The average horizontal electric field across the channel in pinch-off does not depend on the D-S voltage but instead on the voltage across the channel, which is V_{GS} - V_t .

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left[2(V_{GS}-V_t)V_{DS}-V_{DS}^2\right] = \frac{k'}{2}\frac{W}{L}\left[2(V_{GS}-V_t)V_{DS}-V_{DS}^2\right]$$

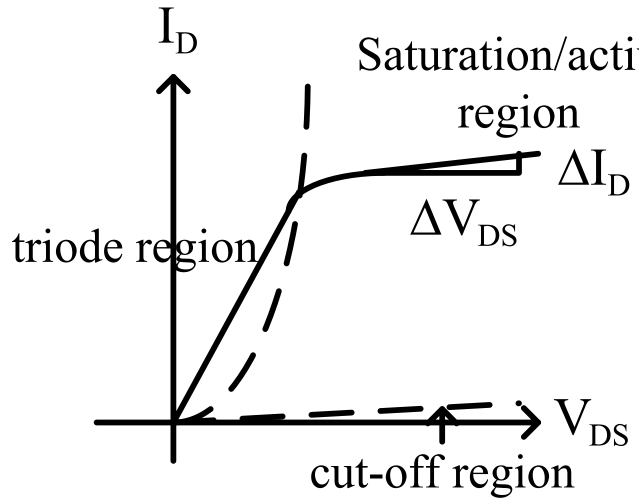
is not valid if $V_{DS} > V_{GS} - V_t$.

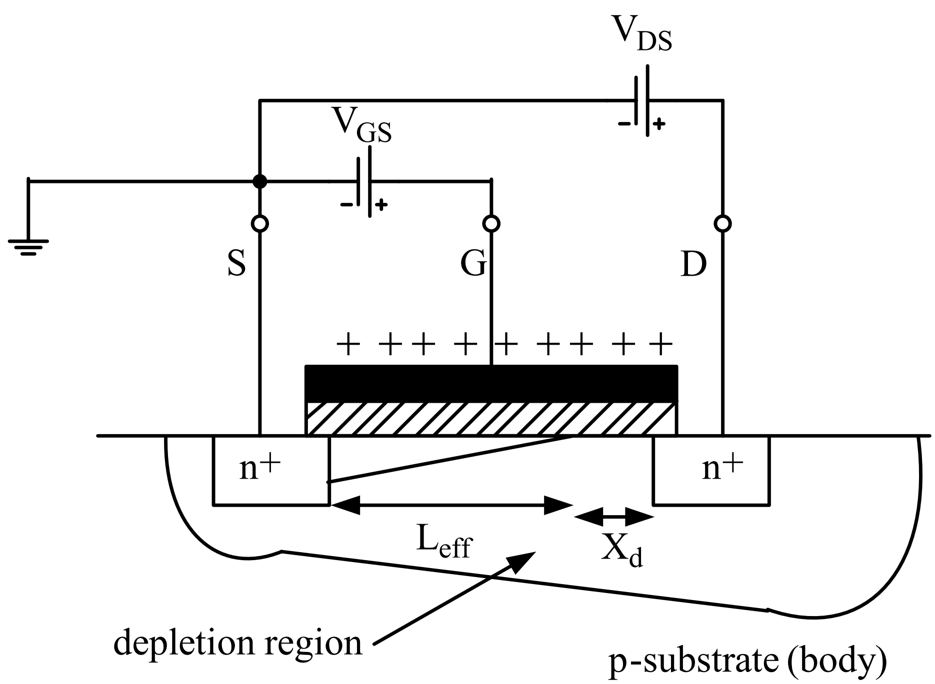
In the pinch-off region where $V_{DS} = V_{GS} - V_t$,

$$\begin{aligned} I_D &= \frac{k'}{2}\frac{W}{L}\left[2(V_{GS}-V_t)(V_{GS}-V_t)-(V_{GS}-V_t)^2\right] \\ &= \frac{k'}{2}\frac{W}{L}(V_{GS}-V_t)^2 \qquad (1.157) \end{aligned}$$

Hence, I_D is independent of V_{DS} in the pinch-off region. In practice, however, I_D in the pinch-off region varies slightly as the drain voltage is varied. This effect is due to the presence of a depletion region between the physical pinch-off point in the channel at the D end and the region itself.

Output characteristic
of NMOS





X_d = depletion layer width between physical pinch-off point in the channel at the D-end and the D region itself.

$L_{eff} = L - X_d$ = effective channel length (1.158)

$$I_D = \frac{k'}{2}\frac{W}{L_{eff}}\left(V_{GS} - V_t\right)^2 \qquad (1.159)$$

Because X_d (and thus L_{eff}) are functions of V_{DS} in the pinch-off region, I_D varies with V_{DS}. This effect is called the channel-length modulation.

$$\frac{\partial I_D}{\partial V_{DS}} = -\frac{k'}{2}\frac{W}{L_{eff}^2}\left(V_{GS} - V_t\right)^2 \frac{dL_{eff}}{dV_{DS}}$$

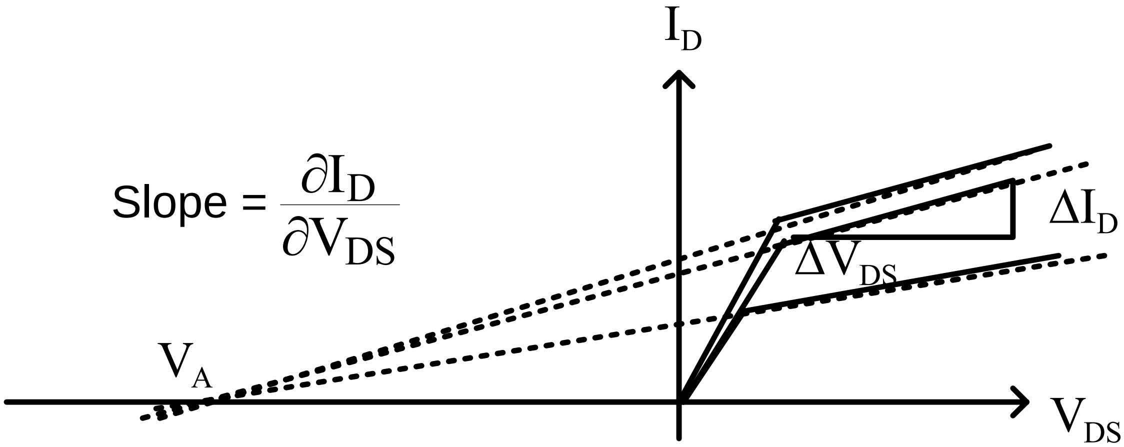
$$\frac{\partial I_D}{\partial V_{DS}} = -\frac{I_D}{L_{eff}}\frac{dL_{eff}}{dV_{DS}}$$

From (1.158), i.e. $L_{eff} = L - X_d$,

$$\frac{dL_{eff}}{dX_d} = -1$$

$$dL_{eff} = -dX_d$$

$$\frac{\partial I_D}{\partial V_{DS}} = -\frac{I_D}{L_{eff}}\left[-\frac{dX_d}{dV_{DS}}\right] = \frac{I_D}{L_{eff}}\left[\frac{dX_d}{dV_{DS}}\right]$$



$$V_A = -\frac{I_D}{\frac{\partial I_D}{\partial V_{DS}}} = \frac{I_D L_{eff}}{I_D \frac{dX_D}{dV_{DS}}} = L_{eff}\left(\frac{dX_D}{dV_{DS}}\right)^{-1} \qquad (1.163)$$

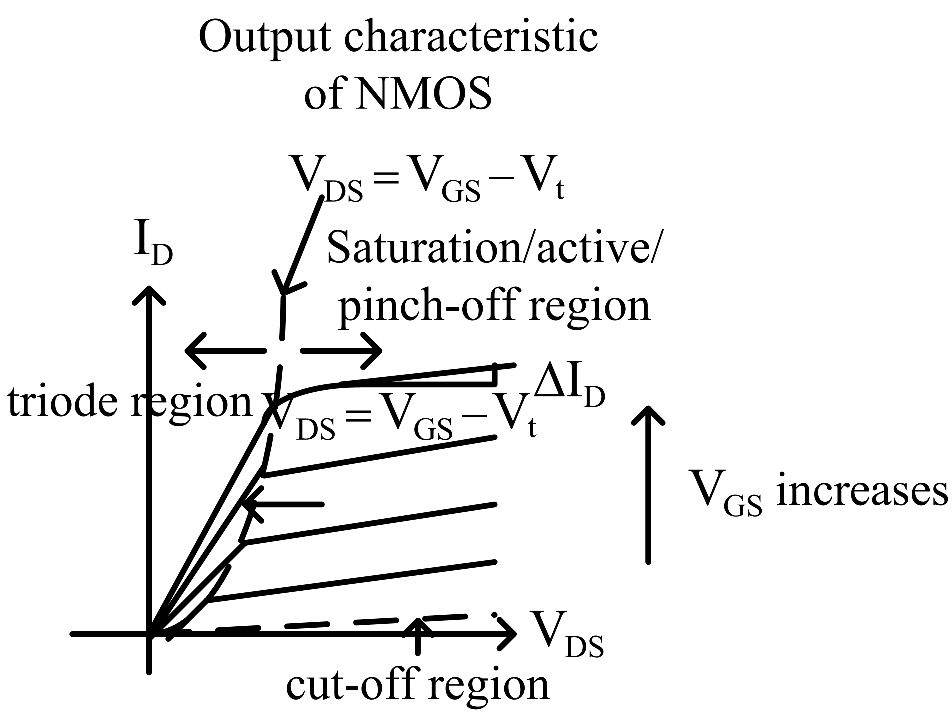
For MOSFETs, a commonly used parameter for the characterization of channel length modulation is:

$$\lambda = \frac{1}{V_A} \qquad (1.164)$$

The large-signal properties of the transistor can be approximated by assuming that λ and V_A are constants independent of the bias conditions. If the channel-length modulation effect is included,

$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2\left(1 + \frac{V_{DS}}{V_A}\right)$$
$$= \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2\left(1 + \lambda V_{DS}\right)$$

From (1.163), i.e. $V_A = L_{eff}\left(\frac{dX_D}{dV_{DS}}\right)^{-1}$ and (1.164), i.e. $\lambda = -\frac{1}{V_A}$, $\lambda \propto -\frac{1}{L_{eff}}$. Typical λ is in the range 0.05 V^{-1} to 0.005 V^{-1}.



$V_{DS} > V_{GS} - V_t \rightarrow$ pinch off / saturation region. In saturation region, output characteristics are almost flat. I_D depends mostly on V_{GS} and only to a small extent on V_{DS}.

If $V_{DS} < V_{GS} - V_t$, the device operates in the ohmic or triode region where the device can be modeled as a non-linear voltage controlled resistor connected between D and S.

In this region,

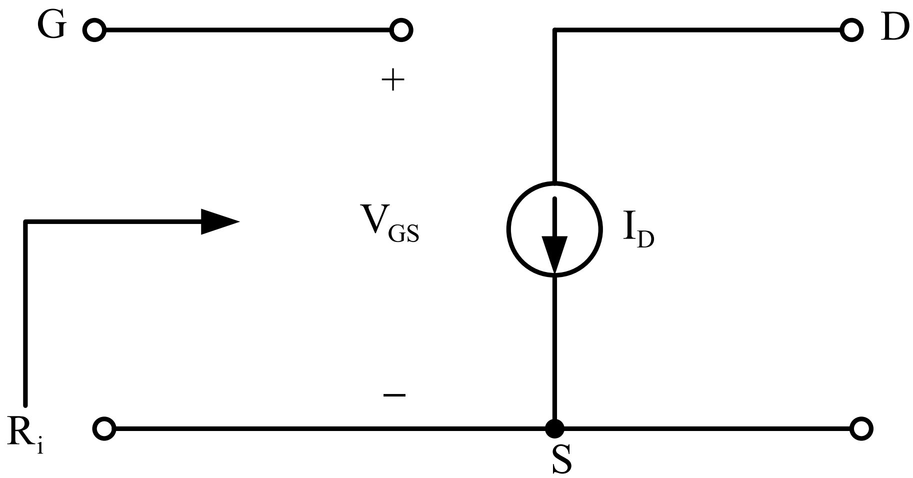
$$I_D = \frac{k'}{2}\frac{W}{L}\left[2\left(V_{GS} - V_t\right)V_{DS} - V_{DS}^2\right] \qquad (1.152)$$

The resistance in this region is non-linear because V_{DS}^2 term in (1.152), causes the resistance to depend on V_{DS}. Since this term is small when V_{DS} is small, the non-linearity is also small:

Triode region is also sometimes called the linear region.

The boundary between the triode and the saturation region occurs when $V_{DS} = (V_{GS} - V_t)$. On this boundary, both (1.152) and (1.157), i.e. $I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$, correctly predict I_D.

Large-signal model for NMOS:



In triode region:

$$I_D = \frac{k'}{2}\frac{W}{L}\left[2\left(V_{GS} - V_t\right)V_{DS} - V_{DS}^2\right]$$

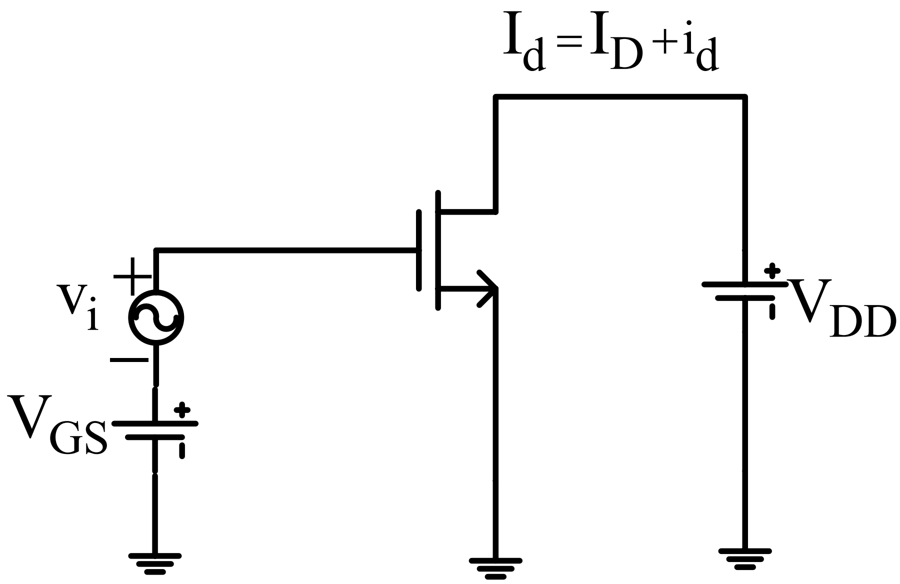
In saturation region:

$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$$

In saturation region with channel-modulation effect included:

$$I_D = \frac{k'}{2}\frac{W}{L_{eff}}\left(V_{GS} - V_t\right)^2$$
$$= \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2\left(1 + \lambda V_{DS}\right)$$

1.6 Small-signal model of MOSFET



Saturation/active mode of operation:

$V_{GS} > V_t$
$V_{DS} > (V_{GS} - V_t)$

V_{GS} and V_{DD} are the bias voltages, producing I_D.
v_i is the small-signal input voltage producing i_d.

1.6.1 Transconductance

$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2\left(1 + \lambda V_{DS}\right)$$
$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{k'}{2}\frac{W}{L}2\left(V_{GS} - V_t\right)\left(1 + \lambda V_{DS}\right)$$
$$g_m = k'\frac{W}{L}\left(V_{GS} - V_t\right)\left(1 + \lambda V_{DS}\right)$$

If $\lambda V_{DS} << 1$, then

$$\left(V_{GS} - V_t\right) = \sqrt{\frac{I_D}{\frac{k'}{2}\frac{W}{L}}}$$
$$g_m = k'\frac{W}{L}\left(V_{GS} - V_t\right)$$
$$= \sqrt{\frac{k'^2\left(\frac{W}{L}\right)^2 I_D}{\frac{k'}{2}\frac{W}{L}}}$$
$$= \sqrt{2k'\frac{W}{L}I_D}$$

$$g_m = k'\frac{W}{L}\left(V_{GS} - V_t\right)$$
$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$$

$$g_m = k' \frac{W}{L}\left(V_{GS} - V_t\right)$$

$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$$

$V_{OV} = V_{GS} - V_t$

$$\frac{g_m}{I_D} = \frac{k' \frac{W}{L}\left(V_{OV}\right)}{\frac{k'}{2}\frac{W}{L}\left(V_{OV}\right)^2} = \frac{2}{V_{OV}}$$

$V_{OV} \approx$ several hundred mV.

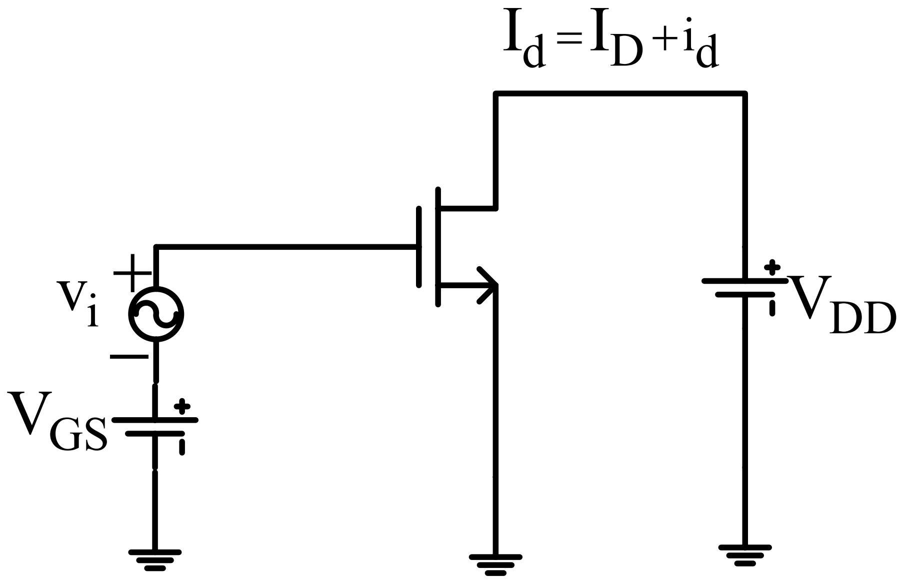
For the BJT: $g_m = \dfrac{I_C}{V_t}$

$$\frac{g_m}{I_C} = \frac{1}{V_t}$$

$V_t = 26$ mV

Hence, $\dfrac{g_m}{I_D} < \dfrac{g_m}{I_C}$

One of the key challenges in MOS analog circuit design is designing high-quality analog circuits with a low transconductance-to-current ratio.



$$I_D = \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$$

$$I_d = \frac{k'}{2}\frac{W}{L}\left(V_{GS} + v_i - V_t\right)^2$$

$$= \frac{k'}{2}\frac{W}{L}\left(\left(V_{GS} - V_t\right)^2 + 2\left(V_{GS} - V_t\right)v_i + v_i^2\right)$$

$$= I_D + \frac{k'}{2}\frac{W}{L}\left(2\left(V_{GS} - V_t\right)v_i + v_i^2\right)$$

$$i_d = I_d - I_D = \frac{k'}{2}\frac{W}{L}\left(2\left(V_{GS} - V_t\right)v_i + v_i^2\right)$$

$$= k'\frac{W(V_{GS} - V_t)v_i}{L}\left(1 + \frac{v_i}{2(V_{GS} - V_t)}\right)$$

$$g_m = k'\frac{W}{L}\left(V_{GS} - V_t\right)$$

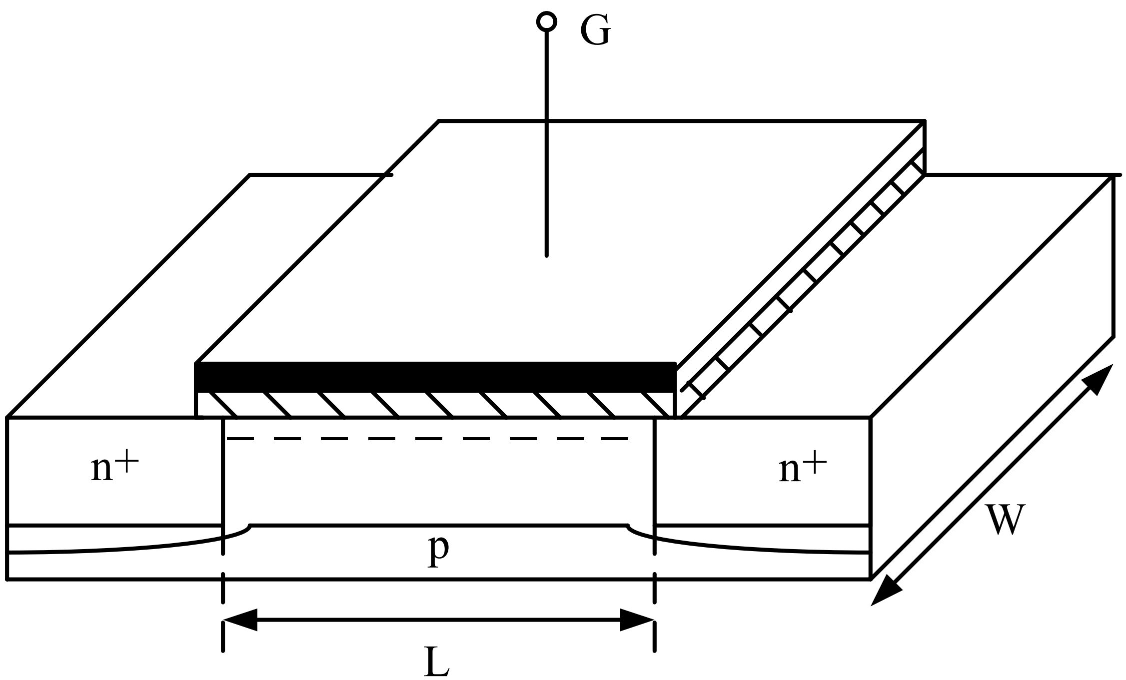
$$i_d = g_m v_i\left(1 + \frac{v_i}{2(V_{GS} - V_t)}\right)$$

$$i_d = g_m v_i \left(1 + \frac{v_i}{2(V_{GS} - V_t)}\right)$$

If $v_i << V_{GS} - V_t$, then $i_d \approx g_m v_i$

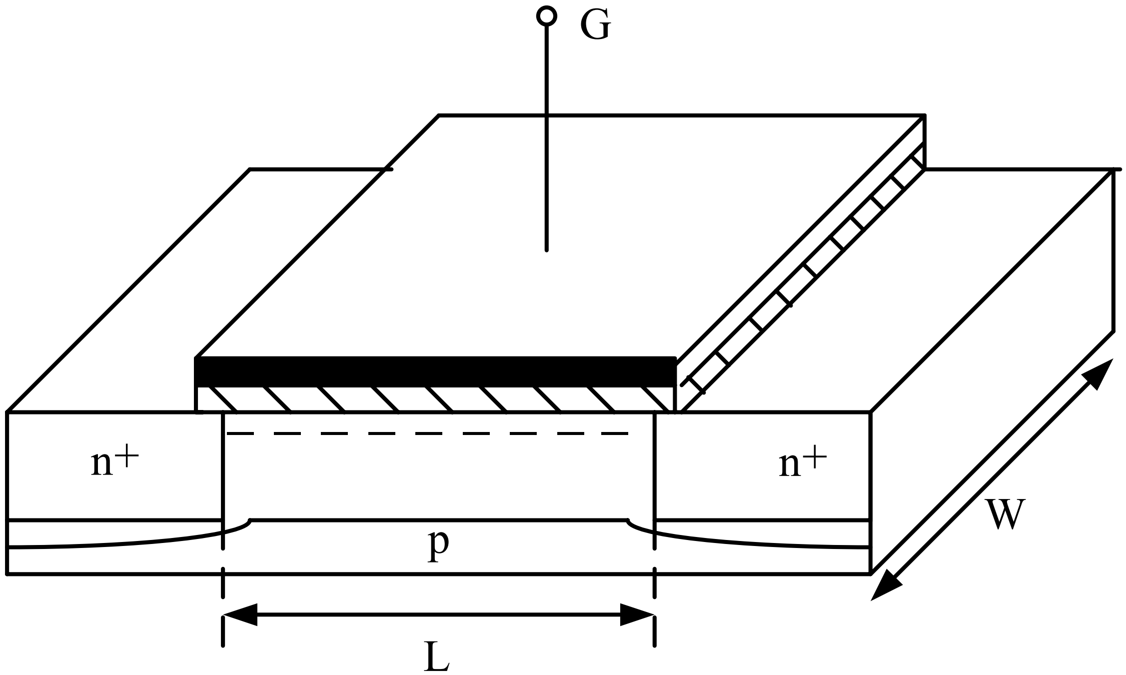
$g_m = k' \frac{W}{L}(V_{GS} - V_t)$ can be used for small-signal analysis if $v_i << V_{OV}$.

1.6.2 Intrinsic Gate-Source and Gate-Drain Capacitance



C_{ox} = oxide capacitance per unit area from G to channel

Hence, total capacitance under the G is $C_{ox}WL$.



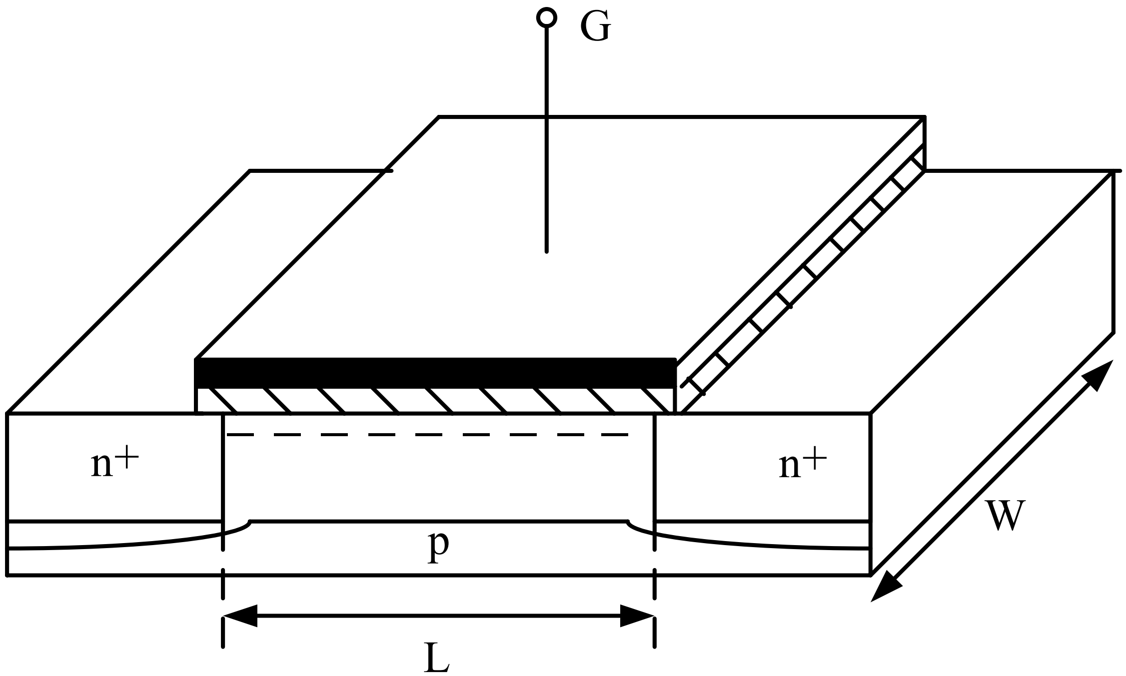
In the <u>triode region</u> of device operation, the channel exists continuously from S to D. The G-channel capacitance is usually lumped into two equal parts at the D and S,

$C_{gs} = C_{gd} = (1/2)C_{ox}WL$

In the <u>saturation or active region</u>, the channel pinches off before reaching the D. The D voltage has little influence on either the channel or the G charge.

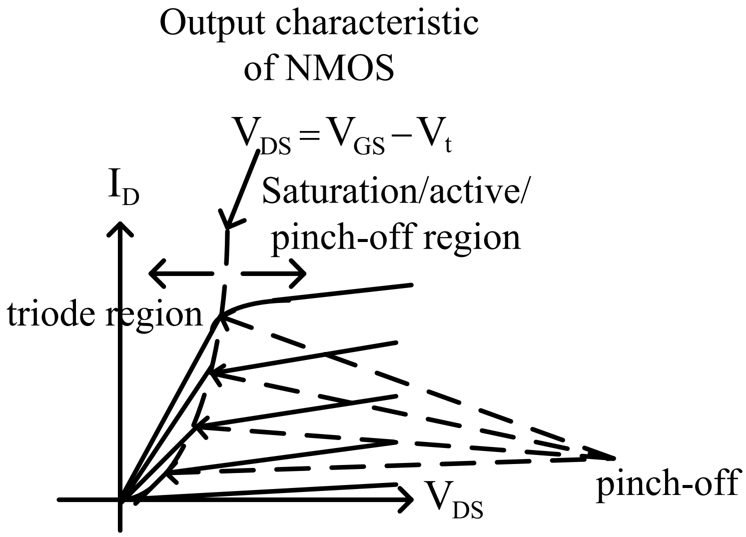
$C_{gd} = 0$
$C_{gs} = (2/3)C_{ox}WL$

1.6.3 Input resistance

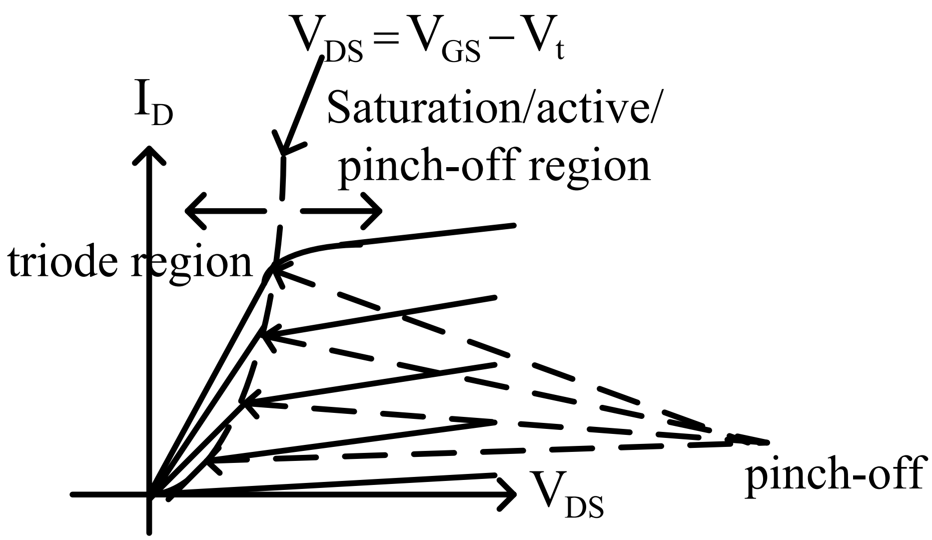


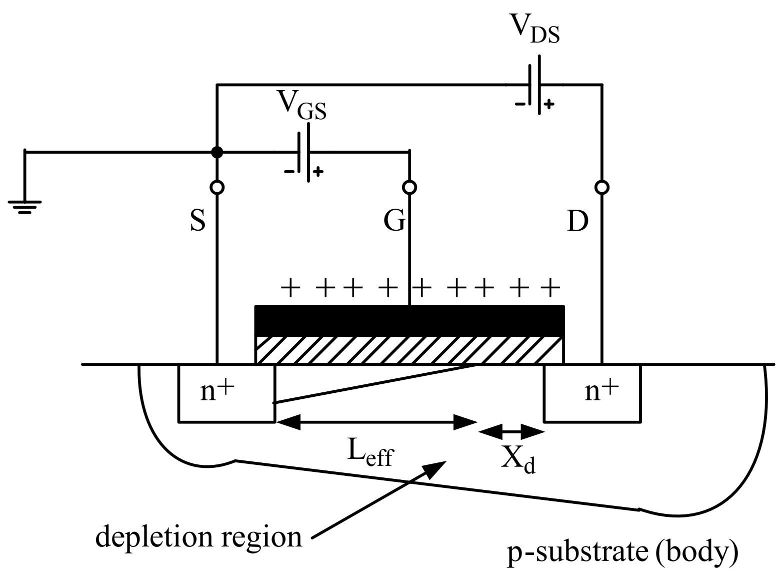
G is insulated from the channel by the SiO_2. At low-frequencies, G current is essentially 0 and input resistance is essentially ∞.

1.6.4 Output resistance

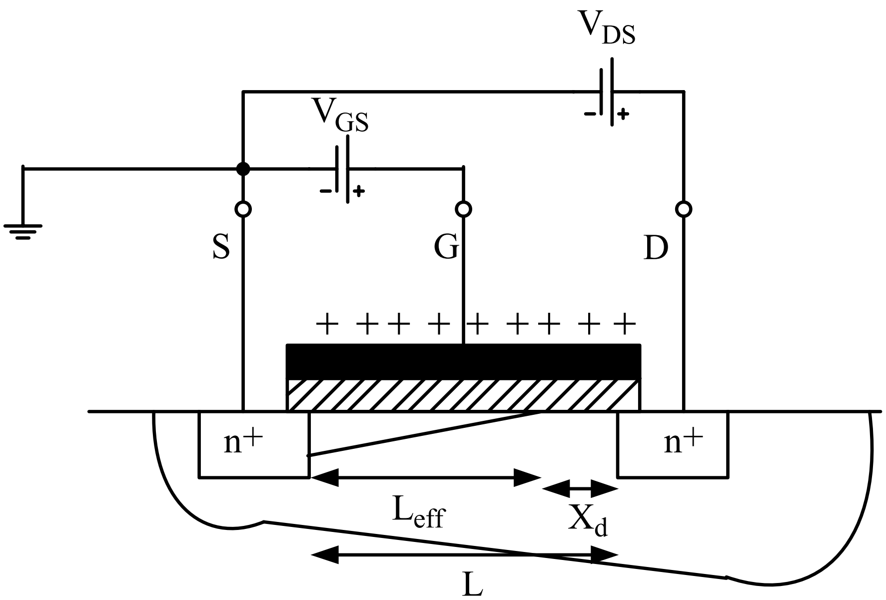


Output characteristic
of NMOS





Increasing V_{DS} will increase the width of the depletion region around the D and reduces the effective channel length of the device in the saturation or active region. This effect is the channel length modulation and causes I_D to increase when V_{DS} is increased.



$$\Delta I_D = \frac{\partial I_D}{\partial V_{DS}} \Delta V_{DS}$$

ΔI_D = change in the D current

ΔV_{DS} = change in the D-S voltage

$L_{eff} = L - X_d$

$$\frac{\partial I_D}{\partial V_{DS}} = \frac{I_D}{L_{eff}} \left[\frac{dX_d}{dV_{DS}} \right]$$

$$V_A = L_{eff} \left(\frac{dX_D}{dV_{DS}} \right)^{-1}$$

Hence, $\frac{\partial I_D}{\partial V_{DS}} = \frac{I_D}{V_A}$

$\lambda = \frac{1}{V_A}$ where V_A is the Early voltage.

$$\frac{\Delta I_D}{\Delta V_{DS}} = \frac{\partial I_D}{\partial V_{DS}} = \frac{I_D}{V_A} = \lambda I_D = \frac{1}{r_o}$$

$$\frac{\Delta I_D}{\Delta V_{DS}} = \frac{\partial I_D}{\partial V_{DS}} = \frac{I_D}{V_A} = \lambda I_D = \frac{1}{r_o}$$

where:

λ = channel length modulation parameter

r_o = small-signal output resistance of the transistor

I_D = drain current without channel length modulation

$$= \frac{k'}{2}\frac{W}{L}\left(V_{GS} - V_t\right)^2$$