Other SC Effects:

Sub-threshold Conduction:

- Long channel devices: If $V_{GS} < V_{T0}$ channel barrier potential prevents I_D circulation. Depends only on V_{GS}.
- Short channel devices: Channel barrier depends on both V_{GS} & V_{DS}

Increasing V_{DS} → Reduces Channel barrier
"Drain-induced barrier lowering (DIBL)" } ⇒ Subthreshold Conduction

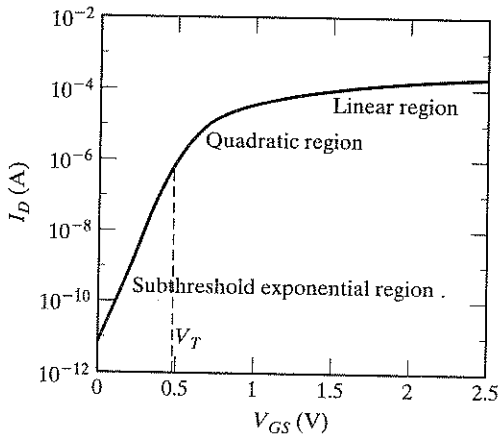
$$I_{D_{sth}} = I_S e^{\frac{V_{GS}}{n\Phi_T}} \left(1 - e^{-\frac{V_{DS}}{\Phi_T}}\right)$$

where:

$$I_S = \frac{q D_n W x_c n_0}{L_B}, \text{ and } \Phi_T = \frac{kT}{q} \quad (n \approx 1.5)$$

x_c = Subthreshold channel depth
D_n = Electron diffusion coefficient
L_B = Length of barrier region

This equation resembles I_C in a BJT; The source-channel-drain is modeled as a BJT to obtain I_{Dsth}. Note the resemblance to a BJT when there is no channel.



I_D current versus V_{GS} (on logarithmic scale), showing the exponential characteristic of the subthreshold region.

- Some authors prefer to write

$$I_{D_{sth}} = I_{D0}\left(\frac{W}{L}\right) e^{\frac{-(V_{GS}-V_T)}{n\Phi_T}}$$

which denotes the dependence of I_{Dsth} on $1/L$ and V_T

- A common measure of device quality related to subthreshold conduction is the Inverse subthreshold slope (S)

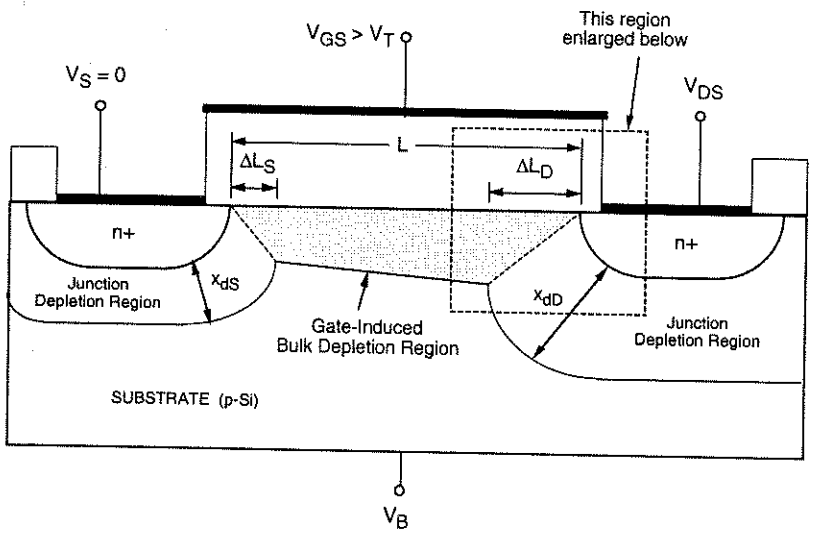
S is just the rate at which I_{Dsth} decays with the reduction of V_{GS}

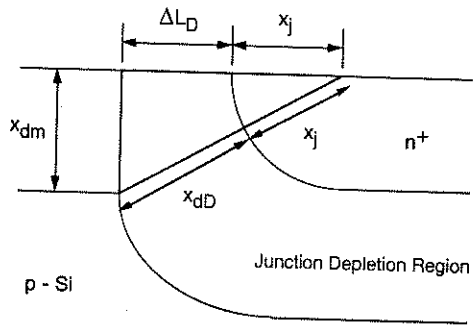
$$S = n\left(\frac{kT}{q}\right) \cdot \ln(10) \approx 60\, n \ \frac{mV}{decade} \ @ 300°K$$

Quality criteria: $I_D\big|_{V_{GS}=0} \leq \varepsilon\, I_D\big|_{V_{GS}=V_T}$ $\quad \varepsilon \approx 10^{-5} \Rightarrow$ Limit in lowest V_T

Threshold Variation

- V_T equation based on long-channel device
  - Channel depletion dependent only on V_{GS}
  - Neglects depletion charge induced by V_{DS}
- In short channel transistors the effect of V_{DS} becomes relevant



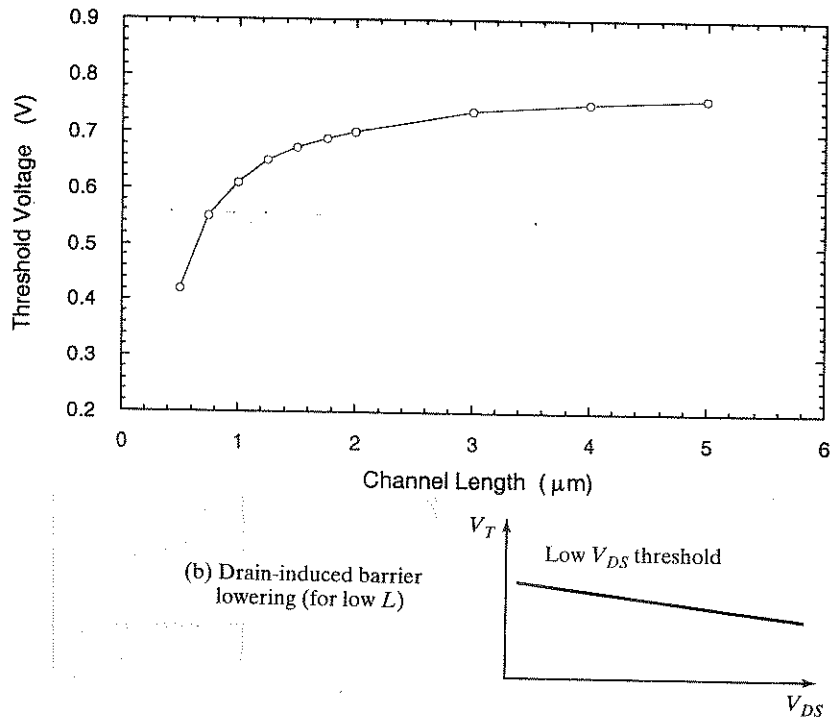


- Under this model, the amount of depletion charge in the channel due to V_{GS} is smaller than that for long-channel model. Thus

$$V_{TO(short)} = V_{TO} - \Delta V_{TO}$$

where

$$\Delta V_{TO} = \frac{1}{C_{ox}} \sqrt{2q\epsilon_{si} N_A |2\phi_F|} \; \frac{x_j}{2L} \left[\sqrt{1 + \frac{2X_{dS}}{X_j}} - 1 + \sqrt{1 + \frac{2X_{dD}}{X_j}} - 1 \right]$$



- ΔV_{TO} variations can be as large as 50% V_{TO} for small L
- Note that $X_{dD} = f(V_{DS})$ (source of DIBL)

$$X_{dD} = \sqrt{\frac{2\epsilon_{si}}{q N_A} (\phi_o + V_{DS})}$$

where ϕ_o is the built-in voltage of the drain-bulk p-n junction

- Net: $V_{TO} = f(V_{DS})$

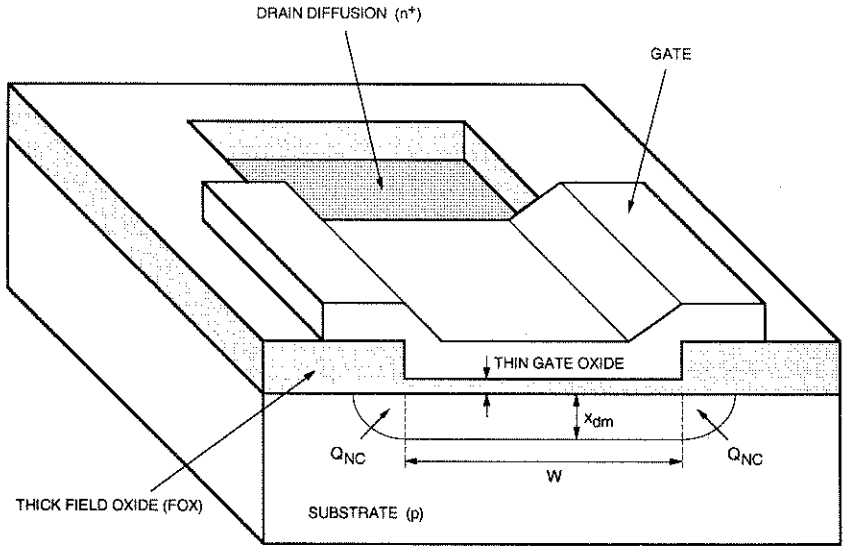
(b) Drain-induced barrier lowering (for low L)

Narrow-Channel Effects

- Arise when W is comparable to X_{dm} (maximum depth of depletion region thickness)
- Induces threshold voltage variations:

$$V_{TO_{(narrow)}} = U_{TO} + \Delta V_{TO}$$

- ΔV_{TO} attributed to overlap of gate electrode on field oxide (FOX)



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

$$\Delta V_{TO} = \frac{1}{C_{ox}} \sqrt{2 q \varepsilon_{si} N_A |2\phi_F|} \cdot \frac{K \cdot X_{dm}}{W}$$

where K is an empirical parameter depending of fringe depletion shape. Assuming quarter circle

$$K = \frac{\pi}{2}$$

- Note that for short, narrow channel devices, threshold variations tend to cancel-out
- However, transistor sizing for specs compliance tend to minimize narrow V_T variation, accentuating short channel induced variations.

Other Effects:

1) Punch-through:
- Large V_{DS} causes drain depletion region to reach source depletion region.
- V_{GS} does not control I_D anymore
- I_{DS} rises sharply and induces transistor melt.

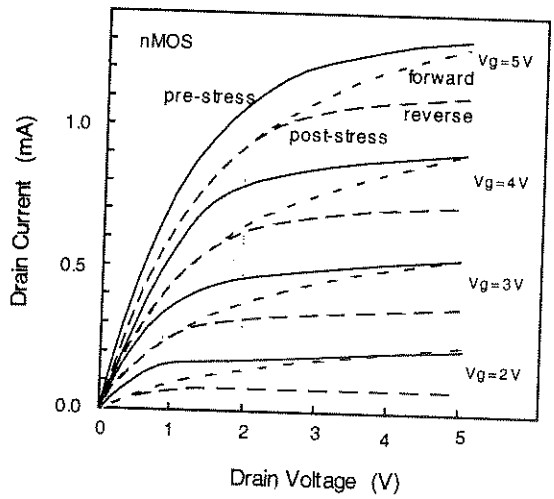
2) Oxide breakdown: Loss of insulating property when the electric field exceeds disruptive level of oxide

3) Oxide pinholes: Caused by oxide defects where uniformity is lost due to small t_{ox}

4) Hot carrier effect: Increase of electric fields in channel region yield high energy carriers (HOT)

Hot carriers penetrate the oxide (upon collisions) permanently changing the oxide characteristics.

- This imposes limits to small device dimensions under conventional fabrication.
- More accentuated in NMOS devices due to electron mobility
- Effect leads to early device aging: Increase V_T, reducing I_D



Stress conditions :

$V_g = 3$ V

$V_d = 8$ V

stress time = 14 h

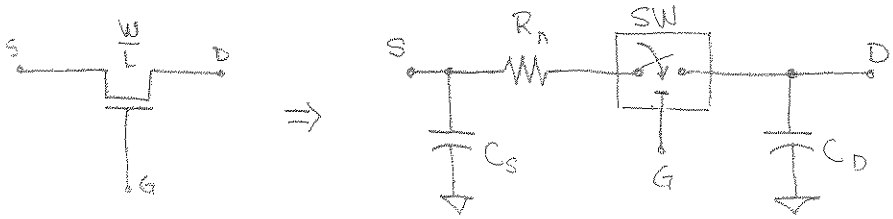
Typical drain current vs. drain voltage characteristics of an n-channel MOS transistor before and after hot-carrier induced oxide damage.

- One of the most compelling reasons to keep V_{DS} small in deep submicron devices.

5) Electromigration: Damage in metal traces caused by mass movement induced by high current densities. Induces voids & hillocks

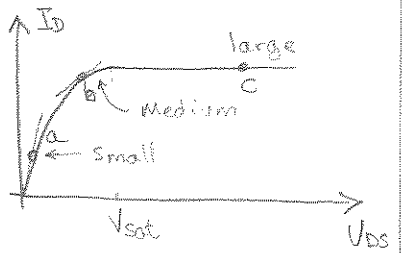
FET R-C Model

Simplest approximation uses a linear model



R: Piece-linear Resistor
SW: Ideal switch

C_S & C_D: Terminal capacitances



MOSFET RESISTANCE

$R_n = V_{DS}/I_D$: In linear region, for small V_{DS} we can neglect $V_{DS}^2/2 \approx 0$

$$I_{Dn} \approx K_n(V_{GS}-V_T)V_{DS} \Rightarrow R_{n_{linear}} \approx \frac{1}{K_n(V_{GS}-V_T)} \quad \text{(Point a)}$$

At point b, $V_{DS}^2/2$ might not be too small. So, a better estimate

$$R_{n_{lin}} = \frac{2}{K_n[2(V_{GS}-V_T)-V_{DS}]} \leftarrow \text{(Point b)}$$

An approximation is $\boxed{R_n = \frac{1}{K_n(V_{DD}-V_T)}}$ ← Understimates R_n by making $V_{GS} = V_{DD}$ Maximum value

This model neglects channel-length modulation and assumes

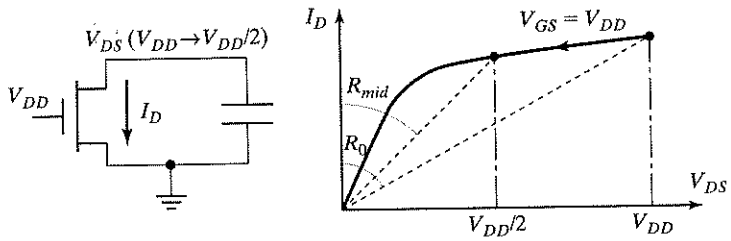
$R_{n_{sat}} = \infty$ ← In reality, $R_{n_{sat}} < \infty$ (Finite) $R_{n_{sat}} \approx \frac{1}{\lambda I_D}$

Although inaccurate, the RC model is very useful for the analysis of complex structures (complex gates and interconnect.)

A better estimate $R_n = \text{Average}(R_n)\Big|_{t_1}^{t_2} = \frac{1}{t_2-t_1}\int_{t_1}^{t_2} R_n(t)\,dt = \frac{1}{t_2-t_1}\int_{t_1}^{t_2}\frac{V_{DS}(t)}{I_D(t)}\,dt$

Consider a velocity saturated, short-channel NMOS discharging a capacitor from V_{DD} to $V_{DD}/2$.

Assuming $V_{DSAT} < V_{DD}/2$



(a) schematic (b) trajectory traversed on ID-VDS curve.

Discharging a capacitor through an NMOS transistor: Schematic (a) and *I-V* trajectory (b). The instantaneous resistance of the transistor equals (V_{DS}/I_D) and is visualized by the angle with respect to the *y*-axis.

$$R_n = \frac{1}{-V_{DD}/2} \int_{V_{DD}}^{V_{DD}/2} \frac{V}{I_{DSAT}(1+\lambda V)} dV \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{7}{9}\lambda V_{DD}\right)$$

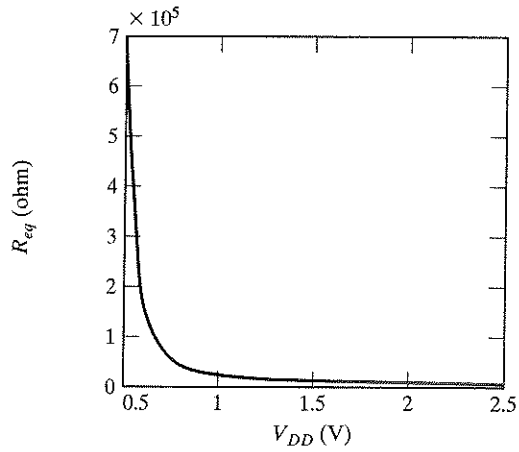
Making an approximation $R_n = \frac{1}{2}\left(R_n(t_1) - R_n(t_2)\right) \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{5}{6}\lambda V_{DD}\right)$

Nearly the same result

This formulation uses $I_{DSAT} = K_n\left[(V_{DD} - V_{DSAT})V_{DSAT} - \frac{V_{DSAT}^2}{2}\right]$

Notice that in every model:

- $R_n \propto \frac{1}{K_n} \propto \frac{1}{W/L}$
- For large V_{DD} R_n is controlled only by λ
- Small value of R_n for V_{DD} around V_T



Simulated equivalent resistance of a minimum size NMOS transistor in 0.25 μm CMOS process as a function of V_{DD} ($V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$).

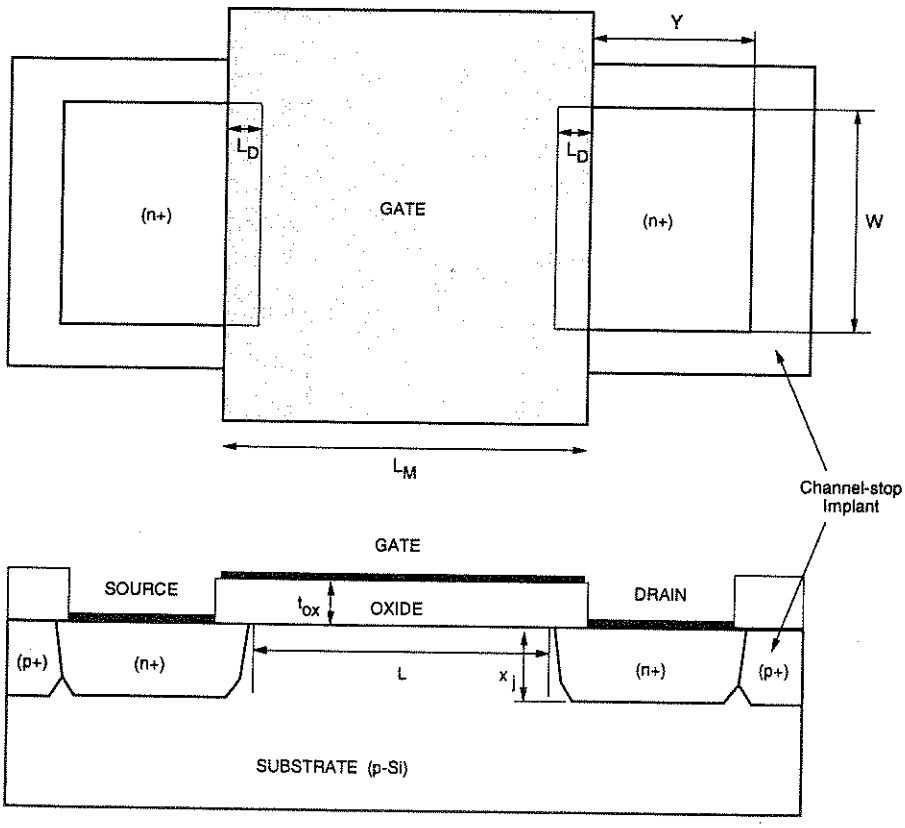
MOSFET CAPACITANCE

MOS capacitances are in general a) Non-linear (voltage dependance)
b) Distributed over the device, rather than lumped, and
c) Three-dimensional (depend on W, L, x_j)

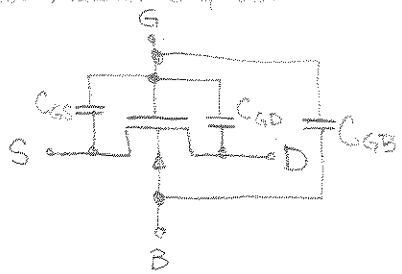
- Capacitance models for manual analysis use lumped approximations. Better models are available through SPICE
- Total capacitance includes gate and source-drain components.

Gate Capacitance:

The gate electrode, due to the lateral diffussion, overlaps with the source and drain regions



The equivalent capacitances resulting can be represented as



The overlap generated capacitances are constant and independent from voltage. Assumin the overlap is the same in the source & drain regions:

$$C_{GS_{ovl}} = C_{GD_{ovl}} = C_{ox} \cdot W \cdot L_D$$

In addition to the overlap, the electrical interaction between gate and source - channel - drain induces additional capacitance components.

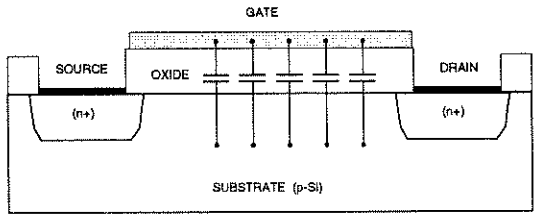
$$C_{G_{volt}} = C_{GS_{volt}} + C_{GD_{volt}} + C_{GB_{volt}}$$

The values of the voltage induced components depend on the state of the channel. (voltage-dependent components)

Cut-off:

- Source and drain are electrically isolated (no inversion)
- Only gate-bulk region contributes to C_{Gvolt}

$$C_{GB_{volt}} = C_{ox} \cdot (W \times L)$$

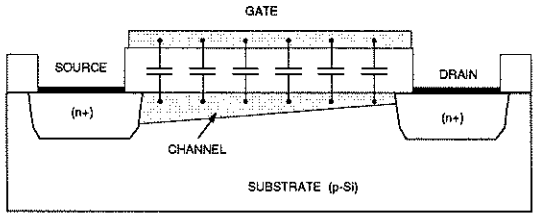


cut-off channel

Linear:

- Inversion layer shields bulk, cancelling $C_{GB_{volt}}$. Only the drain and source regions contribute to C_{Gvolt}

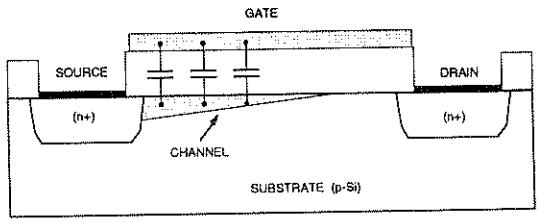
$$C_{GS_{volt}} \approx C_{GD_{volt}} \approx \frac{1}{2} C_{ox} \cdot (W \times L)$$



Linear channel

Saturation:

- Inversion layer covers only part of the channel
- Drain is isolated
- Source linked to conducting channel
- $C_{G_{volt}} = C_{GS} \approx \frac{2}{3} C_{ox} \cdot (W \times L)$



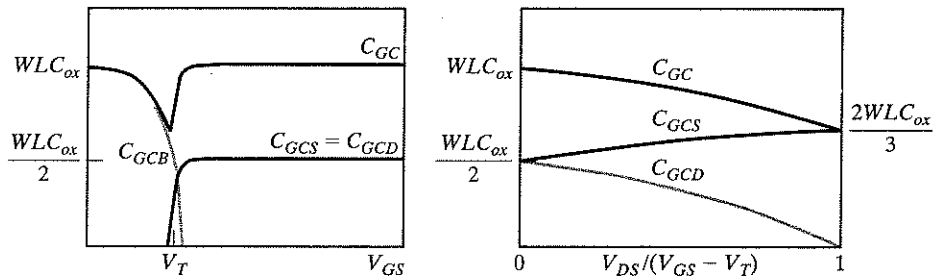
Saturated channel

Considering $C_G = C_{G_{out}} + C_{G_{volt}}$, then its behavior is summarized as:

Capacitance	Cut-off	Linear	Saturation
C_{gb} (total)	$C_{ox}WL$	0	0
C_{gd} (total)	$C_{ox}WL_D$	$\frac{1}{2}C_{ox}WL + C_{ox}WL_D$	$C_{ox}WL_D$
C_{gs} (total)	$C_{ox}WL_D$	$\frac{1}{2}C_{ox}WL + C_{ox}WL_D$	$\frac{2}{3}C_{ox}WL + C_{ox}WL_D$

C_G(total) $C_{ox}(WL + 2WL_D)$ $C_{ox}(WL + 2WL_D)$ $C_{ox}(\frac{2}{3}WL + 2WL_D)$

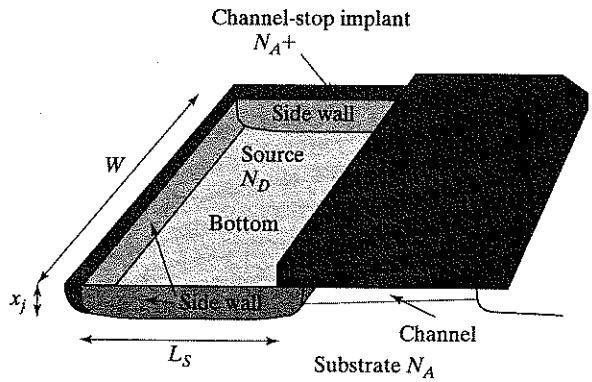
The evolution of C_G as a function of $V_{GS}|_{V_{DS}=0}$ and $V_{DS}|_{V_{GS}>V_T}$ evidences the non-linear nature of C_G.



(a) C_{GC} as a function of V_{GS} (with $V_{DS} = 0$) (b) C_{GC} as a function of the degree of saturation

Source-Drain Capacitances (Junction capacitances)

- Caused by the reverse-biased parasitic p-n junctions between source- and drain-bulk regions.



Detailed view of source junction.

$N_A^+ \approx 10 N_A$

- Junctions are formed with the:

Bottom - Bottom substrate

Side Wall - Channel stop implant

- Side wall capacitance depends on diffusion depth x_j and channel stop doping
- Bottom capacitance depends on bottom area & substrate/diff. doping

- Equations for manual analysis:

$C_{DB} = C_{SB} = C_{Bottom} + C_{sw}$

$C_{Bottom} = A_{Bottom} \cdot C_{j0} \cdot K_{eq}$

where

$$C_{j0} = \sqrt{\frac{\varepsilon_{Si}\, q}{2} \frac{N_A \cdot N_D}{N_A + N_D} \frac{1}{\Phi_0}}$$ ← Zero-bias junction capacitance

$$K_{eq} = \frac{\Phi_0}{(V_2 - V_1)(1-m)} \left[\left(1 - \frac{V_2}{\Phi_0}\right)^{1-m} - \left(1 - \frac{V_1}{\Phi_0}\right)^{1-m} \right]$$ ← Voltage equivalence factor

m = Junction Grading Coefficient, $\Phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right)$, (V_1, V_2) ← limits for bias voltage

$A_{bottom} = W \times L_S$ ← Bottom plate area

For the particular case when $m = 1/2$ (abrupt pn junction)

$$K_{eq} = -\frac{2\sqrt{\Phi_0}}{V_2 - V_1}\left(\sqrt{\Phi_0 - V_2} - \sqrt{\Phi_0 - V_1}\right) \qquad (0 < K_{eq} \leq 1)$$

$K_{eq} = 1$ ← Ignores voltage dependence

$$C_{sw} = A_{sw} \cdot C_{j0sw} \cdot K_{eq(sw)}$$

Since the channel stop implant has a doping concentration higher than the bottom plate C_{j0} & K_{eq} will be differet

$$C_{j0sw} = \sqrt{\frac{\varepsilon_{Si} \cdot q}{2}\left(\frac{N_{A(sw)} \cdot N_D}{N_{A(sw)} + N_D}\right)\frac{1}{\Phi_{0\,sw}}}$$

and

$$K_{eq_{(sw)}} = -\frac{2\sqrt{\Phi_{0sw}}}{V_2 - V_1}\left(\sqrt{\Phi_{0sw} - V_2} - \sqrt{\Phi_{0sw} - V_1}\right)$$

$A_{sw} = (2L_s + W)X_j$ ← No capacitance on the channel side

X_j is constant for a given process. Thus, a zero-bias sidewall junction capacitance is usually defined as

$$C_{jsw} = C_{j0_{sw}} \times X_j$$

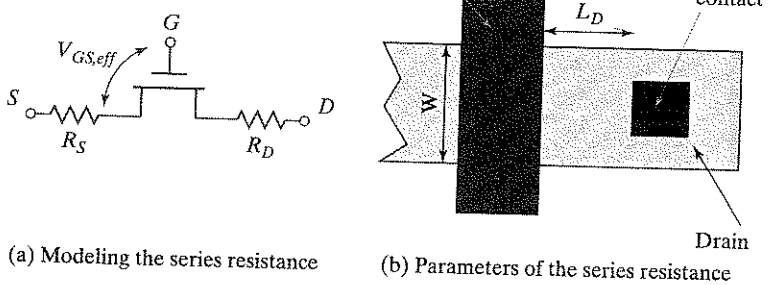
Therefore:

$$C_{sw} = P_{sw} \cdot C_{jsw} \cdot K_{eq(sw)}$$

where $P_{sw} = (2L_s + W)$ ← Sidewall perimeter

SOURCE-DRAIN RESISTANCES: Parasitic resistances of source & drain regions.

$$R_{s(D)} = \frac{L_{s(D)}}{W} R_{\square} + R_C$$



(a) Modeling the series resistance (b) Parameters of the series resistance

Series drain and source resistance.

$R_{\square}$ ← Sheet resistance

R_C ← Contact resistance

These resistances are usually neglected in most hand calculation models.