

*Introduction to VLSI design (EECS 467)*

Project

**Short-Channel Effects in MOSFETs**

December 11<sup>th</sup>, 2000

**Fabio D'Agostino**

**Daniele Quercia**

## Short-Channel Devices

A MOSFET device is considered to be short when the channel length is the same order of magnitude as the depletion-layer widths ( $x_{dD}$ ,  $x_{dS}$ ) of the source and drain junction.

As the channel length  $L$  is reduced to increase both the operation speed and the number of components per chip, the so-called short-channel effects arise.

## Short-Channel Effects

The short-channel effects are attributed to two physical phenomena:

1. the limitation imposed on electron drift characteristics in the channel,
2. the modification of the threshold voltage due to the shortening channel length.

In particular five different short-channel effects can be distinguished:

1. drain-induced barrier lowering and punchthrough
2. surface scattering
3. velocity saturation
4. impact ionization
5. hot electrons

## Drain-induced barrier lowering and punchthrough

The expressions for the drain and source junction widths are:

$$x_{dD} = \sqrt{\left(\frac{2e_{Si}}{qN_A}\right)(V_{DS} + \mathbf{f}_{Si} + V_{SB})}$$

and

$$x_{dS} = \sqrt{\left(\frac{2e_{Si}}{qN_A}\right)(\mathbf{f}_{Si} + V_{DB})}$$

where  $V_{SB}$  and  $V_{DB}$  are source-to-body and drain-to-body voltages.

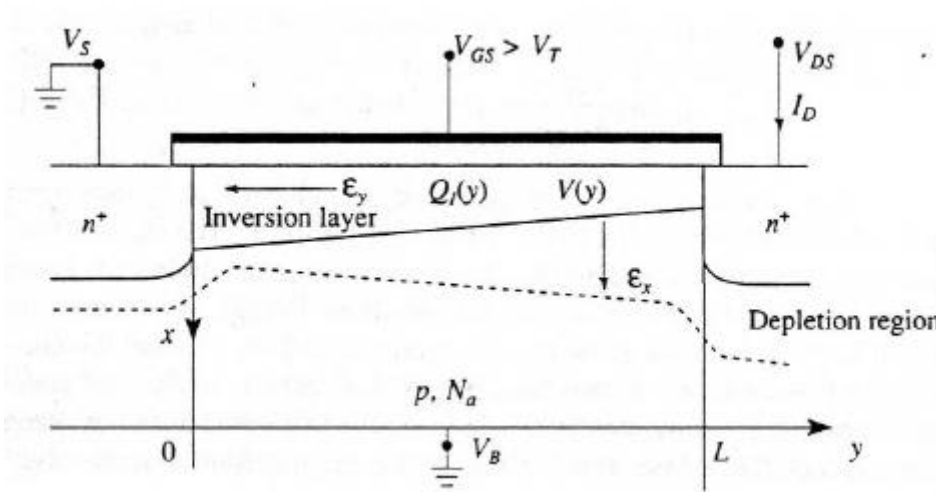
When the depletion regions surrounding the drain extends to the source, so that the two depletion layer merge (i.e., when  $x_{dS} + x_{dD} = L$ ), punchthrough occurs. Punchthrough can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels.

The current flow in the channel depends on creating and sustaining an inversion layer on the surface. If the gate bias voltage is not sufficient to invert the surface ( $V_{GS} < V_{T0}$ ), the carriers (electrons) in the channel face a potential barrier that blocks the flow. Increasing the gate voltage reduces this potential barrier and, eventually, allows the flow of carriers under the influence of the channel electric field. In small-geometry MOSFETs, the potential barrier is controlled by both the gate-to-source voltage  $V_{GS}$  and the drain-to-source voltage  $V_{DS}$ . If the drain voltage is increased, the potential barrier in the channel decreases, leading to *drain-induced barrier lowering* (DIBL). The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage. The channel current that flows under this conditions ( $V_{GS} < V_{T0}$ ) is called the sub-threshold current.

## Surface scattering

As the channel length becomes smaller due to the lateral extension of the depletion layer into the channel region, the longitudinal electric field component  $e_y$  increases, and the surface mobility becomes field-dependent. Since the carrier transport in a MOSFET is confined within the narrow

inversion layer, and the *surface scattering* (that is the collisions suffered by the electrons that are accelerated toward the interface by  $\mathbf{e}_x$ ) causes reduction of the mobility, the electrons move with great difficulty parallel to the interface, so that the average surface mobility, even for small values of  $\mathbf{e}_y$ , is about half as much as that of the bulk mobility.



## Velocity saturation

The performance short-channel devices is also affected by *velocity saturation*, which reduces the transconductance in the saturation mode. At low  $\mathbf{e}_y$ , the electron drift velocity  $v_{de}$  in the channel varies linearly with the electric field intensity. However, as  $\mathbf{e}_y$  increases above  $10^4$  V/cm, the drift velocity tends to increase more slowly, and approaches a saturation value of  $v_{de(sat)}=10^7$  cm/s around  $\mathbf{e}_y=10^5$  V/cm at 300 K.

Note that the drain current is limited by velocity saturation instead of pinchoff. This occurs in short-channel devices when the dimensions are scaled without lowering the bias voltages.

Using  $v_{de(sat)}$ , the maximum gain possible for a MOSFET can be defined as

$$g_m = WC_{ox}v_{de(sat)}$$

## Impact ionization

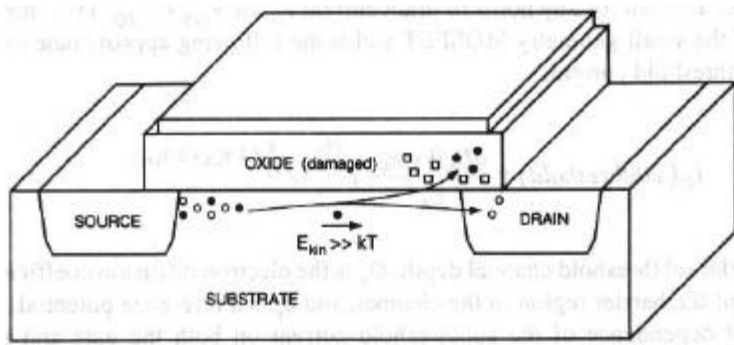
Another undesirable short-channel effect, especially in NMOS, occurs due to the high velocity of electrons in presence of high longitudinal fields that can generate electron-hole (e-h) pairs by *impact ionization*, that is, by impacting on silicon atoms and ionizing them.

It happens as follow: normally, most of the electrons are attracted by the drain, while the holes enter the substrate to form part of the parasitic substrate current. Moreover, the region between the source and the drain can act like the base of an npn transistor, with the source playing the role of the emitter and the drain that of the collector. If the aforementioned holes are collected by the source, and the corresponding hole current creates a voltage drop in the substrate material of the order of .6V, the normally reversed-biased substrate-source pn junction will conduct appreciably. Then electrons can be injected from the source to the substrate, similar to the injection of electrons from the emitter to the base. They can gain enough energy as they travel toward the drain to create new e-h pairs. The situation can worsen if some electrons generated due to high fields escape the drain field to travel into the substrate, thereby affecting other devices on a chip.

## Hot electrons

Another problem, related to high electric fields, is caused by so-called *hot electrons*. This high-energy electrons can enter the oxide, where they can be trapped, giving rise to oxide charging that

can accumulate with time and degrade the device performance by increasing  $V_T$  and affect adversely the gate's control on the drain current.



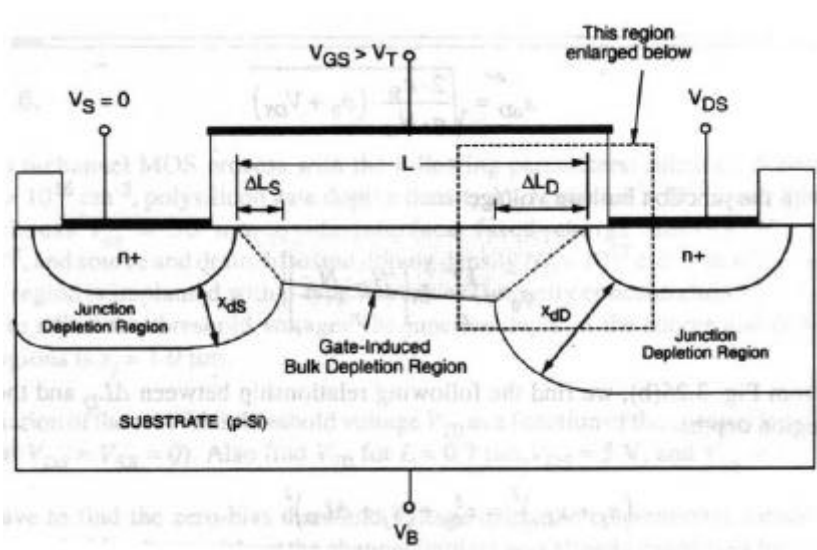
**The modification of the threshold voltage due to Short-Channel Effects (SCE)**

The equation giving the threshold voltage at zero-bias

$$V_{T0} = V_{FB} + 2\phi_F + \left( \frac{1}{C_{ox}} \right) \sqrt{2q \cdot e_{Si} \cdot N_A (2|\phi_F|)} + \frac{qD_I}{C_{ox}}$$

is accurate in describing large MOS transistors, but it collapses when applied to small-geometry MOSFETs. In fact that equation assumes that the bulk depletion charge is only due to the electric field created by the gate voltage, while the depletion charge near  $n^+$  source and drain region is actually induced by pn junction band bending. Therefore, the amount of bulk charge the gate voltage supports is overestimated, leading to a larger  $V_T$  than the actual value.

The electric flux lines generated by the charge on the MOS capacitor gate electrode terminate on the induced mobile carriers in the depletion region just under the gate. For short-channel MOSFETs, on the other hand, some of the field lines originating from the source and the drain electrodes terminate on charges in the channel region. Thus, less gate voltage is required to cause inversion. This implies that the fraction of the bulk depletion charge originating from the pn junction depletion and hence requiring no gate voltage, must be subtracted from the  $V_T$  expression.



The figure shows the simplified geometry of the gate-induced bulk depletion region and the p-n junction depletion regions in a short channel MOS transistor. Note that the bulk depletion region is assumed to have an asymmetric trapezoidal shape, instead of a rectangular shape, to represent

accurately the gate-induced charge. The drain depletion region is expected to be larger than the source depletion region because the positive drain-to-source voltage reversed-biases the drain-substrate junction. We recognize that a significant portion of the total depletion region charge under the gate is actually due to the source and drain junction depletion, rather than the bulk depletion induced by the gate voltage. Since the bulk depletion charge in the short channel device is smaller than expected, the threshold voltage expression must be modified to account for this reduction:

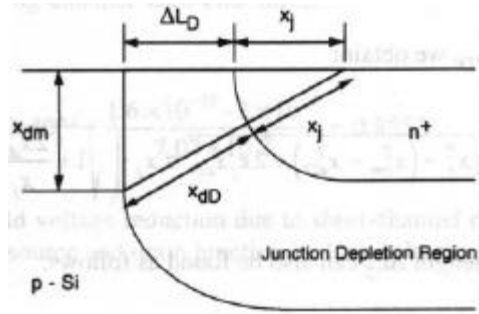
$$V_{T0(\text{short-channel})} = V_{T0} - \Delta V_{T0},$$

where  $V_{T0}$  is the zero-bias threshold voltage calculated using the conventional long-channel formula and  $\Delta V_{T0}$  is the threshold voltage shift (reduction) due to the short-channel effect. The reduction term actually represents the amount of charge differential between a rectangular depletion region and a trapezoidal depletion region.

Let  $DL_S$  and  $DL_D$  represent the lateral extent of the depletion regions associated with the source junction and the drain junction, respectively. Then, the bulk depletion region charge contained within the trapezoidal region is

$$Q_{B0} = - \left( 1 - \frac{\Delta L_S + \Delta L_D}{2L} \right) \sqrt{2q e_{Si} N_A |2f_F|}$$

To calculate  $DL_S$  and  $DL_D$ , we will use the simplified geometry shown in the figure.



Here,  $x_{ds}$  and  $x_{dd}$  represent the depth of the pn-junction depletion regions associated with the source and the drain, respectively. The edges of the source and drain diffusion regions are represented by quarter-circular arcs, each with a radius equal to the junction depth,  $x_j$ . The vertical extent of the bulk depletion region into the substrate is represented by  $x_{dm}$ . The junction depletion region depths can be approximated by

$$x_{dd} = \sqrt{\left( \frac{2e_{Si}}{qN_A} \right) (V_{DS} + f_0)}$$

and

$$x_{ds} = \sqrt{\left( \frac{2e_{Si}}{qN_A} \right) (f_0)}$$

with the junction built-in voltage

$$f_0 = \frac{kT}{q} \ln \left( \frac{N_D N_A}{n_i^2} \right)$$

From figure, we find the following relationship between  $DL_D$  and the depletion region depths.

$$(x_j + x_{dD})^2 = x_{dm}^2 + (x_j + \Delta L_D)^2$$

$$\Delta L_D^2 + 2x_j\Delta L_D + x_{dm}^2 - x_{dD}^2 - 2x_jx_{dD} = 0$$

Solving for  $DL_D$  we obtain:

$$\Delta L_D = -x_j + \sqrt{x_j^2 - (x_{dm}^2 - x_{dD}^2) + 2x_jx_{dD}} \cong x_j \left( \sqrt{1 + \frac{2x_{dD}}{x_j}} - 1 \right)$$

Similarly, the length  $DL_S$  can also be found as follows:

$$\Delta L_S \cong x_j \left( \sqrt{1 + \frac{2x_{dS}}{x_j}} - 1 \right)$$

Now, the amount of the threshold voltage reduction  $\Delta V_{T0}$  due to short-channel effects can be found as:

$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2qe_{si}N_A|2f_F|} \cdot \frac{x_j}{2L} \cdot \left[ \left( \sqrt{1 + \frac{2x_{dD}}{x_j}} - 1 \right) + \left( \sqrt{1 + \frac{2x_{dS}}{x_j}} - 1 \right) \right]$$

The threshold voltage shift term is proportional to  $x_j/L$ . As a result, this term becomes more prominent for MOS transistors with shorter channel lengths, and it approaches zero for long-channel MOSFETs where  $L \gg x_j$ .

## Numerical example

We consider an n-channel MOS process with the following parameters: substrate doping density  $N_A=10^{16} \text{ cm}^{-3}$ , polysilicon gate doping density  $N_D$  (gate) =  $2 \cdot 10^{20} \text{ cm}^{-3}$ , gate oxide thickness  $t_{ox}= 50 \text{ nm}$ , oxide-interface fixed charge density  $N_{ox}=4 \cdot 10^{10} \text{ cm}^{-2}$ , and source and drain diffusion doping density  $N_D= 10^{17} \text{ cm}^{-3}$ . In addition, we assume that the channel region is implanted with p-type impurities (impurity concentration  $N_I= 2 \cdot 10^{11} \text{ cm}^{-2}$ ) to adjust the threshold voltage. Moreover, the junction depth of the source and drain diffusion regions is  $x_j=1.0 \text{ }\mu\text{m}$ .

**Objective:** Plotting the variation of the threshold voltage  $V_T$  as a function of the channel length.

Now, we can calculate the work function difference between the gate and the channel:

$$F_{GC} = F_{GC(substrate)} - F_{F(gate)} = -0.35\text{V} - 0.55\text{V} = -0.90 \text{ V}.$$

The depletion region charge density at  $V_{SB}=0$  is found as follows:

$$Q_{B0} = -\sqrt{2q N_A e_{Si} |-2F_{F(substrate)}|} = -4.82 \cdot 10^{-8} \text{ C/cm}^2;$$

The oxide-interface charge is:

$$Q_{ox} = q N_{ox} = 6.4 \cdot 10^{-9} \text{ C/cm}^2;$$

The gate oxide capacitance per unit area is calculated using the dielectric constant of the silicon dioxide and the oxide thickness  $t_{ox}$ :

$$C_{ox} = e_{ox} / t_{ox} = 7.03 \cdot 10^{-8} \text{ F/cm}^2;$$

Now, we can combine all components and calculate the **threshold voltage without the channel implant**.

$$V'_{T0} = F_{GC} - 2 F_{F(substrate)} - (Q_{B0}/C_{ox}) - (Q_{ox}/C_{ox}) = 0.40 \text{ V}$$

Thus, we find the **long-channel zero-bias threshold voltage** for the process described above as

$$V_{T0} = V'_{T0} + (q N_I)/C_{ox} = 0.855 \text{ V}.$$

Next, the amount of the **threshold voltage reduction** due to short-channel effects must be calculated.

The source and drain junction built-in voltage is

$$F_0 = (kT/q) \ln((N_D N_A)/n_i^2) = 0.76 \text{ V}$$

The depths of source and drain junction depletion regions is found as

$$x_{dS} = \sqrt{(2 e_{Si})/(q N_A) F_0} = 0.314 \text{ }\mu\text{m}$$

$$x_{dD} = \sqrt{\left(\frac{2e_{Si}}{qN_A}\right)(V_{DS} + f_0)}$$

Now, the threshold voltage shift  $DV_{T0}$  can be calculated as a function of the gate length  $L$  and of the drain-to-source voltage  $V_{DS}$  (see the previous paragraph).

$$DV_{T0} = (0.343 / L[\mu\text{m}]) * (-0.724 + \sqrt{1 + 2x_{dD}})$$

where

$$x_{dD} = \sqrt{0.13(0.76 + V_{DS})}$$

The **threshold voltage** of this short-channel MOS transistor is calculated as

$$V_{T0} = 0.855\text{V} - DV_{T0};$$

The Matlab source simulating the previous expression of the threshold voltage is supplied.



```

close all, clear all;
pack;

% arrays for the different values of Vth @ Vds=1V,Vds=3V, Vds=5V.
vector_vt1=[];
vector_vt3=[];
vector_vt5=[];
% arrays for the different values of L @ Vds=1V,Vds=3V, Vds=5V.
vector_l1=[];
vector_l3=[];
vector_l5=[];

% for Vds=1 V
vds=1;
for l=0.5:0.1:6
    rad2=sqrt(0.13*(0.76+vds));
    rad1=sqrt(1+(2*rad2));
    deltavt=(0.343/l)*(rad1-0.724);
    vt=0.855-deltavt;
    vector_l1=[vector_l1,l];
    vector_vt1=[vector_vt1,vt];
end

%for Vds=3 V
vds=3;
for l=0.5:0.1:6
    rad2=sqrt(0.13*(0.76+vds));
    rad1=sqrt(1+(2*rad2));
    deltavt=(0.343/l)*(rad1-0.724);
    vt=0.855-deltavt;
    vector_l3=[vector_l3,l];
    vector_vt3=[vector_vt3,vt];
end

%for Vds=5 V
vds=5;
for l=0.5:0.1:6
    rad2=sqrt(0.13*(0.76+vds));
    rad1=sqrt(1+(2*rad2));
    deltavt=(0.343/l)*(rad1-0.724);
    vt=0.855-deltavt;
    vector_l5=[vector_l5,l];
    vector_vt5=[vector_vt5,vt];
end

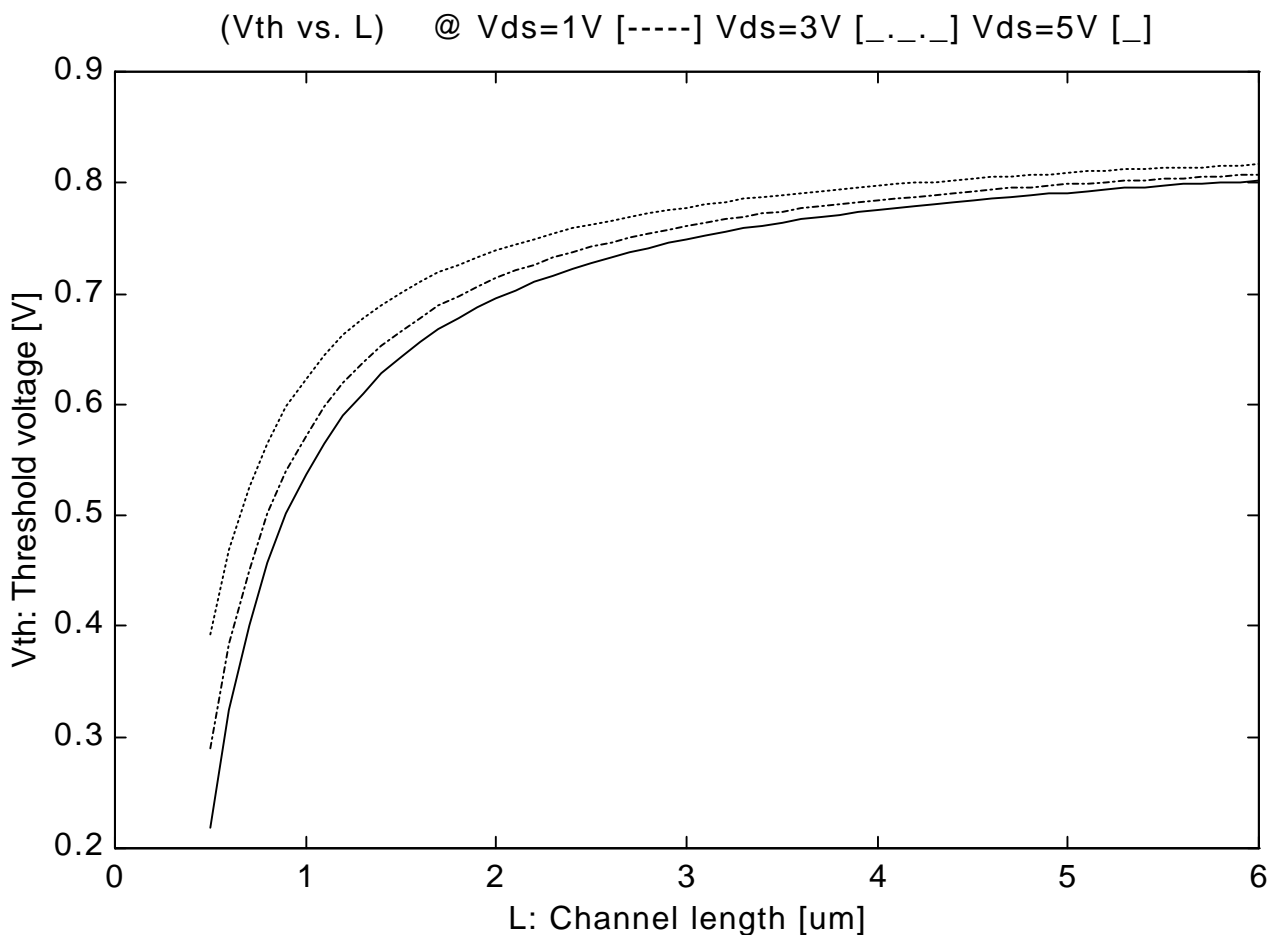
% graps of Vth vs. L @ Vds=1V,Vds=3V, Vds=5V.
plot(vector_l1,vector_vt1,':',vector_l3,vector_vt3,'-
.',vector_l5,vector_vt5,'.');

%comments on the plot
xlabel('L: Channel length [um]'),
ylabel('Vth: Threshold voltage [V]'),
title('(Vth vs. L) @ Vds=1V [-----] Vds=3V [_._._] Vds=5V [.....]');

```

In the next graph, the output of the previous Matlab file is shown. That plot shows the variation of the threshold voltage with the channel length. The threshold voltage decreases by as much as 50% for channel lengths in the submicron range, while it approaches the value of 0.8V for larger channel lengths.

*If the drain voltage is increased, the potential barrier in the channel decreases, leading to drain-induced barrier lowering (DIBL). The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage. The channel current that flows under this conditions ( $V_{GS} < V_{T0}$ ) is called the sub-threshold current.*



## Simulations

### ***Analysis of the short-channel effects on the threshold voltage***

To illustrate the impact of short-channel effects on the threshold voltage, we use SPICE to simulate a circuit with four n-MOSFET connected in parallel. We use LEVEL 2 spice model to generate  $I_D - V_{GS}$  characteristics for  $V_{DS} = 0.1V$ . The transistors have the same parameters except for the channel length and width that have been set to 10, 5, 1 and 0.5 $\mu\text{m}$ .

```
-----Short channel effects on the Threshold Voltage (3.22)
VDS 1 0 0.1V
VGS 2 0

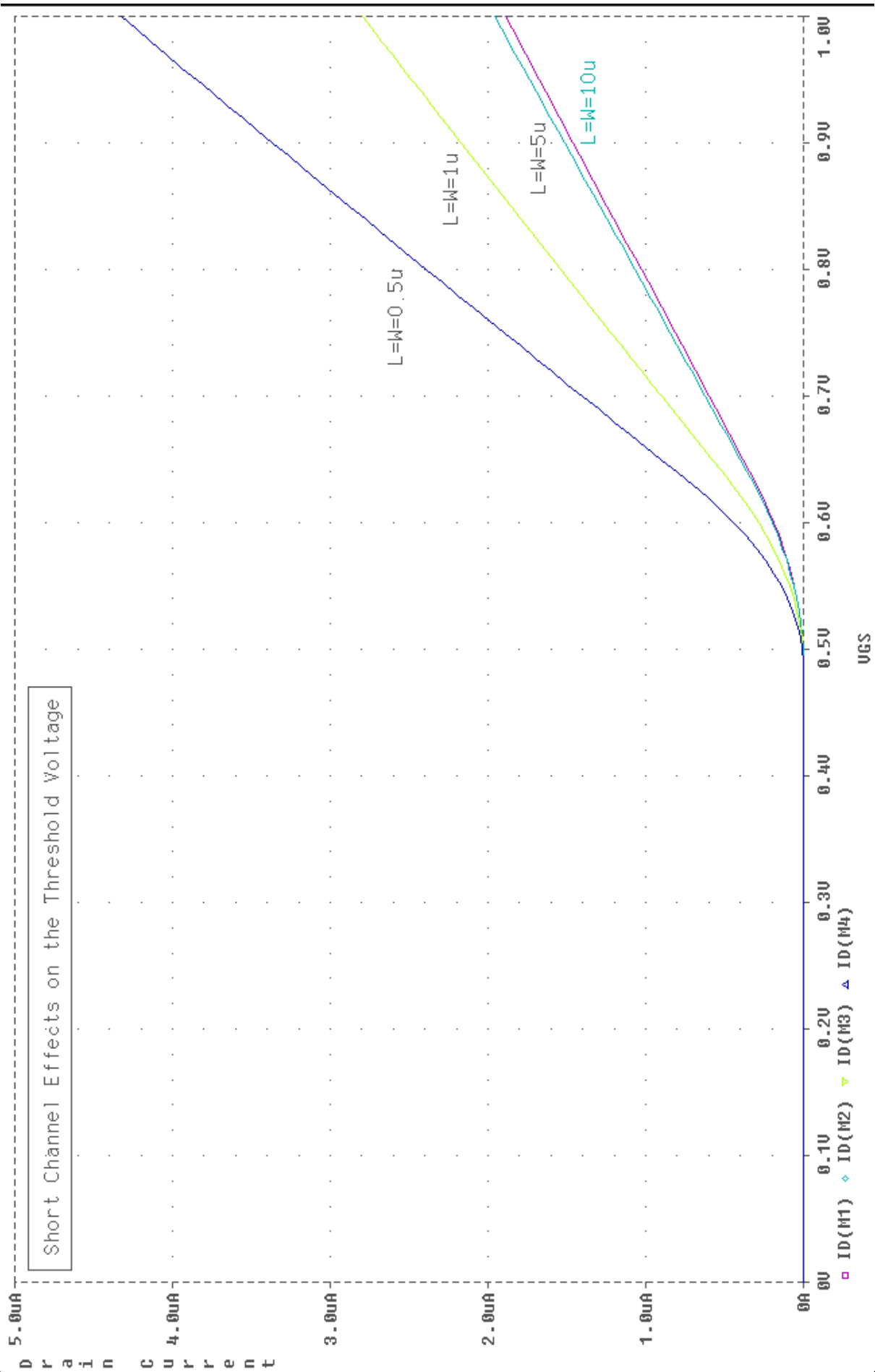
M1  1 2 0 0 N1
M2  1 2 0 0 N2
M3  1 2 0 0 N3
M4  1 2 0 0 N4

.MODEL N1 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=10u  W=10u
.MODEL N2 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=5u   W=5u
.MODEL N3 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=1u   W=1u
.MODEL N4 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=0.5u W=0.5u

.DC VGS 0 1 .01V
.PROBE

.END
```

The resulting plot is reported below. It shows that, those devices with smaller geometry have higher drain currents at the same gate-to-source voltage; hence devices with smaller geometry have lower threshold voltages.



## The limiting effect of the drift velocity

In order to investigate the limiting effect of velocity saturation on the drain current of a short-channel nMOS, we use LEVEL 2 spice model to generate  $I_D - V_{DS}$  characteristics of two transistors connected in parallel. The two transistors have the same parameters except for VMAX, representing the maximum drift velocity, limited in the first transistor to the value  $2 \cdot 10^6$  m/s. In the following lines the Pspice netlist is presented:

```
-----Limiting effect of velocity saturation on the drain current of SC n-MOS
VDS 1 0
VGS 2 0 5V

M1 1 2 0 0 N1
M2 1 2 0 0 N2

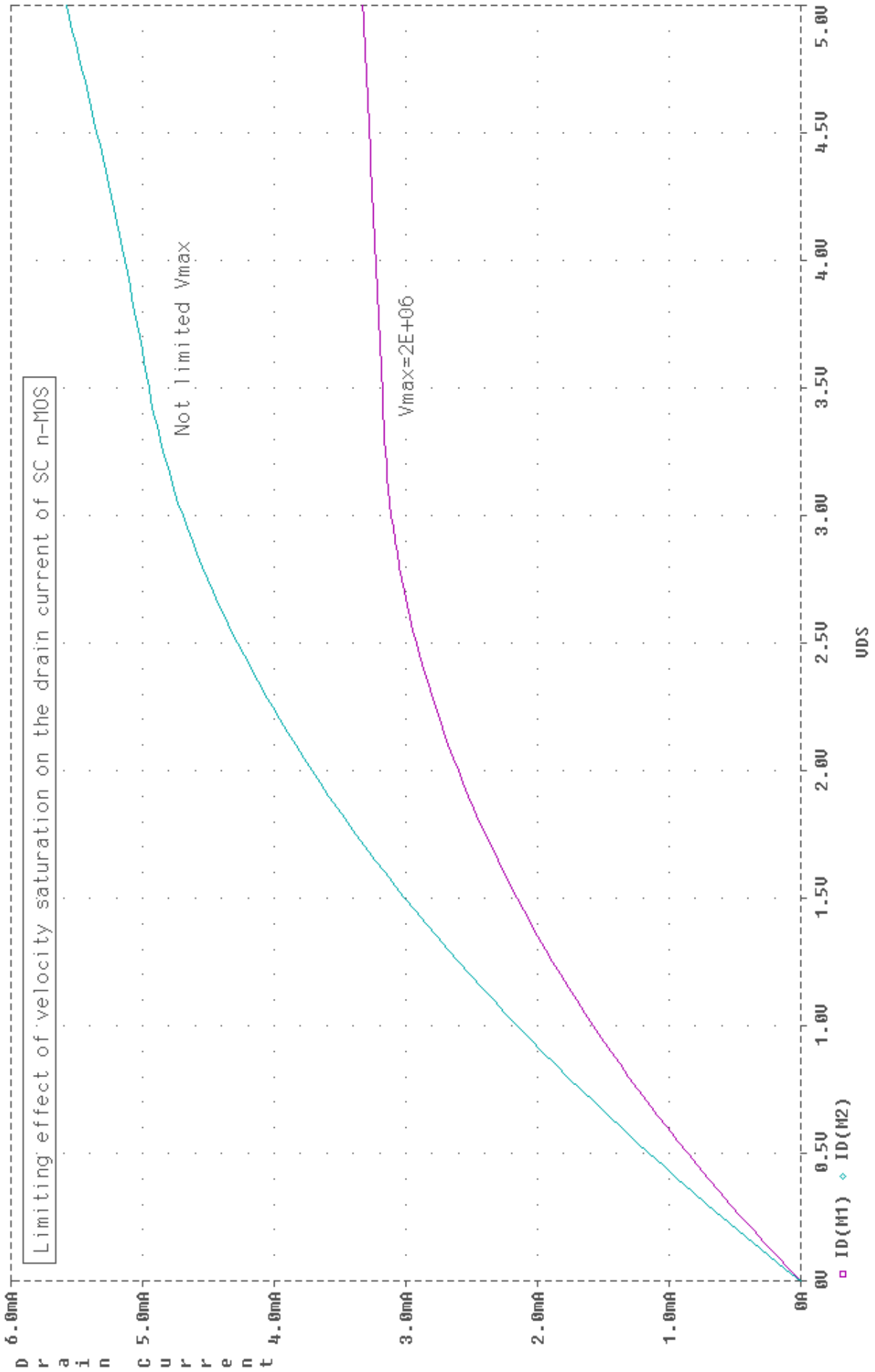
.MODEL N1 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=1u W=10u VMAX=2E+06
.MODEL N2 NMOS LEVEL=2 TOX=50n NSUB=1E16 L=1u W=10u

.DC VDS 0 5 .05
.PROBE

.END
```

The parameters NSUB and TOX represent respectively the substrate doping and the oxide thickness of the n-MOS device.

The plot shows the reduction of the transconductance in the saturation mode.



## Conclusions

The design of Ics was in the past simplified also due to over-designing. But as technology moves deeper into UDSM this is not a viable approach. Too much performance is sacrificed or the area penalty of over-designing leads to decreased yields. However, pushing the edge of the envelope may lead to under-design. Chips that have been under-designed often fail on the test bench or later in the field. Therefore, situations of over-design and under-design must both be identified when evaluating the integrity of a power distribution system.

In the end, the design tradeoffs that satisfy all the necessary constraints are too complex to handle without tools that provide visibility into specific problems and their location on a chip.

Designers are often required to tape out a design, and are left hoping that nothing will go wrong when the chip comes back from the fab. Murphy's Law is apropos for this situation: if something can go wrong, it probably will.

---

## REFERENCES

- Xing Zhou, Khee Yong Lim - *"A general approach to compact threshold Voltage formulation based on 2-D numerical simulation and experimental correlation for deep.submicron ULSI technology development"*
- Kai Chen and Chenming Hu - *"Performance and Vdd scaling in deep submicrometer CMOS"*
- Haldun H. - *"Digital microelectronics"*
- Resve Saleh, Michael Benoit and Pete McCrorie - *"Power distribution planning"*
- Sani R. Nassif - *"Design for Variability in DSM Technologies"*
- Sung-Mo, Yusuf - *"CMOS digital integrated circuits"*
- Lim, Zhou, Zu, Ho, Loiko, Lau, Tse, Choo - *"A predictive semi-analytical threshold voltage model for deep-submicron MOSFET's"*
- Weste, Eshraghian - *"Principles of CMOS VLSI Design"*