VLSI Design (BEC-41)
(Unit-1, Lecture-4)

Presented By:

Prof. R. K. Chauhan

Department of Electronics and Communication Engineering

MOSFET scaling and small-geometry effects

- High density chip
  - The sizes of the transistors are as small as possible
  - The operational characteristics of MOS transistor will change with the reduction of iys dimensions
- There are two basic types of size-reduction strategies
  - Full scaling (constant-field scaling)
  - Constant-voltage scaling
- A new generation of manufacturing technology replaces the previous one about
  - every two or three years
  - The down-scaling factor S about 1.2 to1.5
- The scaling of all dimensions by a factor of S>1 leads to the reduction of the area occupied by the transistor by a factor of S^2

Table 3.1 Reduction of the minimum feature size (minimum dimensions that can be defined and manufactured on chip) over the years, for a typical CMOS gate-array process

Year	1985	1987	1989	1991	1993	1995	1997	1999
Feature size (μm)	2.5	1.7	1.2	1.0	0.8	0.5	0.35	0.25

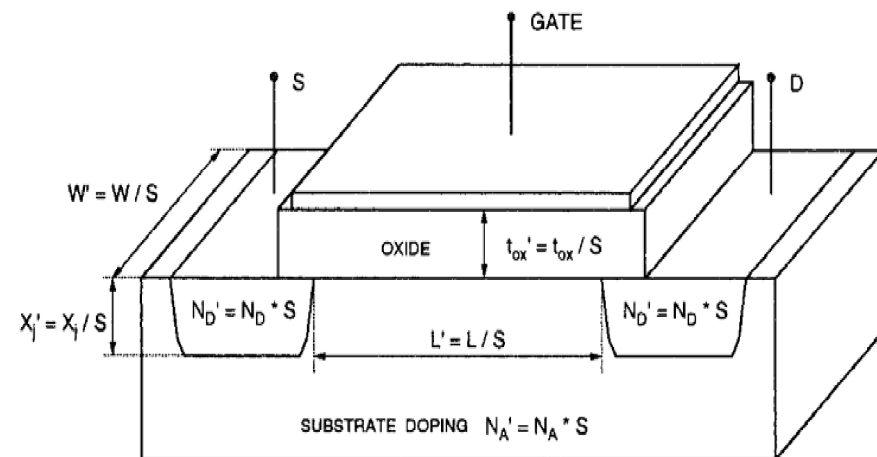


Figure 3.24 Scaling of a typical MOSFET by a scaling factor of *S*.

Full scaling (constant-field scaling)

To achieve this goal, all potentials must be scaled down proportionally, by the same scaling factor

Assuming the surface mobility μ_n is not significantly affected by the scaled doping density

The gate oxide capacitance per unit area

$$C'_{ox} = \frac{\varepsilon_{ox}}{t'_{ox}} = S \cdot \frac{\varepsilon_{ox}}{t_{ox}} = S \cdot C_{ox}$$

The aspect ratio W/L unchanged $\Rightarrow$ the k_n will also scaled by a factor of S

The linear mode drain current

$$I'_D(lin) = \frac{k'_n}{2} \cdot \left[2 \cdot \left(V'_{GS} - V'_T\right) \cdot V'_{DS} - V'^2_{DS}\right]$$

$$= \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot \left[2 \cdot \left(V_{GS} - V_T\right) \cdot V_{DS} - V^2_{DS}\right] = \frac{I_{D(lin)}}{S}$$

The saturation mode drain current

$$I'_D(sat) = \frac{k'_n}{2} \cdot \left(V'_{GS} - V'_T\right)^2 = \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot \left(V_{GS} - V_T\right)^2 = \frac{I_{D(sat)}}{S}$$

The power dissipation

$$P' = I'_D \cdot V'_{DS} = \frac{1}{S^2} \cdot I_D \cdot V_{DS} = \frac{P}{S^2}$$

Table 3.2 Full scaling of MOSFET dimensions, potentials, and doping densities

Quantity	Before scaling	After scaling
Channel length	L	$L' = L/S$
Channel width	W	$W' = W/S$
Gate oxide thickness	t_{ox}	$t'_{ox} = t_{ox}/S$
Junction depth	x_j	$x'_j = x_j/S$
Power supply voltage	V_{DD}	$V'_{DD} = V_{DD}/S$
Threshold voltage	V_{T0}	$V'_{T0} = V_{T0}/S$
Doping densities	N_A	$N'_A = S \cdot N_A$
	N_D	$N'_D = S \cdot N_D$

Table 3.3 Effects of full scaling upon key device characteristics

Quantity	Before scaling	After scaling
Oxide capacitance	C_{ox}	$C'_{ox} = S \cdot C_{ox}$
Drain current	I_D	$I'_D = I_D/S$
Power dissipation	P	$P' = P/S^2$
Power density	$P/Area$	$P'/Area' = P/Area$

The significant reduction of the power dissipation is one of the most attractive features of full scaling

The *power density* per unit area remaining virtually unchanged

C_g is scaled down by a factor of $S \Rightarrow$ the charge - up, and charge - down time improved

A reduction of various parasitic capacitances abd resistances

Constant-voltage scaling

All dimensions of the MOSFETare reduced by a factor of S.

The power supply voltage and the terminal voltages remained unchanged.

The doping densities must be increased by a factor of S^2 in order to preserve the charge - field relations

The gate oxide capacitance per unit area C_{ox} is increased by a factor of S

$\Rightarrow$ The transconductance parameter is also increased by S

The linear mode drain current

$$I_D'(lin) = \frac{k_n'}{2} \cdot \left[2 \cdot \left(V_{GS}' - V_T'\right) \cdot V_{DS}' - V_{DS}'^2\right]$$

$$= \frac{S \cdot k_n}{2} \cdot \left[2 \cdot \left(V_{GS} - V_T\right) \cdot V_{DS} - V_{DS}^2\right] = S \cdot I_D(lin)$$

The saturation mode drain current

$$I_D'(sat) = \frac{k_n'}{2} \left(V_{GS}' - V_T'\right)^2 = \frac{S \cdot k_n}{2} \cdot \left(V_{GS} - V_T\right)^2 = S \cdot I_D(sat)$$

The drain current density increased by a factor of S^3

The power dissipation

$$P' = I_D' \cdot V_{DS}' = (S \cdot I_D) \cdot V_{DS} = S \cdot P$$

The power density incresaed by a factor of S^3

To summarized, constant - voltage scaling may be preferred over full scaling in mamy practical cases because of the external voltage - level constraints.

Disadv. $\Rightarrow$ increasing current density, power density

$\Rightarrow$ electromigration, hot carrier degradation, oxide breakdown, and electrical over - stress

Table 3.5 Effects of constant-voltage scaling upon key device characteristics

Quantity	Before scaling	After scaling
Oxide capacitance	C_{ox}	$C_{ox}' = S \cdot C_{ox}$
Drain current	I_D	$I_D' = S \cdot I_D$
Power dissipation	P	$P' = S \cdot P$
Power density	$P/Area$	$P'/Area' = S^3 \cdot (P/Area)$

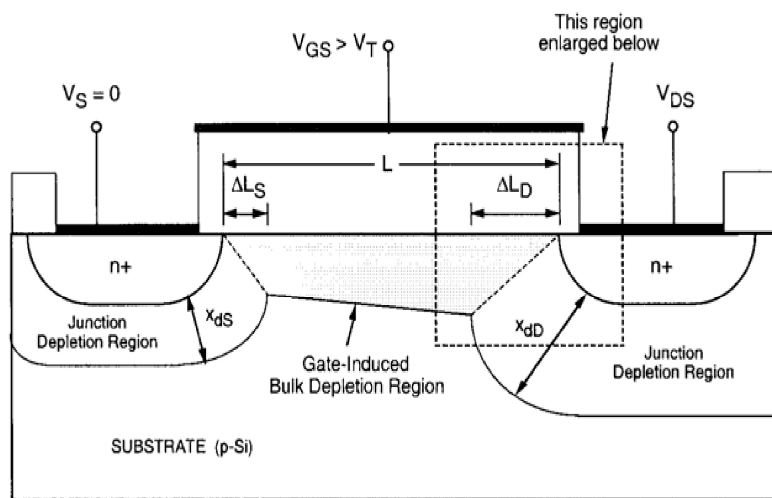
Department of ECE
MMMUT, Gorakhpur

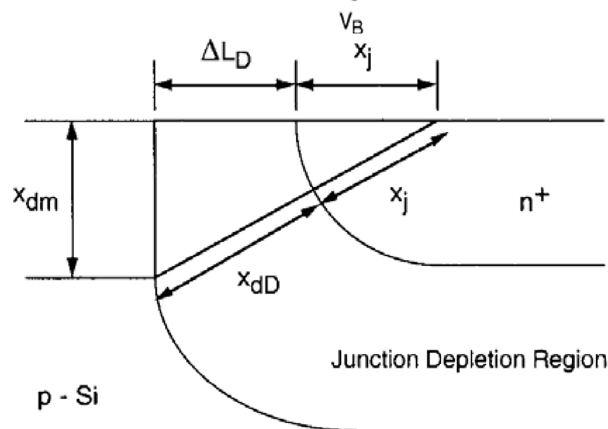
Short-channel effects

- A MOS transistor is called a short-channel device
  - If its channel length is on the same order of magnitude as the depletion region thickness of the *S* and *D* junction
  - The effective channel length $L_{eff} \approx S, D$ junction depth x_j
  - Two physical phenomena arise from short-channel effects
    - The limitations imposed on electron drift characteristics in the channel
      - The lateral electric field E_y increased, v_d reached saturation velocity
      - $$I_{D(sat)} = W \cdot v_{d(sat)} \cdot \int_0^{L_{eff}} q \cdot n(x) \cdot dx = W \cdot v_{d(sat)} \cdot |Q_I| = W \cdot v_{d(sat)} \cdot C_{ox} \cdot V_{DSAT}$$
        - » No longer a quadratic function of V_{GS}, virtually independent of the channel length
      - The carrier velocity in the channel also a function of E_x
        - » Influence the scattering of carriers in the surface
        - » $$\mu_n(eff) = \frac{\mu_{no}}{1 + \Theta \cdot Ex} = \frac{\mu_{no}}{1 + \frac{\Theta \varepsilon_{ox}}{t_{ox} \varepsilon_{Si}} \cdot (V_{GS} - V_c(y))} = \frac{\mu_{no}}{1 + \eta \cdot (V_{GS} - V_T)}$$
    - The modification of the threshold voltage due to the shortening channel length

Short-channel effects-modification of V_T

- The n^+ drain and source diffusion regions in p-type substrate induce a significant amount of depletion charge
  - The long channel VT, overetimates the depletion charge support by the gate voltage
  - The bulk depletion region ⇨ asymmetric trapezoidal shape
    - A significant portion of the total depletion region charge is due the S and D junction depletion





$$V_{T0}(short\ channel) = \mathrm{V_{T0}} - \Delta \mathrm{V_{T0}}$$

$$Q_{B0} = -\left(1 - \frac{\Delta L_S + \Delta L_D}{2L}\right) \cdot \sqrt{2 \cdot q \cdot \varepsilon_{Si} \cdot N_A \cdot |2\varphi_F|}$$

$$x_{dS} = \sqrt{\frac{2 \cdot \varepsilon_{Si}}{q \cdot N_A} \cdot \varphi_0}, \quad x_{dD} = \sqrt{\frac{2 \cdot \varepsilon_{Si}}{q \cdot N_A} \cdot (\varphi_0 + V_{DS})}, \quad \varphi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_D \cdot N_A}{n_i^2}\right)$$

$$(x_j + x_{dD})^2 = x_{dm}^2 + (x_j + \Delta L_D)^2$$

$$\Delta L_D^2 + 2 \cdot x_j \cdot \Delta L_D + x_{dm}^2 - x_{dD}^2 - 2 \cdot x_j \cdot x_{dD} = 0$$

$$\Delta L_D = -x_j + \sqrt{x_j^2 - (x_{dm}^2 - x_{dD}^2) + 2x_j x_{dD}} \cong x_j \cdot \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1\right)$$

$$\Delta L_S \cong x_j \cdot \left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1\right)$$

$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2 \cdot q \cdot \varepsilon_{Si} \cdot N_A \cdot |2\varphi_F|} \cdot \frac{x_j}{2L} \cdot \left[\left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1\right) + \left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1\right)\right]$$

Example 6 (1)

Consider an n-channel MOS process with the following parameters: substrate doping density $N_A = 10^{16}$ cm^{-3}, polysilicon gate doping density N_D (gate) $= 2 \times 10^{20}$ cm^{-3}, gate oxide thickness $t_{ox} = 50$ nm, oxide-interface fixed charge density $N_{ox} = 4 \times 10^{10}$ cm^{-2}, and source and drain diffusion doping density $N_D = 10^{17}$ cm^{-3}. In addition, the channel region is implanted with p-type impurities (impurity concentration $N_I = 2 \times 10^{11}$ cm^{-2}) to adjust the threshold voltage. The junction depth of the source and drain diffusion regions is $x_j = 1.0$ μm.

Plot the variation of the zero-bias threshold voltage V_{T0} as a function of the channel length (assume that $V_{DS} = V_{SB} = 0$). Also find V_{T0} for $L = 0.7$ μm, $V_{DS} = 5$ V, and $V_{SB} = 0$.

First, we have to find the zero-bias threshold voltage using the conventional formula (3.23). The threshold voltage *without* the channel implant was already calculated for the same process parameters in Example 3.2, and was found to be $V_{T0} = 0.40$ V. The additional p-type channel implant will increase the threshold voltage by an amount of qN_I/C_{ox}. Thus, we find the long-channel zero-bias threshold voltage for the process described above as

$$V_{T0} = 0.40 \text{ V} + \frac{q \cdot N_I}{C_{ox}} = 0.40 \text{ V} + \frac{1.6 \times 10^{-19} \cdot 2 \times 10^{11}}{7.03 \times 10^{-8}} = 0.855 \text{ V}$$

Next, the amount of threshold voltage reduction due to short-channel effects must be calculated using (3.88). The source and drain junction built-in voltage is

$$\phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_D \cdot N_A}{n_i^2}\right) = 0.026 \text{ V} \cdot \ln\left(\frac{10^{17} \cdot 10^{16}}{2.1 \times 10^{20}}\right) = 0.76 \text{ V}$$

Example 6 (2)

For zero drain bias, the depth of source and drain junction depletion regions is found as

$$x_{dS} = x_{dD} = \sqrt{\frac{2 \cdot \varepsilon_{Si}}{q \cdot N_A} \cdot \phi_0} = \sqrt{\frac{2 \cdot 11.7 \cdot 8.85 \times 10^{-14}}{1.6 \times 10^{-19} \cdot 10^{16}} \cdot 0.76}$$

$$= 31.4 \times 10^{-6} \text{ cm} = 0.314\ \mu\text{m}$$

Now, the threshold voltage shift ΔV_{T0} due to short-channel effects can be calculated as a function of the gate (channel) length L.

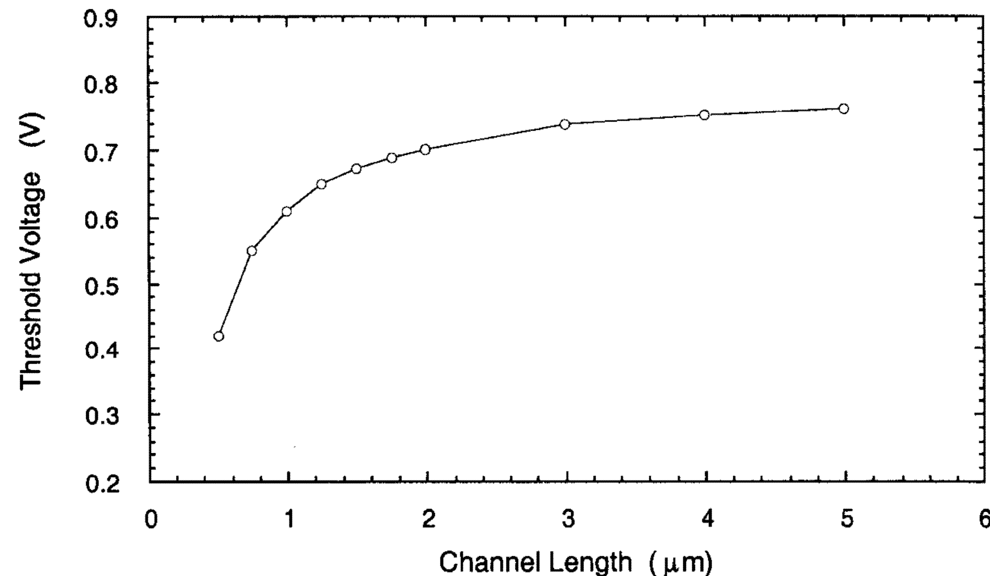
$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2q\varepsilon_{Si}N_A|2\phi_F|} \cdot \frac{x_j}{2L} \cdot \left[\left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1\right) + \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1\right)\right]$$

$$= \frac{4.82 \times 10^{-8} \text{ C/cm}^2}{7.03 \times 10^{-8} \text{ F/cm}^2} \cdot \frac{1.0\ \mu\text{m}}{L} \cdot \left(\sqrt{1 + \frac{2 \cdot 0.314\ \mu\text{m}}{1.0\ \mu\text{m}}} - 1\right)$$

Finally, the zero-bias threshold voltage is found as

$$V_{T0}(\textit{short channel}) = 0.855 \text{ V} - 0.19 \text{ V} \cdot \frac{1}{L[\mu\text{m}]}$$

The following plot shows the variation of the threshold voltage with the channel length. The threshold voltage decreases by as much as 50% for channel lengths in the submicron range, while it approaches the value of 0.8 V for larger channel lengths.



Example 6 (3)

Since the conventional threshold voltage expression (3.23) is not capable of accounting for this drastic reduction of V_{T0} at smaller channel lengths, its application for short-channel MOSFETs must be carefully restricted.

Now, consider the variation of the threshold voltage with the applied drain-to-source voltage. Equation (3.82) shows that the depth of the drain junction depletion region increases with the voltage V_{DS}. For a drain-to-source voltage of $V_{DS} = 5$ V, the drain depletion depth is found as:

$$x_{dD} = \sqrt{\frac{2 \cdot \varepsilon_{Si}}{q \cdot N_A} \cdot (\phi_0 + V_{DS})}$$

$$= \sqrt{\frac{2 \cdot 11.7 \cdot 8.85 \times 10^{-14}}{1.6 \times 10^{-19} \cdot 10^{16}} \cdot (0.76 + 5.0)} = 0.863\ \mu\text{m}$$

The resulting threshold voltage shift can be calculated by substituting x_{dD} found above in (3.88).

$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2q\varepsilon_{Si} N_A |2\phi_F|} \cdot \frac{x_j}{2L} \cdot \left[\left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1\right) + \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1\right)\right]$$

$$= \frac{4.82 \times 10^{-8}}{7.03 \times 10^{-8}} \cdot \frac{1.0}{2 \cdot 0.7} \cdot \left[\left(\sqrt{1 + \frac{2 \cdot 0.314}{1.0}} - 1\right) + \left(\sqrt{1 + \frac{2 \cdot 0.863}{1.0}} - 1\right)\right]$$

$$= 0.45\ V$$

The threshold voltage of this short-channel MOS transistor is calculated as

$$V_{T0} = 0.855 \text{ V} - 0.45 \text{ V} = 0.405 \text{ V}$$

which is significantly lower than the threshold voltage predicted by the conventional long-channel formula (3.23).

Narrow-channel effect

- Channel width W on the same order of magnitude as the maximum depletion region thickness x_{dm}
- The actual threshold voltage of such device is larger than that predicted by the conventional threshold voltage
- Fringe depletion region under field oxide
  - $V_{T0}(\text{narrow channel}) = \mathrm{V_{T0}} + \Delta \mathrm{V_{T0}}$

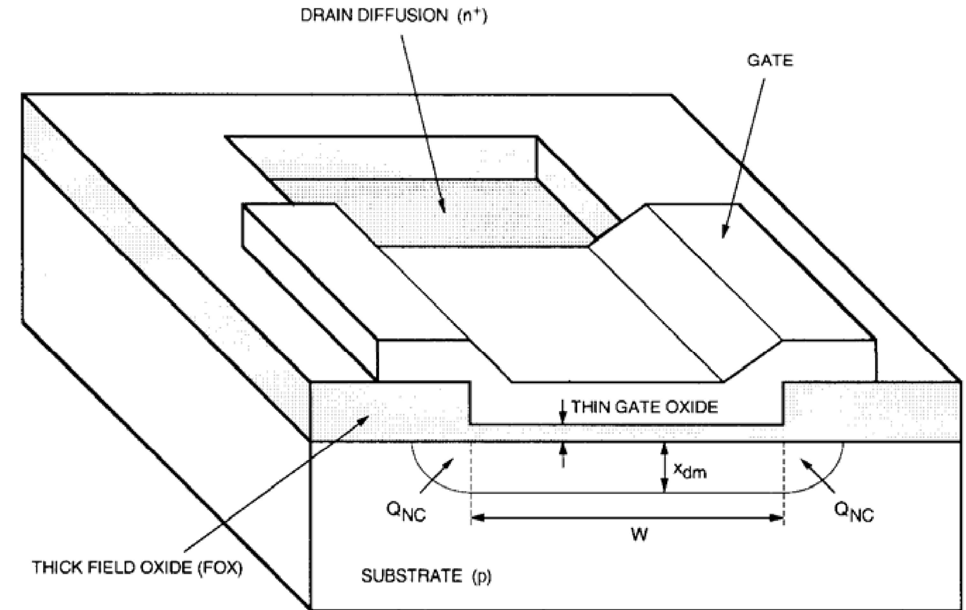


Figure 3.26 Cross-sectional view (across the channel) of a narrow-channel MOSFET. Note that Q_{NC} indicates the extra depletion charge due to narrow-channel effects.

$$\Delta \mathrm{V_{T0}} = \frac{1}{C_{ox}} \cdot \sqrt{2q\varepsilon_{Si} N_A |2\phi_F|} \cdot \frac{\kappa \cdot x_{dm}}{W}$$

$$\kappa = \frac{\pi}{2} \text{ for depletion region modeled by quarter - circular arcs}$$

Other limitations imposed by small-device geometries

- The current flow in the channel are controlled by two dimensional electric field vector
- Subthreshold conduction
  - Drain-induced barrier lowering (DIBL)
  - A nonzero drain current I_D for $V_{GS}<V_{T0}$
  - $$I_D(subthreshold) \cong \frac{qD_n W x_c n_0}{L_B} \cdot e^{\frac{q\phi_r}{kT}} \cdot e^{\frac{q}{kT}(A \cdot V_{GS} + B \cdot V_{DS})}$$
- Punch-through
  - The gate voltage loses its control upon the drain current, and the current rises sharply
- Gate oxide thickness t_{ox} scaled to t_{ox}/S, is restricted by processing difficulties
  - Pinholes, oxide breakdown
- Hot-carrier effect

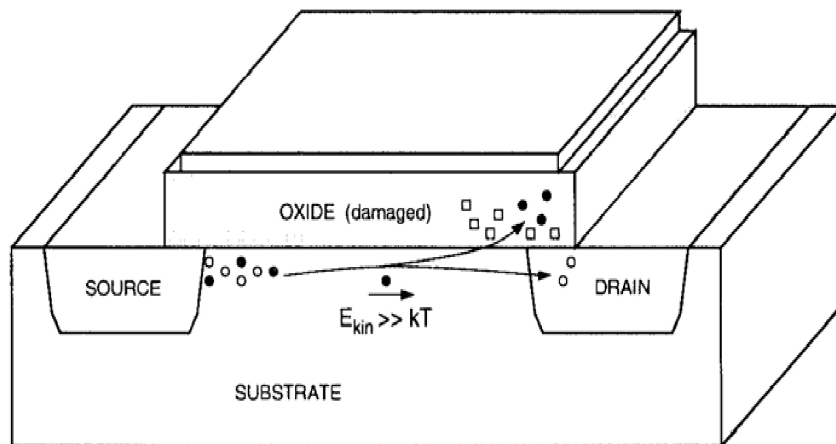


Figure 3.27 Hot-carrier injection into the gate oxide and resulting oxide damage.

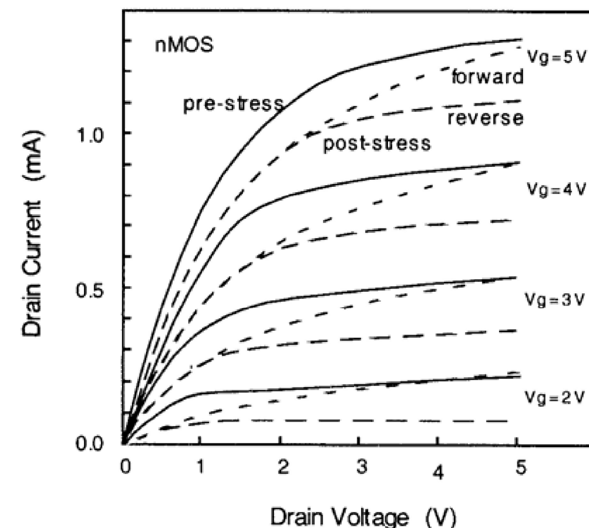


Figure 3.28 Typical drain current vs. drain voltage characteristics of an n-channel MOS transistor before and after hot-carrier induced oxide damage.

Department of ECE
MMMUT, Gorakhpur