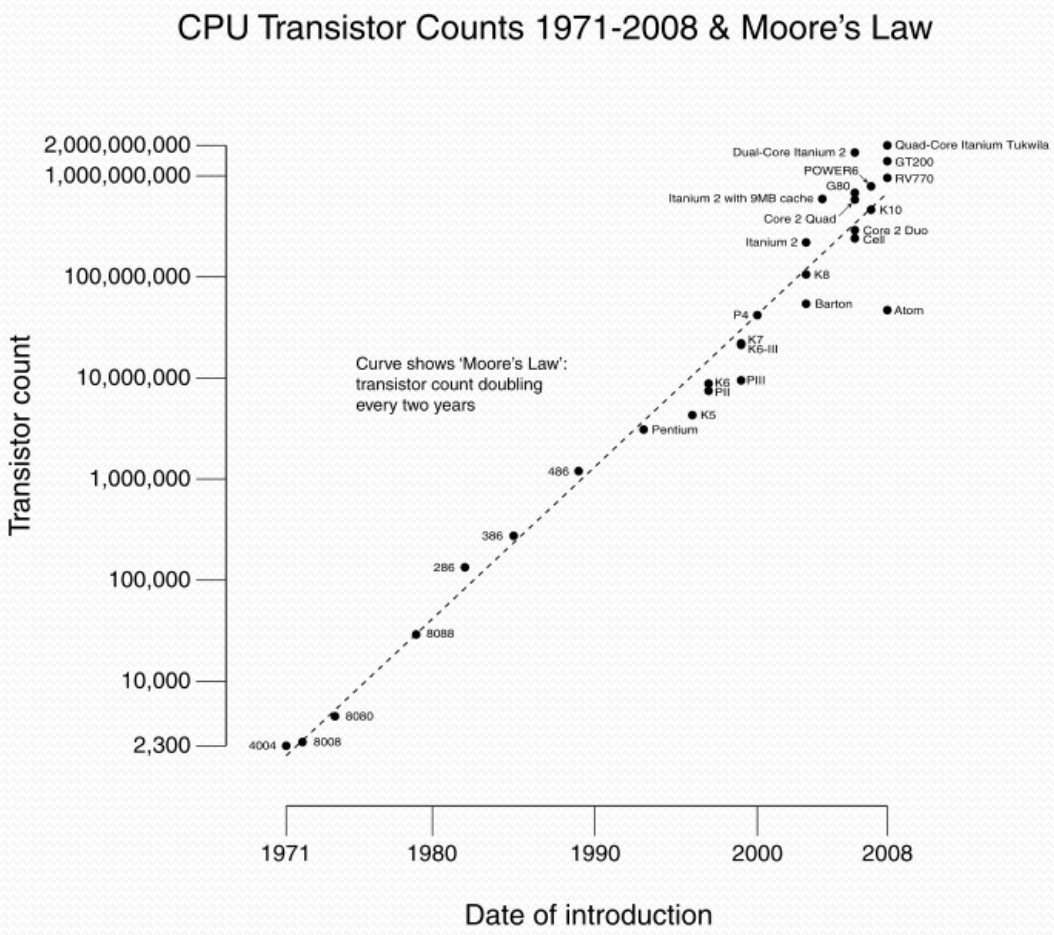
Unit 1
MOSFET Scaling

MOSFET Scaling



CMOS Device Scaling

Parameters	Variables	Scaling Factor
Dimensions	W, L, x_{ox}, x_j	$1/\lambda$
Potentials	V_{ds}, V_{gs}	$1/k$
Doping Concentration	N	λ^2/k
Electric Field	E	λ/k
Current	I_{ds}	λ/k^2
Gate Delay	T	k/λ^2

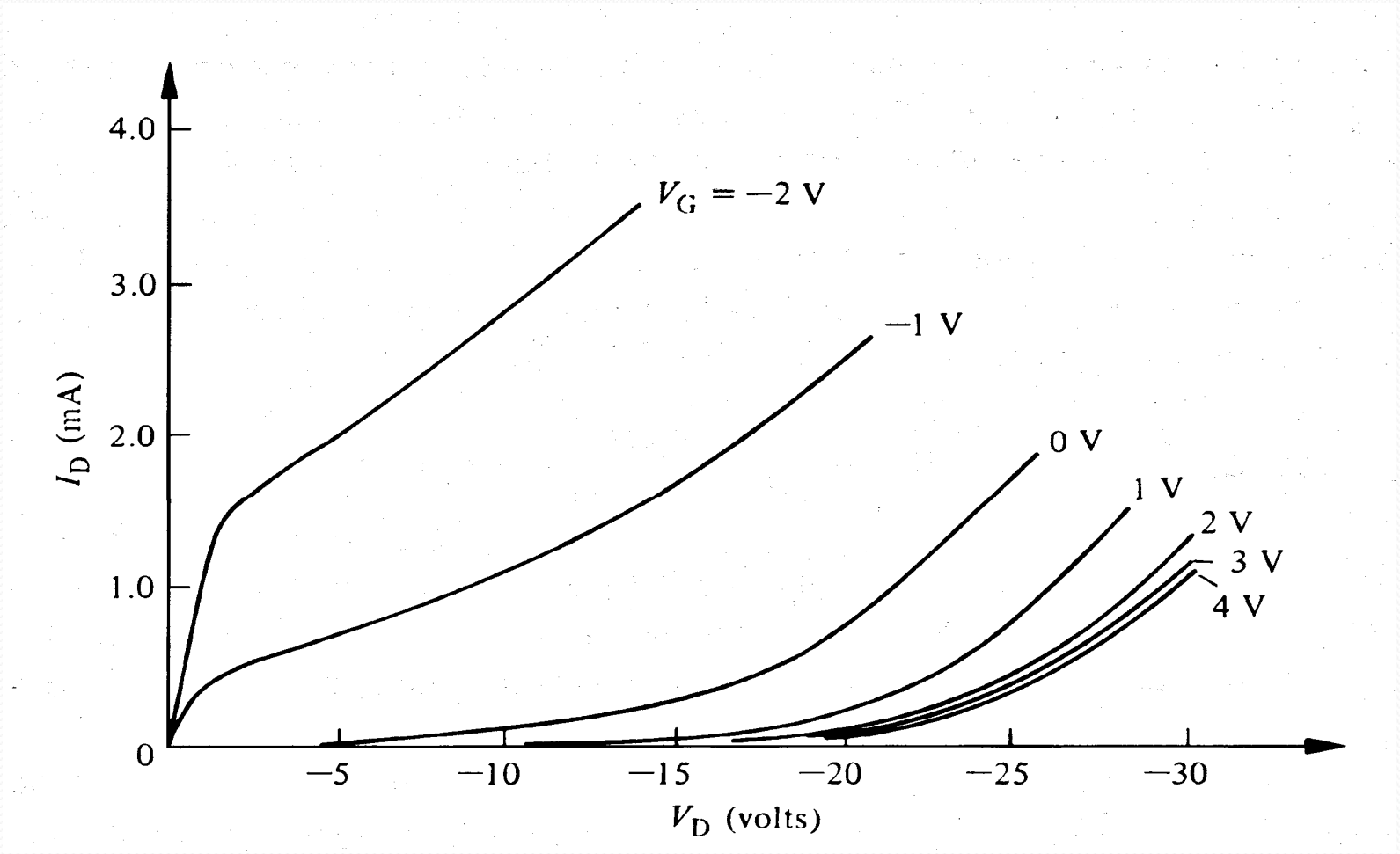
λ=dimensional scaling factor
k=supply voltage scaling factor

Constant Field Scaling: keep E constant in channel **k= λ**

Constant Voltage Scaling: keep supply voltage constant **k=1**(used for submicron scaling)

Parameters	**Const Field**	**Const Volt**
Dimensions	1/λ	1/λ
Potentials	1/λ	1
Doping Concentration	λ	λ^2
Electric Field	1	**λ**
Current	1/λ	λ
Gate Delay	1/λ	**1/ λ^2**

Short Channel Effects – punch through



Electric Fields Effects

- Vertical electric field: $E_{vert} = V_{gs} / t_{ox}$
  - Attracts carriers into channel
  - Long channel: $Q_{channel} \propto E_{vert}$
- Lateral electric field: $E_{lat} = V_{ds} / L$
  - Accelerates carriers from drain to source
  - Long channel: $v = \mu E_{lat}$

Mobility Degradation

- High E_{vert} effectively reduces mobility
  - Collisions with oxide interface

$$\mu_{\text{eff}-n} = \frac{540\frac{\text{cm}^2}{\text{V}\cdot\text{s}}}{1+\left(\frac{V_{gs}+V_t}{0.54\frac{\text{V}}{\text{nm}}t_{\text{ox}}}\right)^{1.85}}$$

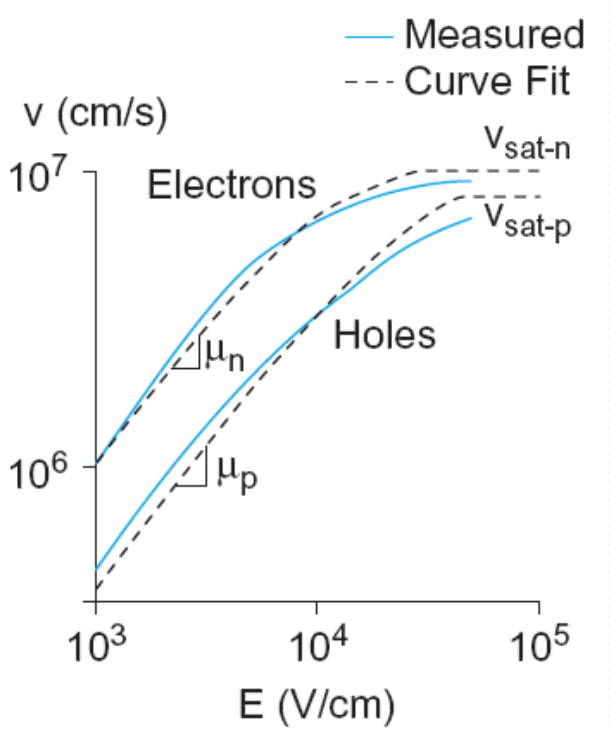
$$\mu_{\text{eff}-p} = \frac{185\frac{\text{cm}^2}{\text{V}\cdot\text{s}}}{1+\frac{\left|V_{gs}+1.5V_t\right|}{0.338\frac{\text{V}}{\text{nm}}t_{\text{ox}}}}$$

Velocity Saturation

- At high E_{lat}, carrier velocity rolls off
  - Carriers scatter off atoms in silicon lattice
  - Velocity reaches v_{sat}
    - Electrons: 10^7 cm/s
    - Holes: 8×10^6 cm/s
  - Better model

$$v = \begin{cases} \dfrac{\mu_{\text{eff}} E}{1 + \dfrac{E}{E_c}} & E < E_c \\ v_{\text{sat}} & E \geq E_c \end{cases}$$

$$E_c = \frac{2 v_{\text{sat}}}{\mu_{\text{eff}}}$$



Vel Sat I-V Effects

- Ideal transistor ON current increases with V_{DD}^2

$$I_{ds} = \mu C_{ox} \frac{W}{L} \frac{\left(V_{gs} - V_t\right)^2}{2} = \frac{\beta}{2}\left(V_{gs} - V_t\right)^2$$

- Velocity-saturated ON current increases with V_{DD}

$$I_{ds} = C_{ox} W \left(V_{gs} - V_t\right) v_{max}$$

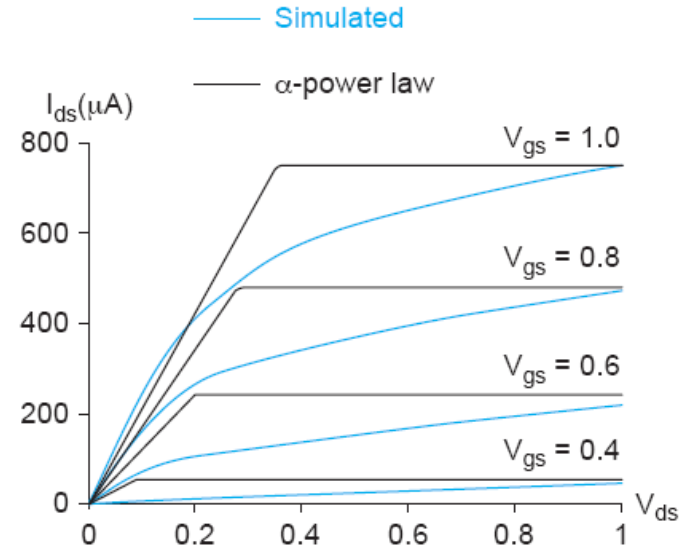
- Real transistors are partially velocity saturated
  - Approximate with α-power law model
  - $I_{ds} \propto V_{DD}^{\alpha}$
  - $1 < \alpha < 2$ determined empirically (≈ 1.3 for 65 nm)

α-Power Model

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ I_{dsat}\dfrac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \quad \text{linear} \\ I_{dsat} & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

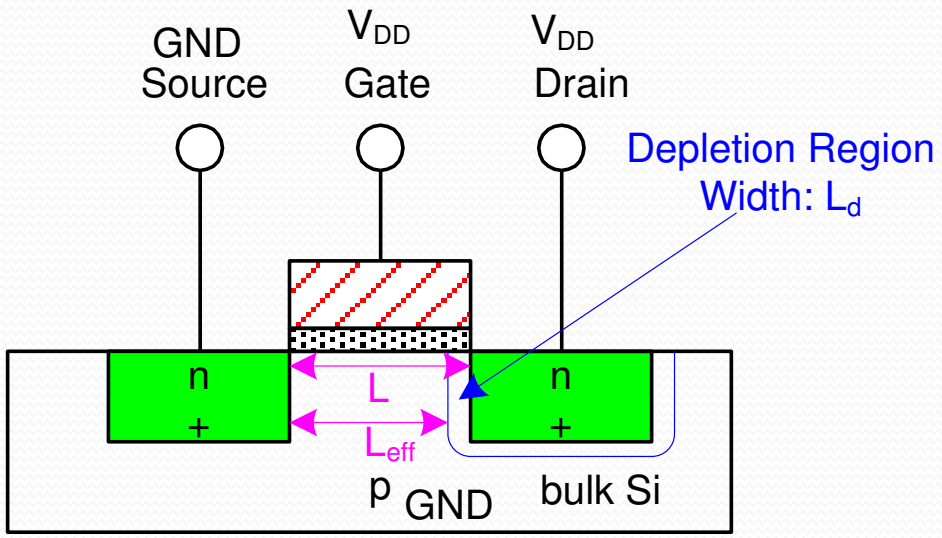
$$I_{dsat} = P_c \frac{\beta}{2}\left(V_{gs} - V_t\right)^{\alpha}$$

$$V_{dsat} = P_v \left(V_{gs} - V_t\right)^{\alpha/2}$$



Channel Length Modulation

- Reverse-biased p-n junctions form a *depletion region*
  - Region between n and p with no carriers
  - Width of depletion L_d region grows with reverse bias
  - $L_{eff} = L - L_d$
- Shorter L_{eff} gives more current
  - I_{ds} increases with V_{ds}
  - Even in saturation



Channel Length Modulation

$$I_{ds} = \frac{\beta}{2}\left(V_{gs} - V_t\right)^2\left(1 + \lambda V_{ds}\right)$$

- λ = *channel length modulation coefficient*
  - not feature size
  - Empirically fit to I-V characteristics

Threshold Voltage Effects

- V_t is V_{gs} for which the channel starts to invert
- Ideal models assumed V_t is constant
- Really depends (weakly) on almost everything else:
  - Body voltage: *Body Effect*
  - Drain voltage: *Drain-Induced Barrier Lowering*
  - Channel length: *Short Channel Effect*

Body Effect

- Body is a fourth transistor terminal
- V_{sb} affects the charge required to invert the channel
  - Increasing V_s or decreasing V_b increases V_t

$$V_t = V_{t0} + \gamma\left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s}\right)$$

- ϕ_s = *surface potential* at threshold

$$\phi_s = 2v_T \ln\frac{N_A}{n_i}$$

  - Depends on doping level N_A
  - And intrinsic carrier concentration n_i
- γ = *body effect coefficient*

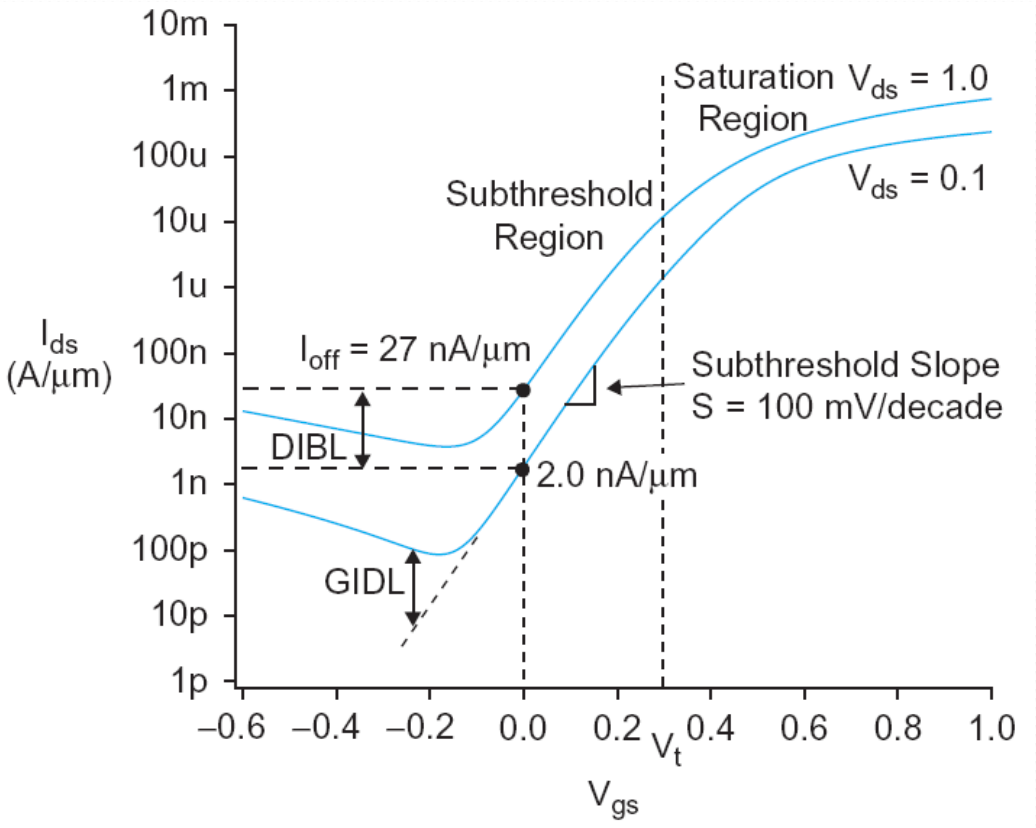
$$\gamma = \frac{t_{ox}}{\varepsilon_{ox}}\sqrt{2q\varepsilon_{si}N_A} = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{ox}}$$

Short Channel Effect

- In small transistors, source/drain depletion regions extend into the channel
  - Impacts the amount of charge required to invert the channel
  - And thus makes V_t a function of channel length
- Short channel effect: V_t increases with L
  - Some processes exhibit a reverse short channel effect in which V_t decreases with L

Leakage

- What about current in cutoff?
- Simulated results
- What differs?
  - Current doesn't go to 0 in cutoff



Leakage Sources

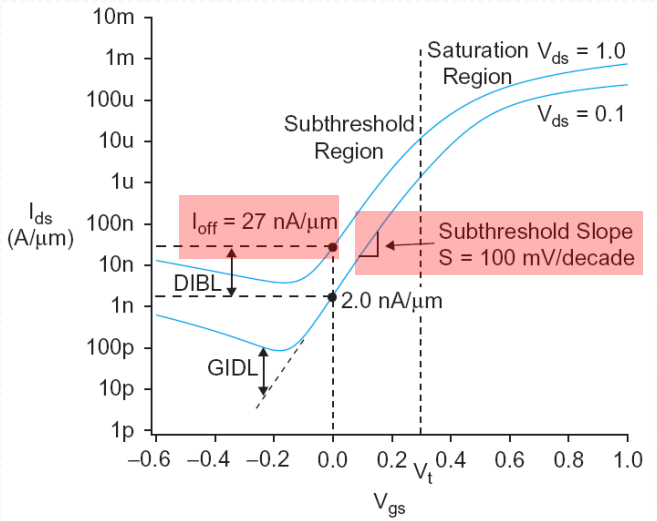
- Subthreshold conduction
  - Transistors can't abruptly turn ON or OFF
  - Dominant source in contemporary transistors
- Gate leakage
  - Tunneling through ultrathin gate dielectric
- Junction leakage
  - Reverse-biased PN junction diode current

Subthreshold Leakage

- Subthreshold leakage exponential with V_{gs}

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_{t0} + \eta V_{ds} - k_\gamma V_{sb}}{n v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right)$$

- n is process dependent
  - typically 1.3-1.7
- Rewrite relative to I_{off} on log scale



$$I_{ds} = I_{off} 10^{\frac{V_{gs} + \eta(V_{ds} - V_{dd}) - k\gamma V_{sb}}{S}} \left(1 - e^{\frac{-V_{ds}}{v_t}}\right)$$

$$S = \left[\frac{d(\log_{10} I_{ds})}{dV_{gs}}\right]^{-1} = n v_T \ln 10$$

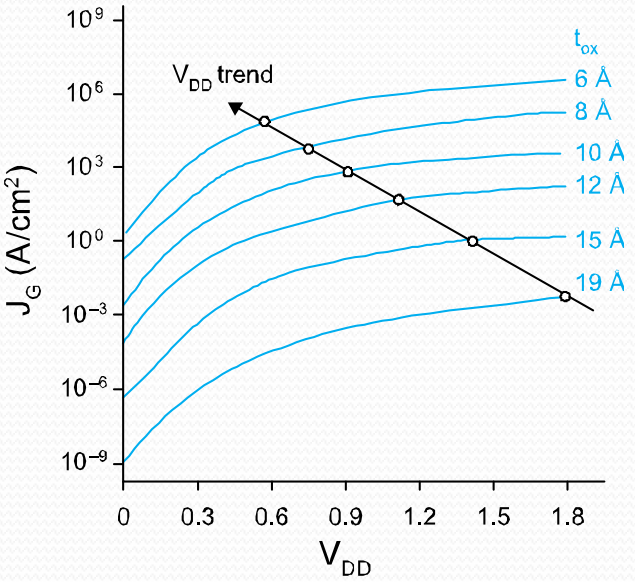
- S ≈ 100 mV/decade @ room temperature

Gate Leakage

- Carriers tunnel thorough very thin gate oxides
- Exponentially sensitive to t_{ox} and V_{DD}

$$I_{gate} = WA\left(\frac{V_{DD}}{t_{ox}}\right)^2 e^{-B\frac{t_{ox}}{V_{DD}}}$$

  - A and B are tech constants
  - Greater for electrons
    - So nMOS gates leak more



- Negligible for older processes (t_{ox} > 20 Å)
- Critically important at 65 nm and below (t_{ox} ≈ 10.5 Å)

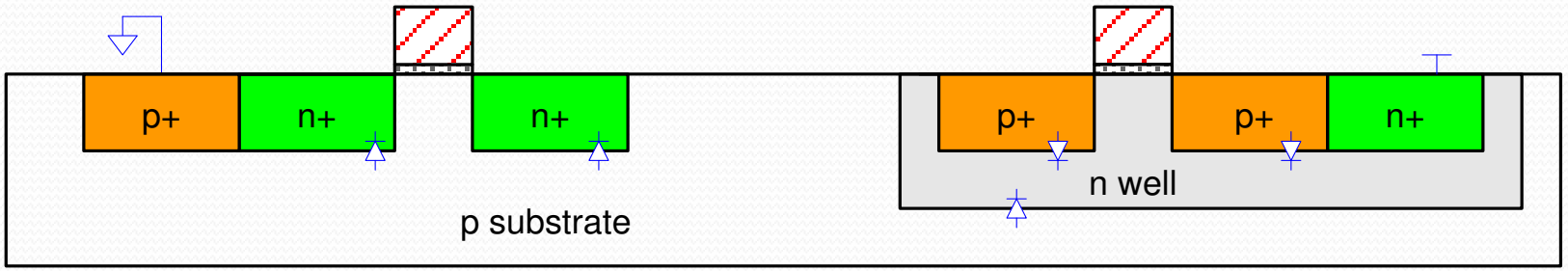
From [Song01]

Junction Leakage

- Reverse-biased p-n junctions have some leakage
  - Ordinary diode leakage
  - Band-to-band tunneling (BTBT)
  - Gate-induced drain leakage (GIDL)

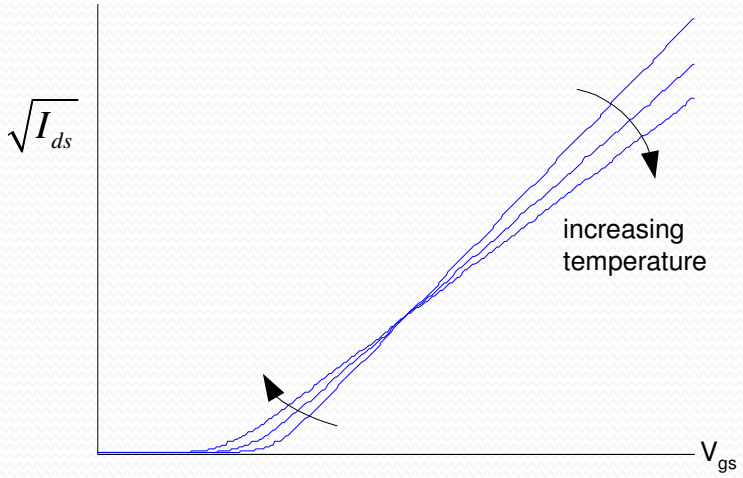
$$I_D = I_S \left(e^{\frac{V_D}{v_T}} - 1 \right)$$

- I_s depends on doping levels
  - And area and perimeter of diffusion regions
  - Typically < 1 fA/μm² (negligible)



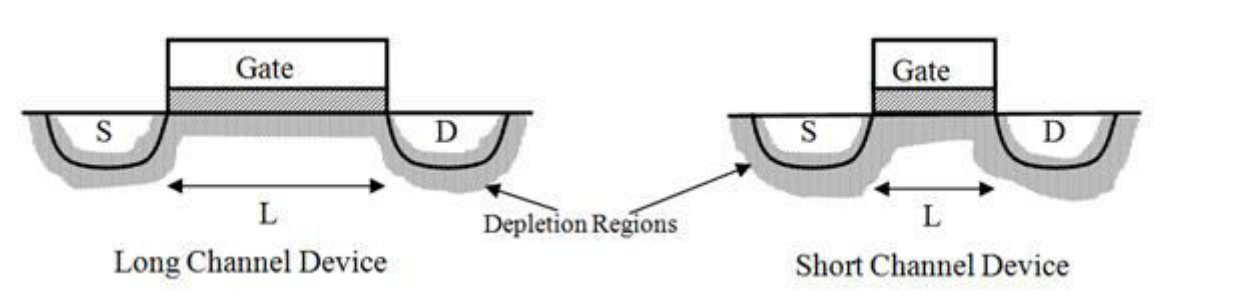
Temperature Sensitivity

- Increasing temperature
  - Reduces mobility
  - Reduces V_t
- I_{ON} decreases with temperature
- I_{OFF} increases with temperature



Small Geometry Effects

Short channel effects start to appear as the channel length becomes less than 10 times the depletion region width of the source/drain regions. The figure below demonstrates the difference between long and short channel transistors.



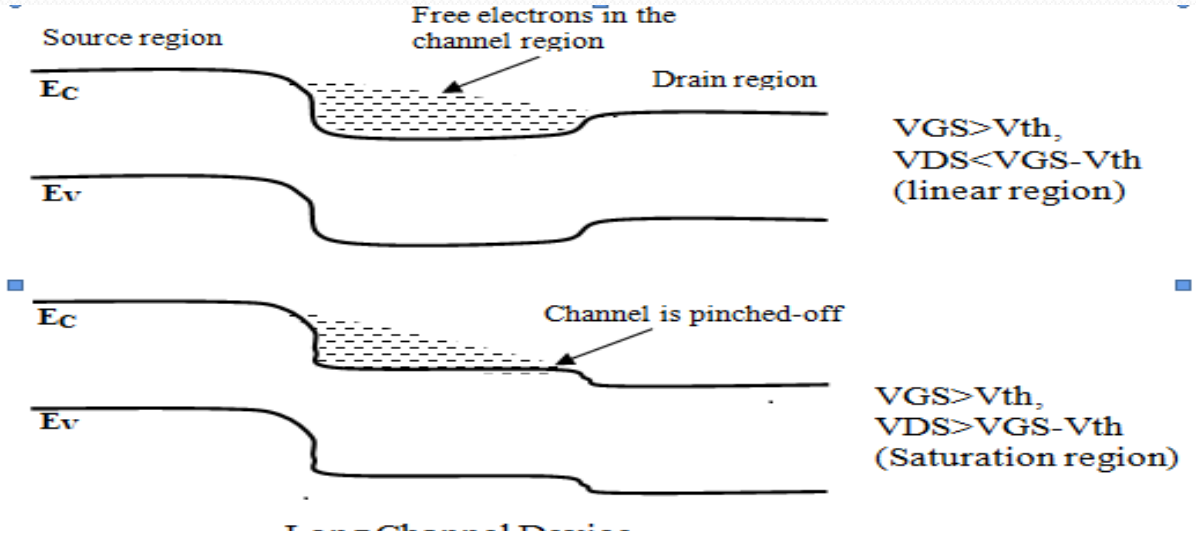
- **The major short-channel effects are:**

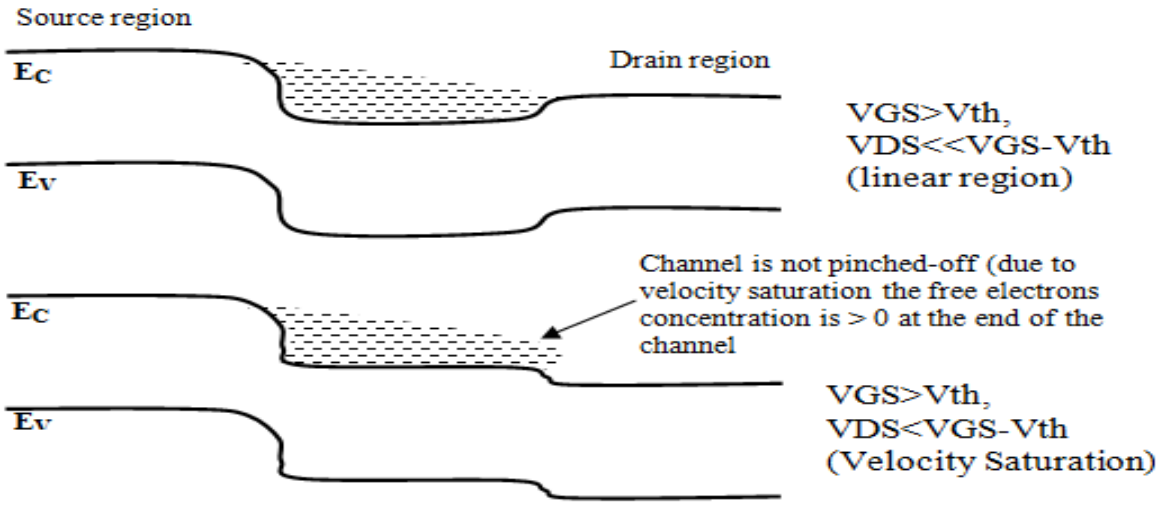
1) **Carrier velocity saturation:**

The mobility determines the increase in carrier velocity as the electric field increases. But this does not continue forever.

As the channel length is reduced, the horizontal electric field between the source and drain increases to a point where the carrier mobility becomes zero, i.e. the carrier velocity won't increase beyond a certain limit (hence the term velocity saturation).

Once velocity saturation sets in, the drain current won't increase as V_{DS} increases even if V_{DS} is still < $V_{GS} - V_{th}$. The figure below shows the effect of velocity saturation using energy band diagram of an NMOS transistor.

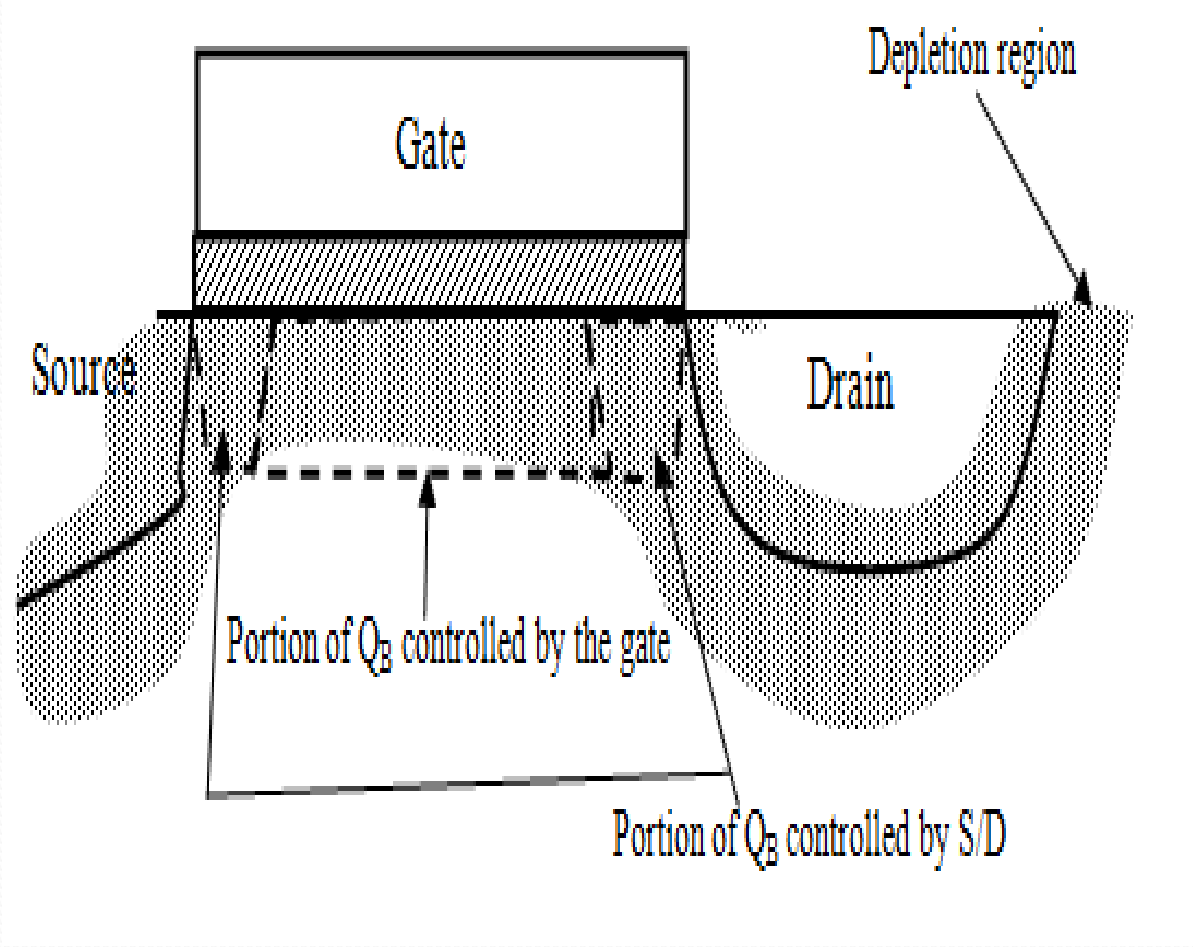




Short Channel Device

2)Threshold voltage reduction:

Since for the short-channel devices, a relatively large portion of the channel depletion charge Q_B is controlled by the drain, a smaller amount of gate voltage is required to achieve strong inversion and create the channel, i.e. the threshold voltage will be smaller. The following figure illustrates this.


Depletion region
Gate
Source
Drain
Portion of Q_B controlled by the gate
Portion of Q_B controlled by S/D

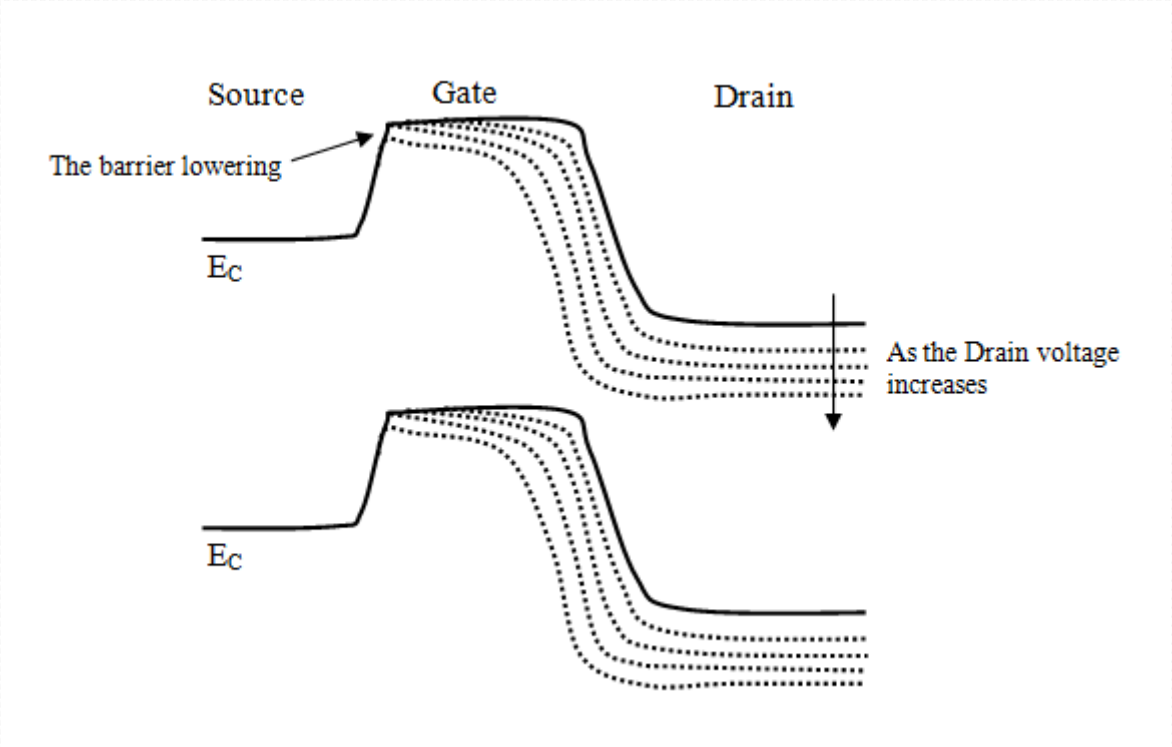
3) Drain-Induced-Barrier-Lowering (DIBL):

Because of the short channel length, the drain voltage will reduce the potential barrier between the source and substrate.

Hence it become easier for carriers at the source to jump over this barrier and drift to the drain even at the absence of a channel (i.e. when the gate to source voltage is less than the threshold voltage). The DIBL effect causes the MOS devices to “leak” currents when they are turned-off.

These currents are called leakage currents or off-currents. If the DIBL effect increases significantly, the MOS switch may become always ON, i.e. it is not controlled by the gate any more.

This condition would occur if the source and drain depletion regions reach one another, a condition called punch-through. Increasing the substrate doping under the gate/channel reduces the DIBL effect. The figure below demonstrates the DIBL effect using energy band diagram.


Source
Gate
Drain
The barrier lowering
E_C
As the Drain voltage increases
E_C

- **Hot Electron Injection into the Gate Oxide:**

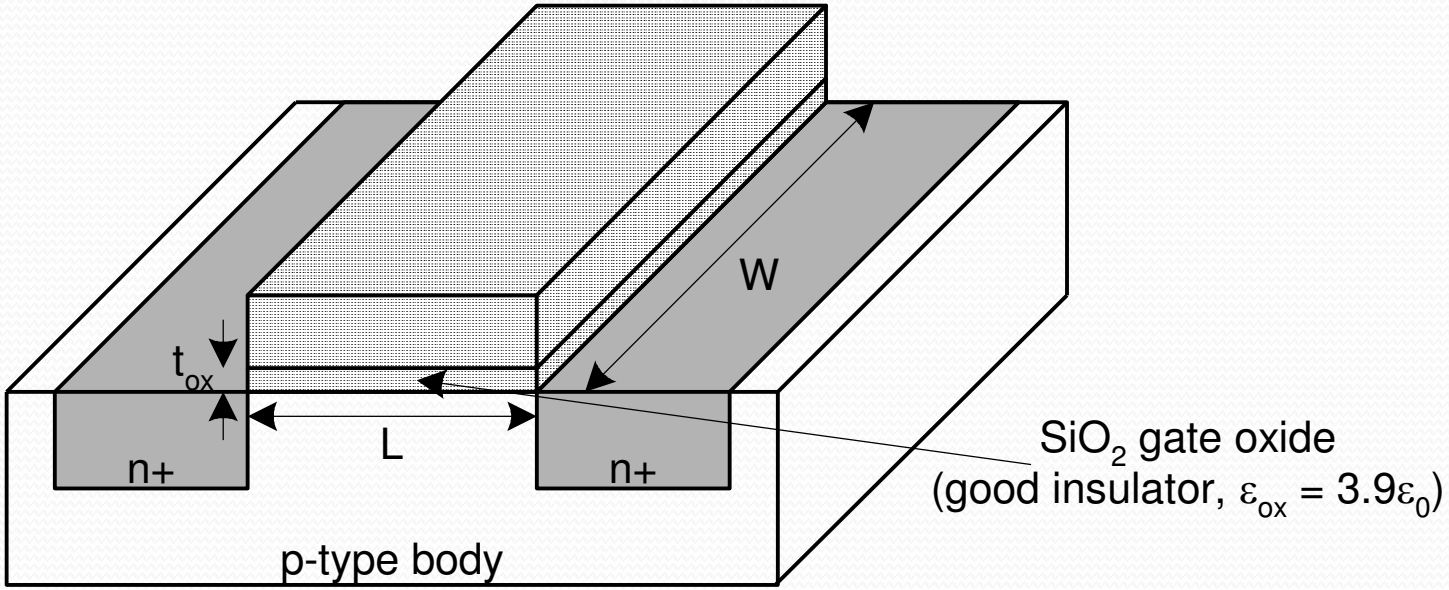
  Due to the high electric fields that result in short-channel devices, electrons can attain enough speed to jump over the energy barrier between the Si channel region and the gate oxide. Once they are injected into the oxide, they become part of its trapped charge. As it is already known, this would alter the device's threshold voltage and may render the device totally useless with time. Also, since the injection occurs near the drain side, where the electrons would have attained the highest speed, the device operation becomes asymmetrical, i.e. the I-V characteristics will exhibit direction dependency. This is because the threshold if the current is flowing from the source to the drain (forward direction) is smaller than the threshold in the reverse direction. The figure below shows the effects of HE injection on the I-V characteristics of an NMOS device.

MOS Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
  - Creates channel charge necessary for operation
- Source and drain have capacitance to body
  - Across reverse-biased diodes
  - Called diffusion capacitance because it is associated with source/drain diffusion

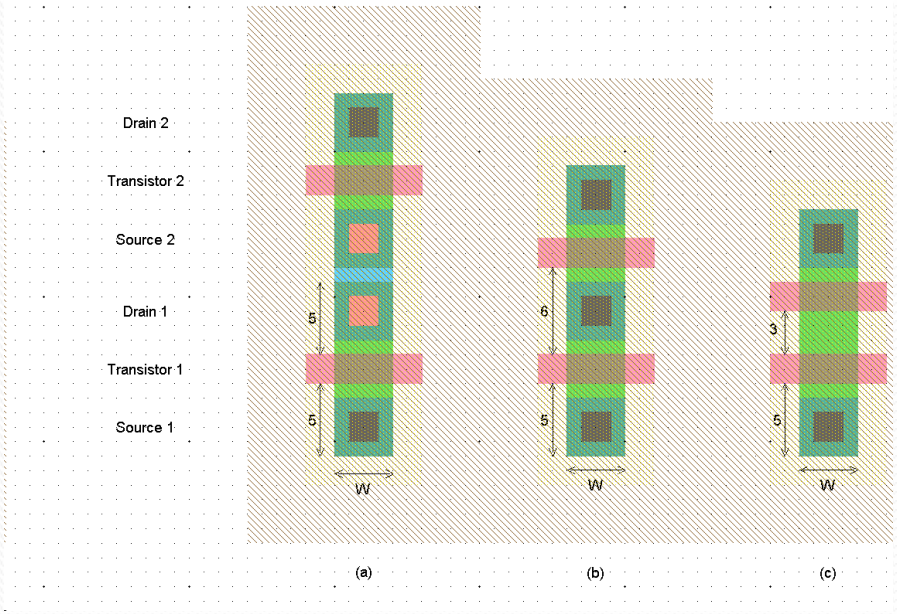
Gate Capacitance

- Approximate channel as connected to source
- $C_{gs} = \varepsilon_{ox} WL/t_{ox} = C_{ox} WL = C_{permicron} W$ (minimum L)
- $C_{permicron}$ is typically about 2 fF/μm



Diffusion Capacitance

- C_{sb}, C_{db}
- Undesirable, called *parasitic* capacitance
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to C_g for contacted diff
  - ½ C_g for uncontacted
  - Varies with process



Electric Fields Effects

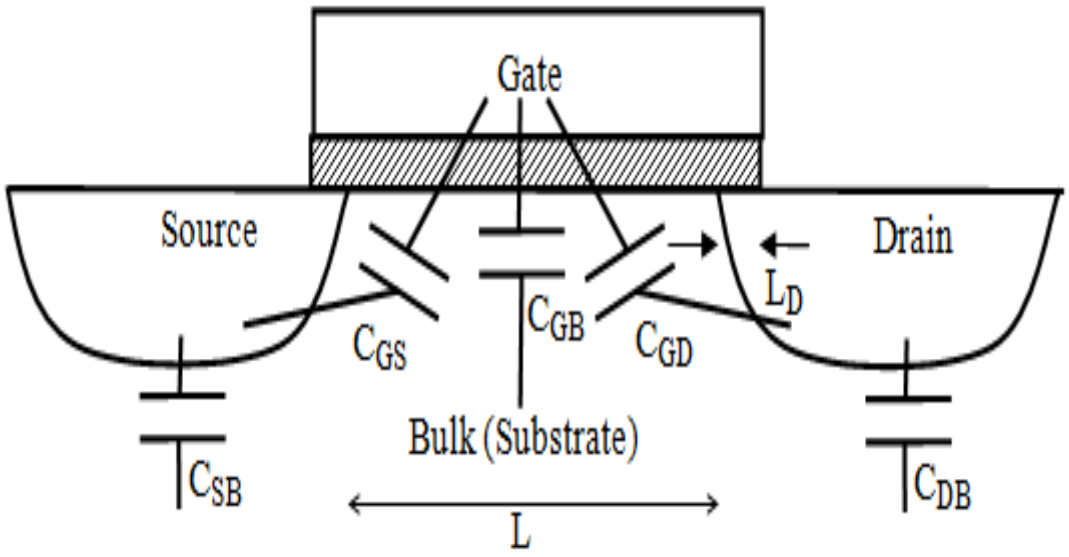
- Vertical electric field: $E_{vert} = V_{gs} / t_{ox}$
  - Attracts carriers into channel
  - Long channel: $Q_{channel} \propto E_{vert}$
- Lateral electric field: $E_{lat} = V_{ds} / L$
  - Accelerates carriers from drain to source
  - Long channel: $v = \mu E_{lat}$

MOS Capacitances

- **Gate Capacitances**
- The channel charge is lumped (modeled) into three capacitances, Gate-to-Bulk capacitance (C_{GB}), Gate-to-Source capacitance (C_{GS}), and Gate-to-Drain capacitance (C_{GD}).
- These capacitances are not constant; their values depend on the region of operation.

- **S/D Junction Capacitances**
- The source and drain forms diodes with the bulk. As seen before these diodes will have junction capacitances that are dependent on the voltage difference between their terminals. Hence, as the source or drain voltages change, these capacitances will be charged or discharged.
- These capacitances are named as C_{SB} and C_{DB}

MOS Capacitances (Contd.)



The MOS Capacitances