETs are very small and they are getting smaller all the time. Because of this, understanding the short MOSFET is of great importance for microelectronics engineers. This is the topic of Chapter 10. Before the short MOSFET can be productively attacked, one has to develop a clear picture of the operation of the "long" MOSFET. This is the purpose of this chapter.

The chapter starts with a definition of the concept of the ideal MOSFET. This is a useful simplification that captures the essence of the MOSFET and allows us to quickly understand its basic operation. The ideal MOSFET is amenable to developing first-order models that provide substantial physical insight. Section 9.2 qualitatively describes the operation of the ideal MOSFET using a simple water analogy. Basic physics of electron transport in the inversion layer are discussed in Section 9.3. Section 9.4 proceeds to describe the current-voltage characteristics of the ideal MOSFET and develops a simple analytical model for it. Section 9.5 presents the charge-voltage characteristics of the ideal MOSFET. After that, Section 9.6 discusses its small-signal behavior in the saturation regime which is the most useful regime of operation. The chapter ends with a section on the most important non-ideal effects of the long MOSFET.

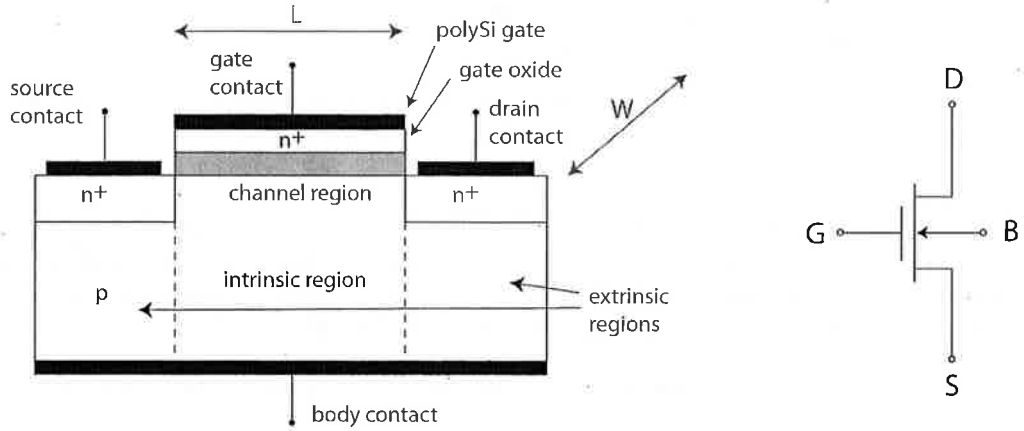


Figure 9.2: Left: sketch of the cross section of an ideal MOSFET. This conceptual device is a good basis for developing first-order understanding and models for MOSFET operation. Right: Circuit symbol for a MOSFET.

9.1 The ideal MOSFET

As in other chapters in this book, it is of great usefulness to define the concept of the ideal MOSFET. This is a model device that captures the essence of the MOSFET but hides away complicating details and second-order effects. The ideal MOSFET is a useful framework on top of which to construct a solid understanding and a useful set of models for the MOSFET. This will be of great value in this chapter about the long MOSFET, but also in Ch. 10 when studying the short MOSFET.

A sketch of the ideal MOSFET is shown in Fig. 9.2. This is a perfectly symmetrical device. It consists of uniformly doped source, drain, gate, and body regions. The edges of the source and drain n^+-regions are perfectly lined up with the edges of the MOS structure. The body contact is made at the bottom of the device.

The ideal MOSFET contains a number of simplifying assumptions:

- The MOS structure is ideal as defined in Ch. 8.
- All carrier flow is one dimensional.
- Doping levels are uniform throughout.
- We assume Maxwell-Boltzmann statistics (non degenerate).
- Electron transport along the inversion layer takes place only by drift (*i.e.*, we neglect diffusion).
- Electrons drift along the inversion layer in the mobility regime, *i.e.*, the electron velocity is proportional to the lateral electric field along the inversion layer. We undo this assumption in Ch. 10.

- We neglect the *body effect*, that is, the dependence of the threshold voltage on the space coordinate along the channel. This is fixed in Sec. 9.7.1 .
- We ignore any resistance effects associated with the quasi-neutral regions or the ohmic contacts to these regions. The impact of parasitic source and drain resistance is discussed in Sec. 9.7.5.
- We ignore any effects associated with the sidewall of the source-body and drain-body junctions. This implies that the extent of the depletion region underneath the MOS structure is assumed to be unaffected by the presence of the source and drain regions. The impact of this on the electrostatics of the channel is discussed in Sec. 10.2.
- We ignore the saturation current of the reverse biased source/body and drain/body PN junctions.
- There are no three-dimensional effects, that is, the device scales perfectly with its width.
- We ignore any effects associated with the substrate that surrounds the transistor (not shown in Fig. 9.2), including the presence of a PN junction with the MOSFET body with rectifying I-V characteristics.
- We neglect impact ionization anywhere in the device. Its role is discussed in Sec. 10.4.2.

Fig. 9.3 shows a cross section of the ideal MOSFET indicating the coordinate axis convention used here. The x axis is selected into the wafer with its origin at the semiconductor surface. The y axis is selected along the channel from source to drain with its origin at the source edge. Indicated is also the inversion layer and the depletion region that under the right conditions exist underneath the source, channel, and drain regions of the device. This figure suggests that the depletion regions of the source and drain merge with that of the channel at the edges of the channel. This is obviously the case. However, in our ideal MOSFET we assume that the depletion region underneath the channel is of uniform thickness throughout. That is, we neglect any effect of the source and drain sidewalls.

Fig. 9.3 also shows the voltage and current notations that are used in this book. All voltages are referred to the source. As usual, entering terminal currents are positive. In our ideal MOSFET, the gate and body currents are assumed zero. Hence, the source current is equal to minus the drain current.

9.2 Qualitative operation of the ideal MOSFET

The operation of the MOSFET can be qualitatively understood by establishing an analogy with a water reservoir system, as sketched in Fig. 9.4. In this analogy, the source and drain can be thought of as water reservoirs with a relative height that can be adjusted. The gate plays the role of barrier separating the two water reservoirs. The relative height of the two reservoirs of water corresponds to the drain to source voltage. The higher V_{DS}, the *lower* the level of the water on the drain is with respect to that of the source. The relative height of the barrier above the water

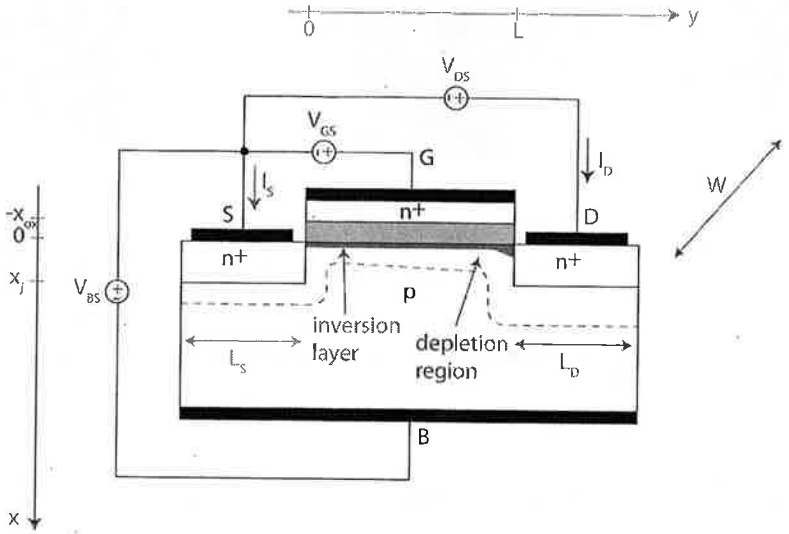


Figure 9.3: Sketch of the cross section of an ideal MOSFET defining spatial coordinates as well as voltage and current notation. Shown are also sketches of the inversion layer and the depletion region that under appropriate conditions exist underneath the source, gate and drain.

level of the source corresponds to the gate to source voltage in excess of threshold. The higher V_{GS}, the lower the gate barrier. At a value of $V_{GS} = V_T$, the top of the barrier is level with the water surface. In this analogy, we ignore the effect of the MOSFET body.

The water analogy of the MOSFET reveals three possible regimes of operation. These are sketched in Fig. 9.5. When the barrier level is above the water level on the source, water cannot flow between source and drain. This is shown on the left of this figure. This occurs regardless of the relative levels of the source and drain water surfaces. In the MOSFET, this corresponds to the *cut-off* regime with $V_{GS} < V_T$. In the cut-off regime, there is no inversion layer under the gate and $I_D = 0$ regardless of the value of V_{DS}.

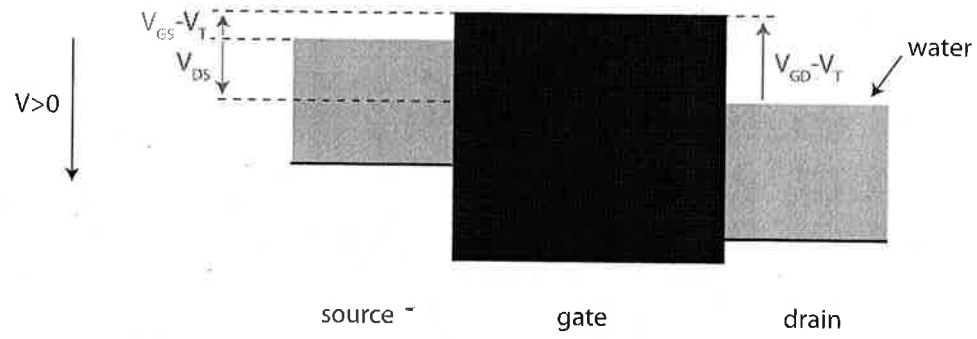


Figure 9.4: Sketch of ideal MOSFET and a water analogy. The source and drain can be thought of as pools of water separated by a barrier. The drain to source voltage is equivalent to the relative heights of the water levels in the two pools of water. The gate to source voltage above threshold is equivalent to the relative height of the barrier above the water level on the source.

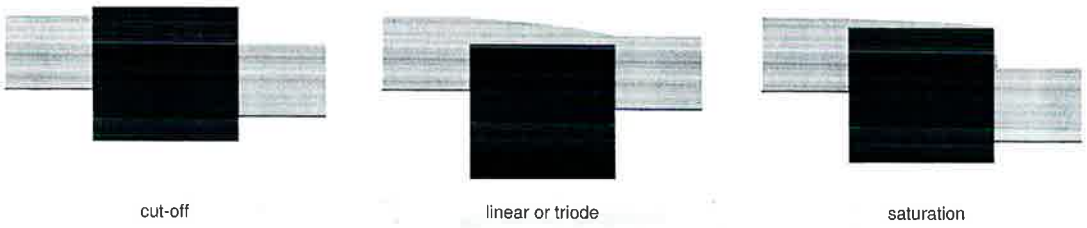


Figure 9.5: Regimes of operation of the water analogy of the MOSFET. The drain water level is always below the source water level. In the cut-off regime (left), the barrier is above the water level on the source and drain and water does not flow. In the triode or linear regime (middle), the barrier is below the water level of both the source and drain. Water flows from source to drain. In the saturation regime (right), the barrier is below the water level of the source but above that of the drain. Water flows from source to drain, but the water current is independent of the relative levels of source and drain.

A second regime of operation corresponds to the case depicted in the middle of Fig. 9.5. When the barrier is below the water levels of the source *and* drain, water flows above the barrier from one reservoir to the other. It is clear that in this case, the higher the water level difference between source and drain, or the lower the barrier level with respect to the water surfaces, the higher the current too. In the MOSFET, this situation corresponds to the *linear* regime, also called the *triode* regime. In this mode of operation, $V_{GS} > V_T$ and $V_{GD} > V_T$. As a consequence, an inversion layer is formed below the gate that constitutes a path that connects the source and drain. With $V_{DS} > 0$, an electric field is created along the inversion layer and electrons drift from source to drain. As in the water analogy, the current flowing through the channel depends both on V_{DS} and V_{GS}. If V_{DS} increases, the lateral electric field along the inversion layer increases and the current increases. If V_{GS} increases, the amount of electrons induced in the inversion layer increases and the current also increases.

Finally, the third regime of operation of the MOSFET is depicted on the right of Fig. 9.5. In this regime, the top of the barrier is below the water level on the source but above the water level on the drain. Water still can flow over the barrier, but there is a key difference with the regime in the middle of the figure. The water flow is independent of the relative height of the water level on the drain side with respect to the source. As the water flows over the barrier, when it reaches the end, it simply falls down freely. Lowering the water level at the drain does not increase the flow of water.

In the MOSFET, this regime of operation corresponds to a situation in which $V_{GS} > V_T$ but $V_{GD} < V_T$. As we will see, this creates an inversion layer at the surface of the semiconductor that thins down along the channel. As in the water analogy, the electron velocity increases along the channel from source to drain. At the drain side of the channel, it is said that the inversion layer is *pinched off*. The electrons simply "drop" into the drain. The remarkable aspect of this regime is that once V_{DS} is high enough for this regime to be established, the drain voltage has no further impact on the electric field that is set up along the channel. Beyond this value of V_{DS}, referred to as V_{DSsat}, the electric field in the channel is only affected by V_{GS}. In consequence, the channel current is independent of V_{DS}. This is called the *saturation* regime and it is the most important regime of operation of the MOSFET. It is in the saturation regime where a MOSFET is biased as an amplifier.

Armed with this qualitative understanding, we are now in a position to start looking in a more rigorous manner at the physics of the MOSFET in its various regimes of operation and start developing models for it. The next section discusses the physics of electron transport along the inversion layer which is at the heart of the MOSFET operation.

9.3 Inversion layer transport in the ideal MOSFET

The first step in the construction of a model for the ideal MOSFET is to develop a suitable formulation for electron transport in the inversion layer. While this is fairly straightforward, there are some subtleties to it. A detailed study is presented in Advanced Topic AT9.1. This section captures the most important results.

In the inversion layer of a MOSFET under normal operation, electrons are trapped in a potential barrier that prevents them from escaping to the gate or the body. In consequence, electron transport only takes place laterally along the surface of the semiconductor from source to drain, right under the gate oxide. In this situation, there is no compelling interest in describing the behavior of electrons as a function of the *normal* direction into the semiconductor (the x direction). The vertical dimension can be abstracted away by integrating the electron concentration across the inversion layer. So, instead of thinking of $n(x, y)$, the electron concentration as a function of location in x and y at any one location in the inversion layer, it is much more productive to think of the *sheet carrier concentration* in the inversion layer, $n_s(y)$, at a particular location along the channel, where:

$$n_s(y) = \int_0^\infty n(x, y)dx \tag{9.1}$$

n_s has units of cm^{-2}. This integrates all electrons in the inversion layer at a location y and ignores their detailed distribution in x. The upper integration limit in this integral can be extended to ∞ without introducing any problems because the electron concentration peaks at the surface and drops very fast with x.

With the MOSFET biased in one of its conducting regimes, the electron current along the inversion layer can be easily expressed in terms of n_s. In the ideal MOSFET, we assume that electron flow takes place by drift and we neglect any electron diffusion (see Advanced Topic AT9.1). Under this assumption, the electron current is simply given by the product of the electron charge, the electron velocity, and the sheet electron concentration, times the width of the device:

$$I_e \simeq -qWv_{ey}(y)n_s(y) \tag{9.2}$$

where v_{ey} is the average velocity for the electrons in the inversion layer in the y direction from source to drain. Note that while v_{ey} and n_s depend on y, I_e does not. This is a consequence of the fact that electrons cannot escape from the channel and must all flow from source to drain.

Eq. 9.2 can also be written in terms of the *sheet charge density* of the inversion layer, Q_i. This is related to n_s through:

$$Q_i(y) = -qn_s(y) \tag{9.3}$$

In terms of Q_i, Eq. 9.2 becomes:

$$I_e \simeq W v_{ey}(y) Q_i(y) \tag{9.4}$$

In more rigorous terms, the key assumption that led us to Eqs. 9.2 and 9.4 is called the *sheet-charge approximation* (SCA). This approximation makes it physically meaningful to define an *average* lateral velocity for all the electrons in the inversion layer. Intuitively, this is possible whenever the distribution of lateral velocities in depth does not change too rapidly in the scale of the changes that are taking place in the electron concentration. It essentially requires that the inversion layer be very thin in the scale of other normal dimensions of the MOS structure (the oxide thickness and the thickness of the depletion region underneath the inversion layer). This is discussed in detail in Advanced Topic AT9.1.

According to our definition of the ideal MOSFET, we consider that the lateral electric field along the inversion layer is small enough for electrons to drift in the mobility regime. In this regime, the lateral electron velocity is proportional to the lateral electric field. As it turns out, this assumption is quite poor in MOSFETs of any gate length. Nevertheless, it is important and useful to proceed with this assumption for a while as the resulting models are an excellent starting point for a more detailed study of the MOSFET. We will undo this assumption and study the impact of velocity saturation in Ch. 10. In the mobility regime, we then have:

$$v_{ey}(y) \simeq -\mu_e \mathcal{E}_y(y) \tag{9.5}$$

where $\mathcal{E}_y(y)$ is the average longitudinal electric field in the inversion layer. In this regime, Eq. 9.4 becomes:

$$I_e \simeq -W \mu_e \mathcal{E}_y(y) Q_i(y) \tag{9.6}$$

An alternate way to express Eq. 9.6 is in terms of the voltage $V(y)$ of the inversion layer. Since the source is our reference for all voltages, we define V as the surface potential at a location y with respect to the surface potential at the source-end of the channel:

$$V(y) = \phi_s(y) - \phi_s(y = 0) \tag{9.7}$$

The lateral electric field in the inversion layer can now be expressed in terms of V:

$$\mathcal{E}_y(y) = -\frac{d\phi_s(y)}{dy} = -\frac{dV(y)}{dy} \tag{9.8}$$

Inserting Eq. 9.8 into 9.6 yields:

$$I_e = W\mu_e Q_i(y)\frac{dV(y)}{dy} \tag{9.9}$$

To make further progress we need to find a way to relate $Q_i(y)$ with $V(y)$. The *gradual-channel approximation* (GCA) allows us to do that. Before we introduce this approximation, it is necessary to remember the fundamental charge control relationship of the inversion layer in inversion. This was studied in Chapter 8. In a two-terminal MOS structure under inversion, the charge in the inversion layer is to the first order determined by the voltage applied to the gate with respect to the source in excess of the threshold voltage:

$$Q_i = -C_{ox}(V_{GS} - V_T) \tag{9.10}$$

This expression was derived for a MOS structure with the source as voltage reference and with the body tied up to the source ($V_S = V_B = 0$). In this case, the inversion layer voltage is zero everywhere and as a consequence, the inversion layer charge is uniform across the structure. In a MOSFET, unlike the simpler two-terminal MOS structure, this expression only applies at the source end of the channel, where $V = 0$. Anywhere else down the channel, V is different from zero and this expression needs to be reformulated.

The GCA allows us to recycle expression 9.10 to situations in which V changes in space but it does so relatively slowly. The key assumption is that at any one location, the inversion layer charge is only set by the local *vertical* electrostatics with the *lateral* electrostatics having a negligible effect on it. This means that we can reuse Eq. 9.10 provided that we subtract the *local* voltage of the inversion layer, $V(y)$, which in general is not zero. This is sketched in Fig. 9.6. This figure shows that the voltage difference between the gate and the inversion layer at a location y is in general $V_{GS} - V(y)$, instead of simply V_{GS}. The inversion layer charge at that location, under the GCA is then given by:

$$Q_i(y) \simeq -C_{ox}[V_{GS} - V(y) - V_T] \tag{9.11}$$

The GCA effectively breaks up the two-dimensional electrostatics of the MOSFET into two simpler one-dimensional problems: the vertical electrostatics control the charge in the inversion layer, while the lateral electrostatics control its lateral flow along the channel. In the GCA, the lateral electrostatics represent a relatively "small" perturbation of the vertical electrostatics. As a consequence, the inversion layer charge at any one location can be computed using the one-dimensional MOS theory developed in the previous chapter, with the simple precaution of using the *local* inversion layer voltage, $V(y)$, instead of $V = 0$.

Advanced Topic AT9.1 discusses the GCA and the SCA in more detail and explores the limits of applicability of these two approximations. It is shown that for the GCA to apply, the lateral electric field must be smaller than an "effective" vertical electric field. In turn, this demands that the aspect ratio of the device be large enough, that is, that the channel length be sufficiently

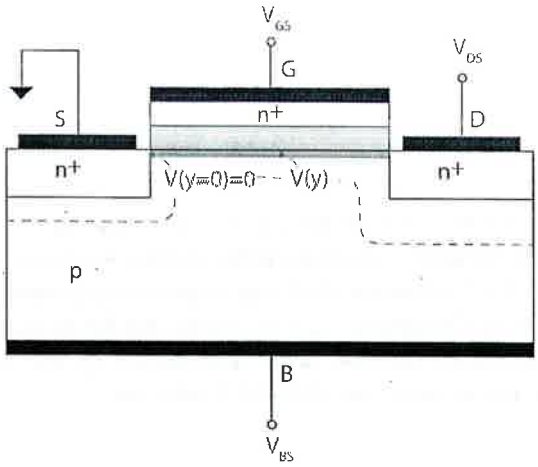


Figure 9.6: Schematic diagram of a MOSFET illustrating local gate voltage overdrive.

longer than a vertical characteristic length. In AT9.1 it is also shown that if the GCA is fulfilled, the SCA also applies.

Eq. 9.11 allows us to formulate the MOSFET current equation in terms of a single variable, $V(y)$. Plugging Eq. 9.11 into 9.9, we get:

$$I_e = -W\mu_e C_{ox}[V_{GS} - V_T - V(y)]\frac{dV(y)}{dy} \tag{9.12}$$

This is a first-order differential equation in terms of $V(y)$. Its solution with appropriate boundary conditions yields the channel current in the ideal MOSFET under a variety of conditions.

9.4 Current-voltage characteristics of the ideal MOSFET

In the previous section, we developed a quasi-1D formulation for inversion layer transport in a MOSFET. This is the basis for a first-order formulation of the current-voltage characteristics of the MOSFET under the cut-off, linear and saturation regimes. This is carried out in the next three subsections.

9.4.1 The cut-off regime

The cut-off regime is defined for values of $V_{GS} < V_T$ and $V_{DS} \geq 0$. In a simple first-order formulation, since V_{GS} is below threshold, there is no inversion layer and the current is zero, regardless of the value of V_{DS}. In the cut-off regime, the transistor is *off*. This is one of the two logic states in which a MOSFET is biased when operating as a switch.

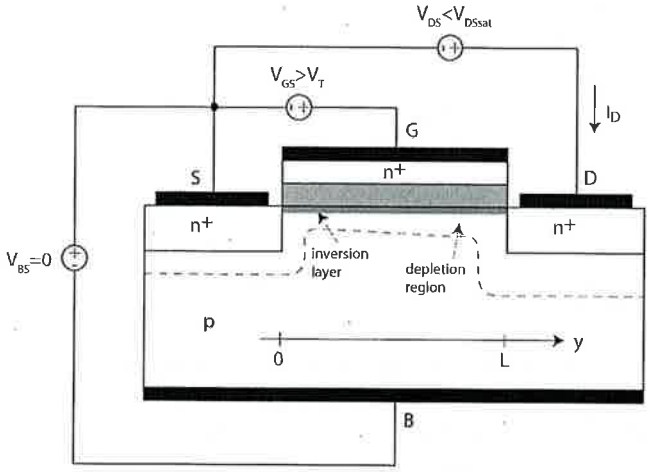


Figure 9.7: Sketch of MOSFET in linear regime and with no back-bias.

To the second order, we know that below threshold, there are still electrons at the surface of the semiconductor. We have referred to it as the weak inversion regime of the MOS structure or the subthreshold regime. As a consequence, even for $V_{GS} < V_T$, there will be a small amount of current flowing through the MOSFET. This is the *subthreshold regime* of operation of the MOSFET and it will be discussed in Section 9.7.4 below.

9.4.2 The linear regime

The linear regime is defined for values of $V_{GS} > V_T$ and $V_{GD} > V_T$. This situation is sketched in Fig. 9.7. Under these conditions, an inversion layer extends all the way under the gate from source to drain, a lateral field is set up along the inversion layer, and current flows.

With our current understanding, we should expect the drain current in the linear regime to increase with V_{GS} and V_{DS}, but for different reasons. All things being equal, if V_{GS} increases, the inversion layer charge increases in absolute magnitude and the channel current must increase along with it. On the other hand, if V_{DS} increases, the electric field along the channel is also enhanced as a result, and more channel current ought to flow. This behavior resembles that of the triode vacuum tube and for this reason, the linear regime is also referred to as the *triode* regime.

Deriving an expression for the current-voltage characteristics of the MOSFET in the linear regime is fairly straightforward. We start where we left off in the previous section, with Eq. 9.12. Solving this differential equation in this case is rather easy. By bringing dy to the left-hand side, variable separation is achieved:

$$I_e dy = -W\mu_e C_{ox}(V_{GS} - V_T - V)dV \tag{9.13}$$

We can then integrate the left-hand side along the channel from $y = 0$ to $y = L$, and the

right-hand side from $V = 0$ to $V = V_{DS}$, to get:

$$I_e \int_0^L dy = -W\mu_e C_{ox} \int_0^{V_{DS}} (V_{GS} - V_T - V)dV \tag{9.14}$$

or, after integration,

$$I_e = -\frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS} \tag{9.15}$$

Going from channel current to drain terminal current is simple. Since electrons flow through the inversion layer from source to drain, this must necessarily result in an *entering* drain terminal current which, under standard sign convention, is positive. Since in our choice of y axis, I_e is negative, we therefore need to apply a negative sign. The drain current is then:

$$I_D = \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS} \tag{9.16}$$

This equation gives the current-voltage characteristics of the ideal MOSFET in the linear regime. Before graphing it for different values of V_{DS} and V_{GS}, we must understand its limits of applicability. Eq. 9.16 was derived under the assumption that the inversion layer extends underneath the entire channel. However, the charge distribution in the inversion layer is not uniform along the channel: it drops in absolute magnitude from source to drain. This is because V increases along the inversion layer from 0 to V_{DS} and, as a result, Q_i is reduced in absolute magnitude (see Eq. 9.11). This is discussed in detail later on in this section. The lowest inversion layer charge occurs at the drain end of the device where it is given by:

$$Q_i(y = L) = -C_{ox}(V_{GS} - V_T - V_{DS}) \tag{9.17}$$

For $|Q_i(y = L)| > 0$, $V_{DS} < V_{GS} - V_T$ must be satisfied. This is the restriction on Eq. 9.16.

Under this constraint, Eq. 9.16 is graphed in Fig. 9.8 which sketches I_D as a function of V_{DS} for different values of V_{GS}. The linear regime is the area of this graph comprised by the vertical axis and the line labeled "$V_{DS} = V_{GS} - V_T$." There are several features in this graph that are worth noting. First, for a given value of V_{GS}, I_D increases with V_{DS}. Initially the increase is relatively fast but then it appears to saturate. The reason for the increase was mentioned before: a higher V_{DS} results in a larger lateral field and a larger current. The reason for the saturation will be discussed below. Second, for a given value of V_{DS}, I_D increases with V_{GS}. This was also expected. A higher V_{GS} induces a higher concentration of electrons in the inversion layer that results in increased current. Finally, for $V_{GS} = V_T$, $I_D = 0$, as it is to be expected since the inversion layer disappears. Notice that for $V_{GS} = V_T$, the model only applies for $V_{DS} = 0$ for which $I_D = 0$ anyway.

In order to better understand the physics of the MOSFET in the linear regime, in particular, its saturating behavior as V_{DS} increases, Fig. 9.9 sketches $|Q_i(y)|$, $|\mathcal{E}_y(y)|$, $V(y)$, and $V_{GS} - V(y)$

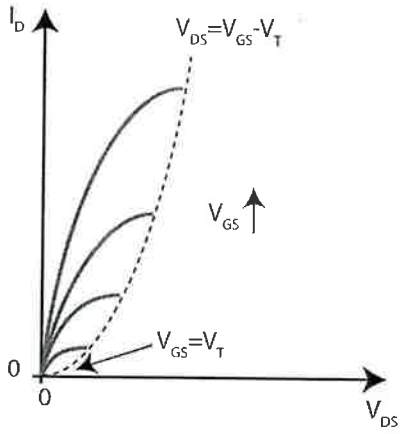


Figure 9.8: Sketch of $I-V$ characteristics of a MOSFET in the linear regime. Each line corresponds to a different value of V_{GS}.

along the channel of the MOSFET from source to drain. The mathematical expressions for these parameters as a function of y are easy to derive (see Problem 9.2). As Fig. 9.9 indicates, Q_i drops in absolute magnitude along the channel from source to drain. This is because V increases along the channel from a value of zero at the source, to V_{DS} at the drain. This "debiases" the inversion layer as we proceed from source to drain. In order to maintain a constant current, $\mathcal{E}_y(y)$ must increase in magnitude from source to drain so that there is an increasing electron velocity. This yields a voltage distribution along the channel that is superlinear in y. The amount of local "gate overdrive," $V_{GS} - V(y) - V_T$, is reduced along the channel from source to drain.

It is illuminating to consider how the profiles of Fig. 9.9 change as V_{DS} increases. This is shown in Fig. 9.10. For $V_{DS} = 0$, the lateral field across the inversion layer is zero and the electron concentration is uniform. For low values of V_{DS}, a nearly uniform electric field is set along the channel and the voltage distribution is nearly linear. As V_{DS} increases, the lateral electric field increases but the drain-end of the device starts getting debiased, that is, the electron concentration towards the drain end of the channel drops as y approaches L. For large V_{DS}, channel debiasing becomes severe. The electron concentration on the drain-side of the channel drops very low and the majority of the voltage is absorbed there.

To understand how channel debiasing results in current saturation, one must focus on its effect on the situation at the source end of the channel. Towards understanding this, it is useful to go back to the very fundamental Eq. 9.4 that expressed the channel current at any one location as the product of the channel charge times the electron velocity at that particular location. At the source end of the channel ($y = 0$), this expression becomes:

$$I_e \simeq W v_{ey}(0) Q_i(0) \tag{9.18}$$

We can write an equation like this for any location of the channel. What is significant about the source end is that the inversion layer charge there is set only by V_{GS} through:

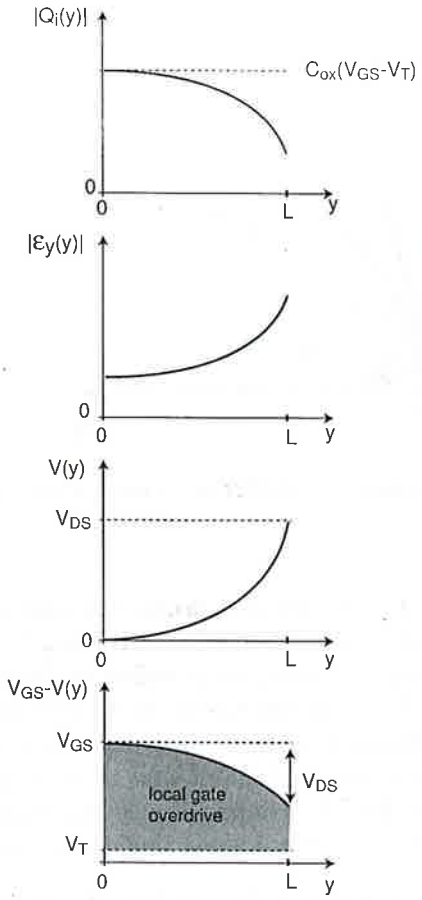


Figure 9.9: Inversion layer charge, lateral electric field, channel voltage and local gate overdrive as a function of location along the channel for a MOSFET biased in the linear regime.

$$Q_i(0) = -C_{ox}(V_{GS} - V_T) \tag{9.19}$$

since $V(0) = 0$. In particular, $Q_i(0)$ is independent of what happens down the channel.

The V_{DS} dependence enters Eq. 9.18 through the electron velocity, $v_{ey}(0)$. This is set by the local lateral electric field. With transport in the mobility regime:

$$v_{ey}(0) \simeq -\mu_e \mathcal{E}_y(0) \tag{9.20}$$

The key to understanding the saturating behavior of I_D with V_{DS} is to focus on the V_{DS} dependence of $\mathcal{E}_y(0)$. As Fig. 9.10 shows, as V_{DS} increases, so does the electric field at the source. However, when channel debiasing becomes prominent at high V_{DS}, the rise of $\mathcal{E}_y(y)$ slows down as higher values of V_{DS} are applied and eventually it does not increase any more. In consequence, at high V_{DS}, the velocity at which electrons are extracted from the source saturates

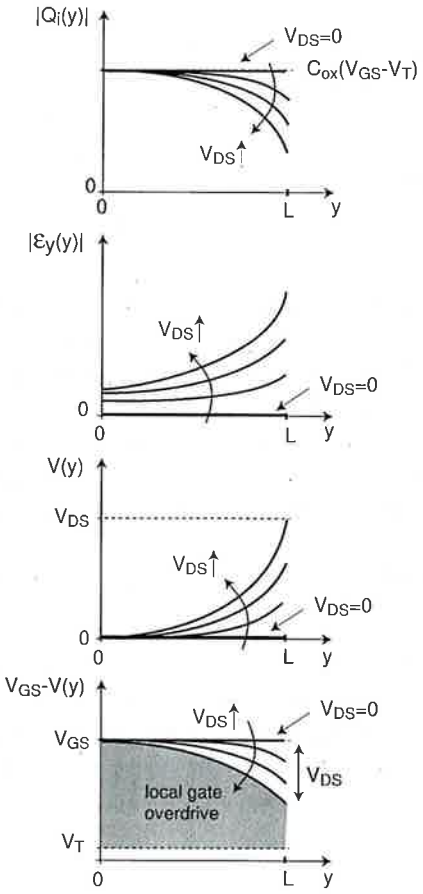


Figure 9.10: Inversion layer charge, lateral electric field, channel voltage and local gate overdrive as a function of location along the channel for a MOSFET biased in the linear regime for increasing values of V_{DS}. For high V_{DS}, channel debiasing at the drain-end of the channel is apparent.

and so does the current.

This line of argumentation is productive in one more and very general way. In the MOSFET, when dealing with transport phenomena, it is often most useful to focus on what is happening on the source side, as opposed to the drain side of the channel. It is at the source end of the channel that we can think in terms of the number of electrons that are extracted from the source and the velocity at which they come out. This is a simple and powerful picture that often provides great physical insight.

It is also interesting to reflect on what is going on at the drain-end of the channel as V_{DS} approaches $V_{GS} - V_T$. This is the point with the lowest concentration of electrons in the inversion layer. As V_{DS} approaches $V_{GS} - V_T$, Eq. 9.17 indicates that Q_i goes to zero at $y = L$. In this limiting case, the location at the drain end of the channel is known as the *pinch-off* point. Complete depletion of electrons is of course not possible, otherwise, the current could not flow. What is happening and how could our formalism deal with it?

There are several problems with our formalism close to $y = L$ for values of V_{DS} that approach $V_{GS} - V_T$. First of all, at $y \simeq L$ for $V_{DS} \simeq V_{GS} - V_T$, the local gate overdrive goes to zero. As a consequence, the fundamental charge control relationship stops applying and the gradual-channel approximation fails. This should make the use of Eq. 9.17 suspect. An additional problem with our formalism around the pinch-off condition is that the assumption of linearity between electric field and electron velocity is unlikely to apply as the lateral field gets high enough for velocity saturation to occur. For these reasons, the current-voltage characteristics given by Eq. 9.16 do not apply when V_{DS} approaches $V_{GS} - V_T$.

Developing a detailed model of current transport at the drain-end of the channel when V_{DS} approaches and exceeds $V_{GS} - V_T$ is a major challenge. Fortunately, for a first-order analysis, we do not need to do it. The reason is that the model embodied in Eq. 9.16 is fairly accurate over nearly the entire range of drain current. This is understood by examining Fig. 9.8 in detail. For small values of V_{DS}, increasing V_{DS} brings about a significant increase in I_D. For higher values of V_{DS}, however, a further increase in V_{DS} produces a proportionally smaller increase in I_D. For V_{DS} values close to $V_{GS} - V_T$, I_D is essentially saturated. The origin of this diminishing return in I_D for high values of V_{DS} is the "channel debiasing" discussed above. Even before we reach pinch-off and the simple model developed above becomes invalid, the drain current has essentially stopped increasing.

It is not difficult to quantify how close V_{DS} can get to $V_{GS} - V_T$ before the simple model developed above fails. As Problem 9.3 shows, for typical MOSFET designs, V_{DS} can get up to about 80% of $V_{GS} - V_T$ before the gradual-channel approximation fails. This means that I_D can build up to about 96% of the maximum value predicted by Eq. 9.16 and still be fairly accurate.

Exercise 9.1: *Consider a long n-channel MOSFET characterized by the following parameters: n^+-polysilicon gate ($W_M = \chi_S = 4.04\ eV$), $x_{ox} = 15\ nm$, uniform $N_A = 10^{17}\ cm^{-3}$, $L = 1\ \mu m$, $W = 10\ \mu m$ at a bias given by $V_{GS} = 2.5\ V$, $V_{DS} = 1\ V$, and $V_{BS} = 0$. Estimate the sheet charge density in the inversion layer at the source- and drain-ends of the channel. Estimate the current flowing through the drain of this transistor. Use $\mu_e = 500\ cm^2/V.s$.*

Proceeding as in Ch. 8, for this MOS structure we can easily find: $C_{ox} = 2.3 \times 10^{-7}\ F/cm^2$, $\gamma = 0.79\ V^{1/2}$, $\phi_{sT} = 0.84\ V$, $V_{FB} = -1\ V$, and $V_T = 0.56\ V$.

To obtain the sheet charge density at the source-end of the channel at the indicated bias, we can use Eq. 9.19:

$$Q_i(0) = -C_{ox}(V_{GS} - V_T) = -2.3 \times 10^{-7}\ F/cm^2\ (2.5 - 0.56)\ V = -4.5 \times 10^{-7}\ C/cm^2$$

Before we can obtain the corresponding value at the drain-end of the channel, we need to assert the regime of operation of the transistor. For the indicated bias, $V_{GS} - V_T = 1.94\ V > V_{DS} = 1\ V$. Hence this MOSFET is biased in the linear regime. We can then use Eq. 9.17 to obtain:

$$Q_i(L) = -C_{ox}(V_{GS} - V_T - V_{DS}) = -2.3 \times 10^{-7}\ F/cm^2\ (2.5 - 0.56 - 1)\ V = -2.2 \times 10^{-7}\ C/cm^2$$

The drain current can be obtained from Eq. 9.16:

$$\begin{aligned} I_D &= \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS} \\ &= \frac{10\ \mu m}{1\ \mu m} 500\ cm^2/V.s \times 2.3 \times 10^{-7}\ F/cm^2 (2.5 - 0.56 - \frac{1}{2}1\ V) \times 1\ V = 1.7\ mA \end{aligned}$$

9.4.3 The saturation regime

In the previous section we have developed a simple model for the linear regime of operation of the MOSFET. We found that I_D depends on V_{DS} and V_{GS} and we derived an expression for I_D that applies up to essentially $V_{DS} = V_{GS} - V_T$. What happens if V_{DS} reaches or exceeds this value? This is the saturation regime.

It is clear from our discussion above that around $V_{DS} = V_{GS} - V_T$, the electron concentration at the drain-end of the channel drops substantially. In fact, it can become rather small in comparison with the doping level of the body. If this happens, we can think about this as a small depletion region appearing right at the drain end of the channel. This is sketched in Fig. 9.11. This depletion region does not represent in any way a barrier to electron flow. On the contrary, the lateral electric field in this region "pulls" electrons into the drain. [2]

What is significant about the appearance of this depletion region is that if V_{DS} is increased beyond $V_{GS} - V_T$, the extra applied voltage drops entirely in this pinch-off region by widening its lateral extension. This is also sketched in Fig. 9.11. What happens here is entirely analogous to the electrostatics of a reverse biased pn junction that widens as the reverse bias increases. If the pinch-off region widening is small in the scale of the channel length, the electrostatics of the

[2] Transport through the pinch-off point in a MOSFET is just like through the base-collector depletion region of a bipolar transistor biased in the forward active regime, as will become clear in Ch. 11.

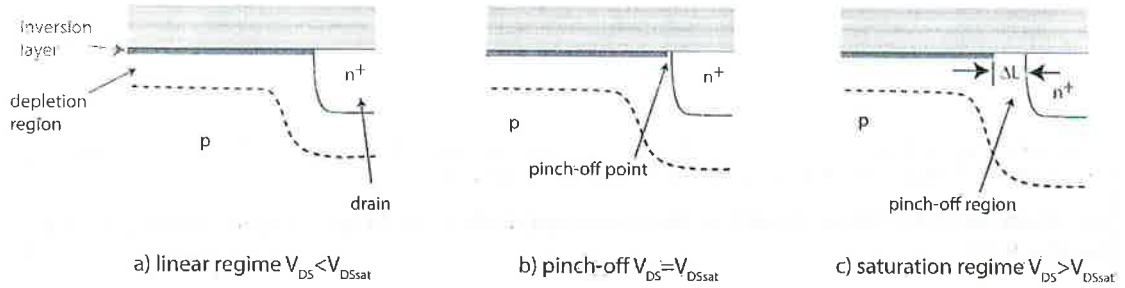


Figure 9.11: Sketch of electrostatics in MOSFET for different values of V_{DS}. a) Linear regime: $V_{DS} < V_{DSsat}$; b) pinchoff $V_{DS} = V_{DSsat}$; c) saturation regime $V_{DS} > V_{DSsat}$.

channel are not significantly affected by an increase in V_{DS} past $V_{GS} - V_T$. Hence the channel current does not change from the value that it has at $V_{DS} = V_{GS} - V_T$. In other words, the drain current has *saturated* and the MOSFET is said to be in the *saturation regime.*

To further clarify the physics of the MOSFET around pinch-off, Fig. 9.12 sketches the inversion layer charge, the lateral electric field, the voltage, and the local gate overdrive along the channel in the linear regime, at pinch-off, and in saturation. The top diagram shows that at pinch-off the inversion charge reaches a minimum at the drain-end of the channel. For higher values of V_{DS}, the pinch-off point widens and extends slightly into the channel. The third diagram from the top shows that all the extra voltage applied past pinch-off drops in the pinch-off region and that the voltage distribution in the rest of the channel is negligibly affected. This results in a negative gate overdrive in the pinch-off region at the drain end of the channel, as seen in the bottom diagram. The second diagram from the top shows a similar picture for the electric field.

Before we write an expression for the drain current in the saturation regime, we need to discuss several aspects of the pinch-off model that we have just introduced. Current saturation in the pinch-off model has been predicated upon the assumption that a depletion region forms at the drain-end of the channel, that is, that the electron concentration drops below the acceptor concentration in the body. This assumption might be satisfied at low current levels, but it is certain to be violated if the current level is high enough. It turns out that the appearance of a depletion region is not an essential condition for current saturation. Electron velocity saturation at the drain-end of the channel also causes it. Velocity saturation of electrons on the drain side of the channel is likely to occur as the electron concentration drops but current continuity must be maintained. We will understand in the next chapter how velocity saturation causes current saturation.

A second ingredient of current saturation is that the extent of the depletion region (or the velocity saturated region) is small in the scale of the channel length. This of course, depends on the design of the MOSFET. For well designed devices, this is always the case, but proving it requires a detailed model of the two-dimensional electrostatics of the drain-end of the channel.

We can now derive a first-order expression for I_{Dsat}, the drain current in the saturation

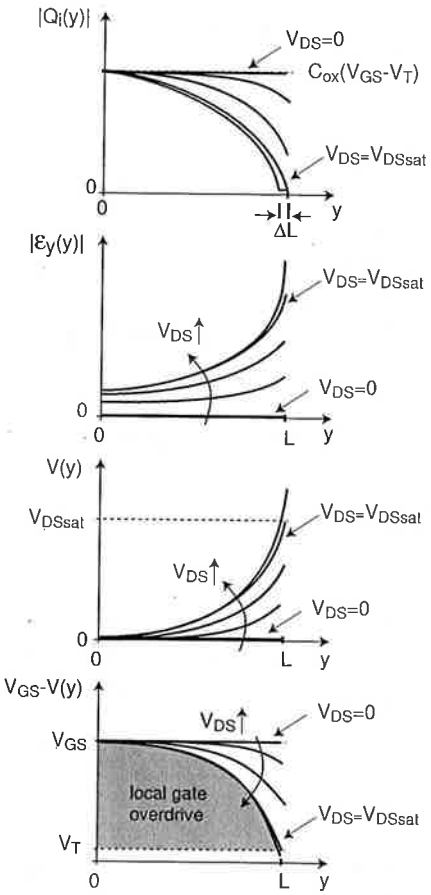


Figure 9.12: Inversion layer charge, lateral electric field, channel voltage and local gate overdrive as a function of location along the channel for a MOSFET for increasing values of V_{DS}. For $V_{DS} > V_{DSsat}$, the electrostatics of the channel are not significantly affected by the increase in V_{DS} and the channel current is saturated.

regime. Since we argued above that Eq. 9.16 predicts I_D with good accuracy nearly all the way to $V_{DS} = V_{GS} - V_T$ and I_D does not increase beyond this value of V_{DS}, then I_{Dsat} can be estimated by substituting $V_{DS} = V_{GS} - V_T$ in Eq. 9.16 to yield:

$$I_{Dsat} \simeq \frac{W}{2L}\mu_e C_{ox}(V_{GS} - V_T)^2 \tag{9.21}$$

Fig. 9.13 shows the so-called *output characteristics* of the MOSFET. In this graph, the drain current is graphed unchanged with V_{DS} for $V_{DS} \geq V_{GS} - V_T$. This is the manifestation of the saturation regime.

A key result that emerges in Eq. 9.21 is that in the saturation regime, the drain current increases with *the square* of the gate overdrive (the excess gate-to-source voltage above threshold). This is sketched in Fig. 9.14. The reason for the square-law dependence deserves some thought.

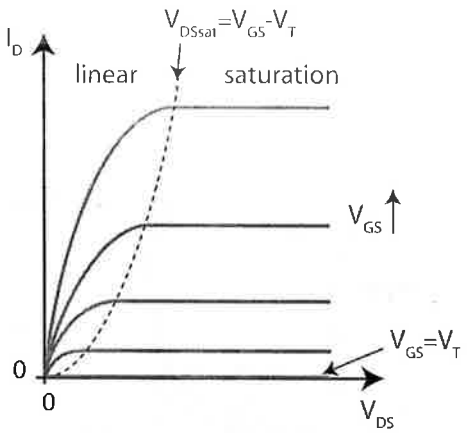


Figure 9.13: Sketch of output $I-V$ characteristics of a MOSFET in the linear and saturation regimes.

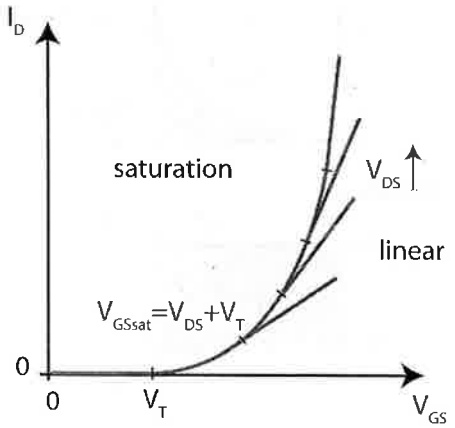


Figure 9.14: Sketch of transfer characteristics of a MOSFET , that is, I_D vs. V_{GS} for different values of V_{DS}. In the saturation regime, I_D increases quadratically with gate overdrive. When V_{GS} exceeds $V_{DS}+V_T$, the transistor enters the linear regime and I_D rises linearly with gate overdrive.

The higher V_{GS}, the higher the inversion charge in the channel that is available for transport. This is a linear dependence. Additionally, the higher V_{GS}, the higher the value of V_{DS} that can be applied before the channel is "pinched-off". As a consequence, when the MOSFET is saturated, the field inside the channel is also higher and the current is also larger. This is also a linear relationship. The combination of both linear dependences results in a square-law dependence for I_D on $V_{GS}-V_T$.

The value of V_{DS} required to pinch-off the device is commonly referred to as V_{DSsat}. In the simple formulation derived here, V_{DSsat} is given by:

$$V_{DSsat} = V_{GS} - V_T \tag{9.22}$$

The locus of V_{DSsat} in the output characteristics is indicated in Fig. 9.13. As a result of the square-law dependence of I_{Dsat} on $V_{GS} - V_T$, this locus has a parabolic shape.

The I-V characteristics of a MOSFET when graphed as in Fig. 9.14 are known as the *transfer characteristics*. For a given value of V_{DS}, as V_{GS} increases above V_T, the transistor turns on directly into the saturation regime and the current rises quadratically with gate overdrive. As V_{GS} increases, eventually the transistor is pushed into the linear regime. This happens when $V_{GSsat} = V_{DS} + V_T$. Beyond this point, I_D rises linearly with gate overdrive.

Fig. 9.15 shows experimental output and transfer characteristics of a 2N7000 MOSFET. The behavior of this transistor is very much as discussed in this section.

Exercise 9.2: *Consider the same long n-channel MOSFET as in Exercise 9.1 but now biased with* $V_{GS} = 2.5\ V$, $V_{DS} = 4\ V$, *and* $V_{BS} = 0$. *Estimate the sheet charge density in the inversion layer at the source- and drain-ends of the channel. Estimate the current flowing through the drain of this transistor. Use* $\mu_e = 500\ cm^2/V.s$.

This transistor is now biased in the saturation regime. This is because $V_{DS} = 4\ V > V_{DSsat} = V_{GS} - V_T = 1.94\ V$.

The expression for the sheet charge density at the source-end of the channel only depends on V_{GS}. Since the value of V_{GS} here is identical to that in Exercise 9.1, $Q_i(0)$ is also the same and given by $4.5 \times 10^{-7}\ C/cm^2$.

With the MOSFET biased in saturation, at the drain-end of the channel there is a pinch-off point. Hence, the sheet charge density of electrons there is quite small and at this moment, we do not know how to estimate it. In Ch. 10 we will learn how to do this.

The drain current can be obtained from Eq. 9.21:

$$I_D = \frac{W}{2L}\mu_e C_{ox}(V_{GS} - V_T)^2 = \frac{10\ \mu m}{2 \times 1\ \mu m} 500\ cm^2/V.s \times 2.3 \times 10^{-7}\ F/cm^2 (2.5 - 0.56\ V)^2 = 2.2\ mA$$

9.4.4 DC large-signal equivalent-circuit model of ideal MOSFET

An equivalent circuit model that captures the DC I-V characteristics of the ideal MOSFET is fairly straightforward. In the ideal MOSFET, the drain current also flows through the source and it is controlled by V_{DS} and V_{GS}. This behavior can be captured by means of a controlled current source. The dependence of the drain current on V_{DS} and V_{GS} is given by Eqs. 9.16 and 9.21 in the linear and saturation regimes, respectively.

In the ideal MOSFET, the gate does not draw any DC current. It is therefore an open circuit. Also, in the ideal MOSFET, we ignore the presence of the source/body and drain/body PN junctions. Hence, the body contact is also an open. The entire DC large-signal equivalent circuit model of the ideal MOSFET is graphed in Fig. 9.16. The diamond shaped current source indicates that this is a dependent source that is controlled by V_{GS} and V_{DS}.

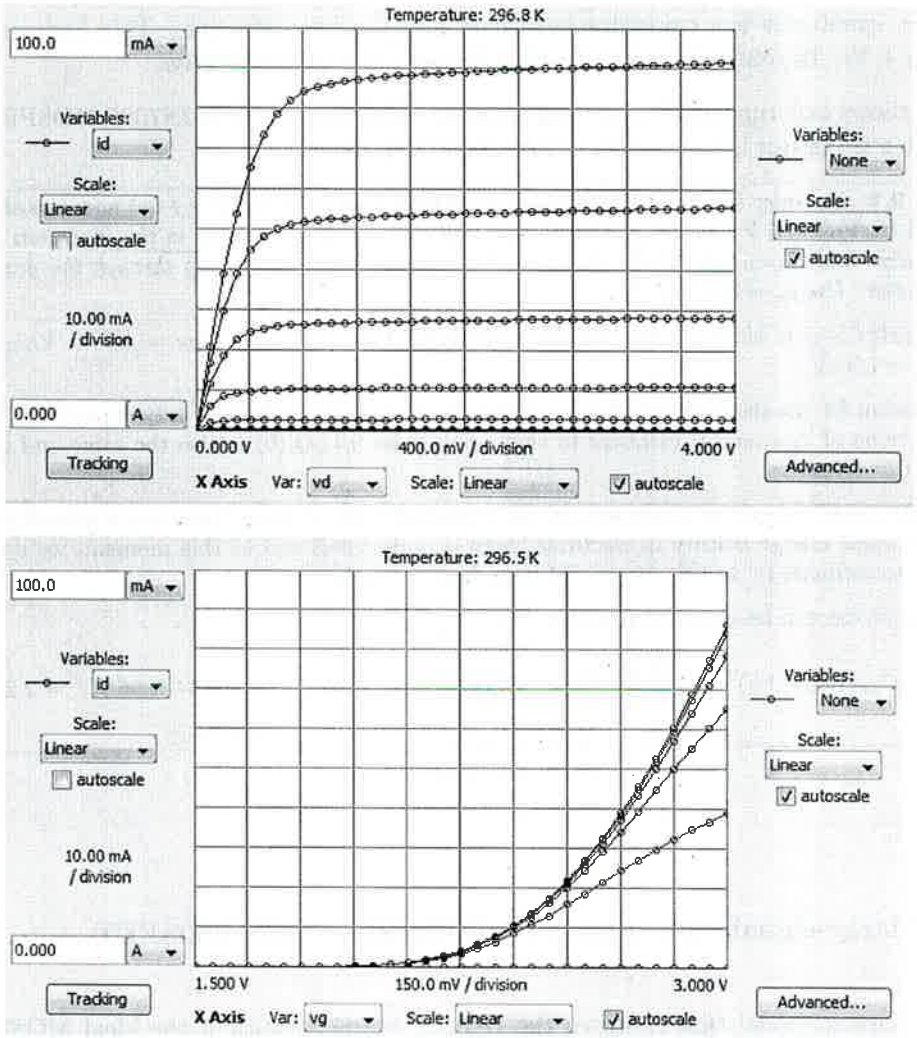


Figure 9.15: Measured I-V characteristics of a 2N7000 MOSFET. Top: output characteristics with V_{GS} stepped from 1 to 3 V in steps of 0.2 V. Bottom: transfer characteristics with V_{DS} stepped from 0 to 1 V in steps of 0.2 V (screen shots from MIT Microelectronics iLab).

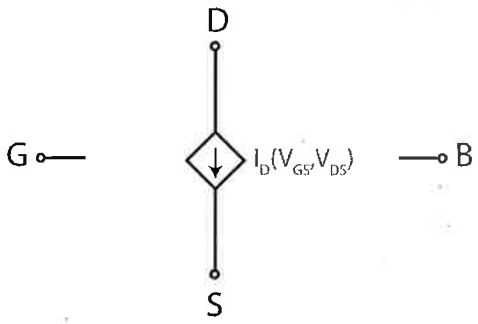


Figure 9.16: DC large-signal equivalent circuit model of ideal MOSFET.

9.4.5 Energy band diagrams

Our understanding of the I-V characteristics of the MOSFET would not be complete without energy band diagrams. Fig. 9.17 sketches energy band diagrams along the semiconductor/insulator interface for a value of V_{GS} in strong inversion and for several values of V_{DS}.

In the linear regime with $V_{DS} = 0$, the Fermi level is flat along the channel from source to drain. The conduction band is close to the Fermi level in the source and drain but a bit further away in the channel. The bands are flat in the channel.

For a value of V_{DS} in the linear regime, the electron Fermi level is lowered at the drain with respect to the source by an amount qV_{DS}. Since the drain is heavily doped, this brings down with it the entire band structure on the drain. A lateral electric field appears in the channel and in consequence, the band structure bends along the channel. Current flows. This also means that the quasi-Fermi level for electrons also tilts along the channel.

At pinch-off, for $V_{DS} = V_{DSsat}$, the bending of the bands and the electron quasi-Fermi level along the channel is exacerbated. As one proceeds from source to drain along the channel, the bending increases, although for different reasons. E_c and E_v bend because the electric field increases towards the drain end of the channel. E_{fe} bends because the electron concentration drops towards the drain but the current is constant along the channel. In consequence, the gradient of E_{fe} must increase as we proceed from source to drain.

In saturation, for $V_{DS} > V_{DSsat}$, the band structure and electron quasi-Fermi level along the channel do not change with respect to the situation observed at $V_{DS} = V_{DSsat}$. The applied voltage beyond V_{DSsat}, that is, $V_{DS} - V_{DSsat}$, drops entirely at the pinch-off point where the bands and the quasi-Fermi level show a steep dive. It is clear from this diagram that the pinch-off point does not represent at all a bottleneck to the drain current. On the contrary, at the pinch-off point, the electrons are in a "free fall" down the channel to the drain.

In saturation, the bottleneck to the drain current is the energy barrier that electrons at the source have to overcome before they can enter the channel and the electric field at that same point. The energy barrier at the source end of the channel is controlled by V_{GS}. The higher V_{GS}, the lower the energy barrier and the higher the concentration of electrons that have enough energy

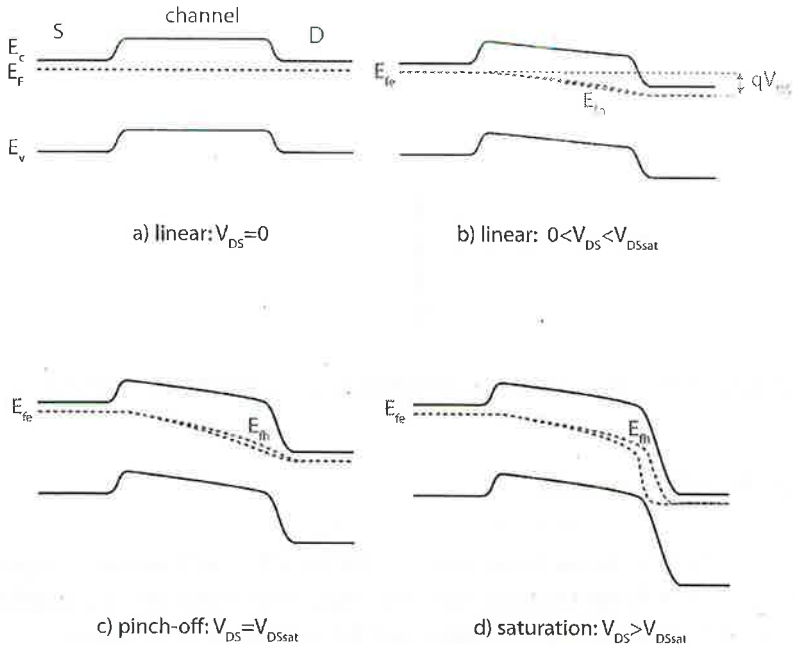


Figure 9.17: Energy band diagram along semiconductor surface from source to drain for four different values of V_{DS} (in all cases $V_{GS} > V_T$).

to get injected into the channel. In particular, the height of this energy barrier is independent of what happens down the channel and hence the electron charge at $y = 0$ is independent of V_{DS}. The tilt of the bands at the source end of the channel determines the electric field there and the velocity at which electrons are extracted from the source. This is set by the potential distribution along the channel. This is unchanged for $V_{DS} > V_{DSsat}$, since all the extra voltage applied beyond V_{DSsat} drops at the pinch-off point.

Fig. 9.17 also shows the quasi-Fermi level for holes along the channel at the different bias conditions. With $V_{DS} = 0$, both quasi-Fermi levels coincide. As V_{DS} increases, the quasi-Fermi levels split with $E_{fh} > E_{fe}$ towards the end of the channel. This is similar to the situation depicted on the right of Fig. 8.29 in which a positive bias is applied between the inversion layer and the body, as will become clear later on in this chapter.

9.5 Charge-voltage characteristics of the ideal MOSFET

The I-V characteristics of a MOSFET, as is the case in most semiconductor devices, represent only a partial view of the behavior that is generally relevant in circuit applications. In particular, to capture the dynamics of a MOSFET, we need to complete the picture with the charge-voltage or capacitance-voltage characteristics. These describe the charge that is stored in the device and that must be provided to it or removed from it every time any terminal voltage changes.

In a MOSFET there are two types of stored charge. There is first the depletion charge in

the depletion regions associated with the source-body and drain-body PN junctions as well as the depletion charge in the body of the semiconductor that is associated with the MOS structure itself. In addition to this, there is the electron charge stored in the inversion layer. If $V_{DS} > 0$, this charge is in transit from source to drain and results in current. Nevertheless, this is charge that must have been provided once and must be removed if the transistor is to be switched off. We construct first-order models for these two types of stored charge in the next two subsections.

9.5.1 Depletion charge

In an ideal MOSFET, as sketched in Fig. 9.3, there are three distinct depletion regions. First, there is the depletion region associated with the source-body PN junction. Second, there is the depletion region associated with the drain-body PN junction. Finally, there is the depletion region associated with the MOS structure itself. The charge-voltage characteristics of these three depletion regions have been studied before in previous chapters.

For the source-body and drain-body junctions, we can recycle results from Sec. 6.4. The charge associated with the source-body depletion region is given by:

$$Q_{jS} = L_S W \sqrt{2qN_A\epsilon_s(\phi_{bij} - V_{BS})} = Q_{jSo}\sqrt{1 - \frac{V_{BS}}{\phi_{bij}}} \tag{9.23}$$

where L_S is the extent of the source in the y dimension, ϕ_{bij} is the built-in potential of the source-body junction. Q_{jSo} is the depletion charge in thermal equilibrium:

$$Q_{jSo} = L_S W \sqrt{2qN_A\epsilon_s\phi_{bij}} \tag{9.24}$$

In writing Eqs. 9.23 and 9.24, we have assumed that this is an asymmetric junction, that is that the doping level of the source is much higher than that of the body. This is a very good assumption in practice.

Similarly, the depletion charge associated with the drain-body junction is given by:

$$Q_{jD} = L_D W \sqrt{2qN_A\epsilon_s(\phi_{bij} - V_{BD})} = Q_{jDo}\sqrt{1 - \frac{V_{BS} - V_{DS}}{\phi_{bij}}} \tag{9.25}$$

with Q_{jDo} given by:

$$Q_{jDo} = L_D W \sqrt{2qN_A\epsilon_s\phi_{bij}} \tag{9.26}$$

This is equal to Q_{jSo} if $L_S = L_D$ as is commonly the case. We have assumed that the doping level on both junctions is the same and that the built-in potential is then identical.

The depletion region charge associated with the MOS structure depends on the regime of operation of the MOSFET. For $V_{GS} < V_T$, the MOSFET is in cut-off and there is no inversion

layer anywhere in the channel. In this case, the charge in this depletion region was calculated in Eq. 8.20 and is given by:

$$Q_{jB} = LW\frac{1}{2}\gamma^2 C_{ox}[\sqrt{1+4\frac{V_{GB}-V_{FB}}{\gamma^2}}-1] - LW\frac{1}{2}\gamma^2 C_{ox}[\sqrt{1+4\frac{V_{GS}-V_{BS}-V_{FB}}{\gamma^2}}-1] \quad (9.27)$$

where V_{FB} is the flatband voltage of the MOS structure. The depletion region charge increases from zero at $V_{GS} = V_{FB}$ to its maximum value when $V_{GS} = V_T$.

For gate voltages in excess of V_T, an inversion layer is formed under the gate and the depletion region thickness does not increase any more. Hence, for the MOSFET in the linear and saturation regimes, the charge associated with the depletion region in the channel is given by:

$$Q_{jBmax} = LW\frac{1}{2}\gamma^2 C_{ox}[\sqrt{1+4\frac{V_T-V_{BS}-V_{FB}}{\gamma^2}}-1] \quad (9.28)$$

When the voltage applied across any of the terminals of a MOSFET changes, the stored depletion charge also changes. This gives rise to capacitive effects. For the source-body and drain-body junctions, the junction capacitances is given, respectively, by:

$$C_{jS} = \frac{C_{jSo}}{\sqrt{1-\frac{V_{BS}}{\phi_{bij}}}} \quad (9.29)$$

$$C_{jD} = \frac{C_{jDo}}{\sqrt{1-\frac{V_{BS}-V_{DS}}{\phi_{bij}}}} \quad (9.30)$$

with

$$C_{jSo} = L_S W\sqrt{\frac{\epsilon_s q N_A}{2\phi_{bij}}} \quad (9.31)$$

$$C_{jDo} = L_D W\sqrt{\frac{\epsilon_s q N_A}{2\phi_{bij}}} \quad (9.32)$$

In the cut-off regime, the depletion region charge underneath the gate is modulated by V_{GB}. Hence there is also a capacitive effect associated with this. This capacitance is given by:

$$C_{jB} = \frac{LWC_{ox}}{\sqrt{1+4\frac{V_{GB}-V_{FB}}{\gamma^2}}} \quad (9.33)$$

This is a result already obtained in Eq. 8.49. It can also be easily obtained by differentiating Eq. 9.27 with respect to V_{GB}. This result applies when $V_{GS} < V_T$.

When $V_{GS} > V_T$, the MOSFET is in the linear or saturation regime and Q_{jB} becomes independent of V_{GB}. The associated capacitance is zero.

Exercise 9.3: *Consider once again the long n-channel MOSFET of Exercises 9.1 and 9.2. The source and drain regions have a doping level* $N_S = N_D = 10^{19}\ cm^{-3}$ *and a length* $L_S = L_D = 5\ \mu m$. *Estimate the values of* C_{jS}, C_{jD} *and* C_{jB} *at the following two bias points:* i) $V_{GS} = 0\ V$, $V_{DS} = 1\ V$, *and* $V_{BS} = 0$, *and* ii) $V_{GS} = 2.5\ V$, $V_{DS} = 1\ V$, *and* $V_{BS} = 0$.

At the first bias point, the transistor is in cut-off. We start by computing the capacitance per unit area of the source and drain junctions in equilibrium. These are identical and given by Eqs. 9.31 and 9.32. Proceeding as in Ch. 6, we start by calculating the built-in potential of these junctions which we find to be $\phi_{bij} = 0.94$. Then:

$$\begin{aligned} C_{jSo} &= C_{jDo} = L_S W\sqrt{\frac{\epsilon_s q N_A}{2\phi_{bij}}} = L_D W\sqrt{\frac{\epsilon_s q N_A}{2\phi_{bij}}} \\ &= 5\times10^{-4}\ cm \times 10\times10^{-4}\ cm\ \sqrt{\frac{1.04\times10^{-12}\ F/cm \times 1.6\times10^{-19}\ C\times10^{17}\ cm^{-3}}{2\times0.94\ V}} \\ &= 4.7\ \times10^{-14}\ F \end{aligned}$$

With $V_{BS} = 0$, the capacitance of the source-body junction (Eq. 9.29) is given by the result just obtained:

$$C_{jS} = C_{jSo} = 4.7\ \times10^{-14}\ F$$

There is a bias applied between the drain and the body. C_{jD} is then given by (Eq. 9.30):

$$C_{jD} = \frac{C_{jDo}}{\sqrt{1-\frac{V_{BS}-V_{DS}}{\phi_{bij}}}} = \frac{4.7\ \times10^{-14}\ F}{\sqrt{1-\frac{0-1\ V}{0.94\ V}}} = 3.3\ \times10^{-14}\ F$$

For the first bias point, the capacitance associated with the depletion region of the MOS structure is given by Eq. 9.33:

$$C_{jB} = \frac{LWC_{ox}}{\sqrt{1+4\frac{V_{GB}-V_{FB}}{\gamma^2}}} = \frac{1\times10^{-4}\ cm\times10\times10^{-4}\ cm\times2.3\times10^{-7}\ F/cm^2}{\sqrt{1+4\frac{0+1\ V}{0.79^2\ V}}} = 1.8\times10^{-14}\ F$$

At the second bias point, V_{GS} has increased to $2.5\ V$ which is above threshold. Hence, the MOSFET is now in the linear regime and an inversion layer forms. This breaks the electrostatic coupling between the gate and the body and $C_{jB} = 0$.

Additionally, V_{BS} and V_{BD} are unchanged from the first bias point so C_{jS} and C_{jD} have the same values that were obtained at the first bias point.

9.5.2 Inversion charge

In addition to the depletion region charge, in a MOSFET in the linear or saturation regimes of operation there is a certain amount of electron charge that is stored in the inversion layer. When

turning on the MOSFET, this charge must be provided, and when turning it off, it has to be removed. The presence of inversion layer charge clearly impacts the dynamics of the MOSFET. In this section, we compute the integrated inversion layer charge for the ideal MOSFET. We also capture its dynamic behavior by means of two intrinsic capacitances.

We seek an expression for the integral of charge in the inversion layer of an ideal MOSFET. This can be expressed mathematically as the integral of Q_i along the entire channel:

$$Q_I = W \int_0^L Q_i(y) dy \tag{9.34}$$

Since we have derived an expression for Q_i in terms of V, rather than y, it is best to change variables in this integral and rewrite it as:

$$Q_I = W \int_0^{V_{DS}} Q_i(V) \frac{dy}{dV} dV \tag{9.35}$$

The term in dy/dV can be obtained from Eq. 9.9:

$$\frac{dy}{dV} = -\frac{W\mu_e}{I_D} Q_i(V) \tag{9.36}$$

Combining this with Eq. 9.35, we get:

$$Q_I = -\frac{W^2\mu_e}{I_D} \int_0^{V_{DS}} Q_i^2(V) dV \tag{9.37}$$

For the ideal MOSFET, $Q_i(V)$ was given in Eq. 9.11. Inserting this in Eq. 9.37 yields:

$$Q_I = -\frac{W^2\mu_e C_{ox}^2}{I_D} \int_0^{V_{DS}} (V_{GS} - V - V_T)^2 dV = \frac{1}{3} W^2 \mu_e C_{ox}^2 \frac{(V_{GS} - V_{DS} - V_T)^3 - (V_{GS} - V_T)^3}{\frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS}} \tag{9.38}$$

After a relatively simple mathematical manipulation [3], we can express Q_I as:

$$Q_I = -\frac{2}{3} WLC_{ox} \frac{(V_{GS} - V_T)^2 + (V_{GS} - V_T)(V_{GD} - V_T) + (V_{GD} - V_T)^2}{(V_{GS} - V_T) + (V_{GD} - V_T)} \tag{9.39}$$

Note that since we have used Eq. 9.11, this expression is only valid in the linear regime. As was the case of the MOSFET current, Q_I in the saturation regime can be obtained from this by setting $V_{DS} = V_{DSsat} = V_{GS} - V_T$, or $V_{GD} = V_T$. This yields:

[3] Note that:

$$(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS} = \frac{1}{2}[(V_{GS} - V_T)^2 - (V_{GD} - V_T)^2]$$

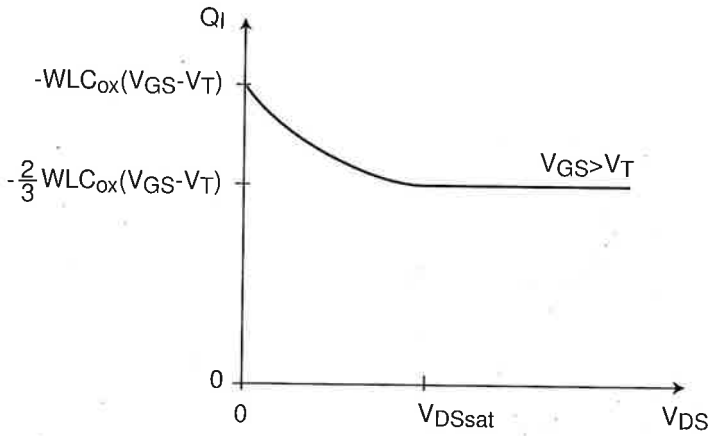


Figure 9.18: Inversion layer charge vs. V_{DS} for an ideal MOSFET in the linear or saturation regime for a given value of $V_{GS} > V_T$.

$$Q_I = -\frac{2}{3}WLC_{ox}(V_{GS} - V_T) \tag{9.40}$$

This interesting and simple result can be best understood by obtaining also Q_I in the limit of small V_{DS}. From Eq. 9.39, we get:

$$Q_I \simeq -WLC_{ox}(V_{GS} - V_T) \tag{9.41}$$

This result makes good sense. In the limit of small V_{DS}, at any one location in the inversion layer, the charge density is given by $-C_{ox}(V_{GS} - V_T)$. Hence, the total charge in the inversion layer can be obtained by multiplying this by the geometrical area of the gate.

The $\frac{2}{3}$ factor that is present in the expression of the inversion layer charge in the saturation regime in Eq. 9.40 deserves some thinking. If we sketch the inversion charge vs. V_{DS} for a given value of $V_{GS} > V_T$, as in Fig. 9.18, we see that as V_{DS} increases, Q_I monotonically decreases from the value given by Eq. 9.41 towards that given by Eq. 9.40. This is the consequence of the channel debiasing that was mentioned earlier. For values of $V_{GS} > V_T$, as V_{DS} increases, the inversion layer charge at any location in the channel, except at the source, gets reduced. This is most prominent towards the end of the channel. This was sketched in the top diagram of Fig. 9.10. Beyond $V_{DS} = V_{DSsat}$, the inversion layer charge profile along the channel does not change anymore and Q_I saturates. This is seen in the top diagram of Fig. 9.12. The factor of $\frac{2}{3}$ comes from the shape of the evolution of Q_i with y along the channel.

As we attempt to draw a large-signal equivalent circuit model for the MOSFET that accounts for the charge storage discussed in this section, we encounter an interesting problem. It is clear that the depletion charge storage associated with the source-body junction depletion region, Q_{jS}, hangs from the source and body terminals. Similarly, Q_{jD} hangs from the drain and body terminals. Additionally, the depletion charge associated with the MOS depletion region, Q_{jB}, hangs from the gate and body terminals. The interesting question is what to do about Q_I?

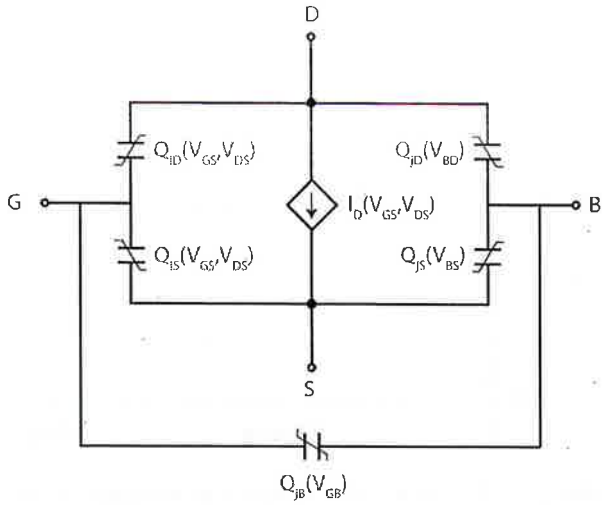


Figure 9.19: Large-signal equivalent circuit model for ideal MOSFET including depletion charge and inversion charge storage.

One of the terminals of Q_I is clear, it is the gate. The other terminal is less obvious, since the inversion layer is in contact with both the source and the drain. It then seems that an appropriate way to proceed is to split Q_I into two components: one that hangs between gate and source, and a second one that hangs between gate and drain. We call these Q_{IS} and Q_{ID}, respectively. It is clear that $Q_I = Q_{IS} + Q_{ID}$. It is not obvious how to compute Q_{IS} and Q_{ID}, individually. In fact, a detailed calculation of Q_{IS} and Q_{ID} is a bit involved. Fortunately, as it will become evident below, in many applications we do not need detailed models for Q_{IS} and Q_{ID}. It suffices to know the fact that they depend, in general, on both V_{GS} and V_{DS}.

The topology of the large-signal equivalent circuit model of the MOSFET incorporating all the storage elements discussed here is shown in Fig. 9.19.

An alternate way to describe the charge-voltage characteristics associated with the inversion layer is through the capacitance. In the case of a microelectronics device with more than two terminals, one has to exercise some caution when defining a capacitance and selecting the terminals across which it is to be placed. The best way to do this is to keep physical insight at all times and to look for the terminals that deliver the charge in both lobes of the charge dipole.

The inversion charge in an ideal MOSFET is supplied by the source and drain. The inversion layer charge is imaged at the gate. The gate charge is delivered by the gate contact. We should therefore expect two capacitors, one placed between the gate and the source and a second one between the gate and the drain. We will refer to these capacitors as C_{gsi} and C_{gdi} where the i refers to the intrinsic device. C_{gsi} and C_{gdi} can be easily obtained by computing suitable derivatives of Q_I in Eq. 9.39 with respect to the appropriate terminal voltages, holding all other voltages constant:

$$C_{gsi} = -\frac{\partial Q_I}{\partial V_{GS}}|_{V_{GD}} = \frac{1}{2}WLC_{ox}(V_{GS}-V_T)\frac{V_{GS}-V_T-\frac{2}{3}V_{DS}}{(V_{GS}-V_T-\frac{1}{2}V_{DS})^2} \quad (9.42)$$

$$C_{gdi} = -\frac{\partial Q_I}{\partial V_{GD}}|_{V_{GS}} = \frac{1}{2}WLC_{ox}(V_{GS}-V_{DS}-V_T)\frac{V_{GS}-V_T-\frac{1}{3}V_{DS}}{(V_{GS}-V_T-\frac{1}{2}V_{DS})^2} \quad (9.43)$$

As was the case of Eq. 9.39, these expressions only hold for the linear regime. For $V_{DS} = V_{DSsat}$, they become:

$$C_{gsi} = \frac{2}{3}WLC_{ox} \quad (9.44)$$

$$C_{gdi} = 0 \quad (9.45)$$

which are the appropriate values for the saturation regime ($V_{DS} > V_{DSsat}$). For very small V_{DS}, it is interesting to see that:

$$C_{gsi} \simeq C_{gdi} \simeq \frac{1}{2}WLC_{ox} \quad (9.46)$$

The evolution of C_{gsi} and C_{gdi} with V_{DS} is shown in Fig. 9.20. This behavior makes good physical sense. For small V_{DS}, we have a very symmetric situation where the charge in the inversion layer splits even between the source and the drain. As V_{DS} increases, the electrostatic influence of V_{GD} wanes, while that of V_{GS} strengthens. At $V_{DS} = V_{DSsat}$, the drain loses control of the inversion layer charge and $C_{gdi} = 0$. The channel has been pinched off. For higher values of V_{DS}, the picture does not change.

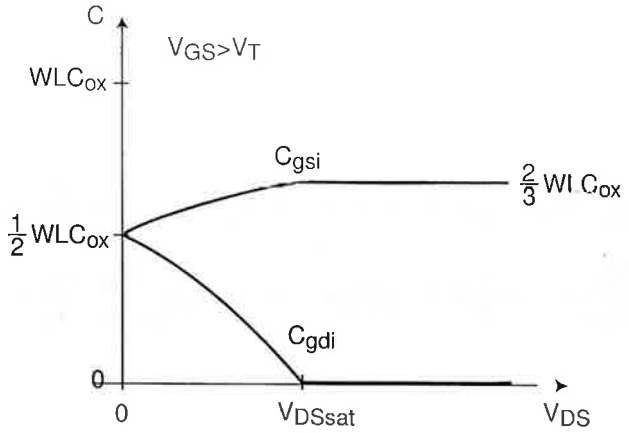


Figure 9.20: C_{gsi} and C_{gdi} vs. V_{DS} for an ideal MOSFET in the linear and saturation regimes for a given value of $V_{GS} > V_T$.

Exercise 9.4: *Consider once again the long n-channel MOSFET of Exercises 9.1-9.3. Estimate the values of C_{gsi} and C_{gdi} at the following three bias points:* i) $V_{GS} = 2.5\ V$, $V_{DS} = 0\ V$, *and* $V_{BS} = 0$; ii) $V_{GS} = 2.5\ V$, $V_{DS} = 1\ V$, *and* $V_{BS} = 0$; *and* iii) $V_{GS} = 2.5\ V$, $V_{DS} = 4\ V$, *and* $V_{BS} = 0$.

We start by computing the geometrical capacitance of the gate. This is given by:

$$C_g = LWC_{ox} = 1 \times 10^{-4}\ cm \times 10 \times 10^{-4}\ cm \times 2.3 \times 10^{-7}\ F/cm^2 = 2.3 \times 10^{-14}\ F$$

At the first bias point, the transistor is in the linear regime with $V_{DS} = 0$. Hence, the geometrical capacitance of the gate is split equally between the source and drain. Then:

$$C_{gsi} = C_{gdi} = \frac{1}{2}LWC_{ox} = 1.2 \times 10^{-14}\ F$$

At the second bias point, the MOSFET is in the linear regime. Hence we have to use expressions 9.42 and 9.43. For C_{gsi}, we have:

$$\begin{aligned} C_{gsi} &= \frac{1}{2}WLC_{ox}(V_{GS} - V_T)\frac{V_{GS} - V_T - \frac{2}{3}V_{DS}}{(V_{GS} - V_T - \frac{1}{2}V_{DS})^2} \\ &= 1.2 \times 10^{-14}\ F(2.5 - 0.56\ V)\frac{2.5 - 0.56 - \frac{2}{3} \times 1\ V}{(2.5 - 0.56 - \frac{1}{2} \times 1\ V)^2} = 1.4 \times 10^{-14}\ F \end{aligned}$$

Proceeding in a similar way with Eq. 9.43, we obtain $C_{gdi} = 8.8 \times 10^{-15}\ F$.

At the third bias point, the MOSFET is in saturation. Hence, according to Eq. 9.44, C_{gsi} is given by:

$$C_{gsi} = \frac{2}{3}WLC_{ox} = \frac{2}{3}2.3 \times 10^{-14}\ F = 1.5 \times 10^{-14}\ F$$

And $C_{gdi} = 0$ (Eq. 9.45).

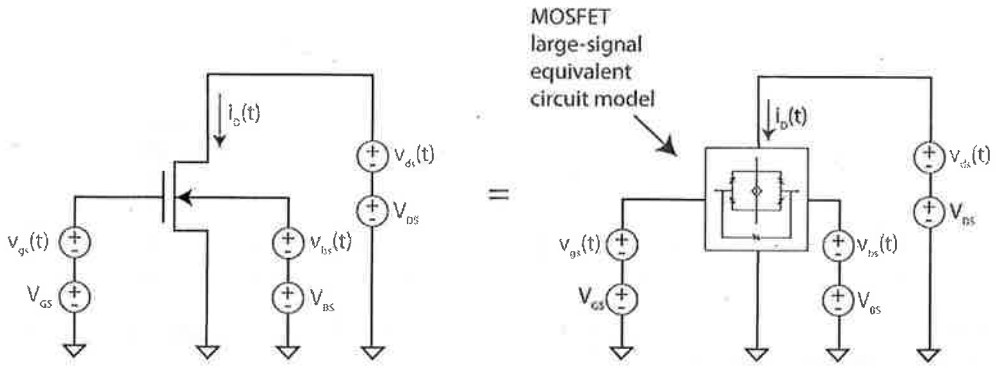


Figure 9.21: Illustration showing how a mixed-signal situation can be studied using the large-signal equivalent circuit model developed in the previous section.

9.6 Small-signal behavior of ideal MOSFET

In analog and mixed-signal applications, MOSFETs are often used in the small-signal mode. In these situations, the device is biased by means of DC sources in some regime, typically saturation, and then a small signal is applied to one or more terminals. There is interest in evaluating the response of the device to the small-signals alone. In a general way, we can consider these kinds of situations as on the left of Fig. 9.21. This figure shows a MOSFET biased through three DC voltage sources, V_{GS}, V_{DS}, and V_{BS}, referred to the source. In addition, three small-signal sources: $v_{gs}(t)$, $v_{ds}(t)$, and $v_{bs}(t)$, are also applied as shown. In general, we are interested in the time dependent drain current, $i_D(t)$, that results.

The equivalent circuit model that we developed in the previous section can certainly be used to solve this problem. This is illustrated in Fig. 9.21. In this approach, the transistor is substituted by its equivalent circuit model and standard circuit solving techniques are used to determine $i_D(t)$. However, since the model elements are non-linear, this approach can become quite tedious. Fortunately, many circuit simulation tools, such as SPICE, are available to solve the problem this way. In this approach, actual waveforms are introduced to a device model and the output waveforms are computed. A problem with this approach is that it is hard to develop an intuitive understanding for the connection among the figures of merit of interest, such as amplitude and phase of the drain current waveform, and the bias point or the device parameters. Fortunately, for a class of problems referred to as "small-signal problems," there is a more expeditious and intuitive approach.

9.6.1 Small-signal equivalent circuit model of ideal MOSFET

If the amplitude of the small signals is small enough, superposition suggests that the time-dependent response of the device, $i_D(t)$, should be the sum of the DC bias current I_D that is imposed by the DC voltage sources plus a small time dependent component $i_d(t)$. This is sketched in Fig. 9.22. Mathematically, this can be expressed as:

$$i_D(V_{GS}, V_{DS}, V_{BS}; v_{gs}, v_{ds}, v_{bs}) \simeq I_D(V_{GS}, V_{DS}, V_{BS}) + i_d(V_{GS}, V_{DS}, V_{BS}; v_{gs}, v_{ds}, v_{bs}) \quad (9.47)$$

Note how the small-signal response in general depends on the bias point that is selected.

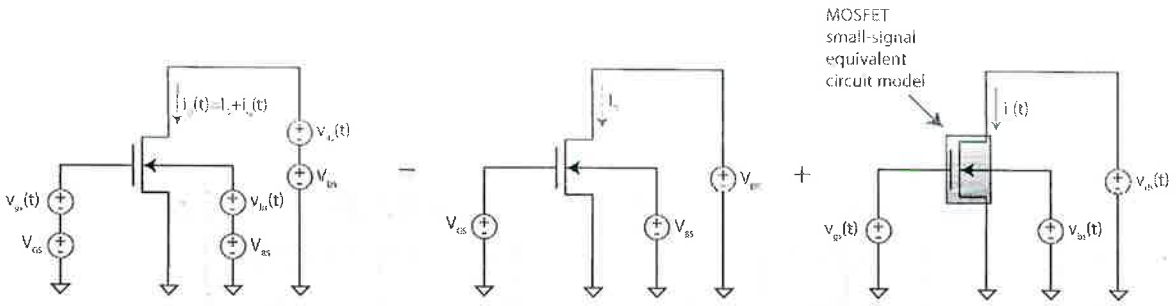


Figure 9.22: Illustration of superposition as a way to break a mixed-signal situation (left) into a DC bias situation (center) plus a small-signal situation (right).

The small-signal equivalent circuit model of the MOSFET is a circuit representation of the set of dependencies of the small-signal current $i_d(t)$ on the small-signal voltages v_{gs}, v_{ds}, and v_{bs}. This is the gray box on the figure on the right of Fig. 9.22. Since the small-signal sources are of small magnitude, the non-linear behavior of the MOSFET gets linearized. Linearity then allows us to express $i_d(t)$ as the sum of three independent terms in each of the small-signal voltage sources:

$$\begin{aligned} & i_d(V_{GS}, V_{DS}, V_{BS}; v_{gs}, v_{ds}, v_{bs}) \\ = \; & i_d(V_{GS}, V_{DS}, V_{BS}; v_{gs}) + i_d(V_{GS}, V_{DS}, V_{BS}; v_{ds}) + i_d(V_{GS}, V_{DS}, V_{BS}; v_{bs}) \end{aligned} \tag{9.48}$$

Depending on the complexity of the small-signal equivalent circuit model, this approach to solving the problem can be fairly fast and yield very meaningful analytical results. So the key question is: what is a suitable small-signal equivalent circuit model for the MOSFET? Since we already have a large-signal equivalent circuit model (Fig. 9.19), this requires only that we linearize it. This is fairly straightforward. In the first step, the charge storage elements become simple capacitors. Their expressions were already derived in the previous section. We are then left to deal with the non-linear voltage-controlled current source.

In a most general way, linearizing I_D demands that we take the linear terms of its Taylor series expansion:

$$\begin{aligned} & i_D(V_{GS}, V_{DS}, V_{BS}; v_{gs}, v_{ds}, v_{bs}) \\ & \simeq I_D(V_{GS}, V_{DS}, V_{BS}) + (\frac{\partial I_D}{\partial V_{GS}}|_{V_{DS}, V_{BS}})v_{gs} + (\frac{\partial I_D}{\partial V_{DS}}|_{V_{GS}, V_{BS}})v_{ds} + (\frac{\partial I_D}{\partial V_{BS}}|_{V_{GS}, V_{DS}})v_{bs} \\ & = I_D + g_m v_{gs} + g_o v_{ds} + g_{mb} v_{bs} \end{aligned} \tag{9.49}$$

The coefficients of v_{gs}, v_{ds}, and v_{bs} receive the names of *transconductance*, *output conductance* (also called "drain conductance"), and *back transconductance*, respectively. In the ideal MOSFET, I_D does not depend on V_{BS}. Hence g_{mb} is zero [4]. There are therefore two terms left in i_d. How do they get captured in the small-signal equivalent circuit model? The first term gives us a sense of remote control. It states that in response to a small wiggle in the gate voltage, the drain current changes a bit. The current

[4] In Sec. 9.7.2 we study the impact of applying a voltage to the MOSFET substrate and we will derive a first-order expression for g_{mb}.

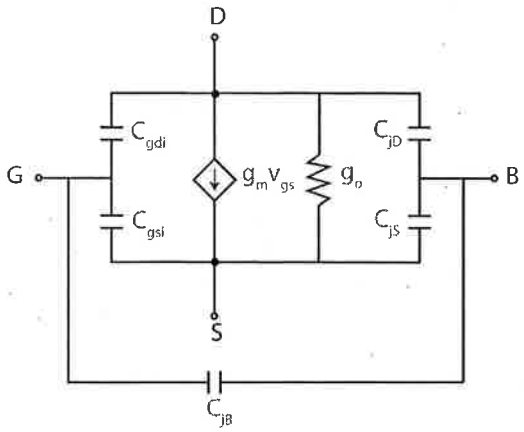


Figure 9.23: Small-signal equivalent circuit model of an ideal MOSFET.

flowing between drain and source is then controlled by the gate-to-source voltage. This is, of course, the transistor effect. To represent this, we need a voltage-controlled current source. In the second term, the drain to source current is modulated by v_{ds}, which is the voltage across the same two terminals. In the ideal MOSFET, this is the case in the linear regime, though this does not happen in the saturation regime. This effect, nevertheless, can be captured by means of a simple resistor of value $1/g_o$, or of a conductance equal to g_o. All together, the small-signal equivalent circuit model for the MOSFET is shown in Fig. 9.23.

We derived above expressions for all the capacitors in Fig. 9.23. We now derive expressions for g_m and g_o for the MOSFET operating in the various regimes.

By definition, the transconductance g_m is:

$$g_m = \frac{\partial I_D}{\partial V_{GS}}|_{V_{DS},V_{BS}} \tag{9.50}$$

In cut-off, $I_D = 0$, and $g_m = 0$. In the linear regime, we go back to Eq. 9.16, and get:

$$g_m = \frac{W}{L}\mu_e C_{ox} V_{DS} \tag{9.51}$$

In the saturation regime, we go back to Eq. 9.21, and get:

$$g_m = \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T) = \sqrt{2\frac{W}{L}\mu_e C_{ox} I_D} \tag{9.52}$$

Fig. 9.24 shows the evolution of g_m with the bias point graphed in various ways. In the linear regime, g_m is independent of V_{GS} and linearly dependent on V_{DS}. For a fixed V_{DS}, g_m in the linear regime is independent of I_D. In the saturation regime, g_m is independent of V_{DS} and is linearly proportional to V_{GS}. Hence, for a fixed V_{DS}, g_m follows a square-root evolution with I_D.

Now we turn our attention to g_o. The output conductance is defined as:

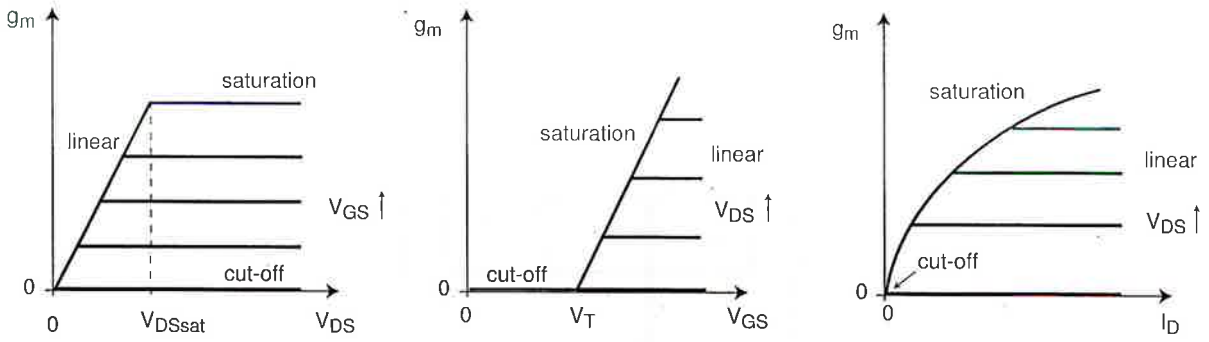


Figure 9.24: Evolution of g_m with V_{GS}, and V_{DS} in ideal MOSFET plotted in various ways.

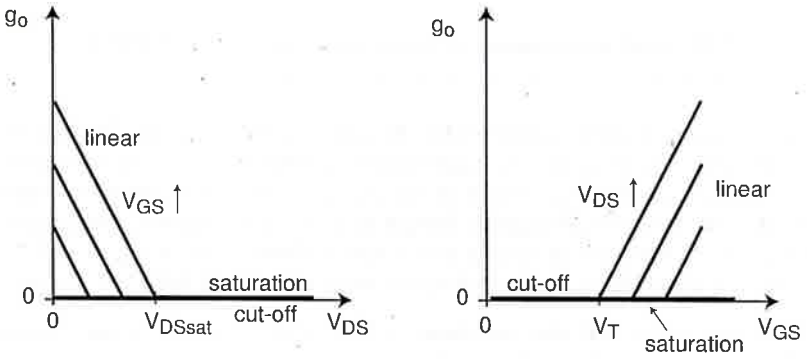


Figure 9.25: Evolution of g_o with V_{GS}, and V_{DS} in ideal MOSFET.

$$g_o = \frac{\partial I_D}{\partial V_{DS}}|_{V_{GS},V_{BS}} \tag{9.53}$$

In cut-off, $I_D = 0$, and $g_o = 0$. In the linear regime, we go back to Eq. 9.16 again, and get:

$$g_o = \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - V_{DS}) \tag{9.54}$$

In the saturation regime, the device is pinched-off and I_D is independent of V_{DS}. Then:

$$g_o = 0 \tag{9.55}$$

The evolution of g_o with V_{DS} and V_{GS} is shown in Fig. 9.25. In the ideal MOSFET, g_o is finite in the linear regime where it increases linearly with V_{GS} and decreases linearly with V_{DS} with an identical slope.

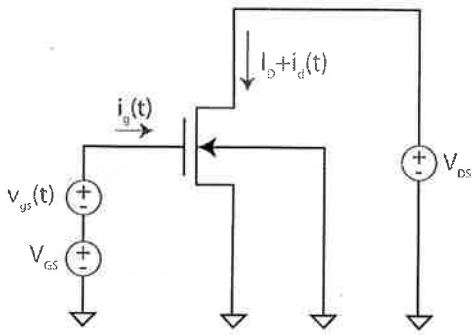


Figure 9.26: Circuit configuration for the evaluation of f_T in a MOSFET.

9.6.2 Short-circuit current-gain cut-off frequency, f_T, of ideal MOSFET in saturation

The short-circuit current-gain cut-off frequency, f_T, of a MOSFET in saturation is a figure of merit that is widely used to assess the suitability of MOSFETs for high-frequency mixed-signal applications. f_T is a widely quoted figure of merit that is indicative of electron transport through the intrinsic device. In this section, we present a definition of f_T, we derive an expression for it for the ideal MOSFET, and we provide a physically intuitive picture for the meaning of f_T.

f_T is defined with the MOSFET biased as shown in Fig. 9.26. Voltage sources on the gate and the drain are selected to bias the device in the saturation regime. The body is shorted out to the source (though this is not strictly necessary). A small signal voltage source is applied on top of the gate bias. We are interested in evaluating the small-signal *current gain* of the MOSFET, i_d/i_g, in these conditions. This is often referred to as h_{21}.

In DC, an ideal MOSFET does not draw gate current. As a result, its DC current gain is infinite. If we now wiggle the gate voltage up and down a little bit, the drain current will follow but, in addition, a displacement current flows into the gate to charge and discharge all the capacitors that hang from the gate terminal. This now results in finite current gain. The magnitude of the gate current increases as the frequency of the input signal increases, and in consequence, the current gain decreases. At some frequency the current gain goes to unity. This is f_T. f_T therefore provides an assessment of the high frequency suitability of MOSFET and, in general, transistors. It is a widely watched and quoted figure of merit. Mathematically, then we have:

$$|h_{21}(f_T)| = |\frac{i_d}{i_g}|_{v_{ds}=0} = 1 \tag{9.56}$$

Deriving an expression for f_T for the ideal MOSFET is fairly straightforward. The first step is to draw a small-signal equivalent circuit model for the circuit configuration. This is shown in Fig. 9.27. In this circuit, the MOSFET is in saturation. Hence $C_{jB} = C_{gdi} = 0$. Also $g_o = 0$. Since the body is shorted to the source, C_{jS} is shorted out. Also, from a small-signal point of view, C_{jD} is also shorted out since it is connected to a DC voltage source.

With $v_{gs}(t)$ applied at the input of this circuit, the gate and drain currents are given by:

$$i_g = j\omega C_{gsi} v_{gs} \tag{9.57}$$

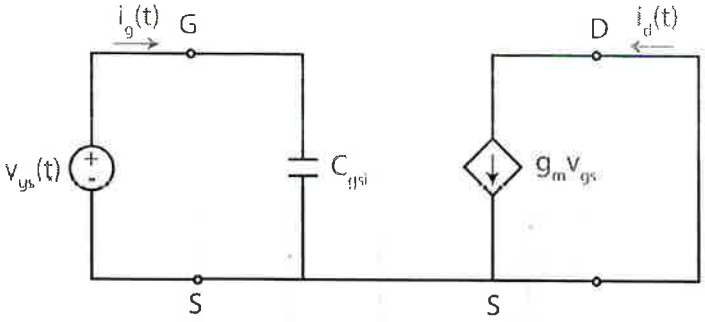


Figure 9.27: Small-signal equivalent circuit model for the computation of f_T in an ideal MOSFET.

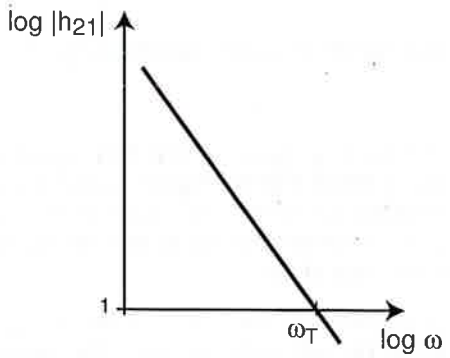


Figure 9.28: Bode plot of magnitude of $|h_{21}|$ versus frequency indicating ω_T.

$$i_d = g_m v_{gs} \tag{9.58}$$

h_{21} is then given by:

$$h_{21} = \frac{g_m}{j\omega C_{gsi}} \tag{9.59}$$

The magnitude of h_{21} is then:

$$|h_{21}| = \frac{g_m}{\omega C_{gsi}} \tag{9.60}$$

The evolution of $|h_{21}|$ with frequency is sketched in a Bode plot in Fig. 9.28. As expected, $|h_{21}|$ decreases with frequency. At a certain frequency, $|h_{21}| = 1$. The frequency at which this happens is given by:

$$f_T = \frac{g_m}{2\pi C_{gsi}} \tag{9.61}$$

This equation makes good physical sense. The higher g_m the higher the drain current that results in response to a wiggle in v_{gs} at the input and hence the higher the frequency at which the transistor exhibits

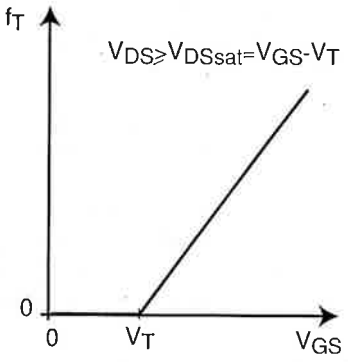


Figure 9.29: In a MOSFET in saturation, f_T is linearly dependent of V_{GS} and is independent of V_{DS}.

current gain. On the other hand, the higher C_{gsi}, all things being equal, the higher the gate current that the device draws, and the lower the frequency up to which the transistor offers current gain.

Let us discuss now the bias dependence of f_T. With an ideal MOSFET in saturation, we know that C_{gsi} is completely independent of the bias point. On the other hand, g_m increases linearly with V_{GS} and is independent of V_{DS}. Hence, f_T increases linearly with V_{GS} above threshold. f_T is also independent of V_{DS}. This can be seen if we substitute Eqs. 9.44 and 9.52 into Eq. 9.61 to get:

$$f_T = \frac{1}{2\pi}\frac{3}{2}\frac{\mu_e(V_{GS}-V_T)}{L^2} \tag{9.62}$$

This is sketched in Fig. 9.29.

Eq. 9.62 also gives a sense of the materials and structural parameters of the MOSFET that affect f_T. We note that f_T is linearly dependent on the electron mobility in the channel. This makes good sense as, all things being equal, a higher value of μ_e results in more drain current. f_T depends inversely on the *square* of the gate length. This is also reasonable since the shorter the gate length, the more drain current the device produces and the smaller the input capacitance.

It is also of interest to reflect on the dependences that are *not* seen in Eq. 9.62. First, f_T is independent of W. This is because both C_{gsi} and g_m are linearly proportional to the gate width of the MOSFET. Note also that f_T is independent of C_{ox}, or x_{ox}. This is again because both C_{gsi} and g_m have the same linear dependence on the capacitance of the MOS structure per unit area.

An intuitive picture of the physical meaning of f_T can be drawn from the realization that f_T has the units of inverse time. We can then define:

$$\tau_d = \frac{1}{2\pi f_T} = \frac{2}{3}\frac{L^2}{\mu_e(V_{GS}-V_T)} \tag{9.63}$$

τ_d is called the *delay time*. What is the physical meaning of τ_d?

We can understand this by asking the following question. How much time does it take for a perturbation applied to the gate of a MOSFET to propagate into the drain current?

Consider an ideal MOSFET biased in saturation with a certain gate voltage V_{GS}. Now consider that at a certain instant, $t = 0$, the bias of the gate is suddenly increased by a small amount ΔV_{GS}. The drain

current at $t = 0^-$ is $I_D(0^-) = \frac{W}{2L}\mu_e C_{ox}(V_{GS} - V_T)^2$. We know that, given enough time, the drain current will increase by an amount $\Delta I_D = \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T)\Delta V_{GS}$. The question is how much time does it take for this to happen?

At $t = 0^+$, the source current suddenly increases by an amount $\Delta I_S = -\frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T)\Delta V_{GS}$. This change is instantaneous. This is because the energy barrier that controls the entrance of electrons from the source into the channel gets instantaneously reduced by the change in V_{GS}. However, before this pulse of current shows up at the drain terminal, the inversion layer needs to be "charged up." At $t = 0^-$, we have $Q_I(0^-) = -\frac{2}{3}WLC_{ox}(V_{GS} - V_T)$. By the time the step in current reaches the drain, the inversion layer charge must have increased by an amount $\Delta Q_I = -\frac{2}{3}WLC_{ox}\Delta V_{GS}$. Since it is the source that is providing these electrons, the time that it takes for this charge to be delivered is:

$$\tau_d = \frac{\Delta Q_I}{\Delta I_S} = \frac{2}{3}\frac{L^2}{\mu_e(V_{GS} - V_T)} \tag{9.64}$$

It is only after this time has passed and the inversion layer has been fully charged up that the drain current increases. Note how this equation is identical to Eq. 9.63.

The delay time is then the time that it takes to charge up the inversion layer after a perturbation has been applied to the gate of a MOSFET. This delay sets the maximum frequency at which the MOSFET exhibits current gain.

Exercise 9.5: *Consider once again the long n-channel MOSFET of Exercises 9.1-9.4. Estimate g_m and f_T at a bias point in saturation for which $I_D = 2\ mA$.*

For g_m we can directly use Eq. 9.52:

$$g_m = \sqrt{2\frac{W}{L}\mu_e C_{ox} I_D} = \sqrt{2\frac{10\ \mu m}{1\ \mu m}500\ cm^2/V.s \times 2.3 \times 10^{-7}\ F/cm^2 \times 2 \times 10^{-3}\ A} = 2.1\ mS$$

For f_T we could use Eq. 9.61:

$$f_T = \frac{g_m}{2\pi C_{gsi}} = \frac{1}{2\pi}\frac{2.1 \times 10^{-3}\ A/V}{1.5 \times 10^{-14}\ F} = 22\ GHz$$

In this, we have used our estimate of C_{gsi} from Exercise 9.4.

9.7 Non-ideal effects in MOSFET

We have completed a first pass through the current-voltage and charge-voltage characteristics of the ideal MOSFET. We have obtained sound physical understanding about its operation and we have developed simple models for its behavior. We are now in an excellent position to evaluate some of the simplifying assumptions that were made in the definition of the "ideal" MOSFET, understand the physics behind a few important "non-ideal" effects and introduce corrections to the models to account for them.

9.7.1 Body effect

The body effect and the impact of back bias were briefly mentioned in Ch. 8. The underlying physics of both of these effects is identical and was explained in Section 8.7. In essence, in a MOS structure when a voltage is applied to the inversion layer with respect to the body underneath, the threshold voltage of the

structure is modified. For an electron inversion layer on a p-type substrate, the application of a positive voltage to the inversion layer with respect to the body results in an increase of the threshold voltage. The body effect and the role of back bias are both manifestations of this piece of physics in a MOSFET.

Let us consider first the *body effect*. Consider a MOSFET biased in the linear regime, as depicted in Fig. 9.7, with the body tied up to the source. As we discussed, there is a voltage drop along the channel from source to drain that we labeled $V(y)$. This was illustrated in Figs. 9.10 or 9.12. A consequence of the appearance of $V(y)$ is that there is a channel-to-body voltage difference. Since the body is tied up the source, this voltage difference is precisely $V(y)$.

On the source side of the channel, the channel-to-body voltage remains always at zero. However, as we advance towards the drain, it increases and it becomes V_{DS} at the drain end of the channel. This rise in channel voltage is what produces the lateral electric field that makes the electrons drift from source to drain. A consequence of the non-zero channel to body voltage is that the *local* threshold voltage in the inversion layer actually depends on channel location. Since V rises along the inversion layer from source to drain, the local V_T shifts positive along the way.

There are obvious consequences to this. At any one location other than the source end of the channel, the inversion layer charge will be smaller in absolute magnitude than in the ideal MOSFET. This is sketched in Fig. 9.30 which depicts the voltage drop along the channel of a MOSFET in the linear regime as well as the gate overdrive with and without the body effect. It is clear from this figure that the positive V_T shift that takes place along the channel reduces the gate overdrive throughout the device. The main consequence of this is a reduction in the current drivability of the MOSFET. Refining the ideal MOSFET model to account for the body effect is rather straightforward.

The starting point is Eq. 9.9 that describes the drift current in the channel of a MOSFET. The inversion layer charge at any one location in the channel is given by expression 9.11. This expresion should explicitly acknowledge the fact that V_T actually depends on position. We rewrite it here:

$$Q_i(y) \simeq -C_{ox}[V_{GS} - V(y) - V_T(y)] \tag{9.65}$$

Now we need to find a way to describe the spatial dependence of V_T. In Ch. 8 we studied the effect of applying a voltage to the inversion layer with respect to the body in a MOS structure. We argued there that this produced a change in the threshold voltage (referred from gate to source) that was given by Eq. 8.68. This is precisely the same situation that we have here. At a location y, there is a potential difference between the inversion layer in a MOSFET and the body. Hence the local value of V_T will also follow Eq. 8.68. Adapted to the notation used in the present chapter, this equation can be rewritten as:

$$V_T(V) = V_{To} + \gamma(\sqrt{\phi_{sT} + V} - \sqrt{\phi_{sT}}) \tag{9.66}$$

where V_{To} refers to the threshold voltage of the two-terminal MOS structure for $V_{SB} = 0$.

Inserting this equation into Eq. 9.65 and the result into Eq. 9.9, we get:

$$I_e = W\mu_e C_{ox}[V_{GS} - V - V_{To} - \gamma(\sqrt{\phi_{sT} + V} - \sqrt{\phi_{sT}})]\frac{dV}{dy} \tag{9.67}$$

This differential equation can be integrated in an identical way as Eq. 9.12. The resulting expression for the drain current is:

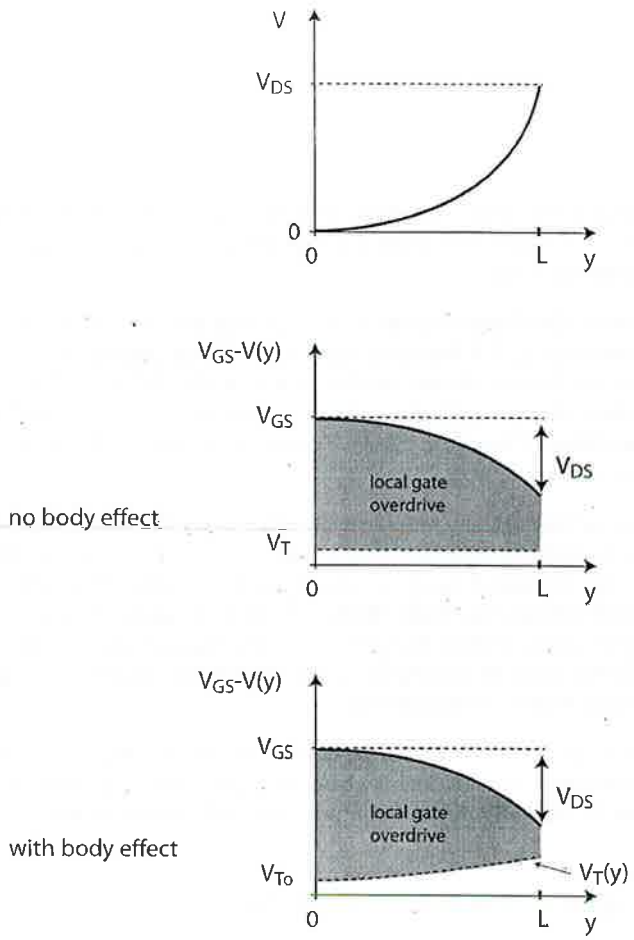


Figure 9.30: From top to botton, voltage, gate overdrive in the absence of body effect, and gate overdrive with body effect for a MOSFET in the linear regime.

$$I_D = \frac{W}{L}\mu_e C_{ox}\{(V_{GS} - V_{To} + \gamma\sqrt{\phi_{sT}} - \frac{1}{2}V_{DS})V_{DS} - \frac{2}{3}\gamma[(\phi_{sT} + V_{DS})^{3/2} - (\phi_{sT})^{3/2}]\} \tag{9.68}$$

If we compare this expression for the drain current with the one derived earlier in Eq. 9.16, we see three new terms that are multiplied by the body parameter γ. Clearly, if γ goes to zero, we revert back to Eq. 9.16. This is to be expected from our presentation in Ch. 8 where we saw that in the limit of $\gamma = 0$, the impact of the back bias in the electrostatics of the MOS structure becomes negligible.

It might not be clear just by looking at Eq. 9.68 that the aggregate effect of all these new terms is to *reduce* the drain current from the ideal value given in Eq. 9.16, but that is what happens. The magnitude of the reduction gets worse as γ increases. In spite of this, the basic behavior of the drain current with V_{DS} and V_{GS} is not seriously affected by the body effect. In particular, channel pinch-off and drain current saturation take place just as before except that the final expression of the saturated drain current and the value of V_{DSsat} are modified.

To get the new expression for V_{DSsat}, we follow an identical procedure as for the ideal MOSFET. We need to look at the expression for the inversion layer charge at $y = L$, and let that go to zero. From Eq. 9.17 and 9.66, we have:

$$Q_i(y = L) = -C_{ox}[V_{GS} - V_{DSsat} - V_{To} - \gamma(\sqrt{\phi_{sT} + V_{DSsat}} - \sqrt{\phi_{sT}})] = 0 \tag{9.69}$$

Solving for V_{DSsat}, we find:

$$V_{DSsat} = V_{GS} - V_{To} + \gamma\sqrt{\phi_{sT}} - \frac{\gamma^2}{2}[\sqrt{1 + \frac{4}{\gamma^2}(V_{GS} - V_{FB})} - 1] \tag{9.70}$$

where we have also used Eq. 8.28. As shown below, the body effect reduces the value of V_{DSsat}. In the limit of $\gamma = 0$, V_{DSsat} reverts to $V_{GS} - V_{To}$, as one should expect.

An expression for the saturated drain current can be obtained by substituting V_{DSsat} for V_{DS} in Eq. 9.68:

$$I_{Dsat} = \frac{W}{L}\mu_e C_{ox}\{(V_{GS} - V_{To} + \gamma\sqrt{\phi_{sT}} - \frac{1}{2}V_{DSsat})V_{DSsat} - \frac{2}{3}\gamma[(\phi_{sT} + V_{DSsat})^{3/2} - (\phi_{sT})^{3/2}]\} \tag{9.71}$$

Fig. 9.31 shows the impact of the body effect on the output I-V characteristics of a typical MOSFET. On this figure, the solid lines represent the I-V characteristics calculated using the first-order model (Eqs. 9.16 and 9.21). The dashed lines are computed using the model that includes the body effect (Eqs. 9.68 and 9.71). Three features are noticeable in this figure. First, for all values of V_{GS} and V_{DS}, including the body effect results in a reduction of the drain current. This was expected. Second, the deleterious effect of the body effect goes away as the device is turned off. This is easily understandable since in this instance, the potential drop along the channel becomes negligible. Third, at a given V_{GS} the body effect reduces the value of the drain to source voltage required to saturate the transistor.

The impact of the body effect on V_{DSsat} is more clearly seen on the left-hand side of Fig. 9.32. This figure makes evident that V_{DSsat} is reduced by the body effect. Interestingly, the dependence of V_{DSsat} on V_{GS} remains rather linear. This suggests that the dependence of I_{Dsat} on $V_{GS} - V_{To}$ might remain quadratic in the presence of the body effect. In fact, the log-log plot of I_{Dsat} as a function of $V_{GS} - V_{To}$ on the right hand side of Fig. 9.32 shows just this. The body effect reduces I_{Dsat}, but does not significantly change the nature of its dependence on gate overdrive.

These observations suggest that simpler but still fairly accurate expressions of V_{DSsat}, I_D and I_{Dsat} exist. These can be obtained by going back to Eq. 9.66 and linearizing the square root in the limit in which V is small with respect to ϕ_{sT}. This yields:

$$V_T(V) \simeq V_{To} + \frac{\gamma}{2\sqrt{\phi_{sT}}}V \tag{9.72}$$

This Taylor series expansion works well for small values of V but it should get progressively inacurate as V increases. Fortunately, V in the channel never gets too high due to pinch-off. [5]

We now retrace our steps one more time and insert Eq. 9.72 into Eq. 9.65 and the result into Eq. 9.9 to get a simple differential equation that when solved yields:

[5] Empirically, it has been found that a factor of 3, instead of 2, in the denominator of the second term in the right-hand side of Eq. 9.72 works better.

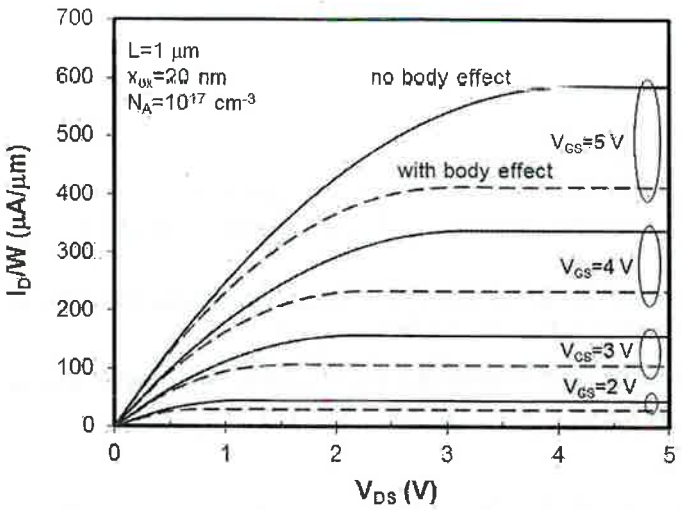


Figure 9.31: Calculated output I-V characteristics of a typical MOSFET using the first-order model without the body effect (solid lines) and including the body effect (dashed lines).

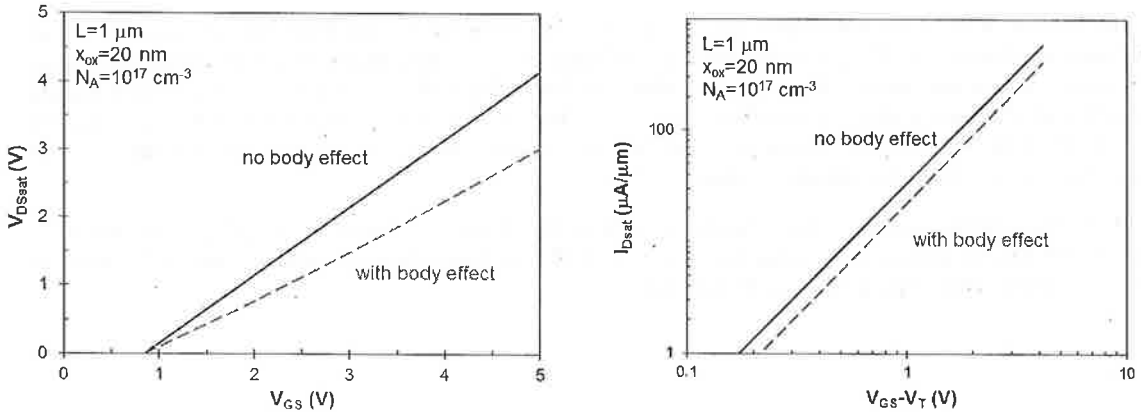


Figure 9.32: Impact of the body effect on V_{DSat} (left) and I_{Dsat} (right) for a typical MOSFET. The device is identical to that of Fig. 9.31. Calculations using the first-order model without the body effect are graphed on solid lines, including the body effect are graphed in dashed lines.

$$I_D \simeq \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_{To} - \frac{m}{2}V_{DS})V_{DS} \tag{9.73}$$

where m is given by:

$$m = 1 + \frac{\gamma}{2\sqrt{\phi_{sT}}} \tag{9.74}$$

The expression of V_{DSsat}, obtained in the same way as before, becomes:

$$V_{DSsat} \simeq \frac{1}{m}(V_{GS} - V_{To}) \tag{9.75}$$

Finally, the expression for the saturated drain current becomes:

$$I_{Dsat} \simeq \frac{W}{2mL}\mu_e C_{ox}(V_{GS} - V_{To})^2 \tag{9.76}$$

Equations 9.73, 9.75, and 9.76 are very similar to the expressions we derived for an ideal MOSFET earlier in this chapter, Eqs. 9.16, 9.22, and 9.21, respectively. The only difference is the appearance of the parameter m, the so-called *body-effect coefficient*. As seen in Eq. 9.74, m is always positive and larger than unity. As a result, all things being equal, I_D, I_{Dsat} and V_{DSsat} get reduced by the incorporation of the body effect. These new simplified equations are fairly accurate and much easier to remember than the exact equations derived earlier in this section. This can be seen in Fig. 9.33.

Interestingly, m has an identical expression to the ideality factor of the subthreshold regime discussed in Section 8.6. This should not be very surprising. In both the subthreshold regime and the body effect there is a competition between the electrostatics of the gate and the electrostatics of the body for control of the inversion layer charge. If the controlling action of the gate dominates over that of the body, then the subthreshold regime is sharp and the body effect is small. The contrary happens if the controlling action of the body becomes significant.

It is common practice when dealing with the long MOSFET to treat the parameter m as an adjustable one. It is found that values of m of the order of 1.1 to 1.4 work quite well.

9.7.2 Effect of back bias

A MOSFET is a *four-terminal* device. In addition to the source, drain and gate, there is also the region right underneath the transistor, also called the back or the *body*. In an integrated device, the body is contacted at the surface of the wafer. So far, we have not considered the impact of applying a body voltage on the I-V characteristics of the transistor. For the most part, any voltage that would be applied to the body with respect to the source will be negative or slightly positive. If a relatively large positive bias is applied, the body-source PN diode turns on adding an undesirable current path to the MOSFET. Hence we should mostly be concerned with applying negative values of V_{BS}.

We learned in Section 8.7 that the application of a negative voltage to the body with respect to the inversion layer results in a positive shift of the threshold voltage. All things being equal, then, if we have a MOSFET that is *ON*, the application of $V_{BS} < 0$ should result in a decrease in the current flowing through the transistor. This is because as V_T shifts positive, the overdrive on the gate of the MOSFET is reduced.

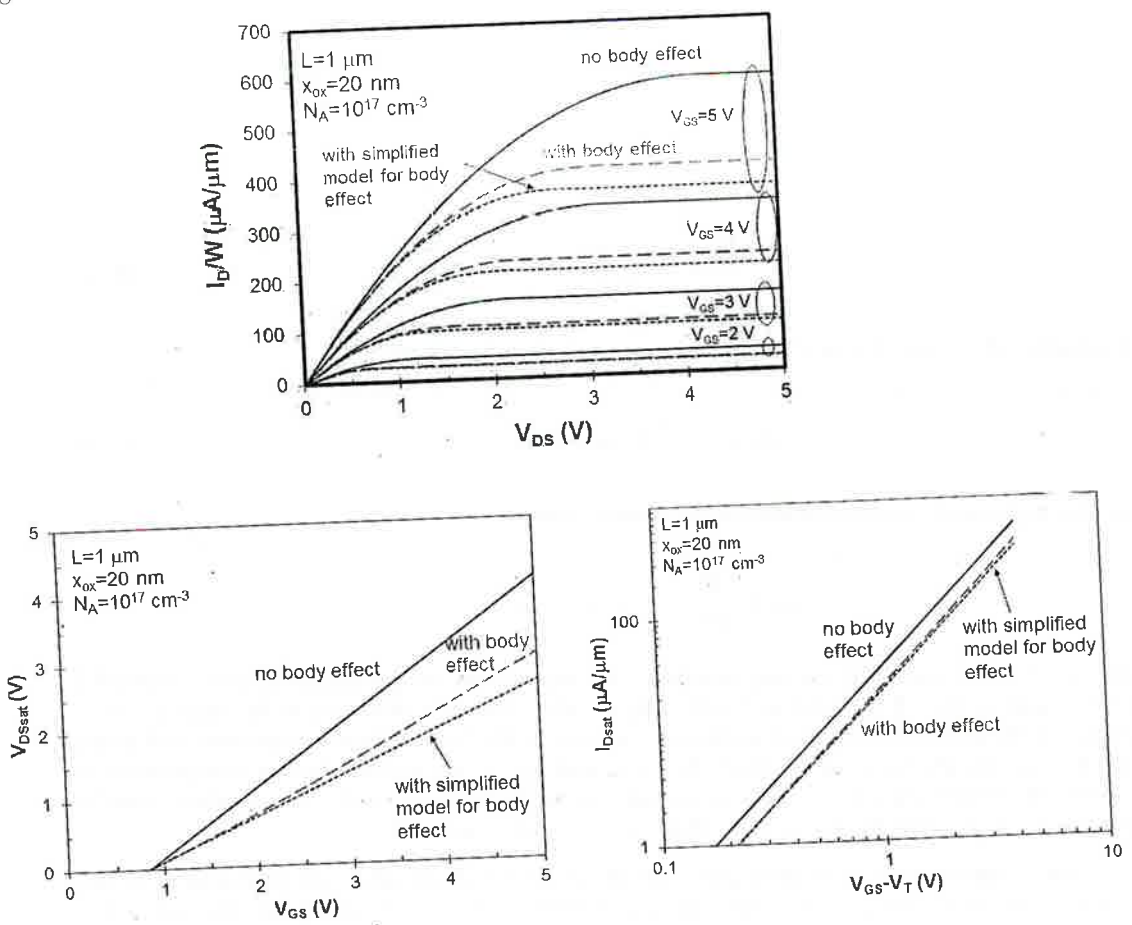


Figure 9.33: Comparison of behavior of simplified equations for output characteristics (top), V_{DSsat} (bottom left) and I_{Dsat} (bottom right) describing the body effect (short dash) against the exact equations (long dash). For reference, calculations without the body effect are also shown (solid). The device is identical to that of Fig. 9.31.

Modifying the simple first-order model developed above to account for this is straightforward. The shift in the threshold voltage as a result of applying a back bias is given by Eq. 8.68. Using the notation of this chapter, we can rewrite it as:

$$V_T(V_{BS}) = V_{To} + \gamma(\sqrt{\phi_{sT} - V_{BS}} - \sqrt{\phi_{sT}}) \tag{9.77}$$

The ideal I-V characteristics of the MOSFET captured in Eqs. 9.16 and 9.21 can then be made to account for the back bias effect provided that Eq. 9.77 is used for the threshold voltage. It is clear in this formulation that as V_{BS} becomes more negative, V_T in Eq. 9.77 goes up along with it, and I_D given by Eqs. 9.16 and 9.21 goes down.

In addition to the straightforward impact of the application of a back bias on the threshold voltage just described, the characteristics of the MOSFET are also affected through the term that accounts for the body effect. It is easy to see why this should be the case. Both the body effect and the back bias effect arise from the electrostatic influence of the body on inversion layer charge. The application of a back bias with $V_{BS} < 0$ results in a widening of the depletion region underneath the inversion layer. This brings

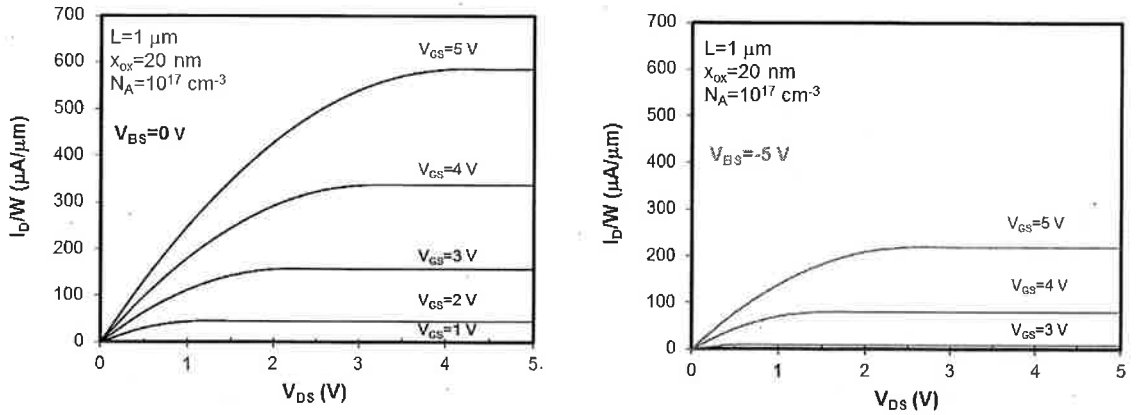


Figure 9.34: Effect of back bias on output characteristics of long MOSFET: $V_{BS} = 0\ V$ (top), $V_{BS} = -5\ V$ (bottom). The device is identical to that of Fig. 9.31.

the "back gate" further away from the inversion layer and weakens the impact of the body effect on the transistor. Let us see how the model that accounts for the body effect is to be modified to include also the impact of back bias.

We start from Eq. 9.66 which gives the expression of the threshold voltage along the inversion layer accounting for the potential drop that occurs along the channel. In the presence of a back bias, this equation becomes:

$$V_T(V_{BS}, V) = V_{To} + \gamma(\sqrt{\phi_{sT} - V_{BS} + V} - \sqrt{\phi_{sT}}) \tag{9.78}$$

To keep the model simple, we can expand this equation around $V = 0$, to get:

$$V_T(V_{BS}, V) \simeq V_T(V_{BS}) + \frac{\gamma}{2\sqrt{\phi_{sT} - V_{BS}}} V \tag{9.79}$$

where $V_T(V_{BS})$ is given by Eq. 9.77.

Eq. 9.79 is nearly identical to Eq. 9.72 except for the addition of V_{BS} to ϕ_{sT} inside the square root. Following the development carried out in the previous subsection, an identical set of equations for I_D emerges if we use a body-effect coefficient that depends on the back bias, as given by:

$$m(V_{BS}) = 1 + \frac{\gamma}{2\sqrt{\phi_{sT} - V_{BS}}} \tag{9.80}$$

It is then clear that the application of $V_{BS} < 0$ increases V_T and reduces m. The dominant effect is a reduction in I_D and I_{Dsat} given by Eqs. 9.73 and 9.76 respectively, as expected. This can be seen in the output characteristics of a MOSFET for a back bias of 0 V and 5 V in Fig. 9.34.

The existence of a dependence of the drain current on the back bias implies that if a small-signal is applied to the substrate terminal with respect to the source, this signal will propagate to the drain current. We actually recognized this when in Sec. 9.6.1 we developed the small-signal model for the MOSFET and defined a *back transconductance*, g_{mb} (Eq. 9.49). Now we are in a position to derive an expression for g_{mb}

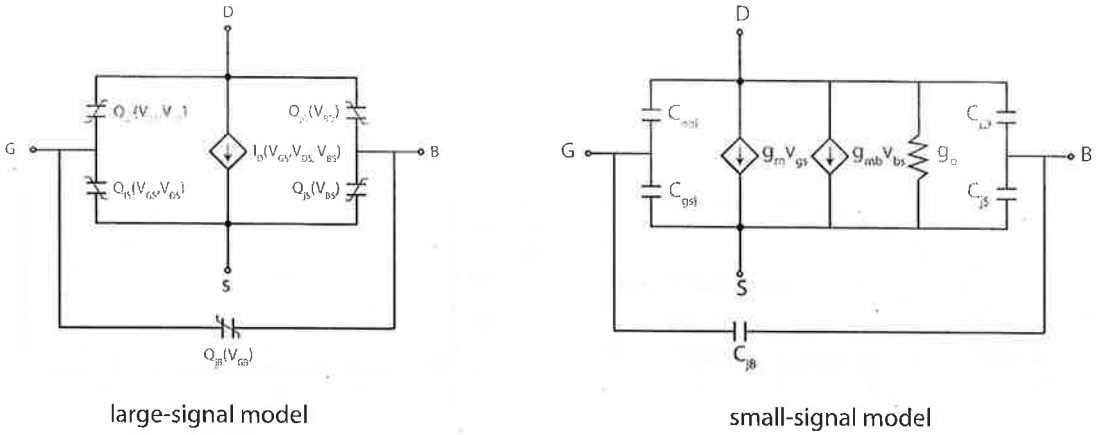


Figure 9.35: MOSFET equivalent circuit models incorporating the impact of the back contract. Left: large-signal model with a V_{BS} dependence made explicit. Right: small-signal model incorporating a new current source driven by v_{bs}.

for the ideal MOSFET. This is straightforward because a small-signal change in V_{BS} impacts the drain current through a small change in V_T. We can then write:

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}}|_{V_{DS},V_{GS}} = \frac{\partial I_D}{\partial V_T}|_{V_{DS},V_{GS}} \times \frac{\partial V_T}{\partial V_{BS}}|_{V_{DS},V_{GS}} \tag{9.81}$$

In an ideal MOSFET in saturation, the first derivative in this equation is precisely $-g_m$. This is because in saturation the drain current depend on $V_{GS} - V_T$, so the dependence of I_D on V_T and $-V_{GS}$ are identical. The second derivative can be easily performed from Eq. 9.77 and expressed in terms of m defined in Eq. 9.80. All together we obtain:

$$g_{mb} = (m-1)g_m \tag{9.82}$$

It is not surprising that the body-effect coefficient, m, makes an appearance here. As discussed above, m reflects the competition between the substrate and the gate for control of the channel surface potential. In general $m > 1$ and g_{mb} is finite. The closer m gets to 1, the more the gate prevails over the body. In the limit of $m = 1$, the body exerts no control over the channel and therefore g_{mb} goes to zero.

Equivalent circuit models incorporating the back bias effect and the associated back transconductance are sketched in Fig. 9.35. In the large-signal model (left), an additional dependence of I_D on V_{BS} is explicitly noted. In the small-signal equivalent circuit model (right), a new current source is added between drain and source that is controlled by v_{bs}.

9.7.3 Channel length modulation

In the ideal MOSFET model developed earlier in this chapter, the current-voltage characteristics are perfectly saturated for $V_{DS} > V_{DSsat}$. This is the saturation regime of operation. However, if we look at a real MOSFET, such as the 2N7000 depicted in Fig. 9.15, we see a small but distinct slope in the I_D-V_{DS} characteristics. The slope also increases with V_{GS}. This imperfectly saturated drain current greatly impacts many circuit applications of MOSFETs, such as the voltage gain in MOSFET amplifiers and the

noise margins of CMOS logic gates, just to cite two examples. This section explains the physical origin of this effect and presents a simple model that accounts for the key dependencies.

The finite output conductance of a MOSFET has its origin in the electrostatics of the pinchoff point. An appropriate treatment is beyond our reach at the moment and has to wait until the short MOSFET is discussed in the next chapter. Nevertheless, reasonable understanding and a first-order model can be derived by borrowing a few results from Ch. 10.

In Sec. 9.4.3 we learned that with the MOSFET in the linear regime, the inversion layer electron velocity increases and the electron concentration drops as we proceed from source to drain. We also learned that the drop in carrier concentration is more pronounced the higher V_{DS} is. We called this channel debiasing. At pinchoff, our very simplistic model predicted that the electron concentration at the drain-end of the channel becomes zero and the electron velocity becomes infinite. Obviously, this is not possible. Electrons cannot travel at velocities higher than the saturation velocity and since the current is finite, the electron concentration cannot drop to zero.

In fact, as we will study in detail in Ch. 10, current saturation in a MOSFET actually occurs when the electron velocity at the drain end of the channel reaches the saturation velocity. At this condition the electron concentration at the pinchoff point is small but finite and the current that flows is also finite. Ideally, a further increase in V_{DS} beyond the bias that yields this condition does not directly increase the electron velocity any further and the current therefore should not change.

An important word in the previous phrase is "directly." To the first order, an increase in the drain voltage beyond V_{DSsat} should not translate into an increased potential build-up along the channel of a MOSFET. Instead, as we will see in Ch. 10, it results in an enlargement of the pinchoff region where the carrier velocity is saturated. This effectively shortens the electrical channel length. Since in this instance the same potential build up (V_{DSsat}) takes place over a shorter length, this implies that the electric field along the channel increases a little bit and so does the drain current. It is because of this effective shortening of the channel length that this effect is given the name of "channel length modulation."

Building a model for channel length modulation essentially means solving this complex two-dimensional electrostatics problem. A simplified treatment is presented in the next chapter. We qualitatively discuss the key findings here.

Fig. 9.36 zooms into the drain end of the channel of a MOSFET in saturation with $V_{DS} > V_{DSsat}$. As discussed, at the drain end of the channel, a velocity saturation region of length l_{sat} appears. Fig. 9.37 sketches the volume charge density, electron velocity, lateral electric field, and lateral potential along the semiconductor surface with the MOSFET in saturation for $V_{DS} = V_{DSsat}$ and $V_{DS} > V_{DSsat}$. This is an exaggerated picture to highlight the key issues.

At the onset of saturation, for $V_{DS} = V_{DSsat}$, we see that the electron concentration drops along the channel and reaches a minimum value right at the drain end. At that same point, the electron velocity peaks at its saturation velocity value. The electric field increases along the channel and peaks also at the drain end at a value that saturates the electron velocity. We label this as $\mathcal{E}_{vs}$. The channel voltage builds up along the channel up to a value of V_{DSsat} at its drain end.

When $V_{DS} > V_{DSsat}$, the additional voltage that is applied is used to extend the region over which the electron velocity is saturated to a length l_{sat}. The effective length of the channel over which the voltage builds up to V_{DSsat} is therefore shortened from L to L'. This results in a small increase of the electric field everywhere including at the source. Since the charge is constant at the source, this implies that the drain current increases a little bit. This is what is responsible for the output conductance of the device.

Modeling this effect requires deriving an expression for l_{sat} as a function of V_{DS}. This is what is done in Ch. 10. To the first order, if we assume that velocity saturation occurs at a constant lateral field $\mathcal{E}_{vs}$,

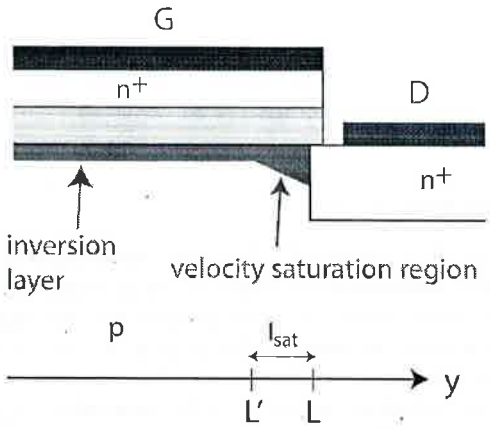


Figure 9.36: Geometry of the pinch-off region in a MOSFET for $V_{DS} > V_{DSsat}$.

as indicated in Fig. 9.37, and that the effect is relatively small, we should expect, approximately:

$$l_{sat} \simeq \frac{V_{DS} - V_{DSsat}}{|\mathcal{E}_{vs}|} \tag{9.83}$$

We can now reuse the drain current model developed earlier in this chapter except that we substitute L for $L - l_{sat}$. Therefore, we have:

$$I_{Dsat} \propto \frac{1}{L - l_{sat}} = \frac{1}{L(1 - \frac{l_{sat}}{L})} \simeq \frac{1}{L}(1 + \frac{l_{sat}}{L}) = \frac{1}{L}(1 + \frac{V_{DS} - V_{DSsat}}{|\mathcal{E}_{vs}|L}) \tag{9.84}$$

where we have assumed that in a well designed device, $l_{sat} \ll L$.

The second term inside the parentheses represents the correction term that is introduced in the MOSFET saturation current model as a result of channel length modulation. It is proportional to $V_{DS} - V_{DSsat}$. The proportionality constant is often written as λ, defined as the *channel length modulation parameter*, with units of inverse volt:

$$\lambda = \frac{1}{\mathcal{E}_{vs}L} \tag{9.85}$$

Using λ, we can write the complete expression of the saturation current of a MOSFET in the presence of channel length modulation as [6]:

$$I_{Dsat} = \frac{W}{2L}\mu_c C_{ox}(V_{GS} - V_T)^2[1 + \lambda(V_{DS} - V_{DSsat})] \tag{9.86}$$

The correction term on the saturation current depends only on $V_{DS} - V_{DSsat}$. However, this term is multiplied by the ideal saturation current. Hence, it becomes more visible in the output characteristics,

[6] In analogy with the Early effect in the bipolar transistor which also results in a finite output conductance, sometimes $1/\lambda$ in a MOSFET is referred to as the *Early voltage*.

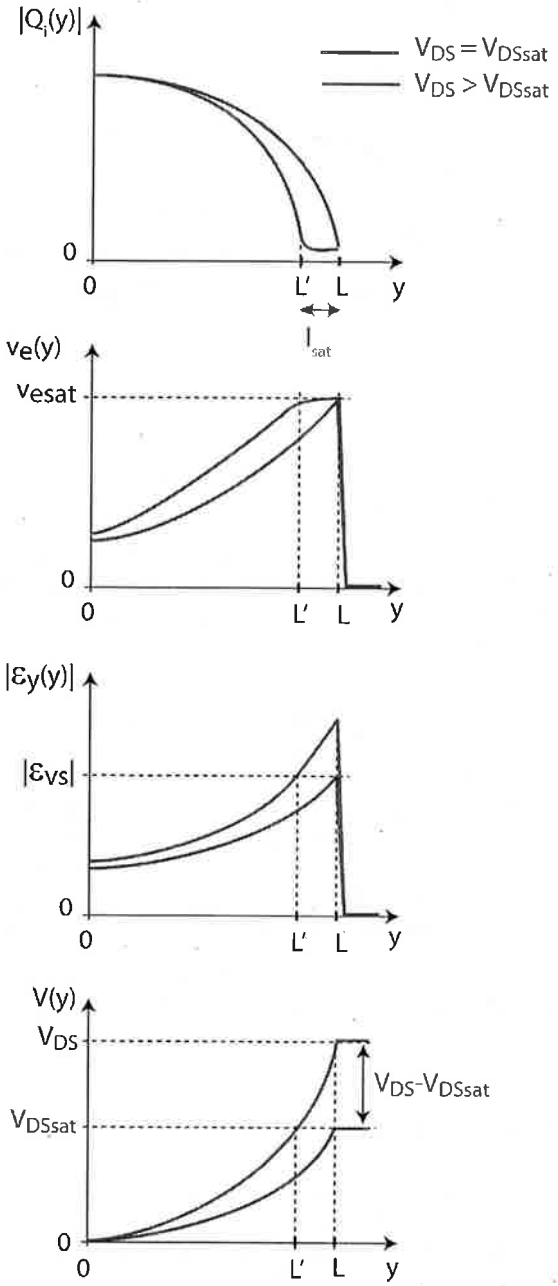


Figure 9.37: Electrostatics along the surface of the semiconductor for the pinch-off region of a MOSFET in saturation. From top to bottom: inversion layer charge density, electron velocity, lateral electric field and channel voltage.

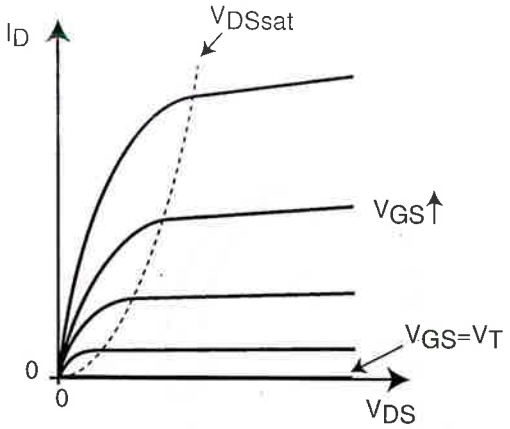


Figure 9.38: Sketch of the output characteristics of the MOSFET in the presence of channel length modulation.

the higher the value of V_{GS}. This is shown in the sketch of the output characteristics of Fig. 9.38 and agrees with what is observed in practice.

Channel length modulation impacts the small-signal equivalent circuit model of the MOSFET. In saturation, the output conductance of the ideal MOSFET model is zero (Eq. 9.55). Obviously, channel length modulation introduces a finite amount of output conductance. This can be easily obtained by applying the definition of Eq. 9.53:

$$g_o = \frac{\partial I_D}{\partial V_{DS}}|_{V_{GS},V_{BS}} \simeq \lambda I_{Dsat} \tag{9.87}$$

The important result here is that in a given device, g_o increases with I_{Dsat}. This matters a lot in analog circuit design where g_o is often a crucial consideration. In a given transistor, the only way to improve g_o (i.e., make it smaller) is to operate it at a lower current level. Alternatively, in applications where low g_o is critical, analog circuit designers often choose devices with longer gate lengths. As seen in Eq. 9.85 this results in a lower value of λ and a reduced output conductance at a certain current level.

Fig. 9.39 sketches the evolution of g_o in a MOSFET now accounting for the impact of channel length modulation. Fig. 9.40 shows the output conductance of the 2N7000 MOSFET whose output characteristics were shown in Fig. 9.15.

9.7.4 The subthreshold regime

In the ideal MOSFET model we assumed that the drain current is zero for $V_{GS} < V_T$. We referred to this as the cut-off regime. In a real MOSFET, the current decreases as V_{GS} drops below V_T, but it never quite becomes zero. For example, Fig. 9.41 shows the output characteristics of the 2N7000 MOSFET plotted in a semilog scale. Graphing the characteristics in this way allows us to examine in more detail the behavior of the MOSFET over a broader range of currents. It is clear from this figure that below a gate-source voltage of about 2 V (which seems a reasonable estimate of the threshold voltage according to Fig. 9.15), the current drops very fast as V_{GS} is reduced. However, down to $V_{GS} = 1$ V, it is still fairly sizable.

There are two interesting features that are noticeable in the data of Fig. 9.41. First, the current drops in what appears to be an exponential way. For values of V_{GS} between 1 and 2 V, the current lines appear

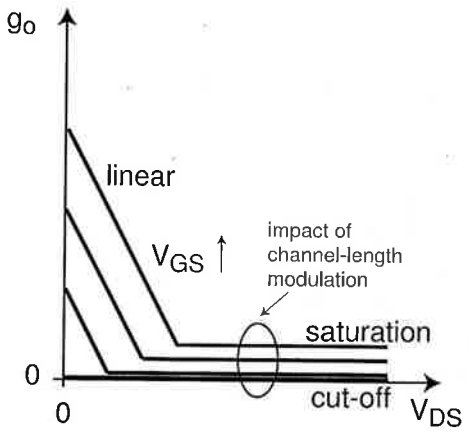


Figure 9.39: Evolution of g_o with V_{GS}, and V_{DS} in MOSFET showing the impact of channel length modulation.

relatively evenly spaced. Since V_{GS} is being stepped in a linear way, this suggest that I_D is evolving in an exponential way with V_{GS}. This can be better seen in the data of Fig. 9.42 which shows the drain current as a function of V_{GS} in a semilog scale. Indeed, below a value of $V_{GS} \simeq 2\ V$, the current-voltage characteristics follow a straight line.

The second interesting feature that can be seen in Fig. 9.41 is that the current is perfectly saturated with V_{DS} down to the lowest value of V_{DS} that can be resolved, which in this measurement is 120 mV. In contrast, for $V_{GS} > V_T$ it takes quite a bit of drain-to source voltage to saturate the drain current.

We are in front of what is termed the subthreshold regime. This is a regime of operation of great significance for analog and logic applications. In logic circuits, for example, we look at the MOSFET as a switch. This means that for $V_{GS} < V_T$, we really want the switch to be open, that is, the drain current to be zero. However, as we can see in Figs. 9.41 and 9.42, the drain current never quite becomes zero, the switch "leaks" a little bit. In consequence, there is some power consumption associated with the imperfectly open switch. Since the current is small, this power can be rather small. However, in a logic IC that integrates billions of transistors, this can add up to a sizable amount of power that needs to be delivered and dissipated.

In this section, we wish to understand the physics of the subthreshold current and develop a simple model that captures the leading dependencies.

At the heart of the subthreshold regime of a MOSFET is the *weak inversion* regime of the MOS structure that we discussed in Section 8.6. Below threshold, we found that electron charge at the surface of the semiconductor (it is not proper to denote this as inversion layer) did not quite go to zero but dropped in an exponential way with V, the gate to body voltage (see Eq. 8.62). This suggests that the exponential dependence of the subthreshold current in the MOSFET must arise from the charge control relationship in weak inversion. The interesting question is how does this charge flow from source to drain to produce drain current?

In a MOSFET below threshold, the dominant electrostatic feature underneath the gate is the presence of a depletion region. To be sure, there are a few electrons inside this depletion region. These are those that are responsible for transport in the subthreshold regime. However, their concentration is much smaller than the acceptor concentration. The electrostatics are then dominated by the volume charge density that arises from the uncompensated acceptors. All this charge is imaged at the gate. In this situation, the drain has no way to manipulate the surface potential under the gate. This is unlike in the linear regime

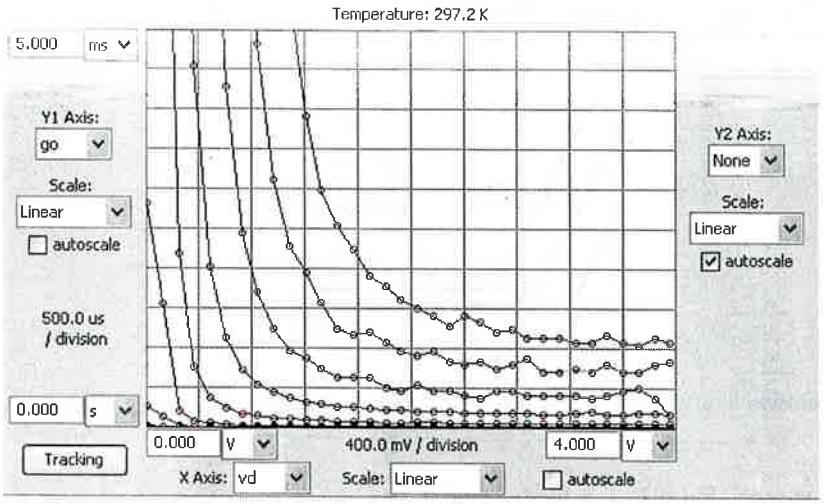


Figure 9.40: Measured output conductance of a 2N7000 MOSFET corresponding to the output characteristics shown in Fig. 9.15. V_{GS} is stepped from 1 to 3 V in steps of 0.2 V (screen shot from MIT Microelectronics iLab).

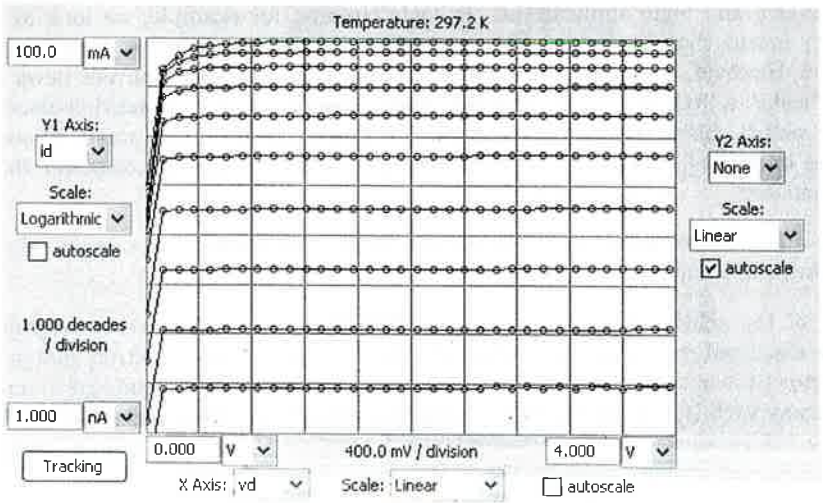


Figure 9.41: Measured output I-V characteristics of a 2N7000 MOSFET. These characteristics are identical to those of Fig. 9.15 except that I_D is shown in a semilogarithmic scale. V_{GS} is stepped from 1 to 3 V in steps of 0.2 V (screen shot from MIT Microelectronics iLab).

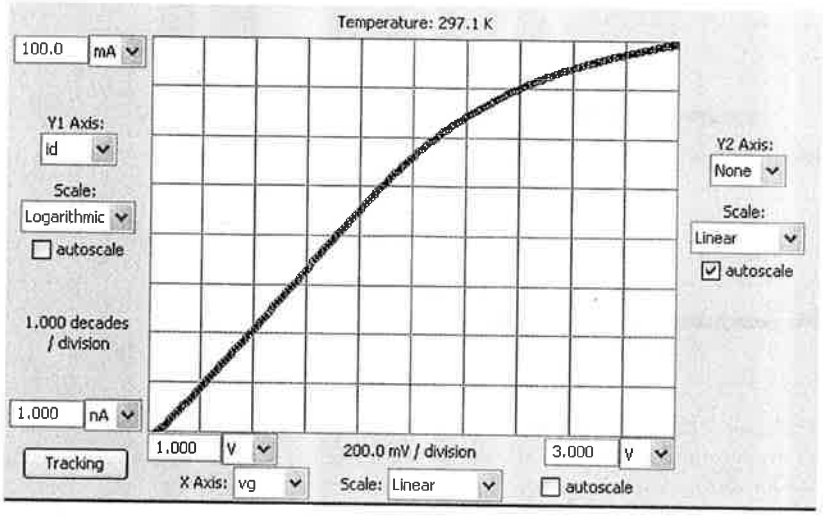


Figure 9.42: Measured drain current vs. gate-source voltage for a 2N7000 MOSFET in a semilog scale. $V_{DS} = 2\,V$ (screen shot from MIT Microelectronics iLab).

in which the drain electrically "grabs" on the inversion layer at the end of the channel and can therefore affect the surface potential along the entire length of the channel. There is no electrical channel to grab on in the subthreshold regime. How, then, can current flow?

We can understand what is happening by looking at Fig. 9.43 which sketches the evolution of the depletion region (left) in a MOSFET biased in subthreshold, and the conduction band edge (right) under the source, gate and drain as a function of V_{DS}.

With $V_{DS} = 0$, the depletion region under the source and drain are identical. The electron quasi-Fermi level is completely flat along the surface of the semiconductor and the electron concentration is uniform everywhere under the channel and given by Eq. ?? where $V = V_{GS} = V_{GD}$.

Increasing V_{DS} widens the depletion region under the drain and slightly sideways under the channel. From an energy point of view, increasing V_{DS} brings the conduction band edge and the quasi-Fermi level for electrons in the drain down with respect to the source by an amount equal to qV_{DS}. Note, however, that the conduction band edge under the gate remains flat. This is because the application of a voltage to the drain with respect to the source has not introduced a lateral electric field on the semiconductor underneath the gate. The electric field in the semiconductor under the gate points vertical.

The interesting question now is what happens to the electron concentration underneath the gate. It is clearly not going to be uniform since the distance between the conduction band edge and the quasi-Fermi level for electrons changes along the surface of the semiconductor from source to drain.

At the source end, the energy band diagram is not modified at all with respect to the situation that we had at $V_{DS} = 0$. Hence, from Eq. 8.62, we have:

$$Q_e(y=0) = -\frac{kT}{q} C_{sT} \exp \frac{q(V_{GS} - V_T)}{nkT} \tag{9.88}$$

where C_{sT} is the capacitance associated with the depletion region in the semiconductor at threshold and was given by Eq. 8.51, and n is the ideality factor given by:

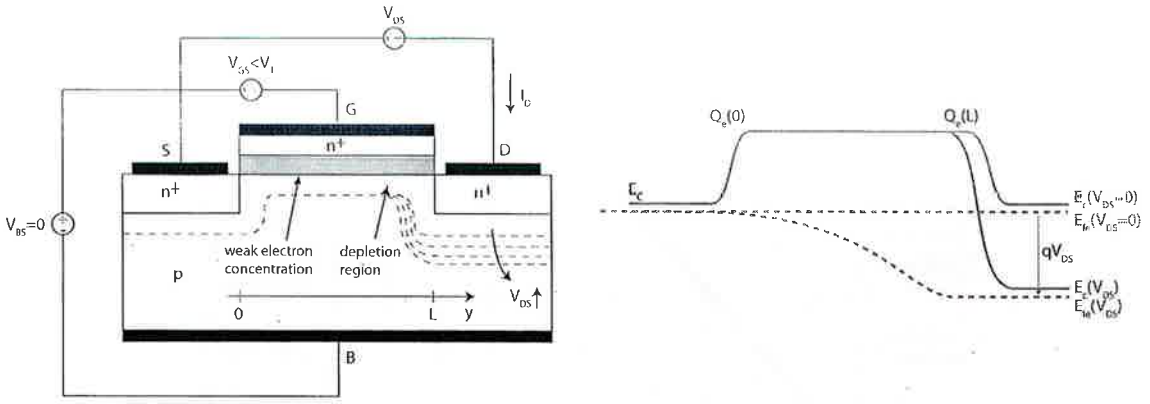


Figure 9.43: Left: sketch of MOSFET in subthreshold regime showing the evolution of the depletion region under the source, gate and drain for different values of V_{DS}. Right: sketch of the conduction band edge along the surface of the semiconductor for different values of V_{DS}.

$$n = 1 + \frac{C_{sT}}{C_{ox}} \tag{9.89}$$

Under the drain the situation is quite different. The application of V_{DS} brings down the conduction band edge and the quasi-Fermi level of electrons inside the drain with respect to the situation at $V_{DS} = 0$. Since the conduction band diagram is flat under the gate, right under the drain edge of the gate, the quasi-Fermi level for electrons has dropped below the conduction band edge by an amount equal to qV_{DS}. This means that the electron concentration must have dropped too. In fact, the electrostatics of the problem at the drain end of the channel are identical to those of a situation that we could obtain if we had a two-terminal MOS structure with a voltage applied between the gate and the body equal to $V_{GS} - V_{DS}$. This implies that the electron charge under the gate at the drain end is given by:

$$Q_e(y = L) = -\frac{kT}{q} C_{sT} \exp \frac{q(V_{GS} - V_{DS} - V_T)}{nkT} \tag{9.90}$$

This is obviously much less than under the source end of the gate.

We have an interesting situation here. The electron concentration at the semiconductor surface under the drain end of the gate is much smaller than under the source end. This is going to result in *electron diffusion* from source to drain and a current flow! This is the subthreshold current.

Obtaining an expression for the subthreshold current is now straightforward. It must be given by Fick's second law which states that electron diffusion current is directly proportional to the gradient of electron concentration. In this situation, this implies that:

$$I_D = W D_e \frac{dQ_e(y)}{dy} \tag{9.91}$$

Since electrons have nowhere to go but from source to drain (they cannot escape to the substrate because there is a potential barrier due to the band bending associated with the depletion region), the electron current must be constant along the surface of the semiconductor. This means that the gradient

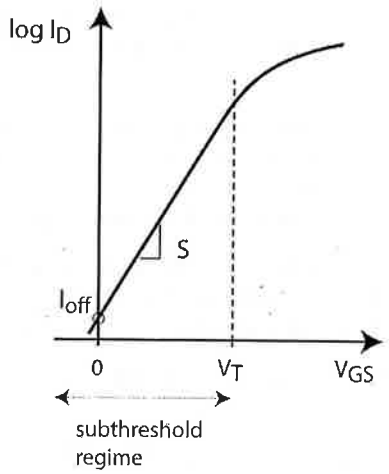


Figure 9.44: Sketch of subthreshold characteristics of a MOSFET.

of electrons must also be a constant or that the electron profile is a straight line. From Eq. 9.91, it then follows that:

$$I_D = WD_e \frac{Q_e(y=L) - Q_e(y=0)}{L} \tag{9.92}$$

Plugging in the results obtained in Eqs. 9.88 and 9.90, we get:

$$I_D \simeq \frac{W}{L} D_e \frac{kT}{q} C_{sT} \exp\frac{q(V_{GS}-V_T)}{nkT}(1-\exp\frac{-qV_{DS}}{nkT}) \tag{9.93}$$

This is the result that we were looking for and is graphed in Fig. 9.44.

Let us examine some of the features of this equation. First, it is clear that I_D evolves with V_{GS} following an exponential law. The ideality factor of the exponential dependence is given by Eq. 9.89. Second, for $V_{DS} = 0$, I_D is zero as it should be. Third, as V_{DS} increases, its impact on the subthreshold current quickly vanishes. This is because the $\exp\frac{-qV_{DS}}{nkT}$ rapidly becomes negligible as V_{DS} increases beyond a few thermal voltages. This is clearly seen in the experimental data of Fig. 9.41.

The slope of the subthreshold characteristics is characterized through the *subthreshold swing* (confusingly, also referred to as "subthreshold slope" or "inverse subthreshold slope") which is defined as:

$$S = \frac{nkT}{q}\ln 10 = (1+\frac{C_{sT}}{C_{ox}})\frac{kT}{q}\ln 10 \tag{9.94}$$

An ideality factor of 1 corresponds to a subthreshold swing of 60 mV/dec at room temperature. This is what the PN diode exhibits. In comparison with this, since in a MOSFET $n > 1$, the subthreshold swing is higher and the subthreshold current rises more gently with V_{GS}.

There are two key dependencies in the subthreshold swing. One is temperature. The higher the temperature, the higher the subthreshold swing. This is just like in the PN diode. The second one is in the ratio C_{sT}/C_{ox}. Anything that minimizes this ratio will sharpen the subthreshold swing of the

MOSFET. What we essentially want is $C_{ox} \gg C_{sT}$, or that the gate exercises a much tighter electrostatic control over the surface potential than the substrate. This suggests the use of thinner gate oxides and lower body doping levels.

One final point to make about the subthreshold swing is that the body voltage, in addition to shifting the subthreshold characteristics by acting on the threshold voltage, also impacts the subthreshold swing. The more negative the body voltage, the deeper the depletion layer becomes underneath the gate and the lower C_{sT} ends up being. In consequence, the subthreshold swing becomes sharper.

The subthreshold swing is important because it sets the value of the off-state current, I_{off}, a key figure of merit in a MOSFET that is used in CMOS logic applications. I_{off} is defined as $I_D(V_{GS} = 0, V_{DS} = V_{DD})$, where V_{DD} is the maximum voltage available in the circuit. I_{off} is marked in Fig. 9.44.

As expression for I_{off} can be easily obtained from Eq. 9.93:

$$I_{off} = I_D(V_{GS} = 0, V_{DS} = V_{DD}) \simeq \frac{W}{L} D_e \frac{kT}{q} C_{sT} \exp \frac{-qV_T}{nkT} \tag{9.95}$$

With the subthreshold swing defined as in Eq. 9.94, this can also be written as:

$$I_{off} \propto 10^{-\frac{V_T}{S}} \tag{9.96}$$

Writing I_{off} this way brings in evidence the two key factors that affect it: the threshold voltage and the subthreshold swing. Increasing V_T decreases I_{off} exponentially. The key trade-off here is that the gate overdrive, $V_{DD} - V_T$, is reduced and the drain current in saturation drops. One can also obtain low I_{off} by lowering the subthreshold swing. The relevant dependencies of S have been studied above.

9.7.5 Source and drain resistance

Our model for the ideal MOSFET so far does not include any resistance effects. In real MOSFETs, there are two resistances associated with the source and drain that are connected in series with the channel. They arise from the contact resistance of the metal contact to the n^+-regions, the finite lateral resistance of these regions and their linkage to the inversion layer under the gate edge. These series resistances represent significant parasitics with often very visible and deleterious impact on the device electrical characteristics.

The most important consequence of the presence of source and drain resistances in a MOSFET is a reduction in the current driving ability of the device. This can be understood through the simple model depicted in Fig. 9.45. This model explicitly places the source and drain resistances in series with the ideal device. When current flows through the MOSFET, the ohmic drops in the source and drain resistances reduce the voltage drive that is applied across the internal nodes of the transistor. This results in a reduction of the drain current below that of the ideal device. We can quantify this through a simple model.

If we denote as V_{DS} and V_{GS} the external terminal voltages, and V_{DSi} and V_{GSi} the internal node voltages (see Fig. 9.45), we can write:

$$V_{GSi} = V_{GS} - I_D R_S \tag{9.97}$$

$$V_{DSi} = V_{DS} - I_D(R_S + R_D) \tag{9.98}$$

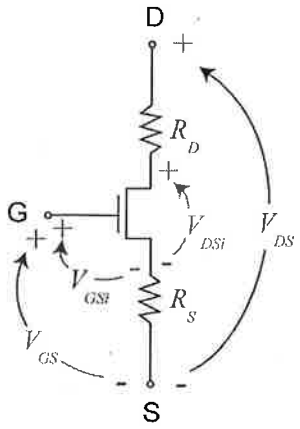


Figure 9.45: Schematic diagram showing parasitic source and drain resistance in MOSFET. The impact of the source and drain resistances is to reduce the internal voltages that are applied to the MOSFET.

In the saturation regime, the drain current of the MOSFET is set by the internal gate overdrive $V_{GSi} - V_T$. Then, using Eq. 9.21 we have:

$$I_{Dsat} = \frac{W}{2L}\mu_e C_{ox}(V_{GSi} - V_T)^2 = \frac{W}{2L}\mu_e C_{ox}(V_{GS} - I_D R_S - V_T)^2 \tag{9.99}$$

For a well designed transistor in which resistance effects are not too large, when we expand the quadratic term in this equation, it is reasonable to assume that $I_D R_S \ll V_{GS} - V_T$. In other words, the ohmic drop on the source resistance is much smaller than the gate overdrive. In this instance, we can easily solve for I_{Dsat}:

$$I_{Dsat} \simeq \frac{\frac{W}{2L}\mu_e C_{ox}(V_{GS} - V_T)^2}{1 + \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T)R_S} \tag{9.100}$$

This equation has the right limit as R_S goes to zero for which we recover the current-voltage characteristics of the ideal device. For non-negligible values of R_S, the term in the denominator reduces the drain current from its ideal value. This term becomes more significant the higher the gate overdrive. A consequence of this is that at high gate overdrive, the transfer characteristics of the transistor are no longer quadratic in $V_{GS} - V_T$ but they exhibit a weaker dependence. The absence of any role for R_D in I_{Dsat} is because the transistor is saturated and the drain current is insensitive to the value of V_{DS}.

In the linear regime, both R_S and R_D play a role. We can derive a first order expression for the linear drain current starting from the ideal equation given in 9.16 and then plugging in Eqs. 9.97 and 9.98. After assuming a well behaved transistor with moderate values of parasitic resistances we obtain:

$$I_D = \frac{W}{L}\mu_e C_{ox}(V_{GSi} - V_T - \frac{1}{2}V_{DSi})V_{DSi} \simeq \frac{\frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS}}{1 + \frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DS})(R_S + R_D)} \tag{9.101}$$

Once again, this equation converges to the ideal expression if $R_S + R_D = 0$. As the parasitic resistances become significant, the linear drain current drops below its ideal value. The drop becomes more prominent the harder we drive the transistor.

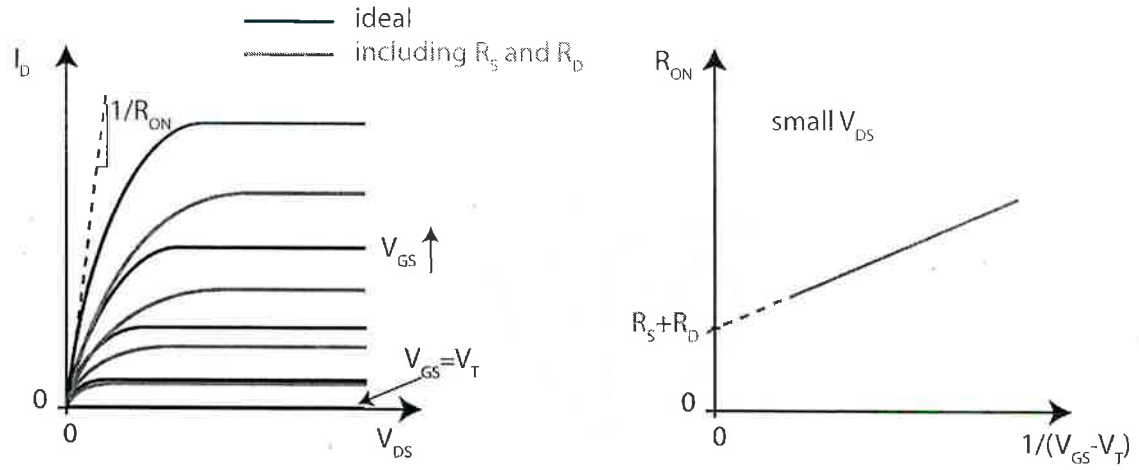


Figure 9.46: Left: Impact of R_S and R_D in the output characteristics of a MOSFET. The definition of R_{ON} is indicated. Right: Illustration of technique to extract $R_S + R_D$ in a MOSFET from measurements of R_{ON} as a function of gate overdrive in the linear regime.

The left of Fig. 9.46 illustrates the changes to the output characteristics of an ideal MOSFET as a result of the presence of source and drain resistance. For a given V_{GS} the drain current is reduced. The higher the gate overdrive, the bigger the effect. The value of V_{DS} that saturates the transistor is also increased. Finally, for very small V_{DS} the slope of the I-V characteristics is reduced.

The change in the linear characteristics of the transistor as a result of the source and drain resistance is of particular importance in certain applications such as power management. It is quantified through the so-called *ON resistance*, R_{ON}. This is defined as the ratio of V_{DS} to I_D in the limit of small V_{DS}. From Eq. 9.101, we can easily get:

$$R_{ON} = \frac{V_{DS}}{I_D}|_{small\ V_{DS}} \simeq R_S + R_D + \frac{1}{\frac{W}{L}\mu_e C_{ox}(V_{GS} - V_T)} \tag{9.102}$$

This result has a simple physical interpretation. The total resistance between the source and drain terminals of a MOSFET for small V_{DS} is given by the sum of the source and drain parasitic resistances plus the channel resistance. Paths to minimize R_{ON} include shrinking R_S and R_D and mitigating the channel resistance by selecting the geometry of the transistor (large W/L) and applying a large gate overdrive.

Eq. 9.102 also suggests a method for extracting $R_S + R_D$ in a MOSFET. If we plot R_{ON} as a function of $1/(V_{GS} - V_T)$, a straight line should result that intersects the y axis at $R_S + R_D$. This is illustrated on the right of Fig. 9.46.

An important consequence of the presence of source and drain resistance is the degradation of the transconductance of the device. It is clear from looking at Fig. 9.46 that g_m is reduced from the ideal value. In saturation, an expression for g_m that accounts for R_S (R_D has no impact) can be easily obtained by differentiating Eq. 9.97, dividing throughout by dI_D and then solving for g_m, to yield:

$$g_m = \frac{g_{mi}}{1 + g_{mi}R_S} \tag{9.103}$$

In this equation, g_{mi} is the *intrinsic transconductance*, or the transconductance of the ideal device:

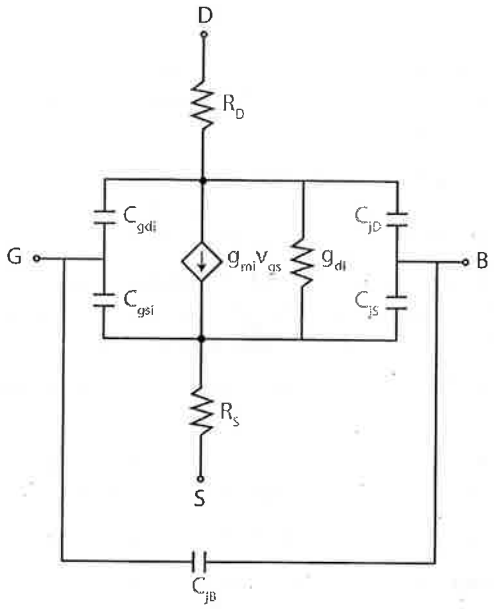


Figure 9.47: Small-signal equivalent circuit model of a MOSFET including R_S and R_D. The transconductance and output conductance generators are those of the ideal MOSFET.

$$g_{mi} = \frac{\partial I_D}{\partial V_{GSi}}|_{V_{DSi}} \tag{9.104}$$

and g_m is the *extrinsic transconductance*, defined as:

$$g_m = \frac{\partial I_D}{\partial V_{GS}}|_{V_{DS}} \tag{9.105}$$

Notice that in these derivatives holding constant V_{DS} in one case and V_{DSi} in the other is not an issue since adding R_S and R_D to the ideal MOSFET does not perturb its perfect saturation. [7]

It is clear from the result of Eq. 9.103 that the transconductance of a MOSFET is reduced as a result of adding source resistance and that the effect is the greatest the higher the intrinsic transconductance of the device.

R_S and R_D should be added to the small-signal equivalent circuit model of the MOSFET. When we do this, we need to be clear that the transconductance and output conductance elements in it represent the intrinsic values characteristic of the ideal MOSFET. A small-signal equivalent circuit model for a MOSFET that includes R_S and R_D is shown in Fig. 9.47. Notice that in this model, C_{jS} and C_{jD} have been placed on the *inside* of R_S and R_D. A better model would break R_S and R_D into a contact resistance component and a lateral component and place the junction capacitances at the common node.

[7] If g_d is included, one has to be more careful. See S. Y. Chou and D. A. Antoniadis, IEEE Trans. Electron Dev. ED-32, p. 448 (1987).

9.8 Summary

- The sheet-charge approximation states that the inversion layer in a MOSFET is very thin in the scale of the vertical dimensions of the device. It allows us to formulate the MOSFET current in terms of the sheet-charge density.
- The gradual-channel approximation breaks the the 2D electrostatic problem into two 1D simpler ones. The vertical electrostatics control the inversion layer charge and the lateral electrostatics control the flow of inversion charge from source to drain.
- In the linear regime, the drain current of an ideal MOSFET increases with V_{GS} and V_{DS}. In the saturation regime, the drain current depends only on V_{GS}. The dependence is quadratic. In the cut-off regime, the drain current through an ideal MOSFET is zero.
- The transconductance of an ideal MOSFET in saturation increases with the square root of the drain current.
- The "body effect" arises from an increase in the local V_T along the channel. It results in premature drain current saturation and reduces the current drive of the transistor.
- The application of a back bias to a MOSFET shifts V_T and affects the drain current.
- In a refined MOSFET model, drain current saturation and channel pinchoff occur when electrons reach velocity saturation at the drain end of the channel. For V_{DS} values higher than V_{DSsat}, the pinchoff point broadens into a region with finite length where the electron velocity is saturated. This effectively reduces the channel length and increases the drain current of the device. Channel length modulation, as this phenomenon is called, is responsible for the finite output conductance of a MOSFET in the saturation regime.
- For V_{GS} values below V_T, the drain current of a MOSFET drops exponentially. This is known as the subthreshold regime. The sharpness of the subthreshold regime is characterized by the subthreshold swing. This reflects a competition between the gate and the body of the transistor for electrostatic control over the surface potential. The sharpness of the subthreshold regime is enhanced the tighter the electrostatic influence of the gate of the transistor.

9.9 Further Reading

Operation and Modeling of the MOS Transistor (Third Edition) by Y. P. Tsividis and C. McAndrew, Oxford, 2011, ISBN 978-0-19-517015-3, TK7871.99.M44T77) is a classic textbook on MOSFET physics and modeling. The treatment of the MOSFET is rather comprehensive and in places goes beyond the treatment in the present book. Chapters relevant here are Ch. 4 that mostly deals with the long-channel MOSFET structure, Ch. 6 describing large-signal dynamic operation, and Ch. 7 and 8 addressing small-signal modeling, including noise. This is an important reference book.

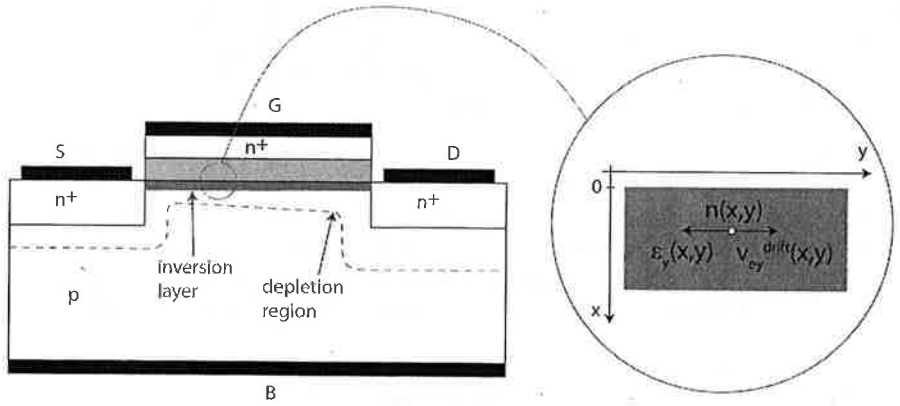


Figure 9.48: Left: sketch of a four-terminal MOS structure biased in the linear regime. Inside the circle: detail of the electrostatics of a small element of the inversion layer.

AT9.1 A more detailed study of inversion layer transport

In general, the inversion layer current has two terms, drift and diffusion. The application of a potential difference across the ends of the inversion layer imposes an electric field that results in a drift current. As we saw in the main body of this chapter, there is also a lateral gradient of electron concentration that is produced. This causes a diffusion current. In general, then, the electron current in a differential of volume inside the inversion layer (see Fig. 9.48) is given by:

$$J_{ey}(x,y) = -qn(x,y)v_{ey}^{drift}(x,y) + qD_e\frac{\partial n(x,y)}{\partial y} \tag{9.106}$$

Understanding the notation used in this equation is important. J_{ey} refers to the electron current along the y-direction. n is the electron concentration at a certain location in the inversion layer. v_{ey}^{drift} is the electron drift velocity along the y direction. All fo them, in general, depend on x and y.

To keep things simple, let us proceed under the assumption that the electric field is small enough so that there is a linear relationship between drift velocity and electric field. In this case, Eq. 9.106 can be written as:

$$J_{ey}(x,y) = qn(x,y)\mu_e\mathcal{E}_y(x,y) + qD_e\frac{\partial n(x,y)}{\partial y} \tag{9.107}$$

In this equation, $\mathcal{E}_y$ refers to the lateral electric field set up in the inversion layer, which in general, depends also on x and y.

We are rarely interested in the details of the electron current distribution vertically in the inversion layer, but on the total longitudinal electron current supported by the inversion layer. In consequence, we can eliminate unnecessary detail by integrating Eq. 9.107 in depth from the oxide-semiconductor interface to the end of the inversion layer:

$$J_e = \int_0^\infty J_{ey}(x,y)dx = q\mu_e\int_0^\infty n(x,y)\mathcal{E}_y(x,y)dx + qD_e\int_0^\infty \frac{\partial n(x,y)}{\partial y}dx \tag{9.108}$$

The integral starts at $x = 0$, the oxide-semiconductor interface and should extend across the entire inversion layer. Since the electrons are all piled up against the interface, we make an insignificant mistake by extending the integral all the way to infinity. Also, since there are no vertical currents in the structure, we have dropped the y subindex in the current expression.

Three comments are proper here. First, the units of J_e in this equation are A/cm or current per unit width of the MOS structure. The *wider* the structure (into the page in Fig. 9.48), the more total current it supports. In contrast, J_{ey} in Eqs. 9.106 and 9.107 above is in A/cm^2. Second, since there are no sources or sinks of electrons inside the inversion layer, J_e does not depend on y. And third, an assumption that has been made also here is that μ_e and D_e are independent of x. We can also view them as suitable average values for the entire electron distribution in depth. μ_e and D_e satisfy Einstein's relation.

AT9.1.1 The sheet-charge approximation

A significant simplification of this formulation is possible if the inversion layer can be considered very thin. As in Ch. 8, we call this the sheet charge approximation. For a thin inversion layer, $\mathcal{E}_y$ in the drift term of Eq. 9.108 does not depend very much on x. We can then drop its x coordinate dependence and bring it out of the integral as $\mathcal{E}_y(y)$. The limits of validity of this assumption are explored later on in this section. Additionally, the term in $\partial/\partial y$ in the diffusion term can be brought out of the second integral since its limits are independent of y (Leibniz rule). All this results in:

$$J_e = q\mu_e n_s(y)\mathcal{E}_y(y) + qD_e\frac{dn_s(y)}{dy} \tag{9.109}$$

where n_s is the *sheet electron concentration* of the inversion layer (in units of cm^{-2}), defined in Eq. 9.1. In general, n_s depends on the lateral coordinate y.

A completely equivalent description can be given in terms of the inversion layer sheet charge density, Q_i, given by Eq. 9.3:

$$J_e = -\mu_e Q_i(y)\mathcal{E}_y(y) - D_e\frac{dQ_i(y)}{dy} \tag{9.110}$$

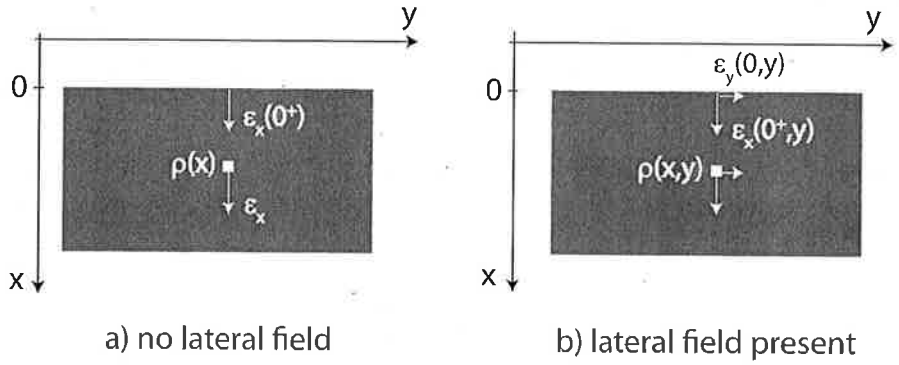
The sheet-charge approximation has allowed us to express the inversion layer current in terms of the inversion layer charge and the lateral electric field. The next step is to find a way to determine $Q_i(y)$.

AT9.1.2 The gradual-channel approximation

Before attempting to find an expression for $Q_i(y)$, it is useful to review the electrostatics of a MOS structure in the case in which no lateral field is applied to the inversion layer. This was studied in great detail in Ch. 8. As sketched on the left of Fig. 9.49, this is strictly a one-dimensional electrostatic problem along the x direction. Gauss' law for this problem can be written as:

$$\frac{d\mathcal{E}_x}{dx} = \frac{\rho(x)}{\epsilon_s} \tag{9.111}$$

where $\mathcal{E}_x$ is the vertical field in the semiconductor. In this situation, $\mathcal{E}_x$ is independent of y. From here, we derived in Ch. 8 a simple relationship between Q_i and the gate voltage V_G that was valid for strong



a) no lateral field b) lateral field present

Figure 9.49: Electrostatics of the inversion layer. Left: In the absence of a lateral field, the electrostatics problem is simply one-dimensional. Right: With a lateral field applied, the electrostatics problem becomes two-dimensional.

inversion and that was given by $Q_i = -C_{ox}(V_G - V_T)$. Note that this equation uses the inversion layer and the substrate as voltage reference ($V_S = V_B = 0$).

With a lateral field applied across the inversion layer, the situation gets more complicated. The electrostatic problem is now of a two-dimensional nature. Gauss' law for this situation is:

$$\frac{\partial \mathcal{E}_x}{\partial x} + \frac{\partial \mathcal{E}_y}{\partial y} = \frac{\rho(x,y)}{\epsilon_s} \tag{9.112}$$

where, in general, $\mathcal{E}_x$ and $\mathcal{E}_y$ depend both on x and y.

In strong inversion and when current flows, $\mathcal{E}_y \neq 0$. In general, then, $\frac{\partial \mathcal{E}_y}{\partial y}$ does not vanish. The two-dimensional problem embodied in Eq. 9.112 is substantially harder than the one-dimensional problem solved in Ch. 8.

A simple approximate solution can be obtained under what is called the *gradual-channel approximation*. If the change of $\mathcal{E}_y$ along the y direction is much slower than the change of $\mathcal{E}_x$ along the x direction, that is, $\frac{\partial \mathcal{E}_x}{\partial x} \gg \frac{\partial \mathcal{E}_y}{\partial y}$, then Eq. 9.112 can be simplified to:

$$\frac{\partial \mathcal{E}_x}{\partial x} \simeq \frac{\rho(x,y)}{\epsilon_s} \tag{9.113}$$

This equation has a similar solution to that of Eq. 9.111. Under strong inversion, the inversion-layer charge at location y, is determined by the voltage overdrive applied to the gate with respect to the inversion layer at the same location, that is,

$$Q_i(y) \simeq -C_{ox}[V_G - V(y) - V_T] \tag{9.114}$$

Q_i now depends on lateral location y through the local inversion layer voltage $V(y)$ that was defined in Eq. 9.7.

The gradual channel approximation allows the break up of the two-dimensional electrostatics problem into two simpler one-dimensional problems: the vertical electrostatics control the charge in the inversion layer, while the lateral electrostatics control the lateral flow of inversion layer charge. In the gradual-channel

approximation, the lateral electrostatics represent a small perturbation of the vertical electrostatics. As a consequence, the inversion layer charge at any one location can be computed using the one-dimensional MOS theory developed in the previous chapter, with the simple precaution of using the *local* inversion layer voltage. At the end of this section, we will discuss the constraints on the applicability of the gradual-channel approximation.

We can now proceed with the derivation of a first-order model for the current. Taking the derivative of $Q_i(y)$ with respect to y in Eq. 9.114 while assuming that V_T does not change in space (no body effect), we can write:

$$\frac{dQ_i(y)}{dy} \simeq C_{ox}\frac{dV(y)}{dy} \tag{9.115}$$

We can plug this into Eq. 9.110, apply Einstein's relation and also use Eq. 9.8 to get:

$$J_e = \mu_e[Q_i(y) - \frac{kT}{q}C_{ox}]\frac{dV(y)}{dy} \tag{9.116}$$

Notice that since Q_i is negative, diffusion actually helps drift in electron transport along the inversion layer. This makes sense since we know that n_s drops along the channel from source to drain.

We can now discuss the relative magnitude of the drift and diffusion terms in Eq. 9.116. Since Q_i is given by Eq. 9.114, it is clear that drift prevails over diffusion if the amount of gate overdrive is much larger than the thermal voltage, that is, if $V_G - V(y) - V_T \gg \frac{kT}{q}$. This occurs when the MOS structure is in strong inversion. In this case, the expression of the inversion current simplifies to:

$$J_e \simeq \mu_e Q_i(y)\frac{dV(y)}{dy} \tag{9.117}$$

This is the equation that we used to derive the I-V characteristics of the MOSFET in the linear and saturation regimes in the main body of this chapter.

Close to threshold, diffusion becomes important. In fact, below threshold, as we will see, current conduction is dominated by electron diffusion. This is the so-called *subthreshold current* (beware, in this regime, Eq. 9.114 does not apply since it was derived for strong inversion; as a consequence, Eq. 9.116 does not apply either).

AT9.1.3 Validity of approximations

We can now discuss the limits of the gradual-channel approximation. As stated above, the gradual channel approximation is valid when:

$$|\frac{\partial \mathcal{E}_x}{\partial x}| \gg |\frac{\partial \mathcal{E}_y}{\partial y}| \tag{9.118}$$

Let us evaluate this inequality at $x = 0^+$, the semiconductor side of the oxide/semiconductor interface. The first term can be estimated by taking the difference in electric field at the surface and at the bottom of the inversion layer:

$$\frac{\partial \mathcal{E}_x}{\partial x}|_{x=0^+} \simeq \frac{\mathcal{E}_x(x=x_{inv}) - \mathcal{E}_x(x=0^+)}{x_{inv}} = \frac{Q_i}{\epsilon_s x_{inv}} \tag{9.119}$$

The second term can be estimated from Eq. 9.8:

$$\frac{\partial \mathcal{E}_y}{\partial y}|_{x=0^+} = -\frac{d^2V}{dy^2} \tag{9.120}$$

The second derivative of V can be obtained from Eq. 9.116 by taking the derivative with respect to y and realizing that $dJ_e/dy = 0$. Hence:

$$\frac{d^2V}{dy^2} \simeq -\frac{1}{Q_i}\frac{dQ_i}{dy}\frac{dV}{dy} = -\frac{C_{ox}}{Q_i}(\frac{dV}{dy})^2 = -\frac{C_{ox}}{Q_i}[\mathcal{E}_y(y)]^2 \tag{9.121}$$

where Eqs. 9.115 and 9.8 have been used (the term on $\frac{kT}{q}C_{ox}$ in Eq. 9.116 was neglected, see justification below).

Combining Eqs. 9.118, 9.119, 9.120, and 9.121, the condition for the gradual channel approximation can be rewritten as:

$$|\mathcal{E}_y(y)| \ll \frac{|Q_i|}{\sqrt{\epsilon_s x_{inv} C_{ox}}} \tag{9.122}$$

And using the expression of Q_i given by Eq. 9.114, it can also be expressed as:

$$|\mathcal{E}_y(y)| \ll \frac{V_G - V(y) - V_T}{\sqrt{\frac{\epsilon_s}{\epsilon_{ox}} x_{inv} x_{ox}}} \tag{9.123}$$

In essence, for the gradual-channel approximation to be fulfilled, the lateral electric field must be much smaller than an effective vertical field given by the voltage overdrive on the gate divided by an effective vertical characteristic length (the denominator in Eq. 9.123). This vertical characteristic length is a function of the oxide thickness, the inversion layer thickness and the permittivities of both materials. The smaller this characteristic length, the easier it is for the gradual channel approximation to be fulfilled.

Let us put some typical numbers to examine the validity of the gradual-channel approximation. For a MOSFET structure with an oxide thickness of $x_{ox} = 4.5\ nm$ and a p-type substrate of $N_A = 6\times10^{17}\ cm^{-3}$, $x_{inv} \simeq 7.5\ nm$, and the vertical characteristic length turns out to be about 10 nm. For a location in which $V_G - V(y) - V_T = 1\ V$, the lateral field must be much smaller than $10^6\ V/cm$, which is an extraordinarily high value. Hence, except right around threshold when $V_G - V(y) - V_T \simeq 0$, the gradual-channel approximation is indeed an excellent one.

We now turn our attention to verifying the sheet-charge approximation that was made at the beginning of this section when $\mathcal{E}_y(x,y)$ was taken out of the integral in Eq. 9.108. This can be done with minimum error if $n(x,y)$ changes with x much faster than $\mathcal{E}_y(x,y)$. Mathematically, this condition can be expressed as:

$$|\frac{1}{\mathcal{E}_y}\frac{\partial \mathcal{E}_y}{\partial x}| \ll |\frac{1}{n}\frac{\partial n}{\partial x}| \tag{9.124}$$

Let us examine this inequality at $x = 0^+$. It is handy to rewrite the left-hand side in the following way. Since $\vec{\nabla} \times \vec{\mathcal{E}} = 0$, it follows that:

$$\frac{\partial \mathcal{E}_y}{\partial x} = \frac{\partial \mathcal{E}_x}{\partial y} \tag{9.125}$$

inside the inversion layer.

In the context of the gradual-channel approximation, at $x = 0^+$:

$$\mathcal{E}_x|_{x=0^+} \simeq -\frac{Q_i + Q_d}{\epsilon_s} \tag{9.126}$$

Hence,

$$\frac{\partial \mathcal{E}_x}{\partial y}|_{x=0^+} \simeq -\frac{1}{\epsilon_s}\frac{dQ_i}{dy} = \frac{C_{ox}}{\epsilon_s}\mathcal{E}_y(y) \tag{9.127}$$

where we have used Eq. 9.115 and 9.8.

Combining 9.127 and 9.125, we can write:

$$|\frac{1}{\mathcal{E}_y}\frac{\partial \mathcal{E}_y}{\partial x}|_{x=0^+} = |\frac{1}{\mathcal{E}_y}\frac{\partial \mathcal{E}_x}{\partial y}|_{x=0^+} \simeq \frac{C_{ox}}{\epsilon_s} \tag{9.128}$$

Now we work on the right hand side of Eq. 9.124. From Ch. 8 we know that:

$$n(x) = N_A \exp\frac{q[\phi(x) - 2\phi_f]}{kT} \tag{9.129}$$

Therefore,

$$\frac{1}{n}\frac{\partial n}{\partial x} = \frac{q}{kT}\frac{\partial \phi}{\partial x} = -\frac{q}{kT}\mathcal{E}_x \tag{9.130}$$

$\mathcal{E}_x$ is given by Eq. 9.126. Hence:

$$|\frac{1}{n}\frac{\partial n}{\partial x}|_{x=0^+} \simeq |\frac{q}{kT}\frac{Q_i + Q_d}{\epsilon_s}| \tag{9.131}$$

Plugging in Eqs. 9.128 and 9.131, the condition 9.124 at $x = 0^+$ can be written as:

$$\frac{C_{ox}}{\epsilon_s} \ll |\frac{q}{kT}\frac{Q_i + Q_d}{\epsilon_s}| \tag{9.132}$$

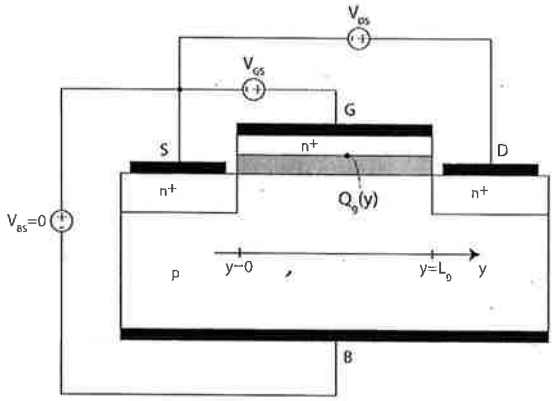
Since Q_i and Q_d have the same sign, this condition is satisfied if:

$$|Q_i| \gg \frac{kT}{q}C_{ox} \tag{9.133}$$

This is exactly the mathematical expression derived above for the gradual channel approximation. Hence, under conditions in which the gradual approximation applies, $\mathcal{E}_y$ changes sufficiently slowly inside the inversion layer that it can safely be taken out of the integral in Eq. 9.108 and the sheet-charge approximation also holds.

Problems

9.1 Consider a long n-channel MOSFET biased in the linear regime. This problem is about answering a simple question: does the sheet charge density imaged on the *gate*, $Q_g(y)$, increase, decrease, or remains constant as one goes from source to drain ($y = 0$ to $y = L_g$)?



You are to answer this question in a quantitative way by calculating the sheet charge density imaged on the gate of a long MOSFET characterized by the following parameters: n^+-polysilicon gate ($W_M = \chi_S = 4.04\ eV$), $x_{ox} = 15\ nm$, uniform $N_A = 10^{17}\ cm^{-3}$, at a bias given by $V_{GS} = 2.5\ V$, $V_{DS} = 1\ V$, and $V_{BS} = 0$.

To solve this problem, you need to consider the impact of the *body effect*. Otherwise, consider this MOSFET to be ideal.

To save you time, for this structure:

$$C_{ox} = 2.3 \times 10^{-7}\ F/cm^2,\ \gamma = 0.8\ V^{1/2},\ \phi_{sT} = 0.84\ V,\ V_{FB} = -1\ V,\ V_{To} = 0.56\ V$$

a) Compute the sheet charge density of electrons in the inversion layer at the *source* end of the channel, $Q_i(y = 0)$.

b) Compute the charge in the depletion region under the channel at the *source* end of the channel, $Q_d(y = 0)$.

c) Compute the sheet charge density of electrons in the inversion layer at the *drain* end of the channel, $Q_i(y = L_g)$.

d) Compute the charge in the depletion region under the channel at the *drain* end of the channel, $Q_d(y = L_g)$.

e) Compute the sheet charge density imaged at the gate/oxide interface at the source and drain ends of the channel, $Q_g(y = 0)$ and $Q_g(y = L_g)$. Which one is highest?

9.2 Consider an long-channel n-MOSFET in the linear regime with $V_{BS} = 0$. Neglecting the body effect, derive analytical expressions for $V(y)$, $\mathcal{E}_{ox}(y)$, $\mathcal{E}_y(x = 0, y)$, $v_e(y)$, and $Q_i(y)$ from source to drain.

9.3 Derive an analytical expression for the condition that gives how close V_{DS} can get to $V_{GS} - V_T$ in a MOSFET biased in the linear regime before the gradual-channel approximation fails. Express your result in terms of $\Delta V_{DS} = V_{GS} - V_T - V_{DS}$. Do not include the body effect. $V_{SB} = 0$. You will need to solve Prob. 9.2 first. Compute ΔV_{DS} for a MOSFET characterized by $L = 0.1\ \mu m$, $x_{ox} = 4.5\ nm$, $N_A = 6 \times 10^{17}\ cm^{-3}$. What fraction of the maximum current predicted by Eq. 9.16 is attained at this maximum tolerable value of V_{DS}?

9.4 Consider an n-channel MOSFET in the saturation regime.

a) Derive expressions and sketch the spatial dependence of $V(y)$, $\mathcal{E}_{ox}(y)$, $\mathcal{E}_y(y)$, $v_e(y)$, and $Q_i(y)$ from source to drain. Explain your results.

b) Derive also an expression for the electron velocity due to diffusion. Compare the diffusion velocity with the drift velocity obtained in part a). Under what conditions is electron diffusion safely ignored?

c) Obtain an expression for the transit time of electrons through the channel from source to drain by computing:

$$\tau_t = \int_0^L dt = \int_0^L \frac{dy}{v_e(y)}$$

d) Compute the transit time again using the following expression:

$$\tau_t = \frac{|Q_I|}{I_D} = \frac{W_g |\int_0^L Q_i(y)dy|}{I_D}$$

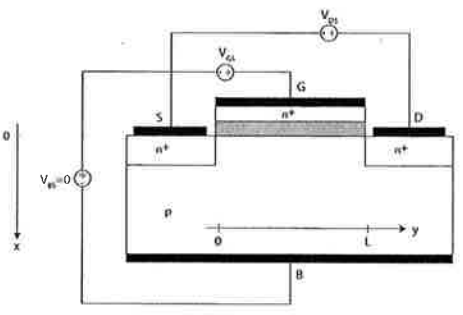
Compare this result with the one obtained on part c).

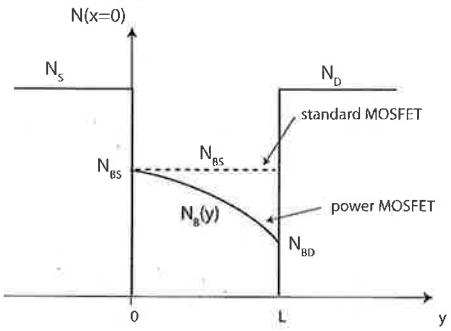
For all parts, neglect the body effect, that is, assume that the current of the transistor is given by Eq. 9.21.

9.5 In power MOSFETs, the body doping is often non-uniform. This problem examines some of the consequences of this.

Consider a power MOSFET as sketched in the diagram below. In this device, the doping level in the body drops along the channel from source to drain, as sketched in the diagram on the right. This diagram shows the doping distribution right along the surface of the device. N_S and N_D refer to the n-type doping level of the source and drain. N_A is the p-type doping level in the body which changes underneath the channel along the y dimension, but not along the x dimension. N_{BS} refers to the body doping at the source-end of the channel ($y = 0$). N_{BD} refers to the body doping at the drain-end of the channel ($y = L$).

In this problem, we will compare the operation of the power MOSFET with that of a regular MOSFET with a uniform doping level in the body equal to N_{BS}, as indicated by the dashed line in the diagram. In this way, these two devices have an identical threshold voltage.





This is a long-channel MOSFET problem. Other than a non-uniform body doping, consider this device, as well as the reference standard MOSFET, as an ideal long-channel MOSFETs as defined in class.

a) Due to the fact that the body doping changes along the channel, the *local* threshold voltage of the inversion layer is also a function of position. How does $V_T(y)$ evolve along the channel? Sketch a diagram of $V_T(y)$ vs. y for the power MOSFET and the regular MOSFET. Explain your graph.

b) Now consider these two devices biased in the linear regime with an identical $V_{GS} > V_T$ and an identical and small V_{DS}. Sketch the absolute of the inversion layer charge $|Q_i(y)|$ along the channel from source to drain for both devices. Explain your graph.

c) Now consider these devices biased in saturation at an identical $V_{GS} > V_T$ and an identical and large $V_{DS} > V_{DSsat}$. Sketch a diagram of the voltage along the channel $V(y)$ vs. y for the power MOSFET and the regular MOSFET. Explain your graph.

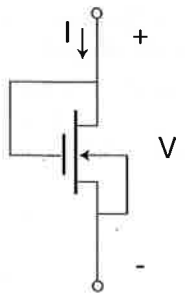
d) Under the bias condition of part c) above, which device has the highest drain current? Why?

e) For an identical value of $V_{GS} > V_T$, sketch a diagram of I_D vs. V_{DS} for the power MOSFET and the regular MOSFET from $V_{DS} = 0$ to $V_{DS} > V_{DSsat}$. Explain your graph.

f) Under the bias condition of part c) above, which device has the highest transconductance? Why?

9.6 You have been given the specs of a foundry's digital CMOS technology that you are considering for one of your company's designs. Some of the values that characterize the n-MOSFET at room temperature are: $I_{off} = 1\ nA/\mu m$, $V_T = 0.5\ V$, and $S = 70\ mV/dec$. For your design, it is crucial that you know I_{off} at 85 °C, which is not given in the spec sheet. You consult some books and you find that V_T changes typically about $-4\ mV/^oC$. Also the inversion layer mobility goes approximately as T^{-2}. With all this, *estimate* I_{off} at 85 °C. State any assumptions that you need to make.

9.7 Analog designers often need to shift DC bias levels in circuits. This is best accomplished using p-n diodes. In standard CMOS, floating p-n diodes (that is, diodes not in direct contact with the power rails) are not readily available. A common way to "synthesize" a diode is to tie up an enhancement-mode MOSFET ($V_T > 0$) with the gate and drain shorted as sketched below. For this, a long MOSFET is typically used.



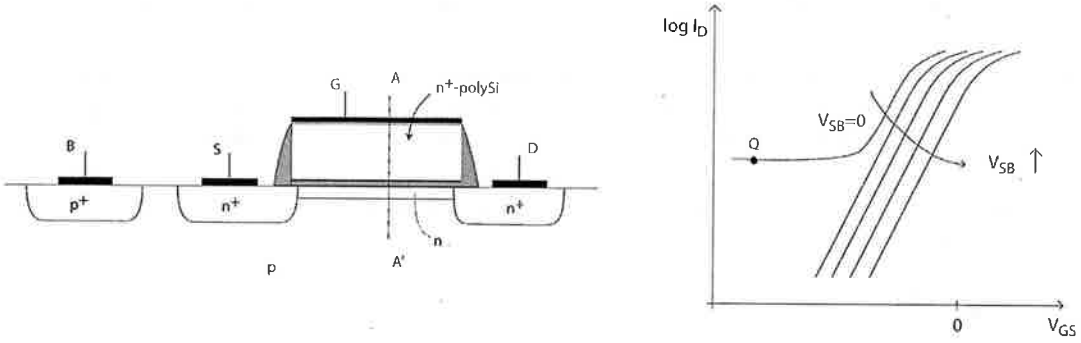
In responding to the questions below, use "long-MOSFET" theory without body effect nor back-bias effect.

a) Derive suitable equations for the I-V characteristics of the "diode", I *vs.* V for $V < V_T$, and $V > V_T$. Sketch the I-V characteristics in a linear scale.

b) Sketch a high-frequency small-signal equivalent circuit model for this diode for situations in which $V > V_T$. Derive expressions for all important small-signal elements as a function of bias.

c) Derive and sketch the I-V characteristics of the "diode" if implemented with a depletion-mode device ($V_T < 0$).

9.8 A depletion-mode MOSFET is an FET that is ON at $V_{GS} = 0\ V$. For an n-MOSFET, this is usually accomplished by purposely introducing an n-type doped region underneath the gate, as sketched below. This is called a *buried-channel MOSFET.*

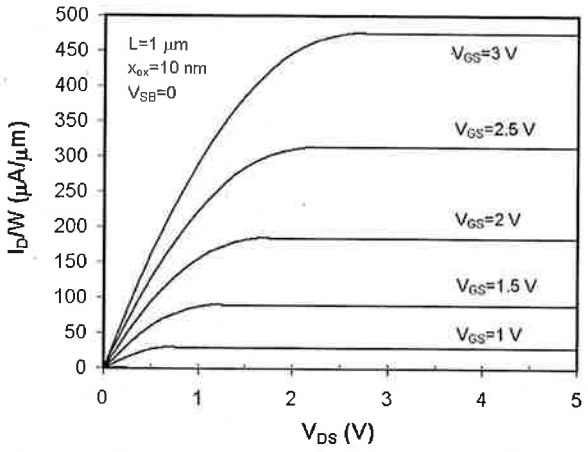


a) Sketch an energy band diagram for $V_{GS} = V_{DS} = V_{BS} = 0$ through the cross-section A-A' shown above. Indicate presence of depletion, accumulation or inversion regions.

b) Sketch an energy band diagram with the MOS structure *at flat-band* across the A-A' section. Provide an expression for the flat-band voltage V_{FB} in terms of appropriate parameters. V_{FB} should be referred to V_{GS}.

c) The subthreshold characteristics of a poorly designed buried-channel n-MOSFET are also sketched above as a function of the back bias. It is found that for V_{SB} sufficiently positive, the transistor turns off nicely. However, for $V_{SB} = 0$, the transistor fails to turn off completely. What is going on? Sketch the energy band diagram across the A-A' section at bias point **Q** indicated in the sketch above.

9.9 Consider the output I-V characteristics of a long-channel n-MOSFET for $V_{SB} = 0$ below. The device has a gate length $L_g = 1\ \mu m$ and a gate oxide thickness $x_{ox} = 10\ nm$. The output characteristics have been normalized for a unity width device.



a) Assuming that this device is an ideal long-channel MOSFET, estimate the threshold voltage of the device.

b) Assuming that this device is an ideal long-channel MOSFET, estimate the mobility of the electron inversion layer.

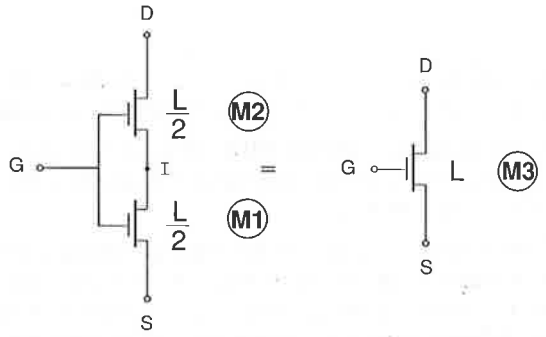
c) Assuming that this device is an ideal long-channel MOSFET, estimate the transconductance g_m at a bias of $V_{GS} = 3\ V$, and $V_{DS} = 3\ V$ for $W_g = 10\ \mu m$.

d) At a bias of $V_{GS} = 3\ V$, and $V_{DS} = 3\ V$, estimate the gate-source capacitance C_{gs} of a $W_g = 10\ \mu m$ device.

e) Now you are given the additional information that the subthreshold swing of this transistor technology at room temperature is $S = 78\ mV/dec$. At a bias of $V_{GS} = 3\ V$, and $V_{DS} = 3\ V$, estimate the back-gate transconductance g_{mb} of a $W_g = 10\ \mu m$ device.

f) At a bias of $V_{GS} = 3\ V$, and $V_{DS} = 3\ V$, how much less current do you expect to have as a result of the body effect? (Considering the additional subthreshold current information, if needed).

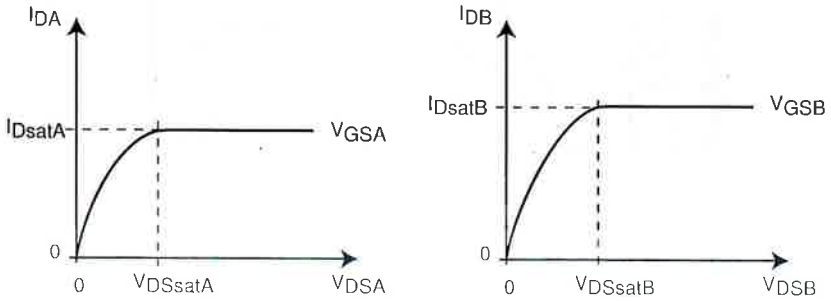
9.10 One of my colleagues asserts that:



configuration A configuration B

This basically means that a configuration of two MOSFETs in series with a common gate voltage is equivalent to a single MOSFET with a channel length that is twice as long, all other aspects of the device unchanged. This problem is about evaluating this assertion for an ideal MOSFET (no body effect, no back bias, no channel length modulation, no short-channel effects).

Consider the following sketches of the I-V characteristics for these two configurations for the same value of $V_{GS} > V_T$:



configuration A configuration B

Label the internal node in configuration A as "I", that is, refer to its voltage with respect to the source as V_{IS}.

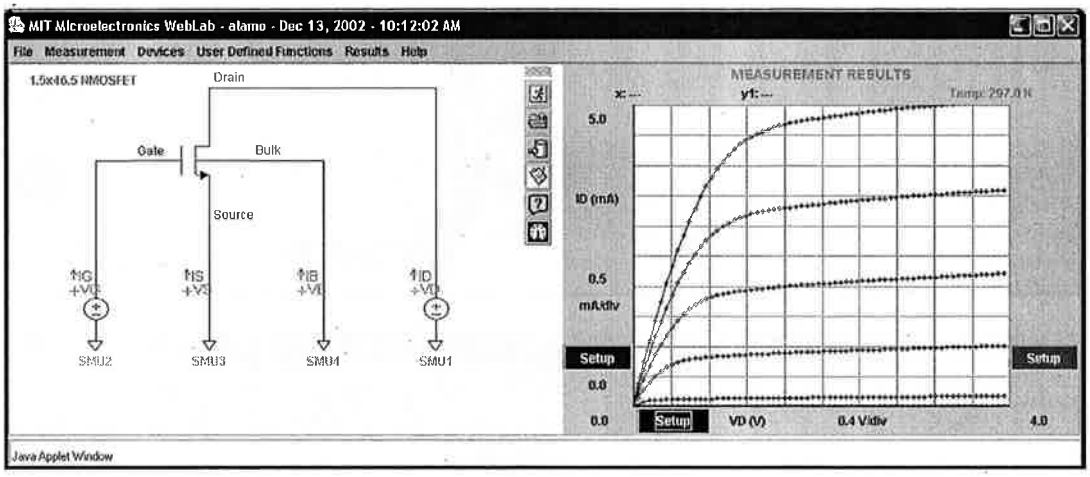
a) For $V_{DS} > V_{DSsat}$, in what regime is each of the transistors in configuration A biased?

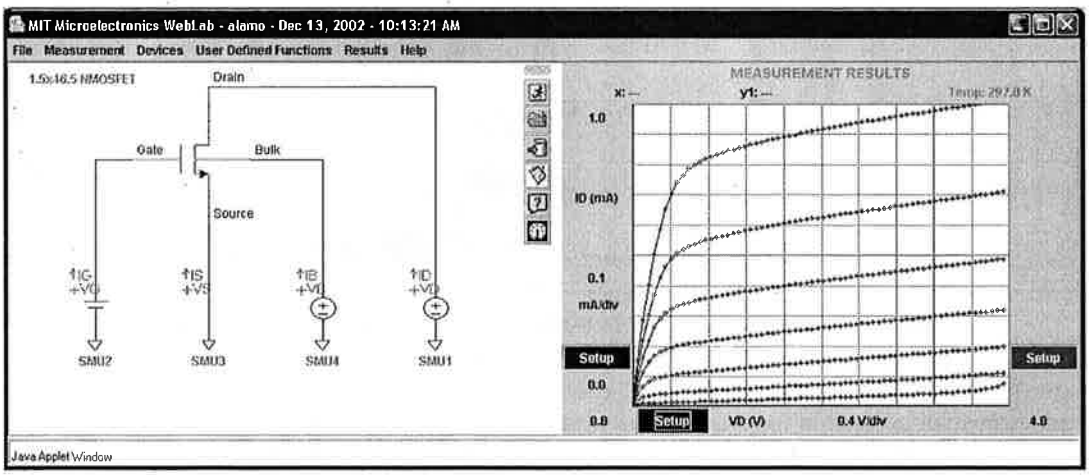
b) For configuration A and for $V_{DS} > V_{DSsat}$, derive equations for the drain current through each transistor as a function of V_{GS}, V_{DS}, V_T, and V_{IS}. From this, derive an expression for V_{IS}.

c) For configuration A, derive an expression for I_{Dsat} as a function of V_{GS}, V_{DS}, V_T and structural parameters. How does this compare with I_{Dsat} of configuration B for the same values of V_{GS} and V_{DS}?

d) For configuration A, derive an expression for V_{DSsat}. How does it compare with that of configuration B for the same V_{GS}?

9.11 Below are two screen shots of weblab measurements taken on a 1.5x46.5 μm NMOSFET. The first one shows the output characteristics obtained for $V_{SB} = 0\ V$, and V_{GS} from 0 to 3 V, in steps of 0.5 V. The second one shows the output characteristics obtained with the roles of the gate and body reversed. In these, $V_{GS} = 1.5\ V$ and V_{SB} is stepped from 0 V to 3 V, in steps of 0.5 V.



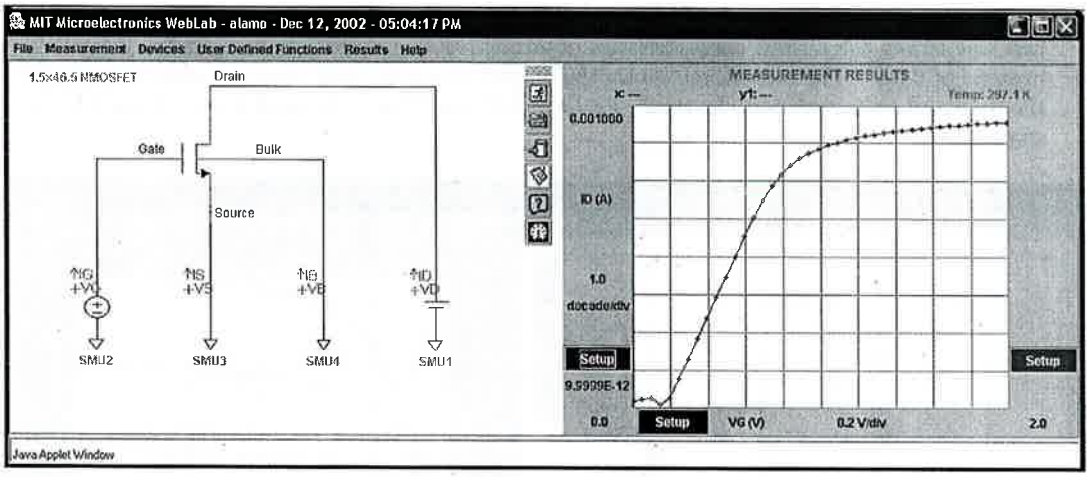


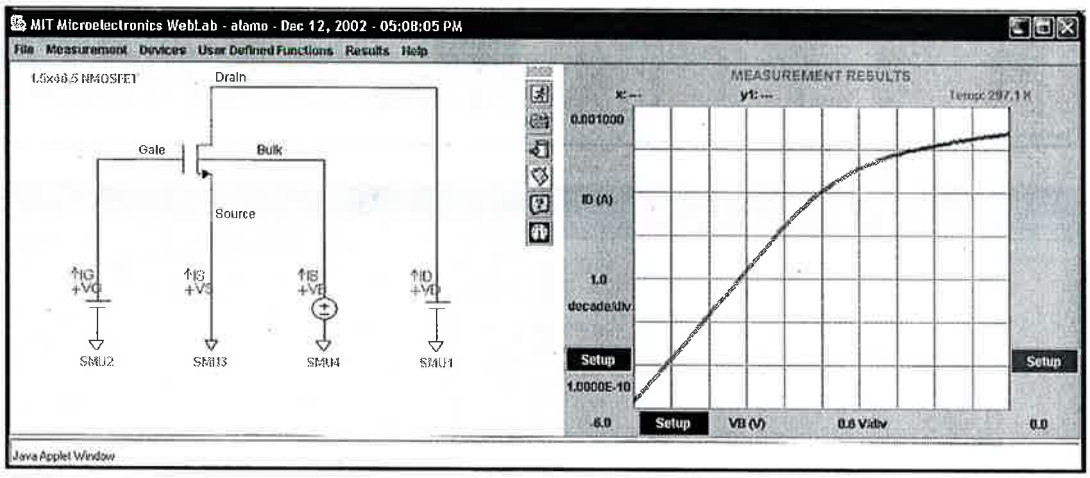
As we discussed in class, the body of the MOSFET behaves as a gate and reasonably looking output characteristics are obtained.

a) One of the most striking differences between the two sets of output characteristics is V_{DSsat}. The characteristics obtained with the body operating as the main gate seem to display remarkably smaller values of V_{DSsat} than in the normal mode. This question is about understanding the origin of this.

Derive a simple expression for V_{DSsat} as a function of V_{SB} for an ideal MOSFET. Do not include the "body effect" (but obviously include the "back bias" effect). Is this expression consistent with the data above? Explain.

Below are the subthreshold characteristics obtained for this MOSFET for both modes of operation. In the normal mode (top), $V_{SB} = 0$. In the mode in which the body is used as gate (bottom), $V_{GS} = 1.5\ V$. In both cases, $V_{DS} = 0.1\ V$.





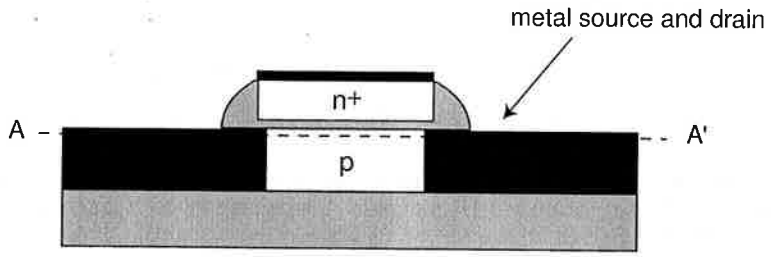
A remarkable difference is also observed here: the subthreshold swing of the device in the normal mode is significantly sharper than in the mode in which the body is used as the main gate.

b) Derive an expression for the relationship between the subthreshold swing of a MOSFET in the normal mode and in the gate-reversed mode. Explain your result.

c) From the experimental characteristics of the previous page, extract the subthreshold swing of this device for both modes of operation. Compare their relationship with that pre-

dicted with the equation derived above. Discuss.

9.12 A Schottky barrier source and drain MOSFET has been under study as an approach to reduce source and drain resistance and eliminate latch up. A schematic of this device is shown below. This problem is about drawing energy band diagrams in order to understand its basic operation.



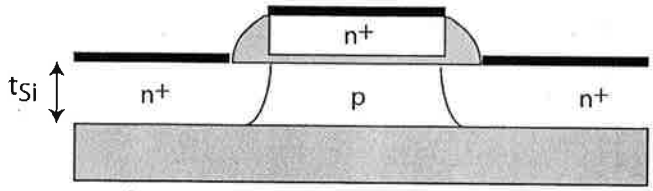
Consider a device with a threshold voltage $V_T = 0.5\ V$, body doping level $N_A = 10^{17}\ cm^{-3}$, and source and drain Schottky barrier height $q\varphi_{Bp} = 0.5\ eV$.

a) Sketch an energy band diagram right along the surface of the device (cross section A-A' above) for $V_{GS} = 1\ V$ and $V_{DS} = 0$. Give as many values of key energy features as you can.

b) Sketch an energy band diagram right along the surface of the device (cross section A-A' above) for $V_{GS} = V_{DS} = 1\ V$. Give as many values of key energy features as you can.

c) How does this device operate? Can you see any drawbacks?

9.13 The figure depicts a fully-depleted Silicon-On-Insulator (SOI) n-channel MOSFET. This is a transistor in which the body is fully depleted in the regular region of operation. For this transistor, $L = 1\ \mu m$, $W = 5\ \mu m$, $x_{ox} = 20\ nm$, $N_A = 10^{17}\ cm^{-3}$. The Si thickness is $t_{Si} = 10\ nm$.



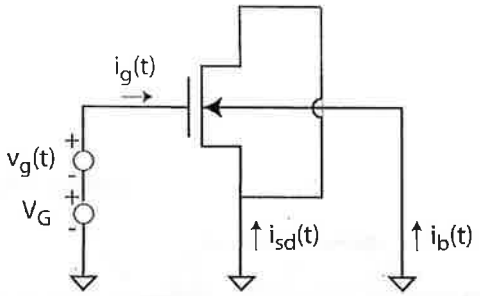
a) With $V_{DS} = 0\ V$, calculate the value of V_{GS} that fully depletes the body of the transistor. Sketch the energy band diagram vertically through the middle of the device at this bias point.

b) Calculate the threshold voltage of this device. Sketch the energy band diagram vertically through the middle of the device at this bias point.

c) Estimate the subthreshold swing of the device at room temperature.

9.14 The "split C-V technique" is widely used to characterize MOSFETs. This problem is about understanding some fundamental aspects of this technique. The configuration for the split

C-V technique is shown in the figure below:

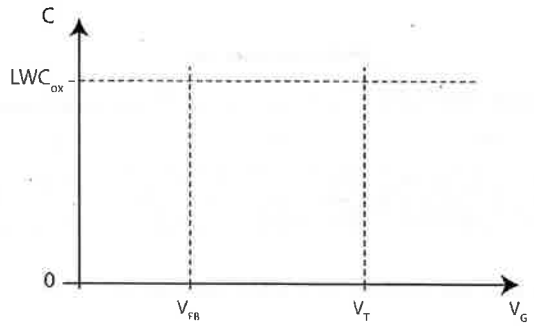


In essence, the source and drain are shorted together and grounded. The body is also grounded. The gate bias is set to V_G through a DC source. A small-signal AC signal v_g is applied on top of V_G. The current through all the terminals is measured as indicated in the figure. From these currents, the gate-to-channel capacitance, C_{gc}, and the gate to body capacitance C_{gb} are obtained. These capacitances are defined as:

$$C_{gc} = -\frac{i_{sd}}{\frac{dv_g}{dt}} \qquad C_{gb} = -\frac{i_b}{\frac{dv_g}{dt}}$$

The physical dimensions of the gate of this transistor are L and W.

a) In the axes shown below, sketch the evolution of C_{gc} with V_G across the entire V_G range for a low frequency AC signal. Explain the key features of your drawing.

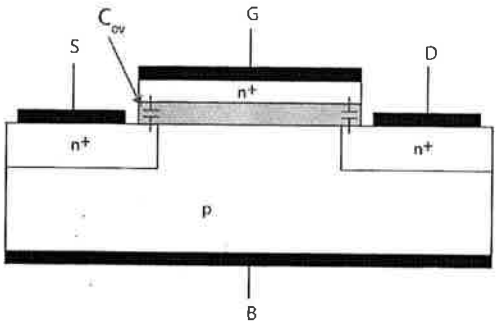


b) In the same axes as in part a), sketch the evolution of C_{gb} with V_G across the entire V_G range for a low frequency AC signal. Explain the key features of your drawing.

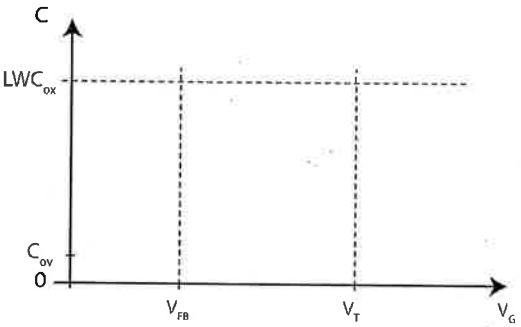
c) In the same axes as in part a), sketch the evolution of C_{gc} and C_{gb} with V_G across the entire V_G range for a *high frequency* AC signal.

d) Now consider the impact of the overlap capacitances between the gate and the source and

drain extensions of the transistor, as sketched below.



In the axes above on the right, sketch again the evolution of C_{gc} and C_{gb} with V_G at low frequency now accounting for the presence of C_{ov}. The relative value of C_{ov} is indicated in the figure below.



e) The split CV technique when combined with DC measurements of the drain current of the MOSFET in the linear regime can be used to experimentally obtain the mobility dependence on $\mathcal{E}_{eff}$. Explain how this would work. Do this ignoring C_{ov}.