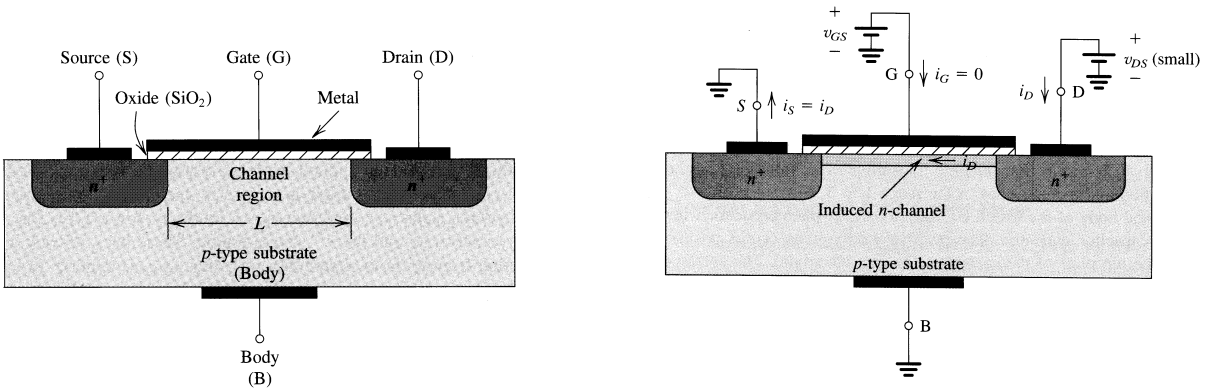
Field-Effect (FET) transistors

References: Barbow (Chapter 8), Rizzoni (chapters 8 & 9)

In a field-effect transistor (FET), the width of a conducting channel in a semiconductor and, therefore, its current-carrying capability, is varied by the application of an electric field (thus, the name field-effect transistor). As such, a FET is a "voltage-controlled" device. The most widely used FETs are Metal-Oxide-Semiconductor FETs (or MOSFET). MOSFET can be manufactured as enhancement-type or depletion-type MOSFETs. Another type of FET is the Junction Field-Effect Transistors (JFET) which is not based on metal-oxide fabrication technique. FETs in each of these three categories can be fabricated either as a n-channel device or a p-channel device. As transistors in these 6 FET categories behave in a very similar fashion, we will focus below on the operation of enhancement MOSFETs that are the most popular.

n-Channel Enhancement-Type MOSFET (NMOS)

The physical structure of a n-Channel Enhancement-Type MOSFET (NMOS) is shown. The device is fabricated on a p-type substrate (or Body). Two heavily doped n-type regions (Source and Drain) are created in the substrate. A thin (fraction of micron) layer of SiO_2, which is an excellent electrical insulator, is deposited between source and drain region. Metal is deposited on the insulator to form the Gate of the device (thus, metal-oxide semiconductor). Metal contacts are also made to the source, drain, and body region.

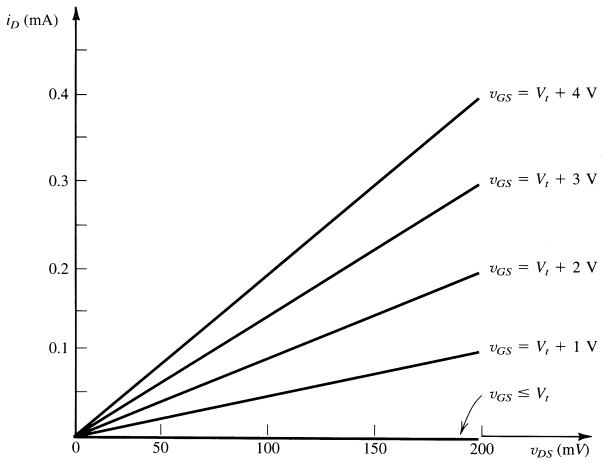


To see the operation of a NMOS, let's ground the source and the body and apply a voltage v_{GS} between the gate and the source, as is shown above. This voltage repels the holes in the p-type substrate near the gate region, lowering the concentration of the holes. As v_{GS} increases, hole concentration decreases, and the region near gate behaves progressively more like intrinsic semiconductor material (when excess hole concentration is zero) and then, finally, like a n-type material as electrons from n^+ electrodes (source and drain) enter this region. As a result, when v_{GS} become larger than a threshold voltage, V_t, a narrow layer

between source and drain regions is created that is populated with n-type charges (see figure). The thickness of this channel is controlled by the applied v_{GS} (it is proportional to $v_{GS} - V_t$).

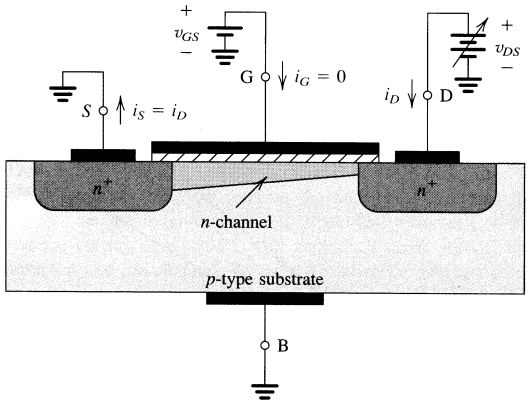
As can be seen, this device works as a channel is induced in the semiconductor and this channel contains n-type charges (thus, n-channel MOSFET). In addition, increasing v_{GS} increases channel width (enhances it). Therefore, this is an n-channel Enhancement-type MOSFET or Enhancement-type NMOS.

Now for a given values of $v_{GS} > V_t$ (so that the channel is formed), let's apply a small and positive voltage v_{DS} between drain and source. Then, electrons from n^+ source region enter the channel and reach the drain. If v_{DS} is increased, current i_D flowing through the channel increases. Effectively, the device acts like a resistor; its resistance is set by the dimension of the channel and its n-type charge concentration. In this regime, plot of i_D versus v_{DS} is a straight line (for a given values of $v_{GS} > V_t$) as is shown.

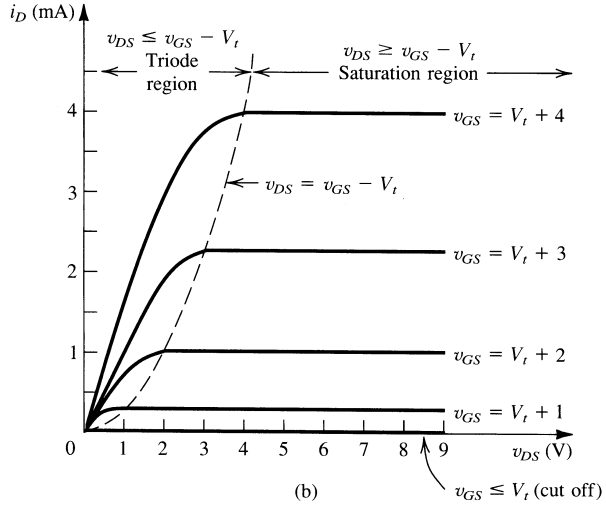


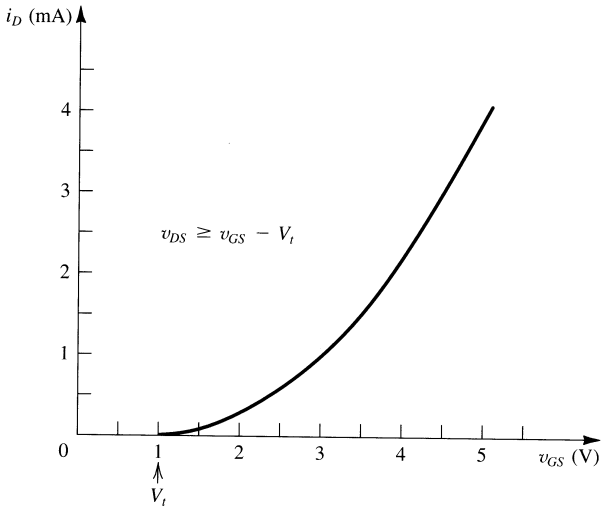
The slope of i_D versus v_{DS} line is the conductance of the channel. For $v_{GS} \le V_t$ no channel exists (zero conductance). As value of $v_{GS} - V_t$ increases, channel becomes wider and its conductivity increases (because its n-type charge concentration increases). Therefore, the channel conductance (slope of i_D versus v_{DS} line) increases with any increase in $v_{GS} - V_t$ as is shown above.

The above description is correct for small values of v_{DS} as in that case, $v_{GD} = v_{GS} - v_{DS} \approx v_{GS}$ and the induced channel is fairly uniform (*i.e.,* has the same width near the drain as it has near the source). For a given $v_{GS} > V_t$, if we now increase v_{DS}, $v_{GD} = v_{GS} - v_{DS}$ becomes smaller than v_{GS}. As such the size of channel near drain becomes smaller compared to its size near the source, as is shown. As the size of channel become smaller, its resistance increases and the curve of i_D versus v_{DS} starts to roll over, as is shown below.



For values of $v_{GD} = V_t$ (or $v_{DS} = v_{GS} - V_t$), width of the channel approaches zero near the drain (channel is "pinched" off). Increasing v_{DS} beyond this value has little effect (no effect in our simple picture) on the channel shape, and the current through the channel remains constant at the value reached when $v_{DS} = v_{GS} - V_t$. So when the channel is pinched off, i_D only depends on v_{GS} (right figure below).





NMOS Characteristic Curves i_D versus v_{GS} in the active (or saturation) regime

In sum, a FET can operate in three regimes:

1) Cut-off regime in which no channel exists ($v_{GS} < V_t$ for NMOS) and $i_D = 0$ for any v_{DS}.

2) Ohmic or Triode regime in which the channel is formed and not pinched off ($v_{GS} > V_t$ and $v_{DS} \le v_{GS} - V_t$ for NMOS) and FET behaves as a "voltage-controlled" resistor.

3) Active or Saturation regime in which the channel is pinched off ($v_{GS} \ge V_t$ and $v_{DS} > v_{GS} - V_t$ for NMOS) and i_D does not change with v_{DS}.

Several important point should be noted:

1) No current flows into the gate, $i_G = 0$ (note the insulator between gate and the body).

2) FET acts as a "voltage-controlled" resistor in the ohmic region. In addition, when $v_{DS} \ll v_{GS}$, FET would act as a linear resistor.

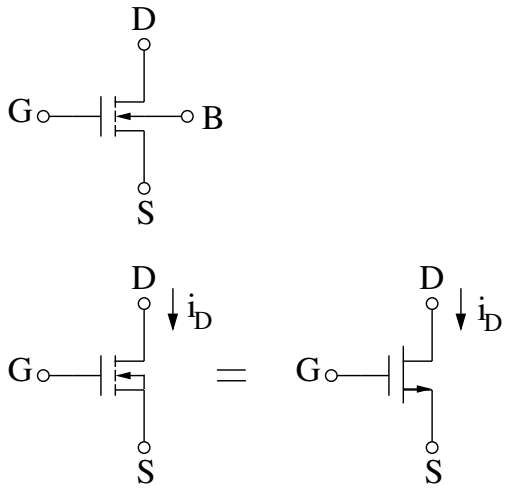
3) When FET is in cut-off, $i_D = 0$. However, if $i_D = 0$, this does not mean that FET is in cut-off. FET is in cut-off only when a channel does not exist ($v_{GS} < V_t$) and $i_D = 0$ for any applied v_{DS}. On the other hand, FET can be in ohmic region, *i.e.,* a channel is formed, but $i_D = 0$ because $v_{DS} = 0$.

4) The third regime is called "saturation" in most electronic books because i_D is "saturated" in this regime and does not increase further. This is a rather unfortunate name as "saturation" regime in a FET means very different thing than it does in a BJT. Some newer books call this regime "active" (as it equivalent to "active-linear" regime of a BJT).

5) The transition between ohmic and active region is clearly defined by $v_{DS} = v_{GS} - V_t$ the point where the channel is pinched off.

The i_D versus v_{DS} characteristic curves of a FET look very similar to i_C versus v_{CE} characteristics curves of a BJT. In fact, as there is a unique relationship between i_B and v_{BE}, the i_C versus v_{CE} characteristic curves of a BJT can be "labeled" with different values of v_{BE} instead of i_B, making the characteristic curves of the two devices even more similar. In FET v_{GS} control device behavior and in BJT v_{BE}. Both devices are in cut-off when the "input" voltage is below a threshold value: $v_{BE} < v_\gamma$ for BJT and $v_{GS} < V_t$ for NMOS. They exhibit an "active" regime in which the "output" current (i_C or i_D) is roughly constant as the "output" voltage (v_{CE} or v_{DS}) is changed. There are, however, major differences. Most importantly, a BJT requires i_B to operate but in a FET $i_G = 0$ (actually very small). These differences become clearer as we explore FETs.

As can be seen from NMOS physical structure, the device is symmetric, that is position of drain and source can be replaced without any change in device properties. The circuit symbol for a NMOS is shown on the right. For most applications, however, the body is connected to the source, leading to a 3-terminal element. In that case, source and drain are not interchangeable. A simplified circuit symbol for this configuration is usually used. By convention, current i_D flows into the drain for a NMOS (see figure). As $i_G = 0$, the same current will flow out of the source.



Direction of "arrows" used to identify semiconductor types in a transistor may appear confusing. The arrows do NOT represent the direction of current flow in the device. Rather, they denote the direction of the underlying *pn* junction. For a NMOS, the arrow is placed on the body and pointing inward as the body is made of *p*-type material. (Arrow is not on source or drain as they are interchangeable.) In the simplified symbol for the case when body and source is connected, arrow is on the source (device is not symmetric now) and is pointing outward as the source is made of *n*-type materials. (*i.,e.* arrow pointing inward for *p*-type, arrow pointing outward for *n*-type).

NMOS i_D versus v_{DS} Characteristics Equations

Like BJT, a NMOS (with source connected to body) has six parameters (three voltages and three currents), two of which (i_S and v_{GD}) can be found in terms of the other four by KVL and KCL. NMOS is simpler than BJT because $i_G = 0$ (and $i_S = i_D$). Therefore, three parameters describe behavior of a NMOS (v_{GS}, i_D, and v_{DS}). NMOS has one characteristics equation that relates these three parameters. Again, situation is simpler than BJT as simple but accurate characteristics equations exist.

Cut-off: $v_{GS} < V_t, \quad i_D = 0 \quad$ for any v_{DS}

Ohmic: $v_{GS} > V_t, \quad i_D = K[2v_{DS}(v_{GS} - V_t) - v_{DS}^2] \quad$ for $v_{DS} < v_{GS} - V_t$

Active: $v_{GS} > V_t, \quad i_D = K(v_{GS} - V_t)^2 \quad$ for $v_{DS} > v_{GS} - V_t$

In the above, K and V_t are constants that depend on manufacturing of the NMOS (given in manufacturer spec sheets). As mentioned above, for small values of v_{DS} ($v_{DS} \ll v_{GS} - V_t$), NMOS behaves as resistor, r_{DS}, and the value of r_{DS} is controlled by $v_{GS} - V_t$. This can be seen by dropping v_{DS}^2 in i_D equation of ohmic regime:

$$r_{DS} = \frac{v_{DS}}{i_D} \approx \frac{1}{2K(v_{GS} - V_t)}$$

How to Solve NMOS Circuits:

Solution method is very similar to BJT circuit (actually simpler because $i_G = 0$). To solve, we assume that NMOS is in a particular state, use NMOS model for that state to solve the circuit and check the validity of our assumption by checking the inequalities in the model for that state. A formal procedure is:

1) Write down a KVL including GS terminals (call it GS-KVL).

2) Write down a KVL including DS terminals (call it DS-KVL).

3) From GS-KVL, compute v_{GS} (using $i_G = 0$)

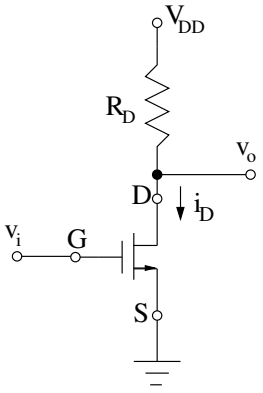
3a) If $v_{GS} < V_t$, NMOS is in cut-off. Let $i_D = 0$, solve for v_{DS} from DS-KVL. We are done.

3b) If $v_{GS} > V_t$, NMOS is not in cut-off. Go to step 4.

4) Assume NMOS is in active region. Compute i_D from $i_D = K(v_{GS} - V_t)^2$. Then, use DS-KVL to compute v_{DS}. If $v_{DS} > v_{GS} - V_t$, we are done. Otherwise go to step 5.

5) NMOS has to be in ohmic region. Substitute for i_D from $i_D = K[2v_{DS}(v_{GS} - V_t) - v_{DS}^2]$ in DS-KVL. You will get a quadratic equation in v_{DS}. Find v_{DS} (one of the two roots of the equation will be unphysical). Check to make sure that $v_{DS} < v_{GS} - V_t$. Substitute v_{DS} in DS-KVL to find i_D.

Example: Consider NMOS circuit below with $K = 0.25$ mA/V^2 and $V_t = 2$ V. Find v_o when $v_i = 0$, 6, and 12 V for $R_D = 1$ KΩ and $V_{DD} = 12$ V.



GS-KVL: $v_{GS} = v_i$

DS-KVL: $V_{DD} = R_D i_D + v_{DS}$

A) $v_i = 0$ V. From GS-KVL, we get $v_{GS} = v_i = 0$. As $v_{GS} < V_t = 2$ V, NMOS is in cut-off, $i_D = 0$, and v_{DS} is found from DS-KVL:

DS-KVL: $v_o = v_{DS} = V_{DD} - R_D i_D = 12$ V

B) $v_i = 6$ V. From GS-KVL, we get $v_{GS} = v_i = 6$ V. Since $v_{GS} = 6 > V_t = 2$, NMOS is not in cut-off. Assume NMOS in active region. Then:

$$i_D = K(v_{GS} - V_t)^2 = 0.25 \times 10^{-3}(6-2)^2 = 4 \text{ mA}$$

DS-KVL: $v_{DS} = V_{DD} - R_D i_D = 12 - 4 \times 10^3 \times 10^{-3} = 8$ V

Since $v_{DS} = 8 > v_{GS} - V_t = 2$, NMOS is indeed in active region and $i_D = 4$ mA and $v_o = v_{DS} = 8$ V.

C) $v_i = 12$ V. From GS-KVL, we get $v_{GS} = 12$ V. Since $v_{GS} > V_t$, NMOS is not in cut-off. Assume NMOS in active region. Then:

$$i_D = K(v_{GS} - V_t)^2 = 0.25 \times 10^{-3}(12-2)^2 = 25 \text{ mA}$$

DS-KVL: $v_{DS} = V_{DD} - R_D i_D = 12 - 25 \times 10^3 \times 10^{-3} = -13$ V

Since $v_{DS} = -13 < v_{GS} - V_t = 12 - 2 = 10$, NMOS is NOT in active region.

Assume NMOS in ohmic region. Then:

$$i_D = K[2v_{DS}(v_{GS} - V_t) - v_{DS}^2] = 0.25 \times 10^{-3}[2v_{DS}(12-2) - v_{DS}^2]$$
$$i_D = 0.25 \times 10^{-3}[20v_{DS} - v_{DS}^2]$$

Substituting for i_D in DS-KVL, we get:

DS-KVL: $V_{DD} = R_D i_D + v_{DS} \quad \rightarrow \quad 12 = 10^3 \times 0.25 \times 10^{-3}[20v_{DS} - v_{DS}^2] + v_{DS}$

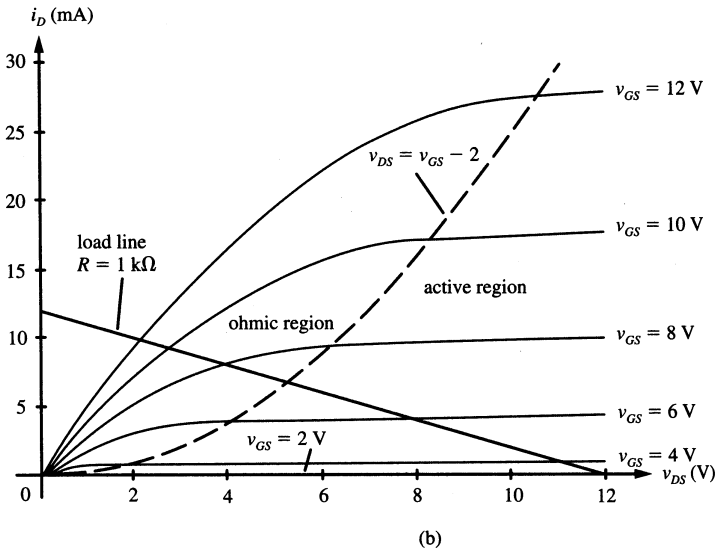
$$v_{DS}^2 - 24v_{DS} + 48 = 0$$

This is a quadratic equation in v_{DS}. The two roots are: $v_{DS} = 2.2$ V and $v_{DS} = 21.8$ V. The second root is not physical as the circuit is powered by a 12 V supply. Therefore, $v_{DS} = 2.2$ V. As $v_{DS} = 2.2 < v_{GS} - V_t = 10$, NMOS is indeed in ohmic region with $v_o = v_{DS} = 2.2$ V and

$$\text{DS-KVL:} \qquad v_{DS} = V_{DD} - R_D i_D \quad \rightarrow \quad i_D = \frac{12 - 2.2}{1,000} = 9.8 \text{ mA}$$

Load Line: Operation of NMOS circuits can be better understood using the concept of load line. Similar to BJT, load line is basically the line representing DS-KVL in i_D versus v_{DS} space. Load line of the example circuit is shown here.

Exercise: Mark the Q-points of the previous example for $v_i = 0$, 6, and 12 V on the load line figure below.



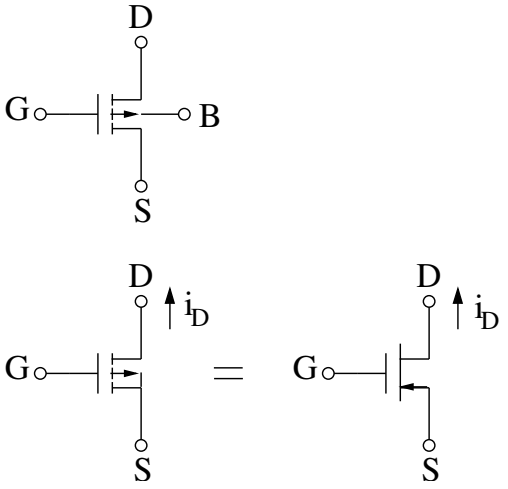
(b)

Body Effect

In deriving NMOS (and other MOS) i_D versus v_{DS} characteristics, we had assumed that the body and source are connected. This is not possible in an integrated chip which has a common body and a large number of MOS devices (connection of body to source for all devices means that all sources are connected). The common practice is to attach the body of the chip to the smallest voltage available from the power supply (zero or negative). In this case, the *pn* junction between the body and source of all devices will be reversed biased. The impact of this to lower threshold voltage for the MOS devices slightly and its called the body effect. Body effect can degrade device performance. For analysis here, we will assume that body effect is negligible.

p-Channel Enhancement-Type MOSFET (PMOS)

The physical structure of a PMOS is identical to a NMOS except that the semiconductor types are interchanged, *i.e.,* body and gate are made of n-type material and source and drain are made of p-type material and a p-type channel is formed. As the sign of the charge carriers are reversed, all voltages and currents in a PMOS are reversed. By convention, the drain current is flowing out of the drain as is shown. With this, all of the NMOS discussion above applies to PMOS as long as we multiply all voltages by a minus sign:



Cut-off: $v_{GS} > V_t, \quad i_D = 0 \quad$ for any v_{DS}

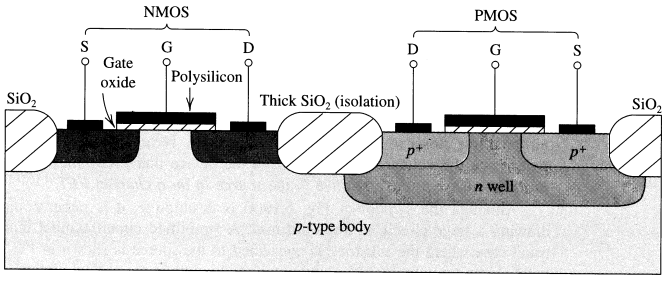
Ohmic: $v_{GS} < V_t, \quad i_D = K[2v_{DS}(v_{GS} - V_t) - v_{DS}^2] \quad$ for $v_{DS} > v_{GS} - V_t$

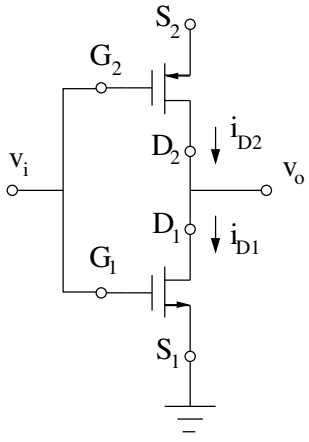
Active: $v_{GS} < V_t, \quad i_D = K(v_{GS} - V_t)^2 \quad$ for $v_{DS} < v_{GS} - V_t$

Note that V_t is negative for a PMOS.

Complementary MOS (CMOS)

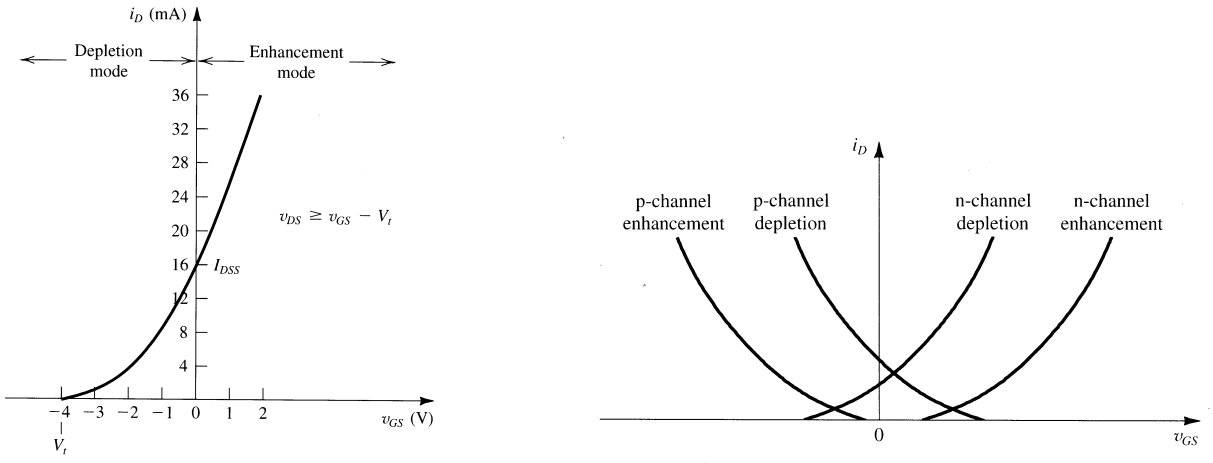
Complementary MOS technology employs MOS transistors of both polarites as is shown below. CMOS devices are more difficult to fabricate than NMOS, but many more powerful circuits are possible with CMOS configuration. As such, most of MOS circuits today employ CMOS configuration and CMOS technology is rapildy taking over many applications that were possible only with bipolar devices a few years ago.

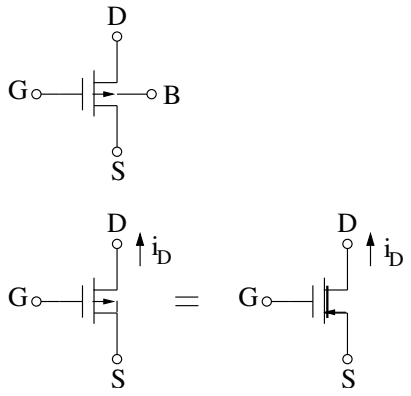




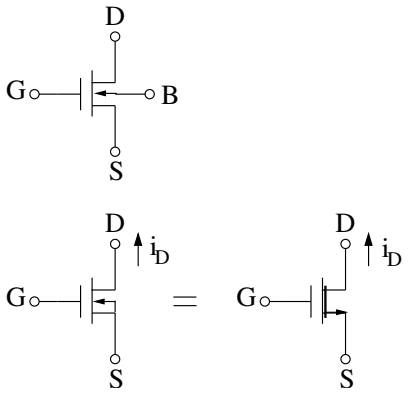
Depletion-Type MOSFET

The depletion-type MOSFET has a structure similar to the enhancement-type MOSFET with only one important difference; depletion-type MOSFET has a physically implanted channel. Thus, a n-type depletion-type MOSFET has already a n-type channel between drain and source. When a voltage v_{DS} is applied to the device, a current $i_D = I_{DSS}$ flows even for $v_{GS} = 0$. (Show $I_{DDS} = KV_t^2$.) Similar to NMOS, if v_{GS} is increased, the channel become wider and i_D increases. However, in a n-type depletion-type MOSFET, a negative v_{GS} can also be applied to the device, which makes the channel smaller and reduces i_D. As such, negative v_{GS} "depletes" the channels from n-type carriers leading to the name depletion-type MOSFET. If v_{GS} is reduced further, at some threshold value V_t (which is negative), the channel disappears and $i_D = 0$, as is seen in the figure. It should be obvious that a depletion-type MOSFET can be operated either in enhancement mode or in depletion mode. p-type depletion MOSFET operate similarly to p-type enhancement MOSFET expect that $V_t > 0$ for depletion type and $V_t < 0$ for the enhancement type. Figure below shows i_D versus v_{GS} of four types of MOSFET devices in the active region. Circuit symbols for depletion-type MOSFET devices are also shown.





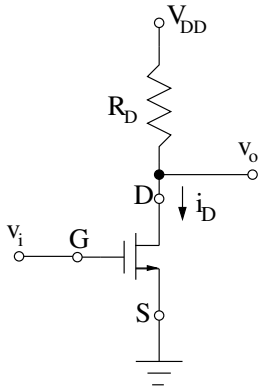
n-type Depletion MOSFET



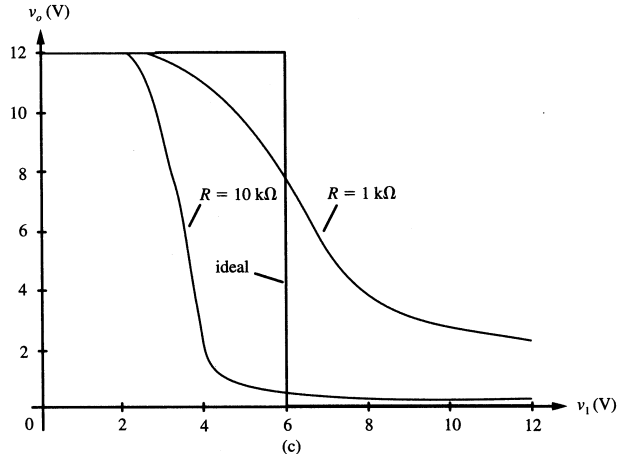
p-type Depletion MOSFET

NMOS Inverter and Switch

The basic NMOS inverter circuit is shown; the circuit is very similar to a BJT inverter. This circuit was solved in page 65 for $V_{DD} = 12$ and $R_D = 1$ kΩ. We found that if $v_i = 0$ (in fact $v_i < V_t$), NMOS will be in cut-off with $i_D = 0$ and $v_o = V_{DD}$. When $v_i = 12$ V, NMOS will be in ohmic region with $i_D = 10$ mA and $v_{DS} = 2.2$ V. Therefore, the circuit is an inverter gate. It can also be used as switch.



There are some important difference between NMOS and BJT inverter gates. First, BJT needs a resistor R_B. This resistor "converts" the input voltages into an i_B and keep $v_{BE} \approx v_\gamma$. NMOS does not need a resistor between the source and the input voltage as $i_G = 0$ and $v_i = v_{GS}$ can be directly applied to the gate. Second, if the input voltage is "high," the BJT will go into saturation with $v_o = v_{CE} = V_{sat} = 0.2$ V. In the NMOS gate, if the input voltage is "high," NMOS is in the ohmic region. In this case, v_{DS} can have any value between 0 and $v_{GS} - V_t$; the value of $v_o = v_{DS}$ is set by the value of the resistor R_D. This effect is shown in the transfer function of the inverter gate for two different values of R_D.



Exercise: Compute v_o for the above circuit with $V_{DD} = 12$ and $R_D = 10$ kΩ when $v_i = 12$ V.

CMOS Inverter

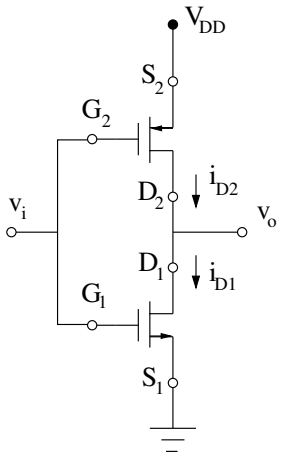
The CMOS inverter, the building block of CMOS logic gates, is shown below. The "low" and "high"states for this circuit correspond to 0 and V_{DD}, respectively. CMOS gates are built on the same chip such that both NMOS and PMOS have the same threshold voltage $V_{tn} = \bar{V}_t$, $V_{tp} = -\bar{V}_t$ and same K (one needs different channel length and width to get the same K for PMOS and NMOS). In this case, the CMOS will have a "symmetric" transfer characteristics, *i.e.,* $v_o = 0.5V_{DD}$ when $v_i = 0.5V_{DD}$ as is shown below. This circuit works only if $V_{DD} > 2V_t$ by a comfortable margin.

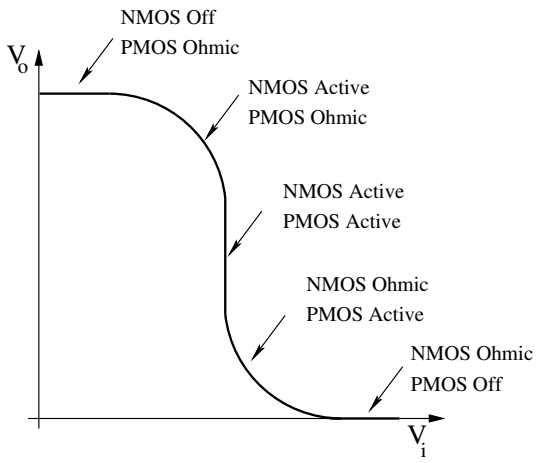
$\underline{v_i = 0}$ Since $v_{GS1} = v_i = 0 < \bar{V}_t$, NMOS will be in cut-off. Therefore, $i_{D1} = 0$. But $v_{GS2} = v_i - V_{DD} = -V_{DD} < -\bar{V}_t$, thus, PMOS will be ON. Recall i_D versus v_{DS} characteristic curves of a NMOS (page 63), each labeled with a values of V_{GS}. As $v_{GS2} = V_{DD}$ is a constant, operating point of PMOS will be on one of the curves. But $i_{D1} = i_{D2} = 0$ and the only point that satisfies this condition is $v_{DS2} = 0$. Note that PMOS is NOT in cut-off, it is in ohmic region and acts like a resistor. $v_{DS2} = 0$ because $i_{D2} = 0$. Output voltage can now be found by KVL: $v_o = V_{DD} - v_{DS2} = V_{DD}$. So, when $v_i = 0$, $v_o = V_{DD}$.

$\underline{v_i = V_{DD}}$ Since $v_{GS1} = v_i = V_{DD} > \bar{V}_t$, NMOS will be ON. But since $v_{GS2} = v_i - V_{DD} = 0 > -\bar{V}_t$, PMOS will be in cut-off and $i_{D2} = 0$. Since $i_{D1} = i_{D2}$, $i_{D1} = 0$. Since NMOS is on and $i_{D1} = 0$, NMOS should be in ohmic region with $v_{DS1} = 0$. Then $v_o = v_{DS1} = 0$. So, when $v_i = V_{DD}$, $v_o = 0$

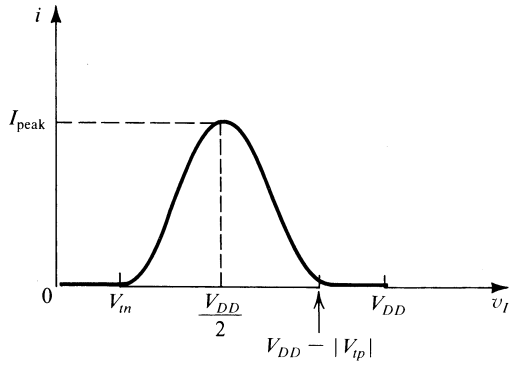
$\underline{v_i = 0.5V_{DD}}$ In this case, $v_{GS1} = v_i = 0.5V_{DD}$ and $v_{GS2} = v_i - V_{DD} = -0.5V_{DD}$. Since, $v_{GS1} > \bar{V}_t$ and $v_{GS2} < -\bar{V}_t$, both transistors will be ON. Furthermore, as transistors have same threshold voltage, same K, $i_{D1} = i_{D2}$, and $v_{GS1} = |v_{GS2}|$, both transistor will be in the same state (either ohmic or active) and will have identical v_{DS}: $v_{DS1} = -v_{DS2}$. Since $v_{DS1} - v_{DS2} = V_{DD}$, then $v_{DS1} = 0.5V_{DD}$, $v_{DS2} = -0.5V_{DD}$, and $v_o = 0.5V_{DD}$. Note that since $V_{DD} > 2V_t$, both transistors are in the active region.

The transfer function of the CMOS inverter is shown below.





CMOS inverter has many advantages compared to the NMOS inverter. Its transfer function is much closer to an ideal inverter transfer function. The "low" and high" states are clearly defined (low state of NMOS depended on the value of R_D). It does not include any resistors and thus takes less area on the chip. Lastly, $i_{D1} = i_{D2} = 0$ for both cases of output being low or high. This means that the gate consumes very little power (theoretically zero) in either state. A non-zero $i_{D1} = i_{D2}$, however, flows in the circuit when the gate is transitioning from one state to another as is seen in the figure.



The maximum value of i_D that flows through the gate during the transition can be easily calculated as this maximum current flows when $v_i = 0.5V_{DD}$ and $v_o = 0.5V_{DD}$.

For example, consider the CMOS inverter with $V_{DD} = 12$ V, $\bar{V}_t = 2$ V, and $K = 0.25$ mA/V^2. Maximum i_D flows when $v_i = v_{GS1} = 0.5V_{DD} = 6$ V. At this point, $v_{DS1} = v_o = 0.5V_{DD} = 6$ V. As, $v_{DS1} = 6 > v_{GS1} - v_t = 4$ V, NMOS is in active regime. Then:

$$i_{D1} = K(v_{GS1} - V_t)^2 = 0.25 \times 10^{-3}(6-2)^2 = 4 \text{ mA}$$

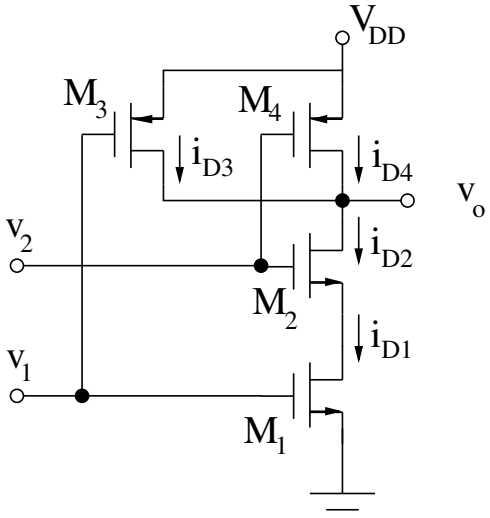
CMOS NAND Gate

As mentioned before CMOS logic gates have "low" and "high" states of 0 and V_{DD}, respectively. We need to consider all possible cases to show that this a NAND gate. To start, we can several general observations:

1) by KVL $v_{GS1} = v_1$, $v_{GS2} = v_2 - v_{DS1}$, $v_{GS3} = v_1 - V_{DD}$, and $v_{GS4} = v_2 - V_{DD}$,

2) by KCL $i_{D1} = i_{D2} = i_{D3} + i_{D4}$

3) by KVL $v_o = v_{DS1} + v_{DS2} = V_{DD} + v_{DS3}$ and $v_{DS3} = v_{DS4}$.



The procedure to solve/analyze CMOS gates are as follows: // A) For a given set of input voltages, use KVLs (similar to 1 above) to find v_{GS} of all transistors and find which ones are ON or OFF.// B) Set $i_D = 0$ for all transistors that are OFF. Use KCLs (similar to 2 above) to find i_D for other transistors.// C) Look for transistors that are ON and have $i_D = 0$. These transistors have to be in Ohmic region with $v_{DS} = 0$. D) Use KVLs (similar to 3 above) to find v_o based on v_{DS}.

$\underline{v_1 = 0,\ v_2 = 0}$

We first find v_{GS} and state of all transistors by using KVLS in no. 1 above.

$$
\begin{array}{lcll}
v_{GS1} = v_1 = 0 < \bar{V}_t & \rightarrow & \text{M1 is OFF} & \rightarrow \quad i_{D1} = 0 \\
v_{GS2} = v_2 - v_{DS1} = -v_{DS1} & \rightarrow & \text{M2 is ?} & \\
v_{GS3} = v_1 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M3 is ON} & \\
v_{GS4} = v_2 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M4 is ON} &
\end{array}
$$

Since $i_{D1} = 0$, by KCLs in no. 2 above , $i_{D2} = i_{D1} = 0$ and $i_{D3} + i_{D4} = i_{D1} = 0$. Since $i_D \geq 0$ for both PMOS and NMOS, the last equation can be only satisfied if $i_{D3} = i_{D4} = 0$. We add the value of i_D to the table above and look for transistors that are ON and have $i_D = 0$. These transistors have to be in Ohmic region with $v_{DS} = 0$.

$$
\begin{array}{lcllll}
v_{GS1} = v_1 = 0 < \bar{V}_t & \rightarrow & \text{M1 is OFF} & \rightarrow & i_{D1} = 0 & \\
v_{GS2} = v_2 - v_{DS1} = -v_{DS1} & \rightarrow & \text{M2 is ?} & & i_{D2} = 0 & \\
v_{GS3} = v_1 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M3 is ON} & & i_{D3} = 0 & \rightarrow \quad v_{DS3} = 0 \\
v_{GS4} = v_2 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M4 is ON} & & i_{D4} = 0 & \rightarrow \quad v_{DS4} = 0
\end{array}
$$

Finally, from KVLs in no. 3. above, we have $v_o = V_{DD} + v_{DS3} = V_{DD}$. So, when both inputs are low, the output is HIGH.

If needed, we can go back and find the state of M2. Assume M2 is ON. This requires $v_{GS2} > \bar{V}_t$. Since $i_{D2} = 0$ and M2 is ON, $v_{DS2} = 0$ (M2 in Ohmic). From KVLs in no. 3. above, we have $v_o = v_{DS1} + v_{DS2} = V_{DD}$. This gives $v_{DS1} = V_{DD}$ and $v_{GS2} = v_2 - v_{DS1} = V_{DD} - V_{DD} = 0 < \bar{V}_t$, a contradiction of our assumption of M2 being ON. Therefore, M2 should be OFF.

$\underline{v_1 = 0,\ v_2 = V_{DD}}$

We first find v_{GS} and state of all transistors by using KVLS in no. 1 above.

$$\begin{array}{lcll} v_{GS1} = v_1 = 0 < \bar{V}_t & \rightarrow & \text{M1 is OFF} & \rightarrow \quad i_{D1} = 0 \\ v_{GS2} = v_2 - v_{DS1} = V_{DD} - v_{DS1} & \rightarrow & \text{M2 is ?} & \\ v_{GS3} = v_1 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M3 is ON} & \\ v_{GS4} = v_2 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M4 is OFF} & \rightarrow \quad i_{D4} = 0 \end{array}$$

Since $i_{D1} = 0$, by KCLs in no. 2 above , $i_{D2} = i_{D1} = 0$. Also, $i_{D3} + i_{D4} = i_{D1} = 0$ leading to $i_{D4} = 0$. We add the value of i_D to the table above and look for transistors that are ON and have $i_D = 0$. These transistors have to be in Ohmic region with $v_{DS} = 0$.

$$\begin{array}{lcllll} v_{GS1} = v_1 = 0 < \bar{V}_t & \rightarrow & \text{M1 is OFF} & \rightarrow \quad i_{D1} = 0 & & \\ v_{GS2} = v_2 - v_{DS1} = V_{DD} - v_{DS1} & \rightarrow & \text{M2 is ?} & \quad\ \ i_{D2} = 0 & & \\ v_{GS3} = v_1 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M3 is ON} & \quad\ \ i_{D3} = 0 & \rightarrow & v_{DS3} = 0 \\ v_{GS4} = v_2 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M4 is OFF} & \rightarrow \quad i_{D4} = 0 & & \end{array}$$

Finally, from KVLs in no. 3. above, we have $v_o = V_{DD} + v_{DS3} = V_{DD}$. So, when v_1 is LOW and v_2 is HIGH, the output is HIGH.

We can go back and find the state of M2. We will find M2 to be ON (left for students).

$\underline{v_1 = V_{DD},\ v_2 = 0}$

We first find v_{GS} and state of all transistors by using KVLS in no. 1 above.

$$\begin{array}{lcll} v_{GS1} = v_1 = V_{DD} > \bar{V}_t & \rightarrow & \text{M1 is ON} & \\ v_{GS2} = v_2 - v_{DS1} = -v_{DS1} < \bar{V}_t & \rightarrow & \text{M2 is OFF} & \rightarrow \quad i_{D2} = 0 \\ v_{GS3} = v_1 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M3 is OFF} & \rightarrow \quad i_{D3} = 0 \\ v_{GS4} = v_2 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M4 is ON} & \end{array}$$

In the above, we used the fact that $v_{DS1} \geq 0$. Since $i_{D2} = 0$, by KCLs in no. 2 above , $i_{D1} = i_{D2} = 0$. Also, $i_{D3} + i_{D4} = i_{D2} = 0$ leading to $i_{D4} = 0$. We add the value of i_D to the

table above and look for transistors that are ON and have $i_D = 0$. These transistors have to be in Ohmic region with $v_{DS} = 0$.

$$
\begin{array}{lllllll}
v_{GS1} = v_1 = V_{DD} > \bar{V}_t & \rightarrow & \text{M1 is ON} & & & & \\
v_{GS2} = v_2 - v_{DS1} = -v_{DS1} < \bar{V}_t & \rightarrow & \text{M2 is OFF} & \rightarrow & i_{D2} = 0 & \rightarrow & v_{DS2} = 0 \\
v_{GS3} = v_1 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M3 is OFF} & \rightarrow & i_{D3} = 0 & \rightarrow & v_{DS3} = 0 \\
v_{GS4} = v_2 - V_{DD} = -V_{DD} < -\bar{V}_t & \rightarrow & \text{M4 is ON} & & & &
\end{array}
$$

Finally, from KVLs in no. 3. above, we have $v_o = V_{DD} + v_{DS3} = V_{DD}$. So, when v_1 is HIGH and v_2 is LOW, the output is HIGH.

$\underline{v_1 = V_{DD},\, v_2 = V_{DD}}$

We first find v_{GS} and state of all transistors by using KVLS in no. 1 above.

$$
\begin{array}{lllll}
v_{GS1} = v_1 = V_{DD} > \bar{V}_t & \rightarrow & \text{M1 is ON} & & \\
v_{GS2} = v_2 - v_{DS1} = V_{DD} - v_{DS1} & \rightarrow & \text{M2 is ?} & & \\
v_{GS3} = v_1 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M3 is OFF} & \rightarrow & i_{D3} = 0 \\
v_{GS4} = v_2 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M4 is OFF} & \rightarrow & i_{D4} = 0
\end{array}
$$

From KCLs in no. 2 above , $i_{D2} = i_{D1} = i_{D3} + i_{D4} = 0$. We add the value of i_D to the table above and look for transistors that are ON and have $i_D = 0$. These transistors have to be in Ohmic region with $v_{DS} = 0$.

$$
\begin{array}{llllllll}
v_{GS1} = v_1 = V_{DD} > \bar{V}_t & \rightarrow & \text{M1 is ON} & & i_{D1} = 0 & \rightarrow & v_{DS1} = 0 \\
v_{GS2} = v_2 - v_{DS1} = V_{DD} - v_{DS1} = V_{DD} > \bar{V}_t & \rightarrow & \text{M2 is ON} & & i_{D2} = 0 & \rightarrow & v_{DS2} = 0 \\
v_{GS3} = v_1 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M3 is OFF} & \rightarrow & i_{D3} = 0 & & \\
v_{GS4} = v_2 - V_{DD} = 0 > -\bar{V}_t & \rightarrow & \text{M4 is OFF} & \rightarrow & i_{D4} = 0 & &
\end{array}
$$

In the above, we used $v_{DS1} = 0$ to find the state of M2 leading to $v_{DS2} = 0$. Finally, from KVLs in no. 3. above, we have $v_o = v_{DS1} + v_{DS1} = 0$. So, when v_1 is HIGH and v_2 is HIGH, the output is LOW.

From the "truth table," the output of this gate is LOW only if both input states are HIGH. Therefore, this is a NAND gate.

CMOS NOR Gate

Exercise: Show that this is a NOR gate.