Asymmetry in effective-channel length of n- and p-MOSFETs

Ron Logan, Yuan Taur, and Emmanuel Crabbé
IBM Semiconductor Research and Development Center,
B640, Z/AE1, Hopewell Junction, NY, 12533, USA

I. INTRODUCTION

The concept of effective-channel length, L_{eff}, is central in describing the current-carrying capability of a MOSFET. This is especially true in the deep-submicron regime where the difference between the physical-gate length, L_{gate}, and the electrically defined effective-length becomes non-negligible. Our ability to understand and characterize L_{eff} is therefore important for the development of semiconductor technology.

A typical manufacturing process for CMOS devices involves using different types of dopants for the source/drain contacts of n- and p-channel devices. For the p-FET, boron is frequently used to dope the source/drain whereas for n-FETs, arsenic is a common choice. Ideally, one would like to construct shallow source/drain-contact extensions having very abrupt p/n junctions with the channel in order to help minimize short-channel effects [1]. While this is reasonably accomplished using the arsenic profile, the boron tends to give rise to junctions having a much more graded character. A simple indicator of this fact can be seen by examining the bias-dependent overlap capacitance [2]. For the abrupt profile, the overlap capacitance varies only slightly while reverse-biasing the source/drain whereas the more graded boron profile shows a substantial bias dependence. Given these facts, one would observe shorter L_{eff} values for p-FETs rather than n-FETs if the peak concentration at the contact surface was the same. As we shall demonstrate, there is an additional mechanism which serves to shorten the p-FET L_{eff} relative to the n-FET. The origin of this can be understood in terms of fundamental differences in electron and hole transport properties.

II. RESULTS AND DISCUSSION

In order to understand effective-channel length, we invoke a simple model containing two distinct types of carrier transport. First, the channel current will be considered to be a quasi-two dimensional flow of carriers that is strongly confined to the Si/SiO_2 interface. In contrast to this, the current in the source/drain is a fully three-dimensional flow extending from the interface into the bulk of the highly-doped contact regions. We describe the boundaries of the effective-channel length as the positions in which the current makes a transition between

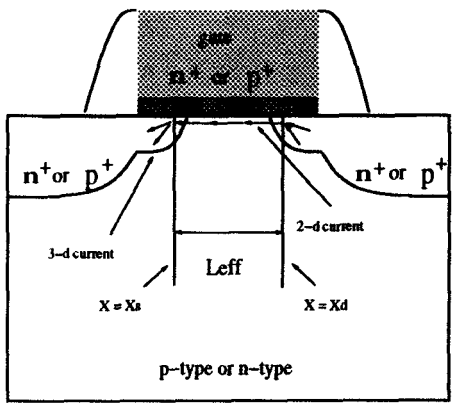

Figure 1: MOSFET cross-section showing the boundaries between the 2- and 3-d current flow regions which define the L_{eff} in our model.

these two distinct types of flow patterns. The present picture does represent a large simplification of real device behavior, since in reality there is no such abrupt transition from 2-d to 3-d current flow. We use this concept only in order to understand some basic, qualitative, differences in L_{eff} for n- and p-FETs related to basic differences in carrier transport. In Figure 1, we schematically show a cross-section of a FET with the L_{eff} boundaries designated as X_s at the source, and X_d at the drain. Between X_s and X_d we assume 2-d channel current and beyond a spread-out 3-d flow pattern.

The location of the points X_d and X_s can be estimated by identifying the position along the interface where the sheet resistance of the channel equals that of the source/drain i.e. for X_s at the source end

$$\rho_{s/d}(X_s) = \rho_{ch}(X_s). \qquad ...(1)$$

The accumulation charge in the extension part of the channel is approximately $C_{ox}V_{gs}$ where C_{ox} is the gate capacitance and V_{gs} is the gate to source voltage. Assuming constant doping in the extension over some

0-7803-3775-1/97/$10.00 © 1997 IEEE.

distance X_j we obtain

$$(q\,\mu_{s/d}\,N(X_s)\,X_j)^{-1} = (\,C_{ox}V_{gs}\,\mu_{ch}\,)^{-1}\ , \qquad ...(2)$$

where $N(X_s)$ is the dopant concentration at the point X_s, $\mu_{s/d}$ and μ_{ch} represent the mobility in the source/drain and channel, respectively. Note that in our simple two-region model the source/drain mobility is assumed to be doping limited. We may solve equation (2) for the dopant concentration at X_s to obtain,

$$N(X_s) = C_{ox}\,V_{gs}\mu_{ch}/(q\,X_j\mu_{s/d})\ . \qquad ...(3)$$

The mobility of holes in the high-doping limited regime is roughly half of that for electrons, whereas in the channel region the hole mobility is smaller by roughly a factor of four (see Fig. 2). Now if we assume that the shapes of the profiles are the same for both the n-FETs and p-FETs then we can estimate the ratio of $N(X_s)^{pfet}$ to $N(X_s)^{nfet}$ as,

$$N\,(X_s)^{pfet}/N(X_s)^{nfet}=\mu_{ch}{}^{pfet}\mu_{s/d}{}^{nfet}/(\mu_{s/d}{}^{pfet}\mu_{ch}{}^{nfet})\approx 1/2\ . \qquad ...(4)$$

The exact number for this ratio is unimportant, the key conclusion is that $N(X_s)^{pfet}/N(X_s)^{nfet} < 1$. The concentration $N(X_s)$ rises rapidly in going from the channel into the source/drain contact region. Therefore equation (4) indicates that the point X_s lies closer to the gate edge for n-FETS than for p-FETs. Thus the fundamental differences in channel and source/drain mobilities inherently give rise to a larger n-FET L_{eff} than p-FET L_{eff} when considering identical doping profiles. In order to investigate the intrinsic transport properties we have constructed "mirror -image" n- and p-FET devices whose (I_{ds},V_{gs}) properties were simulated using FIELDAY. The L_{eff} and series-resistance were then extracted using the Shift-and-Ratio algorithm [3]. Essentially, a generic FET was created and then given doping profiles that were described by Gaussian functions for the deep-source/drain and extension profiles. The well was fairly flat and had a low-10^{17} cm^{-3} type surface concentration. In order to make the n- and p-FETs used in our simulations, we interchanged only the majority/minority types throughout the device. In this way, the shapes of all the profiles in the devices remained the same amongst these devices. Hence the issue of graded versus abrupt extension profiles could be removed from the analysis, and the differences due to the intrinsic transport properties of electrons and holes could be revealed.

As the source/drain profiles become more abrupt the spatial distance over which N(x) varies by a factor of two becomes smaller. Correspondingly, equation (4) implies that the L_{eff} difference for n-FETs and p-FETs becomes smaller. In contrast, if the source/drain profiles are very graded one would expect larger differences in L_{eff}. We have simulated two such cases wherein the lateral grading of the extension profile was set using a σ_x parameter to designate the spreading of the Gaussian function that described the profile. The table below shows the simulation results from the two cases. Figure 3 shows the source profiles along a cut taken at the interface with the origin located at the gate side-wall. Note that $\Delta L = L_{gate} - L_{eff}$ differs significantly only for the p-FET. This implies that the actual value of the ratio $N(X_s)^{pfet}/N(X_s)^{nfet}$

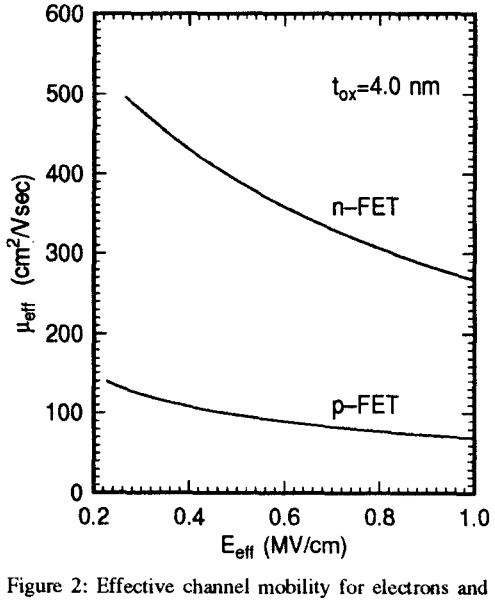

Figure 2: Effective channel mobility for electrons and holes as a function of effective-field normal to the Si/SiO_2 interface.

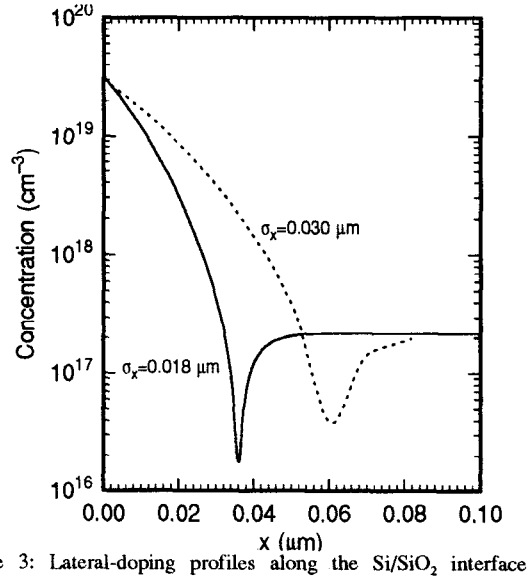

Figure 3: Lateral-doping profiles along the Si/SiO_2 interface for Gaussian extensions having σ_x of 0.18 and 0.03 µm.

Table 1: n-FET and p-FET effective-channel lengths and series resistances for devices having identical profile shapes.

Device	σ_x (μm)	L_{gate}(μm)	L_{eff} (μm)	R_{sd} (Ω μm)
p-FET	0.018	0.2	0.140	496
n-FET	0.018	0.2	0.173	335
p-FET	0.03	0.2	0.107	658
n-FET	0.03	0.2	0.175	334

obtained from simulation is much smaller than the rough estimate given in equation 4 based upon the high-doping limit for $\mu_{s/d}$.

Further insight into the situation may be gained through examining the channel/source-drain sheet resistance, ρ(x). The sheet resistance is related to the L_{eff} through the integral relationship:

$$R_{TOTAL} = \frac{V_{DS}}{I_{DS}} = \int \rho\ (x)\ dx \qquad \ldots (5)$$

where x is in the horizontal direction going from source to drain. Figure 4, below, shows the sheet rho evaluated at several values of gate voltage for a p-FET. The junction grading was modeled using a lateral σ_x of 0.03. There are two limiting cases: first, in the channel the sheet rho forms a roughly constant plateau, second, in the source/drain the sheet rho becomes comparatively small and is gate-voltage invariant. The figure shows three reference lines which indicate the gate edge (solid line), the boundary of L_{eff} (short dash), and the metallurgical junction (long dash). Note that at the L_{eff} boundary, the sheet rho varies substantially at the lower gate voltage values but less so at higher voltages (say $V_{GS} > 2.0$ V). This indicates that our L_{eff} extraction is most sensitive to the high V_{GS}, high current, operating condition. In Figure 5, we shown the same sheet resistance plot but for a n-FET device having identical profile shapes. An immediate, quantitative difference is expectedly seen in that the peak

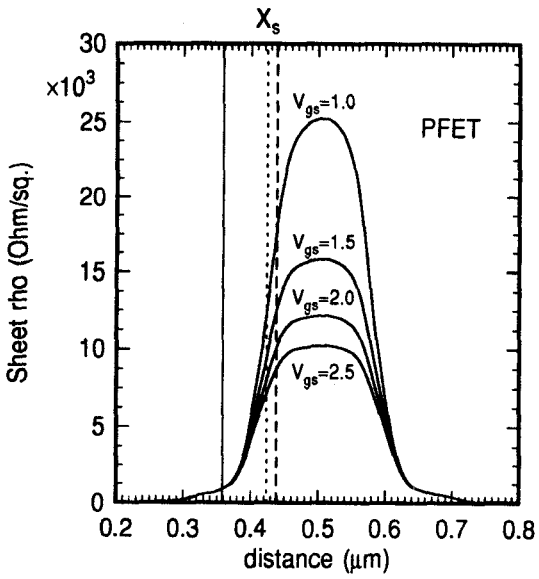

Figure 4: p-FET sheet resistance versus distance from source to drain evaluated at different gate voltages.

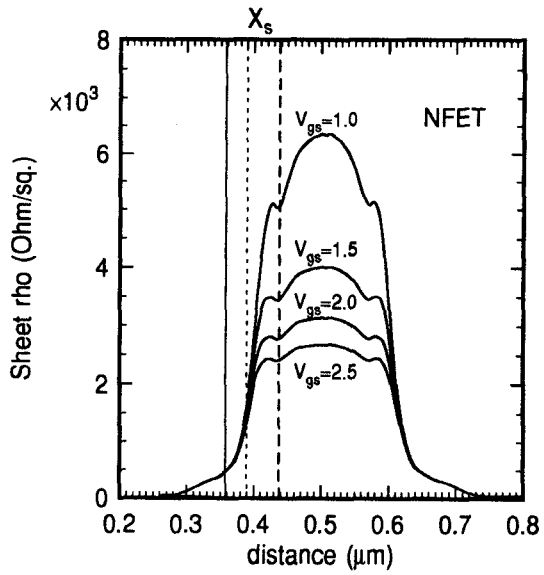

Figure 5: n-FET sheet resistance versus distance from source to drain evaluated at different gate voltages.

values are substantially lower than the p-FET counterpart. This difference scales with the effective mobility. In addition, however, there is also a qualitative difference in the shape of the curves in the transition region between source to channel: the n-FET curves do not display the same, smooth, monotonic decrease from channel to source as do the p-FET curves. Setting aside approximations made for the sheet rho in the 2-d current flow region (common to both cases), a further difference is identified below in Figure 6. Here a plot is shown of the carrier mobilities, evaluated at the semiconductor-insulator interface, along a segment of the device going from the channel into the drain with the position of the metallurgical junction shown (L_{met}). Notice that in the n-FET case there exists a relatively large degradation of mobility once the metallurgical junction is crossed. In contrast, the p-FET mobility varies only slightly over the same distance. Since sheet resistance is inversely proportional to mobility, the sudden drop in μ tends to locally elevate the sheet rho. On the other hand, the doping rises rapidly in going from channel to drain which tends to reduce sheet rho. The net effect of these competing mechanisms is the small oscillation in sheet rho seen in going from channel to source/drain (Figure 5). For the p-FET, such competing effects are absent since the mobility variation over the same region is almost negligible. Hence one observes the smooth, monotonically decreasing sheet rho displayed in Figure 4. We believe that it is this basic difference in mobility variation that gives rise to a qualitatively different sheet rho in the vicinity of the metallurgical junction.

Note, also, that the difference in sheet resistance between n- and p-FET devices studied here is consistent with a longer n-FET L_{eff}. The shoulders on the n-FET sheet rho curves give rise to a larger gate voltage dependence for sheet resistance in the vicinity of L_{met} than is seen for the p-FET in the same location. Once the doping begins to limit the sheet rho, the observed drop is more abrupt compared with the decrease in sheet rho seen for the p-FET. Since the gate substantially modulates the sheet rho over a larger distance in the source/drain for the n-FET, the resultant L_{eff} is longer than that calculated for the p-FET having identical profiles.

III. CONCLUSIONS.

In summary, we have shown that there is a fundamental asymmetry in the effective-channel lengths of n-FET and p-FET devices. Basic differences in electron and hole transport give rise to a larger L_{eff} for n-FETs. This can also be observed through a comparison of the device sheet resistance where the relative difference in channel to source/drain mobility leads to distinct sheet rho patterns near the metallurgical junction. Recent investigations in current flow patterns in the source/drain regions (to be reported elsewhere) has further substantiated these findings.

The grading of the extension profile can further enhance the channel-length asymmetry. In terms of the simplified transport model discussed above, for a given difference in $N(X_s)^{nfet}$ to $N(X_s)^{pfet}$, the spatial separation of these points increases as the junction becomes more graded. Of course, in actual CMOS processes differences in extension dopant diffusivity between n- and p-FETs provides one source of effective-channel length asymmetry. In addition to this, however, the tendency for p-FET to have larger ΔL is thus not only driven by the higher-dopant diffusivity but also by inherent differences in electron and hole mobilities.

IV. REFERENCES.

1. Yuan Taur, Yuh-Jier Mii, Ron Logan, and Hon-Sum Wong, "On Effective-Channel Length in 0.1-μm MOSFETs", IEEE Elect. Dev. Lett. vol. 16, pp. 136-138, April 1995.

2. E. Crabbé, R. Logan, J. Snare, P. Agnello, and J. Sun, "Anomalous Short-Channel Effects in 0.1 μm MOSFETs", IEDM Tech. Dig., pp. 571-574, 1996.

3.Yuan Taur et al., "A New Shift-and-Ratio method for MOSFET-channel length extraction", IEEE Elec. Dev. Lett., vol. 13, pp.267-269, May 1992.

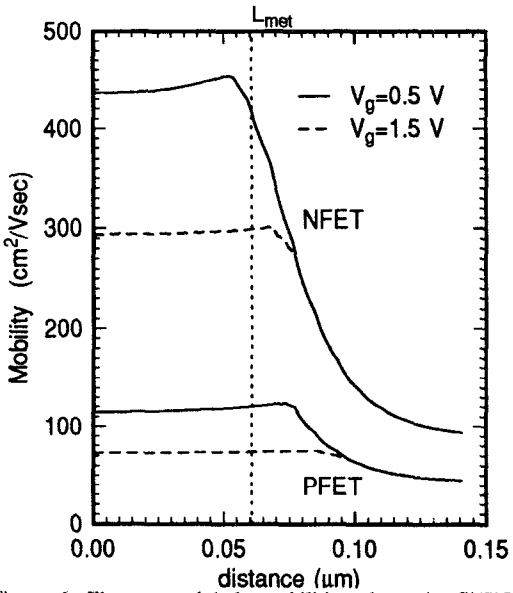

Figure 6: Electron and hole mobilities along the Si/SiO_2 interface going from channel (left) into the drain (right).