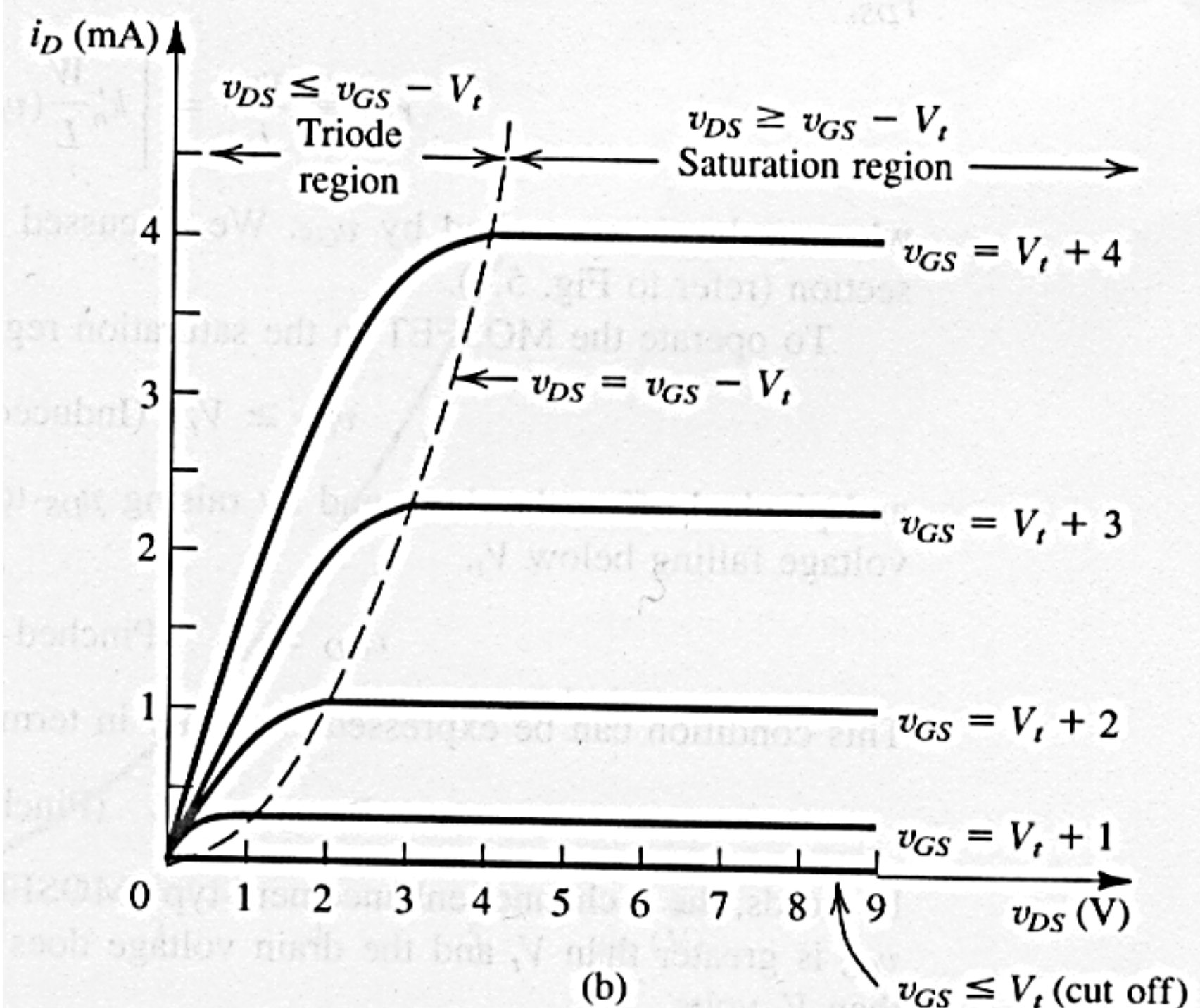

i_D (mA)
$v_{DS} \leq v_{GS} - V_t$
Triode region
$v_{DS} \geq v_{GS} - V_t$
Saturation region
$v_{DS} = v_{GS} - V_t$
$v_{GS} = V_t + 4$
$v_{GS} = V_t + 3$
$v_{GS} = V_t + 2$
$v_{GS} = V_t + 1$
$v_{GS} \leq V_t$ (cut off)
0
1
2
3
4
5
6
7
8
9
v_{DS} (V)

(b)

Voltage-Current Relation: Linear Mode

For long-channel devices (L > 0.25 micron)

- When $V_{DS} \leq V_{GS} - V_T$

$$I_D = k'_n \, W/L \, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$$

where

$k'_n = \mu_n C_{ox} = \mu_n \varepsilon_{ox}/t_{ox}$ = is the process transconductance parameter (μ_n is the carrier mobility (m^2/Vsec))

$k_n = k'_n \, W/L$ is the gain factor of the device

For small V_{DS}, there is a linear dependence between V_{DS} and I_D, hence the name resistive or linear region

Voltage-Current Relation: Saturation Mode

For long channel devices

- When $V_{DS} \geq V_{GS} - V_T$, replacing V_{DS} by V_{GS}-V_T in I_D expression found previously:

$$I_D' = k'_n/2\ W/L\ [(V_{GS} - V_T)^2]$$

since the voltage difference over the induced channel (from the pinch-off point to the source) remains fixed at $V_{GS} - V_T$ and I_D is no longer a function of V_{DS}

- However, the effective length of the conductive channel is modulated by the applied V_{DS}, so

$$I_D = I_D' (1 + \lambda V_{DS})$$

where λ is the channel-length modulation (varies with the inverse of the channel length)

• The transistor in saturation mode acts as a perfect current source, since the current between drain and source is constant and independent of V_{DS}. However, this is not entirely correct because the effective channel length is actually modulated by V_{DS}. (increasing V_{DS} reduces the effective length of the conducting channel). The current increases as L is decreased (L is at the denaminator).

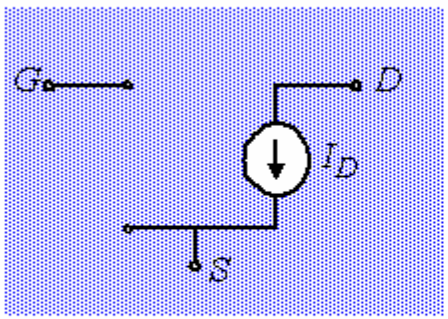
• λ is inversly proportional to the channel length. In shorter transistors, the drain-junction depletion region presents a larger fraction of the channel and the channel-length modulation effect is more pronounced.

• It is therefore advisable to choose long-channel transistors if a high empedance current source is needed.

Current Determinates

- For a fixed V_{DS} and V_{GS} (> V_T), I_{DS} is a function of
  - the distance between the source and drain – L
  - the channel width – W
  - the threshold voltage – V_T
  - the thickness of the SiO_2 – t_{ox}
  - the dielectric of the gate insulator (SiO_2) – ε_{ox}
  - the carrier mobility
    - for nfets: $\mu_n = 500\ cm^2/V\text{-}sec$
    - for pfets: $\mu_p = 180\ cm^2/V\text{-}sec$

A model for manual analysis



$$V_{DS} > V_{GS} - V_T$$

$$I_D = \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS})$$

$$V_{DS} < V_{GS} - V_T$$

$$I_D = k'_n \frac{W}{L}\left((V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right)$$

with

$$V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$$

Velocity Saturation

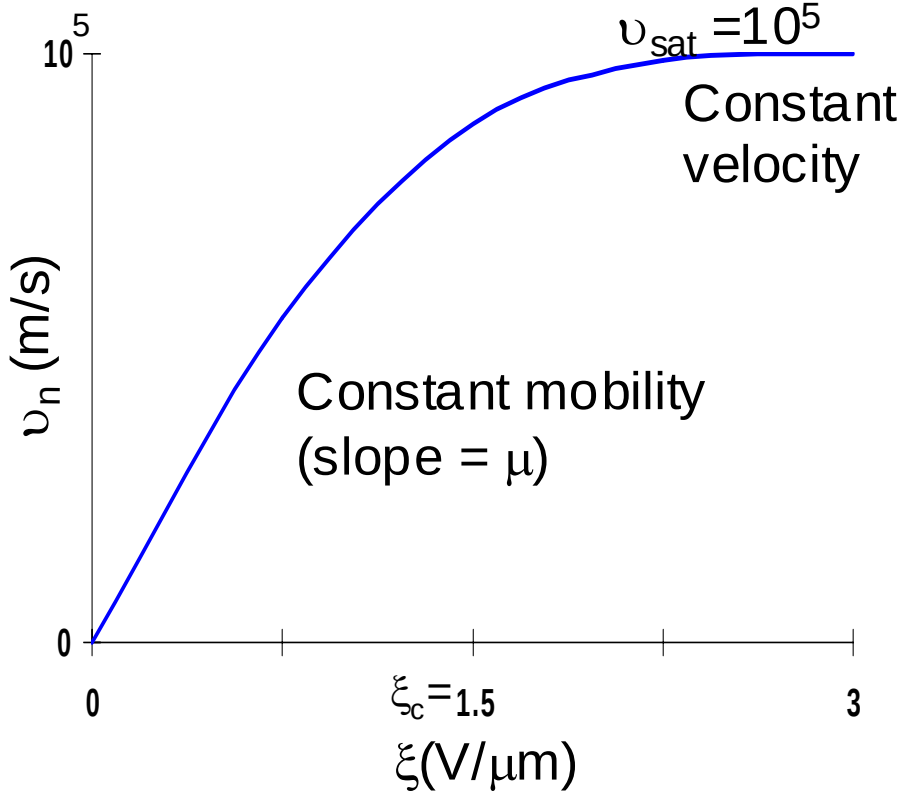
The behavior of transistor with very short channels deviates considerably from the resistive and saturated models. We stated previously that:

$$v_n = \mu_n \xi(x)$$

Where the carrier mobility is constant. However at high electric field strengths (horizontal) when the strength reaches to a critical value ξ_c the velocity of the carriers tend to saturate due to collisions. This condition is easily met in short-channel devices.

Short Channel Effects

- Behavior of short channel device mainly due to



- Velocity saturation – the velocity of the carriers saturates due to scattering (collisions suffered by the carriers)

- For an NMOS device with L of .25μm, only a couple of volts difference between D and S are needed to reach velocity saturation. Velocity saturation effects are less pronounced in PMOS transistors.

A rough approximation for the velocity curve:

$$v = \frac{\mu_n \xi}{1 + \xi / \xi_c} \quad \text{for } \xi \le \xi_c$$

$$v = v_{sat} \quad \text{for } \xi \ge \xi_c$$

continuity requirement of the curve at $v = v_{sat}$ $(\xi = \xi_c)$ point dictates that: $\xi_c = 2v_{sat} / \mu_n$

Reevaluating the I_D equation previously found, with the new velocity formula:

Voltage-Current Relation: Velocity Saturation

For short channel devices

- Linear: When $V_{DS} \leq V_{GS} - V_T$

$$I_D = \kappa(V_{DS})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$$

where $k'_n = \mu_n C_{ox}$ and

$\kappa(V) = 1/(1 + (V/\xi_c L))$ is a measure of the degree of velocity saturation

- Saturation: When $V_{DS} = V_{DSAT} \geq V_{GS} - V_T$

$$I_{DSAT} = \kappa(V_{DSAT})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DSAT} - V_{DSAT}^2/2]$$

in case of long-channel devices (large L) or small values of V_{DS}, κ approaches to 1. For short channel devices κ is less than 1, delivering less I_D than normally expected.

Recalling the dq(x) expression derived for the resistive operation earlier:

$$dq(x) = -C_{ox}Wdx\left[V_{GS} - V(x) - V_T\right]$$

since, I(x) = dq(x)/dt and v_n(x) = -dx/dt

$$I(x) = C_{ox}Wv_n(x)\left[V_{GS} - V(x) - V_T\right]$$

using the expressions for saturation:

$$I_{DSAT} = C_{ox}Wv_{sat}(x)\left[V_{GS} - V_{DSAT} - V_T\right]$$

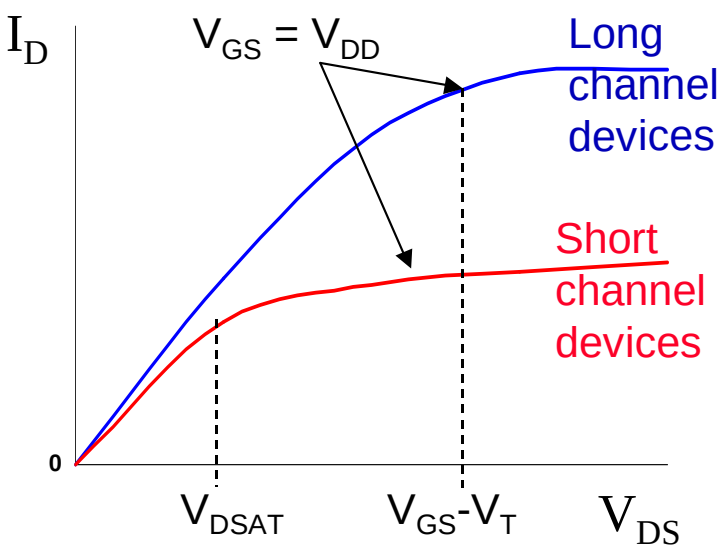
Equating this I_{DSAT} expression to the previous one and after rearranging and cancelling, we find:

$$V_{DSAT} = \kappa(V_{GT})V_{GT}$$

Where, $V_{GT} = V_{GS} - V_T$

For a short channel device and for large enough values of V_{GT}, $\kappa(V_{GT})$ is substantially less than 1 and thus $V_{DSAT} < V_{GT}$. The device enters saturation before V_{DS} reaches V_{GS}-V_T.

Velocity Saturation Effects



For short channel devices and large enough $V_{GS} - V_T$

- $V_{DSAT} < V_{GS} - V_T$ so the device enters saturation before V_{DS} reaches $V_{GS} - V_T$ and operates more often in saturation

- I_{DSAT} has a linear dependence wrt V_{GS} (in contrast with the squared dependence of long-channel device) so a reduced amount of current is delivered for a given control voltage

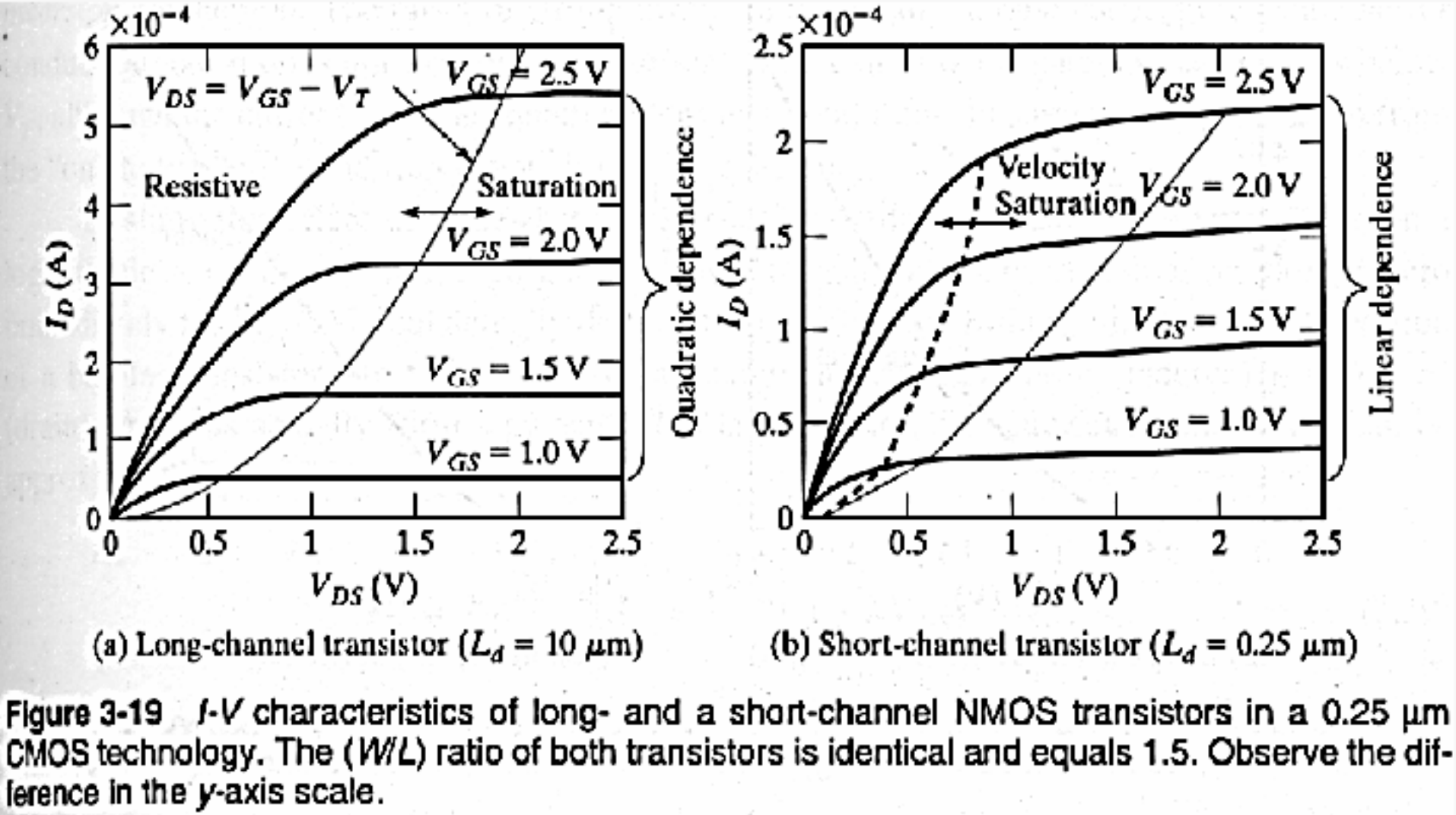
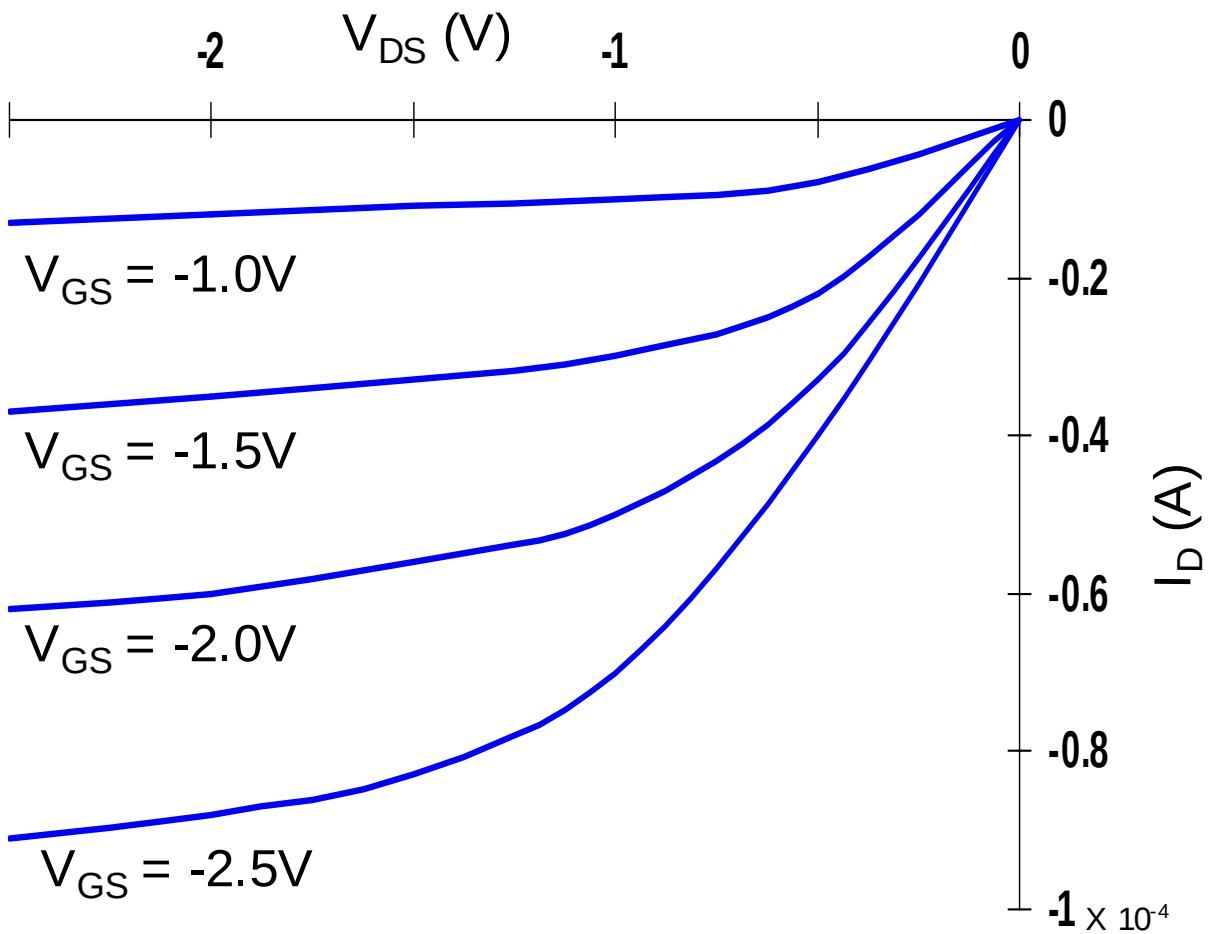


Figure 3-19 *I-V* characteristics of long- and a short-channel NMOS transistors in a 0.25 μm CMOS technology. The (*W*/*L*) ratio of both transistors is identical and equals 1.5. Observe the difference in the *y*-axis scale.

For $V_{GS} = V_{DS} = 2.5V$, the drain current of the short transistor is only 40% of the long one. (220μA versus 540μA)

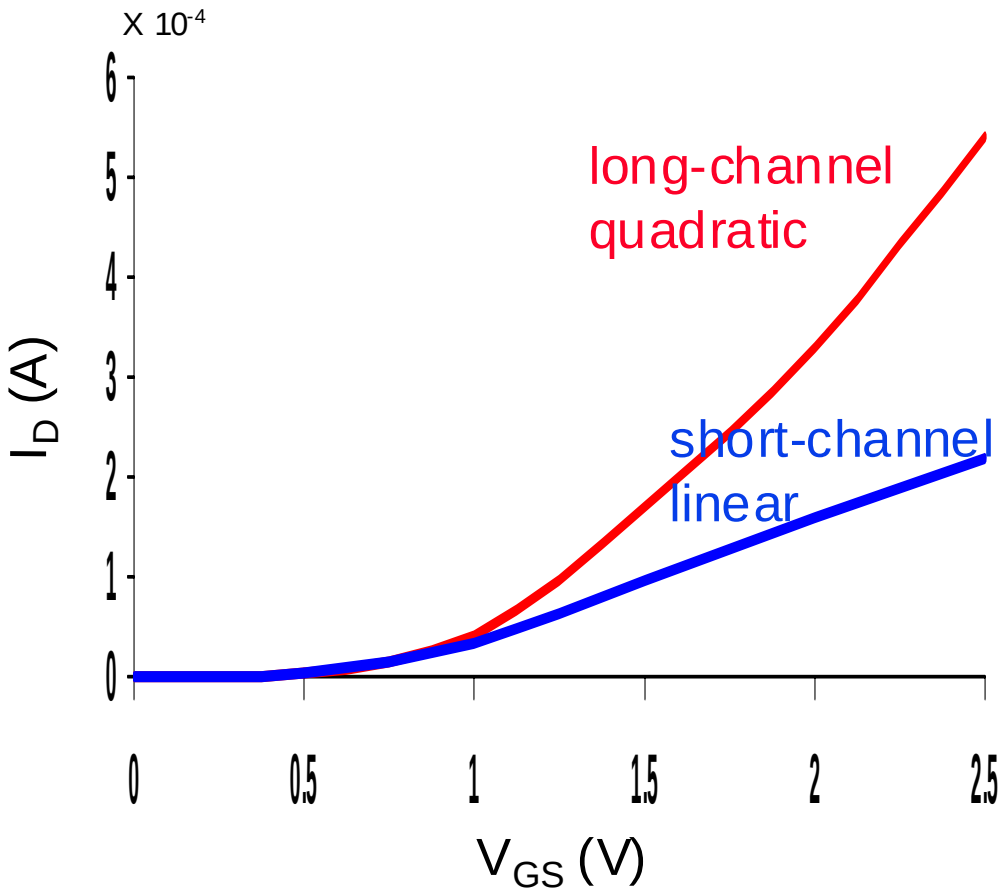
Short Channel I-V Plot (PMOS)

- All polarities of all voltages and currents are reversed



PMOS transistor, 0.25um, L_d = 0.25um, W/L = 1.5, V_{DD} = 2.5V, V_T = -0.4V

MOS I_D-V_{GS} Characteristics



(for V_{DS} = 2.5V, W/L = 1.5)

- Linear (short-channel) versus quadratic (long-channel) dependence of I_D on V_{GS} in saturation

- Velocity-saturation causes the short-channel device to saturate at substantially smaller values of V_{DS} resulting in a substantial drop in current drive

Since the expressions found are complex and unsuitable for first order manual analysis, a simpler model can be obtained by making two assumptions:

1. The velocity saturates abruptly at ξ_c and approximated by:

$$v = \mu_n \xi \qquad \text{for } \xi \le \xi_c$$

$$v = v_{sat} = \mu_n \xi_c \qquad \text{for } \xi \ge \xi_c$$

2. The V_{DSAT} at which the critical electric field is reached and velocity saturation comes into play is constant and approximated by (reasonable for larger values of V_{GT}):

$$V_{DSAT} \approx L\xi_c = \frac{Lv_{sat}}{\mu_n}$$

$$V_{DSAT} = K(V_{GT})V_{GT} = \frac{1}{1+\frac{V_{GT}}{\xi_c L}}V_{GT} = \frac{\xi_c L}{\xi_c L + V_{GT}}V_{GT}$$

where $\xi_c L + V_{GT}$ can be replaced by V_{GT} for large V_{GT}

$$\approx \frac{\xi_c L}{V_{GT}}V_{GT} = \xi_c L$$

Under these assumptions, the current equations for the resistive region remain unchanged from the long-channel model. Once V_{DSAT} is reached, the current abruptly saturates. The value of I_{DSAT} at that point can be derived by inserting the saturation voltage into the current equation of the resistive region:

$$I_{DSAT} = I_D(V_{DS} = V_{DSAT})$$

$$= \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_T)V_{DSAT} - \frac{V^2_{DSAT}}{2}\right] \qquad (1)$$

where $V_{DSAT} \approx L\xi_c = L\frac{v_{sat}}{\mu_n} \rightarrow \frac{\mu_n}{L} = \frac{v_{sat}}{V_{DSAT}}$

inserting into (1) $= \frac{v_{sat}}{V_{DSAT}} C_{ox} W\left[(V_{GS} - V_T)V_{DSAT} - \frac{V^2_{DSAT}}{2}\right]$

$$I_{DSAT} = v_{sat} C_{ox} W\left[V_{GS} - V_T - \frac{V_{DSAT}}{2}\right]$$

This model is truly first-order and empirical. However it causes deviations in the transition zone between linear and velocity-saturated regions.

Subthreshold Conductance

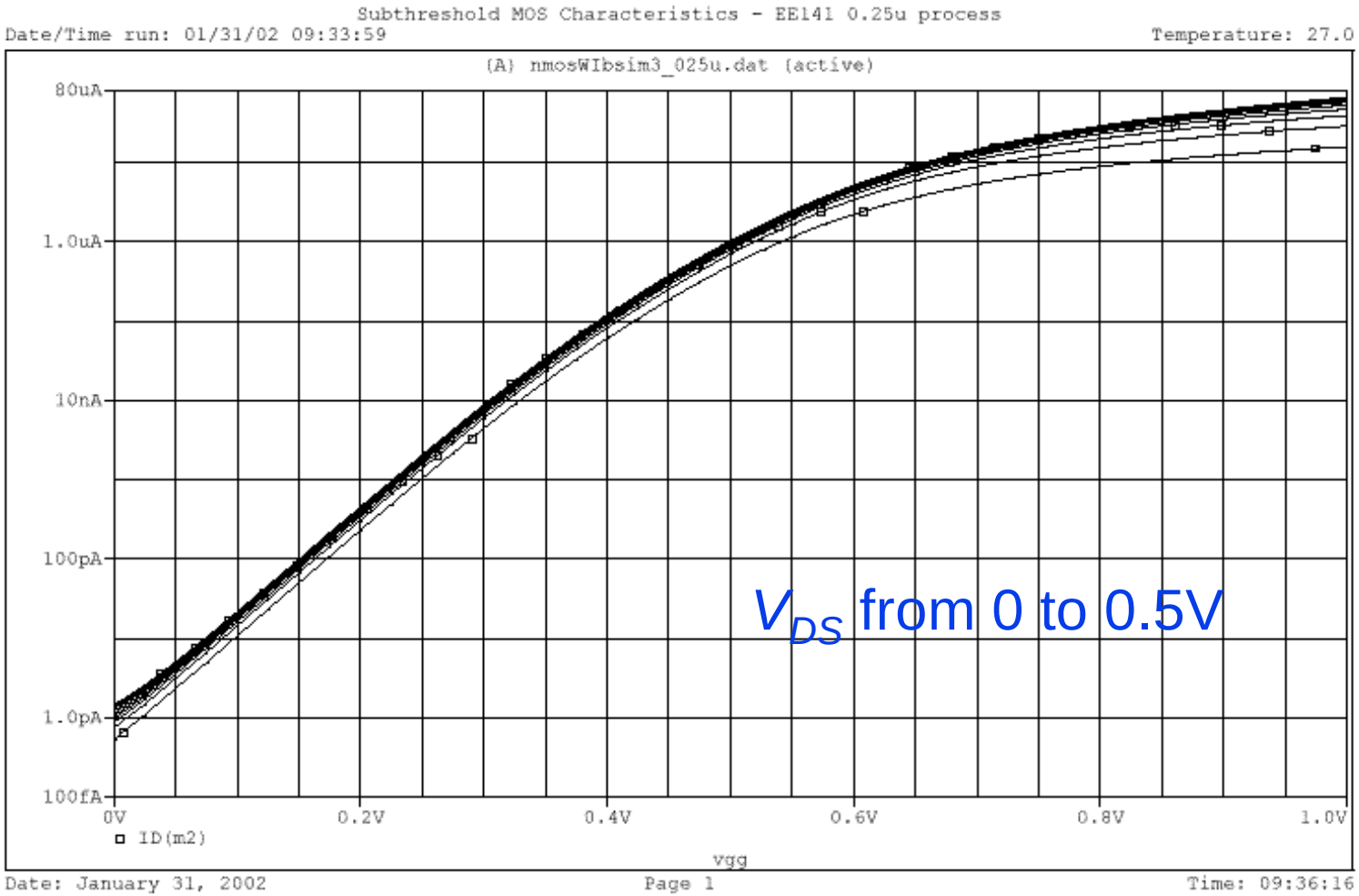
When $V_{GS}<V_T$, the transition from ON to OFF condition is not abrupt, but gradual. The current ($I_{D)}$ does not drop to zero immediately, but actually decays in an exponential fashion, similar to the operation of a bipolar transistor. In the absence of a conducting channel, the source, bulk and drain actually form a parasitic npn bipolar transistor. The current in this condition can be approximated by:

$$I_D = I_S e^{\frac{V_{GS}}{nkT/q}} \left[1 - e^{-\frac{V_{DS}}{kT/q}} \right] (1 + \lambda V_{DS})$$

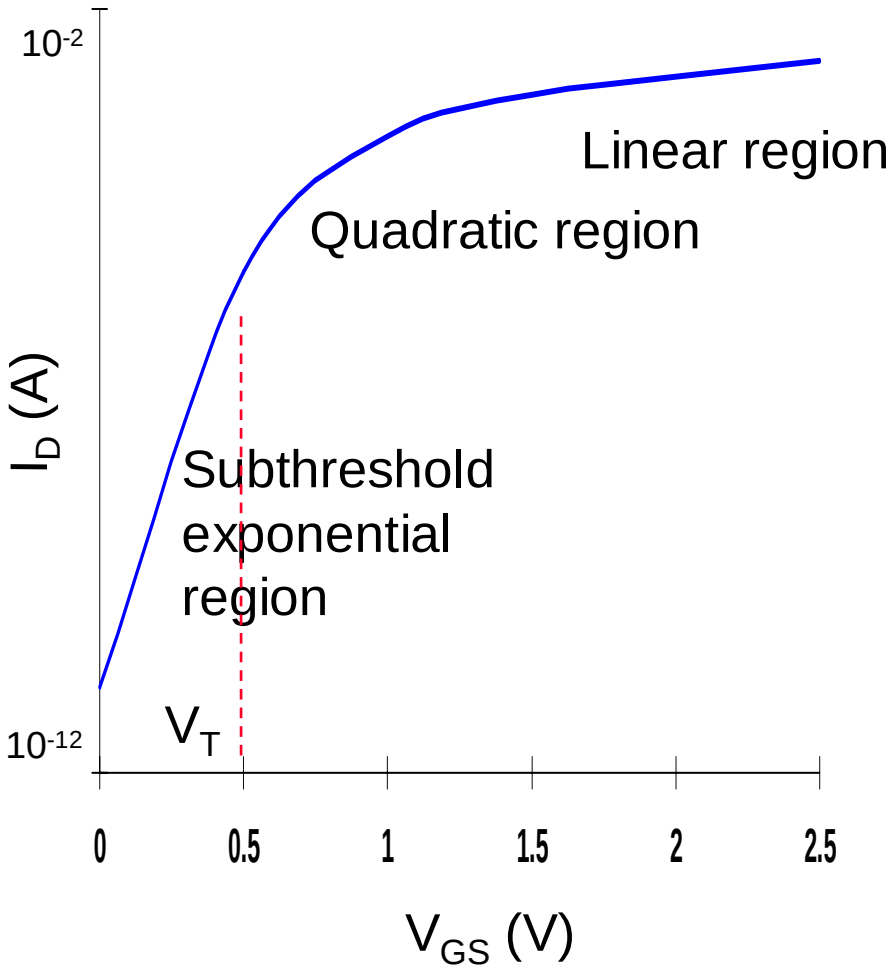
Where I_S and n are empirical parameters, with n typically ranging around 1.5 (≥1)

Subthreshold I_D vs V_{GS}

$$I_D = I_S\, e^{(qV_{GS}/nkT)}\ (1 - e^{-(qV_{DS}/kT)})(1 + \lambda V_{DS})$$



Subthreshold Conductance



$I_D \sim I_S e^{(qV_{GS}/nkT)}$ where $n \geq 1$

- Transition from ON to OFF is gradual (decays exponentially)
- Current roll-off (slope factor) is also affected by increase in temperature

$$S = n\,(kT/q)\,\ln(10)$$

(typical values 60 (n=1) to 100 mV/decade)(how much V_{GS} needs to be dropped, to drop I_D by a factor of 10)

- Creates problems in dynamic circuits and for power consumption

IN SUMMARY:

$I_D = 0 \qquad V_{GS} \le V_T$ (Cut-off)

$I_D = k_n \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^{\ 2}}{2} \right] \qquad V_{GS} > V_T \ , \ V_{DS} < V_{GS} - V_T \ , \ V_{DS} < V_{DSAT}$ (Triode)

$I_D = \frac{1}{2} k_n (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \qquad V_{GS} > V_T \ , \ V_{DS} \ge V_{GS} - V_T \ , \ V_{GS} - V_T < V_{DSAT}$ (Saturation)

$I_D = k_n \left[(V_{GS} - V_T) V_{DSAT} - \frac{V_{DSAT}^{\ 2}}{2} \right] \quad V_{GS} > V_T \ , \ V_{DS} \ge V_{DSAT} \ , \ V_{DSAT} < V_{GS} - V_T$ (Velocity Sat)

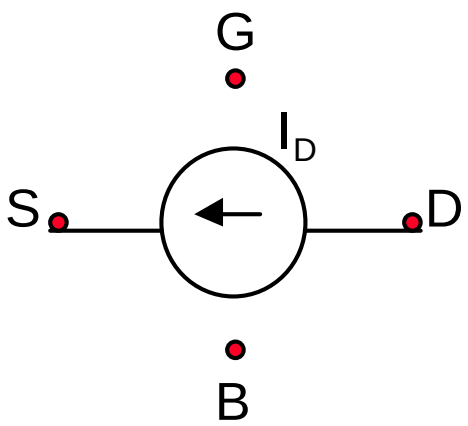
Which can be approximated by: $I_D \approx v_{sat} C_{ox} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$

Where:

$k_n = k_n' \frac{W}{L} \ , \quad k_n' = \mu_n C_{ox} = \mu_n \frac{\varepsilon_{ox}}{t_{ox}}$

$V_{GS} - V_T = V_{GT}$

The MOS Current-Source Model - Unified for manual analysis



$$I_D = 0 \text{ for } V_{GS} - V_T \leq 0$$

$$I_D = k' W/L \left[(V_{GS} - V_T)V_{min} - V_{min}^2/2\right](1+\lambda V_{DS})$$
$$\text{for } V_{GS} - V_T \geq 0$$

with $V_{min} = \min(V_{GS} - V_T, V_{DS}, V_{DSAT})$
and $V_{GT} = V_{GS} - V_T$

$$V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$$

- Determined by the voltages at the four terminals and a set of five device parameters (0.25µm CMOS below)

	V_{T0}(V)	$\gamma(V^{0.5})$	V_{DSAT}(V)	$k'(A/V^2)$	$\lambda(V^{-1})$
NMOS	0.43	0.4	0.63	115×10^{-6}	0.06
PMOS	-0.4	-0.4	-1	-30×10^{-6}	-0.1

This model employs 5 parameters which are determined by device physics and process technology:

$$V_{T0}\ ,\ \gamma,\ V_{DSAT}\ ,\ k',\ \lambda$$

In PMOS devices all of these parameters are negative, while they have positive values for NMOS.

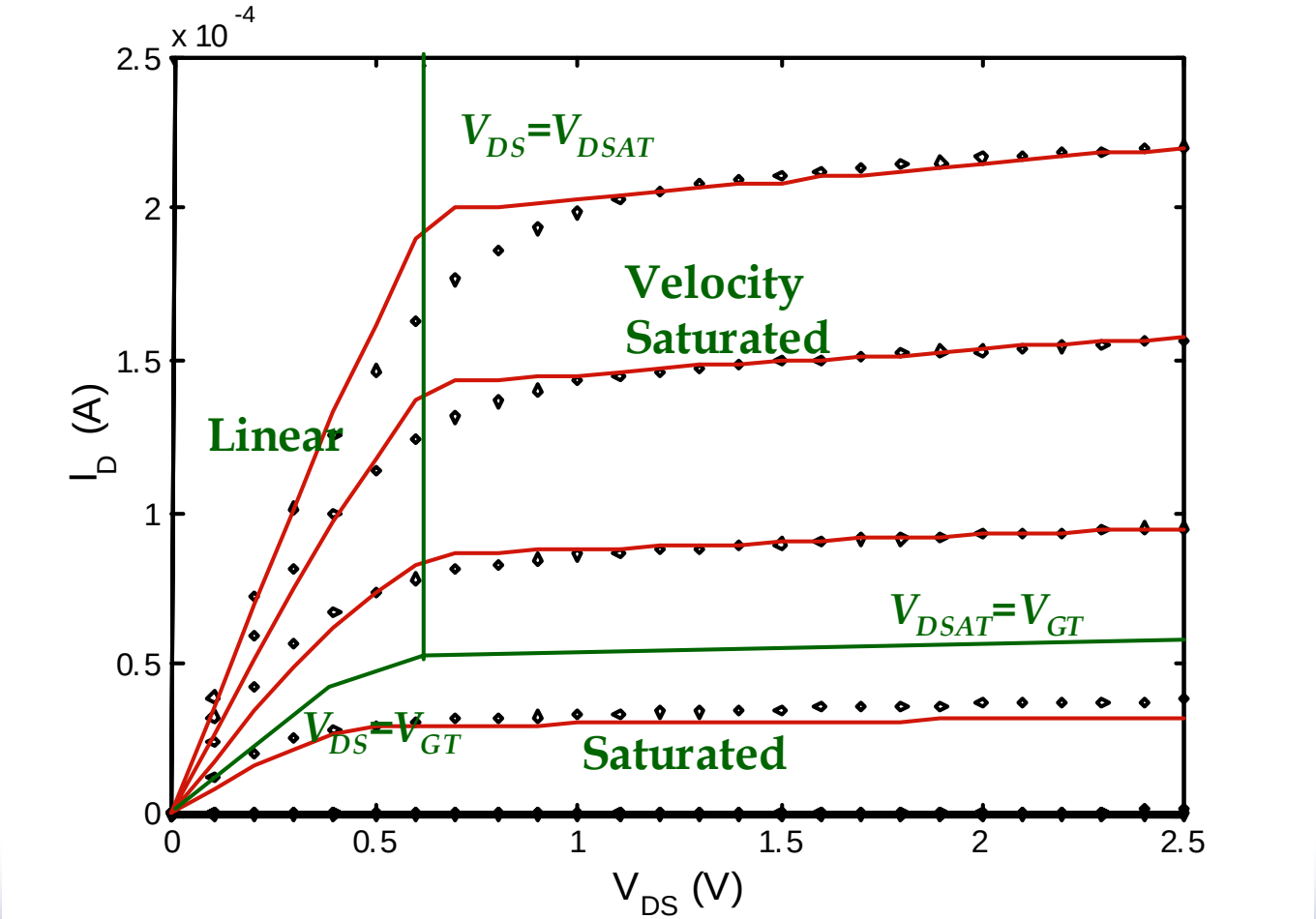
“min” function turns into “max” for PMOS.

V_{GS} and V_{DS} comes from the operating conditions. However, in digital circuits, this is mostly the region of high I_D, that is $V_{GS}=V_{DS}=V_{DD}$ (a good match with the model in these regions is therefore essential)

Simple Model versus SPICE

Simple: Solid line

SPICE: Dotted line



Resistance of Drain-to-Source Channel of NMOS Transistor Operating in Linear Mode

Ignoring the effects of channel-length modulation and velocity saturation:

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{V^2_{DS}}{2}\right]$$

$$R_{DS} = \left[\frac{\partial V_{DS}}{\partial I_D}\right]_{V_{GS}} = \left[\left(\frac{\partial I_D}{\partial V_{DS}}\right)^{-1}\right]_{V_{GS}} = \left[k_n\left((V_{GS} - V_T) - V_{DS}\right)\right]^{-1}$$

$$R_{DS} = \frac{1}{k_n\left(V_{GS} - V_T - V_{DS}\right)}$$

As V_{GS}-$V_T \rightarrow V_{DS}$, $R_{DS} \rightarrow \infty$

Repeating the same calculation for the saturation region (taking into consideration the channel length modulation but ignoring velocity saturation):

$$I_D = (k_n/2)\,[(V_{GS} - V_T)^2](1 + \lambda V_{DS})$$

$$R_{DS} = \left[\frac{\partial V_{DS}}{\partial I_D}\right]_{V_{GS}} = \left[\left(\frac{\partial I_D}{\partial V_{DS}}\right)^{-1}\right]_{V_{GS}}$$

$$R_{DS} = \left[\lambda k_n / 2\left((V_{GS} - V_T)^2\right)\right]^{-1}$$

Which can be approximated by (ignoring $1 + \lambda V_{DS}$ in I_D):

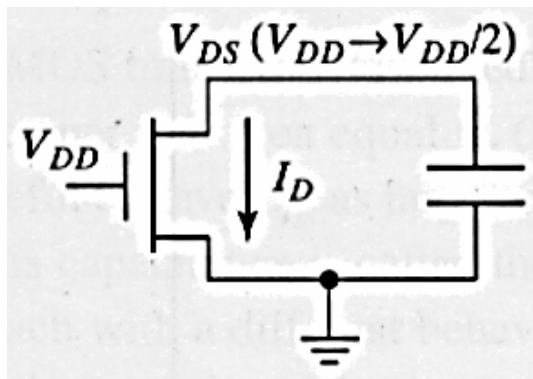
$$R_{DS} \cong \left[\lambda I_D\right]^{-1}$$ For a fixed V_{GS}

R_{on} is time varying, nonlinear and dependent on the operation point of the transistor. An average value of the resistance (R_{eq}) at the end points of the transition (switching between logical states) can be a suitable constant, linear approximation.

$$R_{eq} = average_{t=t_1 \ldots t_2}(R_{on}(t)) = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} R_{on}(t)dt$$

$$= \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} \frac{V_{DS}(t)}{I_D(t)}\,dt \approx \frac{1}{2}(R_{on}(t_1) + R_{on}(t_2))$$

By virtue of the definition of the propagation delay, lets take the discharging of a capacitor from V_{DD} to $1/2V_{DD}$, through an NMOS transistor. (assuming it stays in velocity sat.)



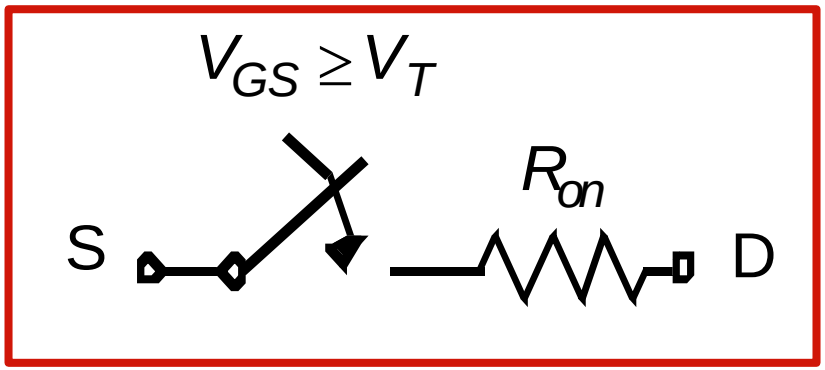
$$R_{eq} = \frac{1}{-V_{DD}/2} \int_{V_{DD}}^{V_{DD}/2} \frac{V}{I_{DSAT}(1+\lambda V)}\, dV = \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{7}{9}\lambda V_{DD}\right)$$

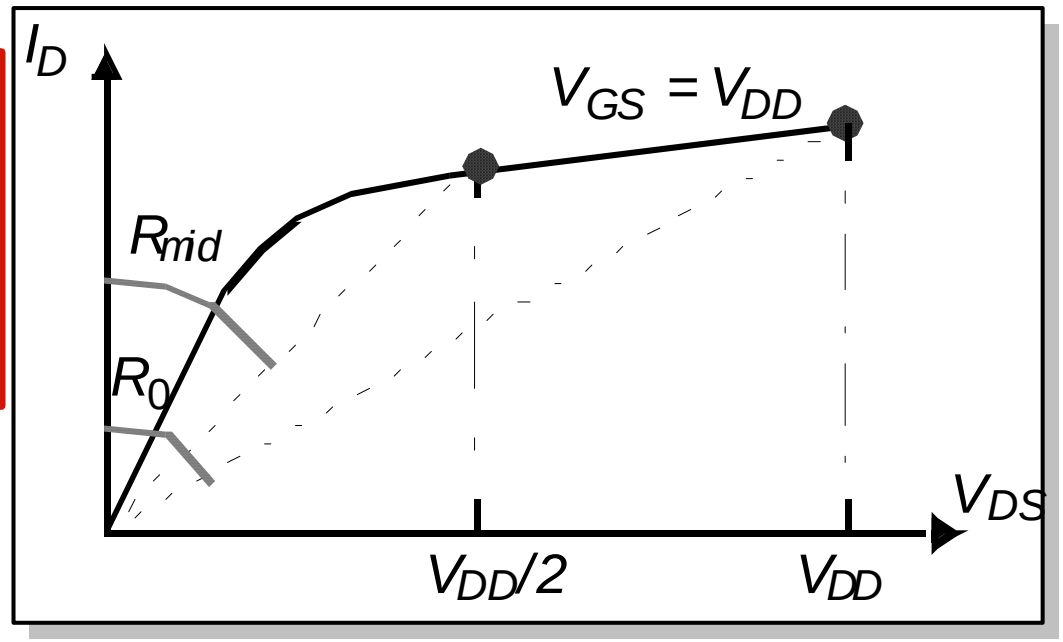
$$\text{with,} \qquad I_{DSAT} = k'\frac{W}{L}\left((V_{DD}-V_T)V_{DSAT} - \frac{V_{DSAT}^2}{2}\right)$$

A similar result can be obtained by just averaging the values of the resistance at the end points of the transition region and simplifying the result using a Taylor expansion:

$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSAT}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1+\lambda V_{DD}/2)}\right) \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{5}{6}\lambda V_{DD}\right)$$

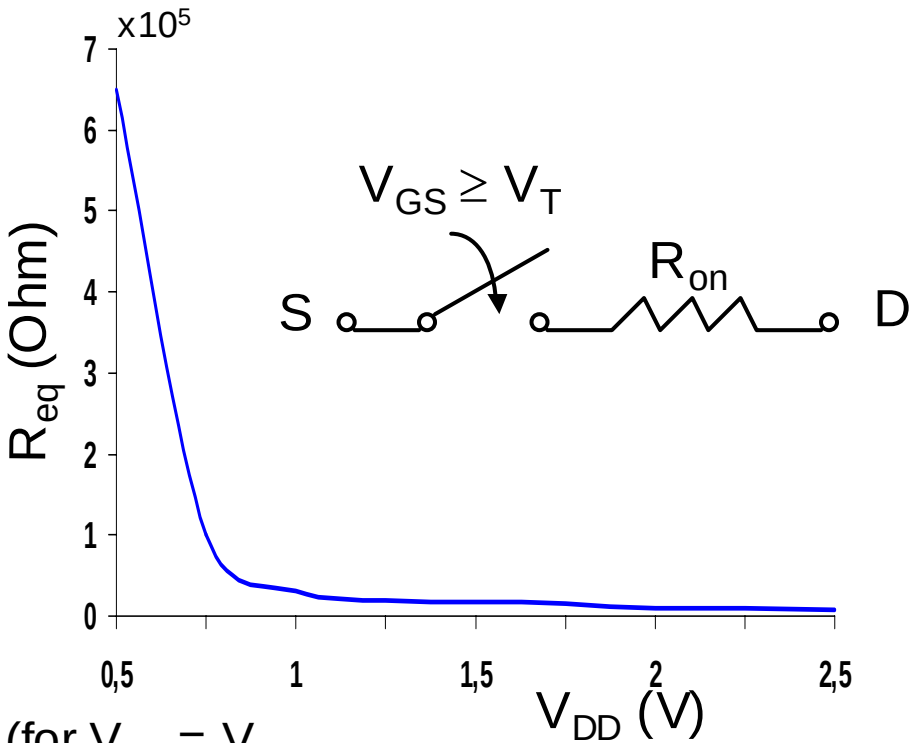
The Transistor as a Switch





$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSAT}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1+\lambda V_{DD}/2)}\right) \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{5}{6}\lambda V_{DD}\right)$$

The Transistor Modeled as a Switch



(for $V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$)

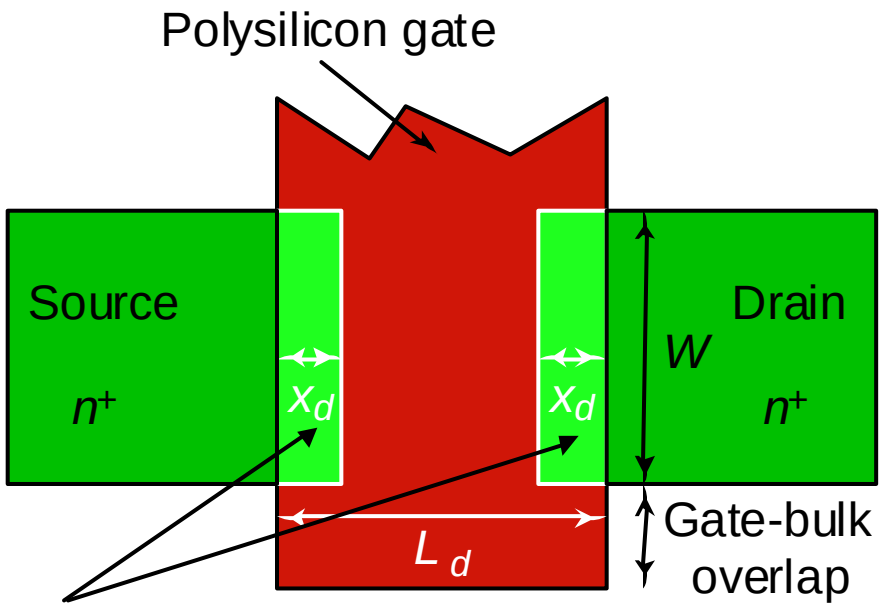
Modeled as a switch with infinite off resistance and a finite on resistance, R_{on}

- Resistance inversely proportional to W/L (doubling W halves R_{on})
- For $V_{DD} >> V_T + V_{DSAT}/2$, R_{on} independent of V_{DD}
- Once V_{DD} approaches V_T, R_{on} increases dramatically

V_{DD}(V)	1	1.5	2	2.5
NMOS(kΩ)	35	19	15	13
PMOS (kΩ)	115	55	38	31

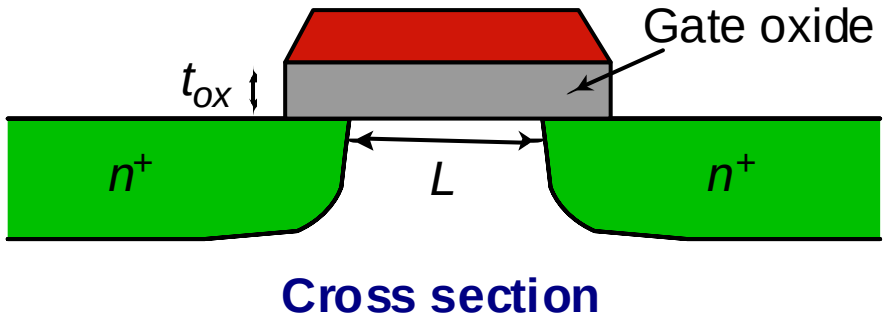
R_{on} (for W/L = 1)
For larger devices divide R_{eq} by W/L

The Gate Capacitance



$C_{ox}=\varepsilon_{ox}/t_{ox}$

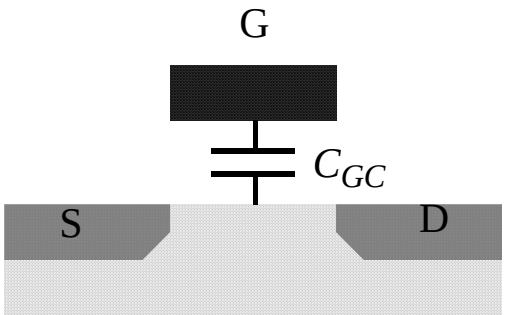
$$C_{gate} = \frac{\varepsilon_{ox}}{t_{ox}} WL$$



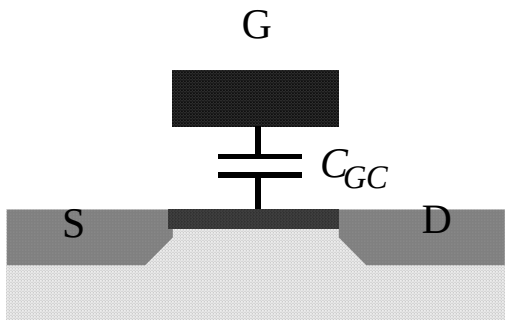
$\Delta L = 2x_d$

Overlap Capacitance: $C_{GSO} = C_{GDO} = C_{ox}x_dW = C_oW$

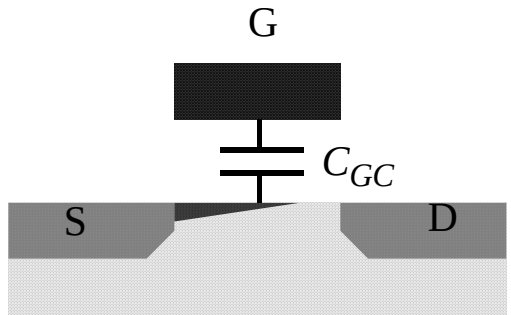
Gate Capacitance



Cut-off



Resistive

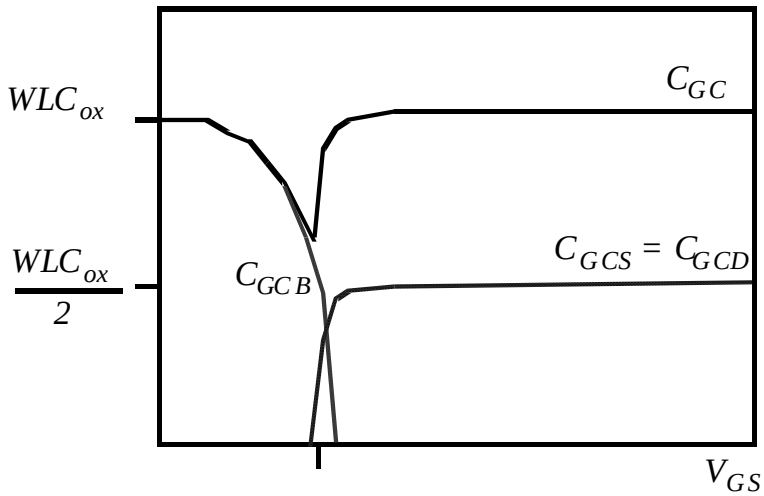


Saturation

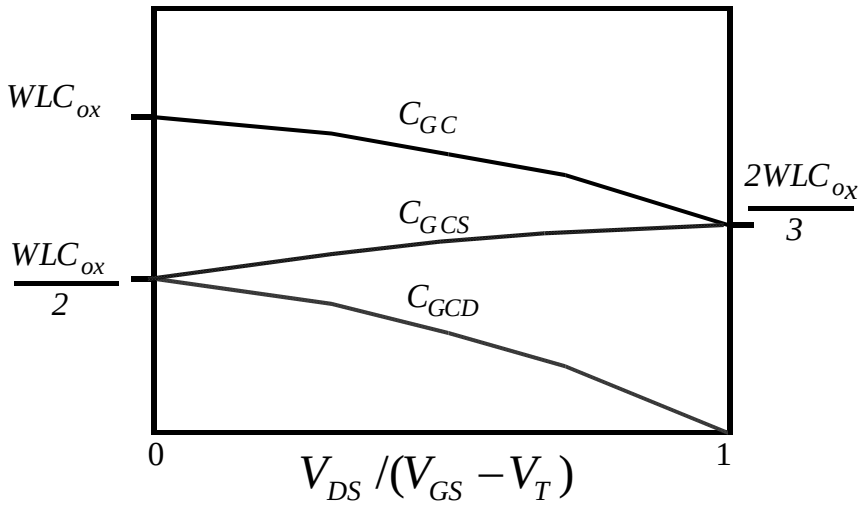
Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Cutoff	$C_{ox}WL$	0	0	$C_{ox}WL$	$C_{ox}WL+2C_0W$
Resistive	0	$C_{ox}WL/2$	$C_{ox}WL/2$	$C_{ox}WL$	$C_{ox}WL+2C_0W$
Saturation	0	$(2/3)C_{ox}WL$	0	$(2/3)C_{ox}WL$	$(2/3)C_{ox}WL+2C_0W$

Most important regions in digital design: saturation and cut-off

Gate Capacitance



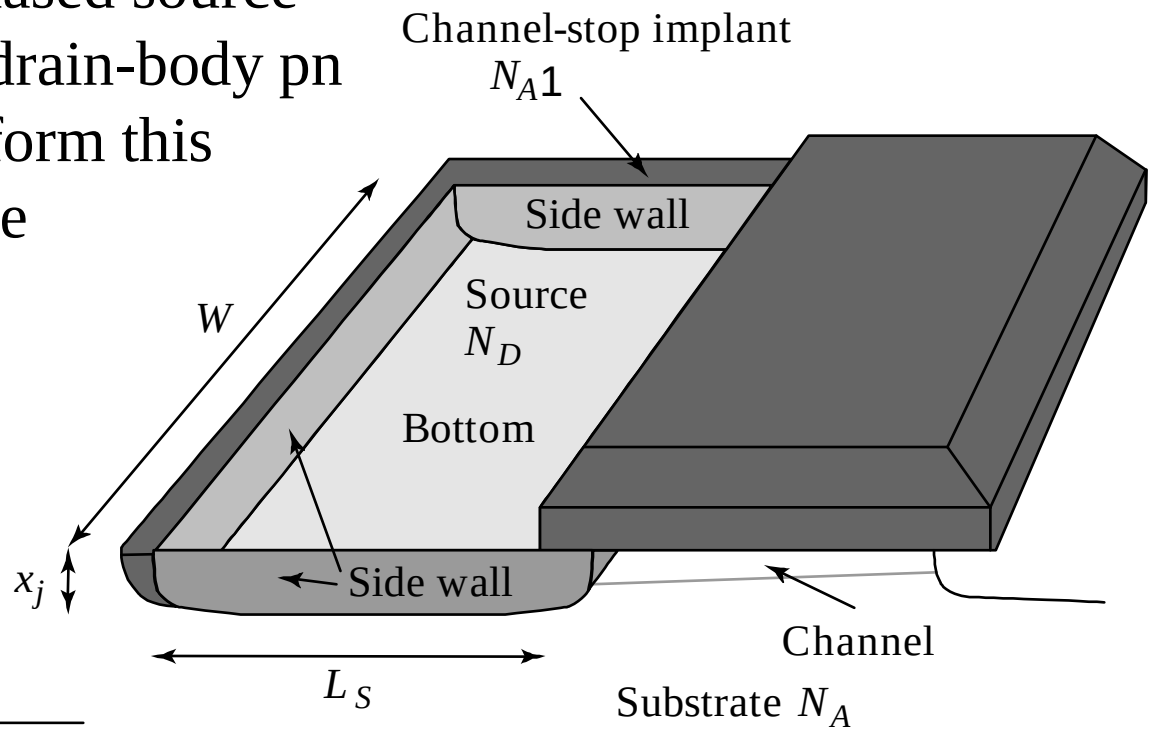
Capacitance as a function of VGS (with VDS = 0)



Capacitance as a function of the degree of saturation

Diffusion Capacitance

Reverse biased source-body and drain-body pn junctions form this capacitance



$$C_j = \frac{C_{j0}}{\left(1 - V_D / \phi_0\right)^m}$$

$$C_{sw} = C'_{sw} x_j (W + 2L_s)$$

$$C_{jsw} = C'_{sw} x_j$$

$$\begin{aligned} C_{diff} &= C_{bottom} + C_{sw} = C_j \times AREA + C_{jsw} \times PERIMETER \\ &= C_j L_S W + C_{jsw}(2L_S + W) \end{aligned}$$

MOSFET Capacitance Model

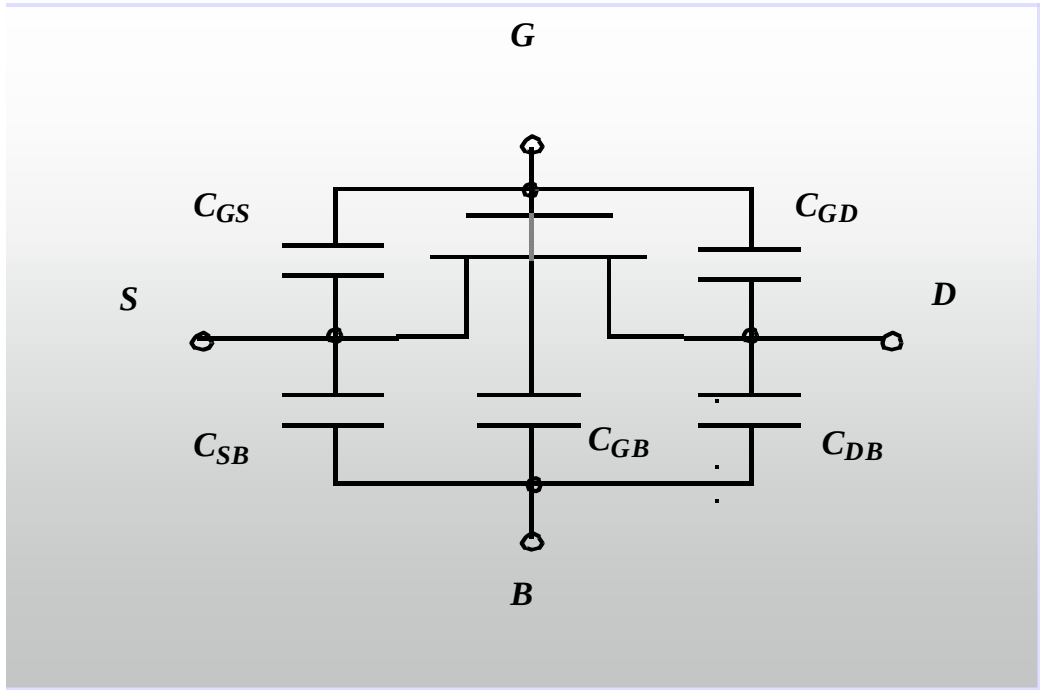
$$C_{GS} = C_{GCS} + C_{GSO}$$

$$C_{GD} = C_{GCD} + C_{GDO}$$

$$C_{GB} = C_{GCB}$$

$$C_{SB} = C_{Sdiff}$$

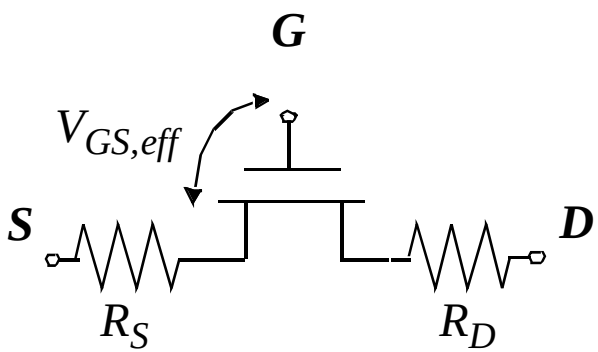
$$C_{DB} = C_{Ddiff}$$

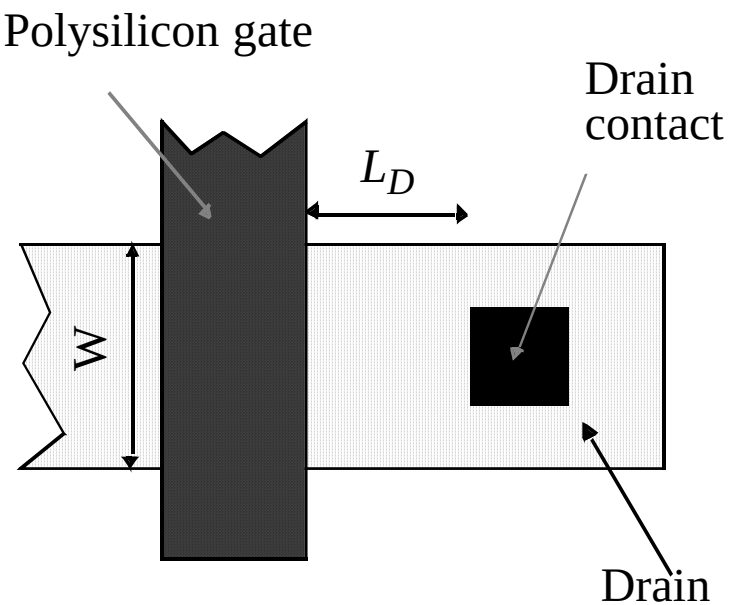


Capacitances in 0.25 μm CMOS process

	C_{ox} (fF/μm^2)	C_O (fF/μm)	C_j (fF/μm^2)	m_j	ϕ_b (V)	C_{jsw} (fF/μm)	m_{jsw}	ϕ_{bsw} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

Parasitic Resistances





HOT-CARRIER EFFECTS

Treshold voltages tend to drift over time in short-channel devices

Device dimensions have been scaled down continuously, but power supply and the operating voltages not.

Resulting an increase in the electrical field, causing increase in the electron velocity.

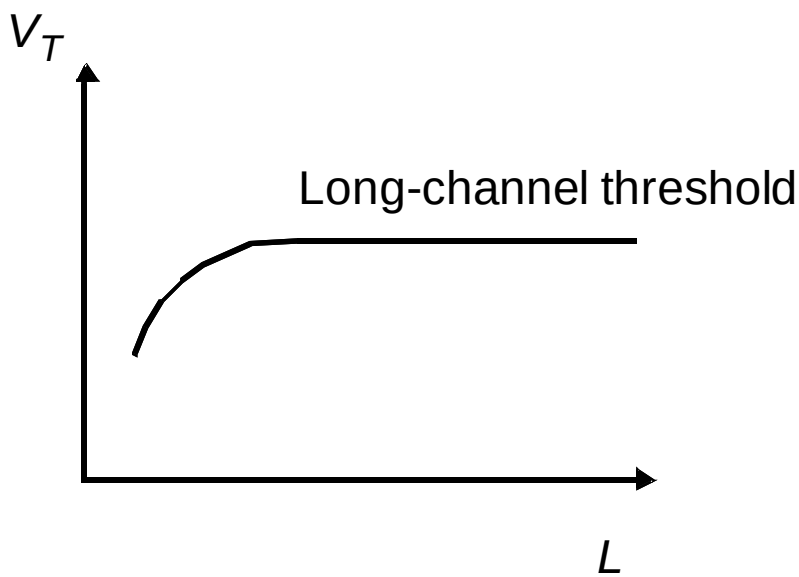
High energy electrons penetrate into gate oxide and get trapped there.

Which cause an increase in treshold voltage (decrease for PMOS device) That is a long term reliability problem.

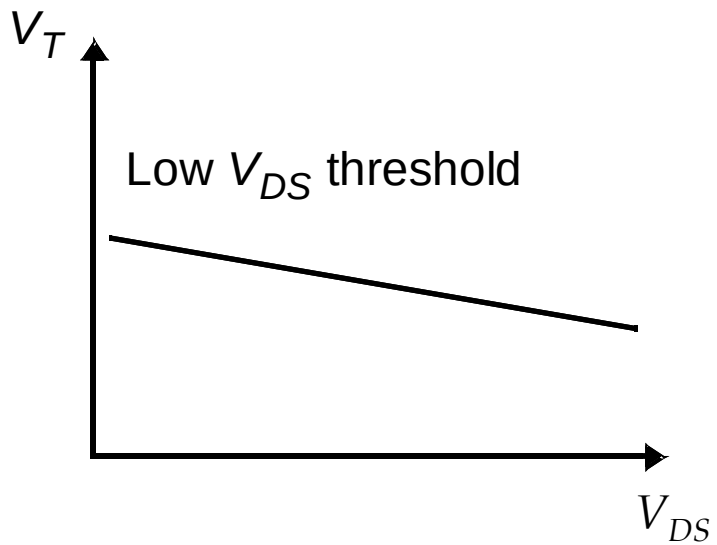
Solution: Specially engineered drain and source regions to limit the peaks of electric field and lowered supply voltages.

Threshold Variations

Part of the region below the gate is already depleted due to drain and source fields, a smaller V_T can cause strong inversion



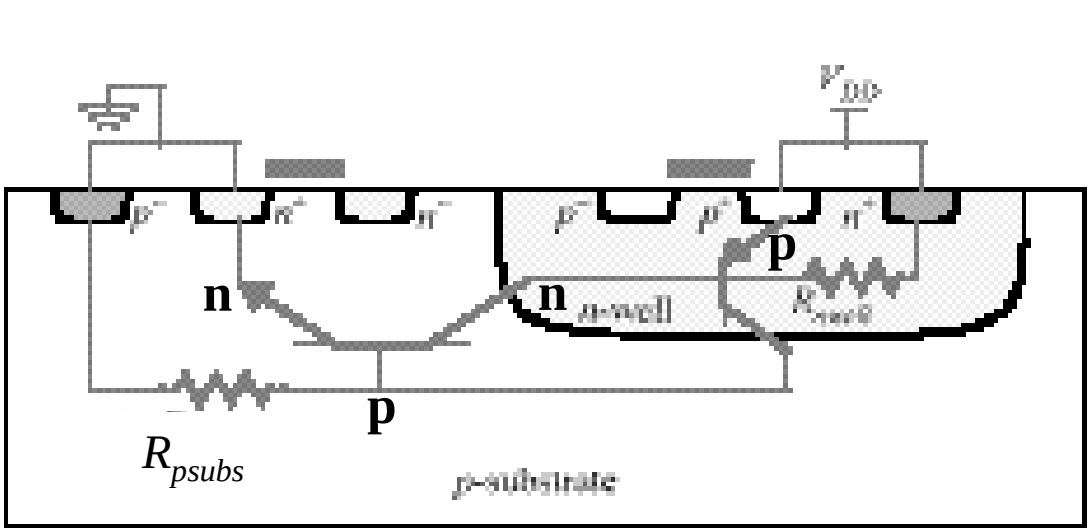
Threshold as a function of the length (for low V_{DS})



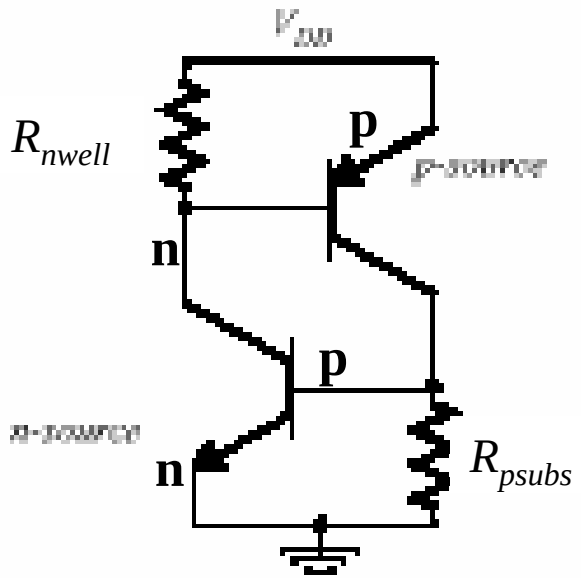
DIBL -Drain-induced barrier lowering *(for low L)*

Data dependent noise: Subtreshold leakage current becomes a function on the bit line and depends upon the data patterns.

Latch-up



(a) Origin of latchup



(b) Equivalent circuit

n (source of NMOS) - **p** (p substrate) - **n** (n well) - **p** (source of PMOS)