Session 2: Solid State Devices

CMOS Technology Review

Outline

1. I
2.
3.
4.
5.

- A
  - B
  - C
  - D
  - E
- F
  - G
- H
- I
- J

Outline

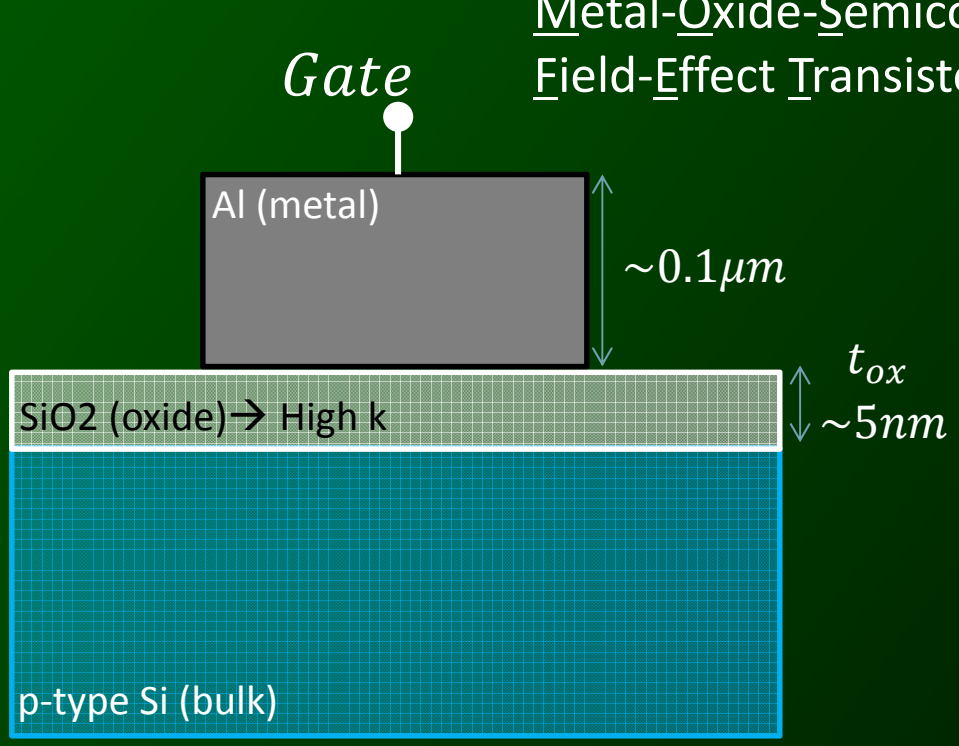
1. I
2.
3.
4.
5.

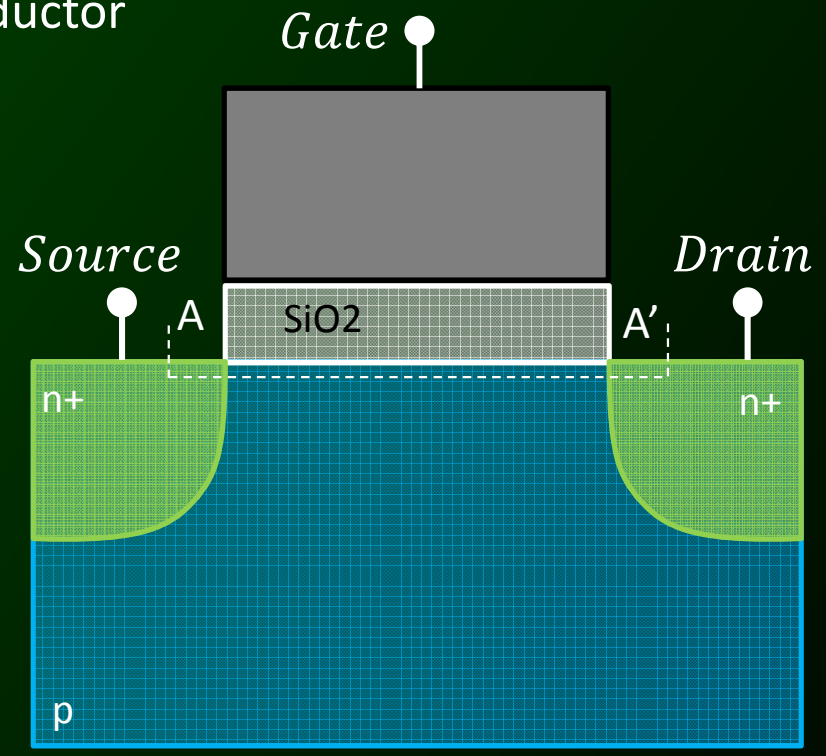
- **MOSFET Basics**
  Ideal MOSFET physics
  Main parameters: threshold, leakage and speed
  Which FET for what application?
  Scaling theory and good design rules of CMOS Devices
- **The Real World**
  Threshold voltage control limitations
  Gate oxide leakage and capacitance scaling
- **Technological Solution?**
  Gate alternative: High-K and Metal Gate
  Channel engineering: Strained-Si
  Alternative devices and substrates
  Basic logic functions

MOSCap → MOSFET

1. I
2.
3.
4.
5.

Metal-Oxide-Semiconductor
Field-Effect Transistor:

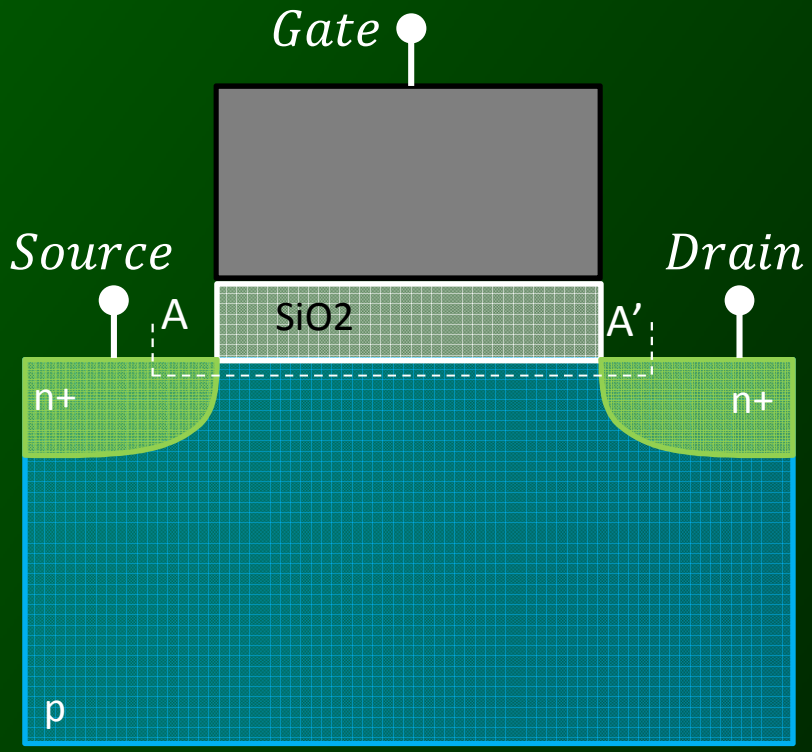


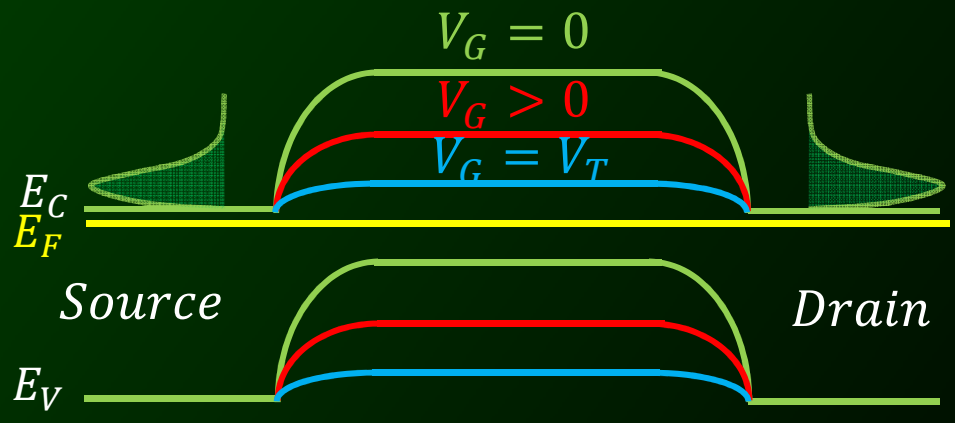


Drift current flowing between 2 doped regions (“source” & “drain”) is modulated by varying the voltage on the “gate” electrode.

Qualitative Theory of the NMOSFET

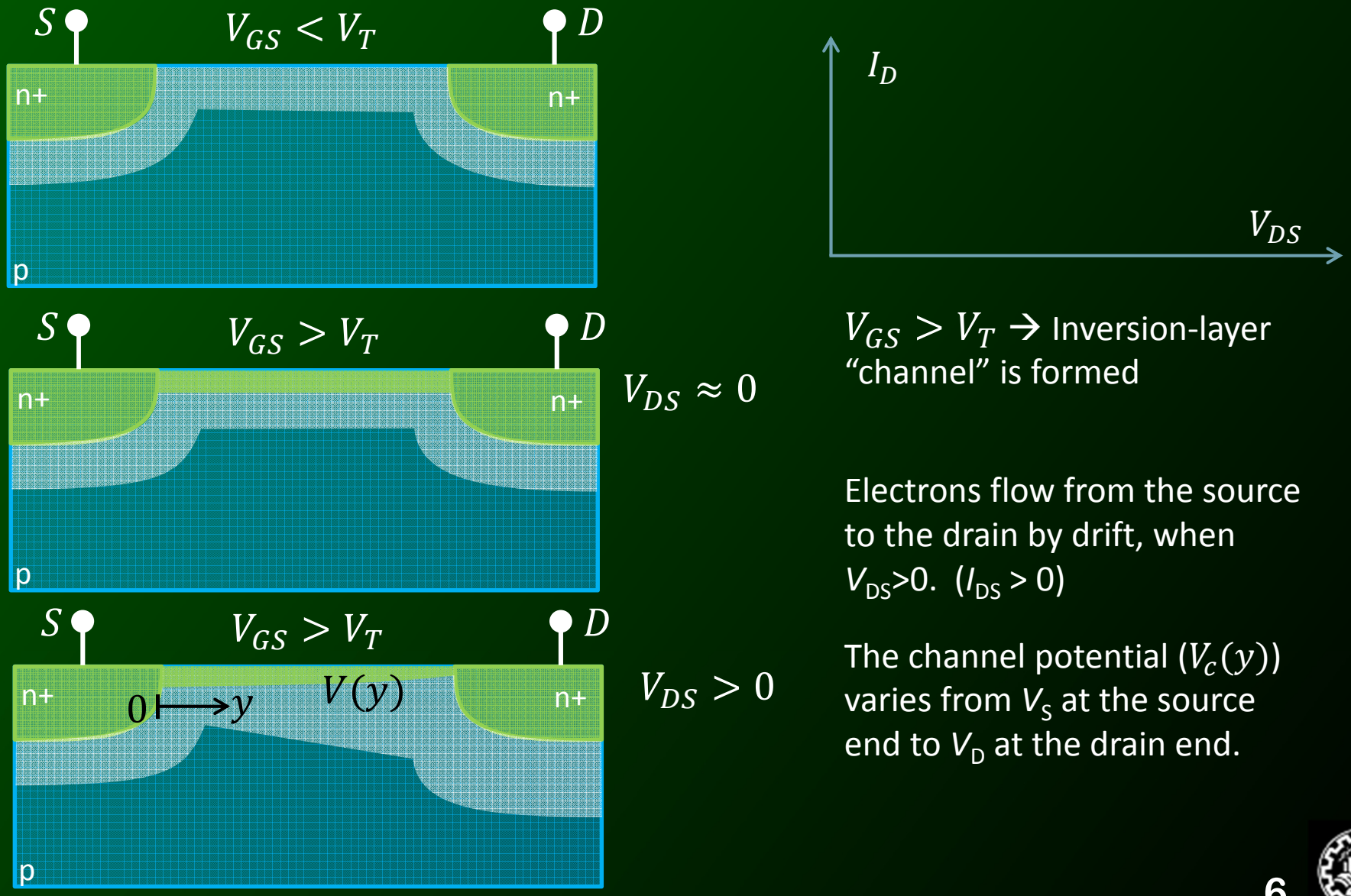
1. I
2.
3.
4.
5.





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



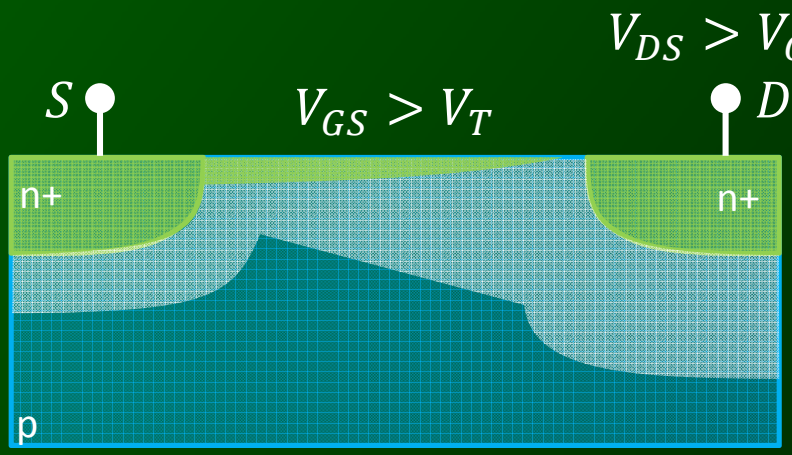
$V_{GS} > V_T$ → Inversion-layer "channel" is formed

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

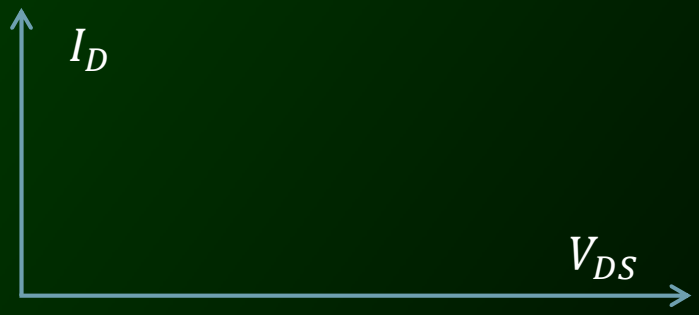
The channel potential ($V_c(y)$) varies from V_S at the source end to V_D at the drain end.

Qualitative Theory of the NMOSFET

1. I
2.
3.
4.
5.



$$V_{DS_{sat}} = V_{GS} - V_T$$



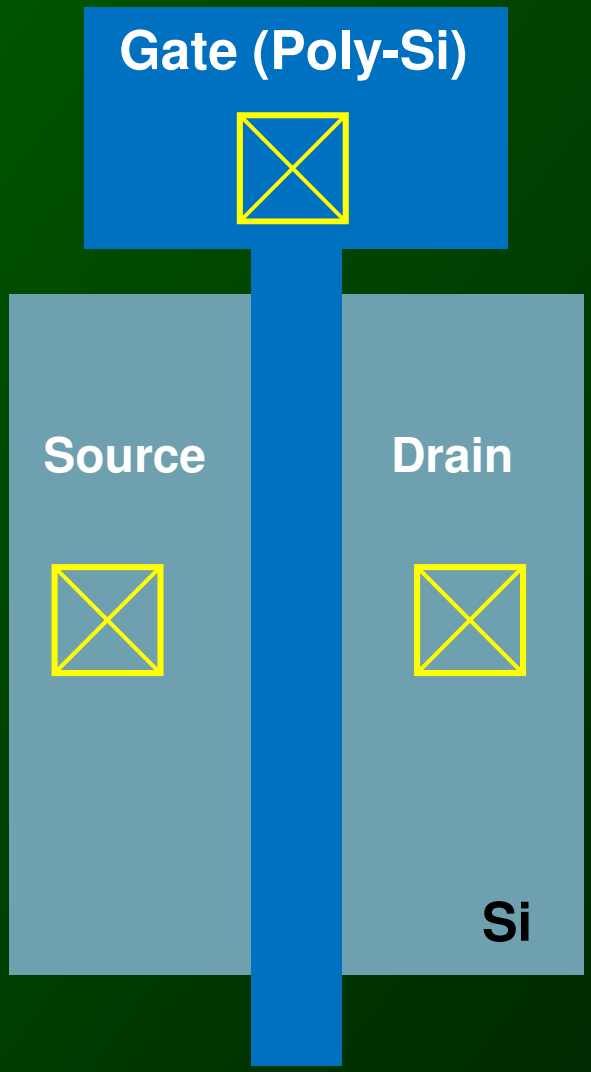
$V_{GS} > V_T$ → Inversion-layer "channel" is formed

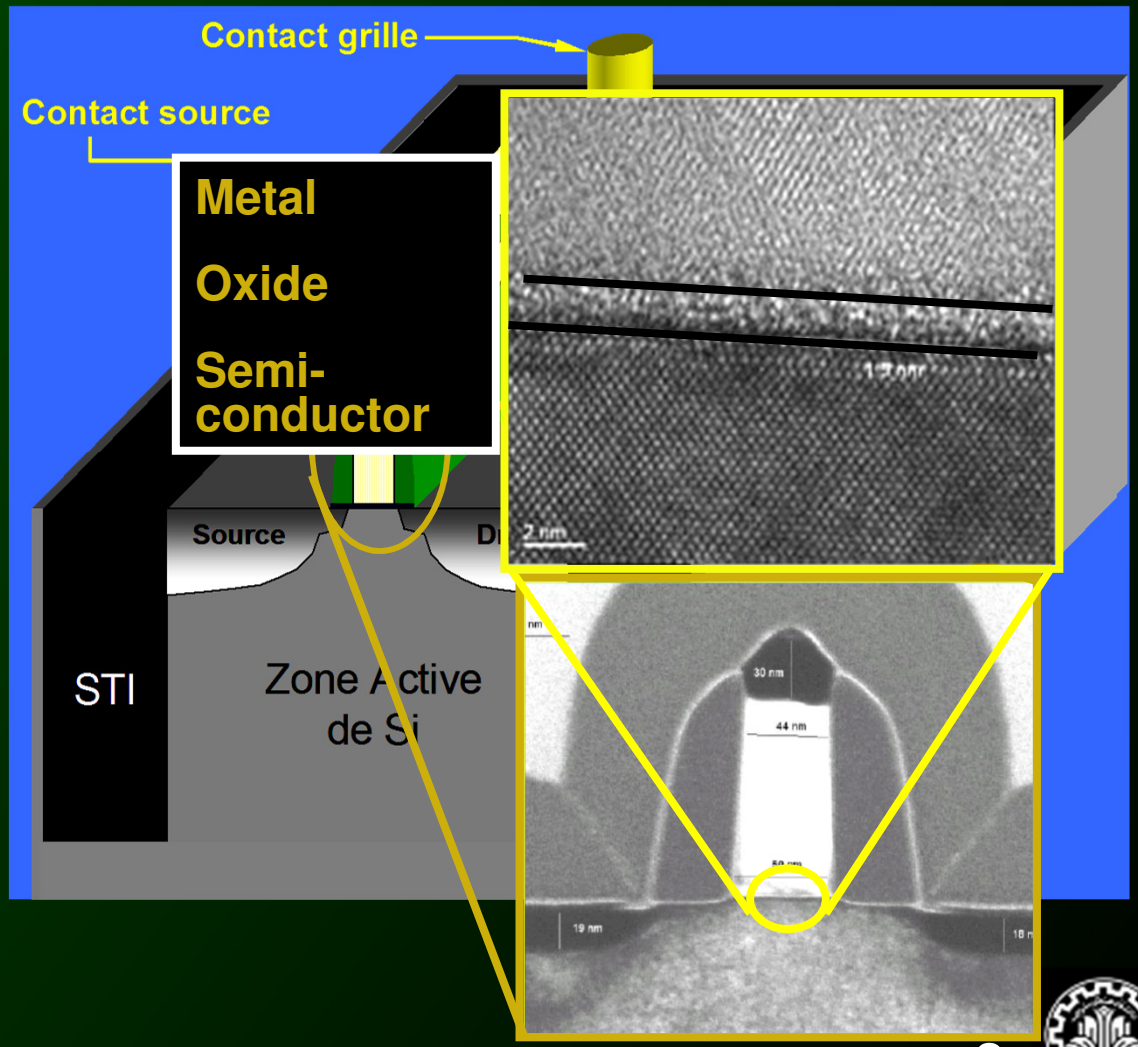
Electrons flow from the source to the drain by drift, when V_{DS}>0. (I_{DS} > 0)

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

MOSFET morphology

1. I
2.
3.
4.
5.

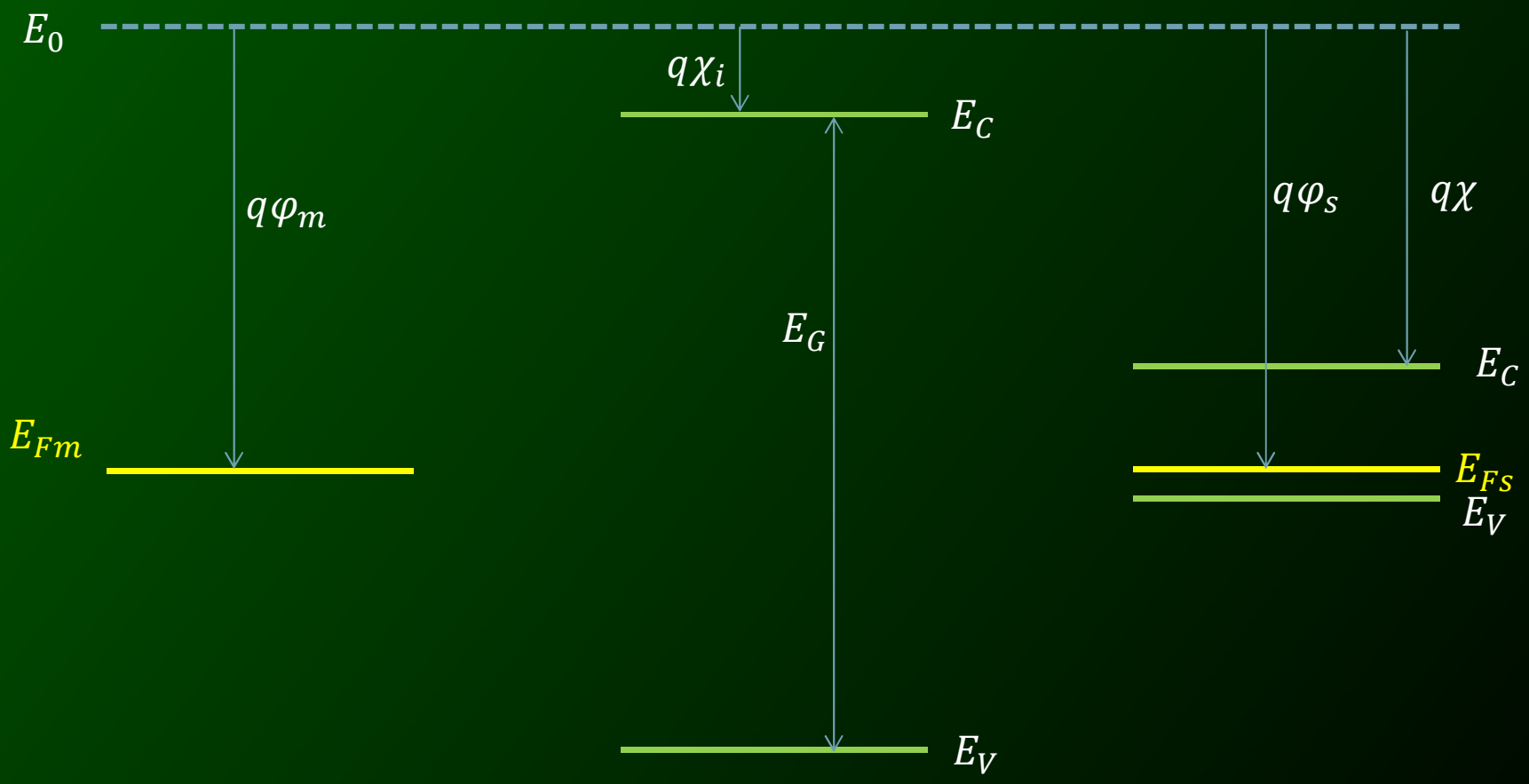




MOS – Special Case

1. I
2.
3.
4.
5.

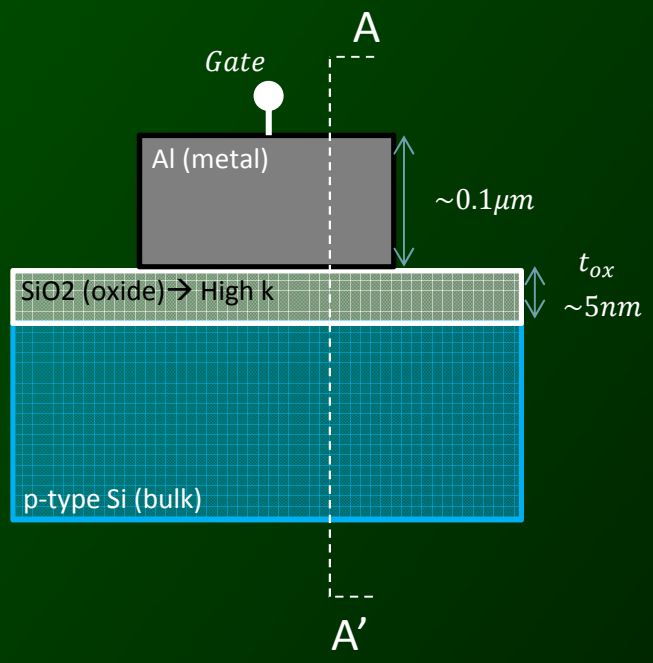
Special case: $\varphi_m = \varphi_s$

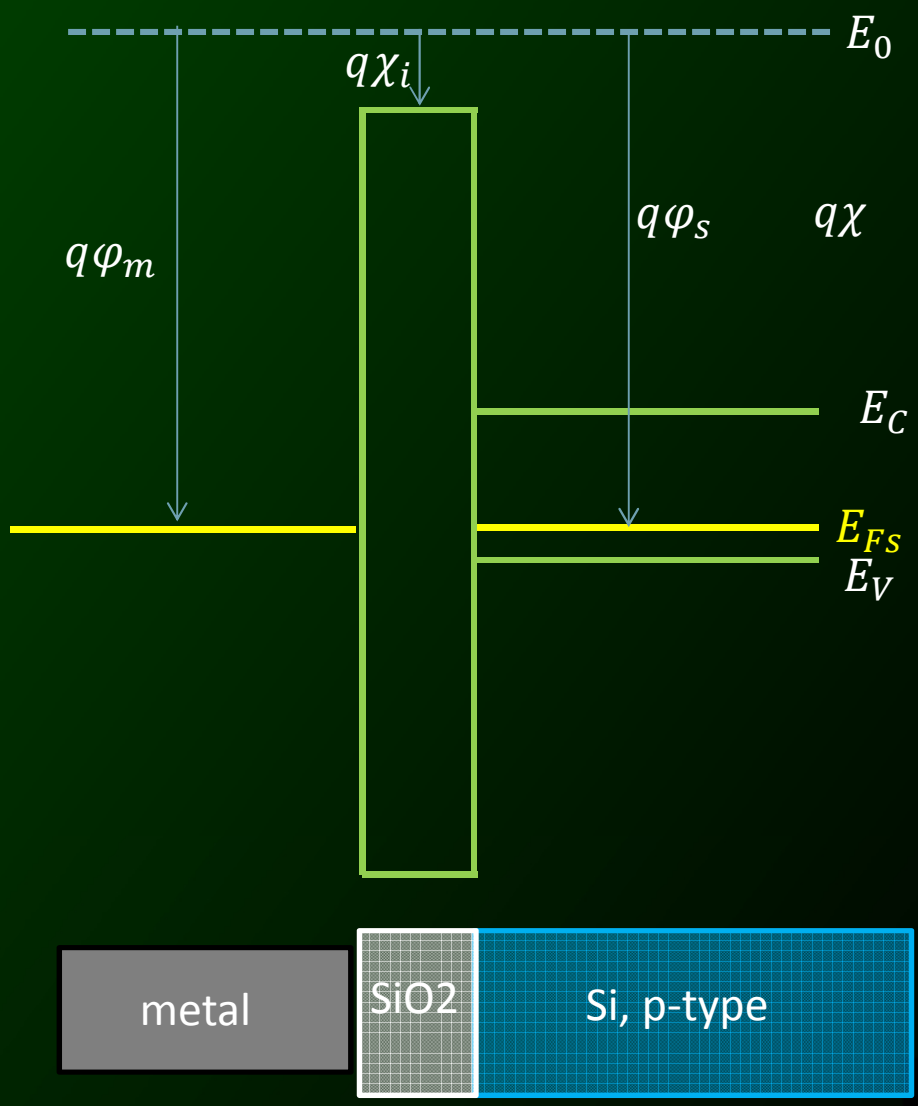


metal SiO2 Si, p-type

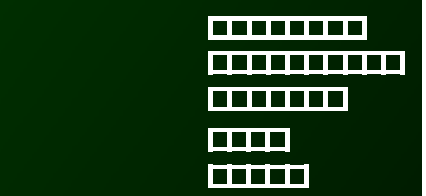
MOS – Special Case

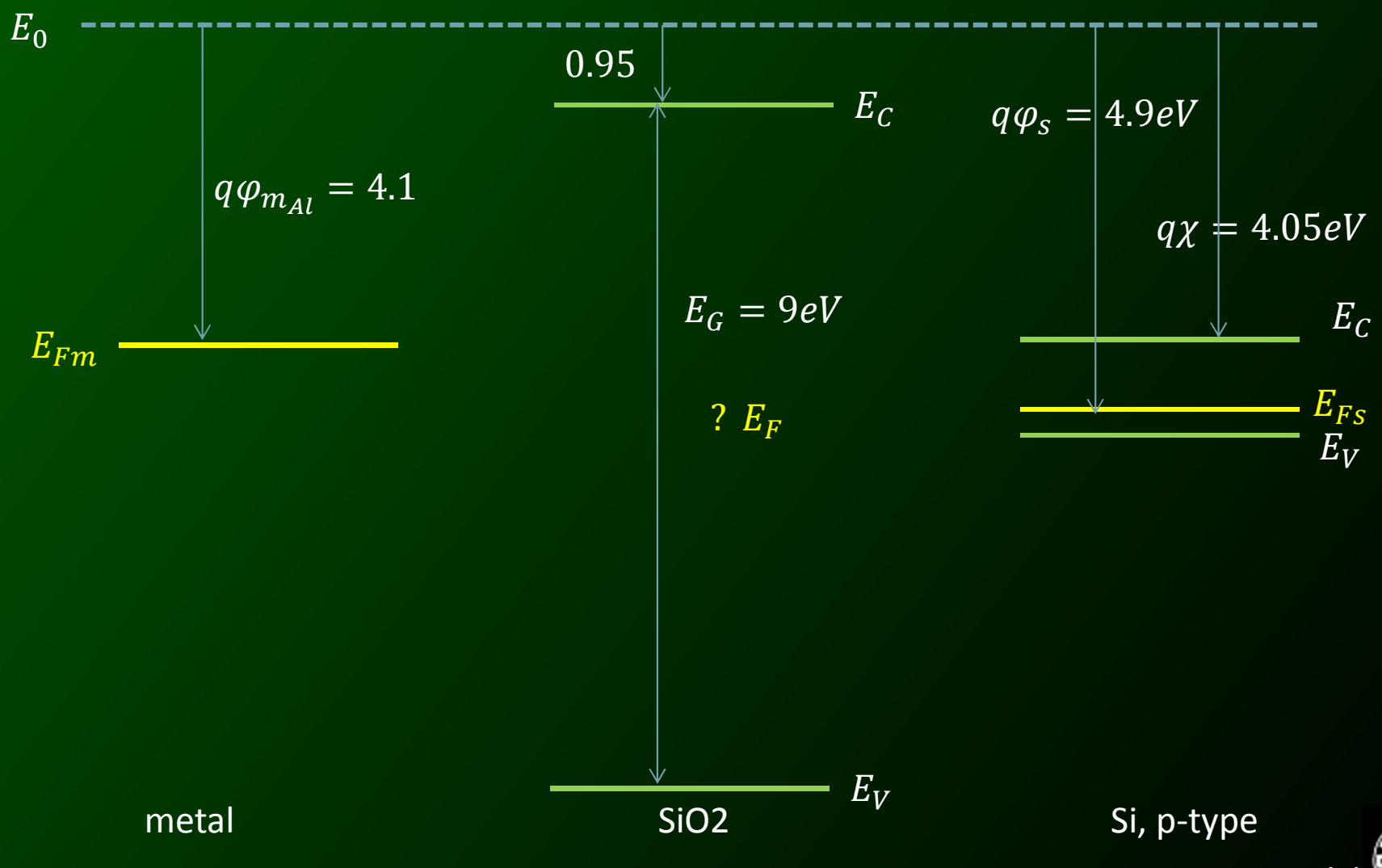
Special case: $\varphi_m = \varphi_s$



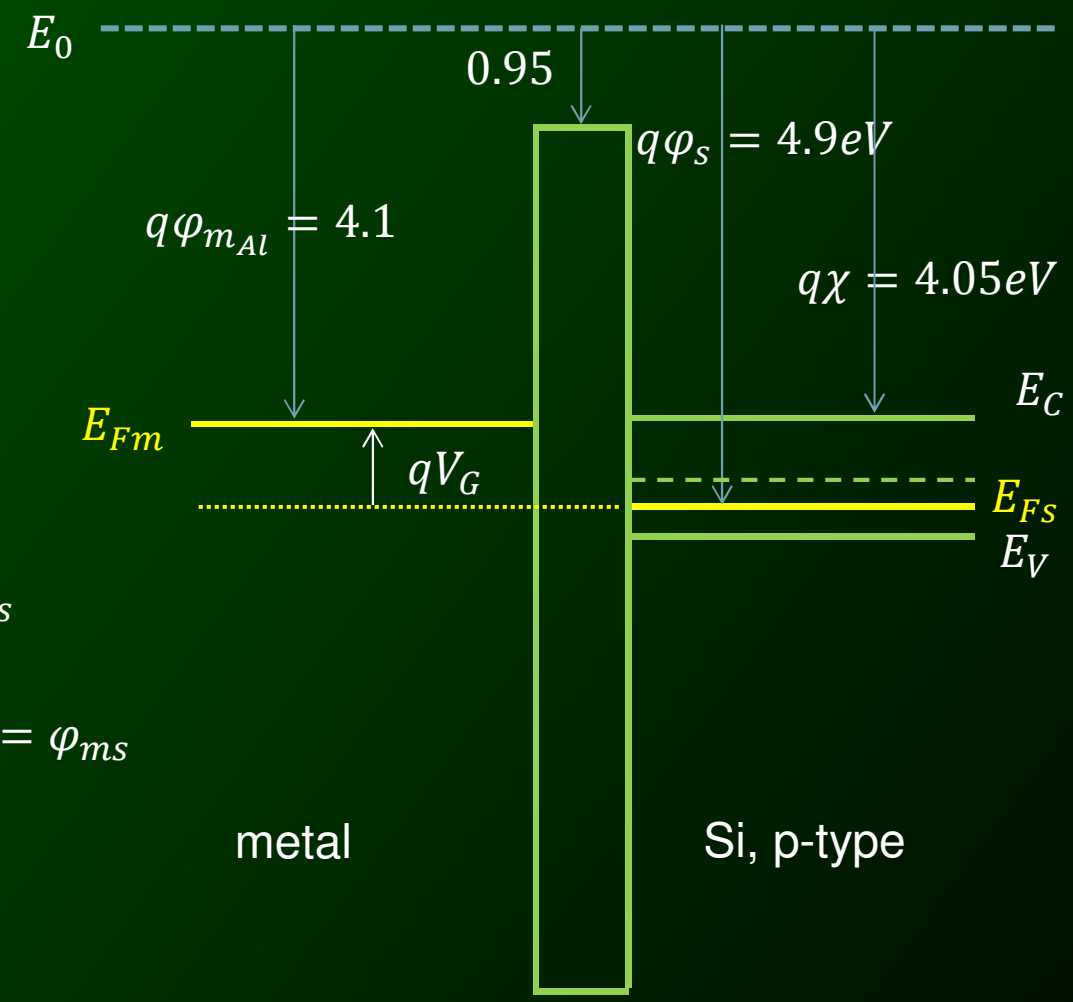


M (Al) , O (SiO2) , S (Si)





Flat Band Voltage

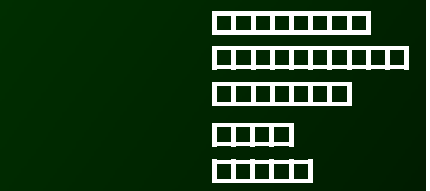


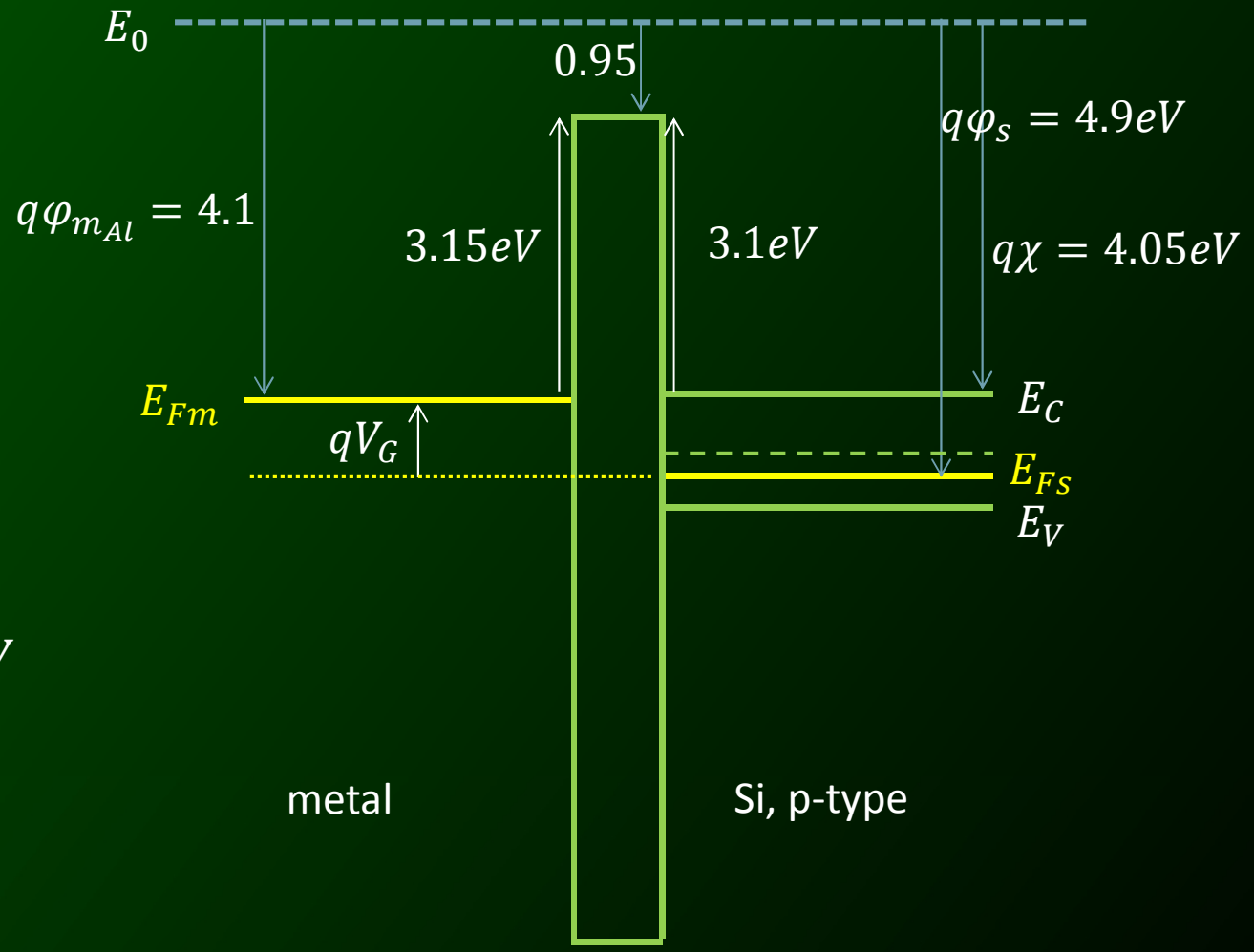
$$qV_G = q\varphi_m - q\varphi_s$$

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

$V_{FB} = -0.8\ V$ metal Si, p-type

Voltage Barrier

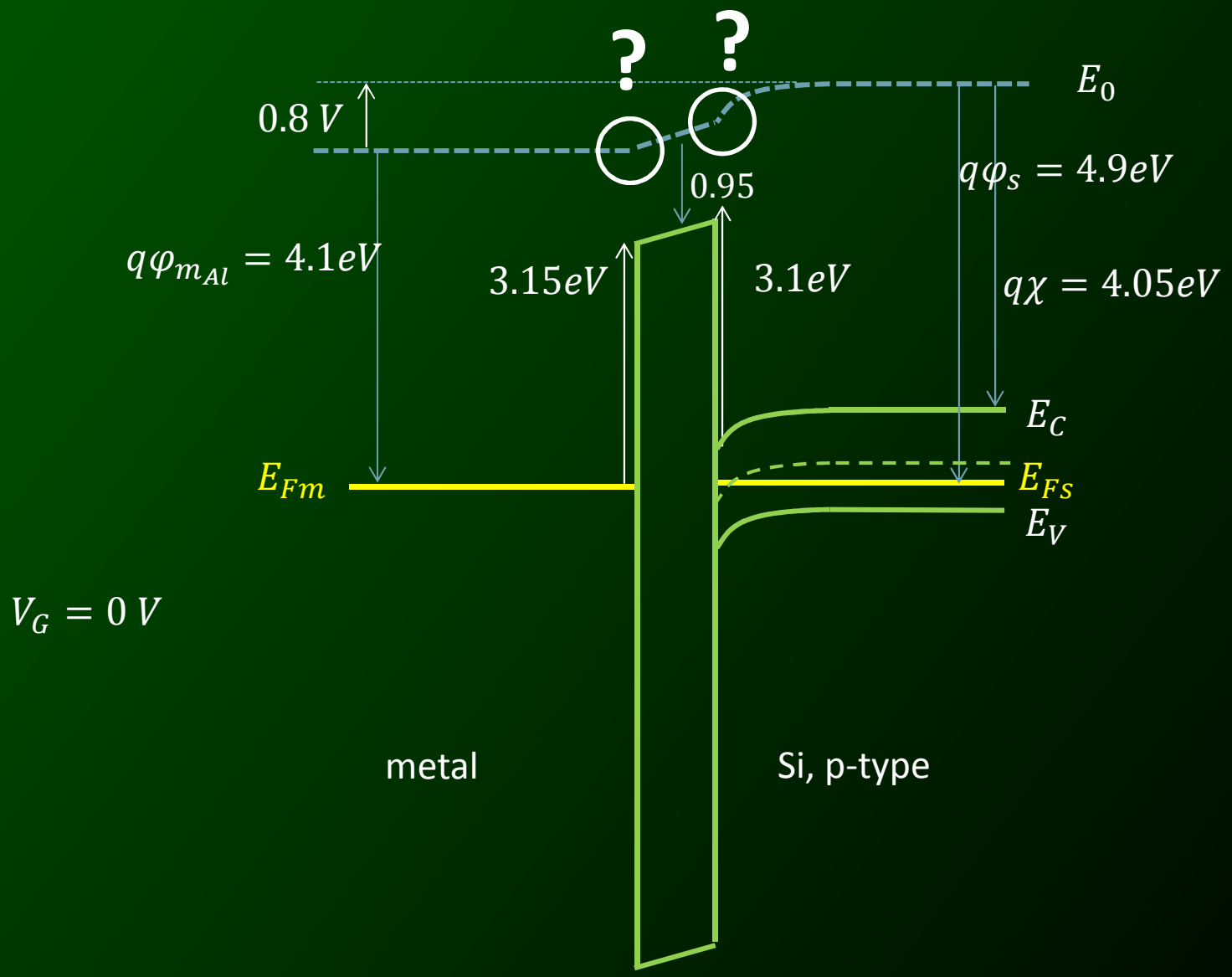




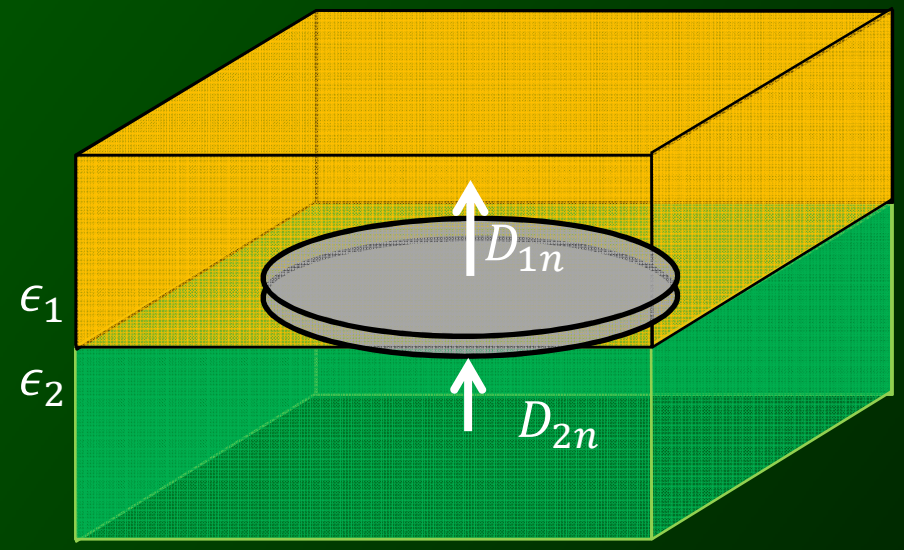
$V_G = V_{FB} = -0.8\ V$

No way electrons might pass the voltage barrier!

No Gate Voltage



Boundary Condition



$$D_{2n} - D_{1n} = \rho_{surfacce}$$

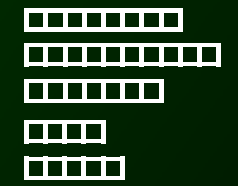
$$\rho_{surfacce} = 0$$

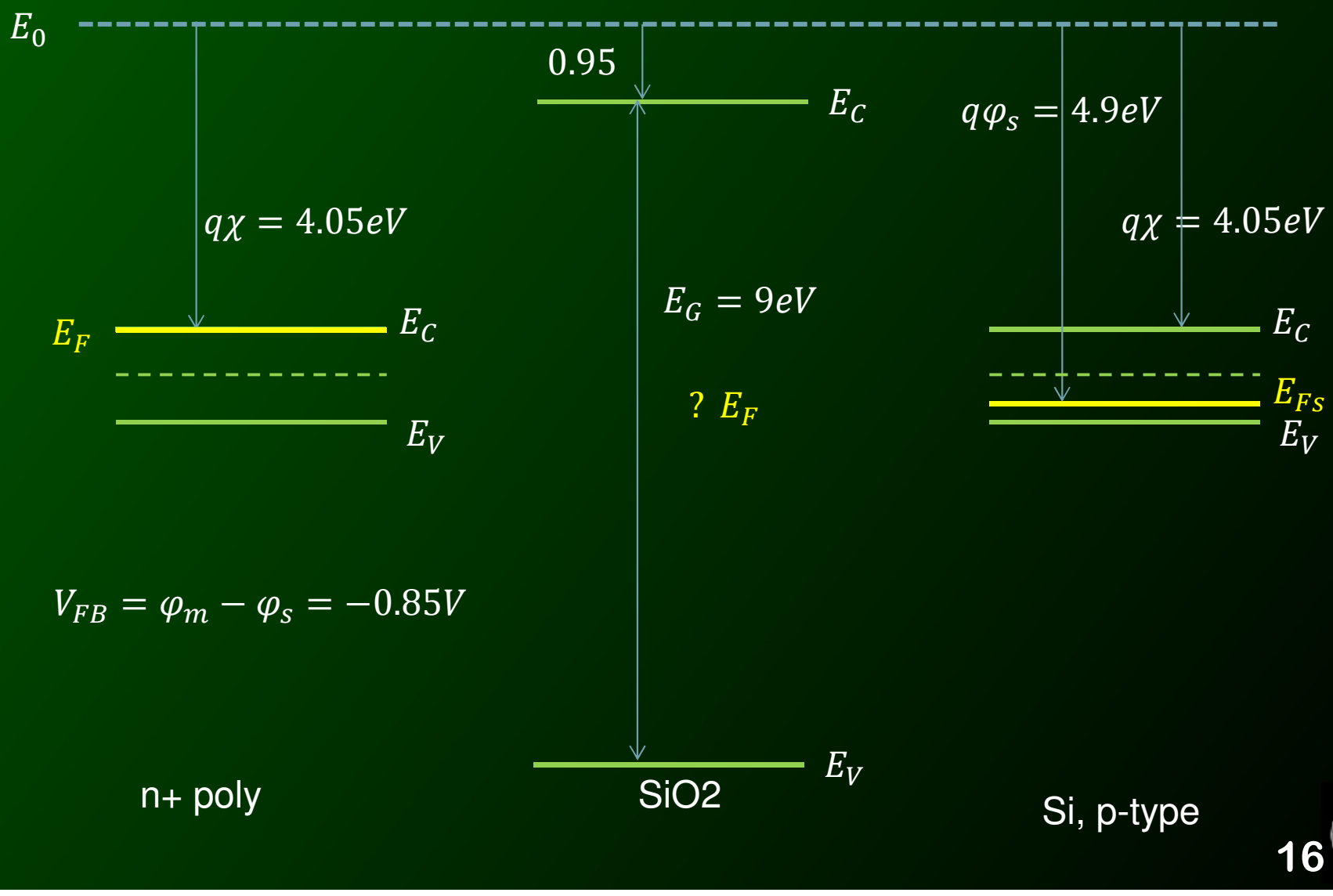
$$\epsilon_1 \mathcal{E}_1 = \epsilon_2 \mathcal{E}_2$$

$$\epsilon_{Ox} \left.\frac{dE_{Ox}}{dx}\right|_{int} = \epsilon_{Si} \left.\frac{dE_{Si}}{dx}\right|_{int}$$

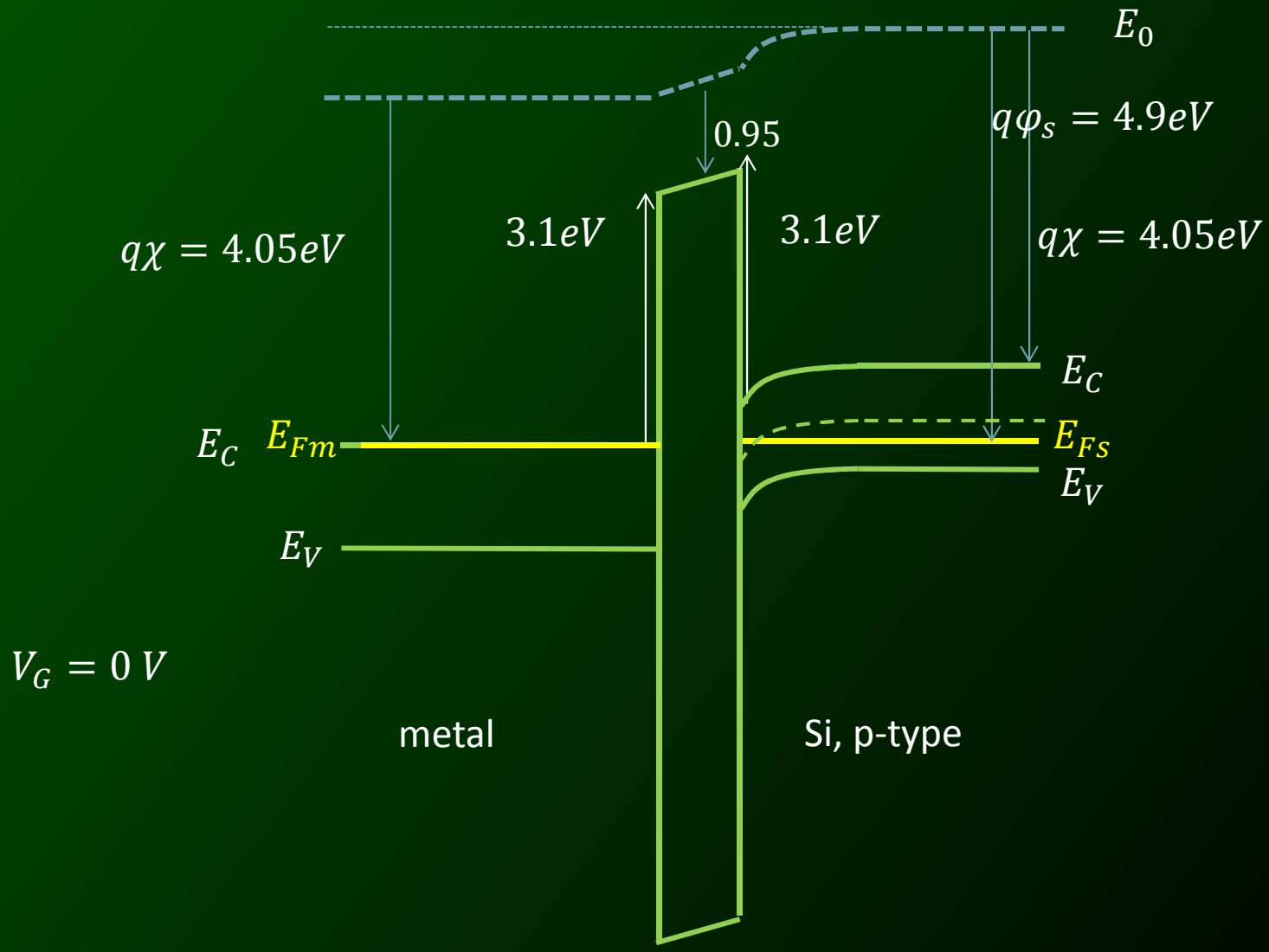
$$\left.\frac{dE_{Ox}}{dx}\right|_{int} \cong 3 \left.\frac{dE_{Si}}{dx}\right|_{int}$$

M (PolyGate) , O (SiO2) , S (Si)

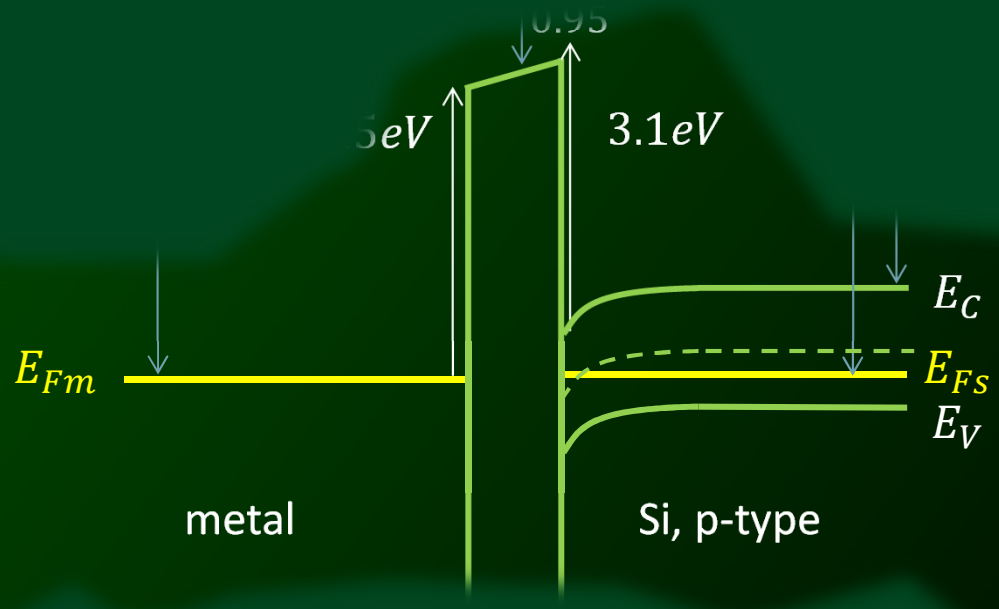




No Gate Voltage – Poly Gate

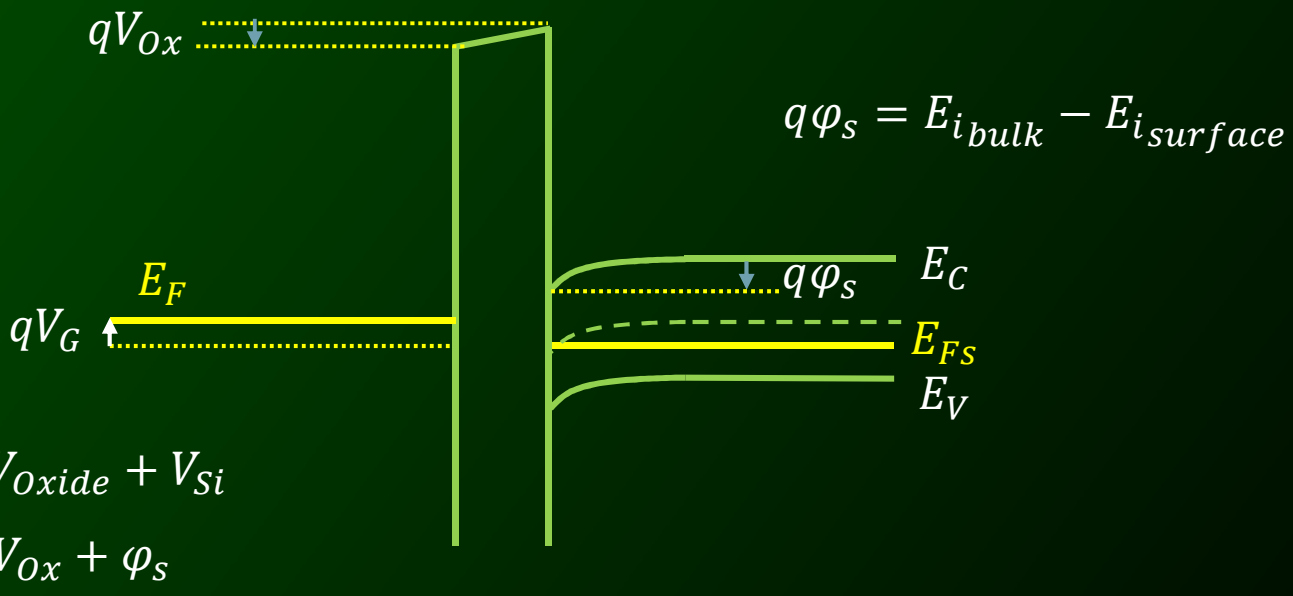


No Gate Voltage



$V_G = 0\ V$

No Gate Voltage



$$q\varphi_s = E_{i_{bulk}} - E_{i_{surface}}$$

$$V_G - V_{FB} = V_{Oxide} + V_{Si}$$
$$= V_{Ox} + \varphi_s$$

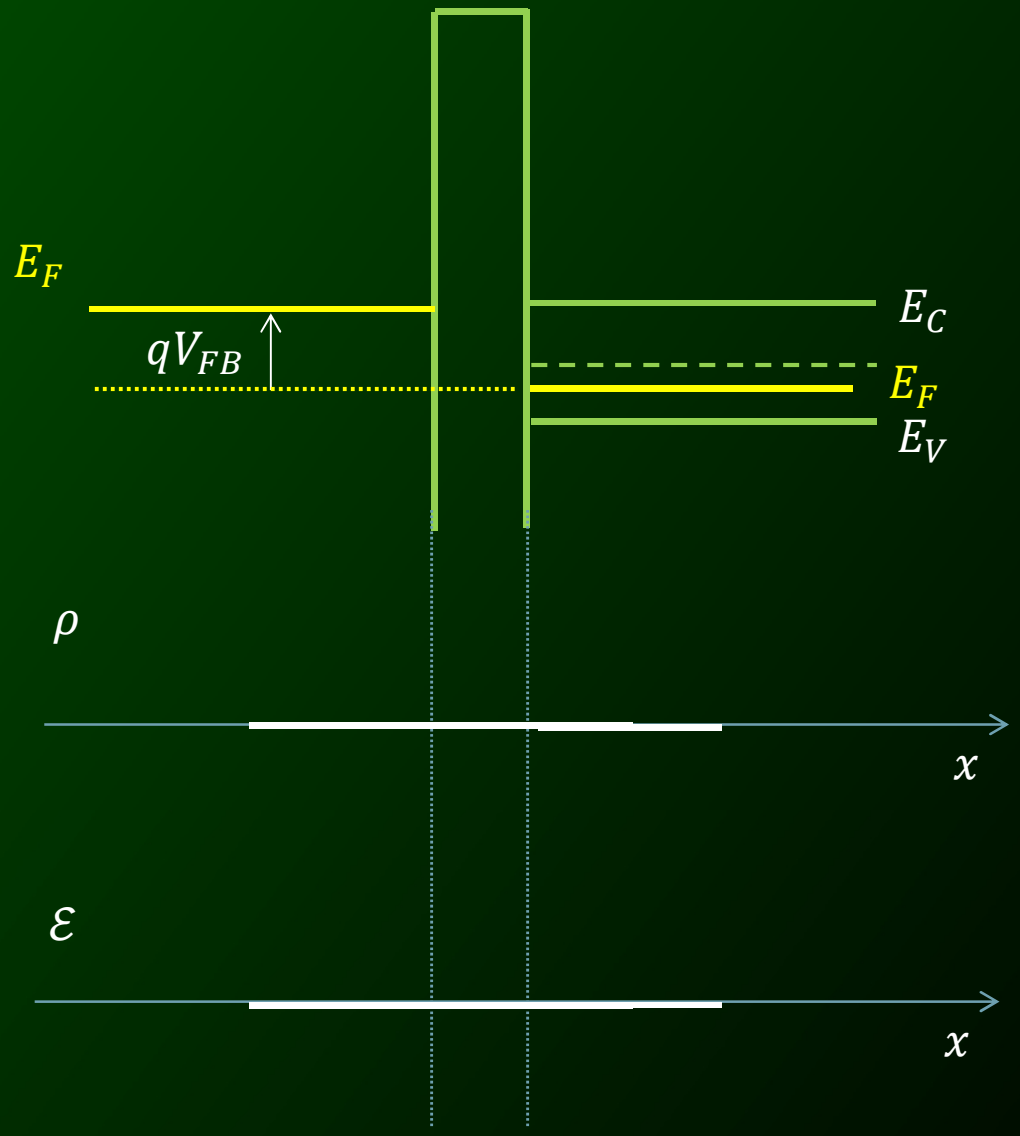
$$\boxed{V_G - V_{FB} = V_{Ox} + \varphi_s}$$

As important as KVL

Flat Band

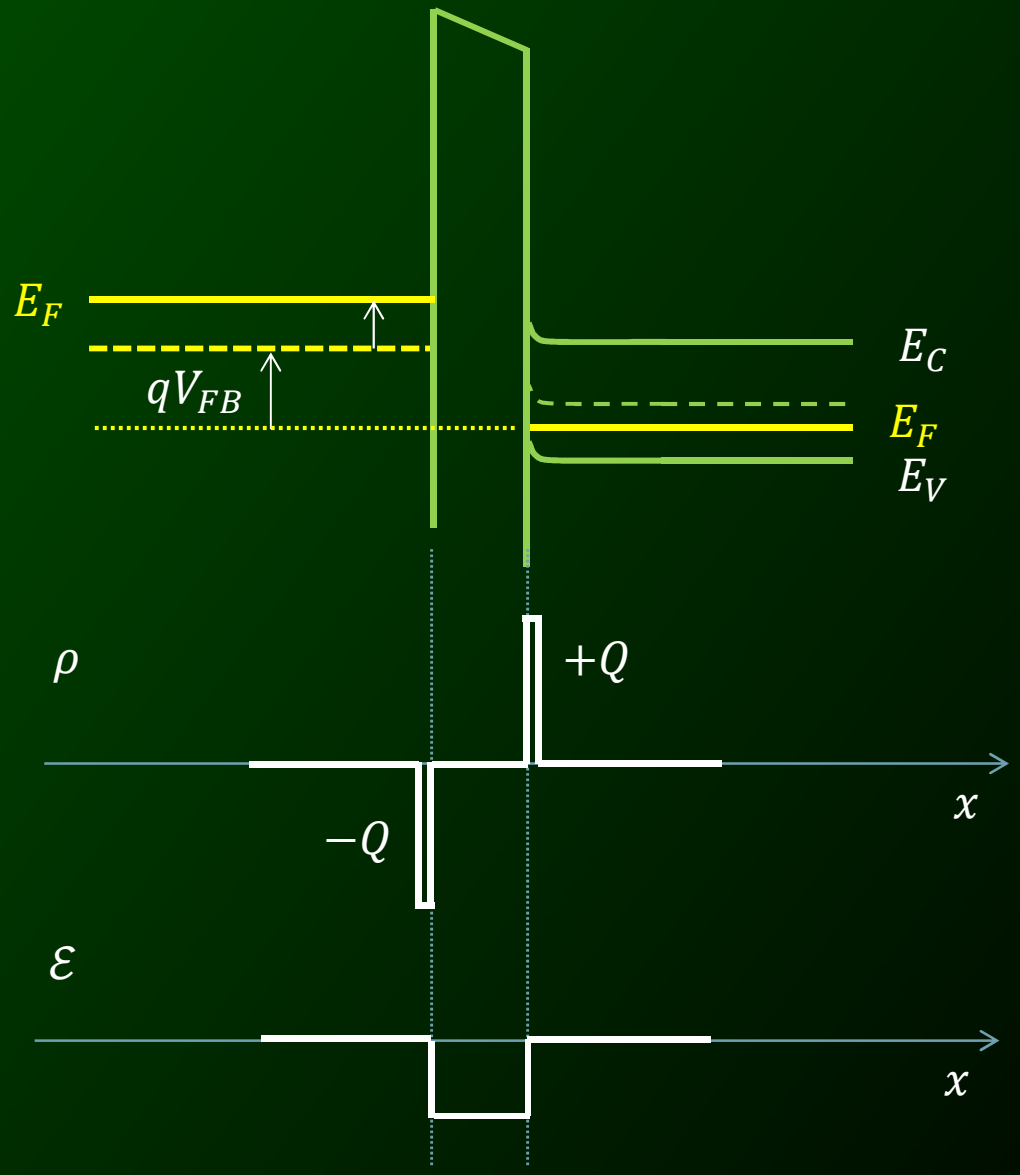
1. I
2.
3.
4.
5.

$V_G = V_{FB}$



Accumulation

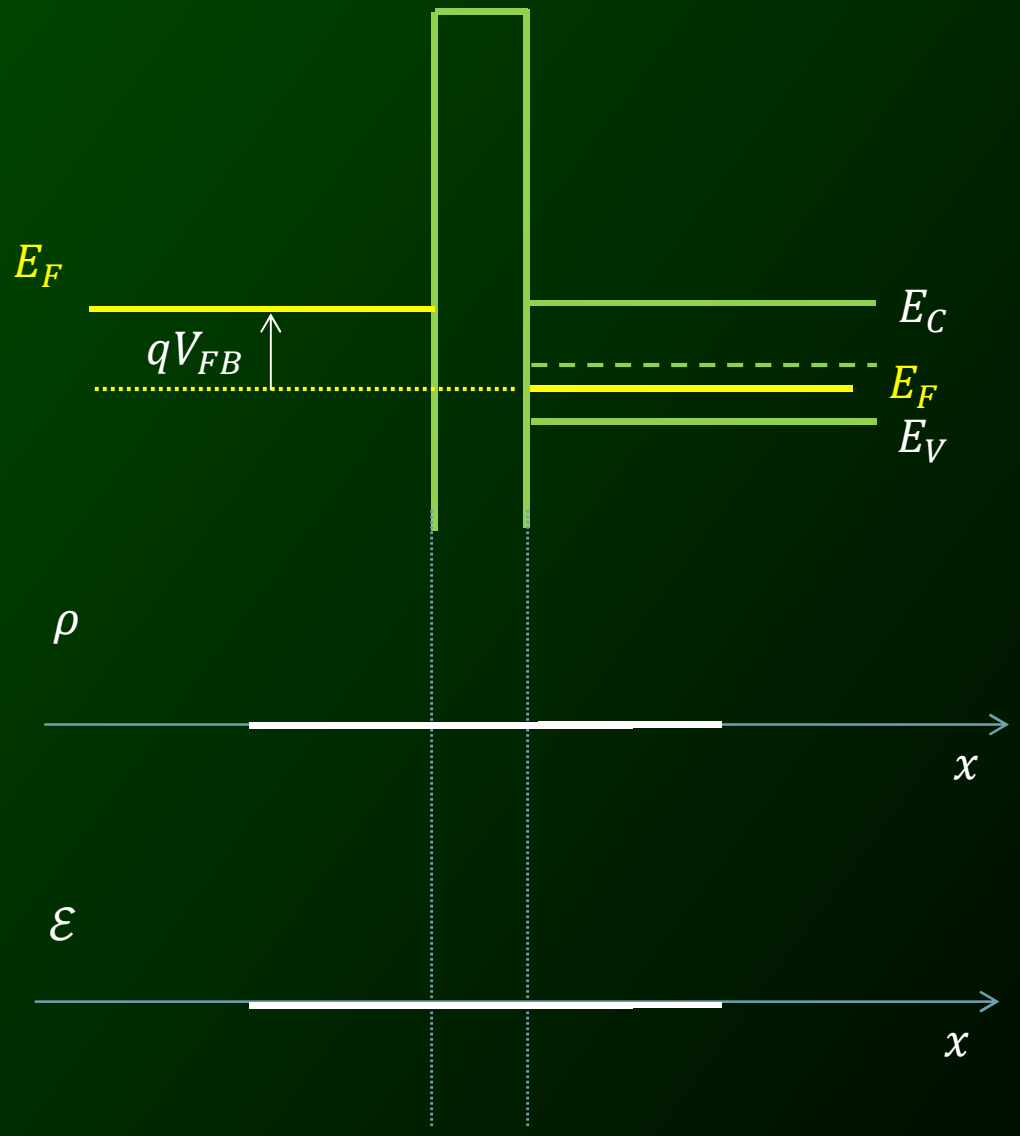
$V_G < V_{FB}$



Flat Band

1. I
2.
3.
4.
5.

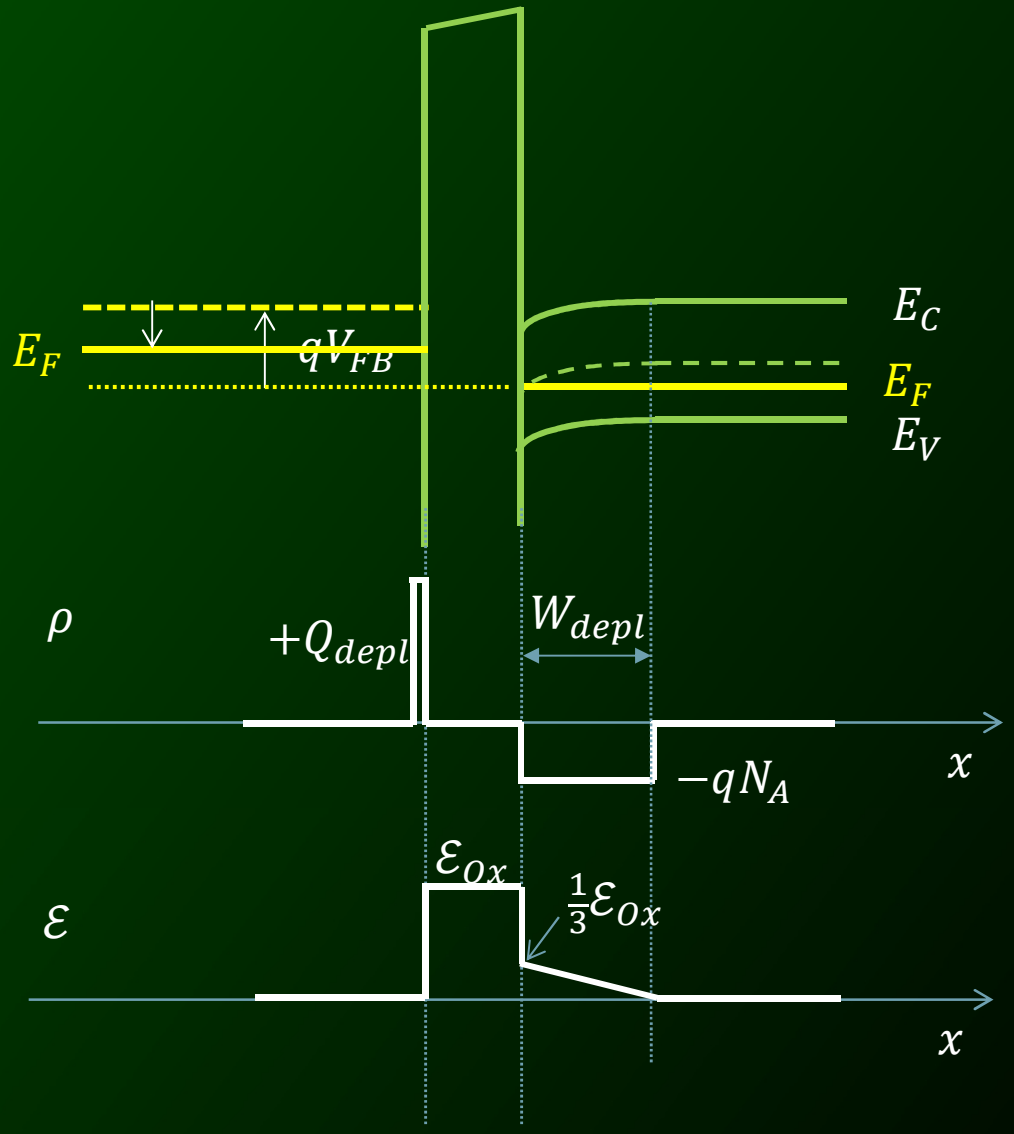
$V_G = V_{FB}$



Depletion (Weak Inversion)

1. I
2.
3.
4.
5.

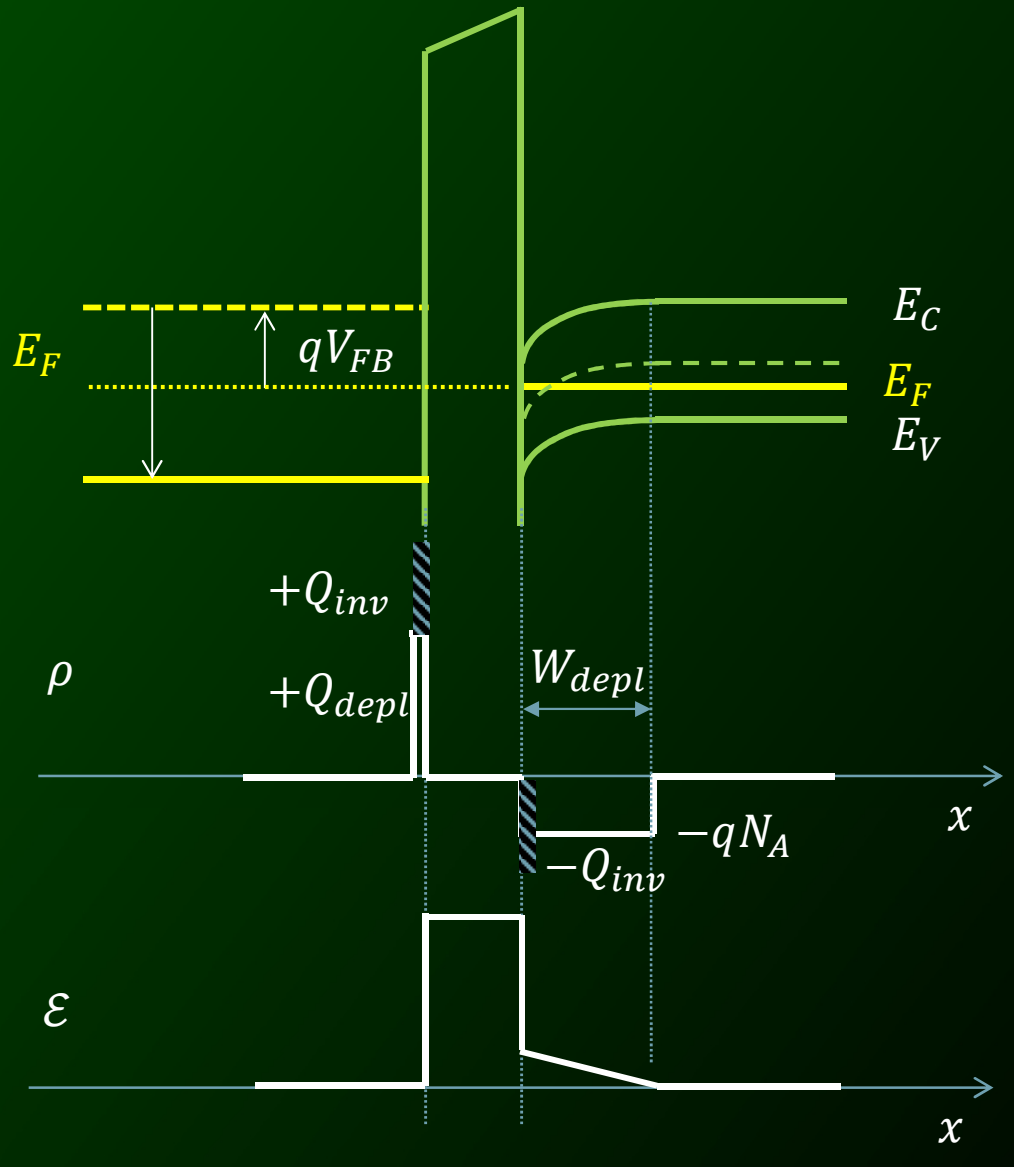
$V_T > V_G > V_{FB}$



(Strong) Inversion

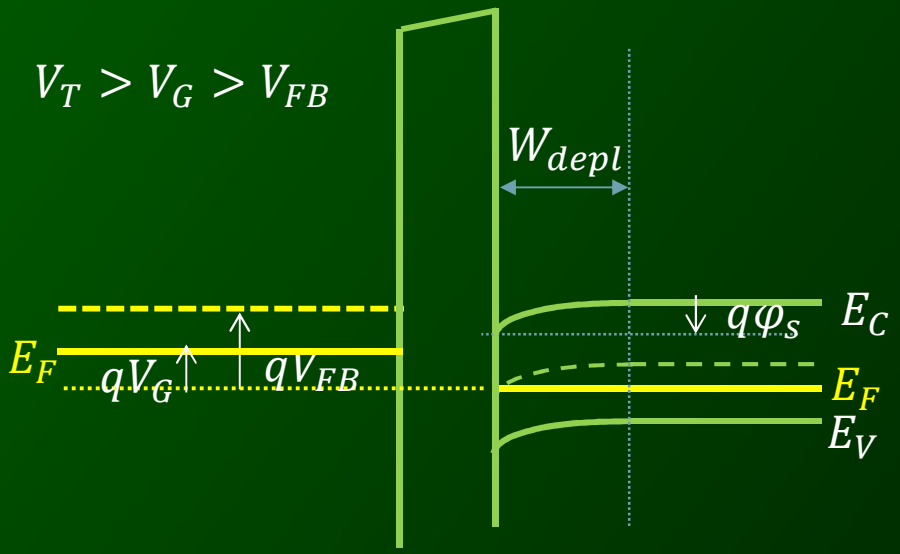
1. I
2.
3.
4.
5.

$V_G > V_T$



Depletion (Weak Inversion)

1. I
2.
3.
4.
5.



$$\frac{d\mathcal{E}}{dx} = \frac{\rho}{\epsilon_{Si}} = \frac{-qN_A}{\epsilon_{Si}} = \frac{d^2\varphi}{dx^2}$$

$$\rightarrow \varphi_s = \frac{qN_A x^2}{2\epsilon_{Si}} \qquad \rightarrow W_d = \sqrt{\frac{2\epsilon_{Si}}{qN_A}\varphi_s}$$

$$Q_{dep} = -qN_A W_d = \sqrt{2qN_A\epsilon_{Si}\varphi_s}$$

$$\mathcal{E}_{Ox} = \frac{-Q_{dep}}{\epsilon_{Ox}} \qquad \rightarrow V_{Ox} = \frac{-t_{ox}Q_{dep}}{\epsilon_{Ox}} = \frac{-Q_{dep}}{C_{Ox}} \quad \begin{matrix}[C/cm^2] \\ [F/cm^2]\end{matrix}$$

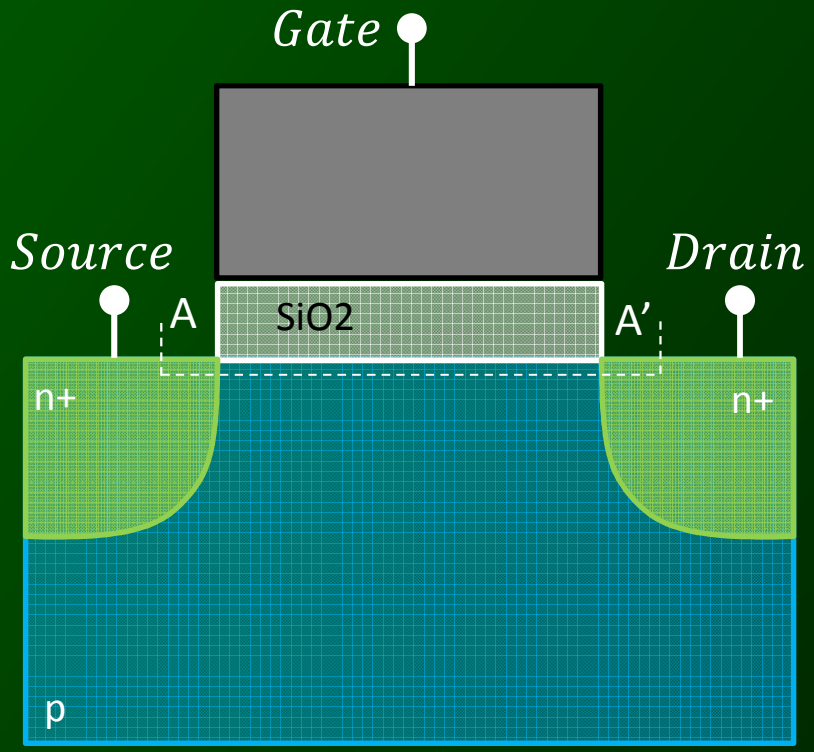
$$V_G = V_{FB} + V_{Ox} + \varphi_s \rightarrow \boxed{V_G = V_{FB} + \varphi_s + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Si}\varphi_s}}$$

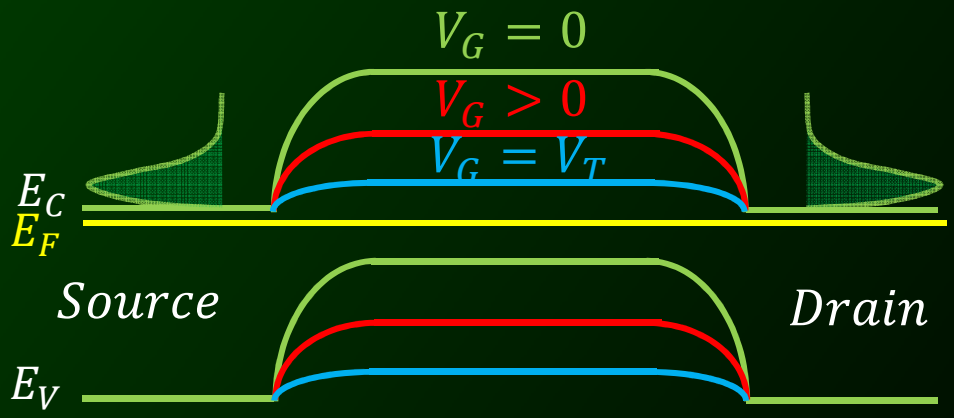
p-type Si (nMOS)
In Depl

$$\varphi_s < 0 \qquad \boxed{V_G = V_{FB} + \varphi_s - \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Si}|\varphi_s|}}$$

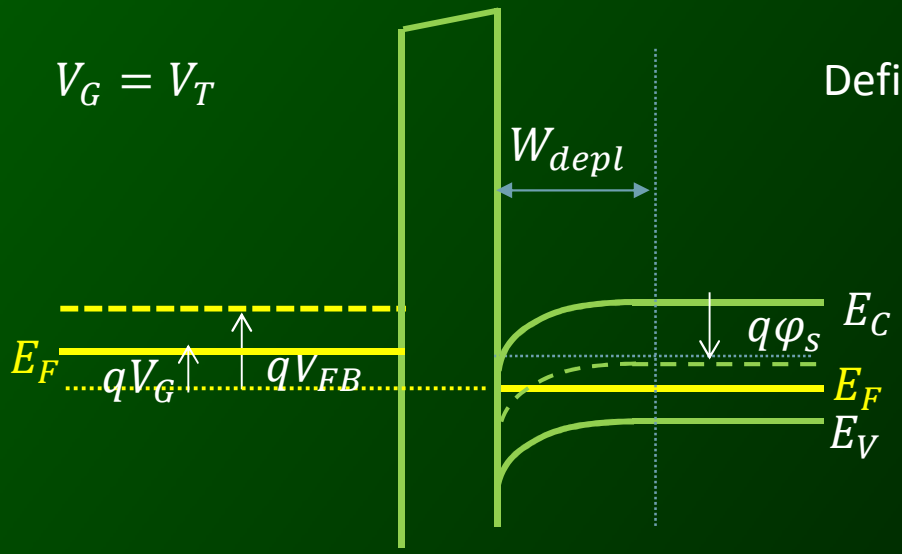
n-type Si (pMOS)
In Depl

Threshold Voltage – Definition!





Threshold Voltage



Definition of Threshold voltage:

$$V_T = V_G\Big|_{\varphi_s=2\varphi_F}$$

$p_{bulk} = N_A \qquad n_{surface} = N_A$

$$W_{max} = W_{depl}\Big|_{\varphi_s=2\varphi_F} = \sqrt{\frac{2\epsilon_{Si}}{qN_A}(2\varphi_F)}$$

p-type

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

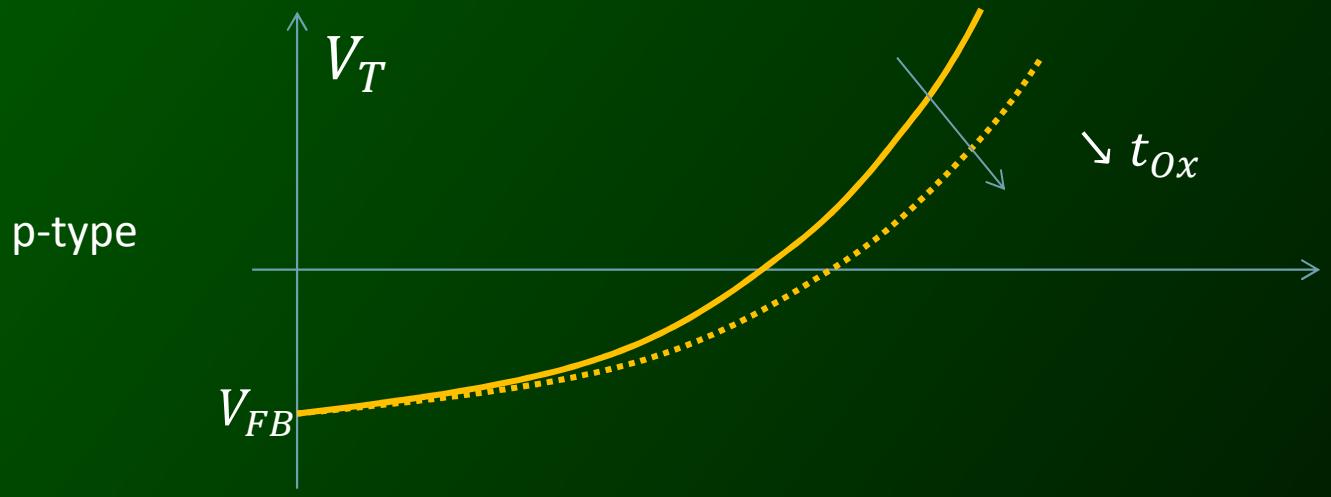
$$q\varphi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) > 0$$

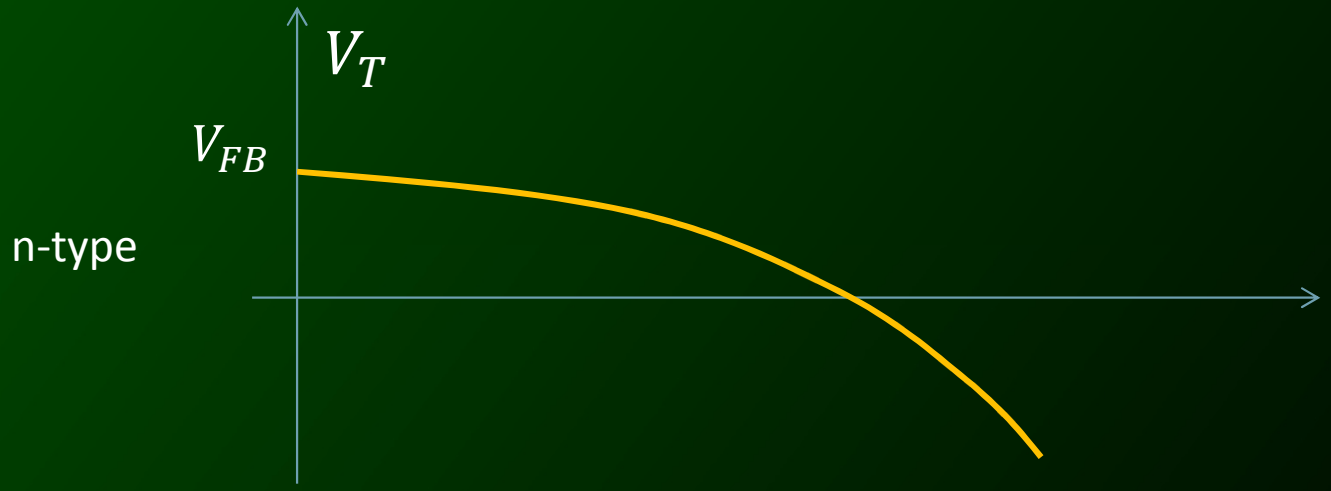
n-type

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

$$q\varphi_F = -\frac{kT}{q}\ln\left(\frac{N_D}{n_i}\right) < 0$$

Threshold Voltage vs. Bulk Doping



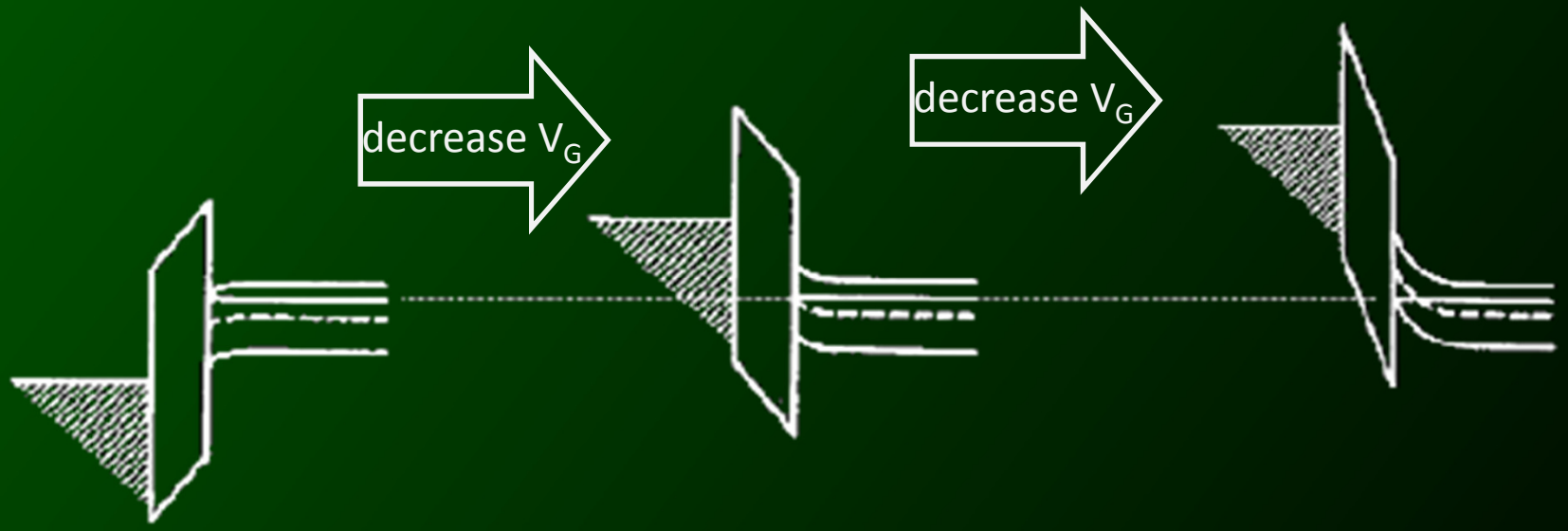


MOS Band Diagram (n-type Bulk)

1. I
2.
3.
4.
5.

Decrease VG (toward more negative values)

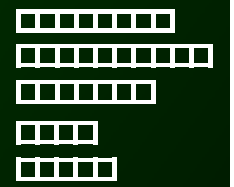
→ move the gate energy-bands up, relative to the Si



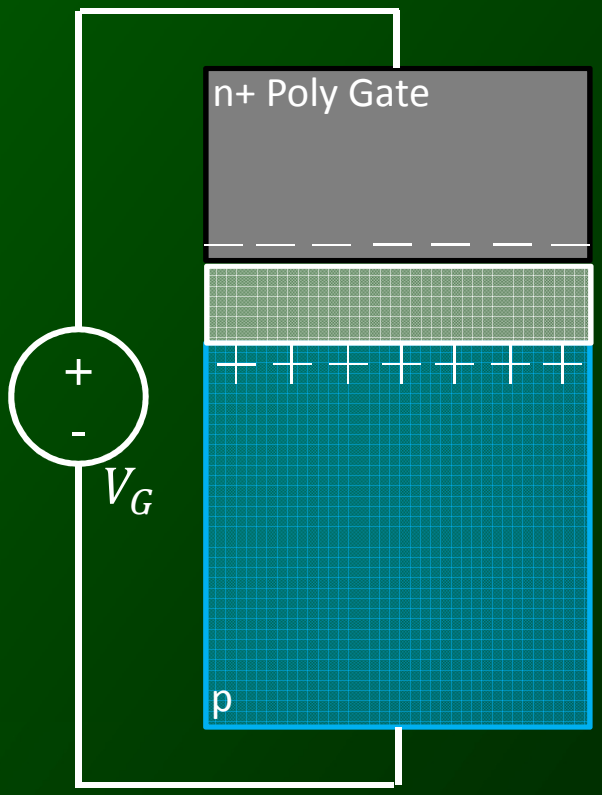
- Accumulation
  - $V_G > V_{FB}$
  - Electrons accumulate at surface
- Depletion
  - $V_G < V_{FB}$
  - Electrons repelled from surface
- Inversion
  - $V_G < V_T$
  - Surface becomes p-type

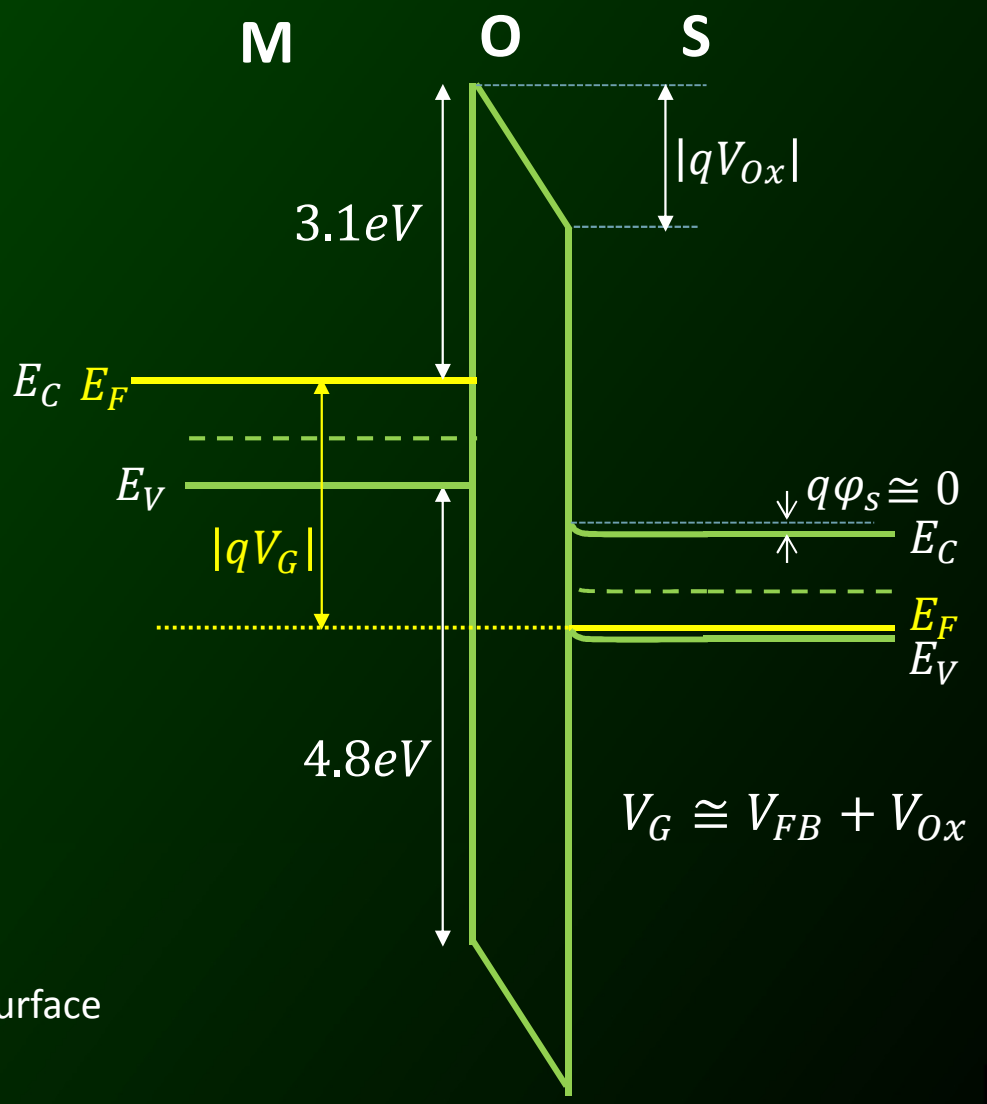
Accumulation, Poly Gate

1. I
2.
3.
4.
5.



$V_G < V_{FB}$

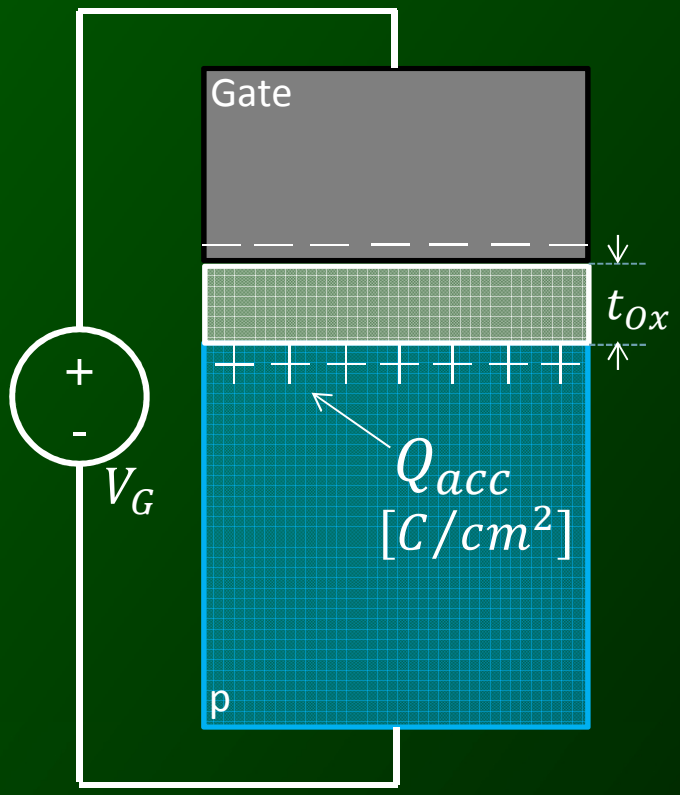




Mobile carriers (holes) accumulate at Si surface

Accumulation, Layer Charge Density

$V_G < V_{FB}$



$$V_{Ox} \cong V_G - V_{FB}$$

From Gauss' Law:

$$\mathcal{E}_{Ox} = -Q_{acc}/\epsilon_{SiO_2}$$

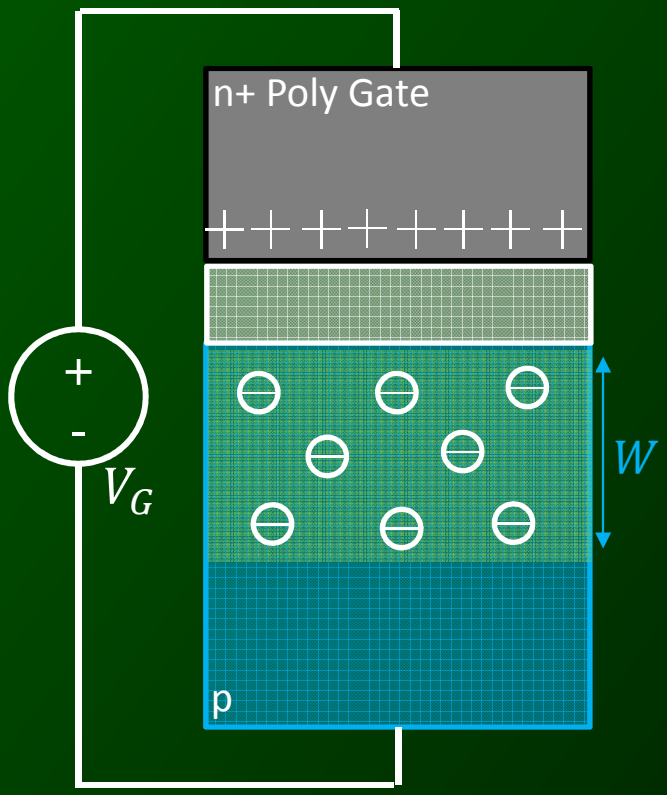
$$V_{Ox} = t_{Ox}\mathcal{E}_{Ox} = -Q_{acc}/C_{Ox}$$

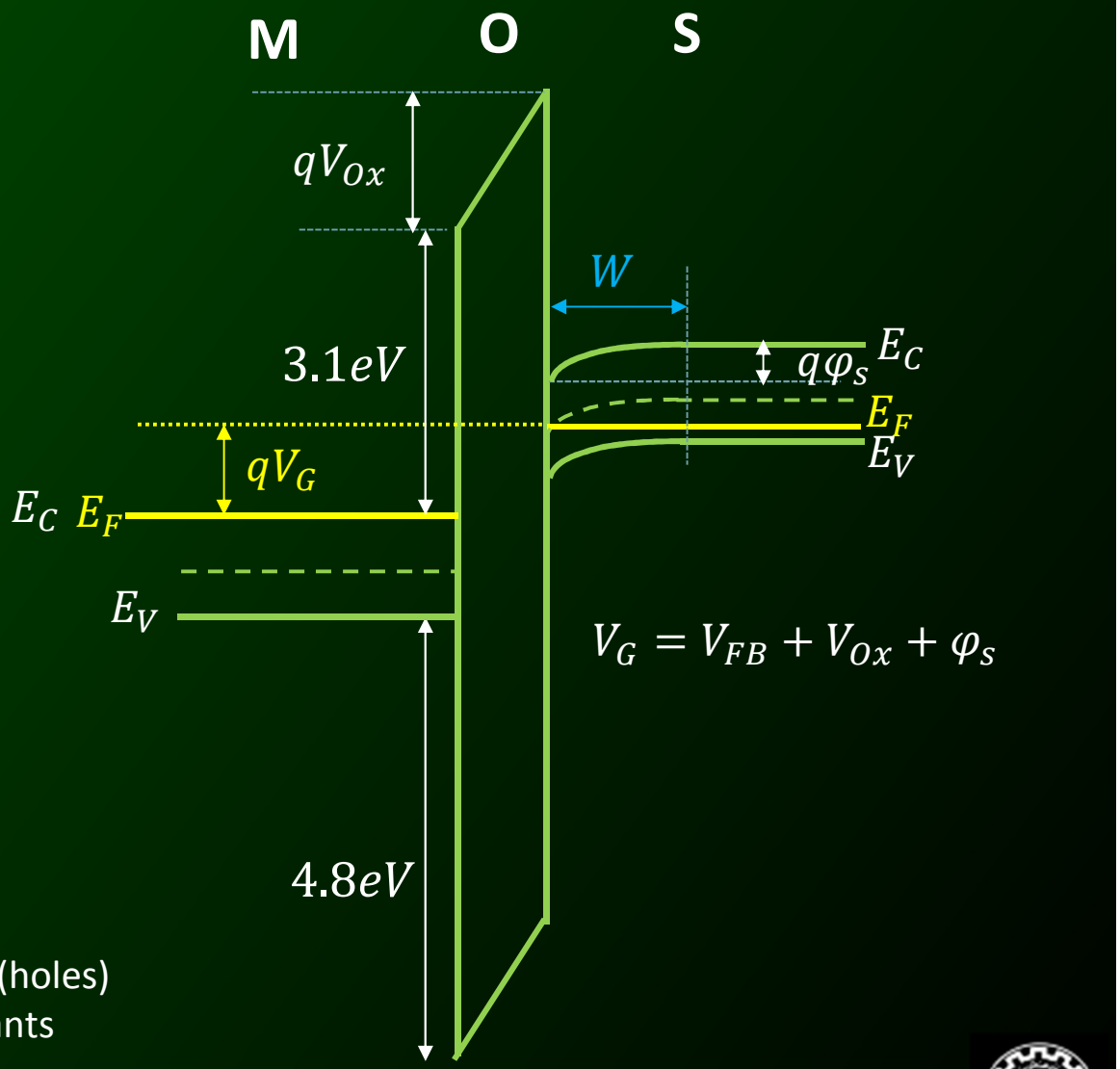
where $C_{Ox} \equiv \epsilon_{SiO_2}/t_{Ox}$ $[F/cm^2]$

$$\rightarrow \boxed{Q_{acc} = -C_{Ox}(V_G - V_{FB}) > 0}$$

Depletion, Poly Gate

$V_G < V_{FB}$

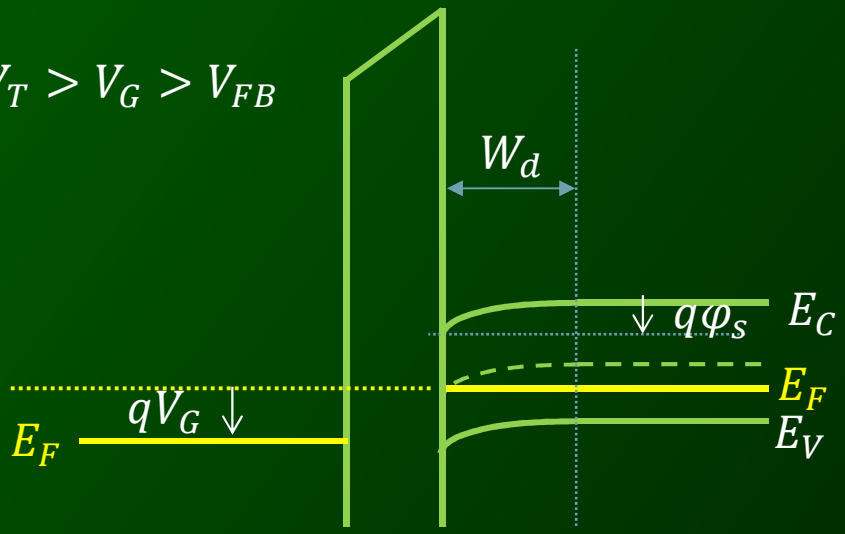




$V_G = V_{FB} + V_{Ox} + \varphi_s$

Si surface is depleted of mobile carriers (holes)
=> Surface charge is due to ionized dopants (acceptors)

Depletion (Weak Inversion)



$$\frac{d\mathcal{E}}{dx} = \frac{-qN_A}{\epsilon_{Si}} = \frac{d^2\varphi}{dx^2} \quad \rightarrow \quad \varphi_s = \frac{qN_A x^2}{2\epsilon_{Si}}$$

$$\rightarrow W_d = \sqrt{\frac{2\epsilon_{Si}}{qN_A}\varphi_s} \qquad V_{Ox} = \frac{-Q_{dep}}{C_{Ox}}$$

$$V_G = V_{FB} + V_{Ox} + \varphi_s \;\rightarrow\; V_G = V_{FB} + \varphi_s + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Si}\varphi_s}$$

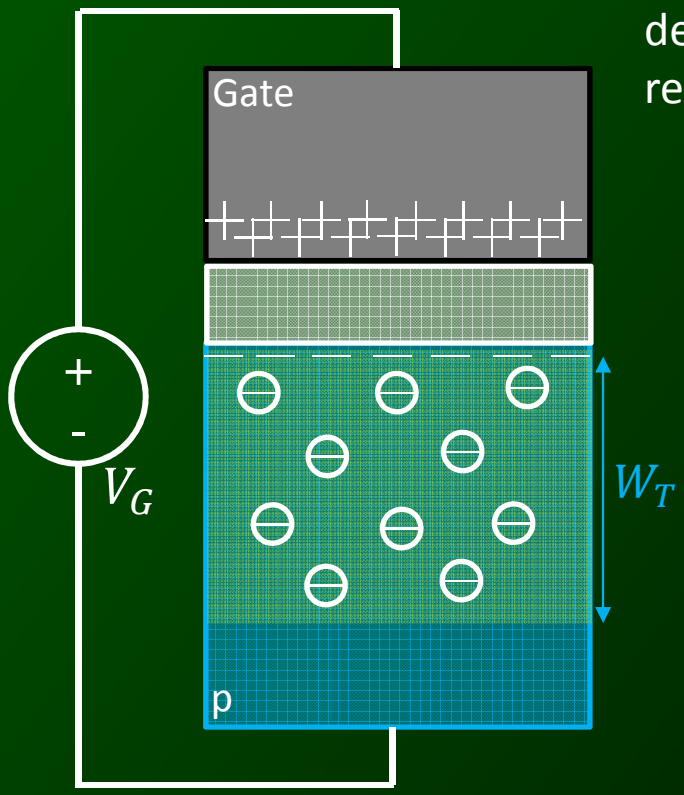
Solving for φ_s :

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

$$Q_{dep} = -qN_A W_d = -\sqrt{2qN_A\epsilon_{Si}\varphi_s}$$

Strong Inversion

$V_G > V_T$



Significant density of mobile electrons at surface (surface is n-type)

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 ~ constant at $W = W_T$

$$\varphi_s \cong 2\varphi_F \quad \rightarrow W \cong W_T = \sqrt{\frac{2\epsilon_{Si}}{qN_A}(2\varphi_F)}$$

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

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

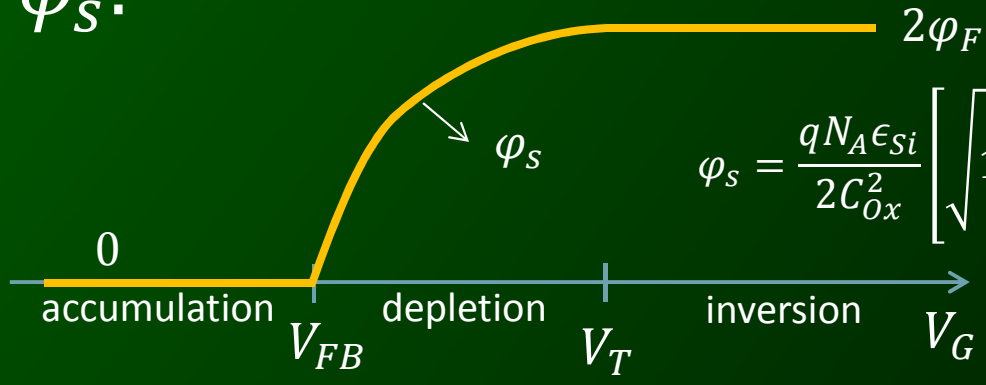
$$= V_{FB} + 2\varphi_F - \frac{\sqrt{2q\epsilon_{Si}N_A(2\varphi_F)}}{C_{Ox}} - \frac{Q_{inv}}{C_{Ox}}$$

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

$$\therefore \quad Q_{inv} = -C_{Ox}(V_G - V_T)$$

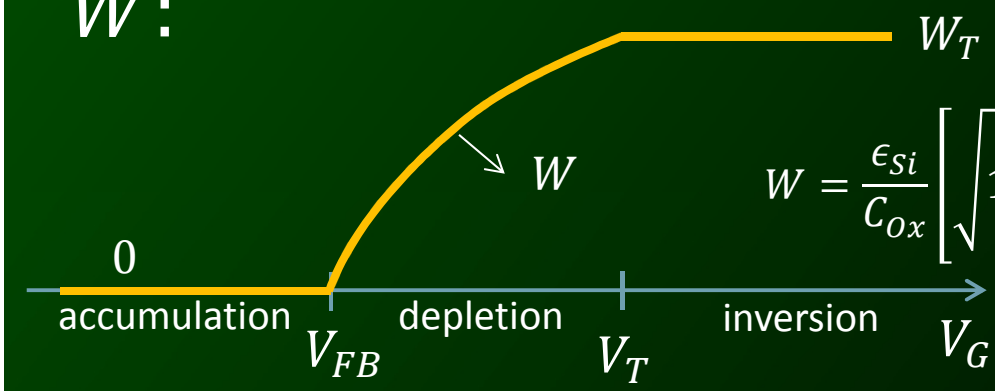
φ_s and W vs. V_G

φ_s:



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

W:



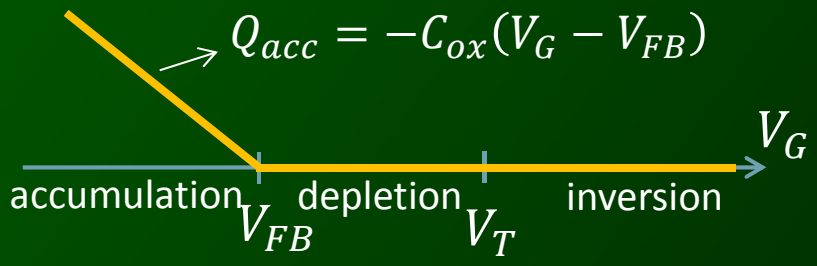
$$W_T = \sqrt{2\epsilon_{Si}(2\varphi_F)/qN_A}$$

$$W = \frac{\epsilon_{Si}}{C_{Ox}}\left[\sqrt{1+\frac{2C_{Ox}^2}{qN_A\epsilon_{Si}}(V_G - V_{FB})} - 1\right] \quad (V_{FB} < V_G < V_T)$$

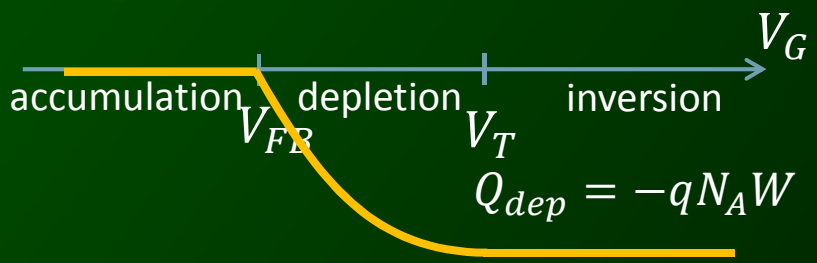
Total Charge Density in Si, Q_S

1. I
2.
3.
4.
5.

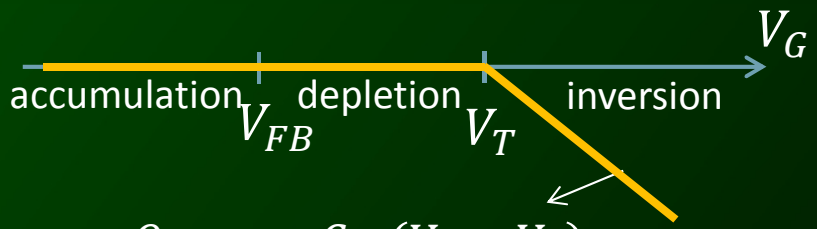
Q_{acc}:



Q_{dep}:



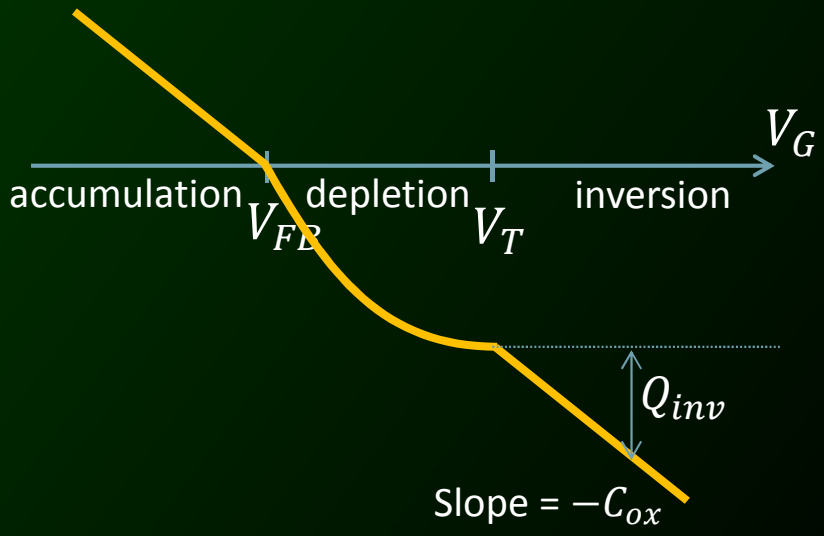
Q_{inv}:



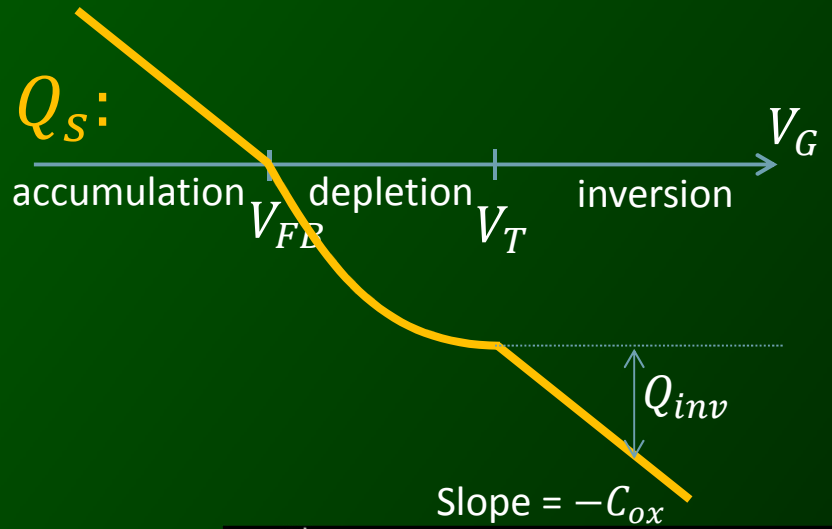
$$Q_{inv} = -C_{ox}(V_G - V_T)$$

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

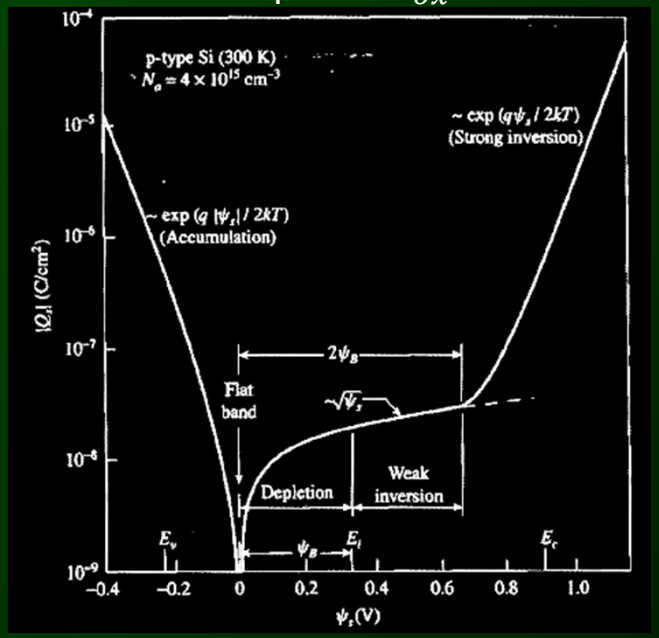
Q_S:

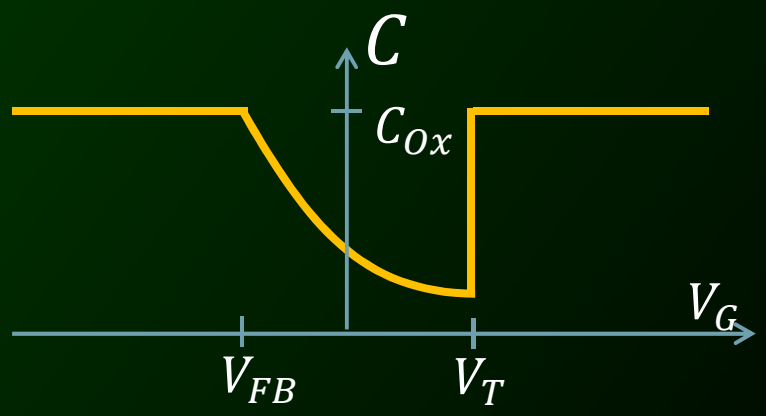


MOS C-V Characteristics



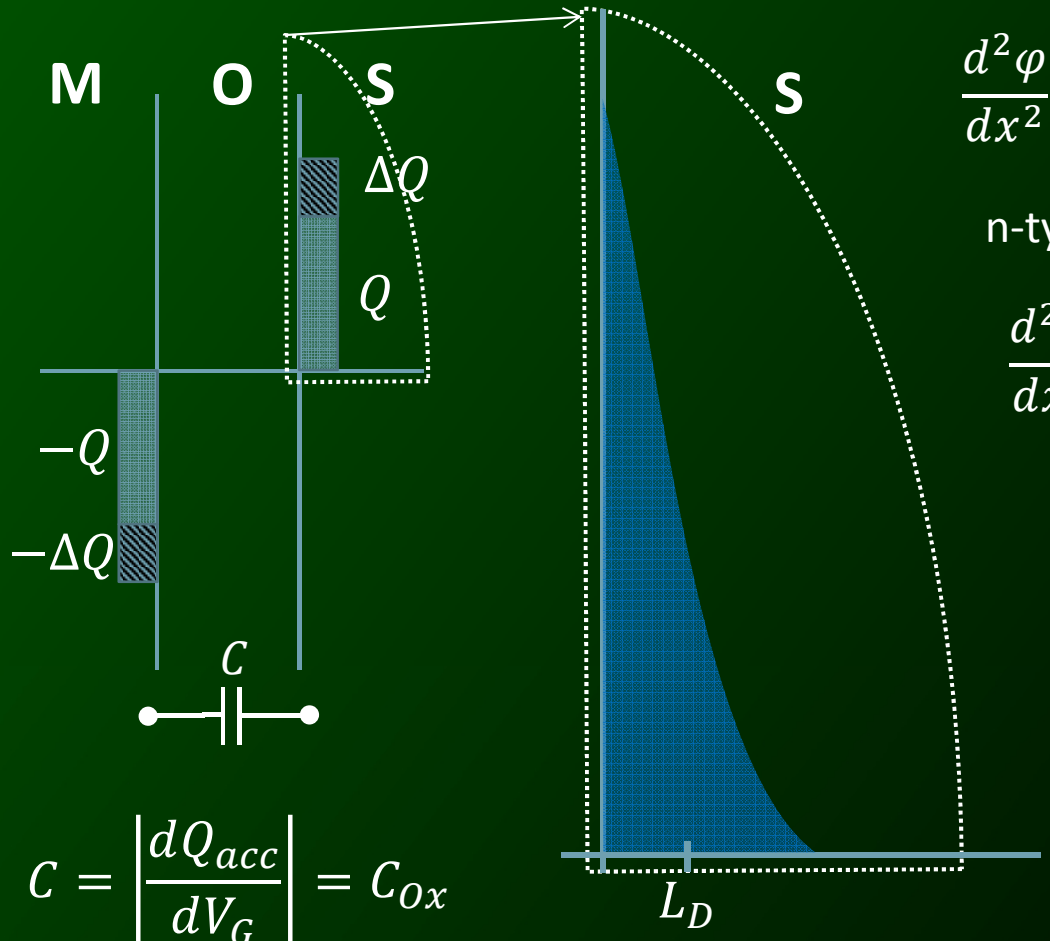
$$C = \left|\frac{dQ_{gate}}{dV_G}\right| = \left|\frac{dQ_S}{dV_G}\right|$$





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.



$$C = \left|\frac{dQ_{acc}}{dV_G}\right| = C_{Ox}$$

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

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

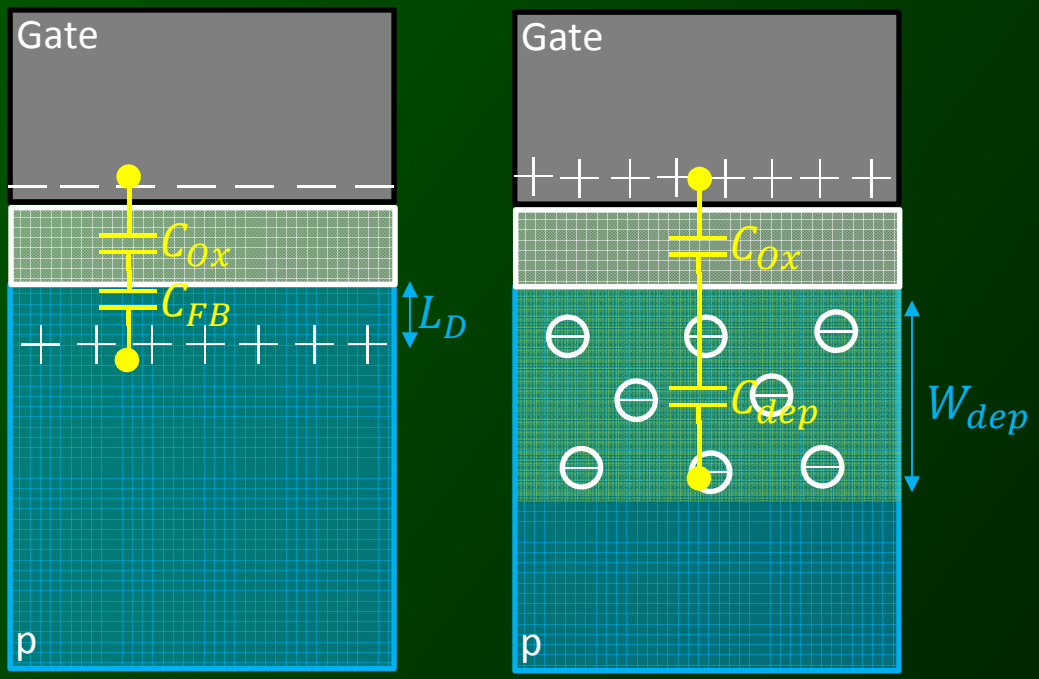
$$\frac{d^2\varphi}{dx^2} = \frac{q}{\epsilon}\left(N_D - n_i e^{-(\varphi_n - \varphi)/\varphi_{th}}\right)$$

$$= \frac{q}{\epsilon} N_D \left(1 - e^{\varphi/\varphi_{th}}\right)$$

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

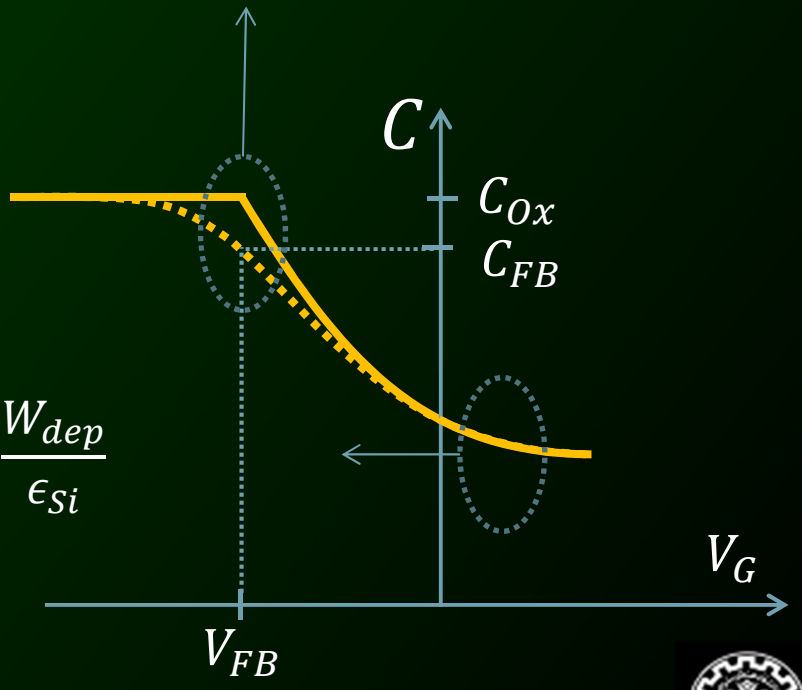
$$L_D = \sqrt{\frac{\epsilon\, kT}{q^2 N_D}}$$

Flat-Band Capacitance



$$C_{FB} = \frac{C_{Ox} C_D}{C_{Ox} + C_D}$$

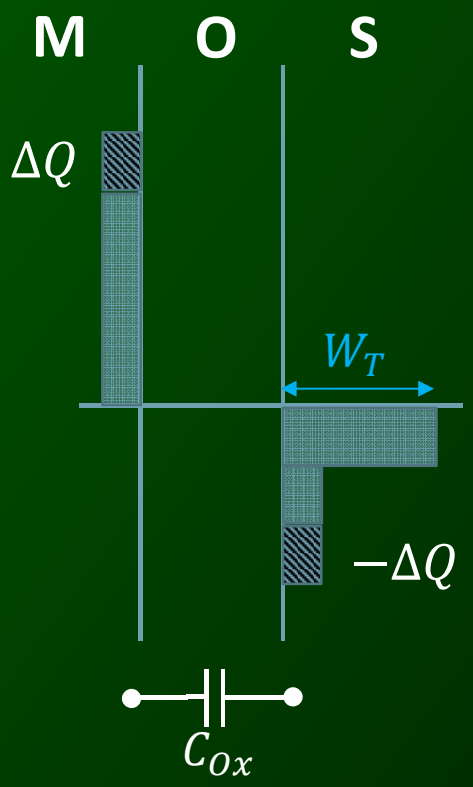
$$\rightarrow \frac{1}{C_{FB}} = \frac{t_{Ox}}{\epsilon_{Ox}} + \frac{L_D}{\epsilon_{Si}}$$



$$\frac{1}{C} = \frac{t_{Ox}}{\epsilon_{Ox}} + \frac{W_{dep}}{\epsilon_{Si}}$$

Capacitance in Inversion

CASE 1: Inversion-layer charge can be supplied/removed quickly enough to respond to changes in the gate voltage.



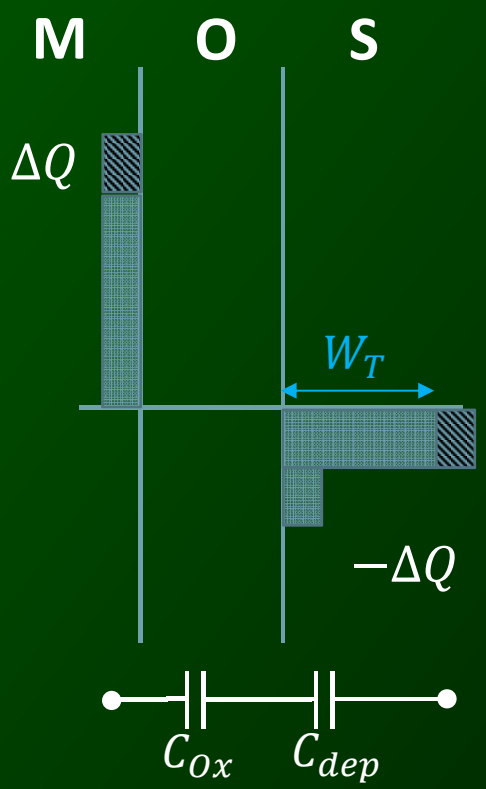
→ Incremental charge is effectively added/subtracted at the surface of the substrate.

Time required to build inversion-layer charge = $2N_A\tau_0/n_i$, where τ_0 = minority-carrier lifetime at the surface

$$C = \left|\frac{dQ_{inv}}{dV_G}\right| = C_{ox}$$

Capacitance in Inversion

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 WT in the substrate.

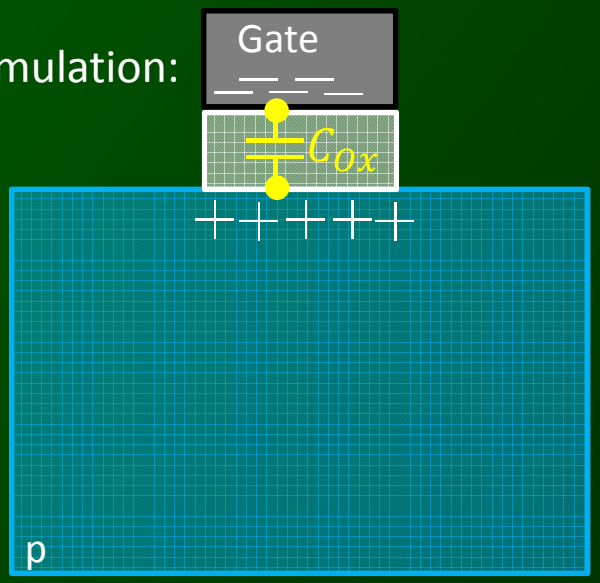
$$\frac{1}{C} = \frac{1}{C_{Ox}} + \frac{1}{C_{dep}}$$

$$= \frac{1}{C_{Ox}} + \frac{W_T}{\epsilon_{Si}}$$

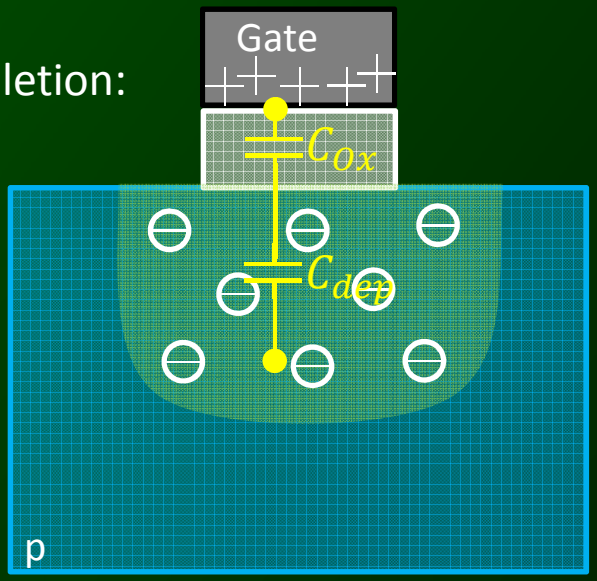
$$= \frac{1}{C_{Ox}} + \sqrt{\frac{2(2\varphi_F)}{qN_A\epsilon_{Si}}} \equiv \frac{1}{C_{min}}$$

Supply of Inversion Charge

Accumulation:

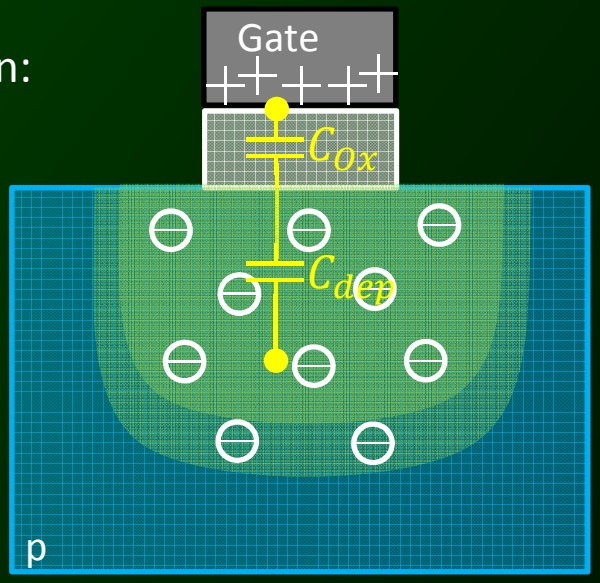


Depletion:

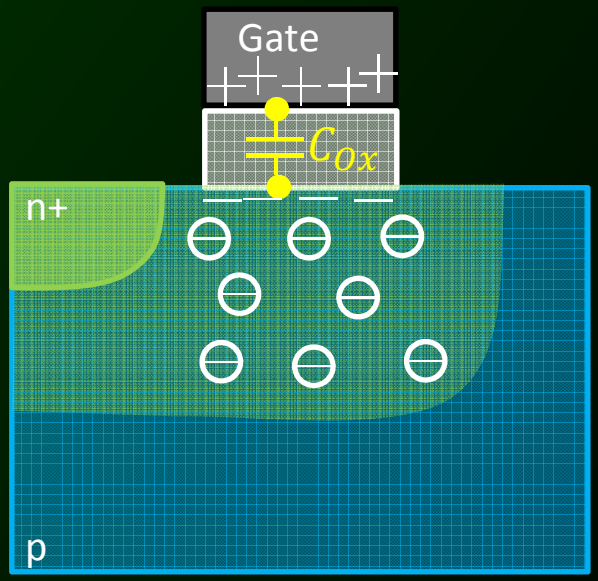


Inversion:

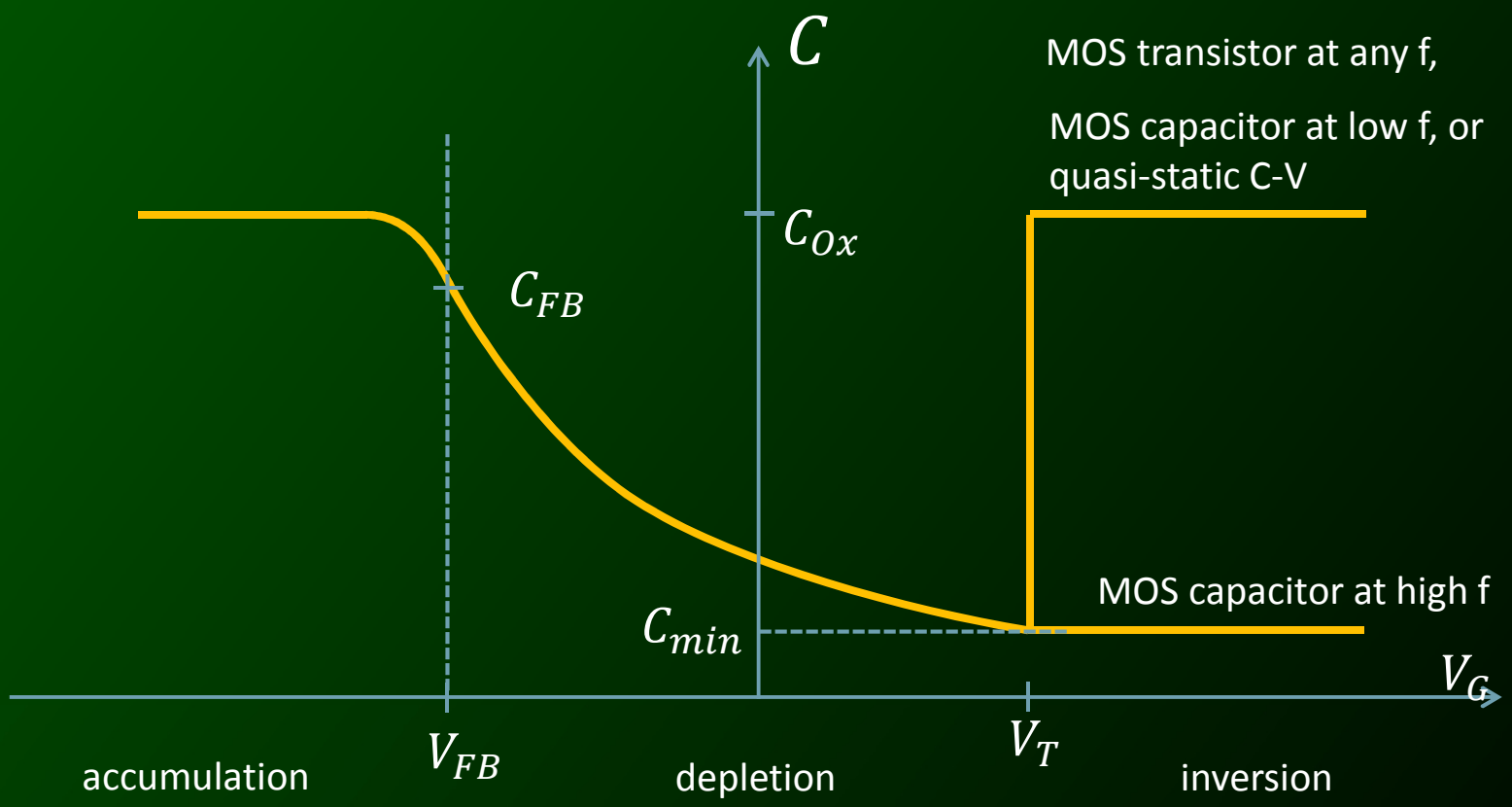
Case 1



Case 2



Boundary Condition

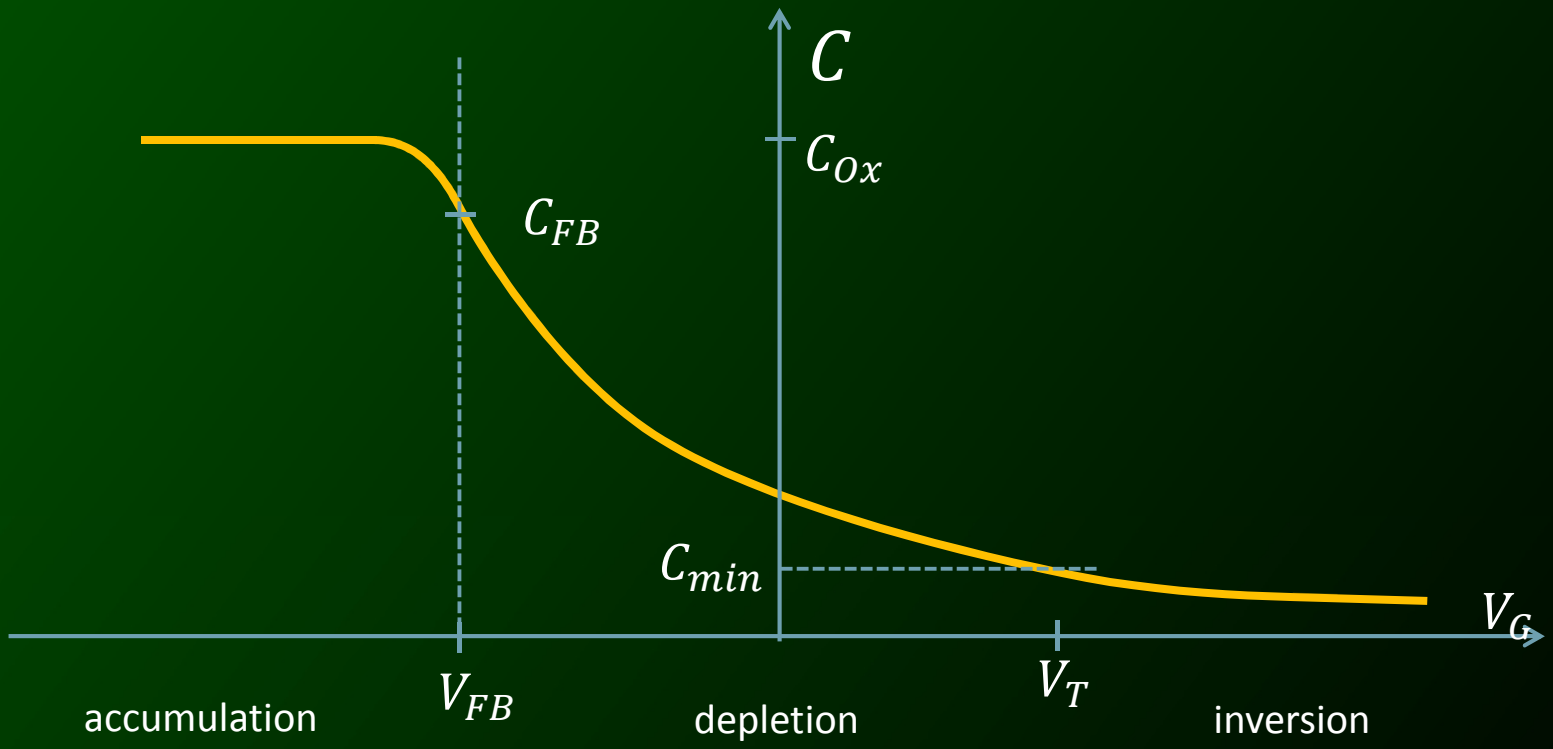


Deep Depletion

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}

⇒ depletion depth W increases as V_G increases

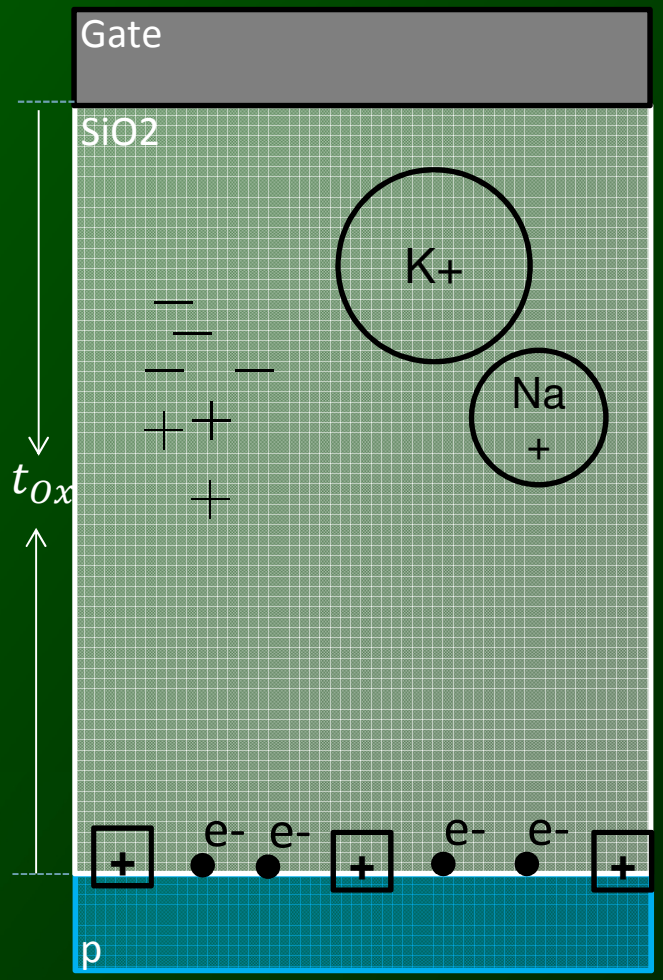
⇒ C decreases as V_G increases



1. I
2.
3.
4.
5.

Oxide Charges

In real MOS devices, there is always some charge in the oxide and at the Si/oxide interface.



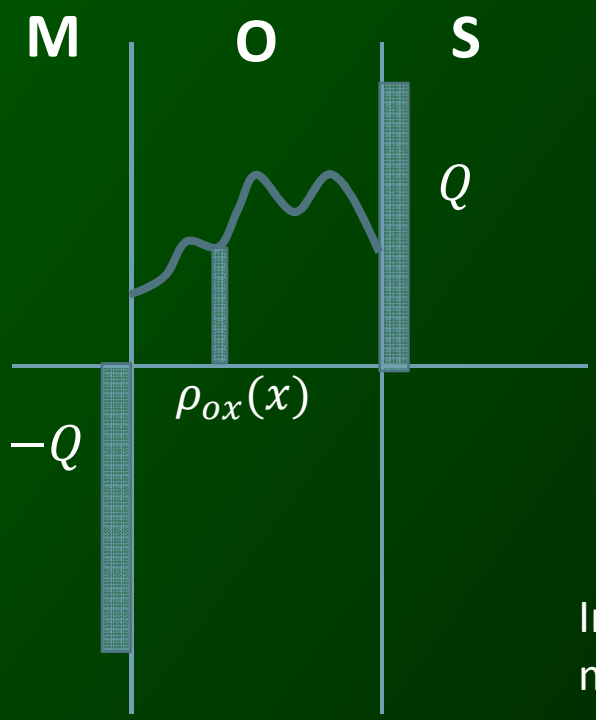
In the oxide:

1. Trapped charge Qox
   High-energy electrons and/or holes injected into oxide
2. Mobile charge QM
   Alkali-metal ions, which have sufficient mobility to drift in oxide under an applied electric field

At the interface:

1. Fixed charge QF
   Excess Si (+)
2. Trapped charge QIT
   Dangling bonds

Effect of Oxide Charges



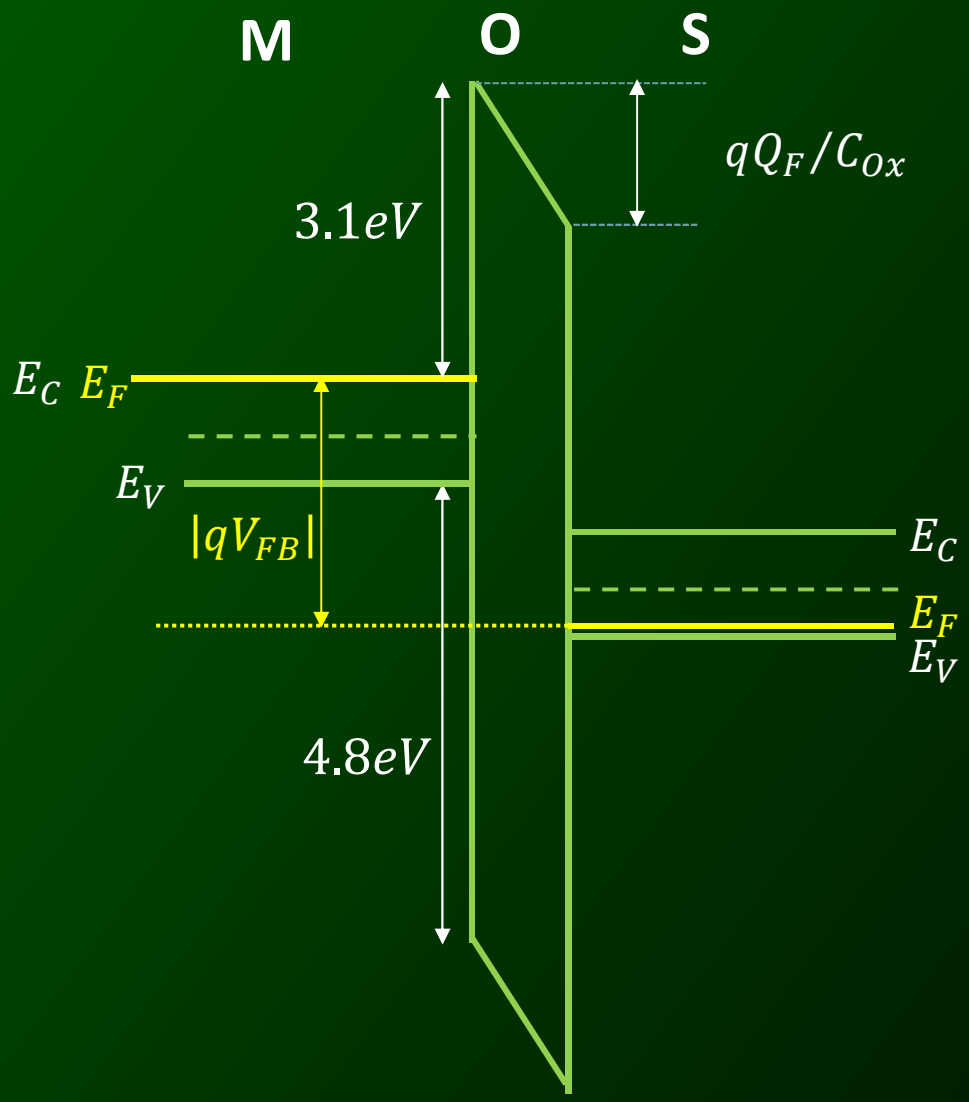
In general, charges in the oxide cause a shift in the gate voltage required to reach the threshold condition:

$$\Delta V_T = -\frac{1}{\epsilon_{SiO_2}} \int_0^{t_{Ox}} x\rho_{0x}(x)dx$$

(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

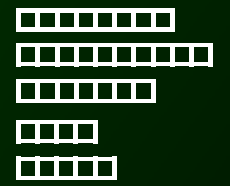
Fixed Oxide Charges Q_F



$$V_{FB} = \varphi_{ms} - \frac{Q_F}{C_{Ox}}$$

Parameter Extraction from C-V

1. I
2.
3.
4.
5.



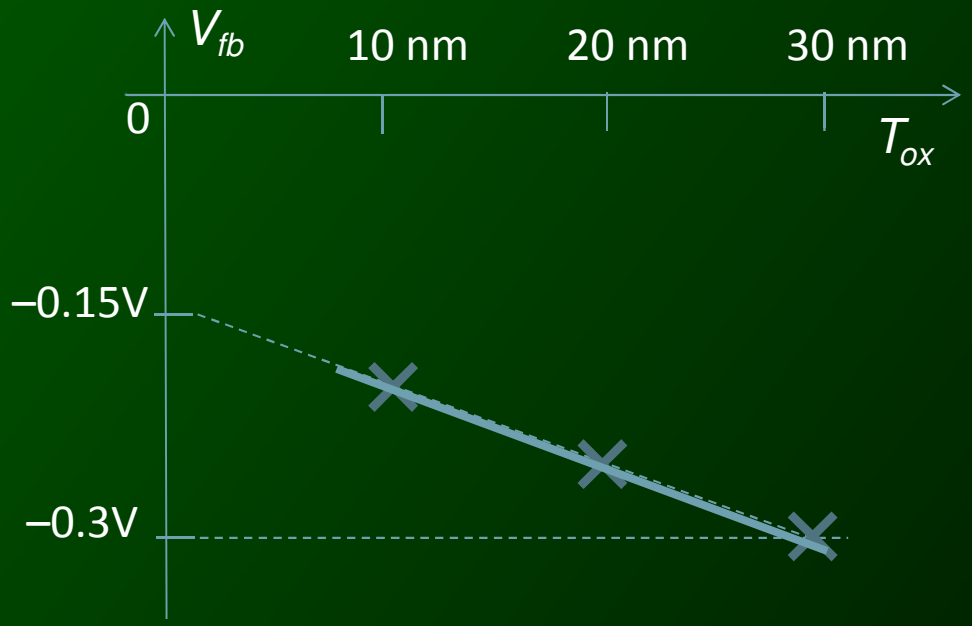
From a single C-V measurement, we can extract much information about the MOS device.

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

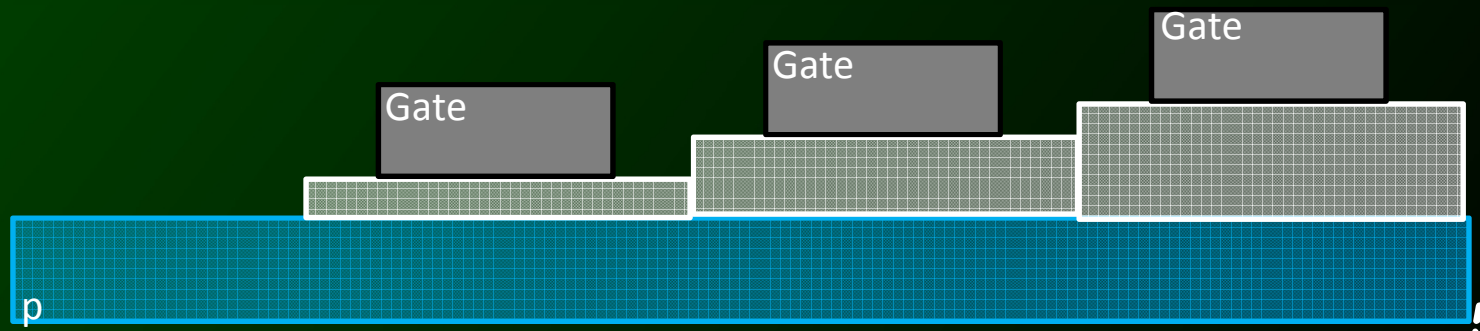
- From $C_{max} = C_{Ox}$ we determine the oxide thickness x_0
- From C_{min} and C_{Ox} we determine substrate doping (by iteration)
- From substrate doping and C_{Ox} we calculate the flat-band capacitance C_{FB}
- From the C-V curve, we can find
- From φ_M, φ_S, C_{Ox}, and V_{FB} we can determine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

Determination of φ_M and Q_F

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

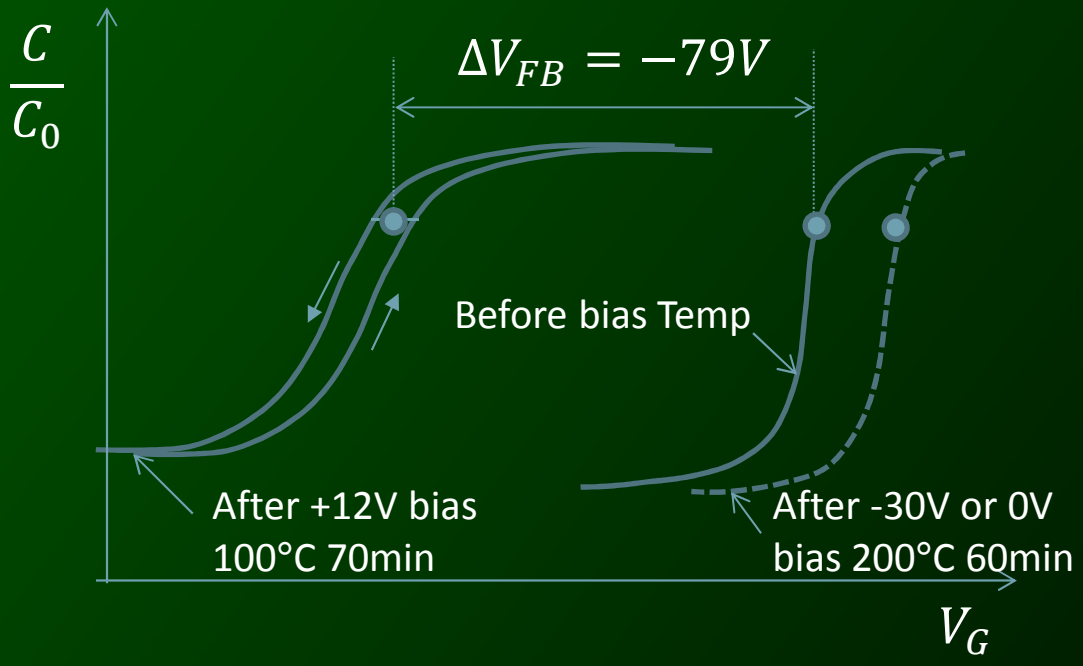


$$V_{FB} = \varphi_{ms} - \frac{Q_F}{\epsilon_{SiO_2}} t_{Ox}$$



Mobile Ions

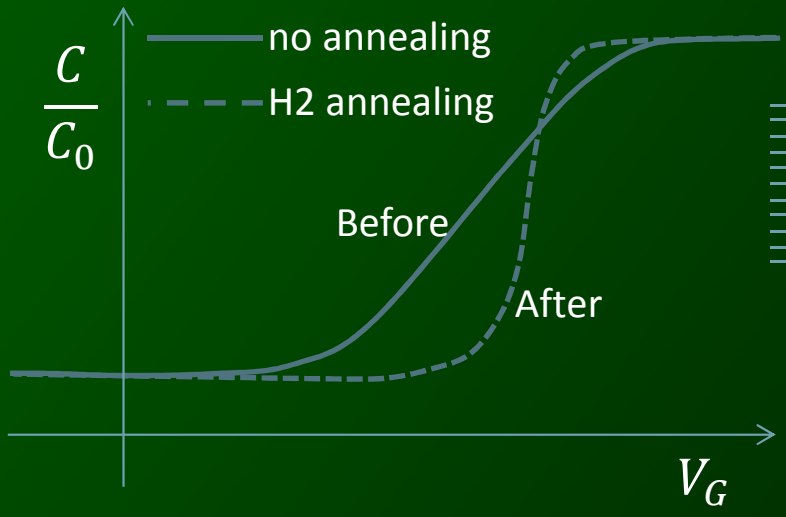
Odd shifts in C-V characteristics were once a mystery:

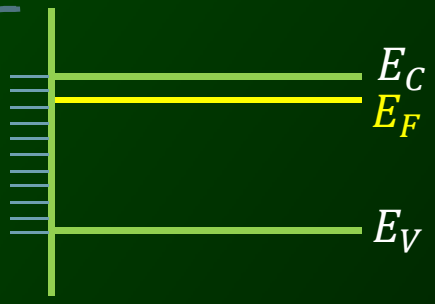


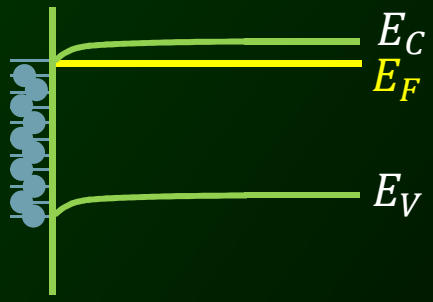
$$\Delta V_{FB} = -\frac{Q_M}{C_{Ox}}$$

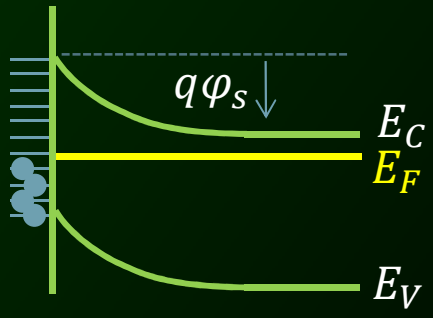
Source of problem: Mobile charge moving to/away from interface, changing charge centroid

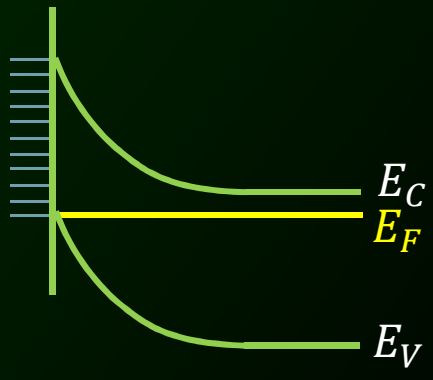
Interface Traps









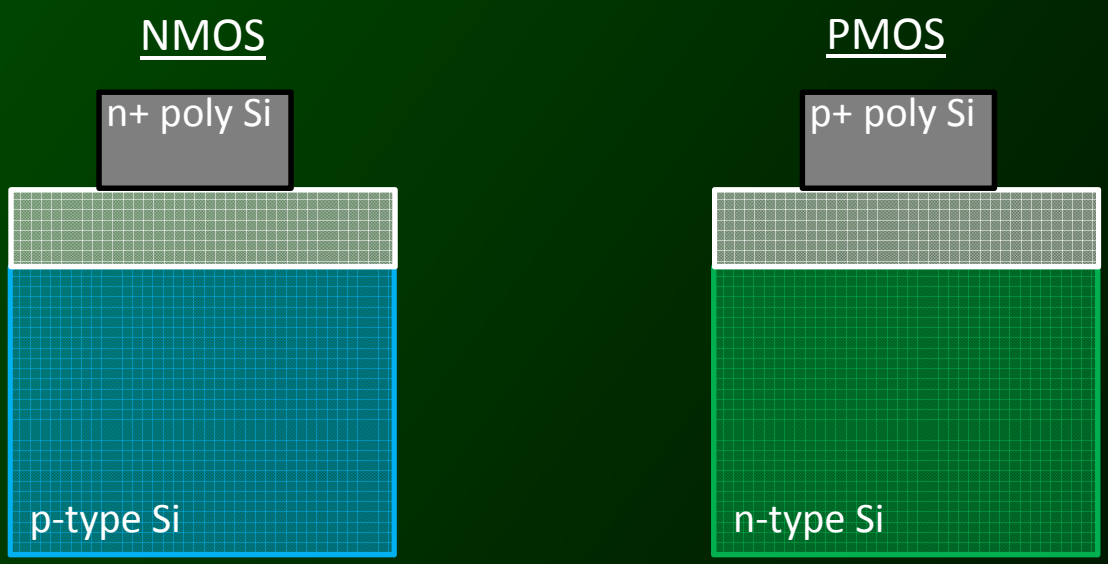


Traps cause “sloppy” C-V and also greatly degrade mobility in channel

$$\Delta V_G = -\frac{Q_{IT}(\varphi_s)}{C_{Ox}}$$

Poly-Si Gate Depletion

A heavily doped film of polycrystalline silicon (poly-Si) is typically employed as the gate-electrode material in modern MOS devices.

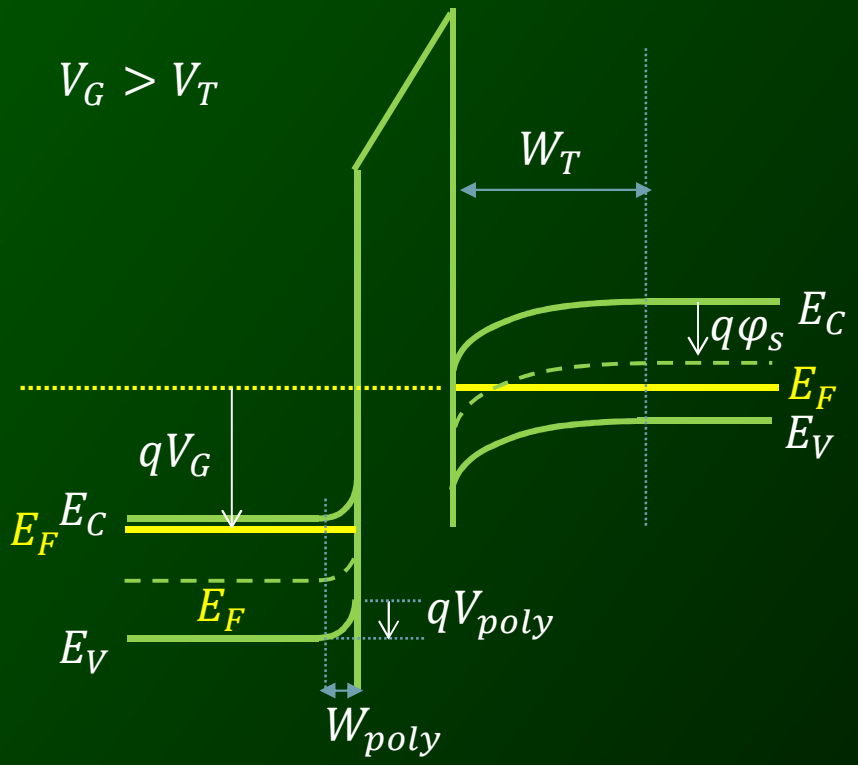


There are practical limits to the electrically active dopant concentration (usually less than 1×10^{20} cm^{-3})

⇒ The gate must be considered as a semiconductor, rather than a metal

MS Junction (Poly Gate)

Si biased to inversion:



n+ poly Si gate p-type Si

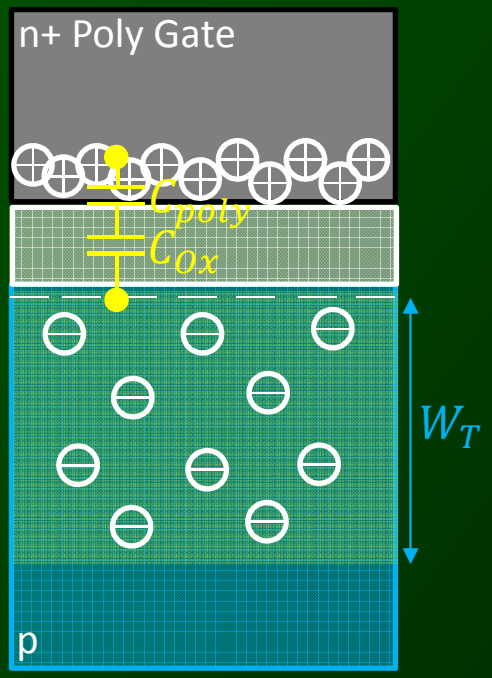
V_G is effectively reduced:

$$Q_{inv} = C_{Ox}(V_G - V_{poly} - V_T)$$

$$W_{poly} = \sqrt{\frac{2\epsilon_{Si}V_{poly}}{qN_{poly}}}$$

How can gate depletion be minimized?

Gate Depletion Effect



Gauss's Law dictates:

$$W_{poly} = \frac{\epsilon_{Ox}\mathcal{E}_{Ox}}{qN_{poly}}$$

t_{Ox} is effectively increased:

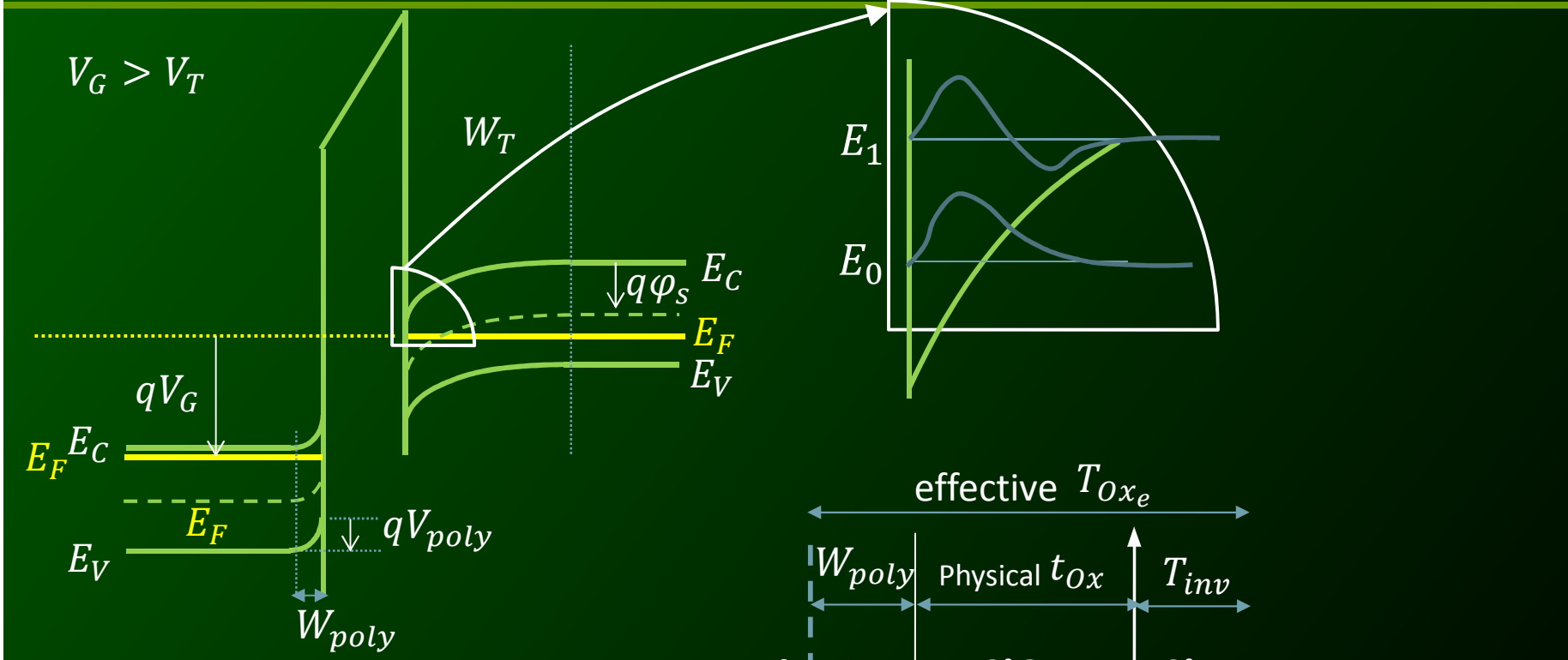
$$C = \left(\frac{1}{C_{Ox}} + \frac{1}{C_{poly}}\right)^{-1} = \left(\frac{t_{Ox}}{\epsilon_{SiO_2}} + \frac{W_{poly}}{\epsilon_{Si}}\right)^{-1}$$

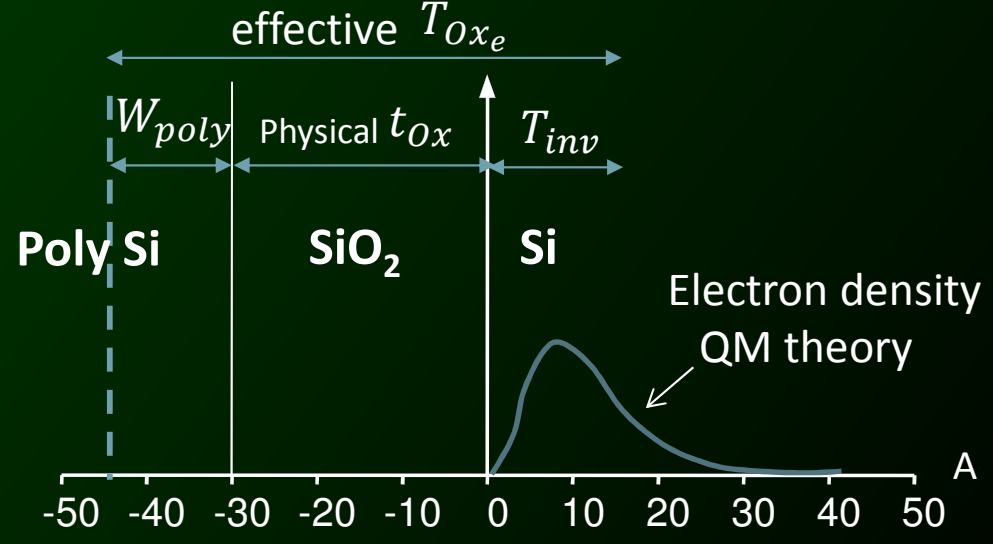
$$= \frac{\epsilon_{SiO_2}}{t_{Ox} + \frac{1}{3}W_{poly}}$$

$$Q_{inv} = C_{Ox}\left(V_G - V_{poly} - V_T\right)$$

$$Q_{inv} = \frac{\epsilon_{SiO_2}}{t_{Ox} + \frac{1}{3}W_{poly}}\left(V_G - V_T\right)$$

Inversion Layer Thickness T_{inv}



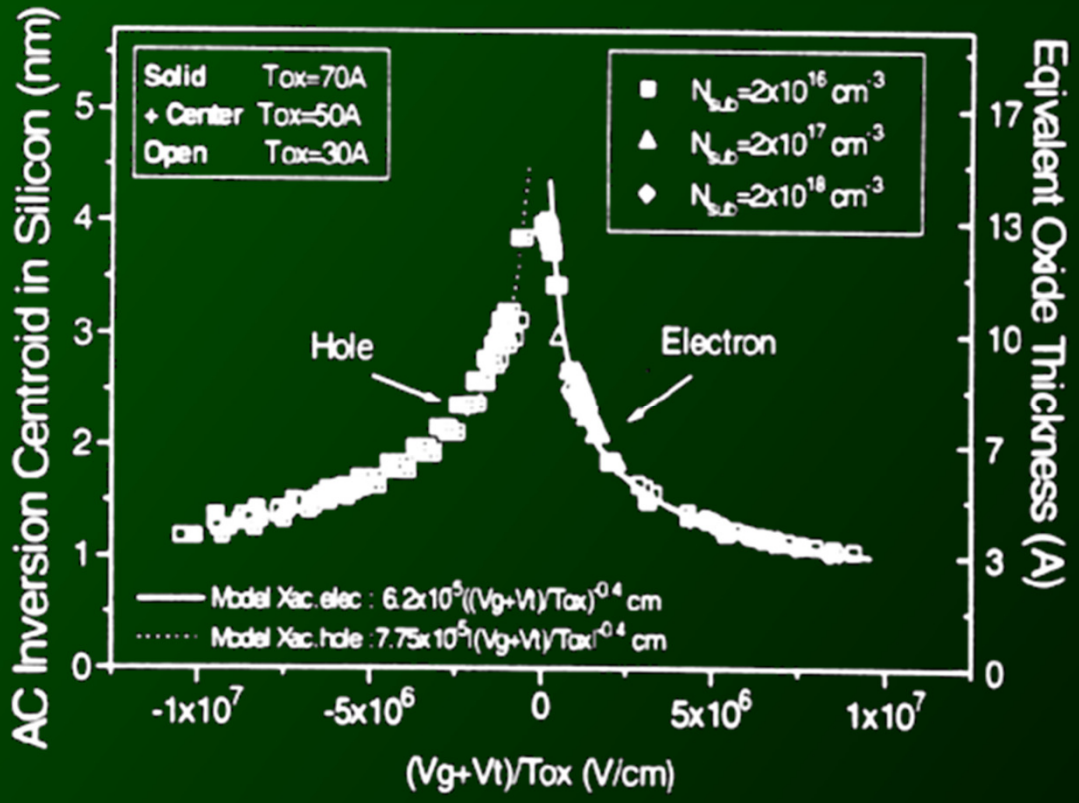


The average inversion-layer location below the Si/SiO2 interface is called the inversion-layer thickness, T_{inv} .

Effective Oxide Thickness , T_{Oxe}

1. I
2.
3.
4.
5.

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

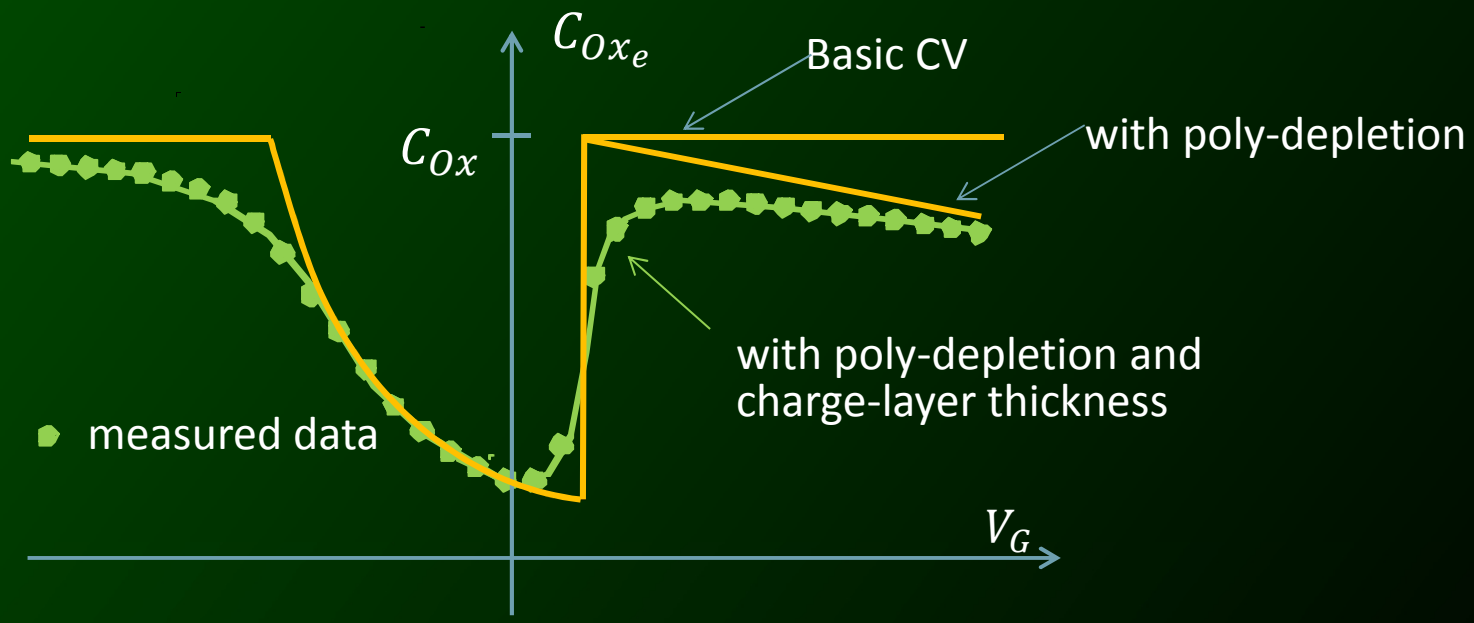


$(V_G + V_T)/T_{Ox}$ can be shown to be the average electric field in the inversion layer. Tinv of holes is larger than that of electrons because of the difference in effective masses.

Effective Oxide Capacitance , C_{Oxe}

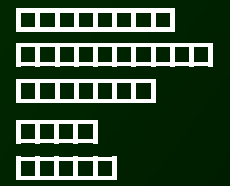
$$Q_{inv} = C_{Ox_e}(V_G - V_T)$$

$$T_{Ox_e} = t_{Ox} + \frac{1}{3}W_{poly} + \frac{1}{3}T_{inv}$$

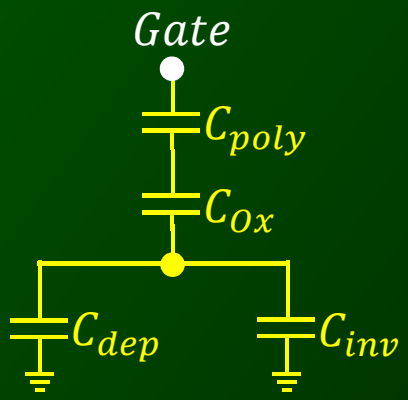


MOS Cap: Equivalent Circuit in Depletion & Inversion

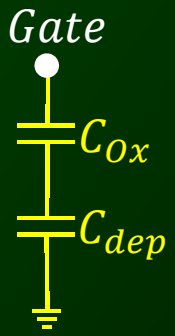
1. I
2.
3.
4.
5.



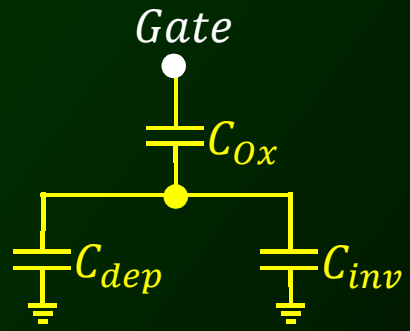
General case for both depletion and inversion regions.



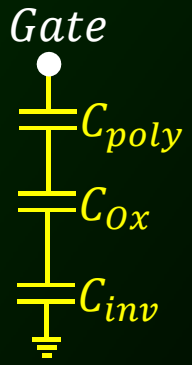
In the depletion regions



$V_G \approx V_T$



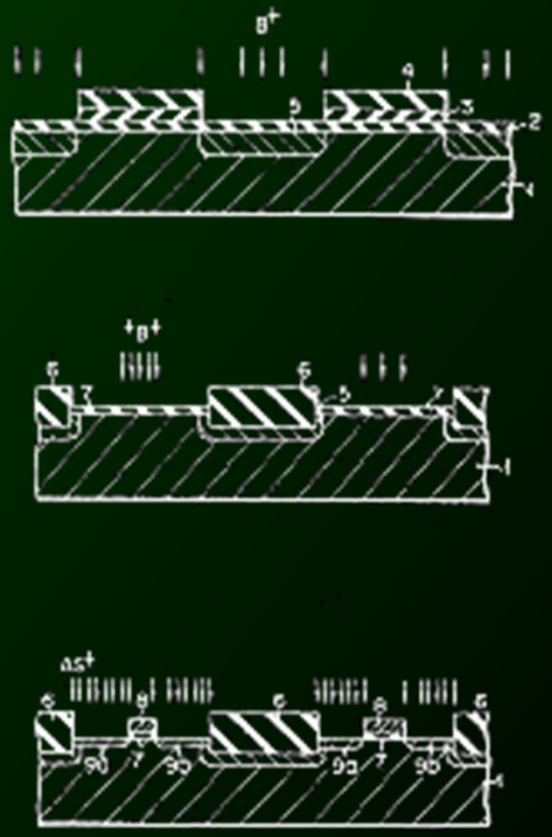
Strong inversion



V_T Adjustment by Ion Implantation

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

- A relatively small dose NI (units: ions/cm2) 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.

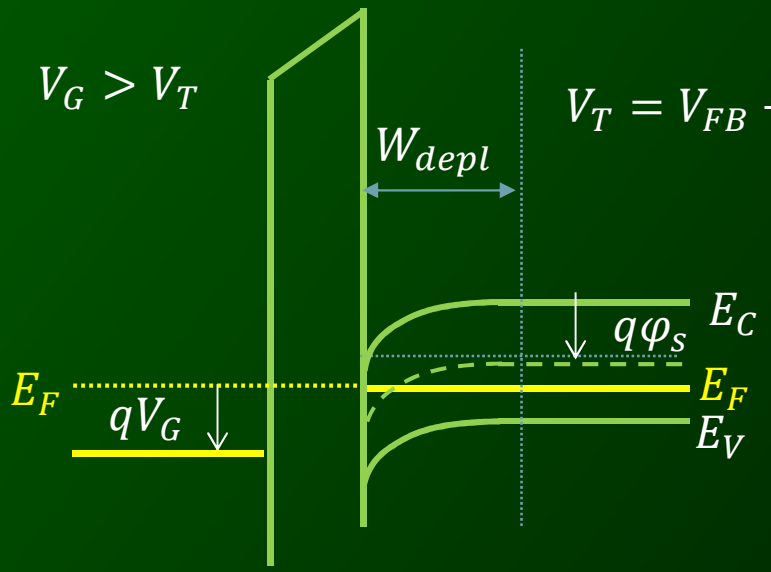


$$\Delta V_T = -\frac{qN_I}{C_{ox}}$$

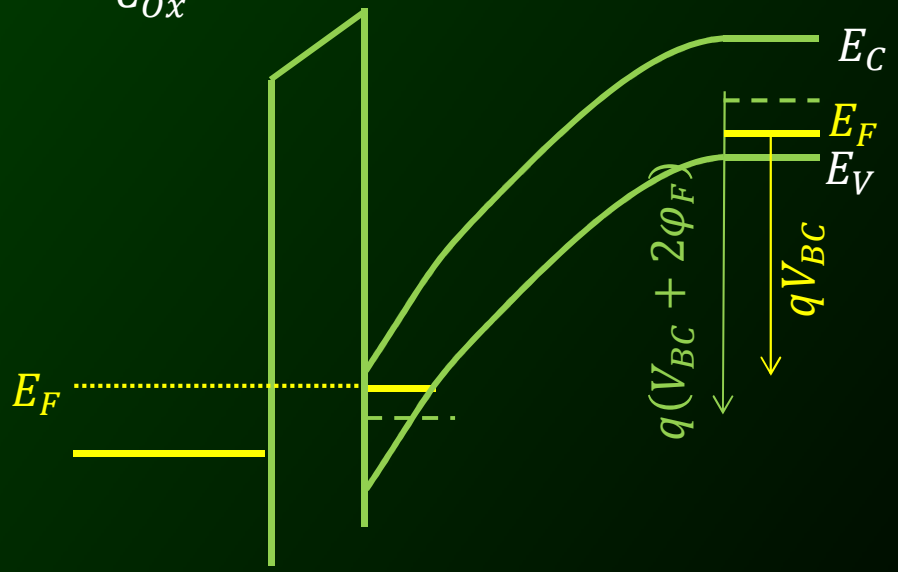
$N_I > 0$ for donor atoms

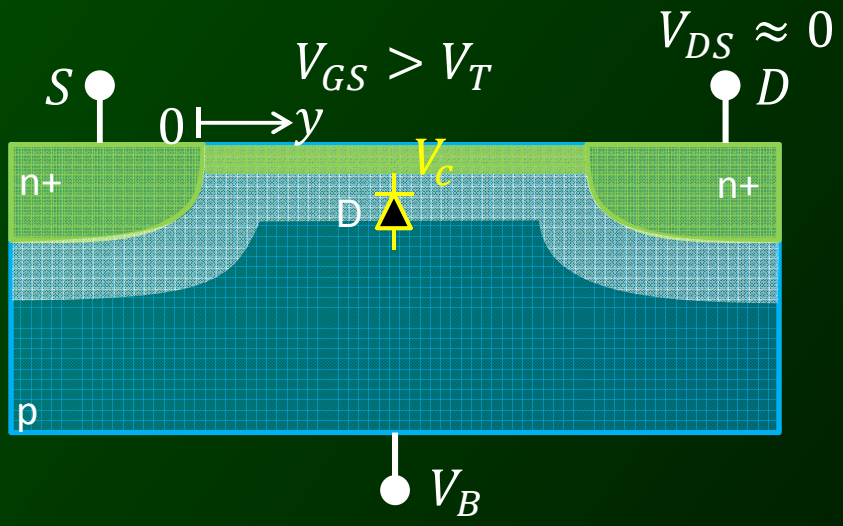
$N_I < 0$ for acceptor atoms

Dynamic V_T Adjustment Bulk Voltage



$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$

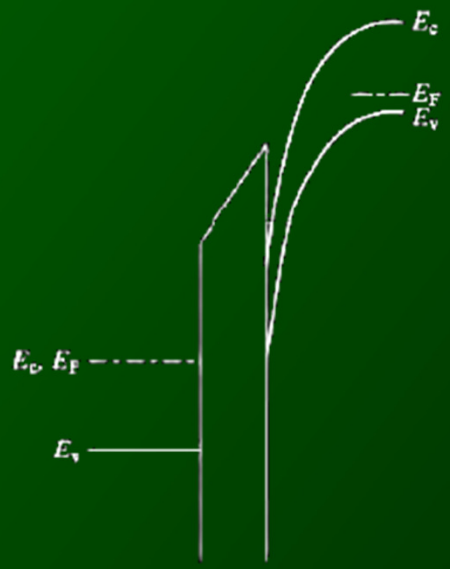




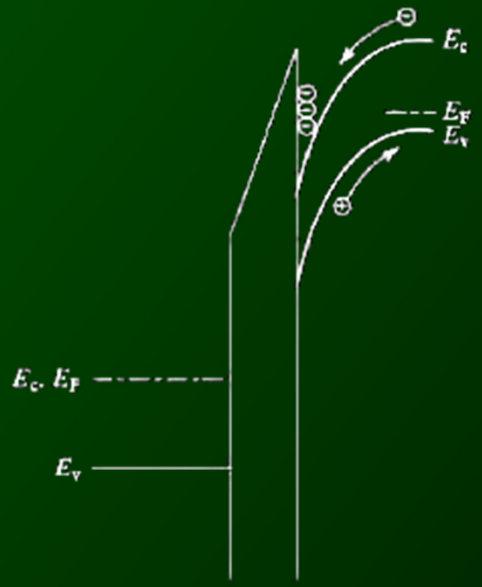
$$V_T = V_{FB} + V_C + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$

$$Q_{depl} = \sqrt{2qN_A\epsilon_{Ox}(V_{BC} + 2\varphi_F)}$$

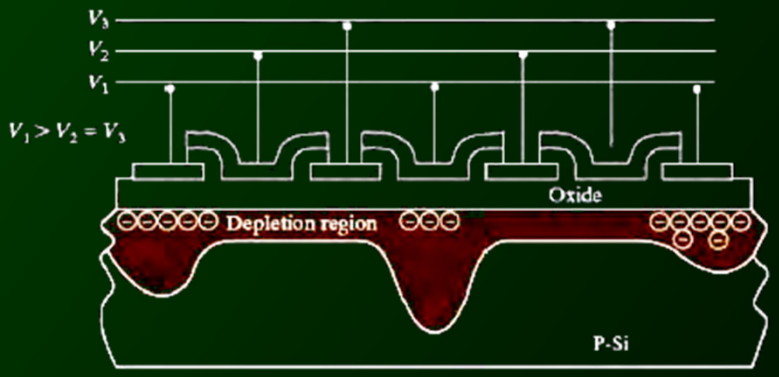
CCD Imager and CMOS Imager

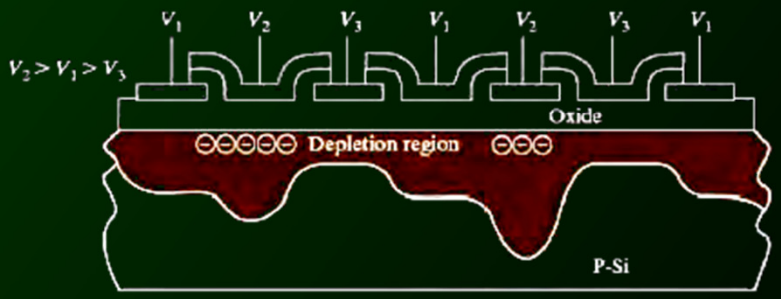


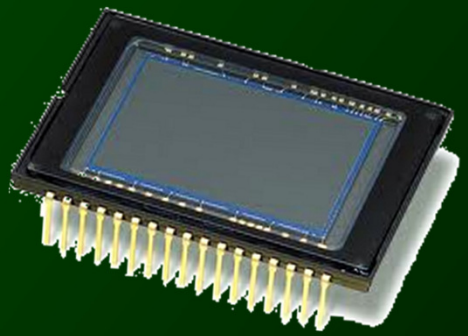
Deep depletion, Q_{inv}= 0

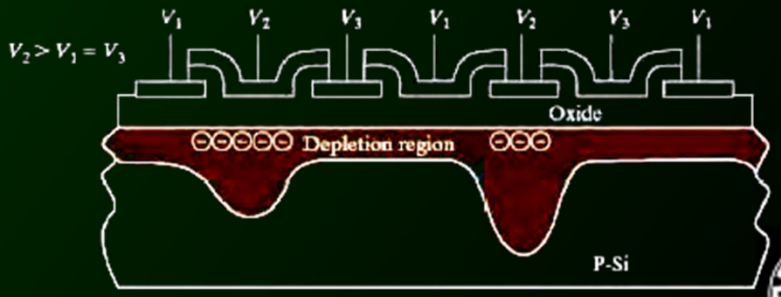


Exposed to light



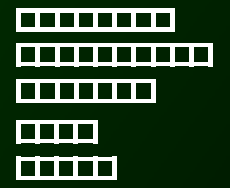






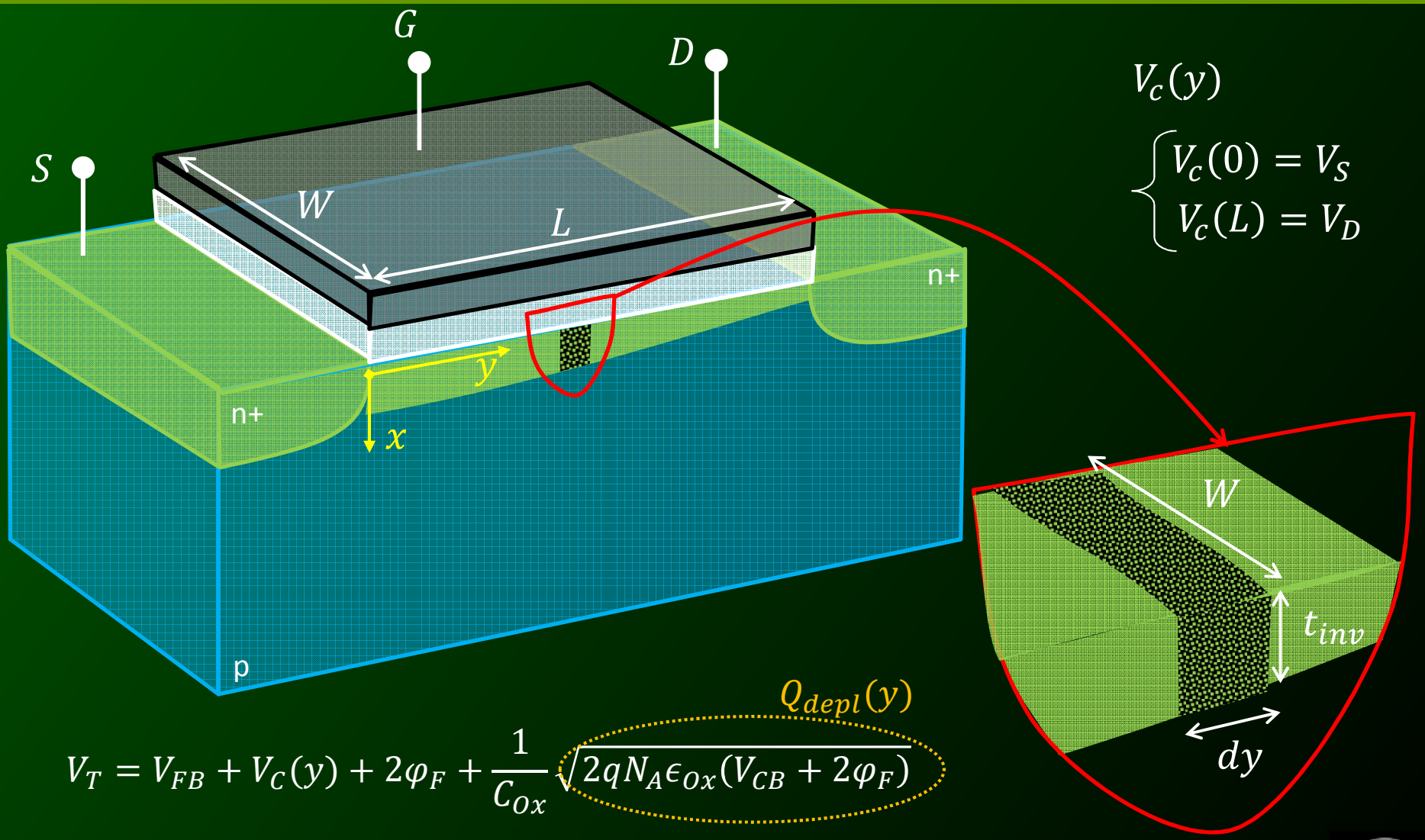
Boundary Condition

1. I
2.
3.
4.
5.



1. I
2.
3.
4.
5.

MOSFET I-V Curve



$$V_c(y)$$

$$\begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases}$$

$$V_T = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}$$

$Q_{depl}(y)$

MOSFET I-V Curve

$$V_T(y) = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\underbrace{\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}}_{Q_{depl}(y)}$$

$V_c(y)$

$$\begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases}$$

$$Q_{inv} = -C_{Ox}(V_G - V_T(y))$$

$$Q_{inv} = -C_{Ox}\left(V_G - V_{FB} - V_C(y) - 2\varphi_F - \frac{Q_{depl}(y)}{C_{Ox}}\right)$$

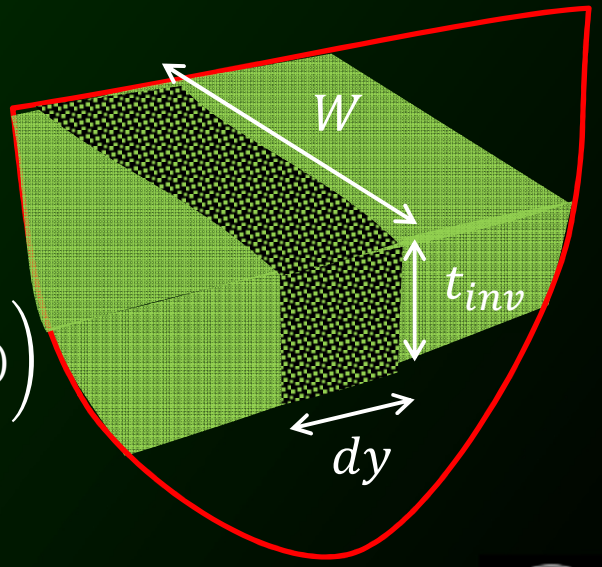
Depletion Region Approximation:

$Q_{inv}(y)$ but $Q_{depl}(y) \approx Q_{depl}(0)$

$$Q_{depl}(y) \approx \sqrt{2qN_A\epsilon_{Ox}(V_{SB} + 2\varphi_F)}$$

$$Q_{inv} = -C_{Ox}\left(V_G - \underbrace{V_{FB} - V_S - 2\varphi_F - \frac{Q_{depl}(0)}{C_{Ox}}}_{V_T(0)} + V_s - V_C(y)\right)$$

$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$



MOSFET I-V Curve

1. I
2.
3.
4.
5.

$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$

Simply call $V_T(0)$ as V_T $\quad Q_{inv}(y) = -C_{Ox}(V_G - V_T + V_s - V_C(y))$

$V_c(y)$

$$\begin{cases} V_c(0) = V_S \\ V_c(L) = V_D \end{cases}$$

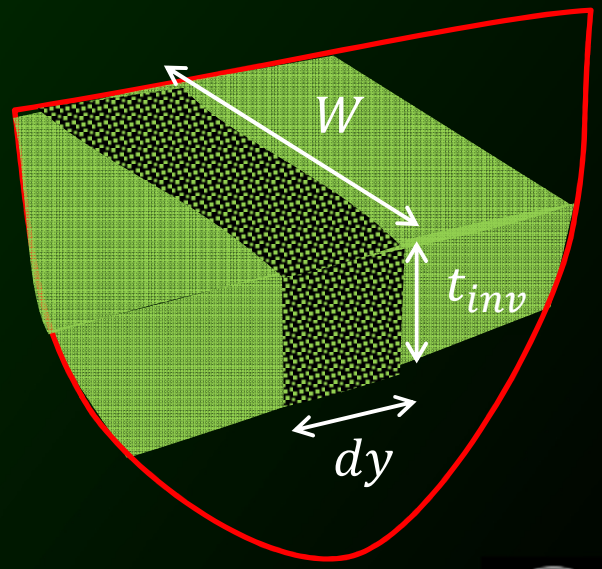
$$dV_c(y) = I_{DS} \cdot dR = I_{DS}\frac{dy}{\sigma W t_{inv}} = \frac{I_{DS}dy}{(q\mu_{eff}n)Wt_{inv}} = \frac{I_{DS}dy}{\underbrace{(qnt_{inv})}_{-Q_{inv}(y)\ [C/cm^2]}\mu_{eff}W}$$

$$\int_0^L I_{DS}dy = \int_{V_S}^{V_D} -\mu_{eff}W Q_{inv}(y)dV_c$$

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

$$I_{DS} = \frac{W}{L}\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \tfrac{1}{2}V_{DS}\right)V_{DS}$$

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



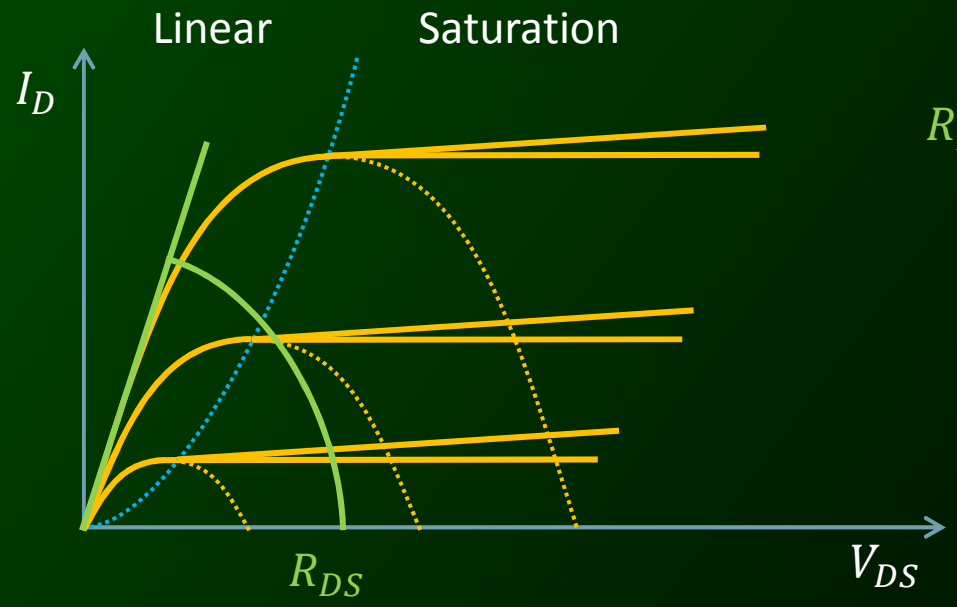
MOSFET I-V Curve

1. I
2.
3.
4.
5.

Linear

$$I_{DS} = \begin{cases} \frac{W}{L}\,\mu_{eff} C_{Ox}\left(V_{GS} - V_T - \frac{1}{2}V_{DS}\right)V_{DS} & V_{DS} < V_{DS_{sat}} \\ \frac{W}{2L}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)^2 & V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \end{cases}$$

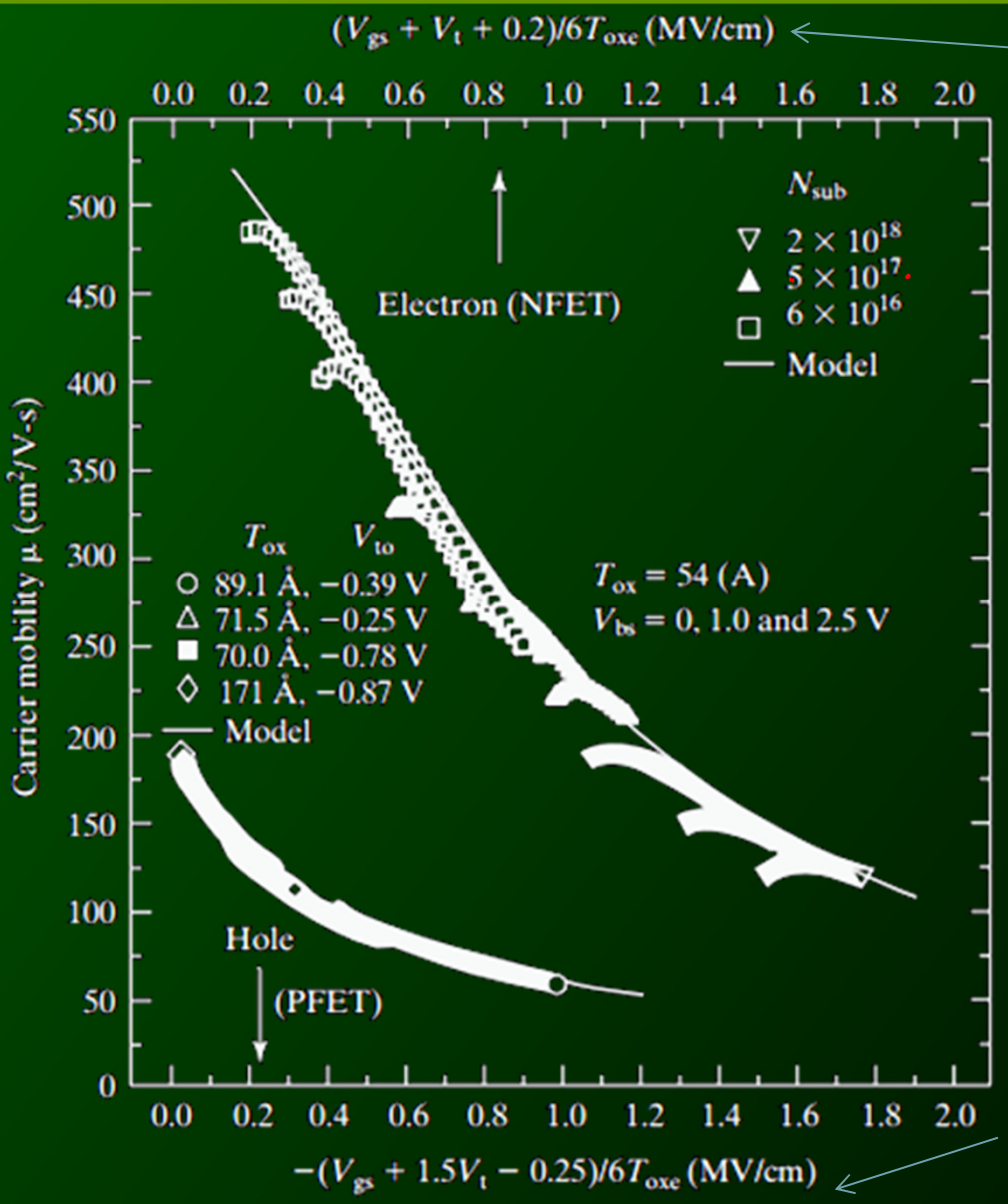
Saturation



$$R_{DS} = \left(\left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{DS}=0}\right)^{-1}$$

$$= \left(\frac{W}{L}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)\right)^{-1}$$

Field-Effect Mobility, μ_{eff}



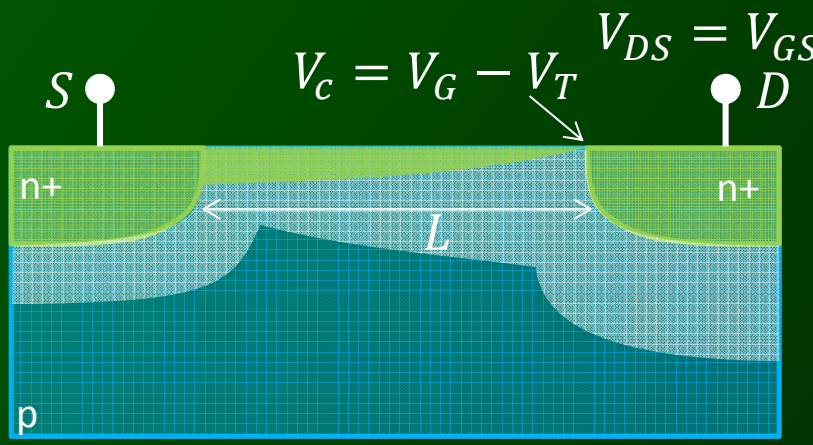
Effective vertical electric filed in the inversion layer for NMOS

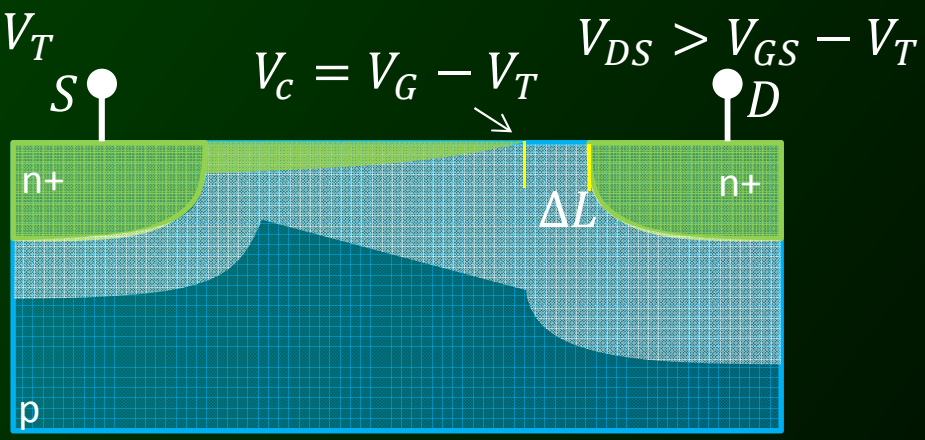
Scattering mechanisms:

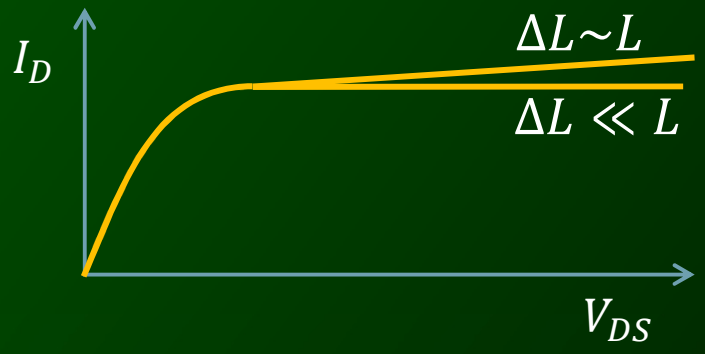
- Columbic scattering
- phonon scattering
- surface roughness scattering

Effective vertical electric filed in the inversion layer for PMOS

MOSFET Saturation Region of Operation







As V_D is increased above V_G-V_T, the length ΔL of the "pinch-off" region increases. The voltage applied across the inversion layer is always V_{Dsat}=V_{GS}-V_T, and so the current saturates.

If ΔL is significant compared to L, then I_{DS} will increase slightly with increasing V_{DS}>V_{Dsat}, due to "channel-length modulation"

"Square Law Theory" ?

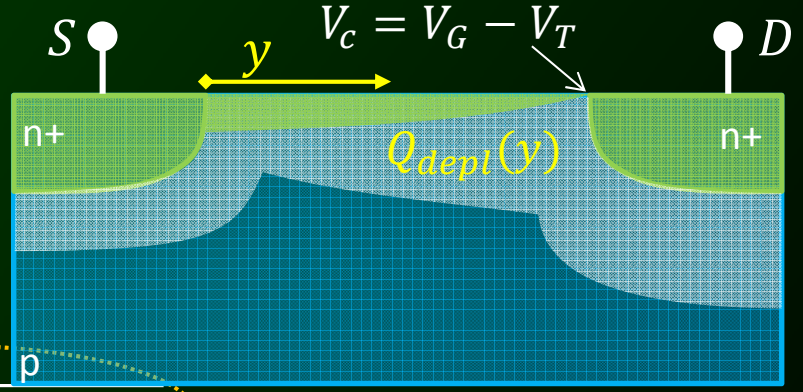
1. I
2.
3.
4.
5.

$$I_{DS} = \begin{cases} \frac{W}{L}\mu_{eff}C_{Ox}\left(V_{GS} - V_T - \frac{1}{2}V_{DS}\right)V_{DS} & V_{DS} < V_{DS_{sat}} \quad \text{Linear} \\ \frac{W}{2L}\mu_{eff}C_{Ox}(V_{GS} - V_T)^2 & V_{DS} < V_{DS_{sat}} = V_{GS} - V_T \quad \text{Saturation} \end{cases}$$

Depletion Region Approximation:

$Q_{inv}(y)$ but $Q_{depl}(y) \approx Q_{depl}(0)$

$$Q_{depl}(y) \approx \sqrt{2qN_A\epsilon_{Ox}(V_{SB} + 2\varphi_F)}$$



$$V_T(y) = V_{FB} + V_C(y) + 2\varphi_F + \frac{1}{C_{Ox}}\sqrt{2qN_A\epsilon_{Ox}(V_{CB} + 2\varphi_F)}$$

$Q_{depl}(y)$

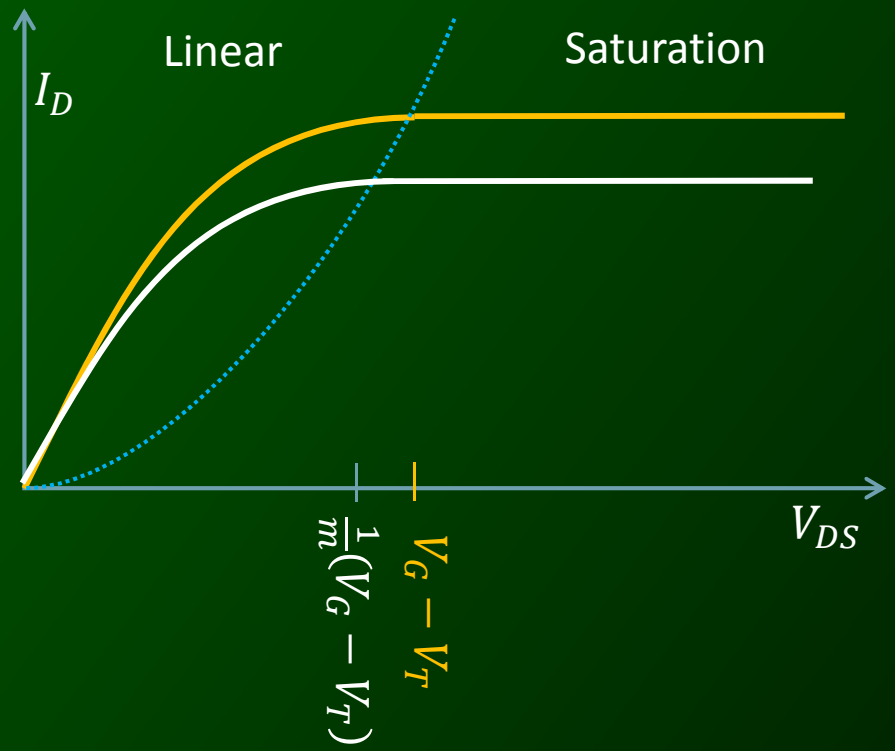
$$Q_{inv}(y) = -C_{Ox}(V_G - V_T(0) + V_s - V_C(y))$$

$Q_{depl}(y) > Q_{depl}(0)$

$V_T(y) > V_T(0)$

$Q'_{inv}(y) < Q_{inv}(y)$ $I'_{DS} < I_{DS}$

Modified (Bulk-Charge) I-V Model



Linear $\quad V_{DS} < V_G - V_T$

$$I_{DS} = \frac{W}{L}\,\mu_{eff} C_{Ox}\left(V_{GS} - V_T - \tfrac{1}{2}V_{DS}\right)V_{DS}$$

Linear $\quad V_{DS} < \frac{1}{m}(V_G - V_T)$

$$I_{DS} = \frac{W}{L}\,\mu_{eff} C_{Ox}\left(V_{GS} - V_T - \tfrac{m}{2}V_{DS}\right)V_{DS}$$

Saturation $\quad V_{DS} > V_G - V_T$

$$I_{DS} = \frac{W}{2L}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)^2$$

Saturation $\quad V_{DS} > \frac{1}{m}(V_G - V_T)$

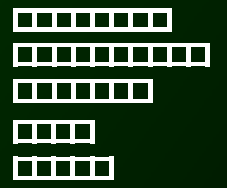
$$I_{DS} = \frac{W}{2mL}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)^2$$

Bulk charge factor

$$m = 1 + \frac{C_{dep}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

Typically $\quad 1.1 < m < 1.4$

The Body Effect



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

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

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

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

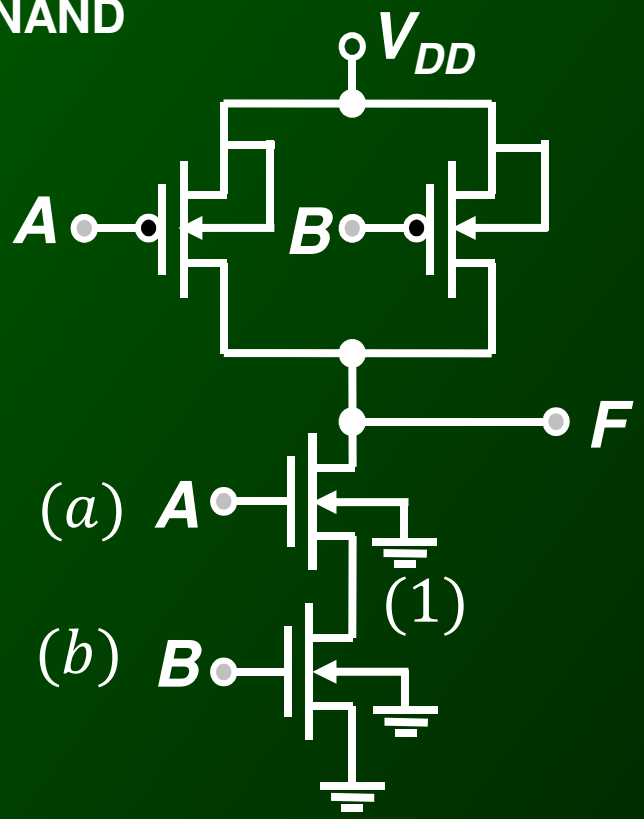
where γ is the *body effect parameter*

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

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

NAND



	A	B	F
	0	0	1
	0	1	1
(*)	1	0	1
	1	1	0

$V_A = V_{DD}$

$V_B = 0$

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

$V_{T_a} > V_{T_b}$

λ: Channel Length Modulation Parameter

1. I
2.
3.
4.
5.

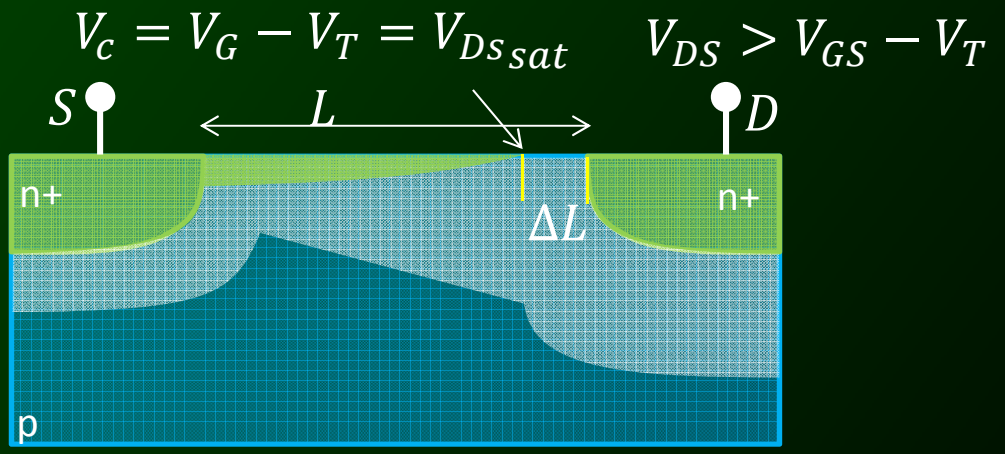
$$I_{D_{sat}} \propto \frac{1}{L-\Delta L} = \frac{1}{L}\left(1+\frac{\Delta L}{L}\right)$$

$$\Delta L \propto V_{DS} - V_{DS_{sat}}$$

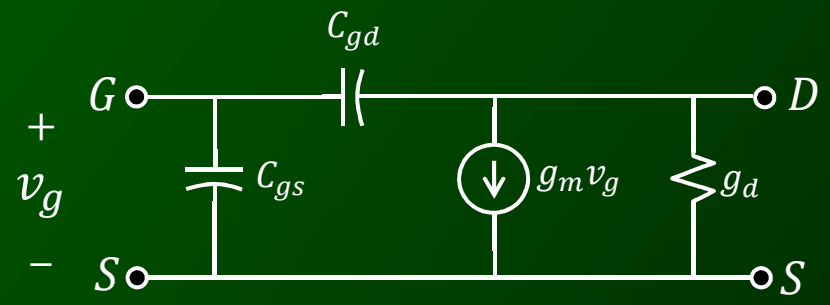
$$\frac{\Delta L}{L} \propto \lambda\left(V_{DS} - V_{DS_{sat}}\right)$$

$$\lambda \sim \frac{1}{L}$$

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



MOSFET: Small Signal Model



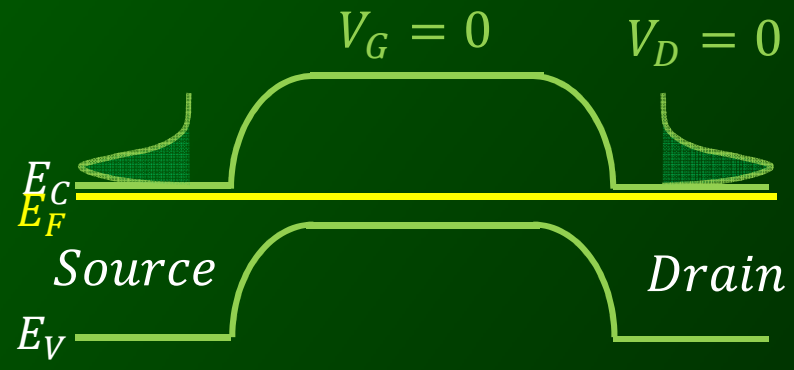
$$g_m = \frac{W}{2mL}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)$$

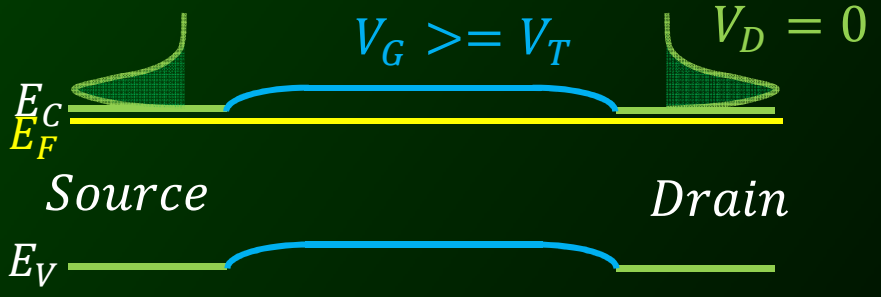
$$g_d = \lambda I_{D\,sat}$$

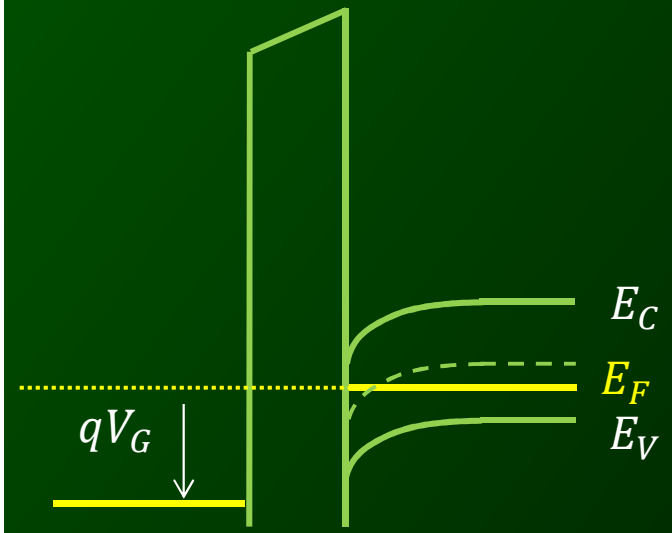
cut-off frequency:

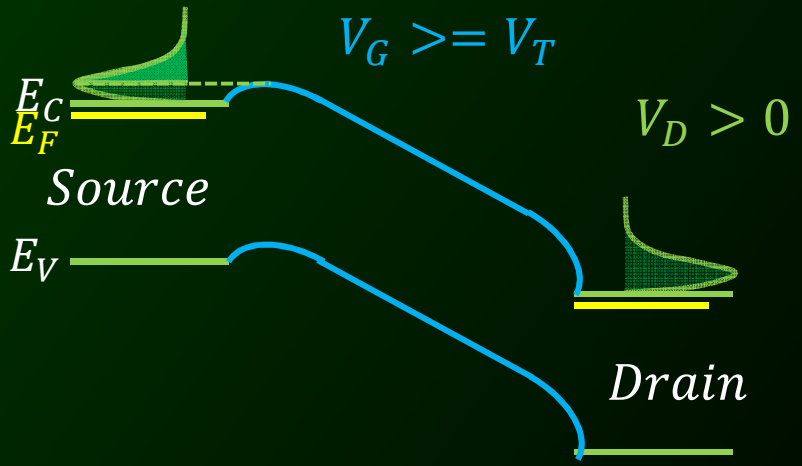
$$f_{max}\ \nearrow \quad \rightarrow \quad \frac{g_m}{2\pi C_{Ox}} \quad \propto \quad \frac{1}{L}\ \searrow$$

Sub-Threshold Current

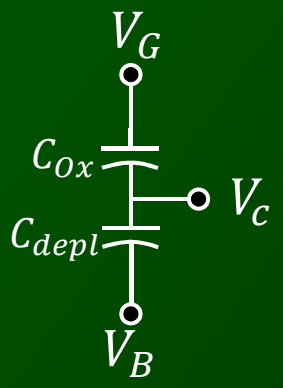








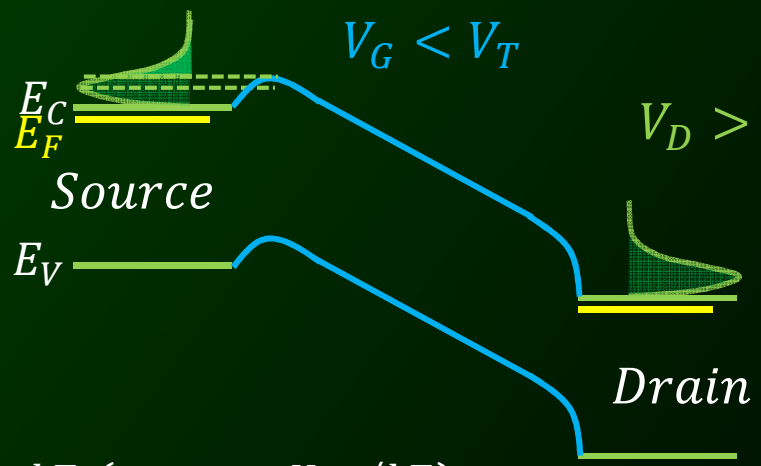
Sub-Threshold Current



Similarly:

$$m = 1 + \frac{C_{depl}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

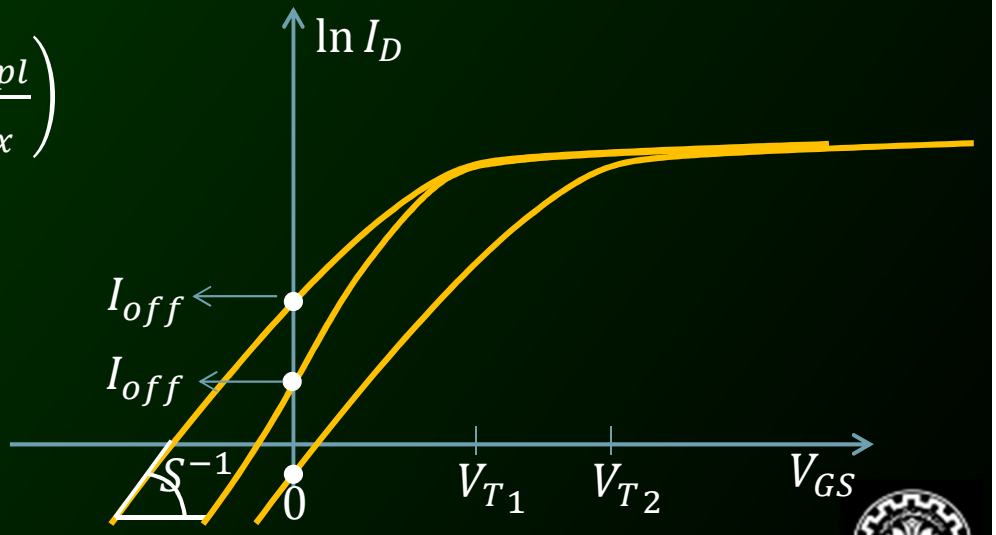
$$V_c = \frac{C_{Ox}}{C_{Ox} + C_{depl}} V_{GB}$$



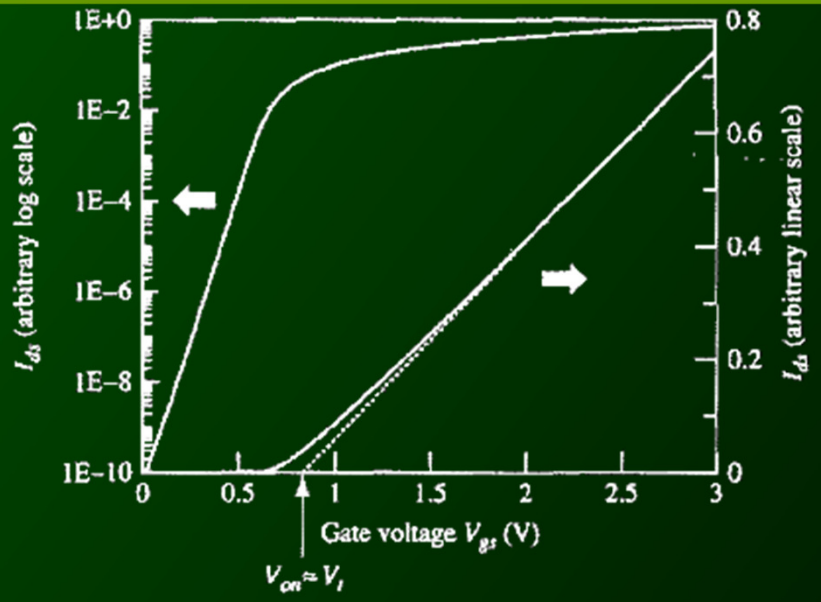
$$I_{DS} = \frac{W}{L}\,\mu_{eff} C_{Ox}(m-1)\left(\frac{kT}{q}\right)^2 e^{q(V_G - V_T)/mkT}\left(1 - e^{-qV_{DS}/kT}\right)$$

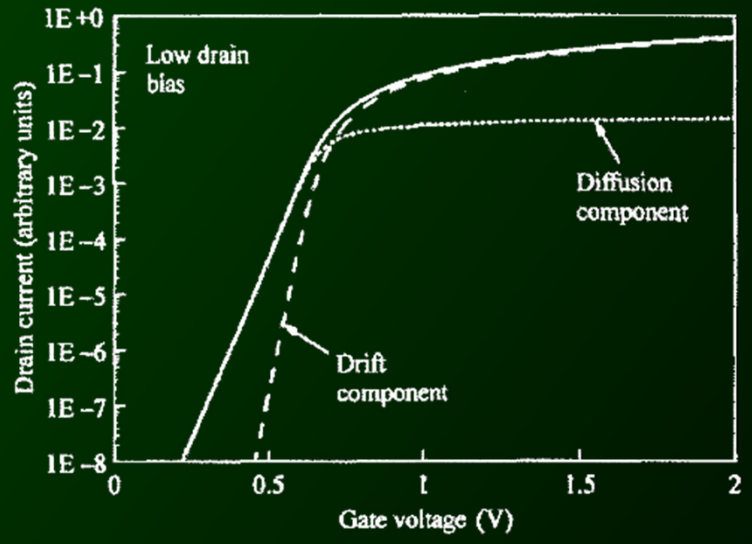
$$S = \left(\frac{d \log_{10} I_{DS}}{dV_{GS}}\right)^{-1} = \underbrace{\frac{kT}{q} \ln 10}_{60\ mV}\left(1 + \frac{C_{depl}}{C_{Ox}}\right)$$

$$m \searrow \begin{cases} N_A \searrow \Longrightarrow\ C_{depl} \searrow \Longrightarrow \text{retrograde} \\ t_{Ox} \searrow \Longrightarrow C_{Ox} \nearrow \\ T \searrow\ \ (\text{low-temperature}) \end{cases}$$

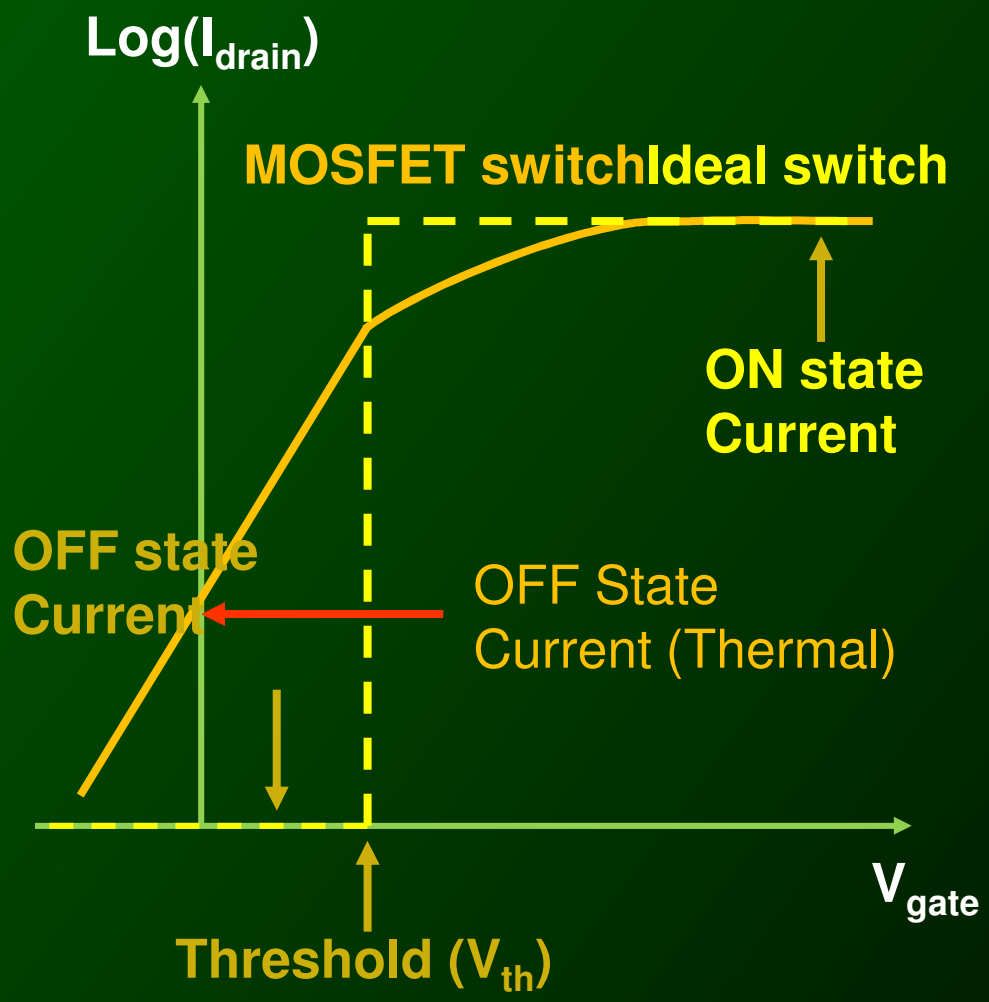


Sub-Threshold Current





Main MOSFET Parameters



3 main parameters

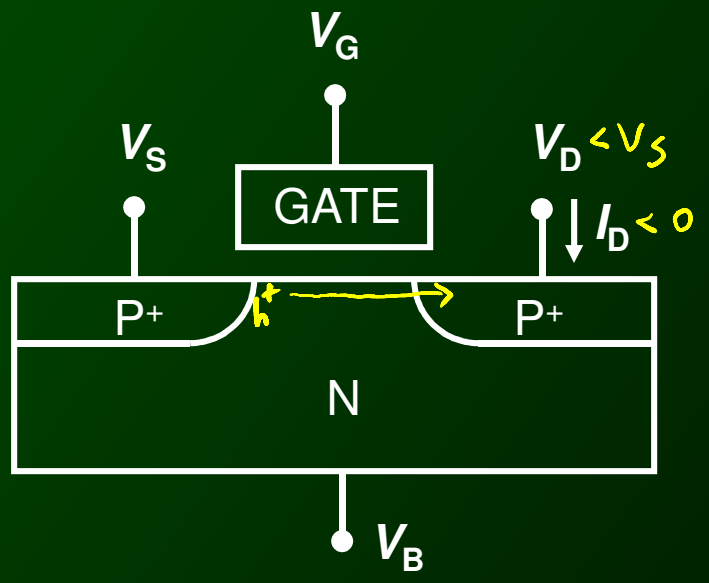
1. Threshold Voltage
2. Ion (=speed)
3. Ioff (=stand-by power)

P-Channel MOSFET

1. I
2.
3.
4.
5.

- The PMOSFET turns on when $V_{GS} < V_T$
  - Holes flow from SOURCE to DRAIN

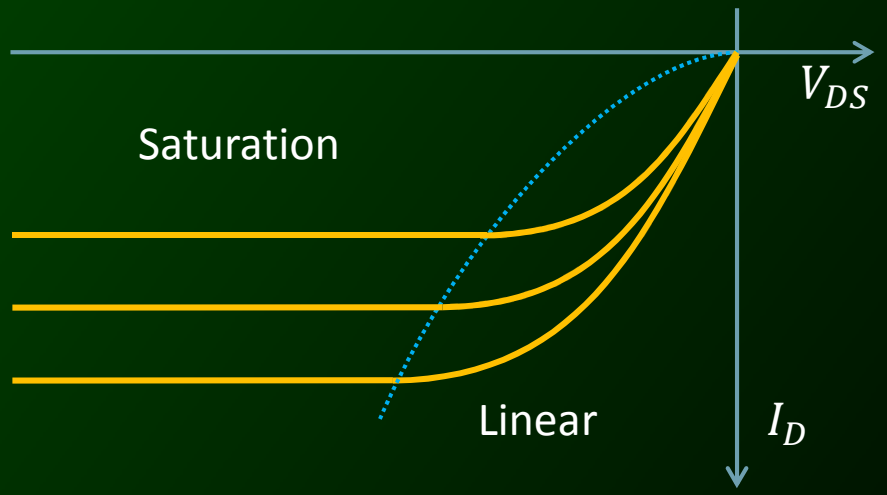
    $\Rightarrow$ DRAIN is biased at a ***lower*** potential than the SOURCE



- $V_{DS} < 0$
- $I_{DS} < 0$
- $|I_{DS}|$ increases with
  - $|V_{GS} - V_T|$
  - $|V_{DS}|$ (linear region)

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

PMOSFET I-V



Linear $\quad 0 < |V_{DS}| < \dfrac{|V_G - V_T|}{m}$

$$I_{DS} = -\frac{W}{L}\,\mu_{eff} C_{Ox}\left(V_{GS} - V_{Tp} - \frac{m}{2}V_{DS}\right)V_{DS}$$

bulk-charge factor

$$m = 1 + \frac{C_{depl}}{C_{Ox}} = 1 + \frac{3t_{Ox}}{W_T}$$

Saturation $\quad |V_{DS}| > \dfrac{|V_G - V_T|}{m}$

$$I_{DS} = I_{D\,sat} = -\frac{W}{2mL}\,\mu_{eff} C_{Ox}(V_{GS} - V_T)^2$$

MOSFET Scaling

MOSFETs have been steadily miniaturized over time

1970s: ~ 10 mm

Today: ~30 nm

Reasons:

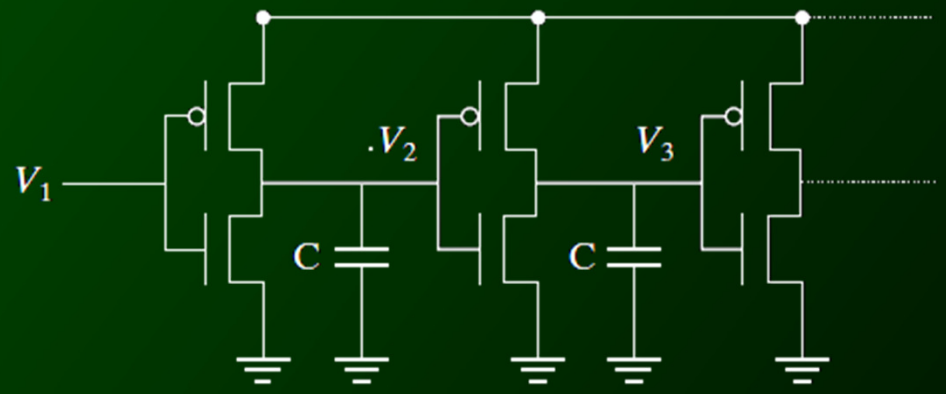
Improved circuit operating speed

Increased device density --> lower cost per function

As MOSFET lateral dimensions (*e.g.* channel length *L*) are reduced:

- I_{Dsat} increases → decreased effective "*R*"
- gate and junction areas decrease → decreased load "*C*"

→ faster charging/discharging (*i.e.* t_d is decreased)



Intrinsic Delay

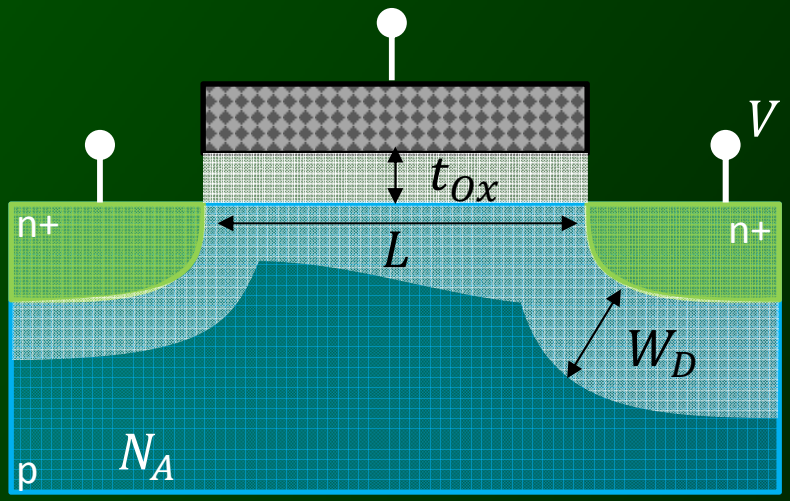
$$\tau = \frac{C_{Ox}\, V_{DD}}{I_{ON}}$$

MOSFET Scaling: Constant-Field Approach

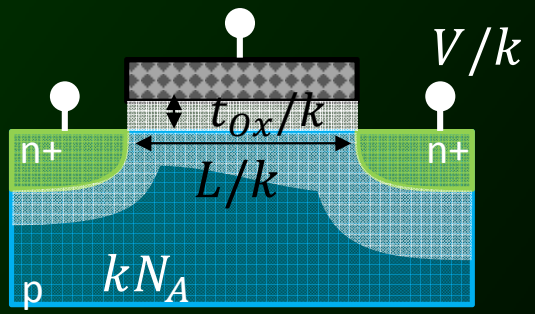
1. I
2.
3.
4.
5.

MOSFET dimensions and the operating voltage (V_{DD}) each are scaled by the same factor $k > 1$, so that the electric field remains unchanged.

Original device



Scaled device



Constant-Field Scaling Benefits

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

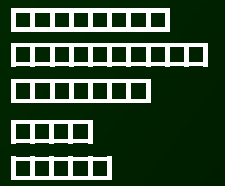
Failure of Constant-Field Scaling

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:

Feature Size (μm)	Power-Supply Voltage (V)	Gate Oxide Thickness (Å)	Oxide Field (MV/cm)
2	5	350	1.1
1.2	5	250	2.0
0.8	5	180	2.8
0.5	3.3	120	2.8
0.35	3.3	100	3.3
0.25	2.5	70	3.6

MOSFET Scaling: Generalized Approach

1. I
2.
3.
4.
5.



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

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

N_{body} must be scaled up by α to suppress short-channel effects

Velocity Saturation

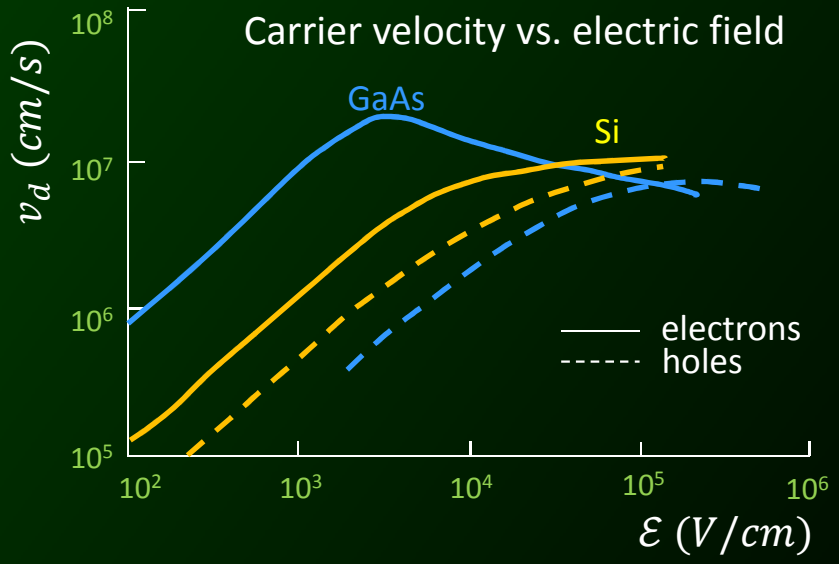
1. I
2.
3.
4.
5.

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

Simple model:

$$v = \begin{cases} \dfrac{\mu\mathcal{E}}{1 + \dfrac{\mathcal{E}}{\mathcal{E}_{sat}}} & for\ \ \mathcal{E} < \mathcal{E}_{sat} \\ v_{sat} & for\ \ \mathcal{E} \geq \mathcal{E}_{sat} \end{cases}$$

$$\mu\,\mathcal{E}_{sat} = 2v_{sat}$$

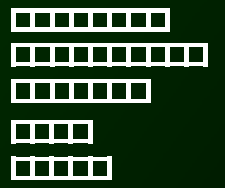


$$v_{sat} = \begin{cases} 8 \times 10^6\ cm/s\ \text{ for } e^- \text{in Si} \\ 6 \times 10^6\ cm/s\ \text{ for } h^+ \text{in Si} \end{cases}$$

If $\mathcal{E} < \mathcal{E}_{sat}$: $\quad \mu \mapsto \dfrac{\mu}{1 + \dfrac{\mathcal{E}}{\mathcal{E}_{sat}}}$

MOSFET I-V with Velocity Saturation

1. I
2.
3.
4.
5.



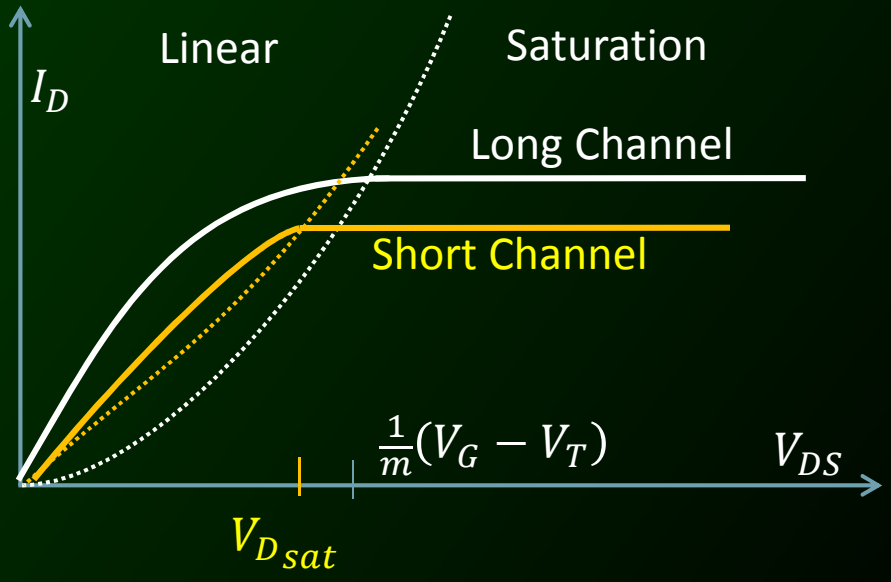
In linear region:

$$I_D = \frac{W}{L}\frac{\mu_{eff}C_{Ox}}{1+\frac{V_{DS}}{L\mathcal{E}_{sat}}}\left(V_{GS}-V_T-\frac{m}{2}V_{DS}\right)V_{DS} = \frac{I_{D\,LongChannel}}{1+\frac{V_{DS}}{L\mathcal{E}_{sat}}}$$

$$\mu \mapsto \frac{\mu}{1+\frac{\mathcal{E}}{\mathcal{E}_{sat}}}$$

MOSFET is Long channel if $L\mathcal{E}_{sat} \gg V_{GS}-V_T$

$$\frac{1}{V_{D\,sat}} = \frac{m}{V_{GS}-V_T} + \frac{1}{L\mathcal{E}_{sat}}$$



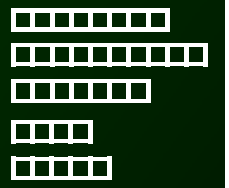
$V_{GS} = 1.8\,V, t_{Ox} = 3\,nm,$
$V_T = 0.25\,V, W = 45\,nm$

μ_{eff} =200 cm^{2}V^{-1}s^{-1} ,$m = 1 + 3T_{oxe}/W_T = 1.2$
$\mathcal{E}_{sat} = 2v_{sat} / \mu_{eff} = 8 \times 10^4$ V/cm

$L = 10\mu m \rightarrow V_{D\,sat} = 1.3V$
$L = 1\mu m \rightarrow V_{D\,sat} = 1.1V$
$L = 100nm \rightarrow V_{D\,sat} = 0.5V$
$L = 30nm \rightarrow V_{D\,sat} = 0.2V$

I_{Dsat} with Velocity Saturation

1. I
2.
3.
4.
5.



In saturation region:

$V_{DS} \mapsto V_{GS} - V_T$

$$I_{D_{sat}} = \frac{W}{2mL} \frac{\mu_{eff} C_{Ox}}{1 + \frac{V_{GS} - V_T}{L\mathcal{E}_{sat}}} (V_{GS} - V_T)^2 = \frac{I_{Dsat_{LongChannel}}}{1 + \frac{V_{GS} - V_T}{L\mathcal{E}_{sat}}}$$

Very short channel length: $L\mathcal{E}_{sat} \ll V_{GS} - V_T$

$$I_{D_{sat}} = \frac{W}{2m} \mu_{eff} \mathcal{E}_{sat} C_{Ox} (V_{GS} - V_T) = \frac{W}{m} v_{sat} C_{Ox} (V_{GS} - V_T)$$

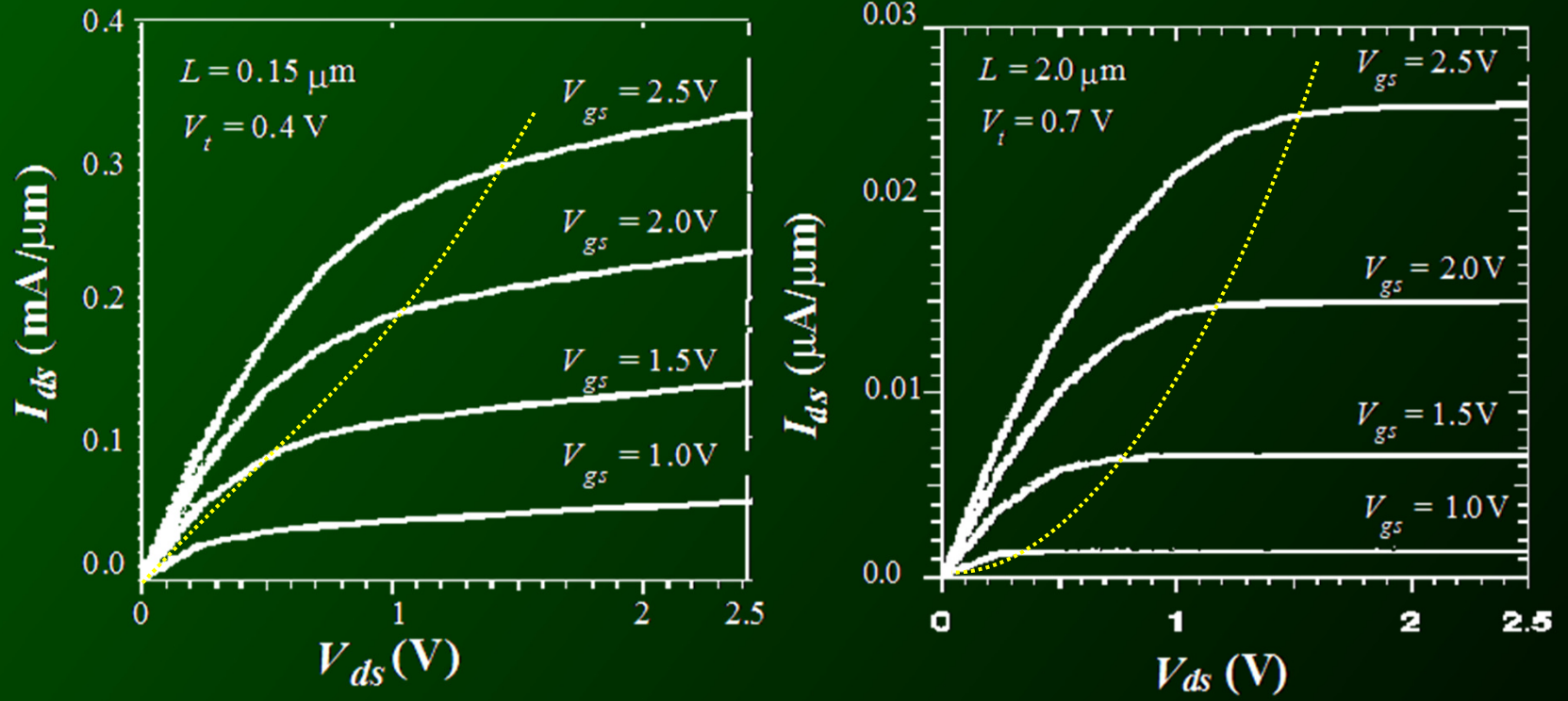
- $I_{D_{sat}}$ is proportional to $V_{GS} - V_T$ rather than $(V_{GS} - V_T)^2$
- $I_{D_{sat}}$ is not dependent on L ☹

To improve modern MOSFETs: $I_{ON} \nearrow$

$C_{Ox} \nearrow$ high-k dielectric

$v_{sat} \nearrow$ strained Si

Short- vs. Long-Channel NMOSFET

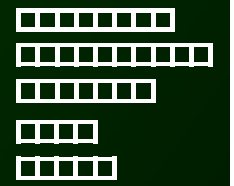


Short-channel NMOSFET:

- I_{Dsat} is proportional to V_{GS}-V_{Tn} rather than $(V_{GS}$-$V_{Tn})^2$
- V_{Dsat} is lower than for long-channel MOSFET
- Channel-length modulation is apparent

Velocity Overshoot

1. I
2.
3.
4.
5.



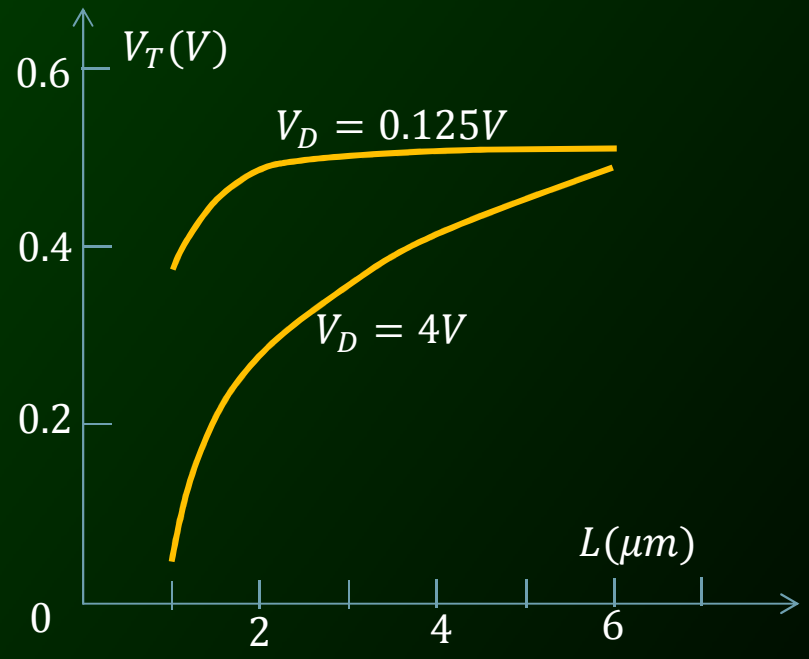
When L is comparable to or less than the mean free path, some of the electrons travel through the channel without experiencing a single scattering event

→ 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 $|V_{DS}|$



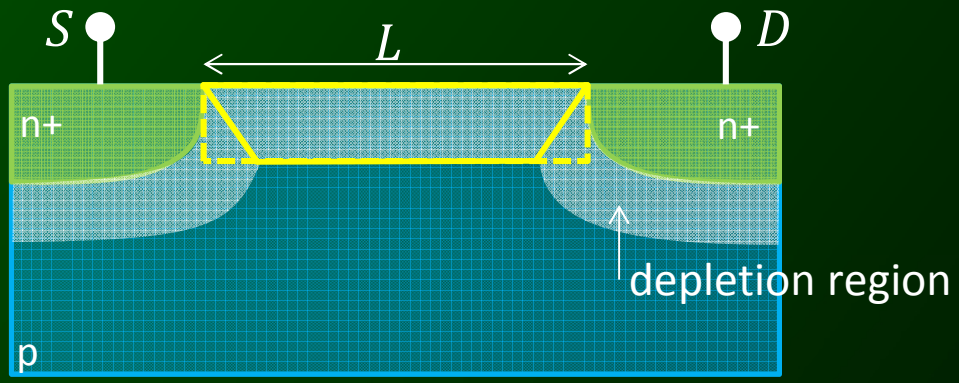
This effect is undesirable (i.e. we want to minimize it!) because circuit designers would like VT to be invariant with transistor dimensions and bias condition

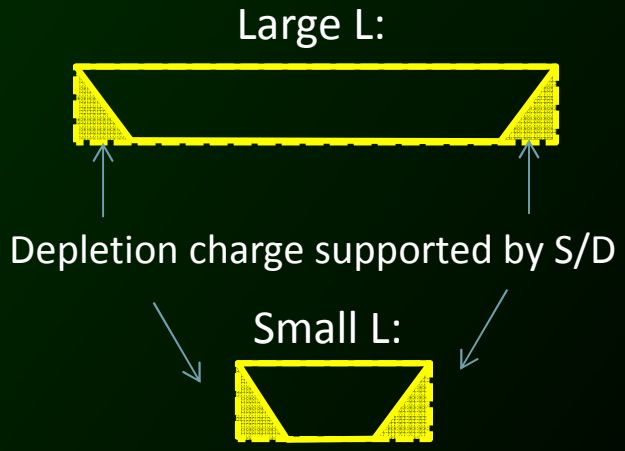
Qualitative Explanation of SCE

Before an inversion layer forms beneath the gate, the surface of the Si underneath the gate must be depleted (to a depth W_T)

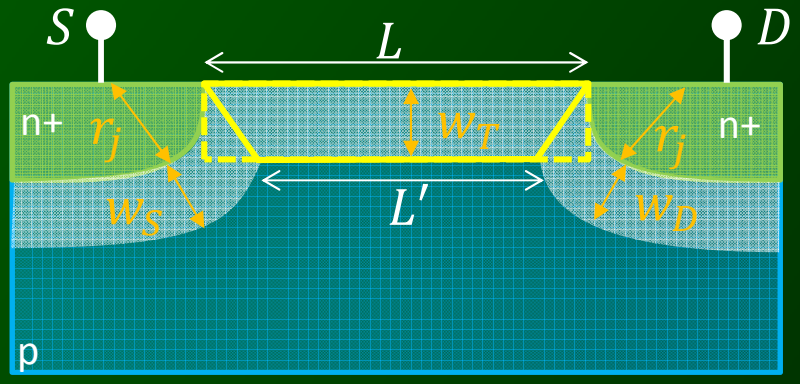
The source & drain pn junctions assist in depleting the Si underneath the gate. Portions of the depletion charge in the channel region are balanced by charge in S/D regions, rather than by charge on the gate. Less gate charge is required to invert the semiconductor surface (i.e. $|V_T|$ decreases)

$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$





V_T Roll-Off: First-Order Model



$$V_T = V_{FB} + 2\varphi_F + \frac{Q_{depl}}{C_{Ox}}$$

$$\Delta V_T \equiv |V_T| - \left|V_{T\,Long\ Channel}\right|$$

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

$$L' = L - 2r_j\left[\sqrt{1 + \frac{2W_T}{r_j}} - 1\right]$$

$$\rightarrow\ \Delta V_T = \frac{-qN_A W_T}{C_{Ox}} \frac{r_j}{L}\left[\sqrt{1 + \frac{2W_T}{r_j}} - 1\right]$$

Minimize ΔV_T by

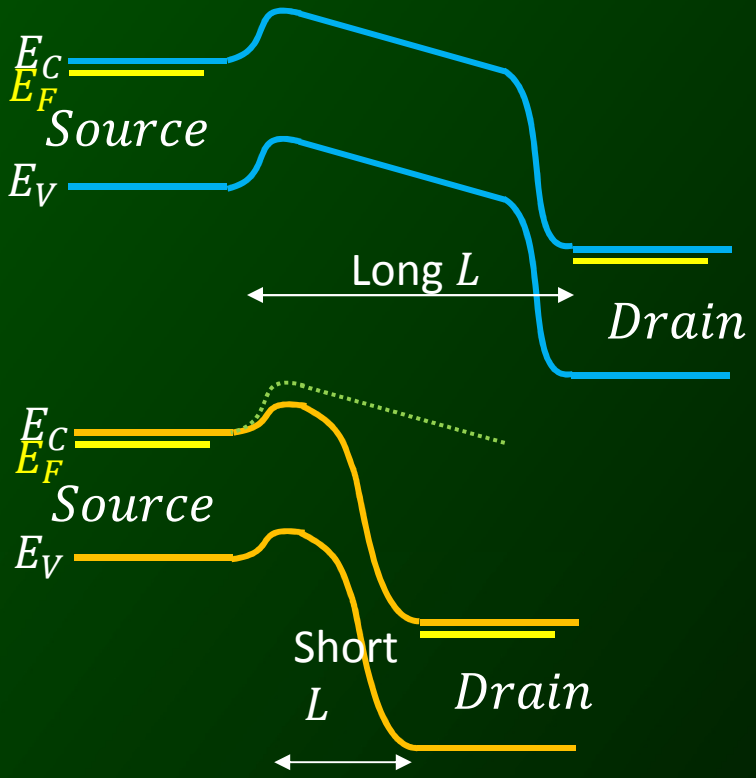
- reducing t_{Ox}
- reducing r_j
- increasing N_A (trade-offs: degraded μ_{eff}, m)

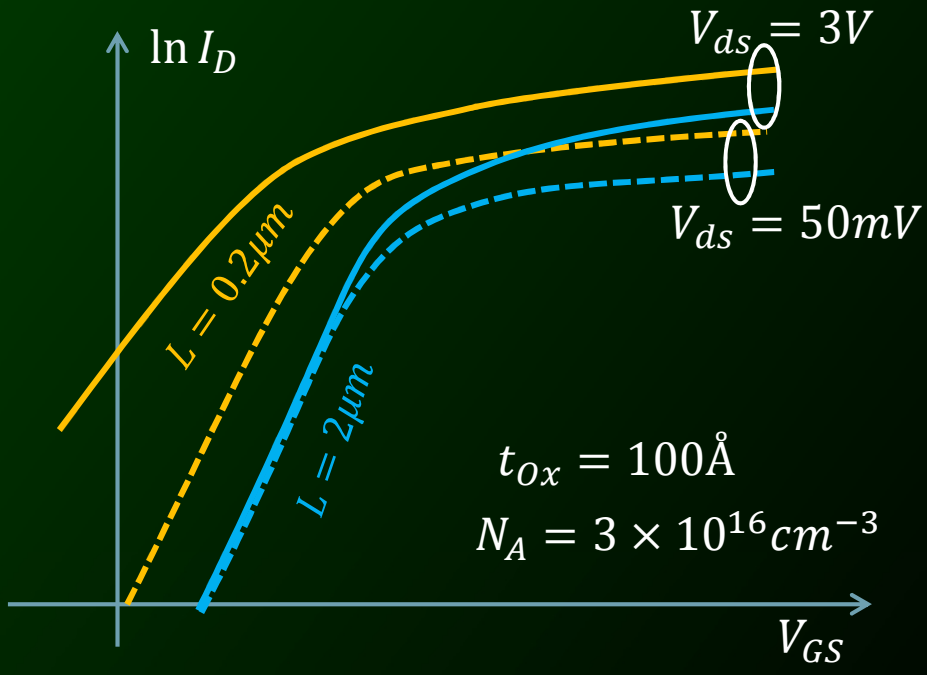
MOSFET vertical dimensions should be scaled along with horizontal dimensions!

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)



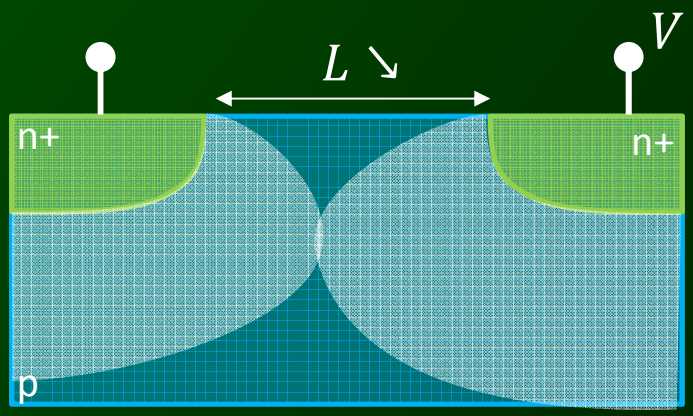


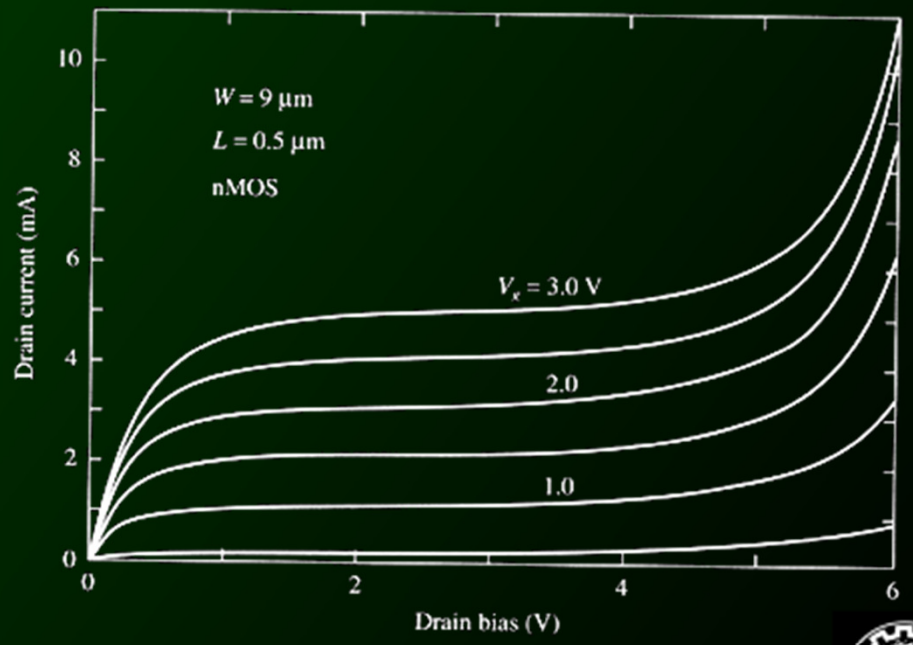
Punchthrough

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_{Dsat} increases rapidly with V_{DS}

This can be mitigated by doping the semiconductor more heavily in the sub-surface region, i.e. using a "retrograde" doping profile.





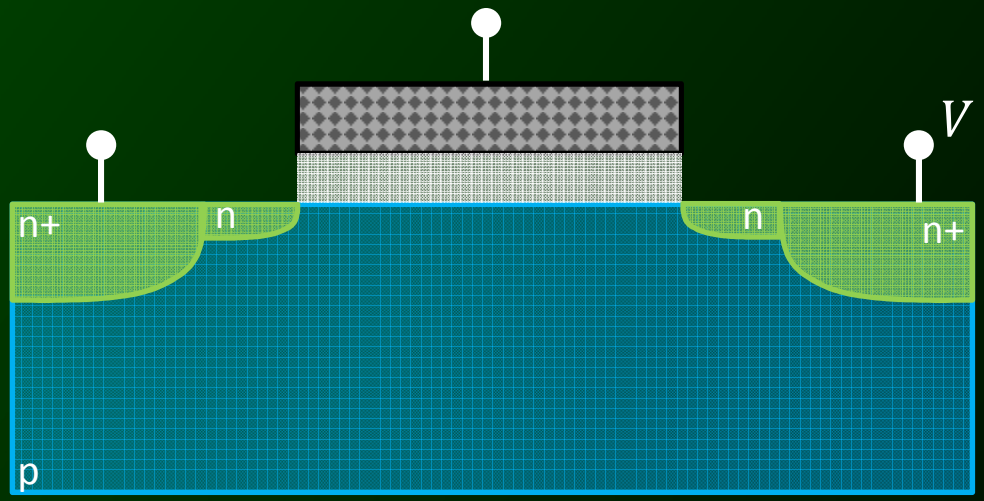
Source and Drain (S/D) Structure

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.

$$R_{source}, R_{drain} \propto \rho / W r_j$$

where ρ = 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



$\mathcal{E}$ - Field Distribution Along the Channel

The lateral electric field peaks at the drain end of the channel.

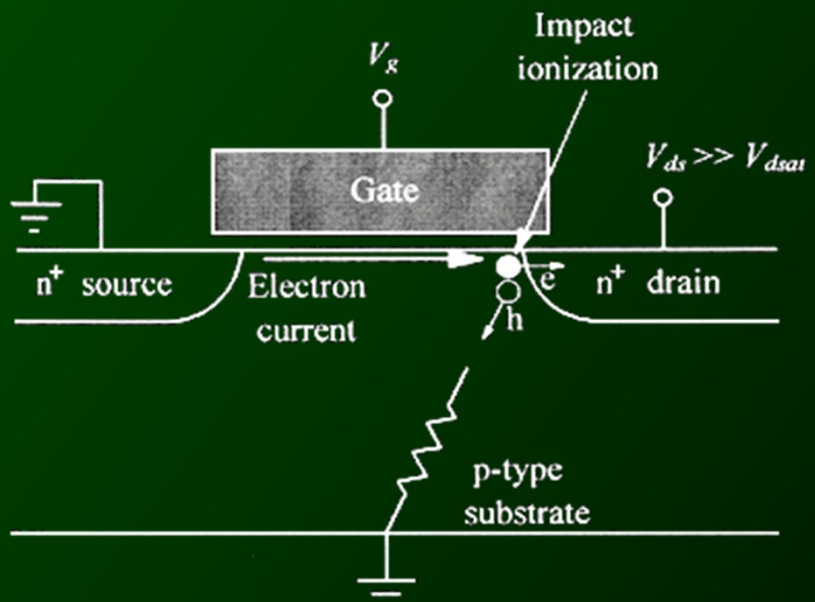
$\mathcal{E}_{peak}$ can be as high as 10^6 V/cm

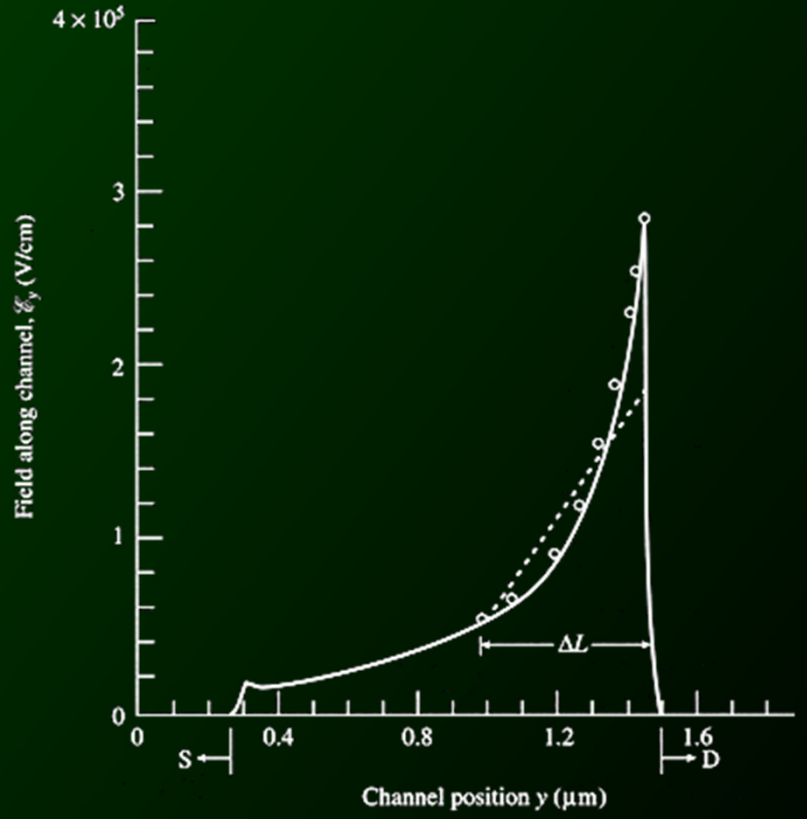
High E-field causes problems:

Damage to oxide interface & bulk

(trapped oxide charge → V_T shift)

substrate current due to impact ionization:



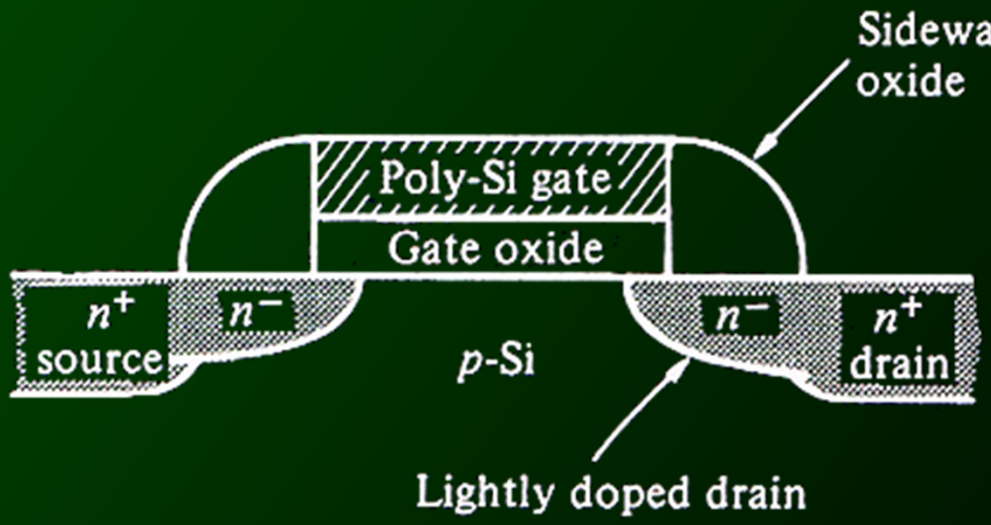


Lightly Doped Drain (LDD) Structure

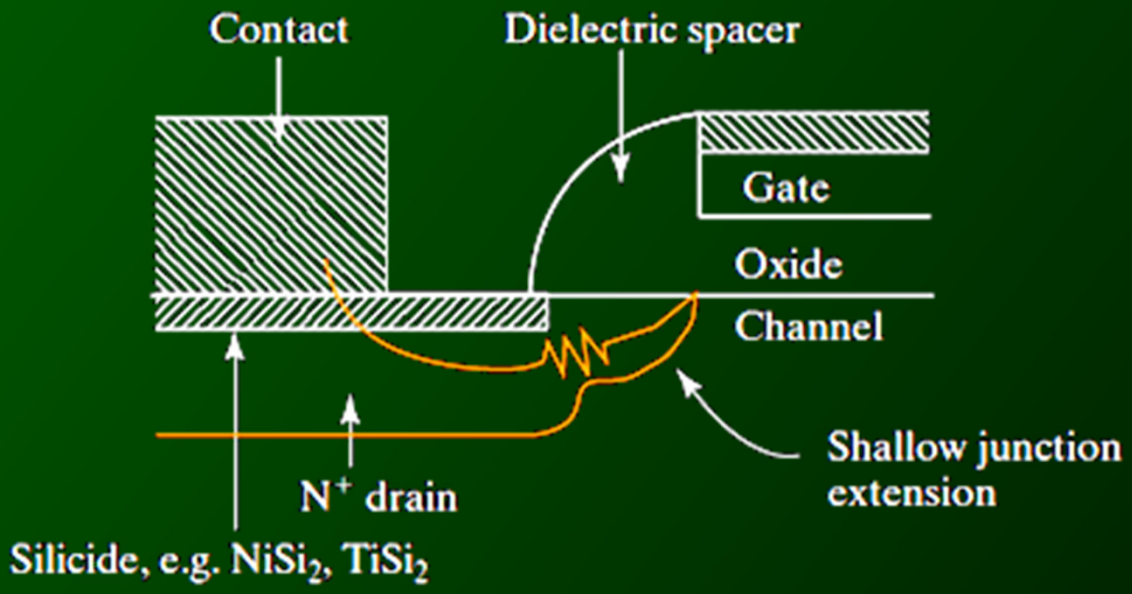
1. I
2.
3.
4.
5.

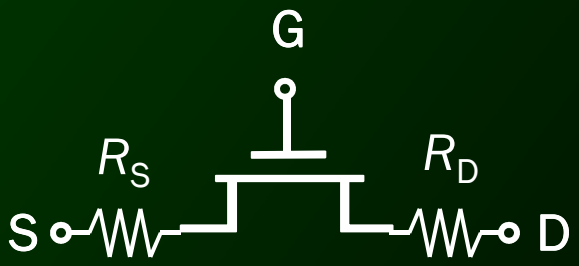
Lower pn junction doping results in lower peak E-field

- ✓ "Hot-carrier" effects are reduced
- × Parasitic resistance is increased



Parasitic Source-Drain Resistance





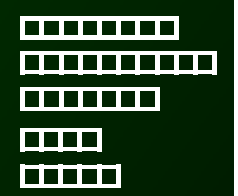
For short-channel MOSFET, $I_{Dsat0} \propto V_{GS} - V_T$, so that

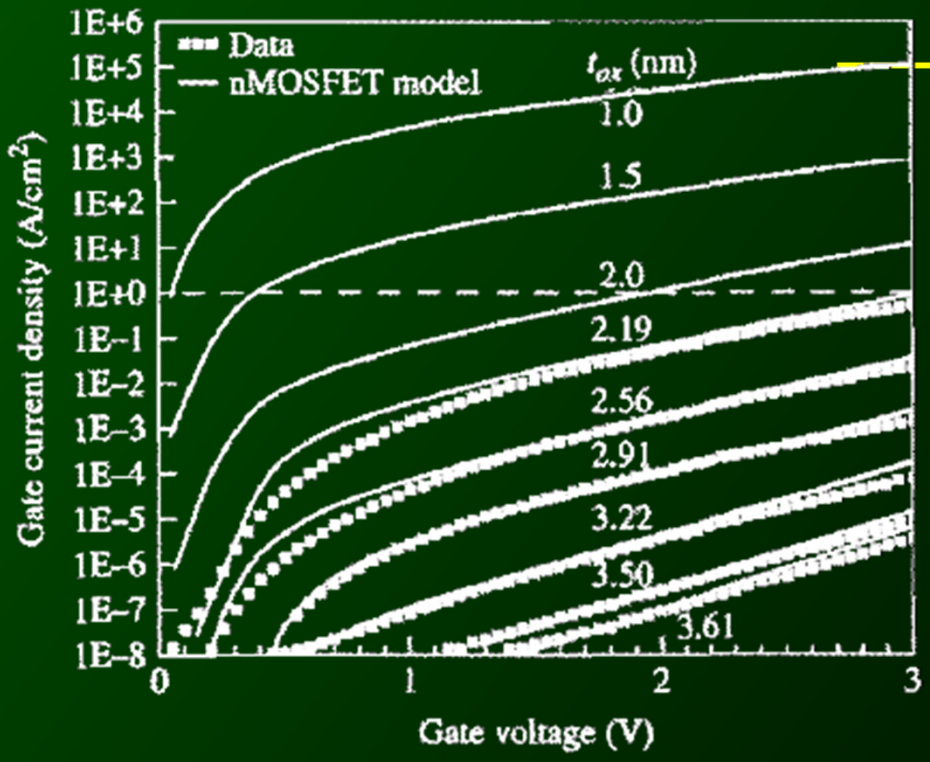
$$I_{Dsat} = \frac{I_{Dsat0}}{1 + \frac{I_{Dsat} R_S}{V_{GS} - V_T}}$$

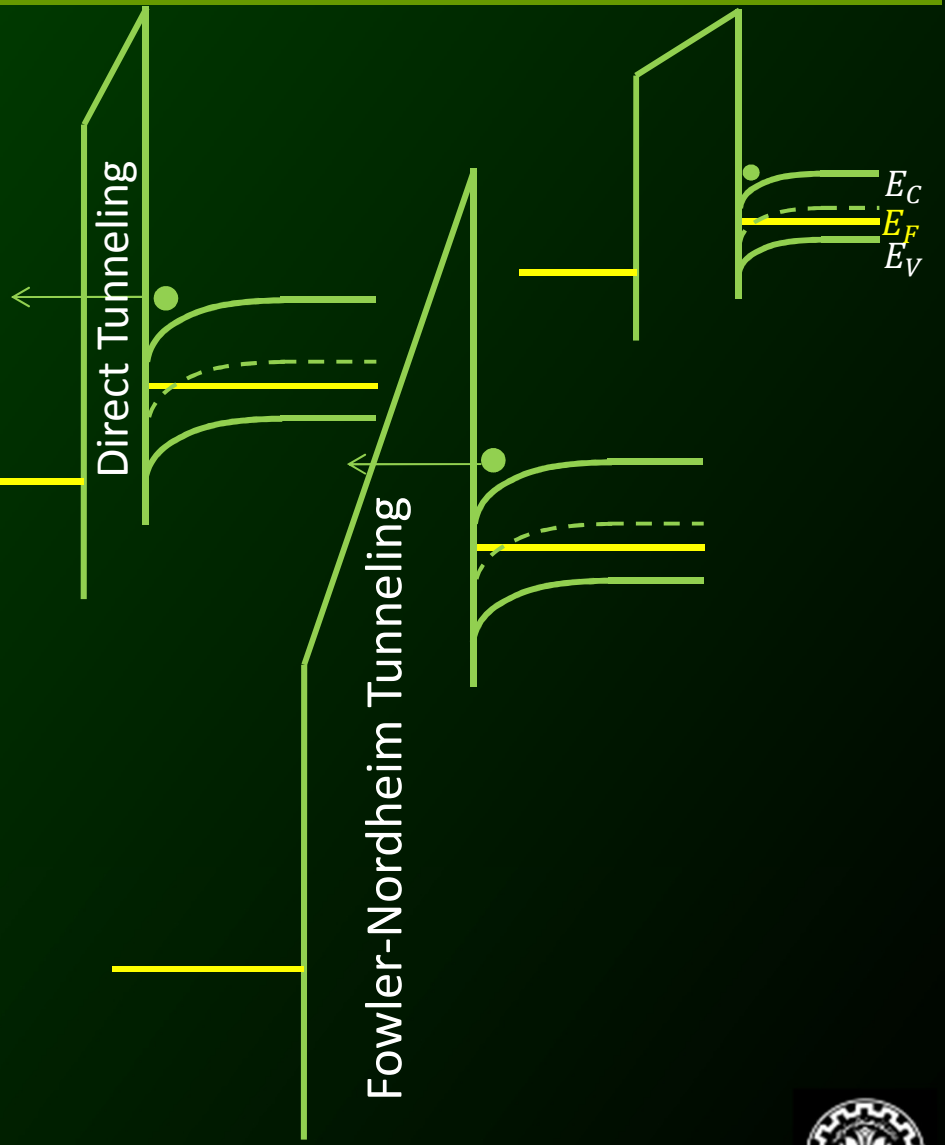
→ I_{Dsat} is reduced by ~15% in a 0.1 mm MOSFET.

V_{Dsat} is increased to $V_{Dsat0} + I_{Dsat}(R_S + R_D)$

Tunneling Into and Through SiO_2

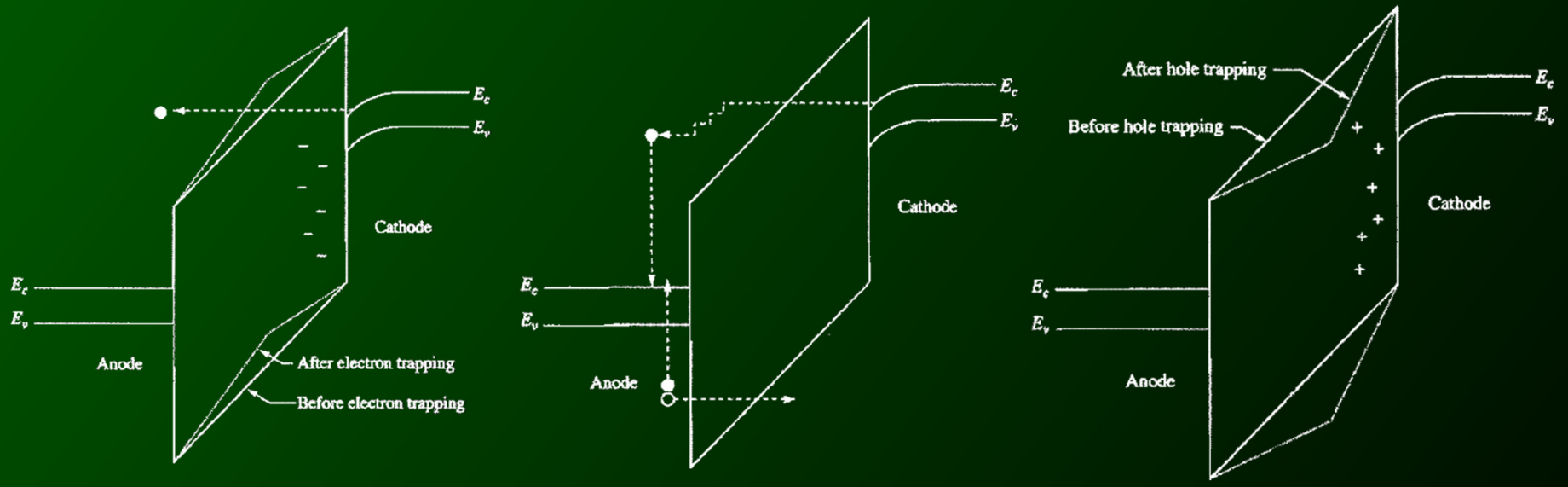






Defect Generation Caused By Tunneling Current

1. I
2.
3.
4.
5.



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.