## Metal Oxide Semiconductor Field Effect Transistors

# MOSFET

EBB424E

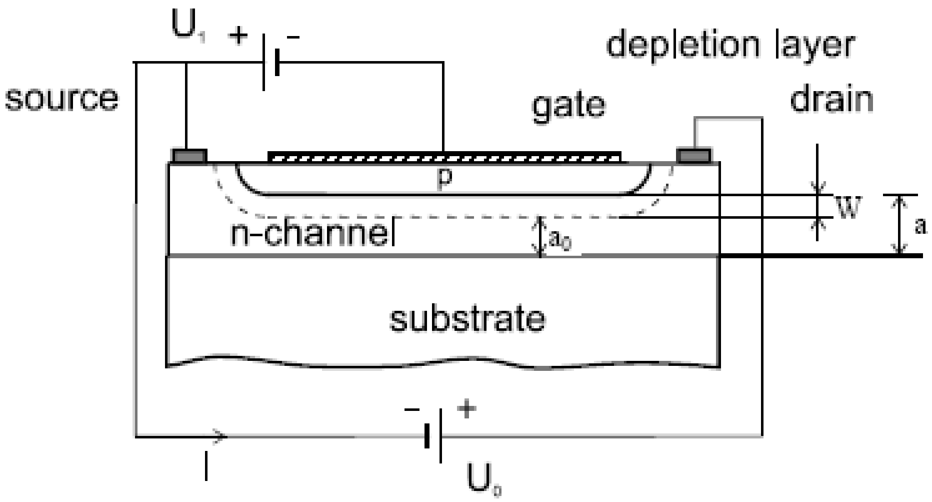
Dr. Sabar D. Hutagalung
School of Materials & Mineral Resources Engineering,
Universiti Sains Malaysia

# Different types of FETs

- Junction FET (JFET)
- Metal-Oxide-Semiconductor FET (MOSFET)
- Metal-Semiconductor FET (MESFET)

# Different types of FETs

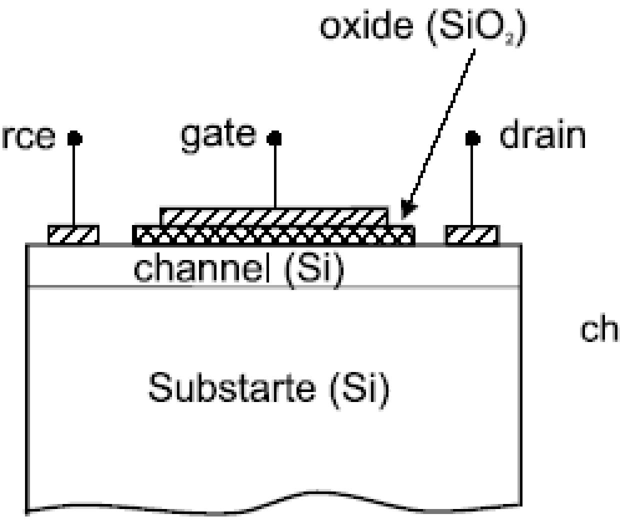
- Junction FET (JFET)

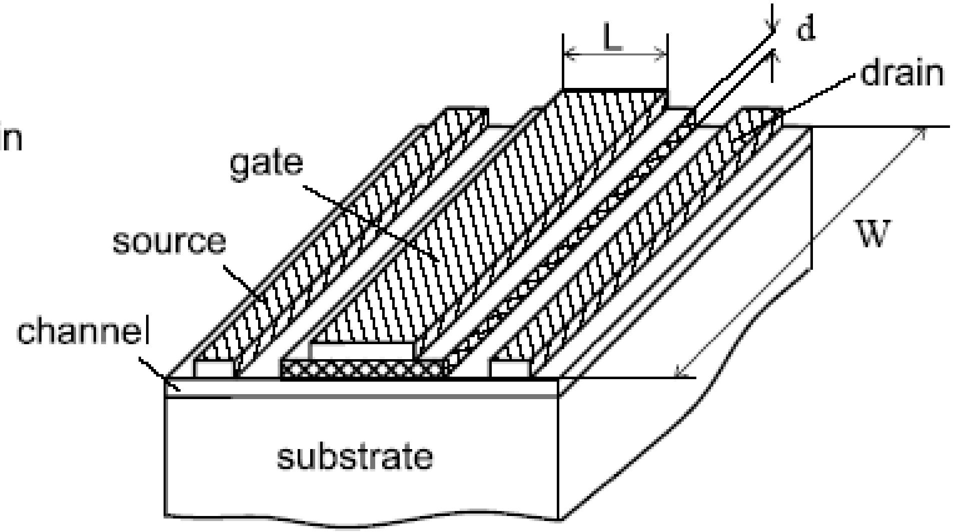


The gate-channel insulator consists of the DEPLETION REGION, i.e. the same material as the channel.

# Different types of FETs

- Metal-Oxide-Semiconductor FET (MOSFET)

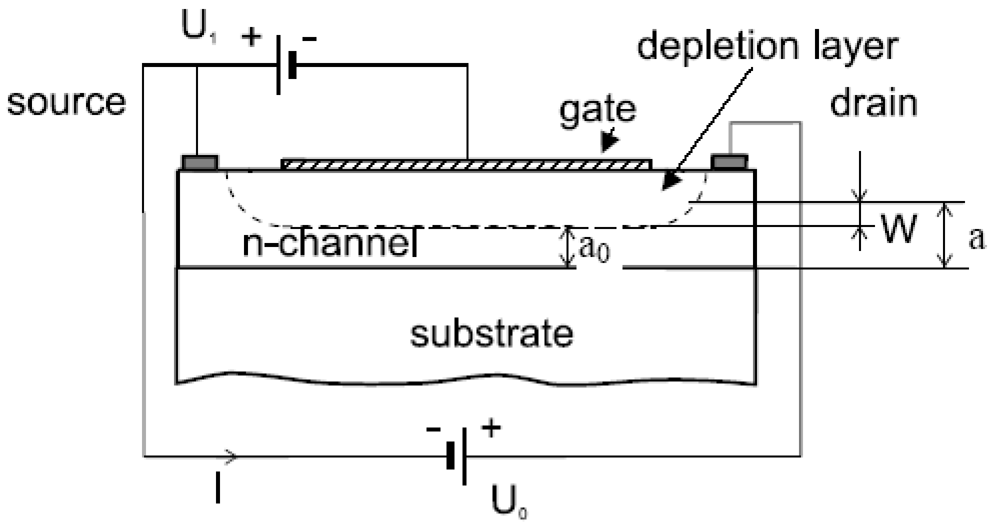




The gate-channel insulator is made out of dielectric (SiO$_2$), $\varepsilon = 3.9$

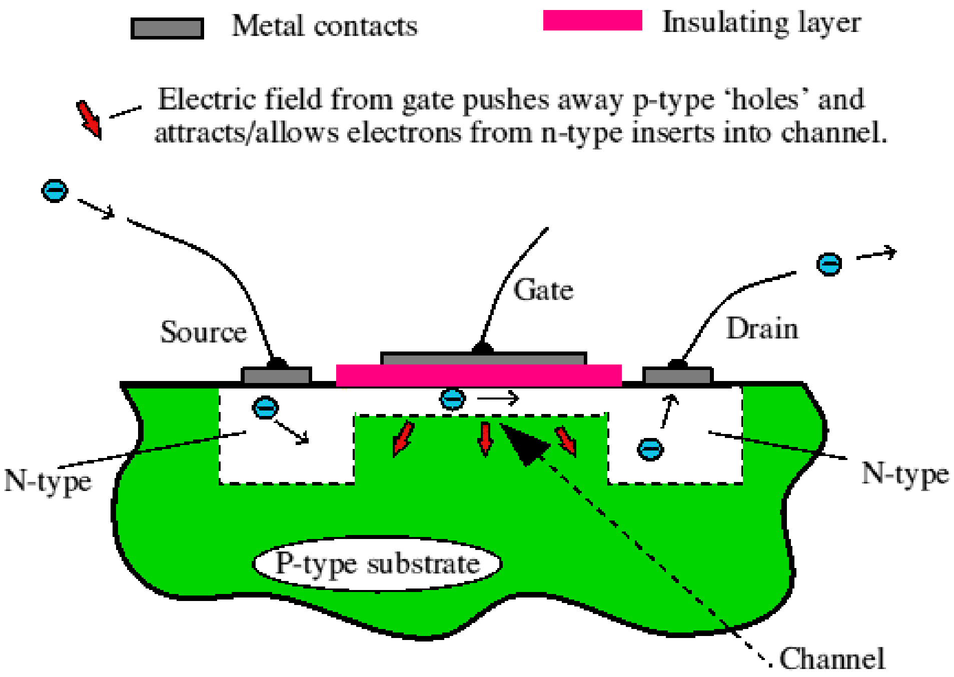
# Different types of FETs

- Metal-Semiconductor FET (MESFET)



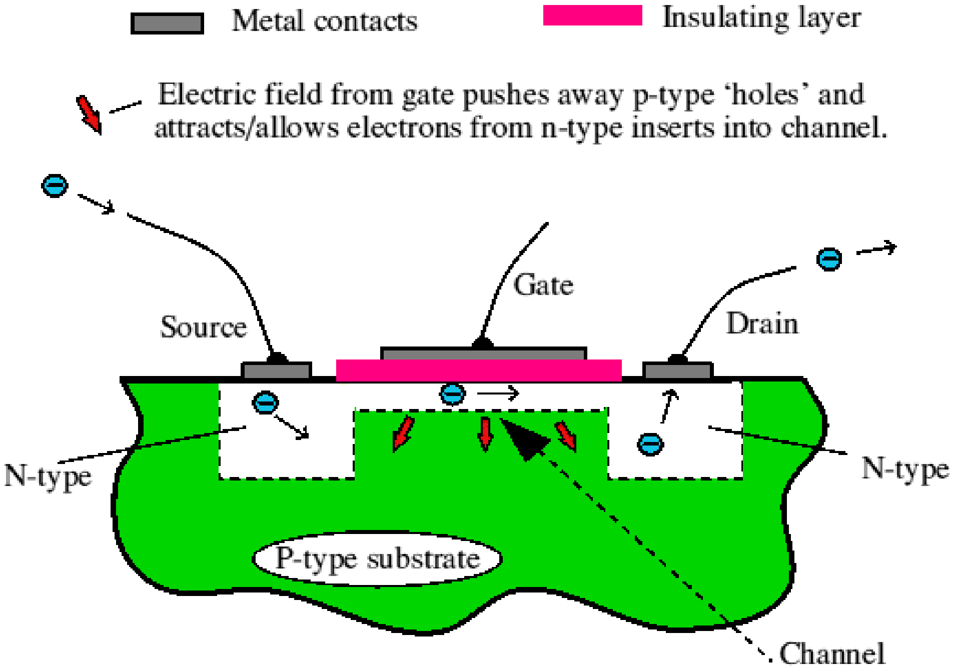
The gate is formed by Schottky barrier to the semiconductor layer. The gate-channel insulator consists of the DEPLETION REGION, i.e. the same material as the channel. Very similar to the JFET

# Basic MOSFET (n-channel)



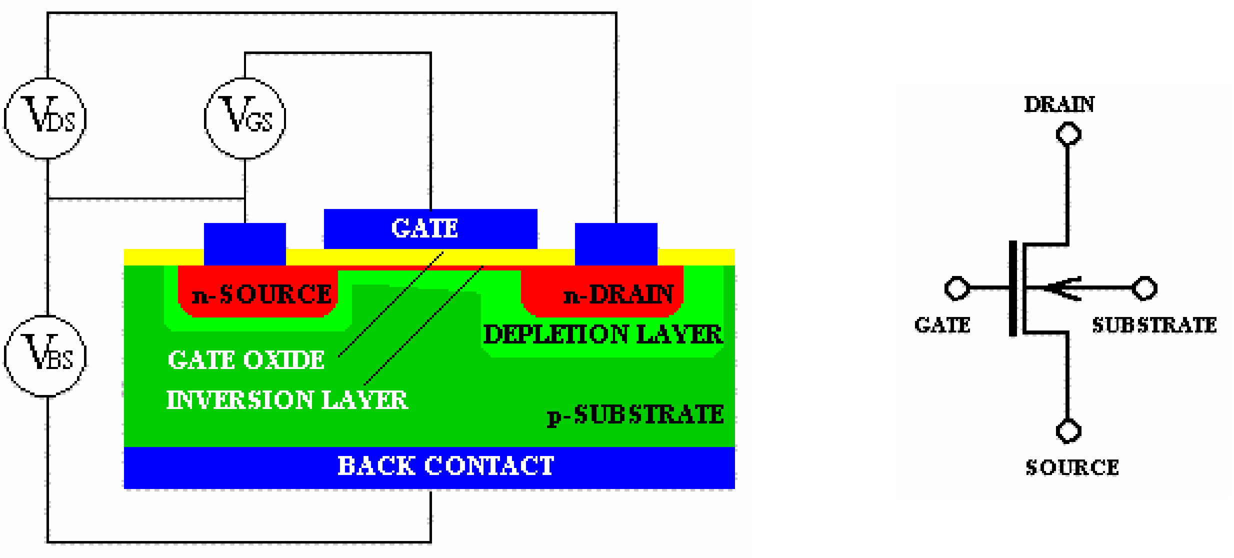
- The gate electrode is placed on top of a very thin insulating layer.
- There are a pair of small n-type regions just under the drain & source electrodes.
- If apply a +ve voltage to gate, will push away the 'holes' inside the p-type substrate and attracts the moveable electrons in the n-type regions under the source & drain electrodes.

# Basic MOSFET (n-channel)

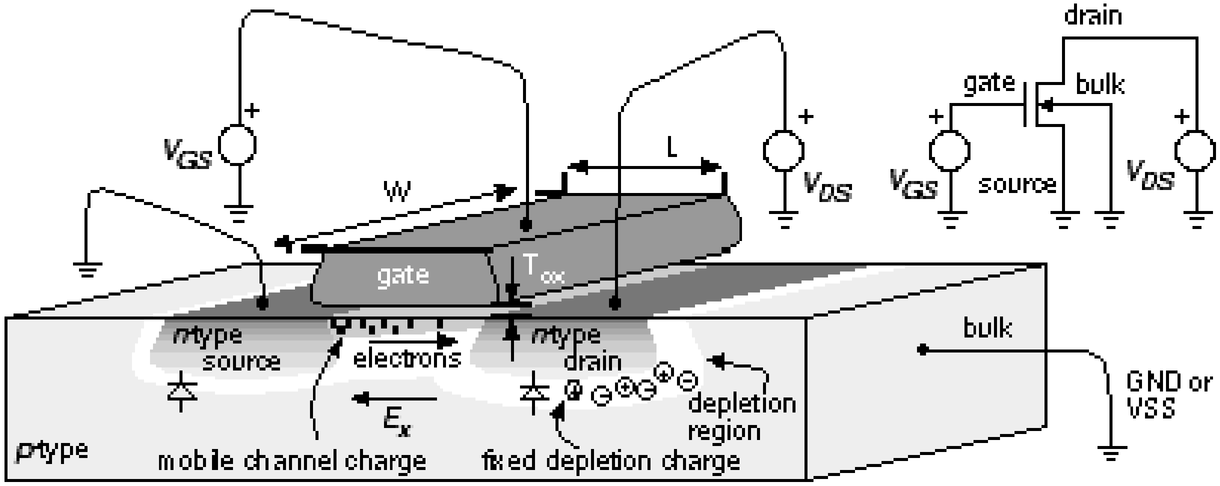


Electric field from gate pushes away p-type 'holes' and attracts/allows electrons from n-type inserts into channel.

- Increasing the +ve gate voltage pushes the p-type holes further away and enlarges the thickness of the created channel.
- As a result increases the amount of current which can go from source to drain — this is why this kind of transistor is called an *enhancement mode* device.



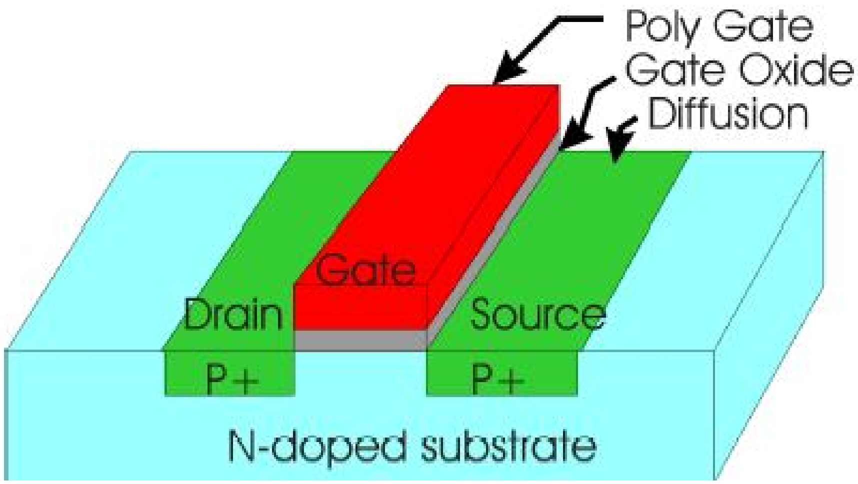
- Cross-section and circuit symbol of an n-type MOSFET.

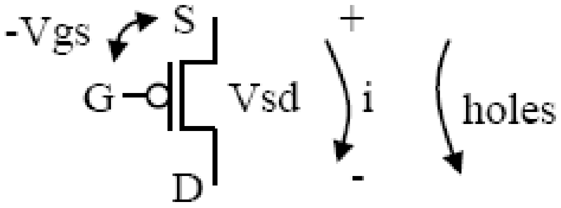


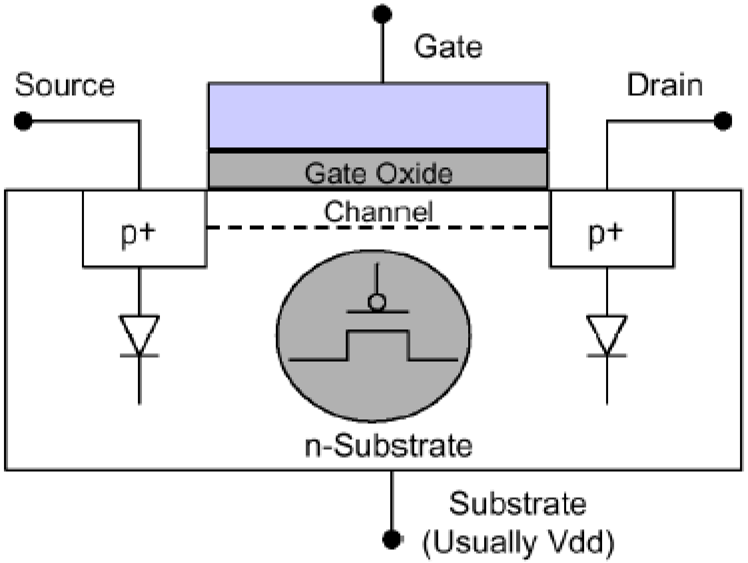
- An n-channel MOS transistor. The gate-oxide thickness, $T_{OX}$, is approximately 100 angstroms (0.01 $\mu$m). A typical transistor length, L = 2 $\lambda$. The bulk may be either the substrate or a well. The diodes represent pn-junctions that must be reverse-biased

# Basic MOSFET (p-channel)

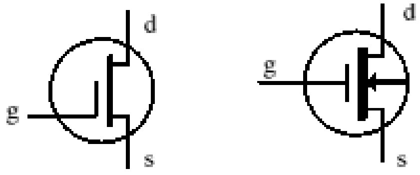
- These behave in a similar way, but they pass current when a -ve gate voltage creates an effective p-type channel layer under the insulator.
- By swapping around p-type for n-type we can make pairs of transistors whose behaviour is similar except that all the signs of the voltages and currents are reversed.
- Pairs of devices like this care called *complimentary* pairs.


Poly Gate
Gate Oxide
Diffusion
Gate
Drain
Source
P+
P+
N-doped substrate

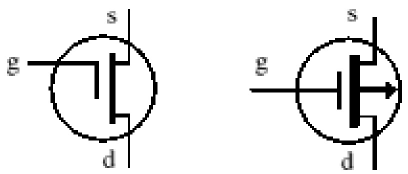

-Vgs
S
G
D
Vsd
+
i
-
holes


Gate
Source
Drain
Gate Oxide
Channel
p+
p+
n-Substrate
Substrate
(Usually Vdd)

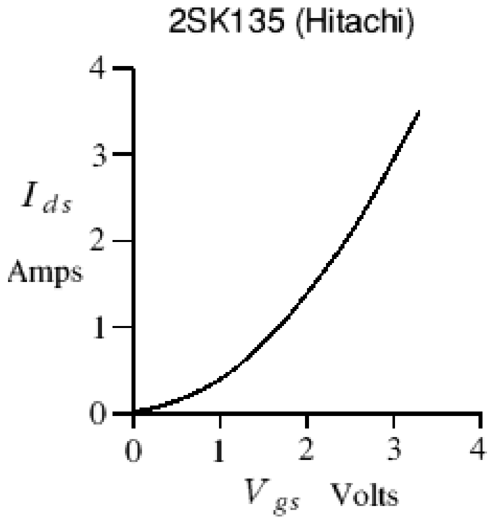
- In an n-channel MOSFET, the channel is made of n-type semiconductor, so the charges free to move along the channel are negatively charged *(electrons)*.
- In a p-channel device the free charges which move from end-to-end are positively charged (*holes*).

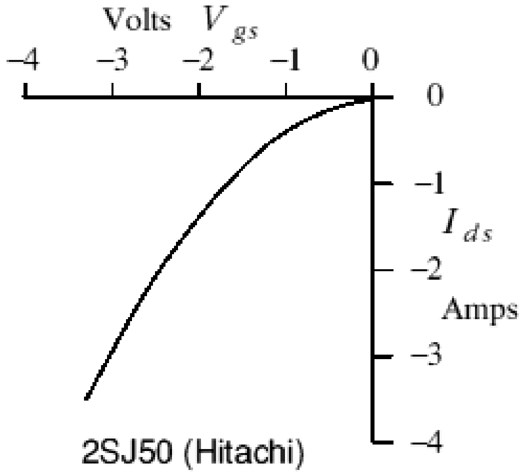


Symbols for N-channel MOSFET



Symbols for P-channel MOSFET



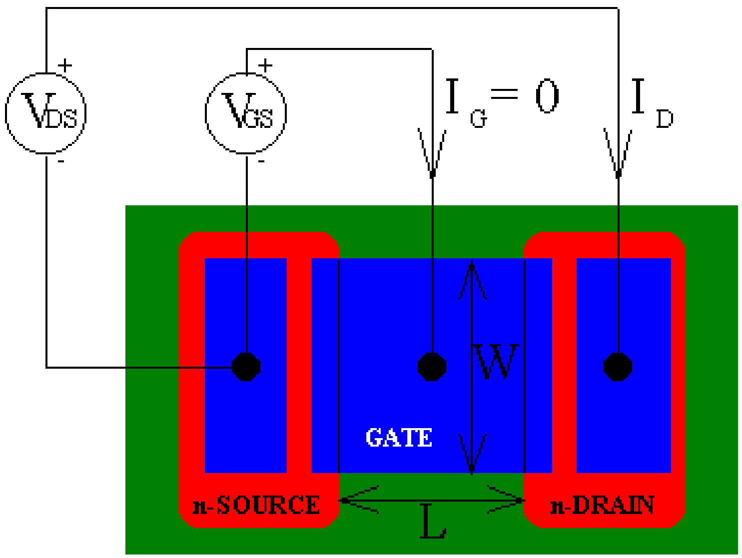


Illustrates the behaviour of a typical complimentary pair of power MOSFETs made by Hitachi for use in hi-fi amplifiers.

- Note that with a n-channel device we apply a +ve gate voltage to allow source-drain current, with a p-channel device we apply a -ve gate voltage.

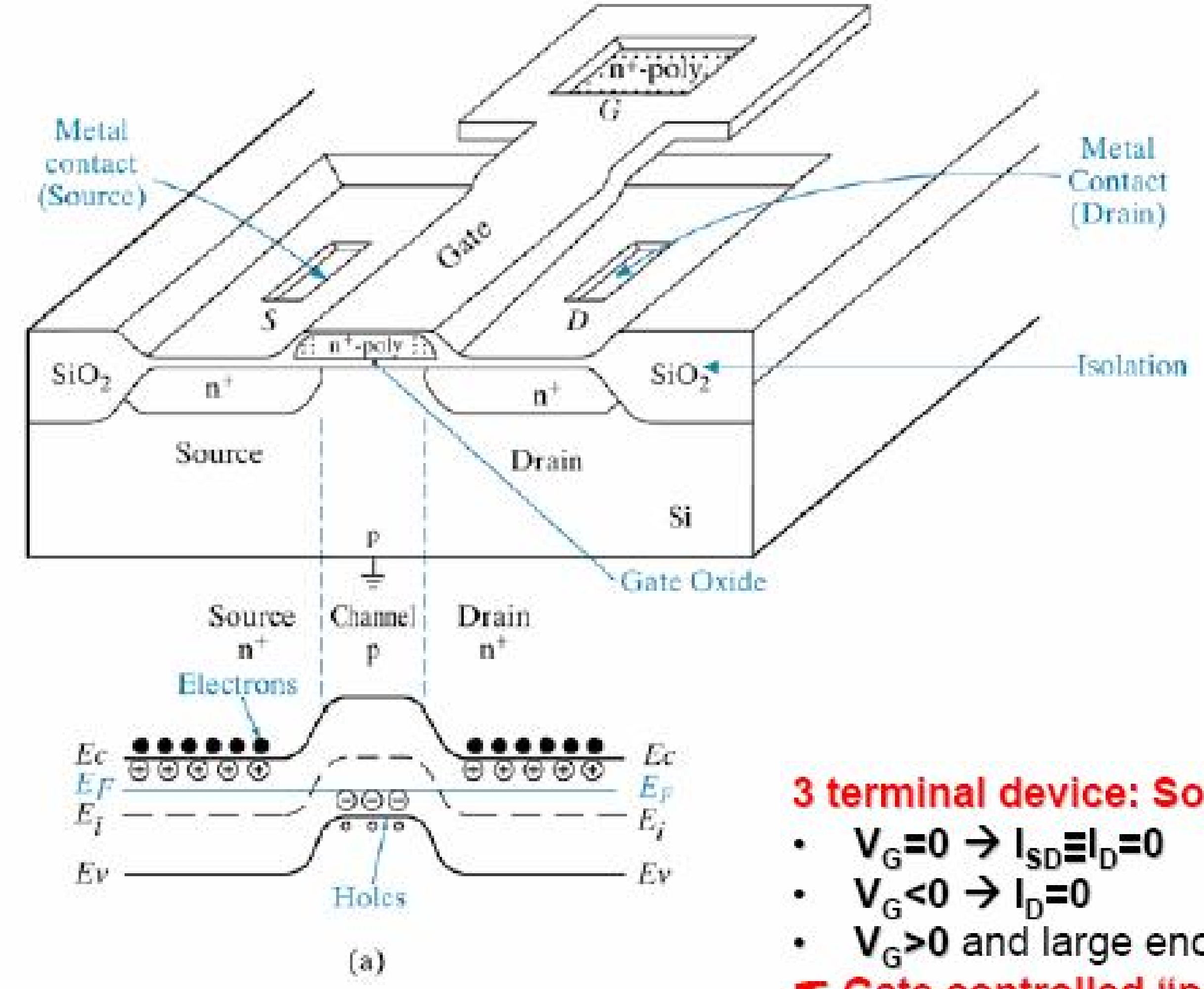
# Structure and principle of operation

- A top view of MOSFET, where the gate length, *L*, and gate width, *W.*
- Note that *L* does not equal the physical dimension of the gate, but rather the distance between the source and drain regions underneath the gate.
- The overlap between the gate and the source/drain region is required to ensure that the inversion layer forms a continuous conducting path between the source and drain region.
- Typically this overlap is made as small as possible in order to minimize its parasitic capacitance.



- Top view of an n-type MOSFET

# MOSFET-Basic Structure



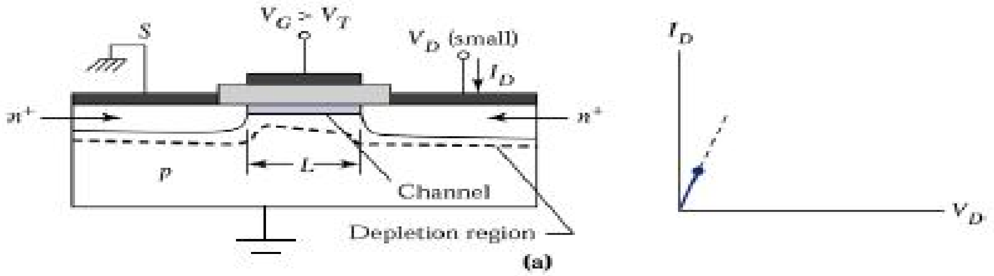
(a)

**3 terminal device: Source, Drain and Gate**

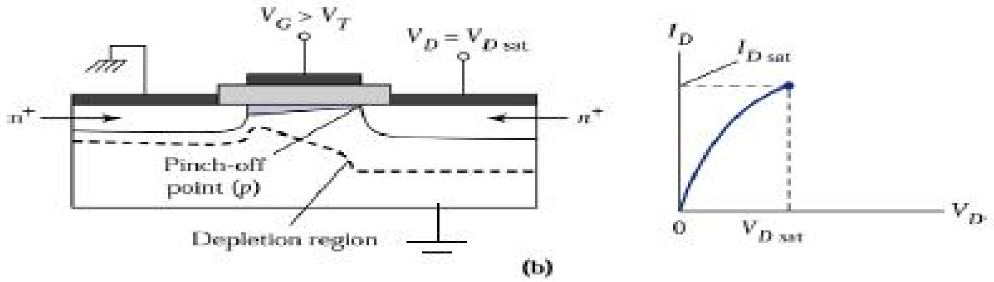
- $V_G=0 \rightarrow I_{SD} \equiv I_D=0$
- $V_G<0 \rightarrow I_D=0$
- $V_G>0$ and large enough → $I_D$ ⬀

☛ **Gate controlled "potential barrier" (or "resistor")**

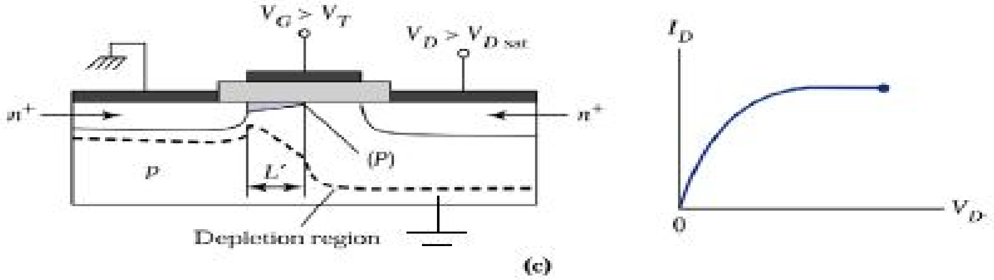
# I-V Characteristics of MOSFET



(*a*) Low drain voltage.



(*b*) Onset of saturation. Point *P* indicates the pinch-off point.



(*c*) Beyond saturation.

# I-V Characteristics of MOSFET

**Objectives: To calculate $I_D=f(V_D, V_G)$**

**−Assumptions:**

- Only drift current is considered
- $\mu_n$=const
- $N_a$=const
- Reverse leakage current is negligible
- Charges in the surface depletion regions are solely created by the gate

**$I_D$ – due to drift of mobile charges → $Q_n$?**

Voltage across oxide

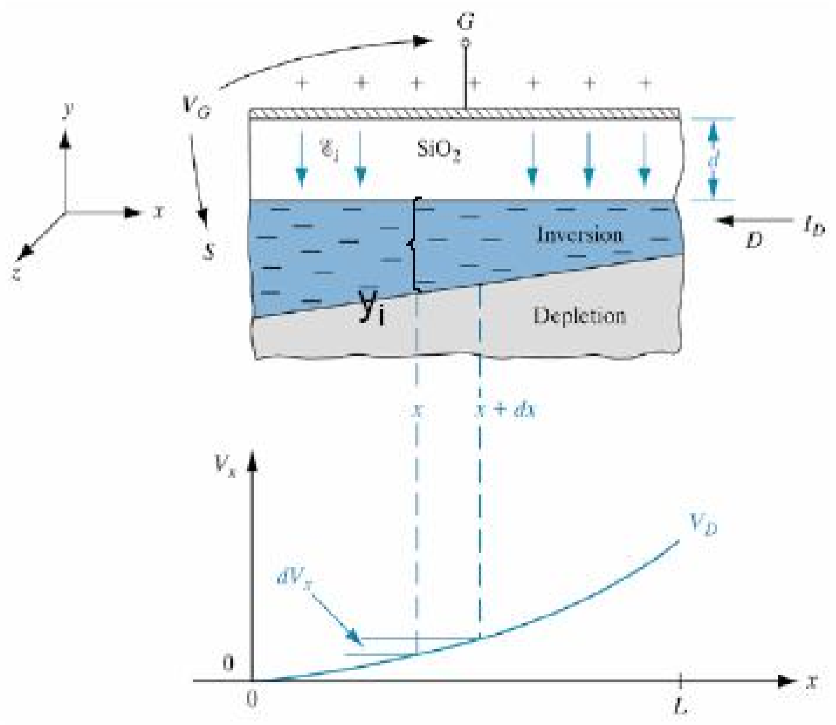
Voltage across depletion region

*$V_D=0$:*

$$V_G = V_{FB} - \frac{Q_d + Q_n}{C_i} + \Psi_s$$

$$\Rightarrow Q_n = -C_i\left(V_G - V_{FB} - \Psi_s - \frac{qN_aW}{C_i}\right)$$

# I-V Characteristics of MOSFET



$V_D \neq 0$: $\Psi_s \rightarrow \Psi_s + V_x \rightarrow W(x)$

$$I_D = g(x)dV_x$$

**Channel width**

$$g(x) = \frac{\sigma(x)Zy}{dx}$$

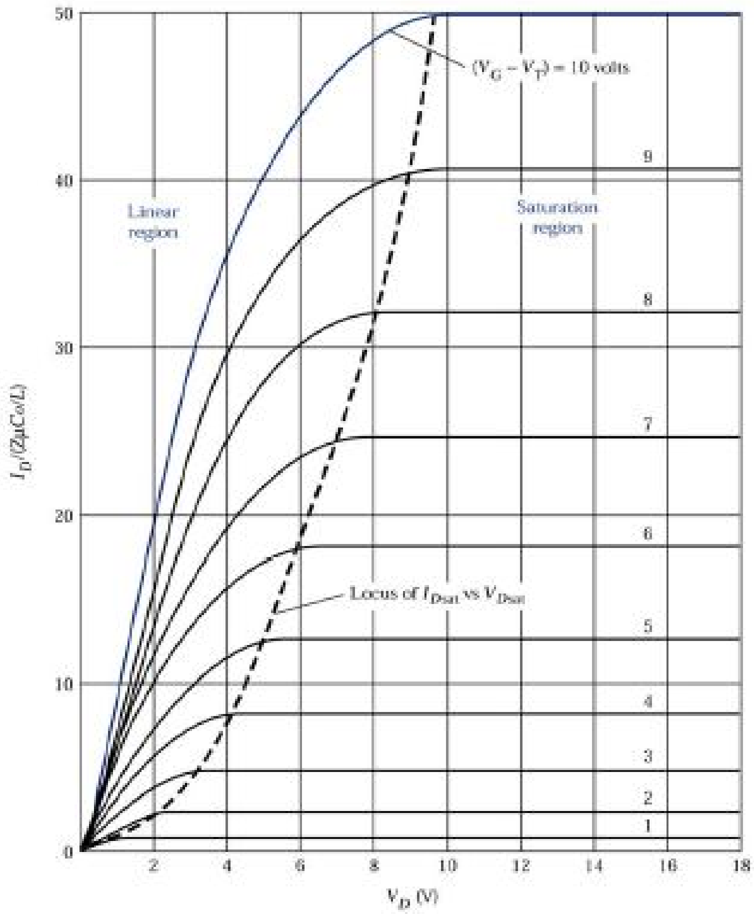
$$\sigma = \sigma(y) = qn(y)\mu_n(y)$$

$$g(x) = \frac{Z\int_0^{y_i}\sigma(y)dy}{dx} = \frac{Z\mu_n|Q_n(x)|}{dx}$$

**Channel conductance**

$$I_D = \frac{z\mu_n C_i}{L}\left[(V_G - 2\Psi_B - {}^{V_D}\!/\!_2)V_D - \frac{2}{3}\frac{\sqrt{2\varepsilon_s qN_a}}{C_i}\left[(V_D + 2\Psi_B)^{\frac{3}{2}} - (2\Psi_B)^2\right]\right]$$

# Ideal Output Characteristics of MOSFET



$I_D$=f($V_D$), $V_G$=const:

**1) Small $V_D$ – linear region**

$$I_D = \frac{z\mu_n C_i}{L}(V_G - V_T)V_D$$

$$g = \frac{\partial I_D}{\partial V_D} = \frac{z\mu_n C_i}{L}(V_G - V_T)$$

☛ Channel conductance *g* is determined by $V_G$

**2) Saturation:**

**$V_D$(sat)→ $Q_n(L)=0$- pinch-off near the drain**

**$V_D$(sat)≈$V_G$-$V_T$**

$$I_D(sat) = \frac{z\mu_n C_i}{2L}(V_G - V_T)^2$$

$$g = 0$$

☛ $I_D$ is determined by $V_G$, does not depend on $V_D$

☛ Changing z → Change in the current capacity

# Ideal Transfer Characteristics of MOSFET

**$I_D = f(V_G)$, $V_D = \text{const}$:**

**1) Small $V_D$**

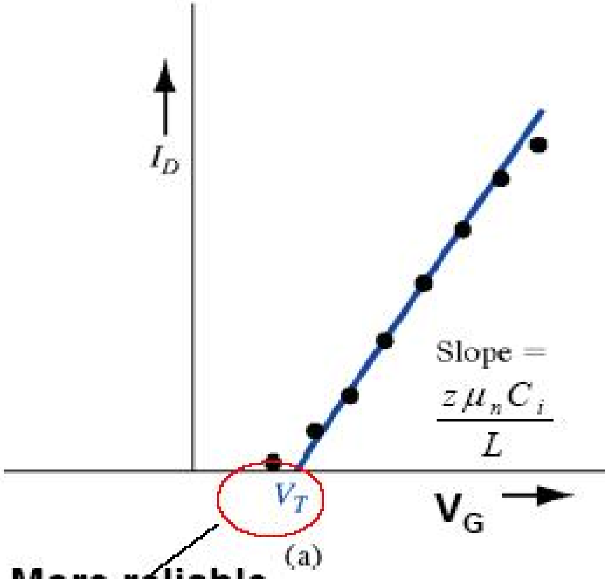
$$I_D = \frac{z\mu_n C_i}{L}(V_G - V_T)V_D$$

$$g_m = \frac{\partial I_D}{\partial V_G} = \frac{z\mu_n C_i}{L} V_D$$

**2) Large $V_D$ (saturation)**

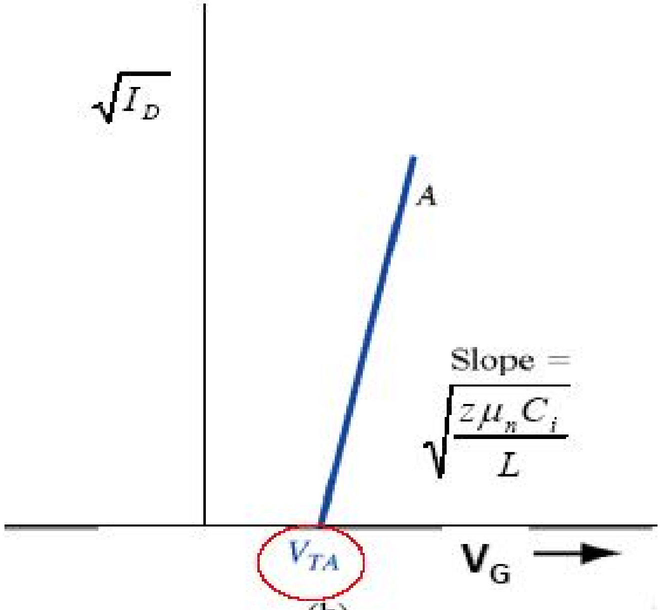
$$I_D(sat) = \frac{z\mu_n C_i}{2L}(V_G - V_T)^2$$

$$g_m(sat) = \frac{\partial I_D(sat)}{\partial V_G} = \frac{z\mu_n C_i}{L}(V_G - V_T)$$



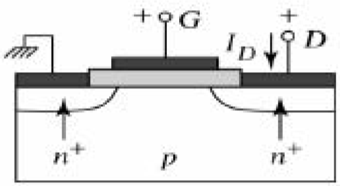
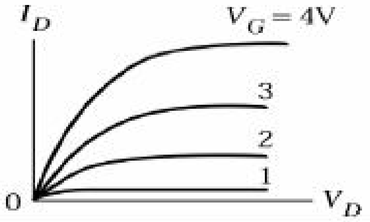
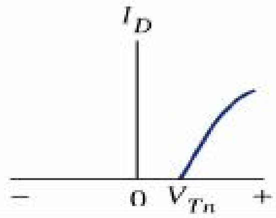
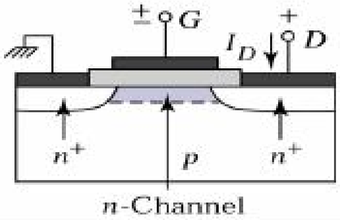
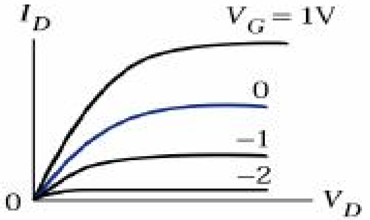
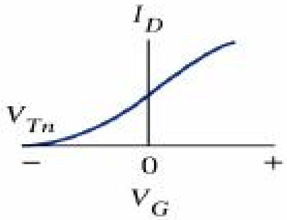
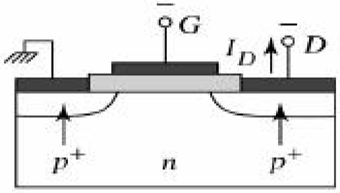
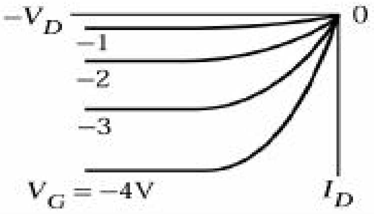
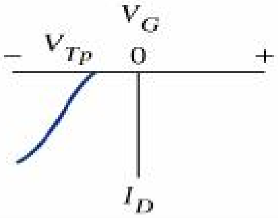
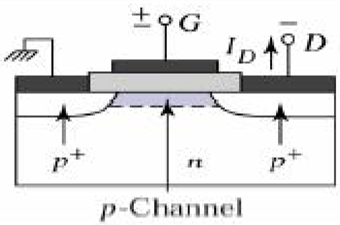
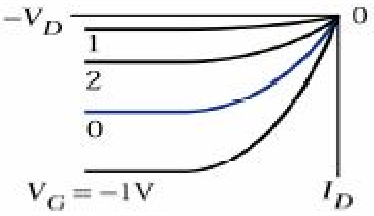
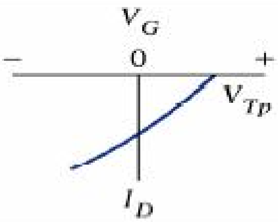
(a)

**More reliable**

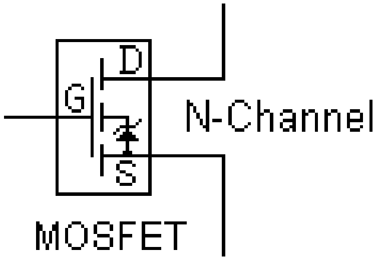


(b)

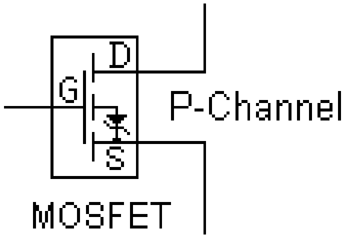
# Types of MOSFET

| Type | Cross Section | Output Characteristics | Transfer Characteristics |
|---|---|---|---|
| *n*-Channel Enhancement (Normally Off) |  |  |  |
| *n*-Channel Depletion (Normally On) |  |  |  |
| *p*-Channel Enhancement (Normally Off) |  |  |  |
| *p*-Channel Depletion (Normally On) |  |  |  |

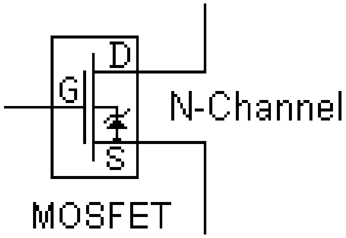
# Types of MOSFET



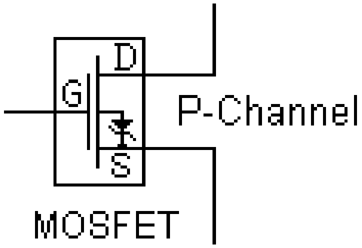
Enhancement Mode



Enhancement Mode



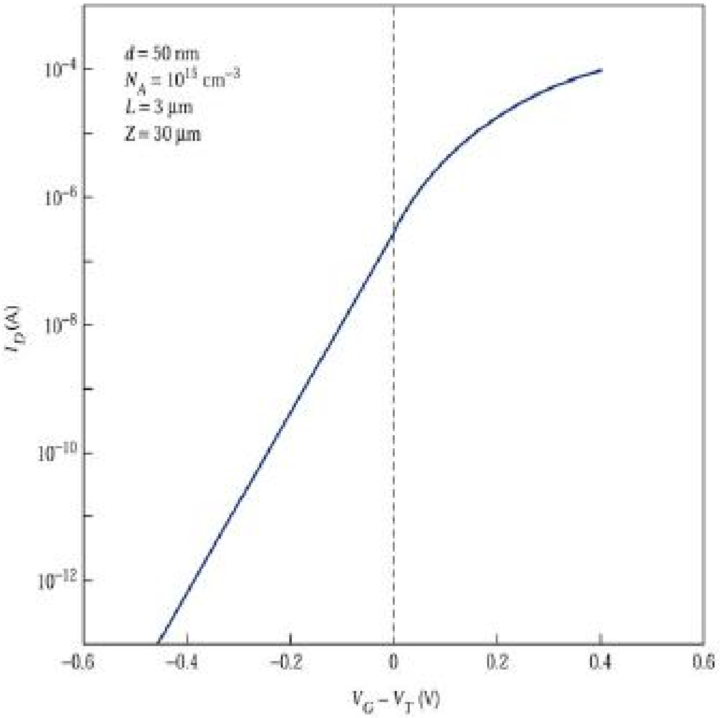
Depletion Mode



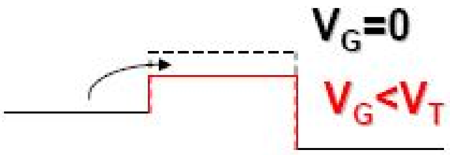
Depletion Mode

# Subthreshold region

**Ideal case: $I_D = 0$ when $V_G < V_T$**
**Real device – subthreshold current:**



**Origin?**



**Diffusion current !:**

$$I_D(sub) \propto \exp(\frac{qV_G}{kT})$$

**Important parameter – subthreshold swing S:**

$$S = \left[\frac{\partial(\log I_D)}{\partial V_G}\right]^{-1}$$

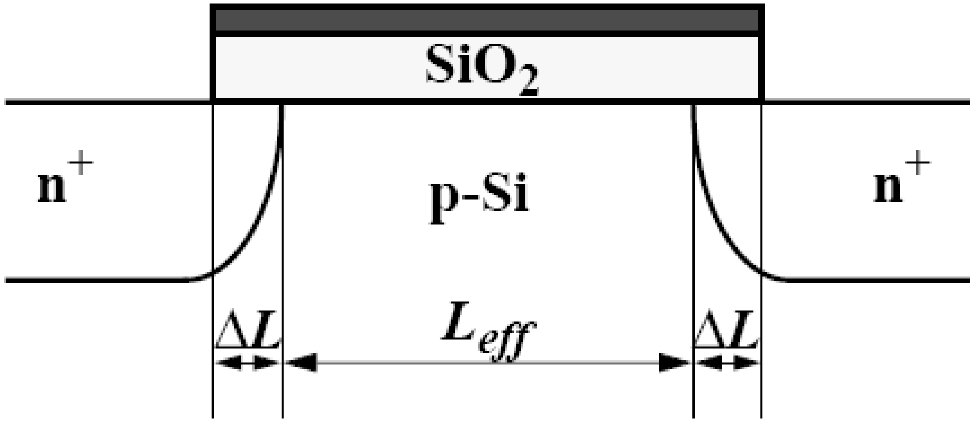
**S should be as small as possible: 70 -100 mV/decade**

**Small $V_T$→ High subthreshold (leakage) current, high power losses**
**High $V_T$→ Low drive current**
**Historically - $V_T$≈ 0.7 V**

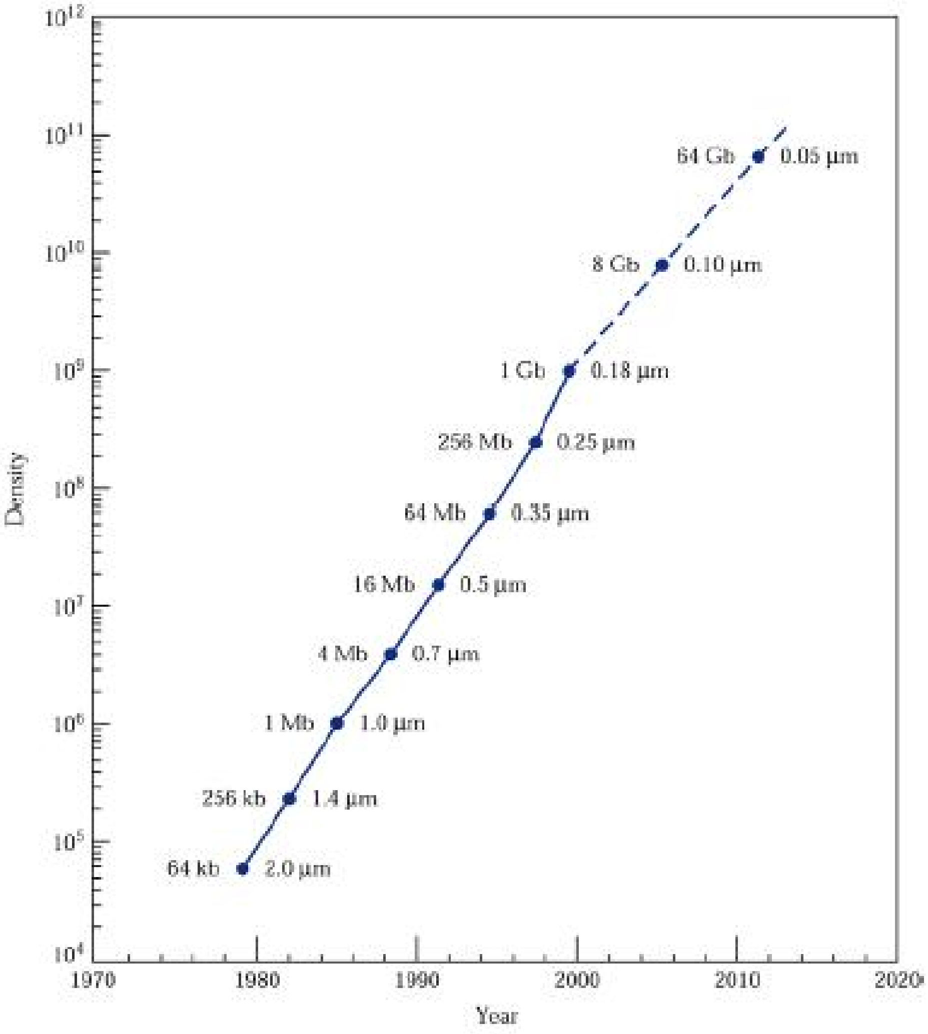
# Channel Length

- Short channel (<1μm) MOSFET
- Effective channel length:

$$L_{eff} = L - 2\Delta L$$



# MOSFET Dimensions - Trend

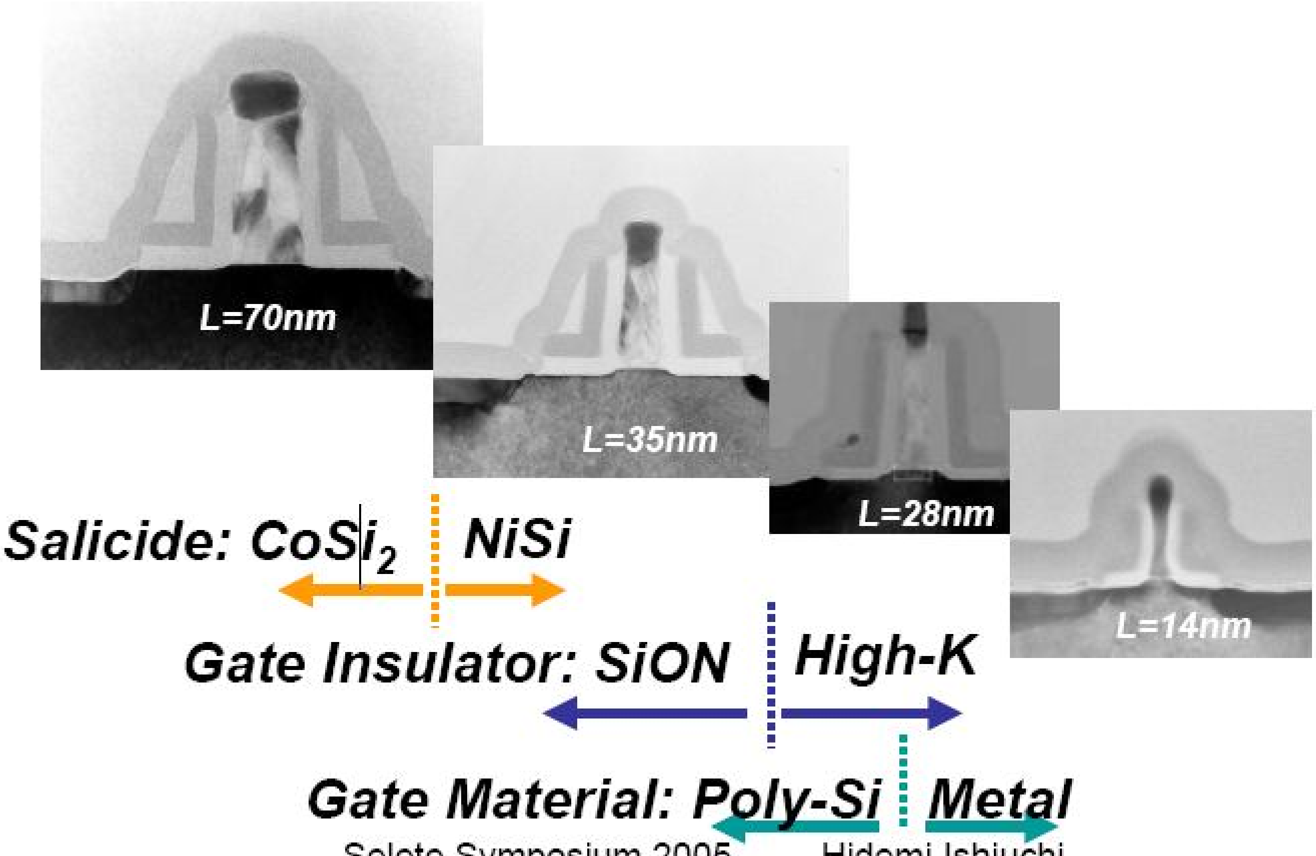


**Technological trend → Scaling down device dimensions**

**Advantages:**

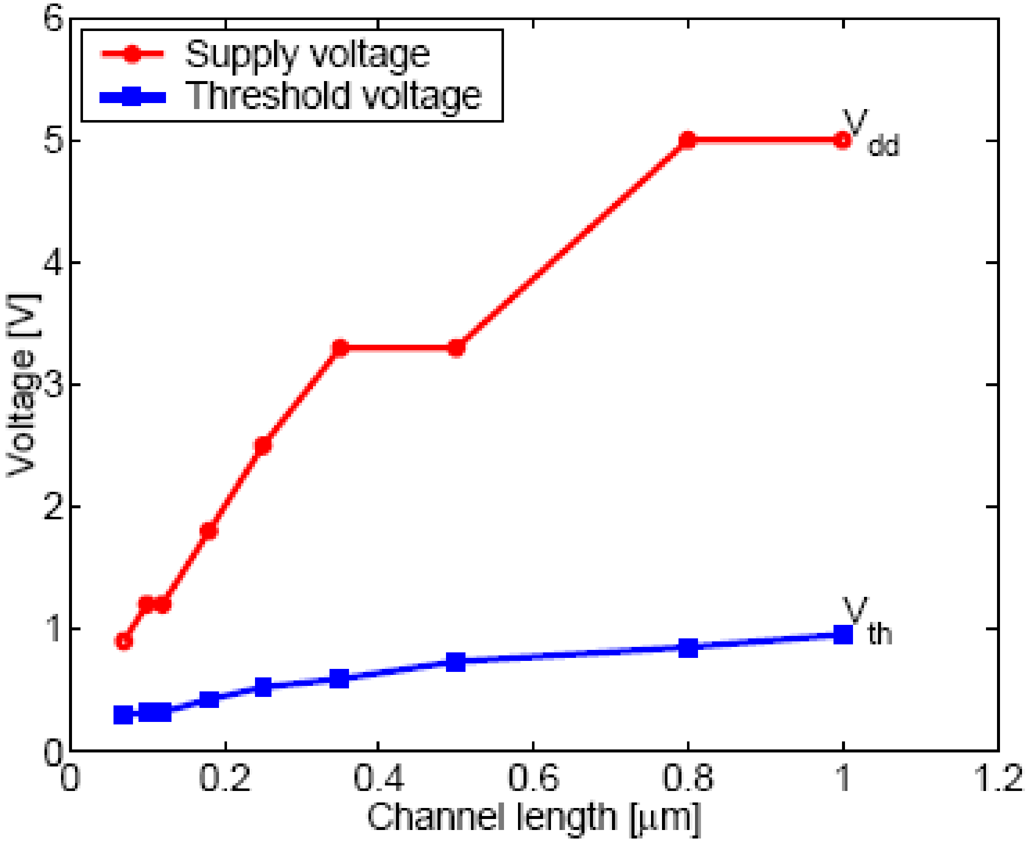
- **Higher device density**
- **Improved driving current $I_D \sim 1/L$**
- **Higher frequency**

# MOSFET scaling scenario



Selete Symposium 2005 Hidemi Ishiuchi

# Voltage Scaling



# Power Supply Voltage

- When scaling down the MOSFET channel length, the supply voltage $V_{dd}$ must be reduced as well to keep device power $P$ in reasonable limits

$$P \propto f V_{dd}^{2}$$

- The lower power supply voltage can increase the gate delay $\tau_d$ ($\alpha$=1.2-1.5 for submicronic process):

$$\tau_d \propto \frac{V_{dd}}{(V_{dd} - V_T)^{\alpha}}$$

# Threshold Voltage

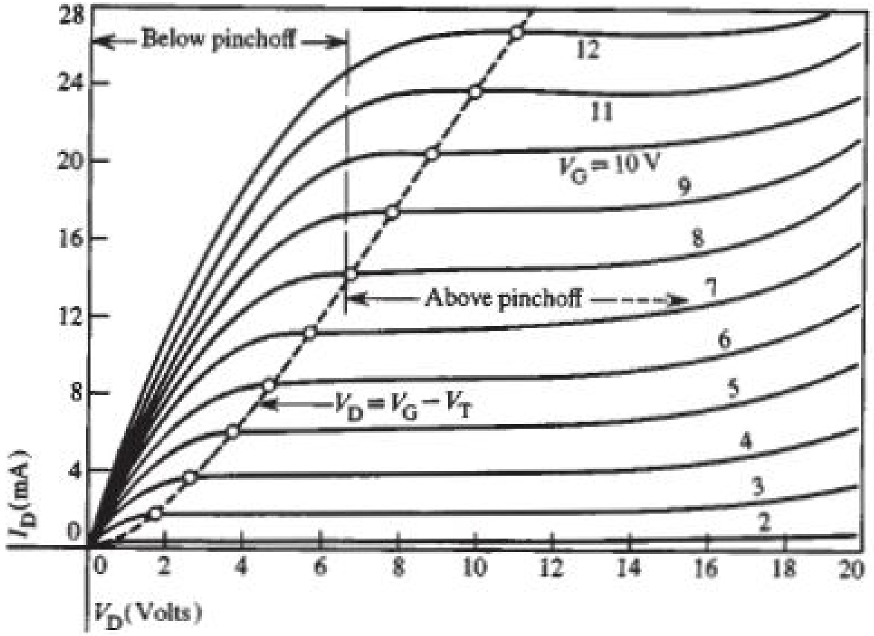
- **Ideal oxide threshold voltage for an n-channel MOSFET:**

$$V_T = \Phi_{ms} + 2|\phi_F| + \sqrt{\frac{4q\varepsilon_{Si}N_a|\phi_F|t_{ox}}{\varepsilon_{ox}}}$$

**The following effects can change the threshold voltage:**

- **Body effect (substrate bias),** $\Delta V_T \propto (V_B)^{1/2}$
- **Drain-induced barrier lowering (DIBL)** $\Delta V_T \propto V_{DS}$

# Threshold Voltage



$V_T$ = Minimum gate to source voltage to turn device on

$\epsilon_{ox}$ = Relative dielectric constant of oxide under gate electrode ($\epsilon_{ox} = 4$ for $SiO_2$)

$\epsilon_{si}$ = Relative dielectric constant of silicon ($\epsilon_{si} = 12$)

$\epsilon_0$ = Dielectric constant of free space ($8.85 \times 10^{-14}$ Farads/cm)

$t_{ox}$ = thickness of oxide under gate

$q$ = electronic charge ($1.6 \times 10^{-19}$ coulombs)

$Q_{ox}$ = oxide charge (considered to be located at Si-$SiO_2$ interface)

$Q_{ss}$ = surface state charges at Si-$SiO_2$ interface

$N_{a,eff}$ = effective impurity concentration level of substrate including allowance for boron depletion

$V_{S\text{-}sub}$ = voltage applied between source and substrate

$\Delta W_F$ = work function difference between aluminum and silicon (−0.8 V for p-type substrate)

$\psi_s$ = voltage across depletion layer at the onset of conduction in the absence of a substrate potential (0.75 V for 2.3 ohm-cm p-type silicon)

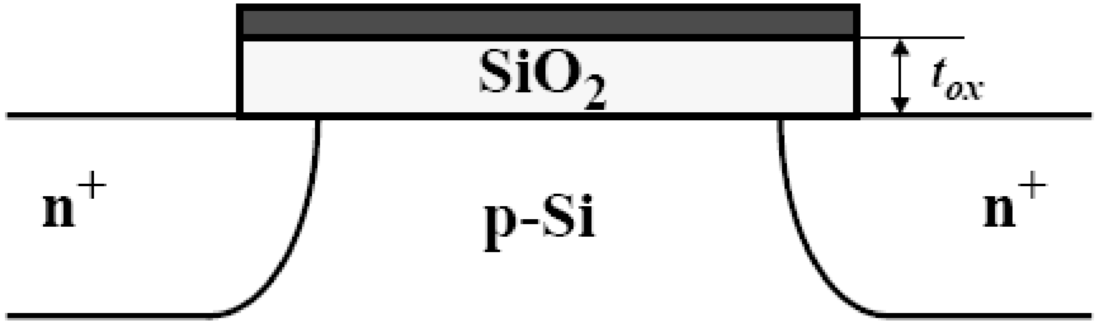
$\Delta V_{DT}$ = shift in threshold due to boron depletion effect

$$V_T = \frac{t_{ox}}{\epsilon_0 \epsilon_{ox}}\left[-Q_{ss} - Q_{ox} + \sqrt{2\epsilon_0 \epsilon_{si} q N_{a,eff}(V_{s\text{-}sub} + \psi_s)}\right] + \Delta W_F + \psi_s + \Delta V_{DT}$$

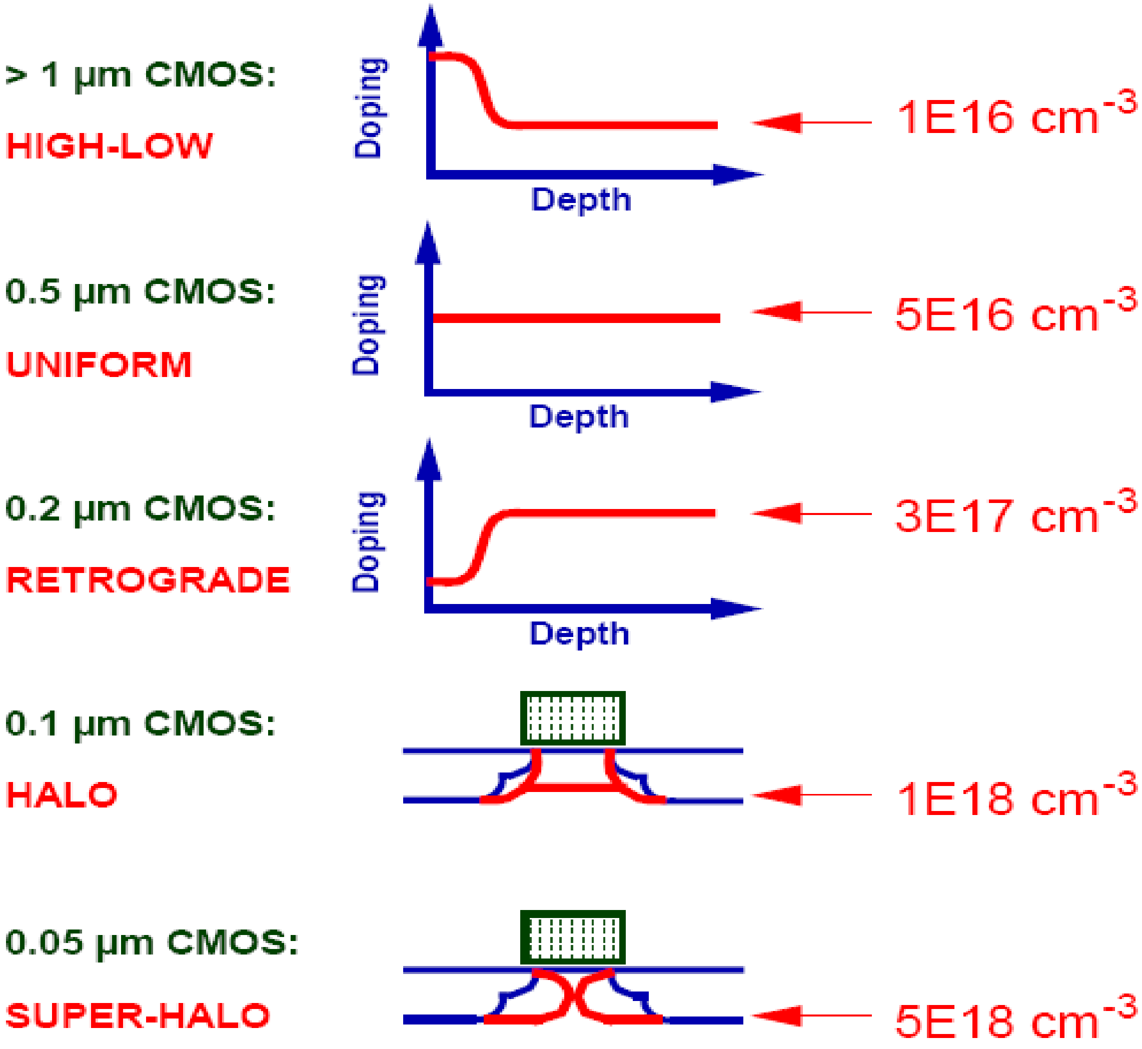
# Gate Oxide Thickness

- If the supply voltage $V_{dd}$ is reduced in a constant electric field $E$ scaling, the gate oxide thickness $t_{ox}$ must be reduced, because $E_{ox}=V_{dd}/t_{ox}$

- The lower limit for the gate oxide thickness is set by tunneling through the gate oxide current



# Channel Profile Evolution



**> 1 µm CMOS:**
**HIGH-LOW**
1E16 cm$^{-3}$

**0.5 µm CMOS:**
**UNIFORM**
5E16 cm$^{-3}$

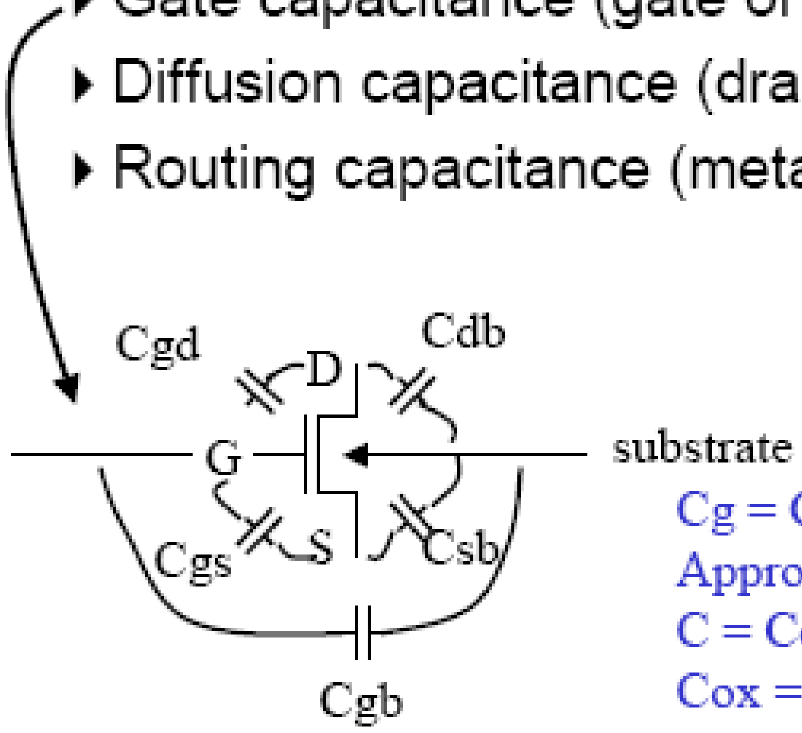
**0.2 µm CMOS:**
**RETROGRADE**
3E17 cm$^{-3}$

**0.1 µm CMOS:**
**HALO**
1E18 cm$^{-3}$

**0.05 µm CMOS:**
**SUPER-HALO**
5E18 cm$^{-3}$

# MOSFET Capacitances

- Three main forms:
  - Gate capacitance (gate of transistor)
  - Diffusion capacitance (drain regions)
  - Routing capacitance (metal, etc.)



$C_g = C_{gb} + C_{gs} + C_{gd}$

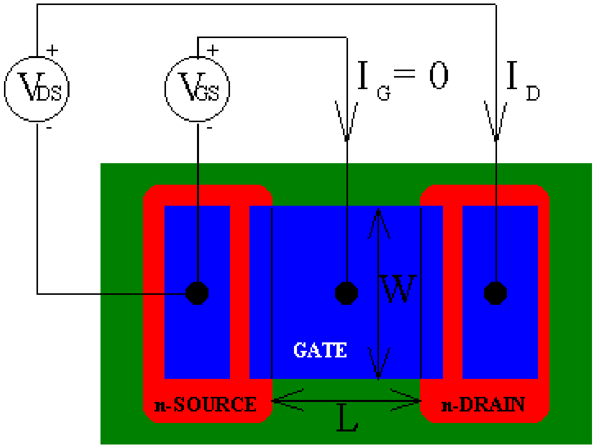
Approximated by

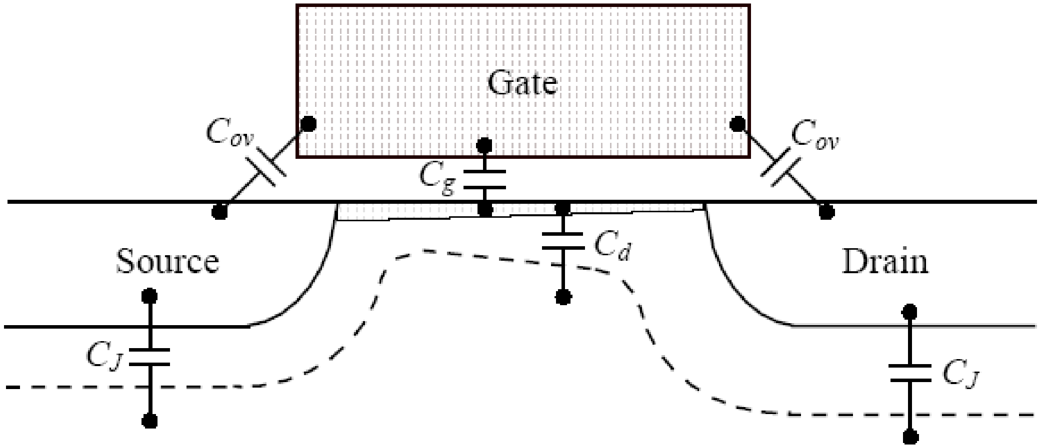
$C = C_{ox}A$

$C_{ox}$ = thin oxide cap

A = area of gate

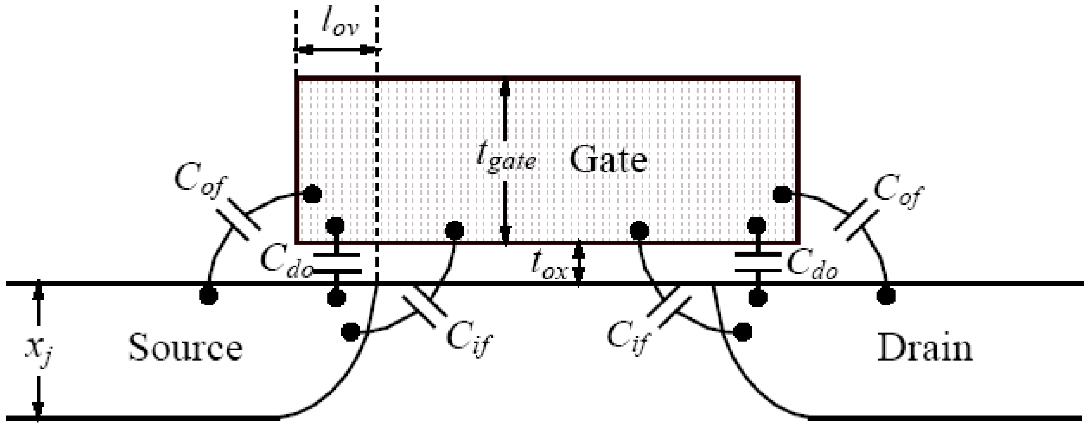
# MOSFET Capacitances





- Intrinsic capacitance: $C_g = WLC_{ox}$ $\quad C_g = \frac{2}{3}WLC_{ox}$
- Parasitic capacitances:
  - Depletion capacitance $C_d = \varepsilon_s WL / W_{dm}$
  - Overlap capacitance
  - Junction capacitance $C_j = \varepsilon_{si} Wd / W_{dj} = Wd\sqrt{\frac{\varepsilon_{si} q N_a}{2(\psi_{bi} + V_j)}}$

# Overlap Capacitance



Direct overlap

$$C_{do} = Wl_{ov}C_{ox} = \frac{\varepsilon_{ox}Wl_{ov}}{t_{ox}}$$

Outer fringe

$$C_{of} = \frac{2\varepsilon_{ox}W}{\pi}\ln\left(1+\frac{t_{gate}}{t_{ox}}\right)$$

Inner fringe

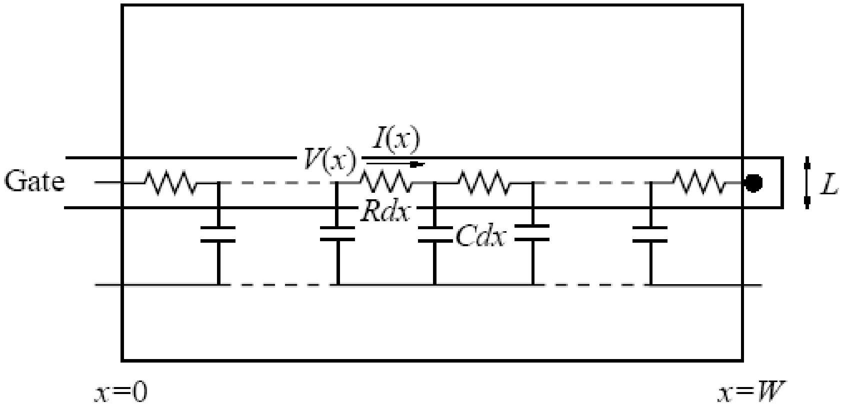
$$C_{if} = \frac{2\varepsilon_{si}W}{\pi}\ln\left(1+\frac{x_j}{2t_{ox}}\right)$$

For typical values of $t_{gate}/t_{ox} \approx 40$ and $x_j/t_{ox} \approx 20$,
$C_{of}/W \approx 2.3\varepsilon_{ox} \approx 0.08$ fF/μm, $C_{if}/W \approx 1.5\varepsilon_{si} \approx 0.16$ fF/μm (off state)

$$C_{ov}(V_g = 0) = C_{do} + C_{of} + C_{if} \approx \varepsilon_{ox}W\left(\frac{l_{ov}}{t_{ox}} + 7\right)$$

For reliability, $l_{ov} \approx (2\text{-}3)t_{ox}$, $C_{ov}/W \approx 10\varepsilon_{ox} \approx 0.3$ fF/μm

# Gate Resistance



The resistance per unit length is $R = \rho_g / L$
where $\rho_g$ is the silicide sheet resistivity (Ω/□).

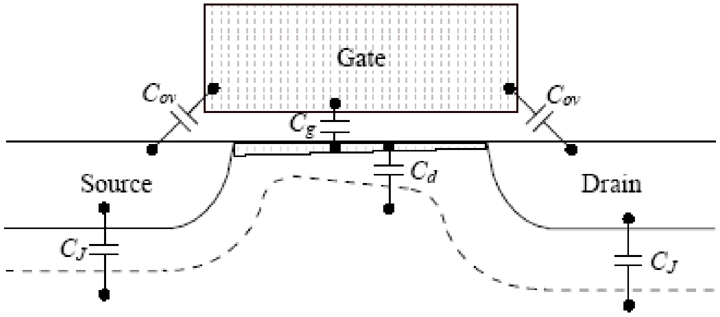
The capacitance per unit length is approximately $C = C_{ox} L = \dfrac{\varepsilon_{ox} L}{t_{ox}}$

Diffusion eq.: $\dfrac{\partial^2 V}{\partial x^2} = RC \dfrac{\partial V}{\partial t}$

The effective *RC* delay is *RCW*²/4 or $\tau_g = \dfrac{\rho_g C_{ox} W^2}{4}$
For $\rho_g$ = 10 Ω/□, $t_{ox}$ = 50 Å; $\tau_g$ < 1 ps if $W$ < 7.6 µm.

***Multiple-finger gate layouts with interdigitated source and drain regions should be used.***

# Components of $C_{in}$ and $C_{out}$



| | Input capacitance | Output capacitance |
|---|---|---|
| Intrinsic gate oxide capacitance (n & p) | 57% | 14% |
| Overlap capacitance | 43% | 35% |
| Junction capacitance (non-folded) | --- | 51% |

# New materials needed for scaling

- Since the early 1980s, the materials used for integrated MOSFET on silicon substrates have not changed greatly.
- The gate "metal" is made from highly-doped polycrystalline Si.
- The gate oxide is silicon dioxide.
- For the smallest devices, these materials will need to be replaced.

# New Gate Oxide

- The capacitance per area of the gate oxide is

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{K\epsilon_o}{t_{ox}}$$

- Scaled MOSFETs require larger $C_{ox}$, which has been achieved with smaller $t_{ox}$.
- Increasing K can also increase $C_{ox}$, and other oxides, "*high-K dielectrics*" are being developed, including for example, mixtures of $HfO_2$ and $Al_2O_3$.

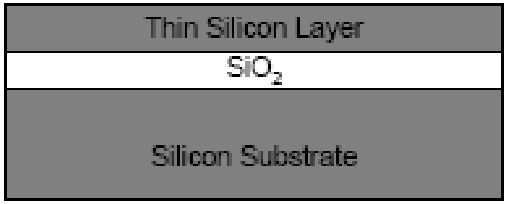
# New Gate Metal

- The doped polycrystalline silicon used for gates has a very thin depletion layer, approximately 1 nm thick, which causes scaling problems for small devices.
- Others metals are being investigated for replacing the silicon gates, including tungsten and molybdenum.

# Removing the substrate: Silicon on Insulator (SOI)

- For high-frequency circuits (about 5 GHz and above), capacitive coupling to the Si substrate limits the switching frequency.
- Also, leakage into the substrate from the small devices can cause extra power dissipation.
- These problems are being avoided by making circuits on insulating substrates (either sapphire or silicon dioxide) that have a thin, approximately 100 nm layer of crystalline silicon, in which the MOSFETs are fabricated.

# Silicon on Insulator (SOI)

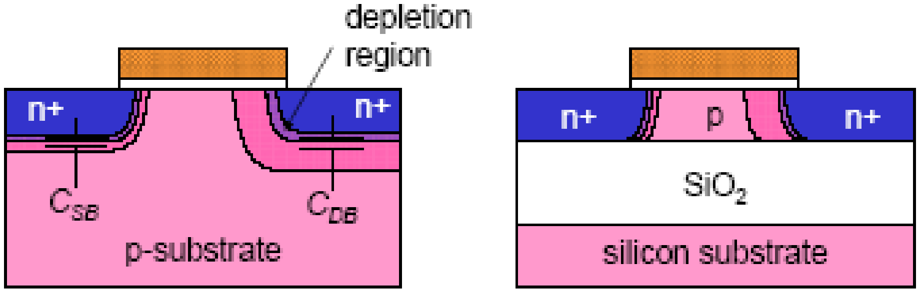


- SOI — silicon on insulator, refers to placing a thin layer of silicon on top of an insulator such as SiO2.
- The devices will be built on top of the thin layer of silicon.
- The basic idea of SOI is to reduced the parasitic capacitance and hence faster switching speed.

# Silicon on Insulator (SOI)

- Every time a transistor is turned on, it must first charge all of its internal (parasitic) capacitance before it can begin to conduct.
- The time it takes to charge up and discharge (turn off) the parasitic capacitance is much longer than the actual turn on and off of the transistor.
- If the parasitic capacitance can be reduced, the transistor can be switched faster — performance.

# Silicon on Insulator (SOI)



- One of the major source of parasitic capacitance is from the source and drain to substrate junctions.
- SOI can reduced the capacitance at the source and drain junctions significantly — by eliminating the depletion regions extending into the substrate.

# SOI CMOS

- Silicon-on-insulator CMOS offers a 20–35% performance gain over bulk CMOS.
- As the technology moves to the 0.13-μm generation, SOI is being used by more companies, and its application is spreading to lower-end microprocessors and SRAMs.
- Some of the recent applications of SOI in high-end microprocessors and its upcoming uses in low-power, radio-frequency (rf) CMOS, embedded DRAM (EDRAM), and the integration of vertical SiGe bipolar devices on SOI are described.

IBM J Res. Dev., Vol. 46, No. 2/3, 2002

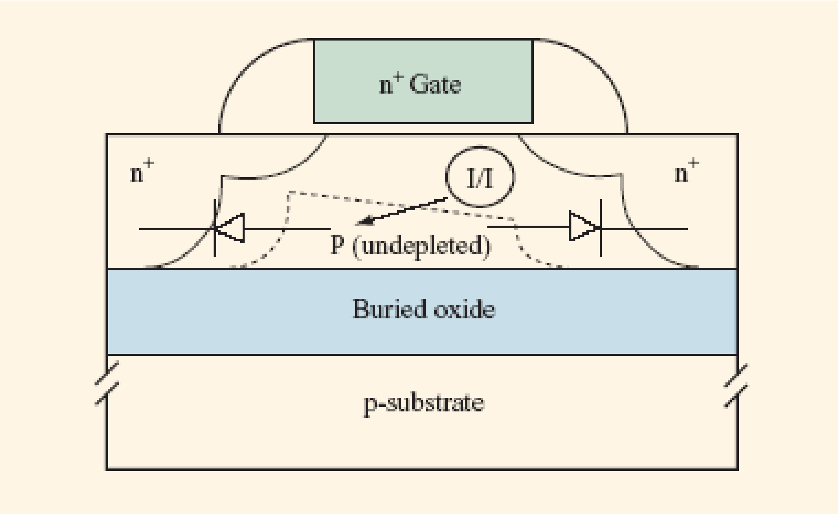


**Figure 1**

Cross section of n-FET on SOI, showing the body charging mechanisms. I/I = ion-implanted region.

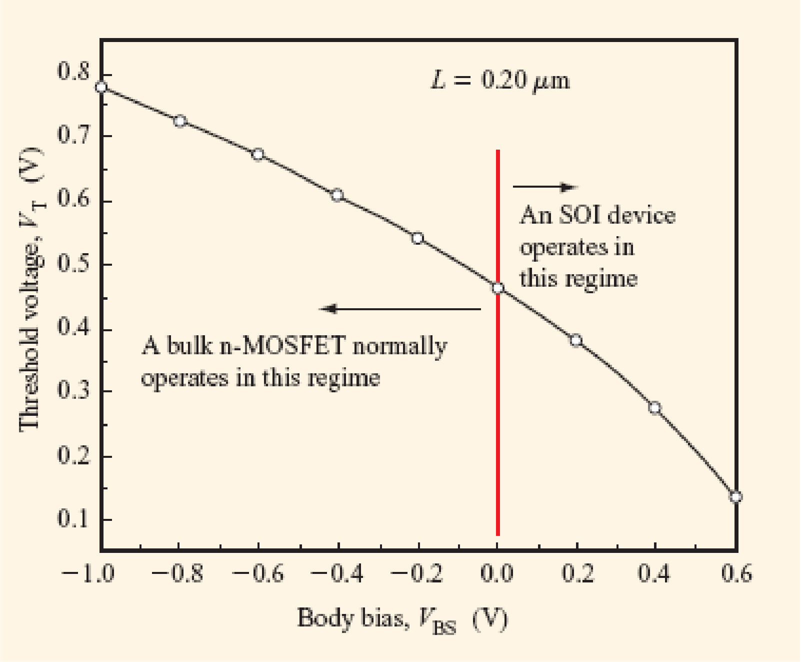
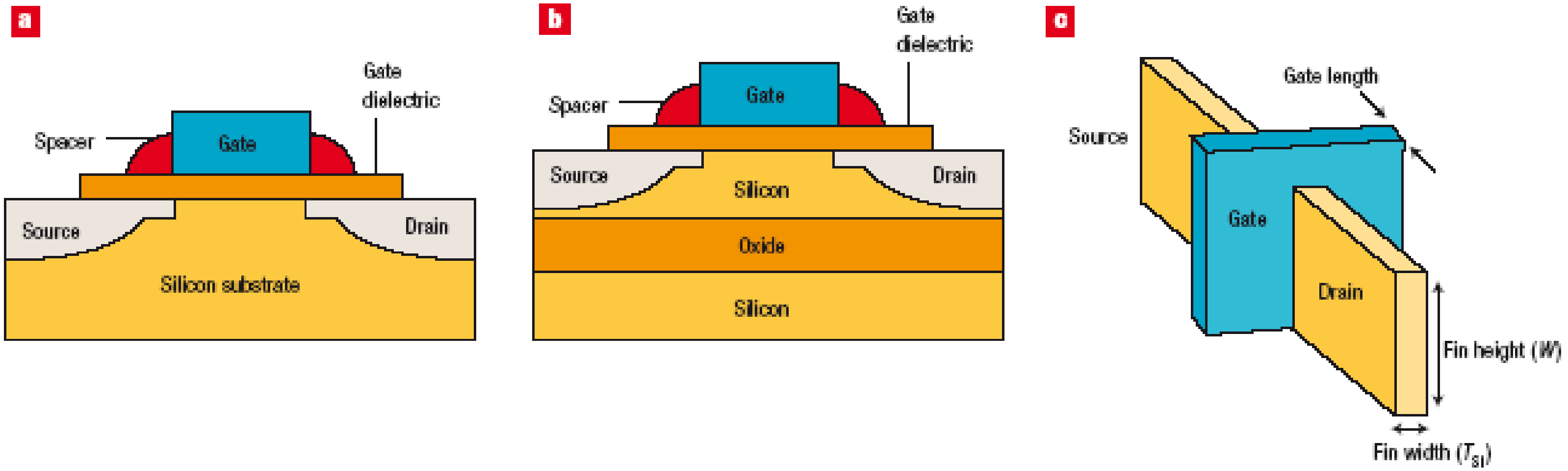


**Figure 2**

Threshold voltage as a function of body bias (i.e., the "body effect") for an n-FET with channel length $L = 20\,\mu$m.

IBM J Res. Dev., Vol. 46, No. 2/3, 2002

# Illustrations of silicon transistors



**a**, A traditional n-channel MOSFET uses a highly doped n-type polysilicon gate electrode, a highly doped n-type source/drain, a p-type substrate, and a silicon dioxide or oxynitride gate dielectric.

**b**, A silicon-on-insulator (SOI) MOSFET is similar to the traditional MOSFET except the active silicon is on a thick layer of silicon dioxide. This electrical isolation of the silicon reduces parasitic junction capacitance and improves device performance.

**c**, A finFET is a three-dimensional version of a MOSFET. The gate electrode wraps around a confined silicon channel providing improved electrostatic control of the channel electrons.

nature nanotechnology, 2 (2007) 25