

Supplementary Reader IV

EECS 40

***Introduction to
Microelectronic Circuits***

Prof. C. Chang-Hasnain

Fall 2006

Table of Contents

Chapter 4.	Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET).....	1
4.1	Introduction	1
4.2	Notation.....	1
4.3	NMOS and PMOS Transistors	2
4.4	N-MOSFET Operating Regions	2
4.4.1	Cut-off	2
4.4.2	Triode	3
4.4.3	Saturation.....	3
4.5	PMOSFET Operating Regions	4
Chapter 5.	Simple MOSFET Circuits.....	5
5.1	Analysis for MOSFET Amplifiers	5
5.1.1	DC Analysis – Load-Line Analysis	5
5.1.2	Small-Signal Equivalent Circuit	5
5.1.3	Finding Voltage Gains, Input, and Output Resistances	6
5.2	The Inverter:	6
5.2.1	Constructing a Logic Gate: the Use of Pull-Down and Pull-Up Networks	7
5.2.2	NMOS Resistor Pull-Up	7
5.2.3	The CMOS Inverter	8
5.3	2-Input NAND Gate:	10
5.4	2-Input NOR Gate	10

Chapter 4. Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET)

4.1 Introduction

A **transistor** is a semiconductor device that uses a small amount of voltage or electrical current to control a larger change in voltage or current. Because of its fast response and accuracy, it may be used in a wide variety of applications, including amplification, switching, signal modulation, and as an oscillator. The transistor is the fundamental building block of both digital and analog circuits — the circuitry that governs the operation of computers, cellular phones, and all other modern electronics.

The **field-effect transistor** (FET) is a transistor that relies on an electric field to control the shape and hence the conductivity of a 'channel' in a semiconductor material. FETs are sometimes used as voltage-controlled resistors. Field-effect transistors are devices that are used in amplifiers and logic gates.

The **metal-oxide-semiconductor field-effect transistor** (MOSFET, MOS-FET, or MOS FET), is by far the most common field-effect transistor in both digital and analog circuits. A MOSFET is a three-terminal device that uses the voltage between two terminals to control the current flowing in the third terminal. Therefore it can be realized as a voltage-controlled current source.

Some of the basic symbols can be found in the following table.

Symbol	Definition
	N-MOSFET symbols. Note that the Drain terminal is on top and the Source terminal on bottom. In the first picture, the arrow points towards the Gate terminal.
	P-MOSFET symbols. Note that the Source terminal is on top and the Drain terminal on bottom. In the first picture, the arrow points away from the Gate terminal.

Table 1. Symbol Information depicting the symbols that will be used throughout the remainder of this text.

4.2 Notation

Superposition is a very important concept while analyzing transistors. For this reason, many types of variables with different subscripts are used. Upper-case letters with upper-case subscripts, e.g. V_{GS} , represent results due to DC analysis. For the value of a single point, it is usually labeled with a subscript "Q", e.g. V_{GSQ} . Lower-case letters with lower-case subscripts, e.g. v_{gs} , represent results due to AC analysis. Finally, lower-case letters with upper-case subscripts, e.g. v_{GS} , represent the general or total, i.e. the result achieved by summing the DC and the AC results.

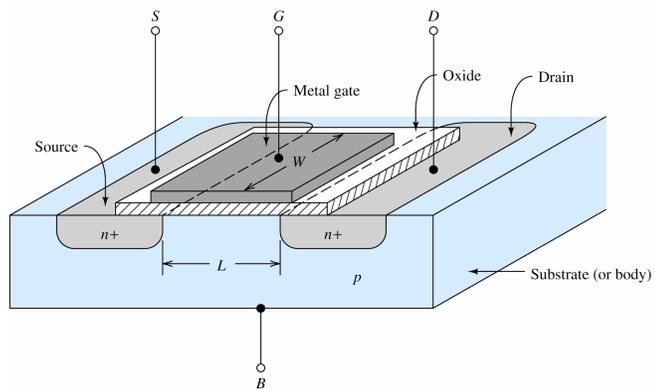


Fig. 1. A device drawing of an NMOS transistor. L represents the length of the channel and W , the width.

4.3 NMOS and PMOS Transistors

Figure 1 depicts the structure of an n-channel enhancement-mode MOSFET, also known as an **NMOS** transistor. The substrate, or body, is doped with acceptors to form p-type silicon. Two regions on the top surface of the substrate are doped to form n-type silicon as indicated by the n^+ regions in the figure. Metal is deposited to form contacts to the n^+ regions. The two contacts are labeled **source (S) and drain (D)**, with another metal to contact the bottom of the substrate, labeled **body (B)**. Between the source and drain, a metal contact is deposited on top of a layer of silicon dioxide, which, in turn, is deposited on top of the p-Si. This contact is labeled **gate (G)**. The gate metal, silicon dioxide and semiconductor underneath form the most important constituents, which is the reason that this type of FET is called MOS-FET.

At a first glance, the source-body and drain-body connections are both n-p junction. And hence, we do not expect different characteristics than a simple diode if a voltage is applied to source-body or drain-body independently. However, by placing a MOS junction in-between, very interesting characteristics is achieved.

Instead of using metal electrode, the gate of modern transistors are typically highly doped poly-silicon. Hence, it is best to explain the MOS as n++-O-p junction. When a positive gate-body voltage is applied, the junction is reverse biased. Due to the oxide being an insulator, no current flows through the gate terminal. As the gate-body voltage increased above a certain positive value (called threshold voltage V_{to}), a thin layer of electrons is formed at the oxide-Si interface. This layer is called the inversion layer, which forms a bridge to conduct electrons from source to drain. As the channel is n-type, with electrons as the conducting carriers, this type of MOSFET is called N-MOSFET. Current can flow into the drain, through the channel, and out the source if a drain to source voltage, v_{DS} , is applied. You may notice that the terms “source” and “drain” seem to be backwards. This is because they are the source and drain of *carriers*, which in the case of a NMOS are electrons. Varying the gate-body voltage changes the inversion layer thickness and hence the resistance of the drain-source channel. Typically we short the body and source contacts, and the gate-source voltage v_{GS} is used to control drain-source current i_{DS} . The device characteristics also depend on device dimensions, such as L , the length of the channel, and W , the width of the channel.

P-MOSFET is similar to NMOS, except all n-doped regions in Fig. 1 are p-doped and the substrate is n-doped. With a negative v_{GS} , the MOS junction is reverse biased. When v_{GS} becomes smaller than a certain V_{to} (typically negative), an inversion layer of holes is formed at the oxide-substrate interface, which forms the channel to conduct between drain-source. The source and body are conventionally shorted, as in NMOS. The drain-source voltage v_{DS} is kept the same sign as v_{GS} , which in this case will be negative. Thus i_{DS} is also negative.

4.4 N-MOSFET Operating Regions

The I-V characteristic for a MOSFET is more complicated than a diode, simply because there are two control voltages, v_{DS} and v_{GS} . Depending on their relative values, the MOSFET can be in one of three modes: cut-off, triode or saturation.

4.4.1 Cut-off

If $v_{GS} < V_{to}$, then no channel forms under the oxide insulator and even if a $v_{DS} > 0$ is applied, virtually no current will flow. Note this is indicated in Fig. 2 in which there appears a dark line at $i_D = 0$. This is known as the cut-off region and even as v_{GS} increases, the device remains in the cut-off region until $v_{GS} > V_{to}$.

$$i_D = 0 \text{ for } v_{GS} < V_{to} \quad (4.1)$$

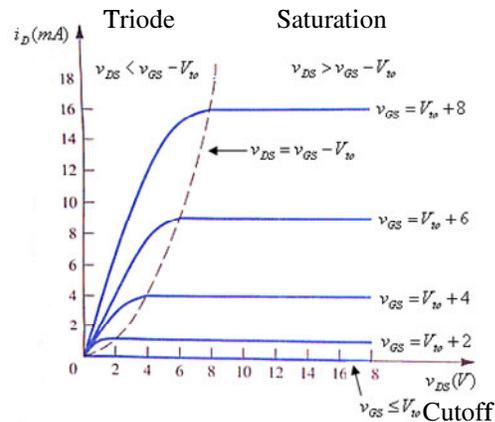


Fig. 2. NMOS characteristic curves. When $v_{GS} < V_{to}$, i_D is zero and the NMOS is in cut-off region. In the triode region, i_D increases for increases in v_{DS} . In the saturation region, i_D increases only if v_{GS} increases. The dashed line shows the boundary between the triode and the saturation region.

4.4.2 Triode

As v_{GS} increases above the threshold voltage and for $v_{DS} < v_{GS} - V_{to}$, the NMOS is in the triode region. The inversion layer is formed under the gate, between the source and the drain. For small positive v_{DS} , the current i_D increases with v_{DS} somewhat linearly. This is because the resistance of the channel is determined by the dimension of the inversion layer, which is independent of v_{DS} for small v_{DS} . As v_{DS} increases, current increases much less rapidly because the channel is getting thinner at the drain side. An equation to approximate the thinning of the inversion channel can be derived, which thus leads to the equation for i_D , which is proportional to v_{DS}^2 .

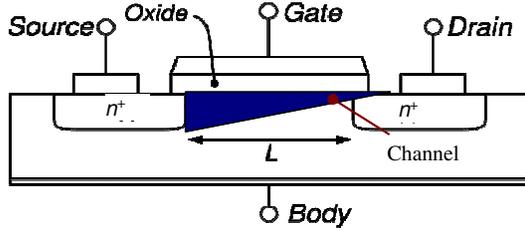


Fig. 3. Channel pinch-off occurs near the drain as v_{DS} increases. The thickness of the channel at the drain is zero when the device goes into saturation region.

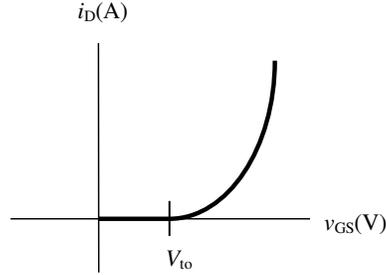


Fig. 4. When $v_{GS} < V_{to}$, the current flowing into the drain is zero. However, after the barrier has been reached, the current follows a quadratic trend with increases in v_{GS} .

In this regime, as v_{GS} increases, the channel becomes thicker, which reduces the resistance and increases i_D . Thus, we can think of the device as a voltage-controlled resistor that decreases its resistance with increases in v_{GS} . The triode region is to the left of the dashed line in Fig. 2.

$$i_D = K \left[2(v_{GS} - V_{to})v_{DS} - v_{DS}^2 \right] \quad \text{for } v_{DS} + V_{to} < v_{GS} \quad (4.2)$$

where $K = \frac{W}{L} \frac{KP}{2}$

where W is the width of the channel, L the length, and KP is a factor depending on the oxide layer.

4.4.3 Saturation

For a fixed v_{GS} , as v_{DS} increases, the gate-to-drain voltage, v_{GD} , will equal the threshold voltage ($v_{GS} - v_{DS} = V_{to}$). The thickness of the channel at the drain end becomes zero at this point. Thus, for further increases in v_{DS} , there are no further changes to i_D . This refers to saturation of the channel, and the curves begin to flatten as shown in Fig. 2. The saturation occurs due to the channel pinch-off caused by the disappearance of the channel layer close to the drain as shown in Fig. 3. Note, in the saturation region, the current is constant for a fixed v_{GS} and independent of v_{DS} . The drain current is given by:

$$i_D = K(v_{GS} - V_{to})^2 \quad \text{for } V_{to} \leq v_{GS} \leq v_{DS} + V_{to} \quad (4.3)$$

where K is the same as above. Fig. 4 shows i_D versus v_{GS} . Note, $i_D=0$ when $v_{GS} < V_{to}$.

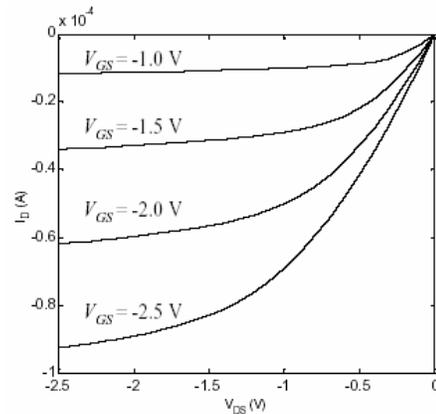


Fig. 5. PMOS characteristic curves. PMOS is in cut-off region when $v_{GS} > V_{to}$ and $i_D=0$. In the triode region, i_D decreases for decreasing v_{DS} . In the saturation region, i_D decreases only if v_{GS} decreases. The dashed line shows the boundary between the triode and the saturation region.

4.5 PMOSFET Operating Regions

PMOS also has three operating regions, cut-off, saturation and triode. In general, all signs are opposite to those of NMOS. Typical PMOS i_D - v_{DS} curves for various v_{GS} values are shown in Fig. 5. Note that in the following as well as in Fig. 5, $i_D=i_{DS}$. This is different from our text book, where $i_D=i_{SD}$ for PMOS.

	NMOS	PMOS (Note $i_D=i_{DS}$)
Channel carrier	Electrons	Holes
V_{to}	Positive	Negative
Cut-off Region	$i_D = 0$ for $v_{GS} < V_{to}$	$i_D = 0$ for $v_{GS} > V_{to}$
Saturation Region	$i_D = K(v_{GS} - V_{to})^2$ for $V_{to} \leq v_{GS} \leq v_{DS} + V_{to}$	$i_D = -K(v_{GS} - V_{to})^2$ for $V_{to} \geq v_{GS} \geq v_{DS} + V_{to}$
Triode Region	$i_D = K[2(v_{GS} - V_{to})v_{DS} - v_{DS}^2]$ for $v_{DS} + V_{to} < v_{GS}$	$i_D = -K[2(v_{GS} - V_{to})v_{DS} - v_{DS}^2]$ for $v_{DS} + V_{to} > v_{GS}$

The following summarizes the regions of operation for NMOS and PMOS.

NMOS Regions of operation	PMOS Regions of operation

Chapter 5. Simple MOSFET Circuits

5.1 Analysis for MOSFET Amplifiers

As previously discussed, we are interested in analyzing MOSFET circuits for analog and digital applications. For analog applications, we are most concerned with small-signal AC response. In particular, we are interested in three parameters: small-signal gain, input impedance and output impedance. The latter two enables us to cascade the amplifiers. The steps include the following:

1. DC analysis to determine the Quiescent (Q) operating point. The results of this analysis are I_{DQ} , V_{DSQ} , and V_{GSQ} , with which we can get the small-signal MOSFET model
2. Replace the original circuit with the small-signal MOSFET model.
3. Analyze the new AC circuit use KCL, KVL or Thevenin (or Norton) equivalent circuits.

In the following, we will show how to obtain these three parameters in a step-by-step fashion.

5.1.1 DC Analysis – Load-Line Analysis

Since it is DC analysis, we turn-off all AC sources and **replace all capacitors with opens and inductors with shorts**.

After this step, we can write two KVL equations.

- Equation (5.1) will be for the loop including V_{DS} and equation will be in the form of I_D as a function of V_{DS} . Remember that there is a voltage drop of V_{DS} from the drain to the source terminal.
- Equation (5.2) will be for the loop V_{GS} . The most crucial point to remember is that **there is NO gate current**. This equation will be in the form of I_D as a function of V_{GS} .

A. If equation 5.2 leads to a constant V_{GS} , we can use equation 5.1 as the load line in conjunction with MOSFET I_D - V_{DS} characteristic curve, e.g. Figure 4.2, to solve for I_D and V_{DS} .

B. If equation 5.2 is a functional equation and V_{GS} is a function of I_D . First, we assume the bias is such that the MOSFET is in saturation region, use load-line analysis for equation (5.2) and MOSFET saturation mode equation (4.3) to solve for I_D and V_{GS} . With these two values, we can then find V_{DS} using equation (5.1).

But if there is no solution for V_{GS} , we know MOSFET is in triode mode and will need to use equation (4.2) to solve for I_D , V_{GS} , and V_{DS} .

After determining the DC bias point, i.e. Q point, we label the values for I_{DQ} , V_{GSQ} , and V_{DSQ} . We can now move on to the small signal model.

5.1.2 Small-Signal Equivalent Circuit

To do AC analysis on the circuit, we need to replace DC sources by ground, capacitors by short circuits, and inductance by open circuits. Next, we replace the MOSFET a voltage controlled current source in parallel with a resistor. This model is valid only when the MOSFET is biased in the saturation mode.

The reason for the resistor is because in the saturation region, i_D is not exactly flat as shown in Fig. 4.2 but slopes slightly upward. Usually, it is a good approximation for the line to be horizontal (in Figure 4.2), in which case r_d is infinite. Note that the current source is connected between the drain and the source terminal. The value, g_m known as transconductance of the MOSFET is a measure of a transistor's sensitivity to the input voltage. If the characteristic curves such as Figure 4.2 are provided, you can find the value of g_m by holding v_{DS} constant at V_{DSQ} and calculating the change in i_D with respect to a change in v_{GS} :

$$g_m = \left(\Delta i_D / \Delta v_{GS} \right) \Big|_{v_{DS}=V_{DSQ}} \quad (5.3)$$

On the other hand, if an equation is provided for i_D , g_m is found by taking the partial derivative of that equation with respect to v_{GS} :

$$g_m = \partial i_D / \partial v_{GS} \quad (5.4)$$

Plugging in equation 4.3 into equation 5.4 yields another way to find g_m :

$$\begin{aligned} g_m &= 2K(V_{GSQ} - V_{to}) \\ &= 2\sqrt{KI_{DQ}} \\ &= \sqrt{2KP} \sqrt{W/L} \sqrt{I_{DQ}} \end{aligned} \quad (5.5)$$

Similarly, through characteristic curves, you can get the value of r_d by holding v_{GS} constant at V_{GSQ} and calculating the change in i_D with respect to a change in v_{DS} :

$$(1/r_d) = (\Delta i_D / \Delta v_{DS}) \Big|_{v_{GS}=V_{GSQ}} \quad (5.6)$$

Remember, you must reciprocate the above equation to find r_d .

Similarly, r_d is found by taking the partial derivative of the i_D equation with respect to v_{DS} :

$$(1/r_D) = \partial i_D / \partial v_{DS} \quad (5.7)$$

Remember, you must reciprocate the above equation to find r_d .

5.1.3 Finding Voltage Gains, Input, and Output Resistances

After drawing the small-signal equivalent circuit, we can find small signal voltage gain, input, and output resistances. There are examples in the text book (pages 556, and 569-570). The terms are defined as follows:

Voltage gain A_v : the ratio of the output voltage to the input (AC) voltage.

Input Impedance R_{in} : the impedance looking into the circuit from the input terminal; it is equal to input voltage, v_{in} , divided by input current, i_{in} .

Output Impedance R_{out} : The output impedance is the impedance looking into the circuit from the load terminals. It can be found most often with steps 1 and 2 below and then by inspection into the load terminals. However, sometimes it requires step 3 to find the value.

1. Remove the load resistance, R_L from the small-signal equivalent circuit.
2. Turn off independent source (remember voltage sources become short circuits when they are zeroed out and current source will be open).
3. Attach a test source, v_x , to the output terminals. The output terminal is where you removed the load resistance from. Find the current coming out of the test source as i_x . Use KCL or KVL to write down an equation relating v_x and i_x . Simplify the equation such that these are the only two unknowns in your equation. The output resistance, R_{out} , is equal to v_x divided by i_x .

5.2 The Inverter:

Fig. 5.2(a) below shows the symbol of an inverter, also known as a NOT Gate. Fig. 5.2(b) depicts a graph of V_{out} vs. V_{in} , which shows the function of an ideal inverter: when the input voltage is low, the output voltage is high and vice versa.

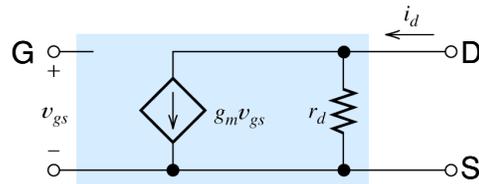


Fig. 5.1. Small signal model for MOSFET under saturation mode. It includes a voltage-controlled current source between drain and source terminals. The direction of the current is always from the drain to the source.

5.2.1 Constructing a Logic Gate: the Use of Pull-Down and Pull-Up Networks

Because an inverter is a type of logic gate, it will be better to first analyze how logic gates are made. To construct a logic gate, we must make use of both pull-down and pull-up networks. A pull-down network is a set of devices used to carry current from the output node to ground. This discharges the output node hence pulling down the voltage. NMOSFETs function as pull-down devices when they are turned on and are used to connect the output to ground. Note that when an NMOS is on, the n-type channel formed within it allows for current to easily flow when a voltage is applied to it. Remember that the channel can only form when the input voltage to the NMOS, V_{GS} , is high and greater than V_{to} . So, you can think of an NMOS as a short circuit when its input voltage is high and an open circuit when its input voltage is low.

PMOSFETs and resistors function as pull-up devices and are used to connect the output to the DC source, V_{DD} . Note that when a PMOS is on, the p-type channel formed within it allows for current to easily flow when a voltage is applied to it. Remember that the channel can only form when the input voltage to the PMOS, V_{GS} , is low and less than V_{to} . So, you can think of a PMOS as a short circuit when its input voltage is low and an open circuit when its input voltage is high.

Fig. 5.3 shows the setup of a typical pull-up network in series with a pull-down network to yield a logic gate with multiple inputs. In this class, we will be analyzing these networks with mostly one input.

Fig. 5.4 and Fig. 5.7 show circuit diagrams using specific devices for the pull-up and pull-down network. The reason for the setup in these figures yielding an inverter is discussed later. Note that the pull-up network is used to connect V_{DD} to F (the output node) and the pull-down network is connecting F to ground. Note that the input signals, (A_1, \dots, A_N) , to each network are the same. So, if the signals are high, the pull-up network is off (it's an open circuit), and the pull-down network is on (it's a short circuit). This means that the voltage drop across the pull-down network which equals F is 0 V or low. This is where the name pull-down comes from, as F is pulled to ground. On the other hand, if the signals are low, the pull-up network is on (it's a short circuit), and the pull-down network is off (it's an open circuit). This means that the voltage drop across the pull-up network is 0 V and there is voltage drop of V_{DD} across the pull-down network. Because F equals the voltage drop across the pull-down network, it is also V_{DD} , or high. Because of a 0 V drop across itself, the pull-up network yields the high output voltage across F . This is where the name pull-up comes from.

5.2.2 NMOS Resistor Pull-Up

Fig. 5.4 shows the NMOS resistor pull-up device, which acts as an inverter. The resistor acts as the pull-up device and the NMOS as the pull-down device. Note that the pull-up device connects V_{DD} to the output, V_{OUT} , and the pull-down device connects V_{OUT} to the ground. The way this device works as an inverter can readily be seen in Fig. 5.5 and in Fig. 5.6.

Fig. 5.5 shows the load-line analysis of this circuit. The load line has been constructed in the same way mentioned earlier and is graphed with the NMOS's characteristic curves. Note that V_{GS} is simply V_{IN} and V_{DS} is equal to V_{OUT} . Each intersecting point gives the value of V_{DS} for a given V_{GS} . Note when V_{GS} or V_{IN} is small, V_{DS} or V_{OUT} is large and vice versa. This also makes sense intuitively. When V_{GS} or V_{IN} is low (below V_{to}), no channel forms within the NMOS and therefore no current flows through it, which means the transistor is off. Because no current flows through the transistor, no current flows through the resistor. This means that V_{OUT} equals V_{DD} , which is high. When V_{IN} is high (above V_{to}), a channel does form and current

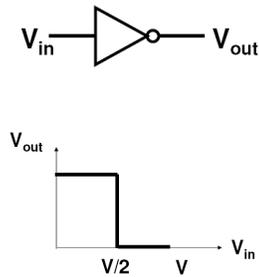


Fig. 5.2(a) above shows the symbol of a basic inverter. (b) shows the ideal characteristic of an inverter. A low input voltage yields a high output voltage and vice versa.

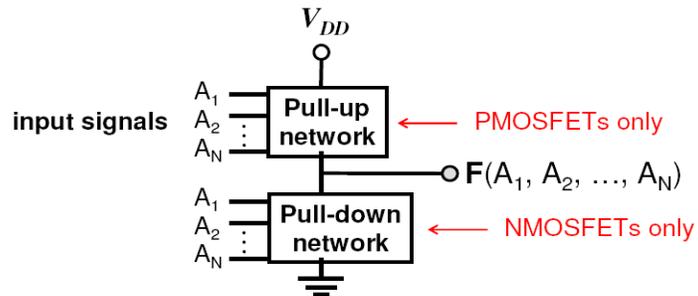


Fig. 5.3. The basic setup of a pull-up network and a pull-down network which yields the basic inverter. F is the output signal and is equal to the voltage drop across the pull-down network.

flows easily through the transistor, which means there is current flowing through the resistor and there is a voltage drop across the resistor. As the current gets larger, most of V_{DD} is dropped across the resistor and the V_{DS} gets smaller. The current gets larger when the channel gets wider, which only occurs with an increase in V_{GS} or V_{IN} . Fig. 5.6 is derived from the intersecting points of Fig. 5.5. For example, when V_{GS} is very high, V_{DD} is very low, which is represented by the green diamond point labeled in Fig. 5.5. This point is also labeled in Fig. 5.6 with a green diamond. As V_{GS} lowers, V_{DD} increases and this point is represented by the square in figures 5.5 and 5.6.

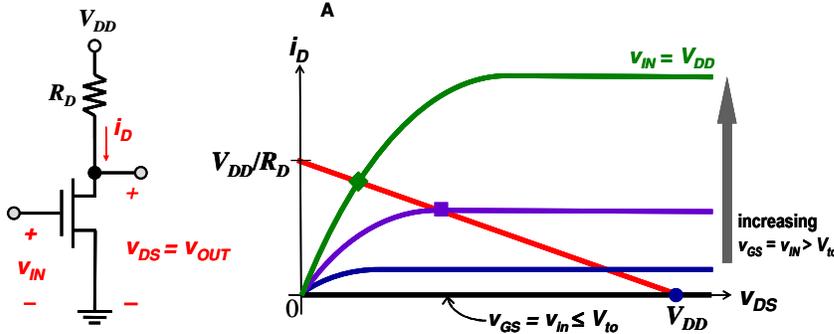


Fig. 5.4. The NMOS resistor pull-up

Fig. 5.5. The load line analysis of the NMOS resistor pull-up. Note V_{GS} increases upward. Note V_{GS} equals V_{IN} and V_{DS} equals V_{OUT} .

Fig. 5.6 A plot of V_{IN} vs. V_{OUT} for the NMOS resistor pull-up. This is derived from Fig. 5.5, so the diamond point in Fig. 5.5 corresponds to the diamond point in this figure.

5.2.3 The CMOS Inverter

Fig. 5.7(a) shows a CMOS inverter, which contains a PMOS and an NMOS. The PMOS acts as the pull-up device and connects the DC source, V_{DD} , to the output node. The NMOS acts as the pull-down device and is connected in the same way as in the NMOS Resistor pull-up device. To see how the inverter works in digital is simple: when V_{IN} is high, both V_{GSN} (V_{GS} of NMOS) and V_{GSP} (V_{GS} of PMOS) are also high. This means that a channel forms in the NMOS, but no channel forms within the PMOS, so the NMOS acts as a short circuit, but the PMOS is an open circuit. Fig. 5.7(b) shows the circuit diagram for when V_{IN} is high. Because V_{OUT} is measured across the NMOS, it becomes 0 V and is therefore low. Conversely, when V_{IN} is low, a channel forms within the PMOS and no channel forms with the NMOS, so the PMOS acts like a short and the NMOS like an open. Fig. 5.7(c) shows the circuit diagram for when V_{IN} is low. Here also, V_{OUT} is measured across the NMOS, which makes V_{OUT} equal V_{DD} and therefore high.

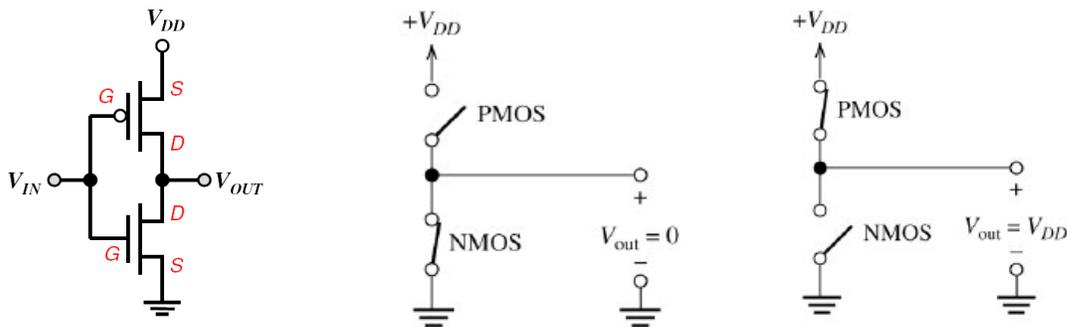


Fig. 5.7(a). The CMOS inverter. This inverter uses a PMOS as the pull-up device and an NMOS as the pull-down device.

Fig. 5.7(b). The circuit diagram for when V_{IN} is high. The PMOS acts like an open circuit while the NMOS acts like a short circuit.

Fig. 5.7(c). The circuit diagram for when V_{IN} is low. The PMOS acts like a short circuit while the NMOS acts like an open circuit.

To analyze this circuit using a more analog method is a bit more complex. We must place the characteristic curves of both the NMOS and the PMOS on the same set of axes. Fig. 5.8(a) shows the I - V characteristic of a PMOS. Note the axes are labeled I_{DSP} and V_{DSP} . Earlier, in Fig. 3, we saw the I - V characteristic of an

NMOS. Even though the axes are labeled I_D and V_{DS} , they should really be labeled as I_{DSN} and V_{DSN} . To plot the two graphs on the same set of axes, we must first find a relationship between these variables. Looking at Fig. 5.7(a), we can obtain the following equations:

$$V_{IN} = V_{DD} + V_{GSP} \tag{5.8}$$

$$V_{OUT} = V_{DD} + V_{DSP} \tag{5.9}$$

Note that V_{IN} is the same as V_{GSN} and V_{OUT} is the same as V_{DSN} , so the above equations become:

$$V_{GSN} = V_{DD} + V_{GSP} \tag{5.10}$$

$$V_{DSN} = V_{DD} + V_{DSP} \tag{5.11}$$

Also note, that:

$$I_{DSN} = -I_{DSP} \tag{5.12}$$

So, to plot the PMOS characteristic curves on an I_{DSN} vs. V_{DSN} , we must change the axes from those labeled in Fig. 5.8(a) to those labeled in Fig. 5.8(c). So, we must invert all the values on the y-axis, which yields a reflection of the PMOS graph across its x-axis. This is shown in the Fig. 5.8(b). Note the axes now read I_{DSN} and V_{DSP} . To change the x-axis of the new PMOS graph to V_{DSN} , we use the equation above which states that $V_{DSN} = V_{DD} + V_{DSP}$. This means that we must shift the x-axis of the new PMOS graph to the right by V_{DD} . This is shown in Fig. 5.8(c). Now, placing the new PMOS graph and the NMOS graph, we obtain Fig. 5.9.

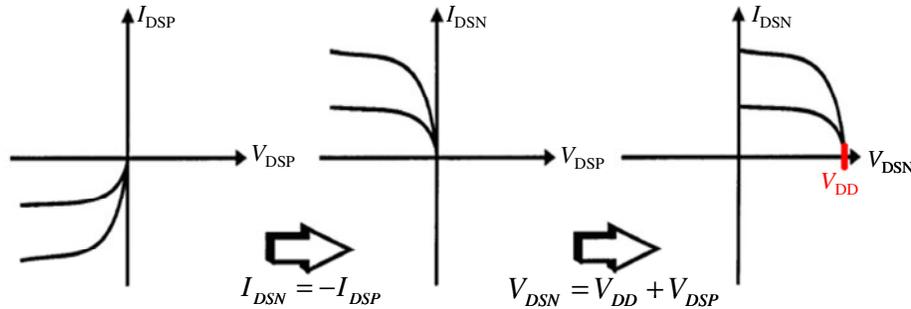


Fig. 5.8(a) shows the I - V characteristic curves of a PMOS. Note that the axes are I_{DSP} and V_{DSP} . In Fig. 5.8(b), the y-axis has been changed to I_{DSN} by the relationship indicated by the arrow. In Fig. 5.8(c), the x-axis has been changed to V_{DSN} by the relationship indicated by the arrow. Fig. 5.8(a) and (c) both show the characteristic curves for a PMOS, however on two different set of axes.

In Fig. 5.9, V_{GSN} increases upward for the NMOS while for the PMOS, V_{GSP} increases downward because for a PMOS, all sign conventions are opposite that of an NMOS. Note that V_{GSN} equals V_{IN} and that there are many intersecting points in the graph, so how do we know which is the right one? The answer is, for a given DC voltage, V_{DD} , and a chosen V_{IN} , the equations above will yield exactly one V_{GSP} . For example, suppose that V_{io} , for the NMOS is 1 V, -1 V for the PMOS, and V_{DD} is fixed at 5 V. Now, let V_{IN} be 1 V, which means that V_{GSN} is also 1 V. Using these values, and the above equations, we find that V_{GSP} equals -4 V. We must find the point where V_{GSN} is 1 V and V_{GSP} is -4 V. This is labeled by the black-colored circle in Fig. 5.9. At this point V_{DSN} equals V_{DD} which equals 5 V. Because V_{DSN} also equals V_{OUT} , V_{OUT} also equals 5 V. So, for V_{IN} equaling 1 V, V_{OUT} equals 5 V. This is labeled in Fig. 5.10 by the same black-colored circle. You should try other values for V_{IN} keeping V_{DD} fixed to see if you obtain the other points labeled in Fig. 5.9. The points labeled in Fig. 5.9 have corresponding points labeled in the same manner in Fig. 5.10, so the white-colored square in Fig. 5.9 corresponds to the white-colored square in Fig. 5.10, etc. By analyzing Fig. 5.9 to yield Fig. 5.10, it becomes clear that the device yields a low output for a high input and vice versa.

5.3 2-Input NAND Gate:

Fig. 5.11 below shows a 2-input NAND gate. It has two NMOS devices in series and two PMOS devices in parallel. To make an M-input NAND gate, you can place M-NMOS devices in series and M-PMOS devices in parallel. To see how this works, remember a NAND gate yields low only when all its inputs are high as indicated in the truth table in Table 2. To see that this works in Fig. 5.11, remember that when the input to an NMOS is high, the NMOS turns on and becomes a short circuit, and when the input is low, the NMOS is off becoming an open circuit. For a PMOS, the conditions are reversed. When both A and B are low, N_1 and N_2 are off, and P_1 and P_2 are on. V_{out} therefore equals V_{DD} and is high. When A is low and B high, P_1 and N_2 are on, and P_2 and N_1 are off. Because N_1 is an open circuit, there is a voltage drop across it which means V_{out} is high. When A is high and B low, N_1 and P_2 are on, and N_2 and P_1 are off. Because N_2 is an open circuit, there is a voltage drop across it which means V_{out} is high. When both A and B are high, N_1 and N_2 are on, and P_1 and P_2 are off. V_{out} therefore equals 0 and is low. This satisfies the truth table shown in Table 2.

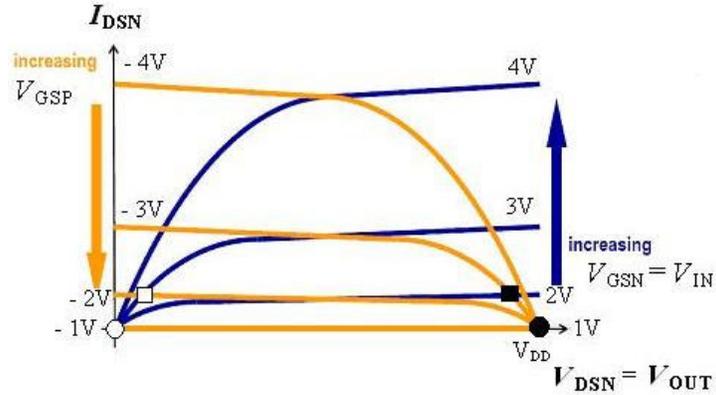


Fig. 5.9. PMOS and NMOS characteristic curves on the same graph with axes I_{DSN} - V_{DSN} . NMOS is in dark and PMOS in light. Note for PMOS, V_{GS} increases as we go down, while for the NMOS, it increases when we go up. The labeled points have corresponding points labeled in the same manner as in Fig. 5.10.

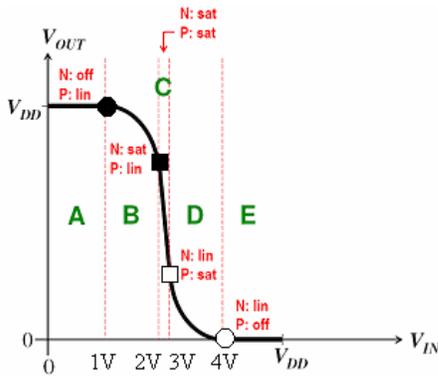


Fig. 5.10. This shows the plot of V_{OUT} vs. V_{IN} for the CMOS inverter. Note when V_{IN} is low, the NMOS is off and the PMOS is in triode region, and V_{OUT} is high. Conversely, when V_{IN} is high, the PMOS is off and the NMOS is in triode region, and V_{OUT} is low. The graph labels the transitions of each transistor as V_{IN} and V_{OUT} change. The labeled points correspond to those in Fig. 5.9.

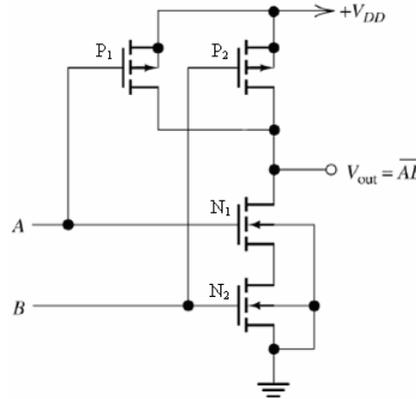


Fig. 5.11. A 2-input NAND gate. There are two NMOS transistors in series and two PMOS transistors in parallel. The output voltage is low only when both NMOS transistors are on, which means that both inputs, A and B, are high.

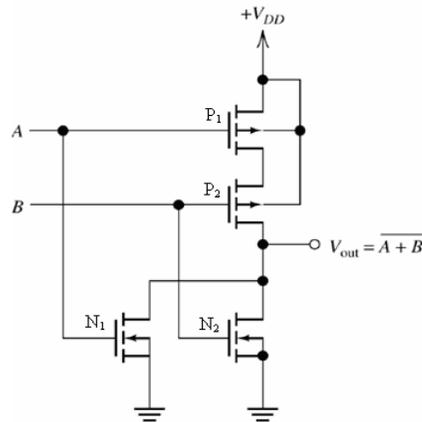
A	B	V_{out}
0	0	1
0	1	1
1	0	1
1	1	0

Table 2. The truth table for a NAND gate. Note that the output voltage is low only when both its inputs are high.

5.4 2-Input NOR Gate

Fig. 5.12 shows a 2-input NOR gate. It has two NMOS devices in parallel and two PMOS devices in series. To make an M-input NOR gate, you can place M-NMOS devices in parallel and M-PMOS devices in series. To see how this works, remember a NOR gate yields high only when all its inputs are low as indicated in the truth table in Table 3. When both A and B are low, N_1 and N_2 are off, and P_1 and P_2 are on. V_{out} therefore equals V_{DD} and is high. When A is low and B high, P_1 and N_2 are on, and P_2 and N_1 are off. Because N_2 is a short circuit, there is no voltage drop across it which means V_{out} is low. When A is high and B low, N_1 and P_2 are on, and N_2 and P_1 are off. Because N_1 is a short circuit, there is no voltage drop across

it which means V_{out} is low. When both A and B are high, N_1 and N_2 are on, and P_1 and P_2 are off. V_{out} therefore equals 0 and is low. This satisfies the truth table shown in Table 3.



A	B	V_{out}
0	0	1
0	1	0
1	0	0
1	1	0

Fig. 5.12. A 2-input NOR gate. There are two NMOS transistors in parallel and two PMOS transistors in series. The output voltage is high only when both NMOS transistors are off, which means that both inputs, A and B, are low.

Table 3. The truth table for a NOR gate. Note that the output voltage is high only when both its inputs are low.