Session 2: Solid State Devices

CMOS Technology Review
Outline

1. A
   - B
   - C
   - D
   - E
2. F
   - G
3. H
4. I
5. J
Outline

1. MOSFET Basics
   Ideal MOSFET physics
   Main parameters: threshold, leakage and speed
   Which FET for what application?
   Scaling theory and good design rules of CMOS Devices

2. The Real World
   Threshold voltage control limitations
   Gate oxide leakage and capacitance scaling

3. Technological Solution?
   Gate alternative: High-K and Metal Gate
   Channel engineering: Strained-Si
   Alternative devices and substrates
   Basic logic functions
Metal-Oxide-Semiconductor Field-Effect Transistor:

Drift current flowing between 2 doped regions ("source" & "drain") is modulated by varying the voltage on the "gate" electrode.
The potential barrier to electron flow from the source into the channel region is lowered by applying $V_{GS} > V_T$. 
Qualitative Theory of the NMOSFET

Electrons flow from the source to the drain by drift, when $V_{DS} > 0$. ($I_{DS} > 0$)

The channel potential ($V_c(y)$) varies from $V_s$ at the source end to $V_D$ at the drain end.

$V_{GS} > V_T \Rightarrow$ Inversion-layer "channel" is formed

$V_{DS} \approx 0$

$V_{DS} > 0$
Electrons flow from the source to the drain by drift, when $V_{DS} > 0$. ($I_{DS} > 0$)

The channel potential ($V_c(y)$) varies from $V_S$ at the source end to $V_D$ at the drain end.
MOS – Special Case

Special case: \( \varphi_m = \varphi_s \)

\[
\begin{align*}
E_0 & \quad q\chi_i & \quad E_C \\
q\varphi_m & \quad & \quad q\varphi_s \\
E_{Fm} & \quad & \quad q\chi \\
& \quad & \quad E_G \\
& \quad & \quad E_C \\
& \quad & \quad E_{FS} \\
& \quad & \quad E_V \\
\text{metal} & \quad \text{SiO}_2 & \quad \text{Si, p-type}
\end{align*}
\]
MOS – Special Case

Special case: $\varphi_m = \varphi_s$
$M$ (Al), $O$ (SiO2), $S$ (Si)

$E_0$  

$q \varphi_{m_{Al}} = 4.1$

$E_{Fm}$

$E_G = 9eV$

$E_C$

$0.95$

$q \varphi_s = 4.9eV$

$q \chi = 4.05eV$

$E_{FS}$

$E_{FS}$

$E_V$

metal

SiO2

$E_V$

Si, p-type
Flat Band Voltage

\[ q\varphi_m = 4.1 \]

\[ q\varphi_m - q\varphi_s = qV_G \]

\[ V_{FB} \equiv \varphi_m - \varphi_s = \varphi_{ms} \]

\[ V_{FB} = -0.8 \text{ V} \]

metal

Si, p-type

\[ q\varphi_s = 4.9\text{eV} \]

\[ q\chi = 4.05\text{eV} \]
Voltage Barrier

$E_0$

$q\varphi_{m_{Al}} = 4.1$

$E_{Fm}$

$qV_G$

$V_G = V_{FB} = -0.8\ V$

$E_{Fm}$

$E_C$

$E_{FS}$

$E_V$

No way electrons might pass the voltage barrier!
No Gate Voltage

$V_G = 0 \text{ V}$

$q\varphi_{m_{Al}} = 4.1eV$

$q\varphi_s = 4.9eV$

$q\chi = 4.05eV$

$E_{Fm}$

metal

Si, p-type

$E_F$
Boundary Condition

\[ D_{2n} - D_{1n} = \rho_{surface} \]

\[ \rho_{surface} = 0 \]

\[ \epsilon_1 \epsilon_1 = \epsilon_2 \epsilon_2 \]

\[ \epsilon_{ox} \left. \frac{dE_{ox}}{dx} \right|_{int} = \epsilon_{si} \left. \frac{dE_{si}}{dx} \right|_{int} \]

\[ \left. \frac{dE_{ox}}{dx} \right|_{int} \approx 3 \left. \frac{dE_{si}}{dx} \right|_{int} \]
M (PolyGate) , O (SiO2) , S (Si)

\[ E_0 \]
\[ q\chi = 4.05\text{eV} \]
\[ E_F \]
\[ 0.95 \]
\[ E_C \]
\[ q\varphi_s = 4.9\text{eV} \]
\[ E_G = 9\text{eV} \]
\[ ? E_F \]
\[ V_{FB} = \varphi_m - \varphi_s = -0.85\text{V} \]
\[ \text{n+ poly} \]
\[ \text{SiO2} \]
\[ \text{Si, p-type} \]
\[ E_V \]
No Gate Voltage – Poly Gate

$q\chi = 4.05\text{eV}$

$E_C, E_{Fm}$

$E_V$

$V_G = 0\text{ V}$

metal

Si, p-type

$E_0$

$q\Phi_s = 4.9\text{eV}$

$q\chi = 4.05\text{eV}$
No Gate Voltage

$V_G = 0 \, V$

metal

Si, p-type

$E_{Fm}$

$E_F$ ( Fermi level in metal)

$E_F$ (Fermi level in Si)

$E_C$

$E_V$

$3.1 \, eV$

$5 \, eV$

$0.93$
No Gate Voltage

\[ V_G - V_{FB} = V_{Oxide} + V_{Si} = V_{Ox} + \phi_s \]

As important as KVL
Flat Band

\[ V_G = V_{FB} \]
Accumulation

$V_G < V_{FB}$

$E_F$ $E_C$

$qV_{FB}$

$\rho$

$\pm Q$

$\varepsilon$

$x$

$x$
Flat Band

\[ V_G = V_{FB} \]
Depletion (Weak Inversion)

\[ V_T > V_G > V_{FB} \]
(Strong) Inversion

\[ V_G > V_T \]

\[ E_F \]

\[ qV_{FB} \]

\[ +Q_{inv} \]

\[ +Q_{depl} \]

\[ W_{depl} \]

\[ -Q_{inv} \]

\[ -qN_A \]

\[ \rho \]

\[ \varepsilon \]

\[ x \]

\[ E_C \]

\[ E_F \]

\[ E_V \]
Depletion (Weak Inversion)

\[ V_T > V_G > V_{FB} \]

\[ W_{depl} \]

\[ E_F \uparrow qV_G \uparrow qV_{FB} \]

\[ E_F \quad E_C \quad E_F \quad E_V \]

\[ \varepsilon_{Ox} = \frac{-Q_{dep}}{\varepsilon_{Ox}} \quad \rightarrow \quad V_{Ox} = \frac{-t_{ox} Q_{dep}}{\varepsilon_{Ox}} = \frac{-Q_{dep}}{C_{Ox}} \quad [C/cm^2] \]

\[ V_G = V_{FB} + V_{Ox} + \varphi_s \rightarrow \]

\[ V_G = V_{FB} + \varphi_s + \frac{1}{C_{Ox}} \sqrt{2qN_A \varepsilon_{Si} \varphi_s} \]

\[ \varphi_s < 0 \]

\[ Q_{dep} = -qN_A W_d = \sqrt{2qN_A \varepsilon_{Si} \varphi_s} \]

p-type Si (nMOS)
In Depl

n-type Si (pMOS)
In Depl
Threshold Voltage – Definition!
Threshold Voltage

Definition of Threshold voltage:

\[ V_T = V_G \bigg|_{\varphi_s=2\varphi_F} \]

\[ p_{bulk} = N_A \quad n_{surface} = N_A \]

\[ W_{max} = W_{depl} \bigg|_{\varphi_s=2\varphi_F} = \sqrt{ \frac{2\epsilon_{Si}}{qN_A} (2\varphi_F) } \]

p-type

\[ V_T = V_G \bigg|_{\varphi_s=2\varphi_F} = V_{FB} + 2\varphi_F + \frac{1}{C_{Ox}} \sqrt{2qN_A\epsilon_{Si}(2\varphi_F)} \]

q \varphi_F = \frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right) > 0

n-type

\[ V_G = V_{FB} + 2\varphi_F - \frac{1}{C_{Ox}} \sqrt{2qN_A\epsilon_{Si}|2\varphi_F|} \]

q \varphi_F = -\frac{kT}{q} \ln \left( \frac{N_D}{n_i} \right) < 0
Threshold Voltage vs. Bulk Doping

- $V_T$ vs. $\ln N_D$
- $V_F B$ vs. $\ln N_A$

- p-type
- n-type
Decrease VG (toward more negative values)
→ move the gate energy-bands up, relative to the Si

• Accumulation
  – $V_G > V_{FB}$
  – Electrons accumulate at surface

• Depletion
  – $V_G < V_{FB}$
  – Electrons repelled from surface

• Inversion
  – $V_G < V_T$
  – Surface becomes p-type
Accumulation, Poly Gate

Mobile carriers (holes) accumulate at Si surface
Accumulation, Layer Charge Density

\[ V_G < V_{FB} \]

\[ V_{Ox} \approx V_G - V_{FB} \]

From Gauss’ Law:

\[ \varepsilon_{Ox} = -\frac{Q_{acc}}{\varepsilon_{SiO_2}} \]

\[ V_{Ox} = t_{Ox} \varepsilon_{Ox} = -\frac{Q_{acc}}{C_{Ox}} \]

where \( C_{Ox} \equiv \frac{\varepsilon_{SiO_2}}{t_{Ox}} \quad [F/cm^2] \)

\[ Q_{acc} = -C_{Ox}(V_G - V_{FB}) > 0 \]
Depletion, Poly Gate

1. $V_G < V_{FB}$

2. Si surface is depleted of mobile carriers (holes)

3. Surface charge is due to ionized dopants (acceptors)

$V_G = V_{FB} + V_{Ox} + \phi_s$

$W$ is the width of the depletion layer.
Depletion (Weak Inversion)

\[ V_T > V_G > V_{FB} \]

\[ V_G = V_{FB} + V_{Ox} + \varphi_s \rightarrow V_G = V_{FB} + \varphi_s + \frac{1}{C_{Ox}} \sqrt{2qN_A \varepsilon_{Si} \varphi_s} \]

Solving for \( \varphi_s \):

\[ \varphi_s = \frac{qN_A \varepsilon_{Si}}{2C_{Ox}^2} \left[ \sqrt{1 + \frac{2C_{Ox}^2}{qN_A \varepsilon_{Si}} (V_G - V_{FB}) - 1} \right]^2 \]

\[ Q_{dep} = -qN_A W_d = -\sqrt{2qN_A \varepsilon_{Si} \varphi_s} \]
Strong Inversion

As $V_G$ is increased above $V_T$, the negative charge in the Si is increased by adding mobile electrons (rather than by depleting the Si more deeply), so the depletion width remains $\sim$ constant at $W = W_T$

$$W \approx W_T = \sqrt{\frac{2\varepsilon_{Si}}{qN_A}(2\varphi_F)}$$

$$V_G = V_{FB} + \varphi_s + V_{OX}$$

$$V_G = V_{FB} + 2\varphi_F - \frac{Q_{dep} + Q_{inv}}{C_{OX}}$$

$$V_G = V_{FB} + 2\varphi_F - \frac{\sqrt{2q\varepsilon_{Si}N_A(2\varphi_F)}}{C_{OX}} - \frac{Q_{inv}}{C_{OX}}$$

$$V_G = V_T - \frac{Q_{inv}}{C_{OX}}$$

$\therefore$ $Q_{inv} = -C_{OX}(V_G - V_T)$
\( \varphi_s \) and \( W \) vs. \( V_G \)

**\( \varphi_s \):**
\[
\varphi_s = \frac{q N_A \varepsilon_{Si}}{2 C_0^2} \left[ 1 + \frac{2 C_0^2}{q N_A \varepsilon_{Si}} (V_G - V_{FB}) - 1 \right]^{2} \quad (V_{FB} < V_G < V_T)
\]

**\( W \):**
\[
W = \frac{\varepsilon_{Si}}{C_{ox}} \left[ 1 + \frac{2 C_0^2}{q N_A \varepsilon_{Si}} (V_G - V_{FB}) - 1 \right] \quad (V_{FB} < V_G < V_T)
\]
Total Charge Density in Si, $Q_S$

- $Q_{acc}$:
  
  $Q_{acc} = -C_{ox}(V_G - V_{FB})$

- $Q_{dep}$:
  
  $Q_{dep} = -qN_A W$

- $Q_{inv}$:
  
  $Q_{inv} = -C_{ox}(V_G - V_T)$

$Q_S = Q_{acc} + Q_{dep} + Q_{inv}$

$Q_S$: 

Slope = $-C_{ox}$
MOS C-V Characteristics

\[ C = \left| \frac{dQ_{gate}}{dV_G} \right| = \left| \frac{dQ_S}{dV_G} \right| \]

\( Q_S : \) accumulation, depletion, inversion

\( V_{FE} \)

\( V_T \)

Slope = \(-C_{ox}\)

\( Q_{inv} \)

\( C_{ox} \)

accumulation, depletion, inversion

\( V_{FB} \)

\( V_T \)
Debye Length

- As the gate voltage is varied, incremental charge is added/subtracted to/from the gate and substrate.
- The incremental charges are separated by the gate oxide.

\[
\frac{d^2 \varphi}{dx^2} = -\frac{\rho}{\epsilon} = \frac{q}{\epsilon} (N_D - N_A + p - n)
\]

n-type bulk:
\[
\frac{d^2 \varphi}{dx^2} = \frac{q}{\epsilon} (N_D - n)
\]

\[
\frac{d^2 \varphi}{dx^2} = \frac{q}{\epsilon} \left( N_D - n_i e^{-(\varphi_n - \varphi)/\varphi_{th}} \right)
\]

\[
= \frac{q}{\epsilon} N_D (1 - e^{\varphi/\varphi_{th}})
\]

\[
\cong \frac{q}{\epsilon} N_D \frac{\varphi}{\varphi_{th}} = \frac{\varphi}{L_D^2}
\]

\[
L_D = \sqrt{\frac{\epsilon kT}{q^2 N_D}}
\]
Flat-Band Capacitance

\[ C_{FB} = \frac{C_{OX}C_D}{C_{OX} + C_D} \]

\[ \frac{1}{C_{FB}} = \frac{t_{OX}}{\varepsilon_{OX}} + \frac{L_D}{\varepsilon_{Si}} \]

\[ \frac{1}{C} = \frac{t_{OX}}{\varepsilon_{OX}} + \frac{W_{dep}}{\varepsilon_{Si}} \]

\[ V_G \]

\[ V_{FB} \]
CASE 1: Inversion-layer charge can be supplied/removed quickly enough to respond to changes in the gate voltage.

 Incremental charge is effectively added/subtracted at the surface of the substrate.

Time required to build inversion-layer charge = $\frac{2N_A \tau_0}{n_i}$, where $\tau_0 = \text{minority-carrier lifetime at the surface}$

$$C = \left| \frac{dQ_{\text{inv}}}{dV_G} \right| = C_{Ox}$$
CASE 2: Inversion-layer charge *can not* be supplied/removed quickly enough to respond to changes in the gate voltage.

→ Incremental charge is effectively added/subtracted at depth $W_T$ in the substrate.

\[
\frac{1}{C} = \frac{1}{C_{Ox}} + \frac{1}{C_{dep}} \\
= \frac{1}{C_{Ox}} + \frac{W_T}{\varepsilon_{Si}} \\
= \frac{1}{C_{Ox}} + \sqrt{\frac{2(2\phi_F)}{qN_A\varepsilon_{Si}}} \equiv \frac{1}{C_{min}}
\]
Supply of Inversion Charge

Accumulation:

Inversion:

Depletion:

Case 1

Case 2

n+
Boundary Condition

- MOS transistor at any f,
- MOS capacitor at low f, or quasi-static C-V
- MOS capacitor at high f

Accumulation

Depletion

Inversion

$V_{FB}$

$V_T$

$C_{FB}$

$C_{min}$

$C_{Ox}$

$C$
Deep Depletion

If $V_G$ is scanned quickly, $Q_{inv}$ cannot respond to the change in $V_G$. The increase in substrate charge density $Q_s$ must then come from an increase in depletion charge density $Q_{dep}$.

$\Rightarrow$ depletion depth $W$ increases as $V_G$ increases.

$\Rightarrow$ $C$ decreases as $V_G$ increases.
In real MOS devices, there is always some charge in the oxide and at the Si/oxide interface.

In the oxide:

1. Trapped charge $Q_{ox}$
   High-energy electrons and/or holes injected into oxide

2. Mobile charge $Q_{M}$
   Alkali-metal ions, which have sufficient mobility to drift in oxide under an applied electric field

At the interface:

1. Fixed charge $Q_{F}$
   Excess Si (+)

2. Trapped charge $Q_{IT}$
   Dangling bonds
Effect of Oxide Charges

In general, charges in the oxide cause a shift in the gate voltage required to reach the threshold condition:

\[ \Delta V_T = -\frac{1}{\varepsilon_{SiO_2}} \int_0^{t_{Ox}} x \rho_{ox}(x) dx \]

(x defined to be 0 at metal-oxide interface)

In addition, they may alter the field-effect mobility of mobile carriers (in a MOSFET) due to Coulombic scattering.
Fixed Oxide Charges $Q_F$

$$V_{FB} = \varphi_{ms} - \frac{Q_F}{C_{Ox}}$$
From a single C-V measurement, we can extract much information about the MOS device.

Suppose we know that the gate-electrode material is heavily doped n-type poly-Si ($\varphi_M=4.05\text{eV}$), and that the gate dielectric is SiO2 ($\varepsilon_r = 3.9$):

- From $C_{\text{max}} = C_{Ox}$ we determine the oxide thickness $x_0$
- From $C_{\text{min}}$ and $C_{Ox}$ we determine substrate doping (by iteration)
- From substrate doping and $C_{Ox}$ we calculate the flat-band capacitance $C_{FB}$
- From the C-V curve, we can find
- From $\varphi_M$, $\varphi_S$, $C_{Ox}$, and $V_{FB}$ we can determine $Q_F$
Determination of $\varphi_M$ and $Q_F$:

Measure C-V characteristics of capacitors with different oxide thicknesses. Plot $V_{FB}$ as a function of $x_0$:

$$V_{FB} = \varphi_{ms} - \frac{Q_F}{\varepsilon_{SiO_2}} t_{ox}$$
Odd shifts in C-V characteristics were once a mystery:

$$\Delta V_{FB} = -79V$$

Source of problem: Mobile charge moving to/away from interface, changing charge centroid

$$\Delta V_{FB} = -\frac{Q_M}{C_{OX}}$$
Interface Traps

Traps cause “sloppy” C-V and also greatly degrade mobility in channel

$$\Delta V_G = -\frac{Q_{IT}(\varphi_s)}{C_{Ox}}$$
A heavily doped film of polycrystalline silicon (poly-Si) is typically employed as the gate-electrode material in modern MOS devices.

There are practical limits to the electrically active dopant concentration (usually less than $1\times10^{20} \text{ cm}^{-3}$)

$\Rightarrow$ The gate must be considered as a semiconductor, rather than a metal
How can gate depletion be minimized?
Gate Depletion Effect

Gauss’s Law dictates:

\[ W_{poly} = \frac{\varepsilon_{Ox} \varepsilon_{ox}}{q N_{poly}} \]

\( t_{Ox} \) is effectively increased:

\[ C = \left( \frac{1}{C_{Ox}} + \frac{1}{C_{poly}} \right)^{-1} = \left( \frac{t_{Ox}}{\varepsilon_{SiO_2}} + \frac{W_{poly}}{\varepsilon_{Si}} \right)^{-1} \]

\[ = \frac{\varepsilon_{SiO_2}}{t_{Ox} + \frac{1}{3} W_{poly}} \]

\[ Q_{inv} = C_{Ox} (V_G - V_{poly} - V_T) \]

\[ Q_{inv} = \frac{\varepsilon_{SiO_2}}{t_{Ox} + \frac{1}{3} W_{poly}} (V_G - V_T) \]
The average inversion-layer location below the Si/SiO2 interface is called the inversion-layer thickness, $T_{inv}$. 
Effective Oxide Thickness, $T_{Oxe}$

$$T_{Oxe} = t_{Ox} + \frac{1}{3} W_{poly} + \frac{1}{3} T_{inv}$$ @ $V_G = V_{DD}$

$(V_G + V_T)/T_{Ox}$ can be shown to be the average electric field in the inversion layer. $T_{inv}$ of holes is larger than that of electrons because of the difference in effective masses.
Effective Oxide Capacitance, $C_{Oxe}$

\[ Q_{inv} = C_{Oxe} (V_G - V_T) \]

\[ T_{Oxe} = t_{Ox} + \frac{1}{3} W_{poly} + \frac{1}{3} T_{inv} \]
MOS Cap: Equivalent Circuit in Depletion & Inversion

General case for both depletion and inversion regions.

In the depletion regions

\[ V_G \approx V_T \]

Strong inversion
In modern IC fabrication processes, the threshold voltages of MOS transistors are adjusted by ion implantation:

- A relatively small dose $N_I$ (units: ions/cm$^2$) of dopant atoms is implanted into the near-surface region of the semiconductor.
- When the MOS device is biased in depletion or inversion, the implanted dopants add to the dopant-ion charge near the oxide-semiconductor interface.

$$\Delta V_T = -\frac{qN_I}{C_{ox}}$$

- $N_I > 0$ for donor atoms
- $N_I < 0$ for acceptor atoms
**Dynamic $V_T$ Adjustment Bulk Voltage**

\[ V_G > V_T \]

\[ V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{ox}} \]

\[ q\varphi_s \]

\[ qV_G \]

\[ E_F \]

\[ D \]

\[ V_{DS} \approx 0 \]

\[ V_{GS} > V_T \]

\[ V_{B} \]

\[ V_{BC} \]

\[ E_C \]

\[ E_F \]

\[ E_V \]

\[ q(V_{BC} + 2\varphi_F) \]

\[ qV_{BC} \]

\[ V_T = V_{FB} + V_C + 2\varphi_F + \frac{Q_{depl}}{C_{ox}} \]

\[ Q_{depl} = \sqrt{2qN_A\varepsilon_{ox}(V_{BC} + 2\varphi_F)} \]
CCD Imager and CMOS Imager

Deep depletion, $Q_{inv} = 0$

Exposed to light
Boundary Condition
MOSFET I-V Curve

\[ V_T = V_{FB} + V_C(y) + 2\phi_F + \frac{1}{C_{Ox}} \sqrt{2qN_A\varepsilon_{ox}(V_{CB} + 2\phi_F)} \]

\[ Q_{depl}(y) \]

\[ V_c(y) \]

\[ \begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases} \]
MOSFET I-V Curve

\[ V_T(y) = V_{FB} + V_C(y) + 2\phi_F + \frac{1}{C_{ox}} \sqrt{2qN_A\varepsilon_{ox}(V_{CB} + 2\phi_F)}. \]

\[ Q_{inv} = -C_{ox}(V_G - V_T(y)) \]

\[ Q_{inv} = -C_{ox}\left(V_G - V_{FB} - V_C(y) - 2\phi_F - \frac{Q_{depl}(y)}{C_{ox}}\right) \]

Depletion Region Approximation:

\[ Q_{inv}(y) \quad \text{but} \quad Q_{depl}(y) \approx Q_{depl}(0) \]

\[ Q_{depl}(y) \approx \sqrt{2qN_A\varepsilon_{ox}(V_{SB} + 2\phi_F)} \]

\[ Q_{inv} = -C_{ox}\left(V_G - V_{FB} - V_S - 2\phi_F - \frac{Q_{depl}(0)}{C_{ox}} + V_s - V_C(y)\right) \]

\[ Q_{inv}(y) = -C_{ox}(V_G - V_T(0) + V_s - V_C(y)) \]

\[ V_C(y) \]

\[ \begin{cases} V_C(0) = V_S \\ V_C(L) = V_D \end{cases} \]
MOSFET I-V Curve

\[ Q_{\text{inv}}(y) = -C_{\text{Ox}}(V_G - V_T(0) + V_s - V_C(y)) \]

Simply call \( V_T(0) \) as \( V_T \)

\[ Q_{\text{inv}}(y) = -C_{\text{Ox}}(V_G - V_T + V_s - V_C(y)) \]

\[ dV_C(y) = I_{DS} \cdot dR = I_{DS} \frac{dy}{\sigma W t_{\text{inv}}} = \frac{I_{DS} dy}{(q \mu_{\text{eff}} n) W t_{\text{inv}}} = \frac{I_{DS} dy}{(q n t_{\text{inv}}) \mu_{\text{eff}} W} \]

\[ \int_0^L I_{DS} dy = \int_{V_S}^{V_D} -\mu_{\text{eff}} W Q_{\text{inv}}(y) dV_C \]

\[ I_{DS}L = \mu_{\text{eff}} W \int_{V_S}^{V_D} [C_{\text{Ox}}(V_G - V_T + V_s - V_C(y))] dV_C \]

\[ I_{DS} = \frac{W}{L} \mu_{\text{eff}} C_{\text{Ox}}(V_{GS} - V_T - \frac{1}{2}V_{DS})V_{DS} \]

\[ \left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_{DS_{\text{sat}}}} = 0 \]

\[ V_C(y) \]

\[ \begin{cases} V_C(0) = V_S \\ V_C(L) = V_D \end{cases} \]
MOSFET I-V Curve

\[ I_{DS} = \begin{cases} 
\frac{W}{L} \mu_{eff} C_{Ox} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \\
\frac{W}{2L} \mu_{eff} C_{Ox} (V_{GS} - V_T)^2 
\end{cases} \]

Linear
\[ V_{DS} < V_{DS_{sat}} \]
Saturation
\[ V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \]

\[ R_{DS} = \left( \frac{\partial I_{DS}}{\partial V_{DS}} \bigg|_{V_{DS}=0} \right)^{-1} = \left( \frac{W}{L} \mu_{eff} C_{Ox} (V_{GS} - V_T) \right)^{-1} \]
Field-Effect Mobility, $\mu_{\text{eff}}$

Scattering mechanisms:
- Columbic scattering
- phonon scattering
- surface roughness scattering

Effective vertical electric field in the inversion layer for NMOS

Effective vertical electric field in the inversion layer for PMOS
MOSFET Saturation Region of Operation

As $V_D$ is increased above $V_G-V_T$, the length $\Delta L$ of the “pinch-off” region increases. The voltage applied across the inversion layer is always $V_{D_{Sat}}=V_{GS}-V_T$, and so the current saturates.

If $\Delta L$ is significant compared to $L$, then $I_{DS}$ will increase slightly with increasing $V_{DS}>V_{D_{Sat}}$ due to “channel-length modulation”
“Square Law Theory”?

\[
I_{DS} = \begin{cases} 
\frac{W}{L} \mu_{eff} C_{OX} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \\
\frac{W}{2L} \mu_{eff} C_{OX} (V_{GS} - V_T)^2
\end{cases}
\]

\[V_{DS} < V_{DS_{sat}} \quad \text{Linear}\]

\[V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \quad \text{Saturation}\]

Depletion Region Approximation:

\[Q_{inv}(y) \quad \text{but} \quad Q_{depl}(y) \approx Q_{depl}(0)\]

\[Q_{depl}(y) \approx \sqrt{2qN_A \varepsilon_{OX} (V_{SB} + 2 \varphi_F)}\]

\[V_T(y) = V_{FB} + V_C(y) + 2 \varphi_F + \frac{1}{C_{OX}} \sqrt{2qN_A \varepsilon_{OX} (V_{CB} + 2 \varphi_F)}\]

\[Q_{inv}(y) = -C_{OX} (V_G - V_T(0) + V_S - V_C(y))\]

\[Q_{depl}(y) > Q_{depl}(0) \quad Q'_{inv}(y) < Q_{inv}(y) \quad I'_{DS} < I_{DS}\]
Modified (Bulk-Charge) I-V Model

Linear: \( V_{DS} < V_G - V_T \)

\[
I_{DS} = \frac{W}{L} \cdot \mu_{eff} C_{OX} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS}
\]

Saturation: \( V_{DS} > V_G - V_T \)

\[
I_{DS} = \frac{W}{2L} \cdot \mu_{eff} C_{OX} (V_{GS} - V_T)^2
\]

Typically: \( 1.1 < m < 1.4 \)

Bulk charge factor:

\[
m = 1 + \frac{C_{dep}}{C_{OX}} = 1 + \frac{3t_{OX}}{W_T}
\]
The Body Effect

Note that $V_T$ is a function of $V_{SB}$:

$$V_T = V_{FB} + 2\varphi_F + \frac{1}{C_{Ox}} \sqrt{2qN_A\varepsilon_{Ox} (V_{SB} + 2\varphi_F)}$$

$$V_T = V_{T0} + \frac{1}{C_{Ox}} \sqrt{2qN_A\varepsilon_{Ox}} \left(\frac{1}{\sqrt{V_{SB} + 2\varphi_F}} - \sqrt{2\varphi_F}\right)$$

$$= V_{T0} + \gamma \left(\sqrt{V_{SB} + 2\varphi_F} - \sqrt{2\varphi_F}\right)$$

$$\gamma = \frac{1}{C_{Ox}} \sqrt{2qN_A\varepsilon_{Ox}}$$

where $\gamma$ is the \textit{body effect parameter}

When the source-body pn junction is reverse-biased, $|V_T|$ is increased. Usually, we want to minimize $g$ so that $I_{Dsat}$ will be the same for all transistors in a circuit.
The Body Effect

\[
\begin{array}{ccc}
A & B & F \\
0 & 0 & 1 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

\((*)\)

\(V_A = V_{DD}\)

\(V_B = 0\)

\(V_{(1)} = V_{DD}\)

\(V_{T_a} > V_{T_b}\)
\( \lambda: \text{Channel Length Modulation Parameter} \)

\[ I_{D_{sat}} \propto \frac{1}{L - \Delta L} = \frac{1}{L} \left( 1 + \frac{\Delta L}{L} \right) \]

\[ \Delta L \propto V_{DS} - V_{DS_{sat}} \]

\[ \frac{\Delta L}{L} \propto \lambda (V_{DS} - V_{DS_{sat}}) \]

\[ \lambda \sim \frac{1}{L} \]

\[ I_{DS} = \frac{W}{2mL} \mu_{eff} C_{ox} (V_{GS} - V_T)^2 \left( 1 + \lambda (V_{DS} - V_{DS_{sat}}) \right) \]

\[ V_c = V_G - V_T = V_{DS_{sat}} \]

\[ V_{DS} > V_{GS} - V_T \]
MOSFET: Small Signal Model

\[ g_m = \frac{W}{2mL} \mu_{\text{eff}} C_{\text{ox}} (V_{GS} - V_T) \]

\[ g_d = \lambda I_{D\text{sat}} \]

cut-off frequency:

\[ f_{\text{max}} \rightarrow \frac{g_m}{2\pi C_{\text{ox}}} \propto \frac{1}{L} \]
Sub-Threshold Current

\[ V_G = 0 \quad V_D = 0 \]

\[ V_G \geq V_T \quad V_D = 0 \]

\[ V_G \geq V_T \quad V_D > 0 \]
Sub-Threshold Current

Similarly:

\[ m = 1 + \frac{C_{\text{depl}}}{C_{\text{ox}}} = 1 + \frac{3t_{\text{ox}}}{W_{T}} \]

\[ V_c = \frac{C_{\text{ox}}}{C_{\text{ox}} + C_{\text{depl}}} V_{GB} \]

\[ I_{DS} = \frac{W}{L} \mu_{\text{eff}} C_{\text{ox}} (m - 1) \left( \frac{kT}{q} \right)^2 e^{q(V_G - V_T)/mkT} \left( 1 - e^{qV_D/kT} \right) \]

\[ S = \left( \frac{d \log_{10} I_{DS}}{dV_{GS}} \right)^{-1} = \frac{kT}{q} \ln 10 \left( 1 + \frac{C_{\text{depl}}}{C_{\text{ox}}} \right) \]

\[ S^{-1} = 60 \text{ mV} \]

\[ m \begin{cases} N_A \downarrow \Rightarrow C_{\text{depl}} \downarrow \Rightarrow \text{retrograde} \\ t_{\text{ox}} \downarrow \Rightarrow C_{\text{ox}} \uparrow \\ T \downarrow \Rightarrow \text{low-temperature} \end{cases} \]

\[ V_G < V_T \]

\[ V_D > 0 \]

\[ \text{Source} \]

\[ \text{Drain} \]

\[ \text{ln} I_D \]

\[ I_{\text{off}} \]

\[ V_{T_1} \]

\[ V_{T_2} \]
Sub-Threshold Current

![Graph showing sub-threshold current characteristics.](image)
Main MOSFET Parameters

- Log($I_{\text{drain}}$)
- $V_{\text{gate}}$

MOSFET switch Ideal switch

ON state Current

OFF state Current

OFF State Current (Thermal)

Threshold ($V_{th}$)

3 main parameters:

1. Threshold Voltage
2. Ion (=speed)
3. Ioff (=stand-by power)
1. The PMOSFET turns on when $V_{GS} < V_T$
   - Holes flow from SOURCE to DRAIN
   $\Rightarrow$ DRAIN is biased at a **lower** potential than the SOURCE

2. In a CMOS technology, the PMOS & NMOS threshold voltages are usually symmetric about 0, *i.e.* $V_{Tp} = -V_{Tn}$
PMOSFET I-V

Linear

\[ 0 < |V_{DS}| < \frac{|V_G - V_T|}{m} \]

\[ I_{DS} = -\frac{W}{L} \mu_{eff} C_{Ox} (V_{GS} - V_{TP} - \frac{m}{2}V_{DS})V_{DS} \]

Saturation

\[ |V_{DS}| > \frac{|V_G - V_T|}{m} \]

\[ I_{DS} = I_{DSat} = -\frac{W}{2mL} \mu_{eff} C_{Ox} (V_{GS} - V_T)^2 \]

bulk-charge factor

\[ m = 1 + \frac{C_{depl}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T} \]
MOSFETs have been steadily miniaturized over time
1970s: ~ 10 mm
Today: ~30 nm

Reasons:
Improved circuit operating speed
Increased device density --> lower cost per function

As MOSFET lateral dimensions (e.g. channel length $L$) are reduced:
- $I_{D_{sat}}$ increases $\rightarrow$ decreased effective “$R$”
- gate and junction areas decrease $\rightarrow$ decreased load “$C$”
  $\rightarrow$ faster charging/discharging (i.e. $t_d$ is decreased)

\[
\tau = \frac{C_{ox} \cdot V_{dd}}{I_{on}}
\]
MOSFET dimensions and the operating voltage ($V_{DD}$) each are scaled by the same factor $k > 1$, so that the electric field remains unchanged.
# Constant-Field Scaling Benefits

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Multiplication factor ((k&gt;1))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scaling assumptions</strong></td>
<td></td>
</tr>
<tr>
<td>Device dimensions ((t_{ox}, L, W, r_j))</td>
<td>(1/k)</td>
</tr>
<tr>
<td>Doping concentration ((N_A, N_D))</td>
<td>(k)</td>
</tr>
<tr>
<td>Voltage ((V))</td>
<td>(1/k)</td>
</tr>
<tr>
<td><strong>Derived scaling behavior of device parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Electric field ((\mathcal{E}))</td>
<td>1</td>
</tr>
<tr>
<td>Carrier velocity ((\nu))</td>
<td>1</td>
</tr>
<tr>
<td>Depletion-layer width ((W_D))</td>
<td>(1/k)</td>
</tr>
<tr>
<td>Capacitance ((C = \varepsilon A/t))</td>
<td>(1/k)</td>
</tr>
<tr>
<td>Inversion charge density ((Q_{inv}))</td>
<td>1</td>
</tr>
<tr>
<td>Current drift ((I))</td>
<td>(1/k)</td>
</tr>
<tr>
<td>Channel resistance ((R_{ch}))</td>
<td>1</td>
</tr>
<tr>
<td><strong>Derived scaling behavior of circuit parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Circuit delay time ((\tau \sim CV/I))</td>
<td>(1/k)</td>
</tr>
<tr>
<td>Power diss. per circuit ((P \sim VI))</td>
<td>(1/k^2)</td>
</tr>
<tr>
<td>Power-delay product per circuit ((P\tau))</td>
<td>(1/k^3)</td>
</tr>
<tr>
<td>Circuit density ((\propto 1/A))</td>
<td>(k^2)</td>
</tr>
<tr>
<td>Power density ((P/A))</td>
<td>1</td>
</tr>
</tbody>
</table>
Since $V_T$ cannot be scaled down aggressively, the operating voltage ($V_{DD}$) has not been scaled down in proportion to the MOSFET channel length:

<table>
<thead>
<tr>
<th>Feature Size ($\mu m$)</th>
<th>Power-Supply Voltage ($V$)</th>
<th>Gate Oxide Thickness ($Å$)</th>
<th>Oxide Field ($MV/cm$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>350</td>
<td>1.1</td>
</tr>
<tr>
<td>1.2</td>
<td>5</td>
<td>250</td>
<td>2.0</td>
</tr>
<tr>
<td>0.8</td>
<td>5</td>
<td>180</td>
<td>2.8</td>
</tr>
<tr>
<td>0.5</td>
<td>3.3</td>
<td>120</td>
<td>2.8</td>
</tr>
<tr>
<td>0.35</td>
<td>3.3</td>
<td>100</td>
<td>3.3</td>
</tr>
<tr>
<td>0.25</td>
<td>2.5</td>
<td>70</td>
<td>3.6</td>
</tr>
</tbody>
</table>
# MOSFET Scaling: Generalized Approach

Electric field intensity increases by a factor $\alpha > 1$

$N_{body}$ must be scaled up by $\alpha$ to suppress short-channel effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Multiplication factor ($k&gt;1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scaling assumptions</strong></td>
<td></td>
</tr>
<tr>
<td>Device dimensions ($t_{Ox}, L, W, r_j$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td>Doping concentration ($N_A, N_D$)</td>
<td>$\alpha k$</td>
</tr>
<tr>
<td>Voltage ($V$)</td>
<td>$\alpha/k$</td>
</tr>
<tr>
<td><strong>Derived scaling behavior of device parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Electric field ($\mathcal{E}$)</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Depletion-layer width ($W_D$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td>Capacitance ($C = \epsilon A / t$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td>Inversion charge density ($Q_{inv}$)</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Carrier velocity ($v$)</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Current drift ($I$)</td>
<td>$1$</td>
</tr>
<tr>
<td>Long ch.</td>
<td>$\alpha^2/k$</td>
</tr>
<tr>
<td>Vel Sat.</td>
<td>$\alpha/k$</td>
</tr>
<tr>
<td><strong>Derived scaling behavior of circuit parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Circuit delay time ($\tau \sim CV / I$)</td>
<td>$1/\alpha k$</td>
</tr>
<tr>
<td>Power diss. per circuit ($P \sim VI$)</td>
<td>$\alpha^3 / k^2$</td>
</tr>
<tr>
<td>Power-delay product per circuit ($P\tau$)</td>
<td>$\alpha^2 / k^3$</td>
</tr>
<tr>
<td>Circuit density ($\propto 1/A$)</td>
<td>$k^2$</td>
</tr>
<tr>
<td>Power density ($P/A$)</td>
<td>$\alpha^3$</td>
</tr>
<tr>
<td></td>
<td>$\alpha^2$</td>
</tr>
</tbody>
</table>
Velocity Saturation

Velocity saturation limits $I_{Dsat}$ in sub-micron MOSFETS

Simple model:

$$v = \begin{cases} 
\frac{\mu \varepsilon}{1 + \frac{\varepsilon}{\varepsilon_{sat}}} & \text{for } \varepsilon < \varepsilon_{sat} \\
v_{sat} & \text{for } \varepsilon \geq \varepsilon_{sat} 
\end{cases}$$

$$\mu \varepsilon_{sat} = 2v_{sat}$$

$$v_{sat} = \begin{cases} 
8 \times 10^6 \text{ cm/s for } e^- \text{in Si} \\
6 \times 10^6 \text{ cm/s for } h^+ \text{in Si}
\end{cases}$$

If $\varepsilon < \varepsilon_{sat}$:

$$\mu \rightarrow \frac{\mu}{1 + \frac{\varepsilon}{\varepsilon_{sat}}}$$
MOSFET I-V with Velocity Saturation

In linear region:
\[
I_D = \frac{W}{L} \frac{\mu_{\text{eff}} C_{\text{ox}}}{1 + \frac{V_{DS}}{L \varepsilon_{\text{sat}}}} (V_{GS} - V_T - \frac{m}{2} V_{DS}) \frac{V_{DS}}{L \varepsilon_{\text{sat}}}
\]

MOSFET is Long channel if \( L \varepsilon_{\text{sat}} \gg V_{GS} - V_T \)

\[
\frac{1}{V_{D_{\text{sat}}}} = \frac{m}{V_{GS} - V_T} + \frac{1}{L \varepsilon_{\text{sat}}}
\]

\[
V_{GS} = 1.8 \text{ V}, t_{Ox} = 3 \text{ nm}, \quad V_T = 0.25 \text{ V}, W/T = 45 \text{ nm}
\]

\[
\mu_{\text{eff}} = 200 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}, m = 1 + 3 T_{\text{ox}}/W_T = 1.2
\]

\[
\varepsilon_{\text{sat}} = 2 \nu_{\text{sat}} / \mu_{\text{eff}} = 8 \times 10^4 \text{ V/cm}
\]

\[
L = 10 \mu m \quad \rightarrow \quad V_{D_{\text{sat}}} = 1.3 \text{ V}
\]

\[
L = 1 \mu m \quad \rightarrow \quad V_{D_{\text{sat}}} = 1.1 \text{ V}
\]

\[
L = 100 \text{ nm} \quad \rightarrow \quad V_{D_{\text{sat}}} = 0.5 \text{ V}
\]

\[
L = 30 \text{ nm} \quad \rightarrow \quad V_{D_{\text{sat}}} = 0.2 \text{ V}
\]
\( I_{D_{\text{sat}}} \) with Velocity Saturation

In saturation region:

\[
V_{DS} \rightarrow V_{GS} - V_T \\
I_{D_{\text{sat}}} = \frac{W}{2mL} \frac{\mu_{\text{eff}} C_{OX}}{1 + \frac{V_{GS} - V_T}{L \varepsilon_{sat}}} (V_{GS} - V_T)^2 = \frac{I_{D_{\text{sat,LongChannel}}}}{1 + \frac{V_{GS} - V_T}{L \varepsilon_{sat}}}
\]

Very short channel length: \( L \varepsilon_{sat} \ll V_{GS} - V_T \)

\[
I_{D_{\text{sat}}} = \frac{W}{2m} \mu_{\text{eff}} \varepsilon_{sat} C_{OX} (V_{GS} - V_T) = \frac{W}{m} v_{sat} C_{OX} (V_{GS} - V_T)
\]

- \( I_{D_{\text{sat}}} \) is proportional to \( V_{GS} - V_T \) rather than \( (V_{GS} - V_T)^2 \)
- \( I_{D_{\text{sat}}} \) is not dependent on \( L \)

To improve modern MOSFETs:

\[
I_{ON} \uparrow \quad C_{OX} \uparrow \quad \text{high-k dielectric} \\
v_{sat} \uparrow \quad \text{strained Si}
\]
Short-channel NMOSFET:

- $I_{D_{\text{sat}}}$ is proportional to $V_{GS} - V_{\text{Tn}}$ rather than $(V_{GS} - V_{\text{Tn}})^2$
- $V_{D_{\text{sat}}}$ is lower than for long-channel MOSFET
- Channel-length modulation is apparent
Velocity Overshoot

When L is comparable to or less than the mean free path, some of the electrons travel through the channel without experiencing a single scattering event

→ projectile-like motion ("ballistic transport")
The Short Channel Effect (SCE)

“$V_T$ roll-off”

$|V_T|$ decreases with L
Effect is exacerbated by high values of $|V_{DS}|$

This effect is undesirable (i.e. we want to minimize it!) because circuit designers would like $V_T$ to be invariant with transistor dimensions and bias condition.
Before an inversion layer forms beneath the gate, the surface of the Si underneath the gate must be depleted (to a depth $W_T$).

The source & drain pn junctions assist in depleting the Si underneath the gate. Portions of the depletion charge in the channel region are balanced by charge in S/D regions, rather than by charge on the gate. Less gate charge is required to invert the semiconductor surface (i.e. $|V_T|$ decreases).

$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$
**$V_T$ Roll-Off: First-Order Model**

\[ V_T = V_{FB} + 2\phi_F + \frac{Q_{depl}}{C_{OX}} \]

\[ \Delta V_T \equiv |V_T| - |V_{T Long Channel}| \]

\[ \Delta V_T \propto \frac{qN_A}{C_{OX}} W_T \left(1 - \frac{L + L'}{2L}\right) \]

\[ \Delta V_T = \frac{-qN_A W_T r_j}{C_{OX}} \sqrt{1 + \frac{2W_T}{r_j} - 1} \]

Minimize $\Delta V_T$ by
- reducing $t_{Ox}$
- reducing $r_j$
- increasing $N_A$ (trade-offs: degraded $\mu_{eff}, m$)

MOSFET vertical dimensions should be scaled along with horizontal dimensions!
Drain Induced Barrier Lowering (DIBL)

As the source and drain get closer, they become electrostatically coupled, so that the drain bias can affect the potential barrier to carrier diffusion at the source junction.

\[ V_T \] decreases (i.e. OFF state leakage current increases)

\[ \rightarrow V_T \text{ decreases (i.e. OFF state leakage current increases)} \]
Punchthrough

A large drain bias can cause the drain-junction depletion region to merge with the source-junction depletion region, forming a sub-surface path for current conduction.

\[ I_{\text{Dsat}} \text{ increases rapidly with } V_{\text{DS}} \]

This can be mitigated by doping the semiconductor more heavily in the sub-surface region, i.e. using a “retrograde” doping profile.
Source and Drain (S/D) Structure

To minimize the short channel effect and DIBL, we want shallow (small $r_j$) S/D regions – but the parasitic resistance of these regions increases when $r_j$ is reduced.

$$R_{source}, R_{drain} \propto \rho/Wr_j$$

where $\rho = $ resistivity of the S/D regions

Shallow S/D “extensions” may be used to effectively reduce $r_j$ with a relatively small increase in parasitic resistance.
The lateral electric field peaks at the drain end of the channel. \( \mathcal{E}_{\text{peak}} \) can be as high as \( 10^6 \) V/cm.

High E-field causes problems:
- Damage to oxide interface & bulk
- (trapped oxide charge \( \rightarrow V_T \) shift)
- Substrate current due to impact ionization:
Lightly Doped Drain (LDD) Structure

Lower pn junction doping results in lower peak E-field

✓ “Hot-carrier” effects are reduced

✗ Parasitic resistance is increased
Parasitic Source-Drain Resistance

For short-channel MOSFET, \( I_{D,\text{sat}0} \propto V_{GS} - V_T \), so that

\[
I_{D,\text{sat}} = \frac{I_{D,\text{sat}0}}{1 + \frac{I_{D,\text{sat}}}{} R_S \frac{R_D}{V_{GS} - V_T}}
\]

\( \rightarrow I_{D,\text{sat}} \) is reduced by \(~15\%\) in a 0.1 mm MOSFET.

\( V_{D,\text{sat}} \) is increased to \( V_{D,\text{sat}0} + I_{D,\text{sat}} (R_S + R_D) \)
Tunneling Into and Through SiO$_2$
Defect Generation Caused By Tunneling Current

trapping of tunneling electrons. As electrons are trapped, the oxide field near the cathode (electron source) is decreased, while the oxide field near the anode (electron sink) is increased.

generation of an electron-hole pair in the anode by a tunneling electron. The hole thus generated can then be injected (by tunneling in this example) into the oxide layer.

trapping of holes in the oxide layer. The trapped holes enhance the electric field near the cathode, and decrease the electric field near the anode.