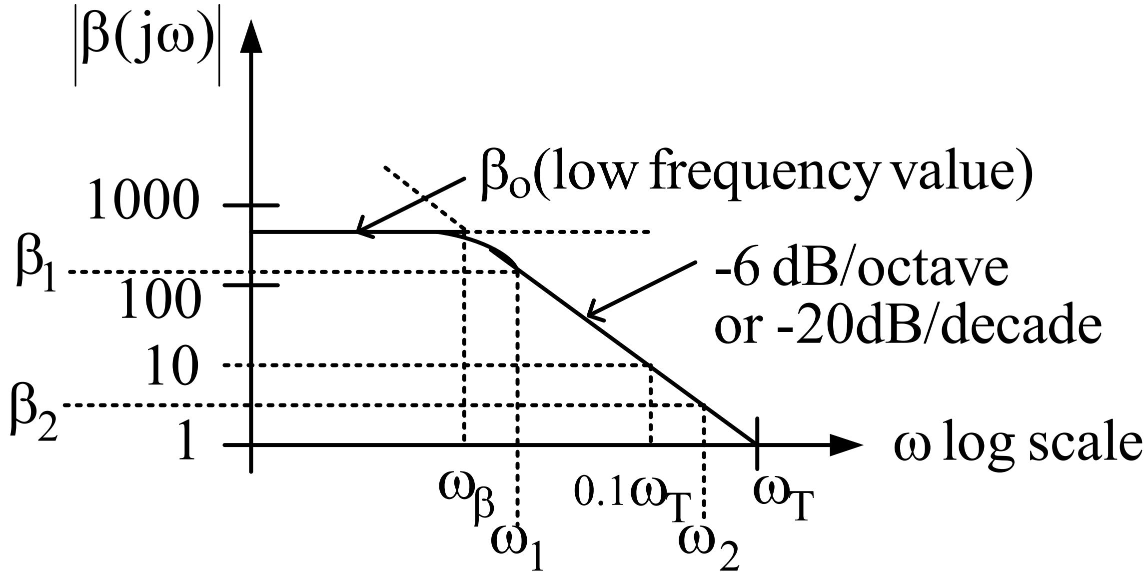
From (1.123), $\beta(j\omega)=\frac{g_m}{j\omega\left(C_\pi+C_\mu\right)}$

At ω_β, $\left|\beta(j\omega)\right|=\frac{\beta_o}{\sqrt{2}}$. Then $\frac{\beta_o}{\sqrt{2}}=\frac{g_m}{\omega_\beta\left(C_\pi+C_\mu\right)}$.

Therefore, $\omega_\beta=\frac{g_m\sqrt{2}}{\beta_o\left(C_\pi+C_\mu\right)}=\frac{g_m\sqrt{2}}{\beta_o\left(C_\pi+C_\mu\right)}=\frac{\omega_T\sqrt{2}}{\beta_o}$



At a high frequency (where the gain is in the -6 dB/octave slope region):

$\left|\beta(j\omega_x)\right|=\frac{g_m}{\omega_x\left(C_\pi+C_\mu\right)}=\frac{\omega_T}{\omega_x}$

Practically, ω_T might be a very large frequency and deviations from the ideal behavior tend to occur as $|\beta(j\omega)|$ approaches unity. So, in order to determine ω_T, the practice is to measure $|\beta(j\omega)|$ at some lower frequency, ω_x, where the $|\beta(j\omega_x)|$ is still falling at 6 dB/octave and using the expression:

$$\omega_T = \omega_x |\beta(j\omega_x)| \tag{1.127}$$

$|\beta(j\omega_x)|$ is typically measured at some frequency where its magnitude is about 5 or 10, and then using (1.127) to determine ω_T.

Time constant:

$$\tau_T = \frac{1}{\omega_T} = \frac{C_\pi + C_\mu}{g_m} = \frac{C_b + C_{je}}{g_m} + \frac{C_\mu}{g_m} \tag{1.128}$$

$$\tau_T = \frac{\tau_F g_m}{g_m} + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m}$$

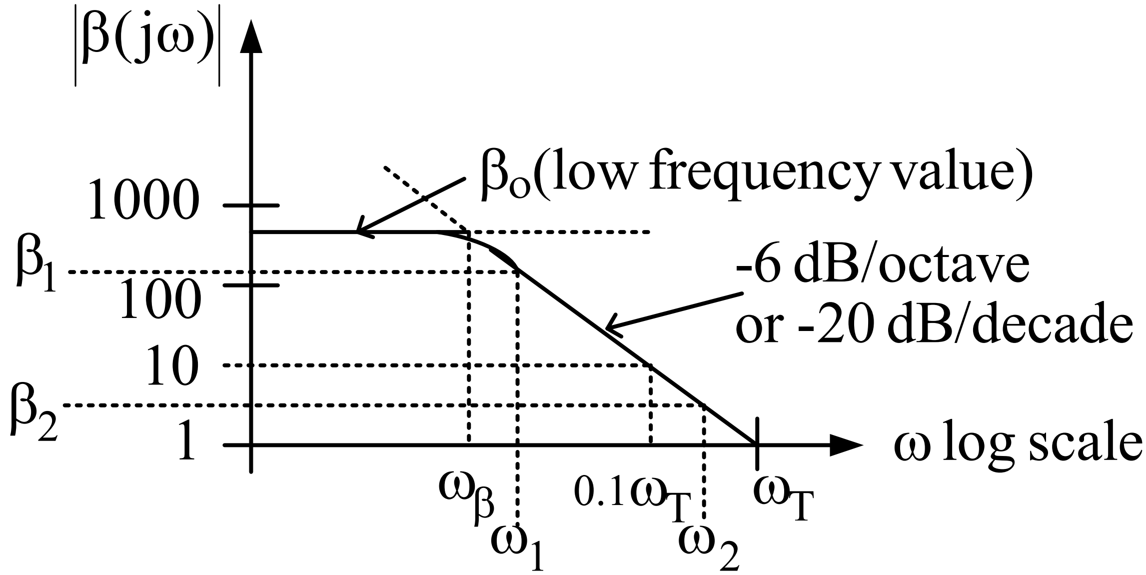
$$\tau_T = \frac{\tau_F g_m}{g_m} + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m}$$

where τ_F = B transit time in the forward direction.

τ_T is dependent on I_C (through g_m) and approaches a constant value of τ_F at high C bias currents ($I_C \uparrow g_m \uparrow \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} \downarrow \tau_T \approx \tau_F$).

At low I_C, the terms C_{je} and C_μ dominate ($I_C \downarrow g_m \downarrow \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} \uparrow \tau_T \approx \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m}$).

To explain why $|\beta(j\omega)| = \frac{\beta_o}{\sqrt{2}}$ is 3 dB down from the low frequency value:



At low frequency, $|\beta(j\omega)| = \beta_o = 20\log\beta_o$ dB

At $\omega = \omega_\beta$, $|\beta(j\omega)| = \frac{\beta_o}{\sqrt{2}} = 20\log\frac{\beta_o}{\sqrt{2}}$ dB

The difference in $|\beta(j\omega)|$ (in dB) for $\omega = \omega_\beta$ and at low frequencies is:

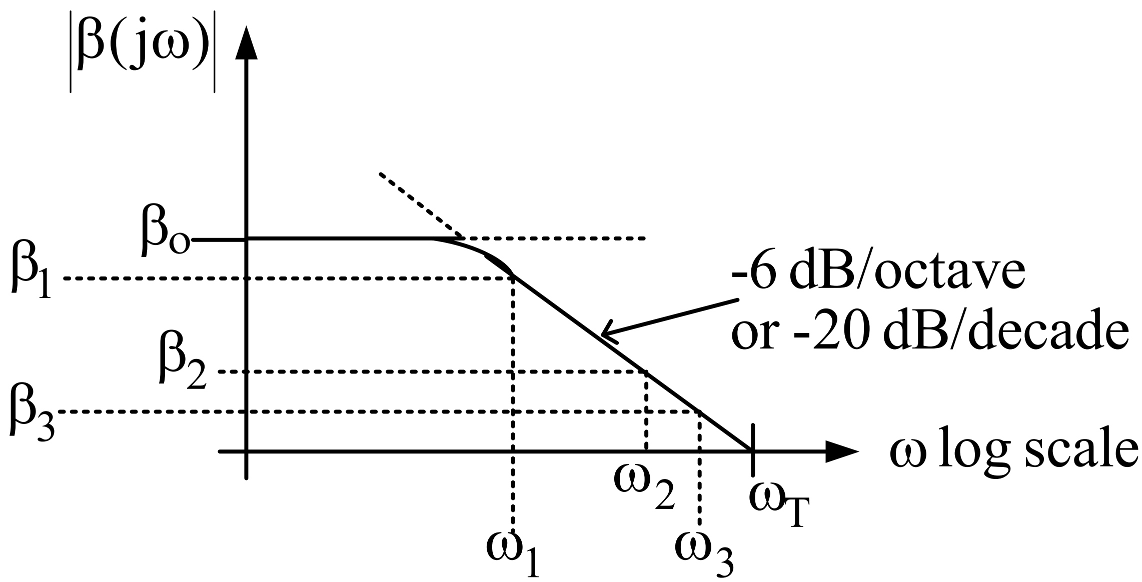
$$20\log\beta_o - 20\log\frac{\beta_o}{\sqrt{2}}$$

$$= 20\log\frac{\beta_o}{\frac{\beta_o}{\sqrt{2}}}$$

$$= 20\log\sqrt{2}$$

$$= 3\ \text{dB}$$

To explain -6 dB/octave is equivalent with -20 dB/decade:



$\omega_2 = 2\omega_1$

$\omega_3 = 10\omega_1$

$$\text{slope(per decade)} = \frac{20\log\beta_3 - 20\log\beta_1}{\log\omega_3 - \log\omega_1}$$

$$= \frac{20\log\dfrac{\beta_3}{\beta_1}}{\log\dfrac{\omega_3}{\omega_1}}$$

$$= 20\log\frac{\beta_3}{\beta_1}$$

$$\text{slope(per octave)} = \frac{20\log\frac{\beta_2}{\beta_1}}{\log\frac{\omega_2}{\omega_1}}$$

$$= \frac{20\log\frac{\beta_2}{\beta_1}}{0.3}$$

If slope is -6 dB/octave,

$$\frac{20\log\frac{\beta_2}{\beta_1}}{0.3} = \frac{-6}{0.3} = -20$$

If slope is -20 dB/decade,

$$20\log\frac{\beta_3}{\beta_1} = -20$$

Hence, -6 dB/octave = -20 dB/decade.

1.5 Large-signal behavior of MOSFET (pg. 38)

MOSFETs have become dominant in the area of digital ICs because they allow high density and low power dissipation

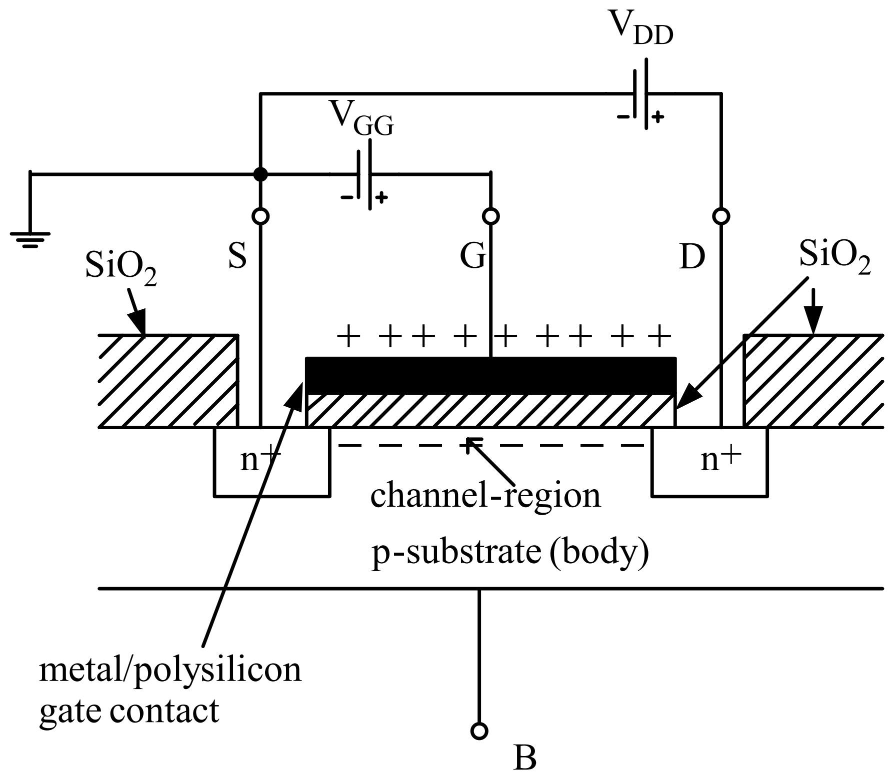
On the other hand, BJT provides many advantages in stand-alone ICs. An example is the g_m per unit bias current in BJT is usually much higher than in the MOS transistors.

Hence, bipolar technologies are often preferred for the analog ICs and MOS technologies for the digital.

To reduce system cost and increase portability, both increased levels of integration and reduced power dissipation are required, forcing the associated analog circuits to use MOS compatible technologies. One way is to use BiCMOS, but BiCMOS is more expensive than MOS process.

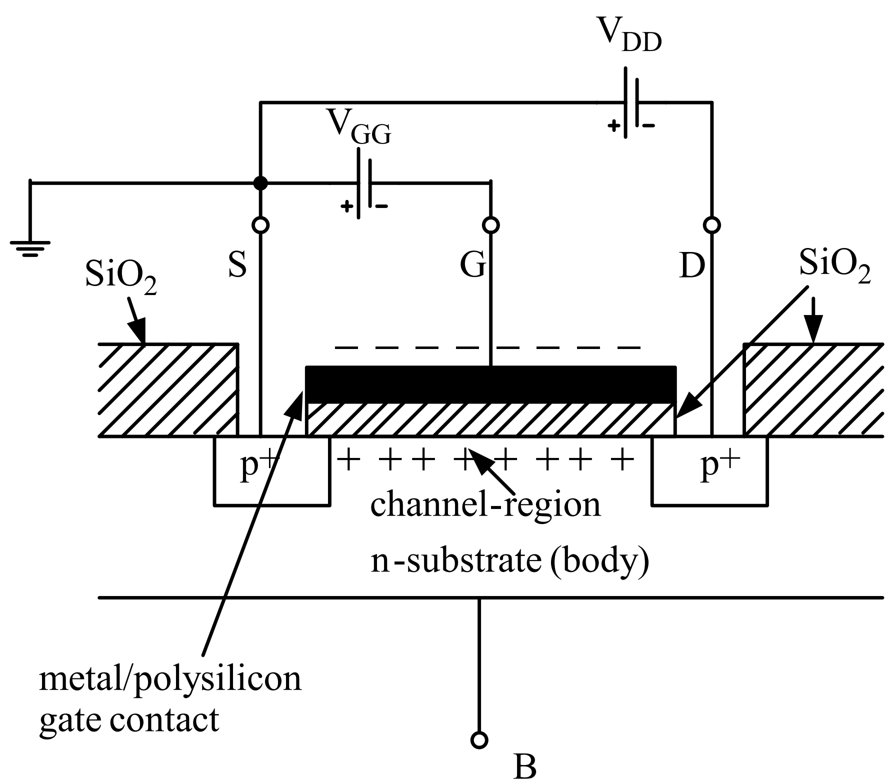
Due to economic reasons, IC manufacturers are driven to use all-MOS processes. Hence, the study of the MOS transistor's characteristics is important.

NMOS



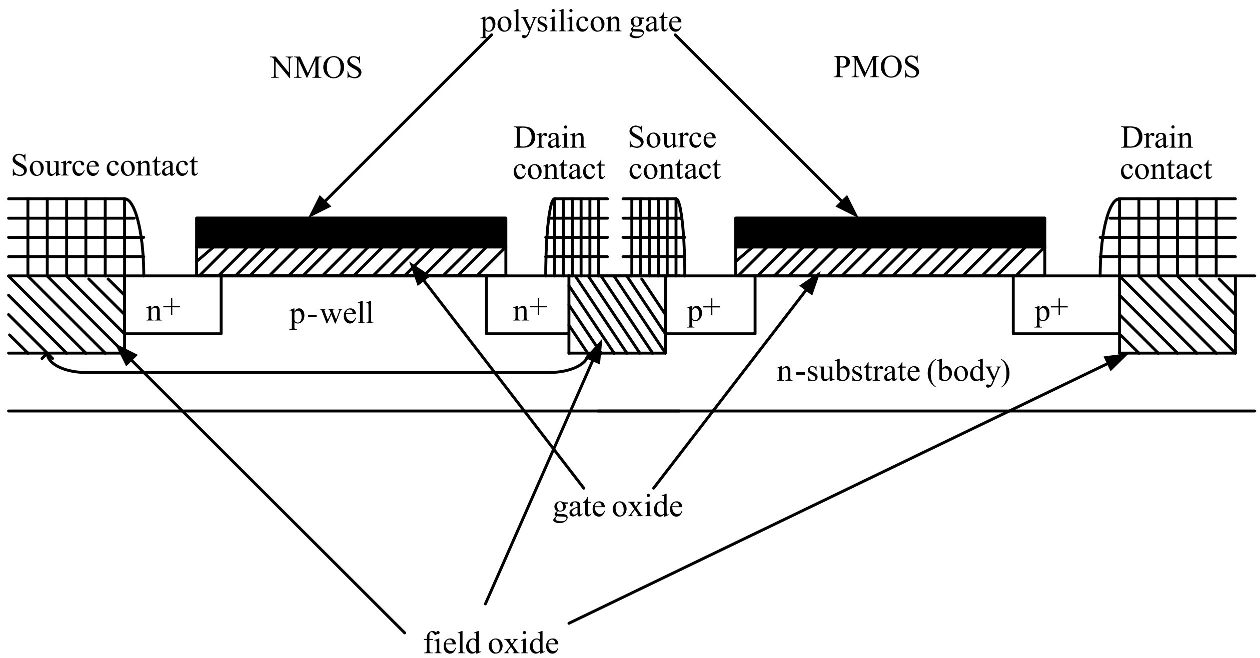
V_{DD}
V_{GG}
S
G
D
SiO_2
SiO_2
n+
n+
channel-region
p-substrate (body)
metal/polysilicon
gate contact
B

PMOS

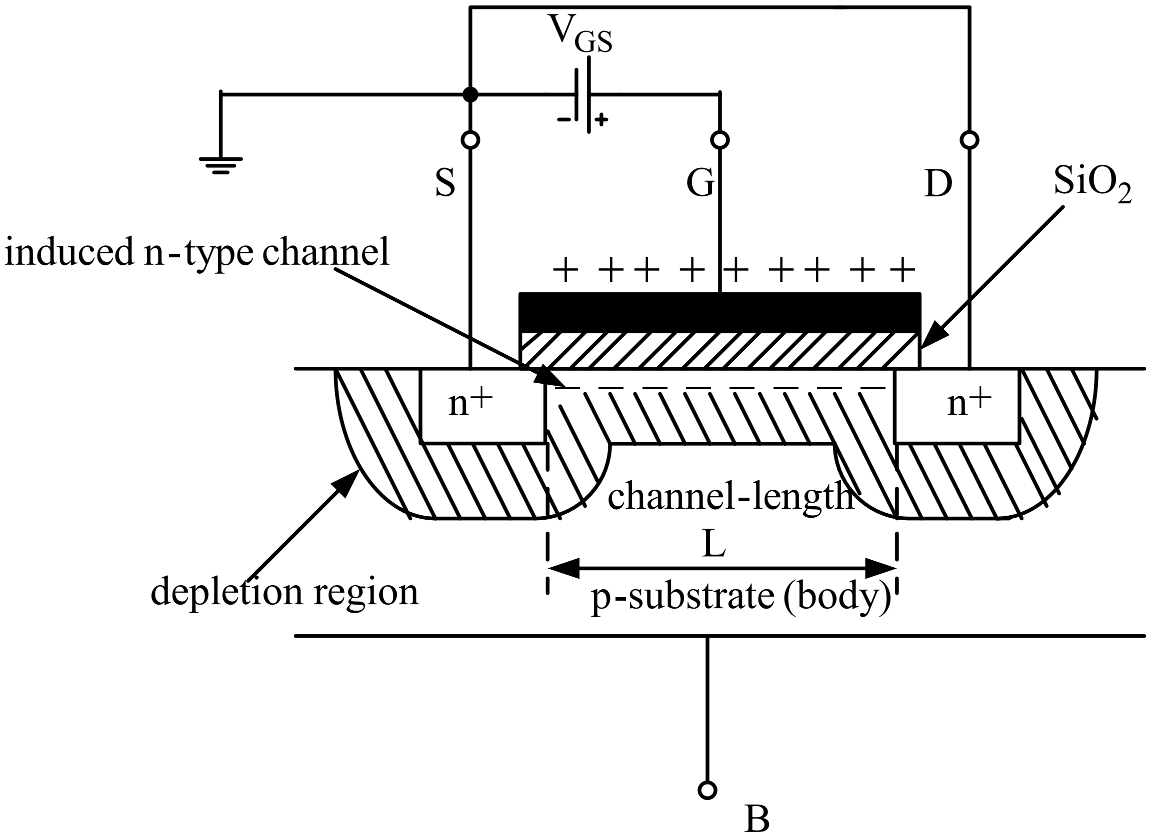


V_{DD}
V_{GG}
S
G
D
SiO_2
SiO_2
p+
p+
channel-region
n-substrate (body)
metal/polysilicon
gate contact
B

CMOS

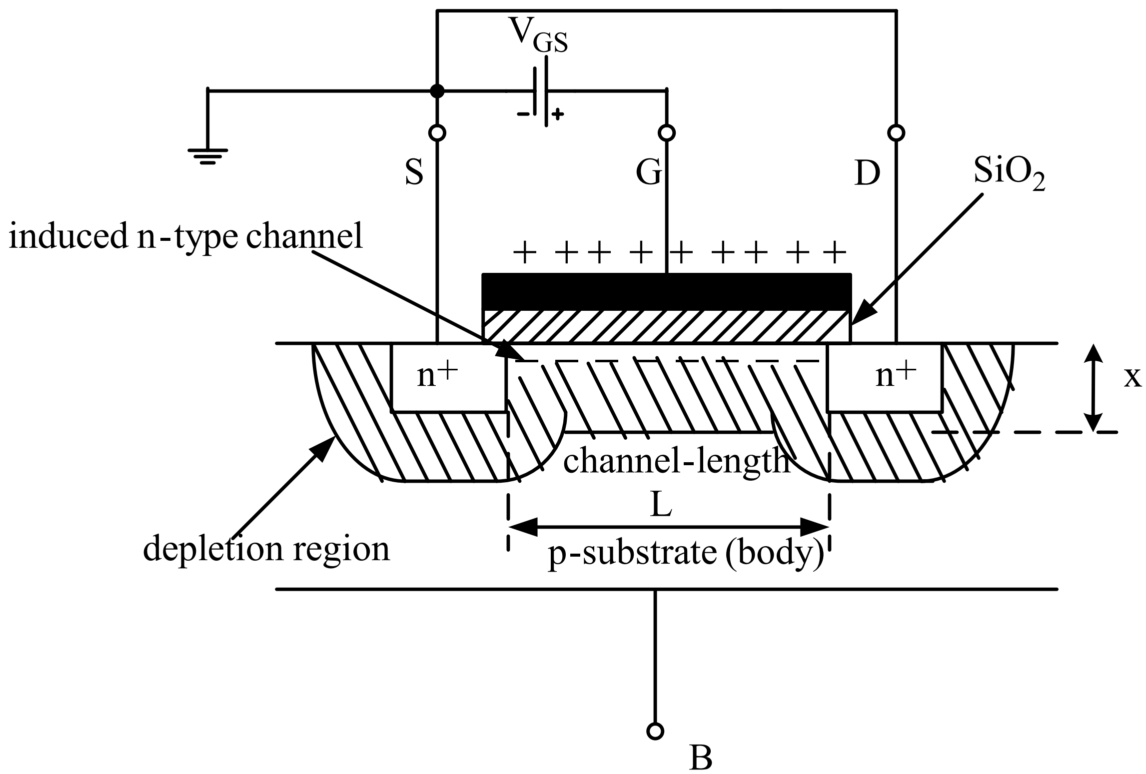

polysilicon gate
NMOS
PMOS
Source contact
Drain contact
Source contact
Drain contact
n+
p-well
n+
p+
p+
n-substrate (body)
gate oxide
field oxide

1.5.1 Transfer characteristics of MOS



V_{GS} modifies the conductance of the channel, hence, the current flowing between S and D. This also means that V_{GS} controls the gain in analog circuit and switching characteristics in digital circuits. No conduction occurs when V_{GS} = 0.

When V_{GS} is +ve, G and substrate form the plates of a capacitor with SiO_2 as a dielectric. +ve charges on the G and –ve charges in the substrate. Holes repelled from the channel region.



$$x = \left(\frac{2\varepsilon\varphi}{qN_A}\right)^{\frac{1}{2}}$$

=depletion-layer width under the oxide

φ=potential in the depletion layer at the oxide-Si interface.

N_A=doping density of the p-type substrate in atoms/cm^3

ε=permittivity of the Si

The charge per unit area in the depletion region:

$$Q = qN_A x = \sqrt{2qN_A\varepsilon\varphi}$$

When the surface potential in the Si reaches a critical value equal to twice the Fermi level, φ_f, a phenomenon known as inversion occurs.

Fermi level, $\varphi_f = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \approx 0.3\,V$

k = Boltzmann constant = $1.38x10^{-23}$ J/K

n_i = intrinsic carrier concentration

$$n_i = \sqrt{N_C N_V}\exp\left(\frac{-E_g}{2kT}\right)$$

E_g = bandgap of Si at T=0K

N_C = density of allowed states near the edge of conduction band

N_V = density of allowed states near the edge of valence band

After the potential in the depletion region reaches $2\varphi_f$, further increase in the G voltage produces no further change in the depletion-layer width but instead induce a thin layer of electrons in the depletion layer at the surface of the Si directly under the oxide. Inversion produces a continuous n-type region with the S-D regions and forms the conducting channel between the S and D. The G-S voltage required to produce an inversion layer is called the threshold voltage, V_t.

$$V_t = V_{t0} + \gamma\left(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right)$$

V_{t0} = threshold voltage with $V_{SB} = 0$

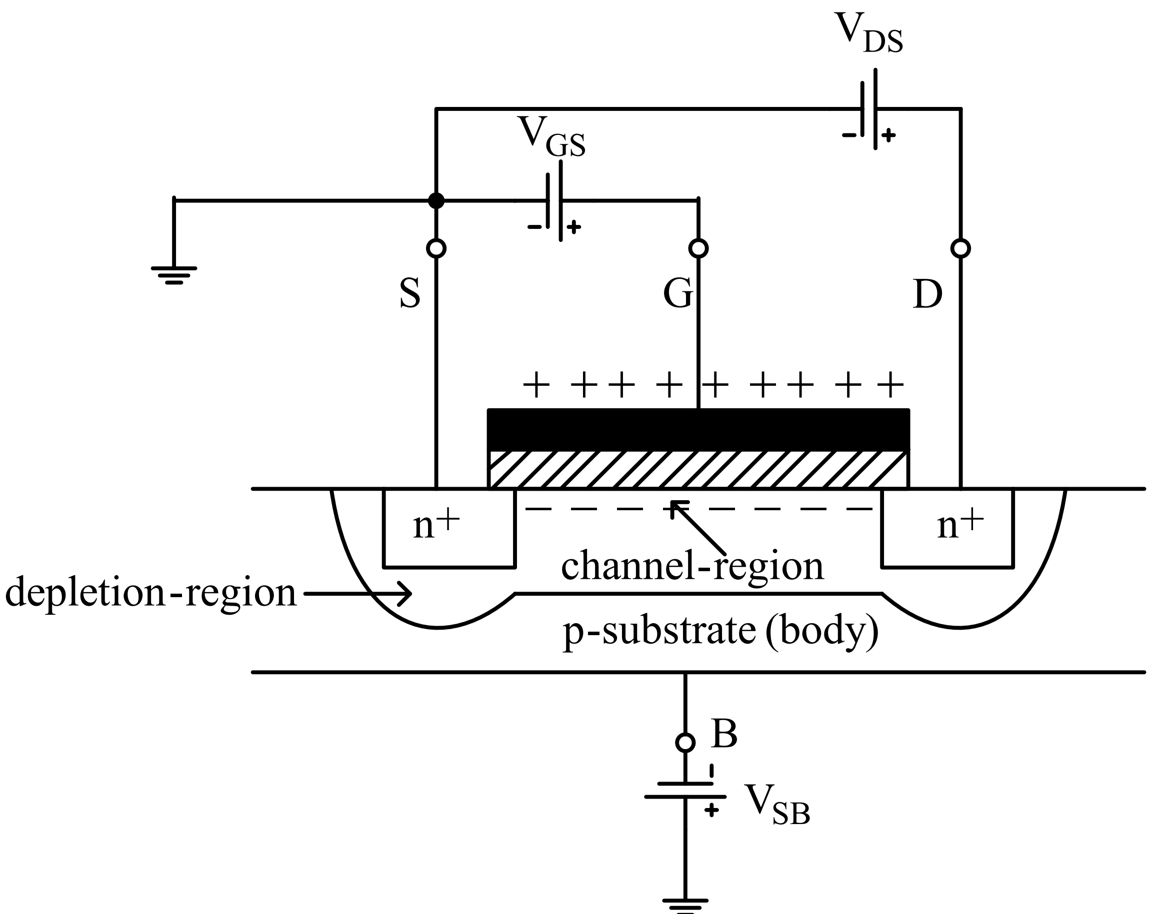
$$\gamma = \frac{1}{C_{ox}}\sqrt{2q\varepsilon N_A}$$

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

where ε = permittivity of the Si

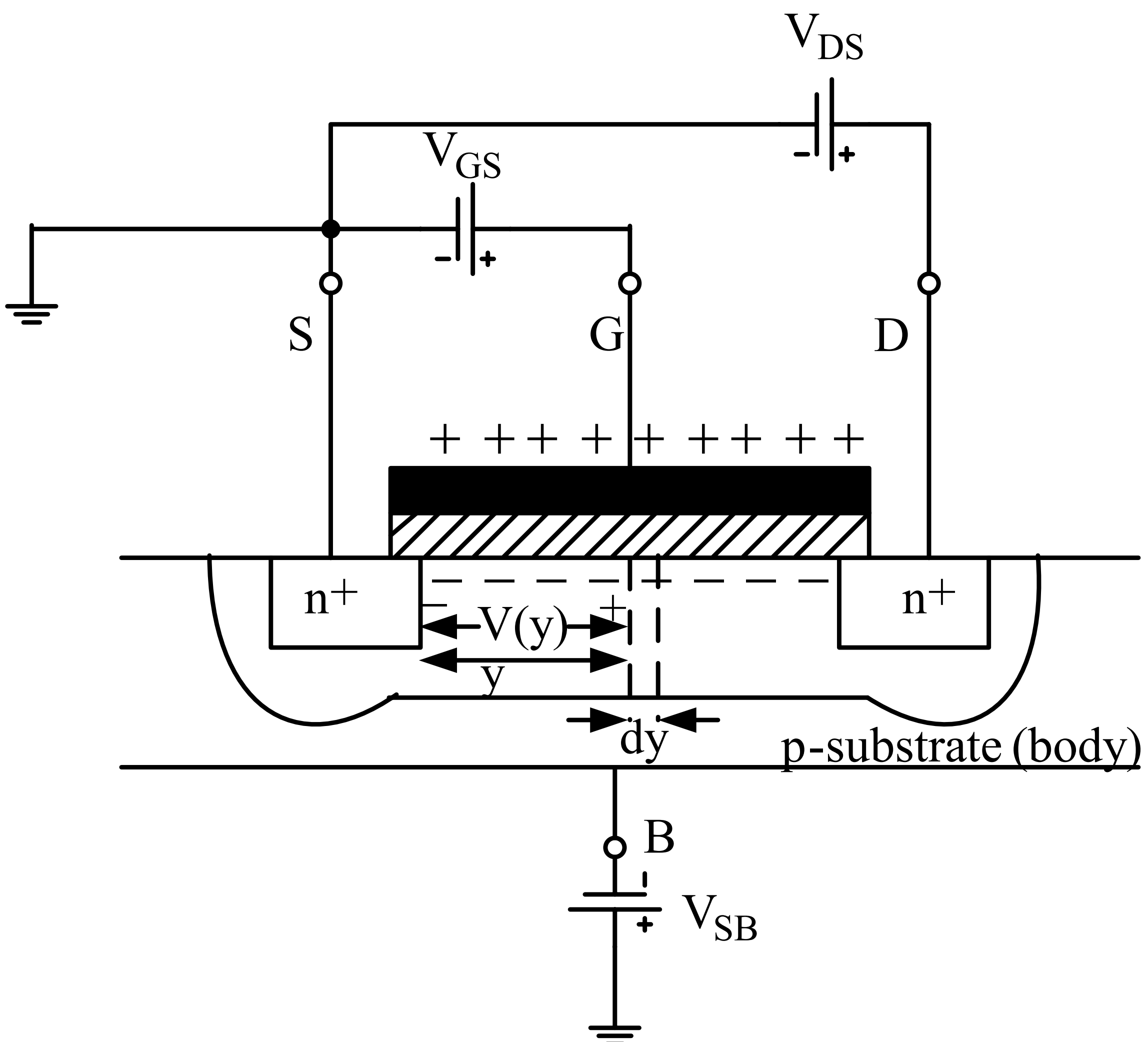
ε_{ox} = permittivity of the oxide

t_{ox} = thickness of the oxide



If $V_{GS} > V_t$, inversion occurs and a conducting channel exists. The channel conductivity is determined by the vertical electric field, which is controlled by the value V_{GS} - V_t . If $V_{DS} = 0$, $I_D = 0$ because the horizontal electric field is 0. The value of the current depends on both the horizontal and the vertical electric fields, explaining the term field-effect transistor.

+ve V_{DS} causes the rb from the D to the substrate to be larger than the S to the substrate. Hence, widest depletion region exists at the D.

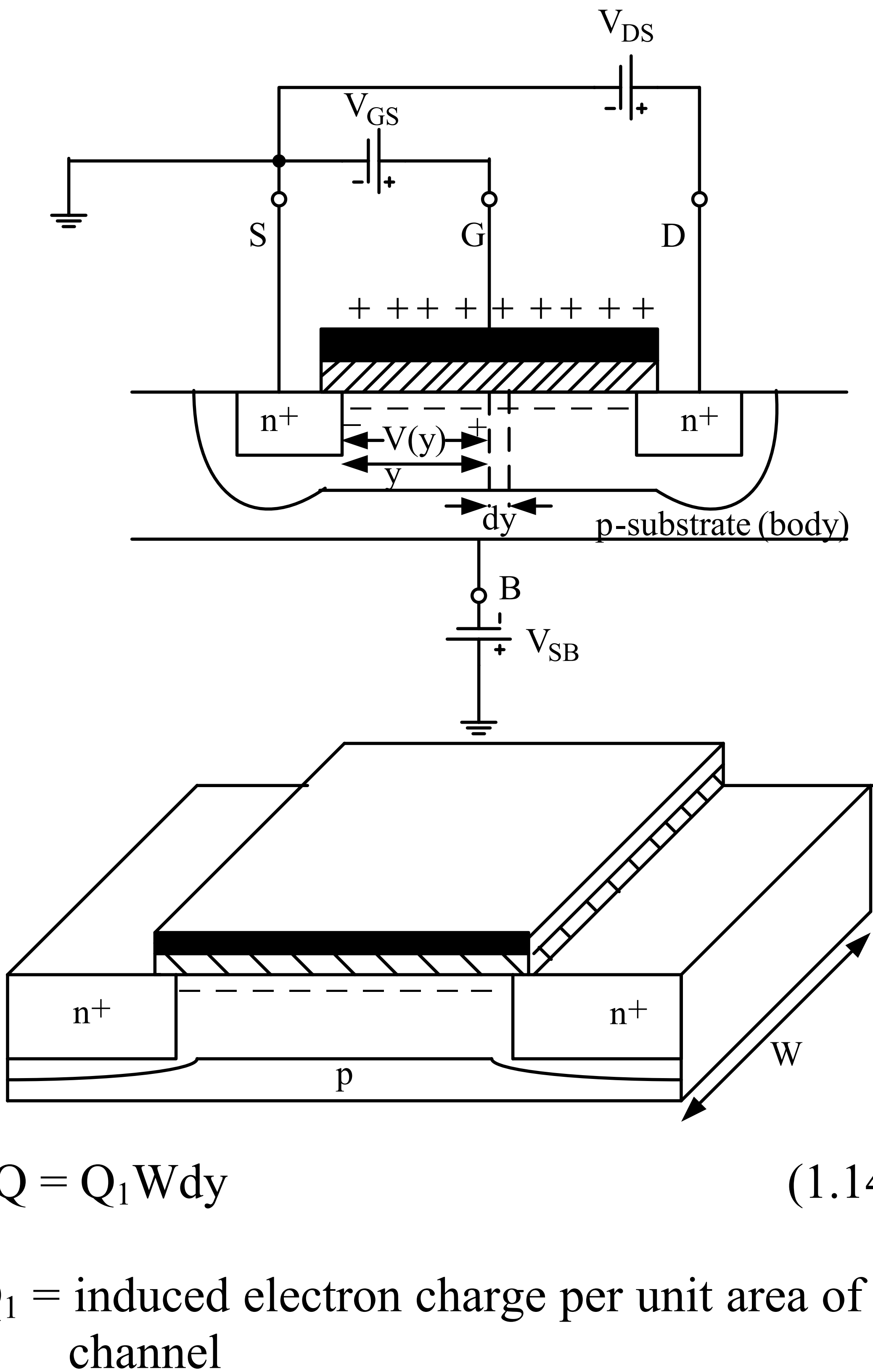


For simplicity, assume that the voltage drop along the channel is small so that the depletion-layer width is constant along the channel.

$$I_D = \frac{dQ}{dt}$$

dQ = incremental channel charge at a distance y from the S in an incremental length dy of the channel.

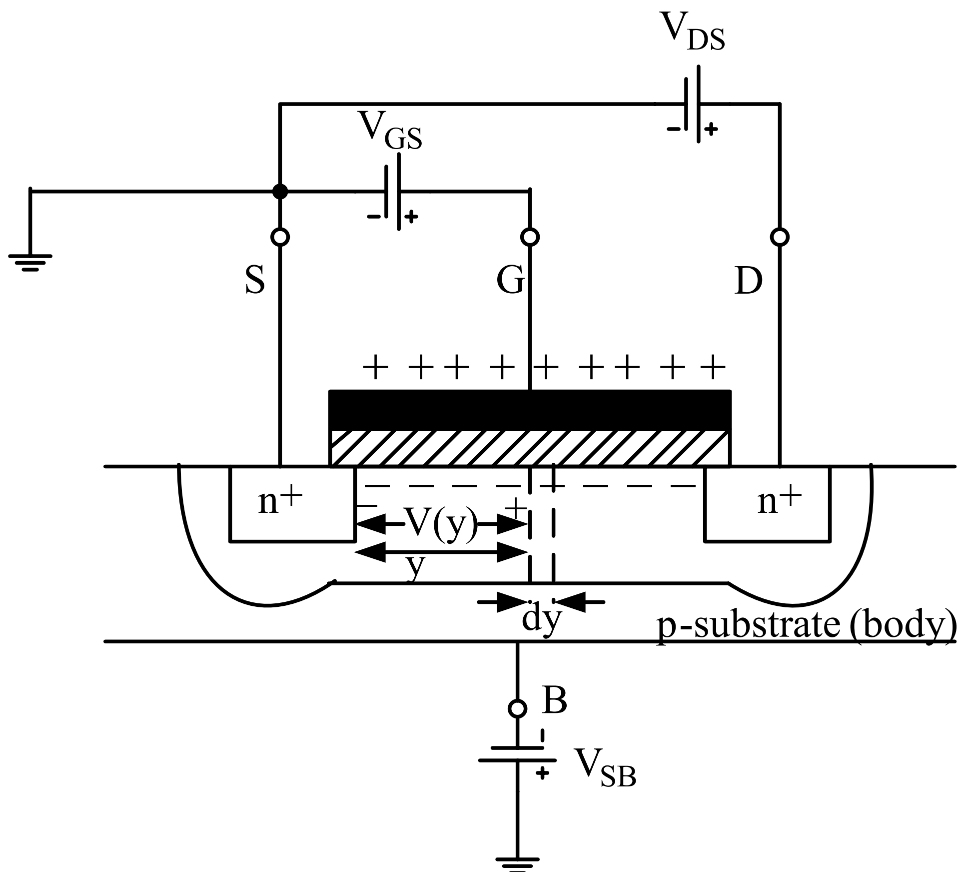
dt = time required for the charge to cross length dy



$$dQ = Q_1 W dy \tag{1.144}$$

Q_1 = induced electron charge per unit area of the channel

W = width of device



$Q_1(y) = C_{ox} [V_{GS} – V(y) - V_t]$ (1.145)

$V(y)$ = voltage at a distance y along the channel wrt S

$V_{GS} – V(y)$ = G to channel voltage at point y from S

$$dt = \frac{dy}{v_d(y)} \quad (1.146)$$

$v_d(y)$ = electron drift velocity at a distance y from the S.

From (1.144) and (1.146),

$dQ = Q_1 W dy$

$$dt = \frac{dy}{v_d(y)}$$

Hence, $dQ = Q_1 W\, v_d(y)\, dt$

Since $I_D = \frac{dQ}{dt}$, then $I_D = Q_1(y) W v_d(y)$ (1.147)

$v_d(y)$ is determined by the horizontal electric field. When then horizontal electric field E(y) is small,

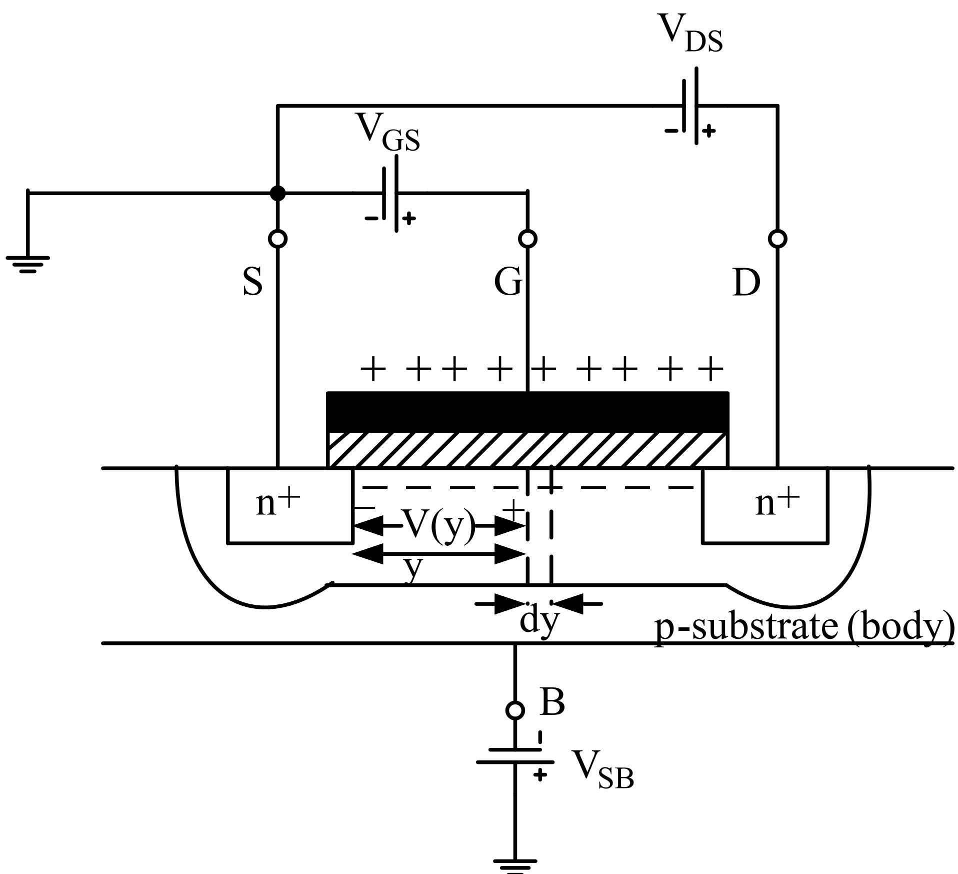
$v_d(y) = \mu_n E(y)$ (1.148)

where μ_n = average electron mobility in the channel
= also known as surface mobility for electron because the channel forms at the surface of the Si

μ_n depends on the temperature and doping level but is almost constant for a wide range of normally used doping levels. Typically:

$\mu_n = 500\ cm^2 / Vs$ to $700\ cm^2 / Vs$

Mobility of electron in the bulk of Si = 1400 cm^2 / Vs. The reason for the difference in the surface and bulk electron mobility is the surface defects that impede the flow of electrons in MOS.



$$E(y) = dV / dy \tag{1.149}$$

dV = incremental voltage drop along the length of channel dy at a distance y from S.

Substituting (1.145), (1.148) and (1.149) into (1.147) gives:

$Q_1(y) = C_{ox}\,[V_{GS} - V(y) - V_t]$

$v_d(y) = \mu_n E(y)$

$E(y) = dV/dy$

$I_D = Q_1(y) W v_d(y)$
$\quad = \mu_n W C_{ox}\,[V_{GS} - V(y) - V_t]\, dV/dy$

$$\int_0^L I_D dy = \int_0^{V_{DS}} W \mu_n C_{ox} \left[V_{GS} - V(y) - V_t \right] dV$$

$$I_D L = W \mu_n C_{ox} \left\{ V_{GS} \left[V \right]_0^{V_{DS}} - \left[\frac{V^2}{2} \right]_0^{V_{DS}} - V_t \left[V \right]_0^{V_{DS}} \right\}$$

$$= W \mu_n C_{ox} \left[V_{GS} V_{DS} - \frac{V_{DS}^2}{2} - V_t V_{DS} \right]$$

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

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

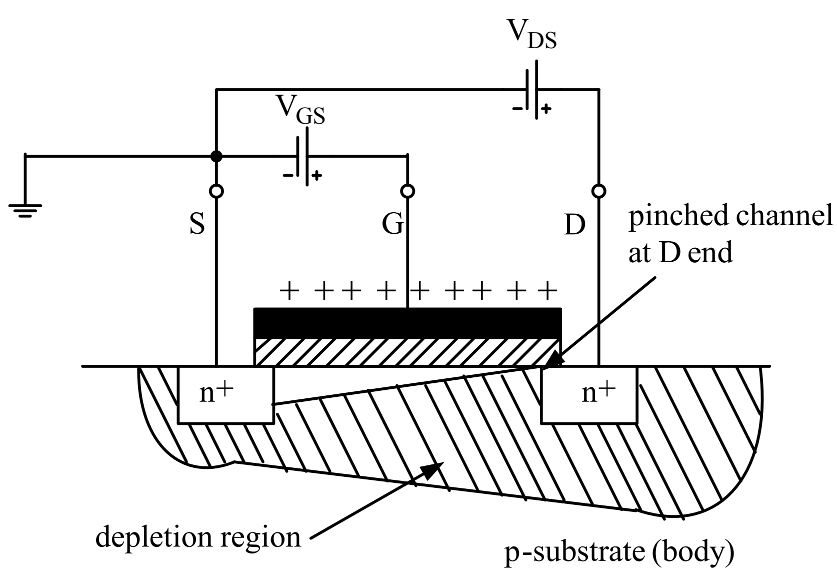
$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left[2(V_{GS}-V_t)V_{DS}-V_{DS}^2\right]$$
$$= \frac{k'}{2}\frac{W}{L}\left[2(V_{GS}-V_t)V_{DS}-V_{DS}^2\right]$$

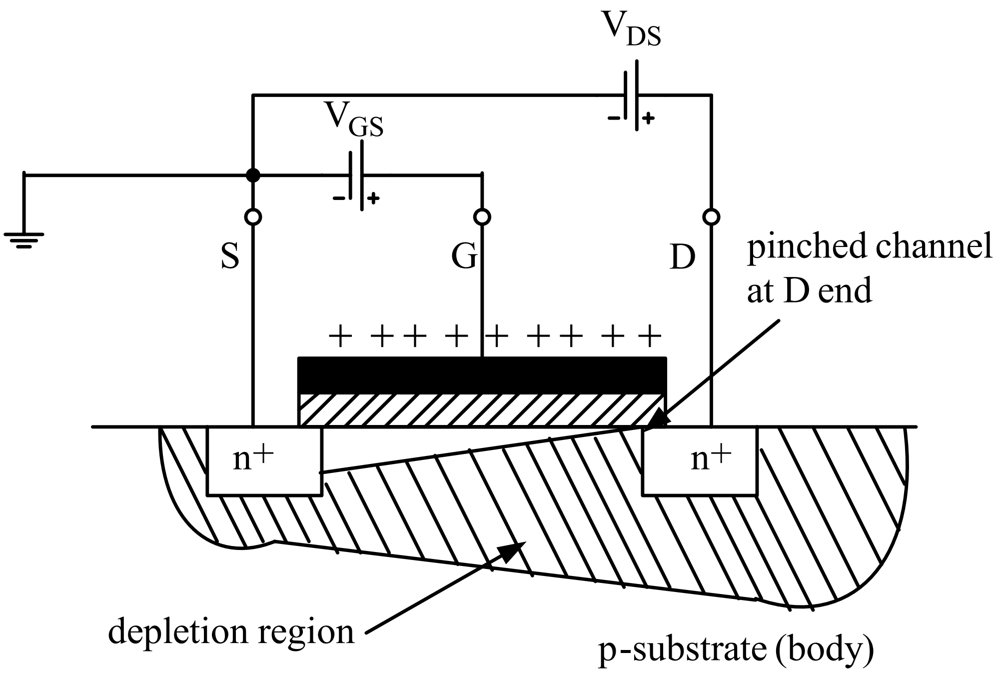
where $k' = \mu_n C_{ox} = \mu_n \varepsilon_{ox} / t_{ox}$

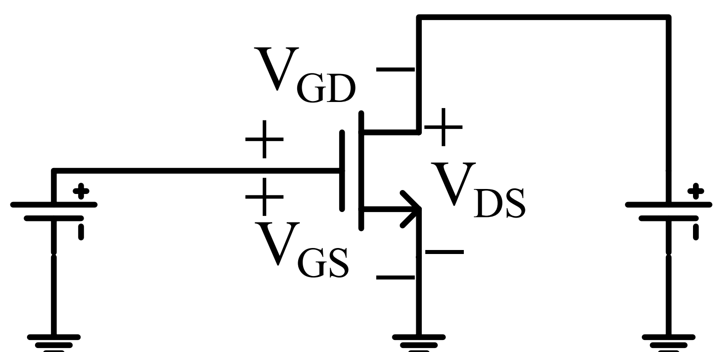
If $V_{DS} << 2(V_{GS} - V_t)$, $I_D \propto V_{DS}$.

$Q_1(y) = C_{ox}\,[V_{GS} - V(y) - V_t]$

If V_{DS} is increased to $V_{DS} = V_{GS} - V_t$, then $Q_1(y) = 0$ at the D end as $V(y) = V_{DS}$ at the D end. Hence, the channel is no longer connected to the D when $V_{DS} > V_{GS} - V_t$. This phenomenon is called pinch-off.







KVL around the transistor:

$-V_{GS} + V_{GD} + V_{DS} = 0$

$V_{DS} = V_{GS} - V_{GD} = V_{DG} + V_{GS}$

When $V_{DS} > V_{GS} - V_t$, $V_{DG} + V_{GS} > V_{GS} - V_t$.

Hence, $V_{GD} < V_t$.

The channel no longer exists at D. At the point where the channel pinches off, the channel voltage is $V_{GS} - V_t$.