Halo Doping: Physical Effects and Compact Modeling

S. Mudanai, R. Rios, W. Shih, P. Packan and S.W. Lee

Abstract

In this paper, we present the physical effects observed when halo implants are used to control short channel effects (SCE). The impact of using halo implants was studied using a 2 or 3 segmented transistor. The multi-segmented transistor model approach show that (1) the mobility needed to match IV data shows an artificial length dependence, (2) the threshold voltage needed to match the IV data is different from the threshold voltage needed to match CV data and (3) the NQS effects are more severe for halo transistors. Two dimensional numerical devices simulations were also performed to comprehend the observed output resistance degradation. Based on these numerical simulations a 2-transistor model is proposed to describe output resistance in compact models.

1. Introduction

As the devices are scaled to ultra short channel lengths, pocket or halo implants have been used widely to reduce Drain Induced Barrier Lowering (DIBL) and other short channel effects [1-3]. Although, the use of halo implants helps with control of the short transistors in a given technology, the performance of the analog transistors which are typically longer is degraded severely because of these implants. The reason being that transistors with halo implants exhibit reduced output resistance [6,7] and long channel drain induced barrier lowering [8]. This issue is becoming more relevant with the increase in integration of analog and RF components with digital circuits on the same chip [4,5]. Halo implants also impact the extraction of mobility for long channel transistors, aside from causing the well known reverse short channel effect [9]. The compact modeling equation for drain current is usually derived assuming that the doping is uniform in the channel. Hence, the effects of laterally and vertically varying dopants are typically accounted for by the use of dimension and bias dependent doping [9-11]. However, these models do not capture other halo effects, such as output resistance degradation. Compact modeling of these halo effects require a comprehensive understanding of how the implants affect the output resistance, dibl and threshold voltages. In this paper we focus on the physics of output resistance degradation due to halo implants. The analysis presents for the first time a comprehensive explanation of the output resistance degradation due to halo implants.

2. Multi-segmented model description

The effect of halo doping on device characteristics can be explained by splitting the device into three transistors in series, with the end transistors having a higher threshold voltage while the middle transistor has a lower threshold voltage.

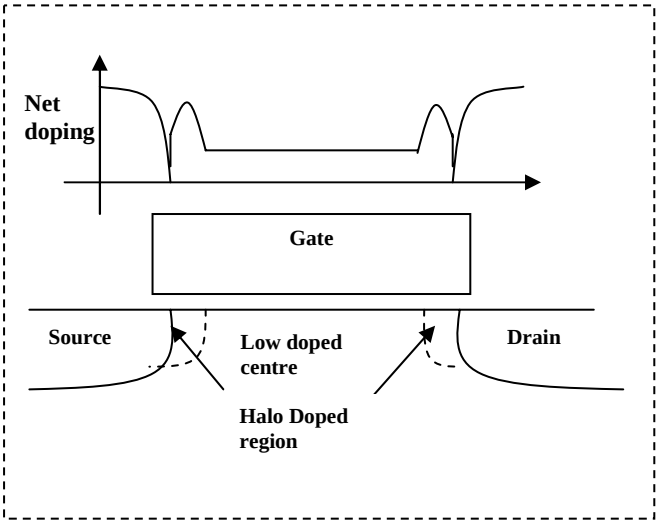


Figure 1: Illustration of the doping profile in a halo doped long channel device.

In Fig 2. a composite transistor of channel length 1um with a halo doping of 2e+18cm^{-3} over a length of 0.1um and a bulk doping 3e+17cm^{-3} is considered. The linear region threshold voltage of the composite transistor is matched to that of a uniformly doped transistor at 1.3e+18cm^{-3}. For the sake of comparison the IV curves of uniformly doped devices with just the bulk doping and halo doping are also shown.

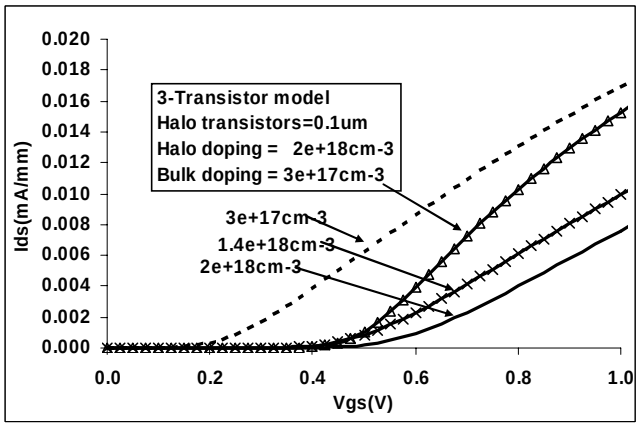


Figure 2: The halo doped transistor has a higher linear region current than the uniformly doped transistor but the same threshold voltage.

As described earlier [9] the apparent mobility enhancement for the composite 3-Transistor is due to the fact that in strong inversion, the lower doped middle region

dominates and increases the effective channel conductance. This means that during the parameter extraction process, a higher mobility is needed for the longer channel transistors. Also, as the gate length is reduced the halo regions from the source and drain regions begin to merge. As a result the artificial increase in mobility needed will reduce as a function of channel length.

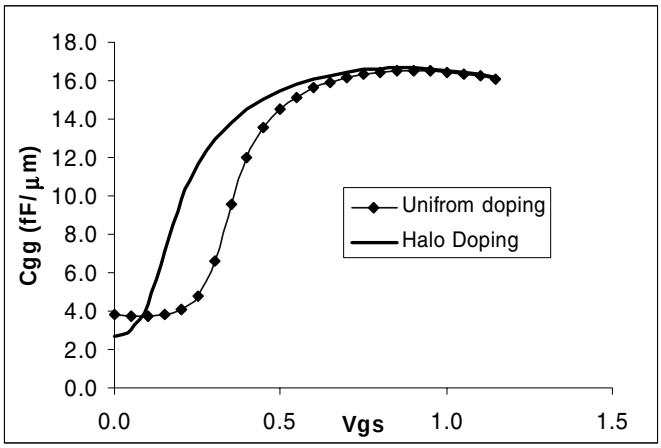


Figure 3: The halo transistor exhibits a lower Vth since the center of the channel inverts at a lower gate bias.

Another outcome of having a lower V_{th} for the middle segment is that the center starts building inversion charge before the overall transistor starts conduction. This is because the end transistors are still in sub-threshold and prevent current conduction. This implies that the threshold voltage for the IV and CV are different. This behavior is illustrated in Fig. 3, where the gate capacitance for two devices with different doping profiles but with the same drain current based threshold voltage. A simple and elegant approach to handle this difference in V_{th} between IV and CV was described in [9].

The CV curves described above were obtained at low frequency or under quasi-static conditions. At higher frequencies, though, the high V_{th} and hence higher channel resistance regions at the source and drain ends of a halo doped transistor will impede inversion layer formation

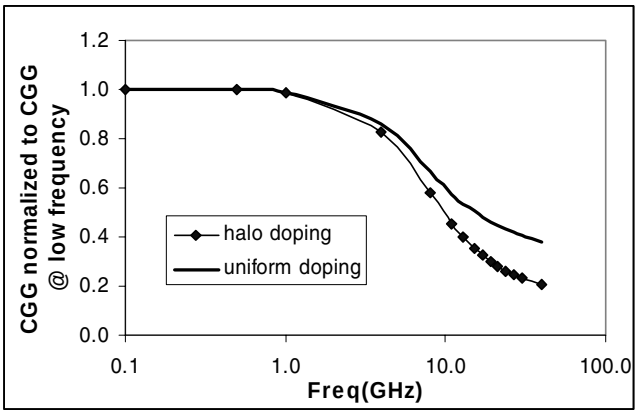


Figure 4: The halo doped transistor shows worse NQS behavior at VGS = 0.6.

. In figure 4, this phenomenon can be observed from the CGG vs frequency simulations using a 2-D device simulator. The higher channel resistance of the halo aggravates the NQS effect at higher frequencies

3. Understanding Rout degradation

2-D numerical device simulations were performed for two devices: (A) with uniform doping of 9.7e17 cm^{-3} and (B) with a laterally non-uniform doping using a bulk doping of 4e17 cm^{-3} and a Gaussian halo with a peak of 3.9e18 cm^{-3}. Figure 3, shows the id vs. vds curves for the devices A and B biased at the same gate voltage. The linear region current is higher for the halo doped device, as explained earlier. It is also distinctly clear that the output resistance is much lower for device B. An interesting point to observe is that although both devices were designed to have the same threshold voltage, the device B pinches off sooner than the uniformly doped device A. Both devices are long enough that the drain saturation occurs because of channel pinch off. Under pinch off conditions the drain saturation voltage, V_{dsat}, is approximately given by the expression:

$$V_{dsat} = \frac{V_{gs} - V_t}{\alpha} \qquad (1)$$

$$\alpha = 1 + \frac{dQ_b}{d\phi_s} \qquad (2)$$

In the above equations V_t stands for the threshold voltage of the device, Q_b is the depletion charge normalized to the gate capacitance and ϕ_s is the surface potential.

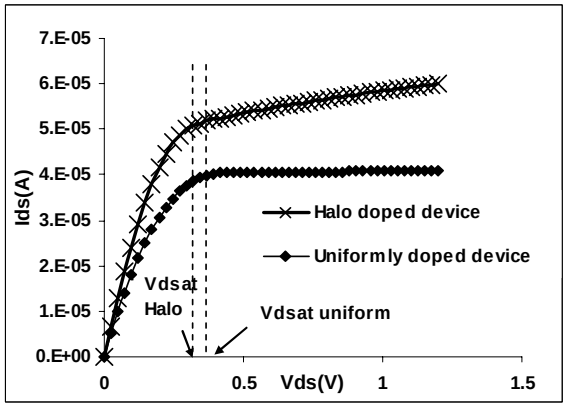


Figure 3: Ids vs. Vds curves at Vgs = 0.6V for the halo doped device and the uniformly doped device. The Vdsat for the halo doped device is less than the uniformly doped device. The parameters used for the device simulation are: W=1um, L =1um, tox =1.2nm.

It should be noted that the Vt used in Equation (1) is the threshold voltage corresponding to the doping at the drain end of the channel. So even though the overall threshold voltage of the devices A and B are the same, the threshold voltage corresponding to the halo regions of device B is much higher than the uniformly doped regions of device A. Thus, device B pinches off sooner than device A because the threshold voltage corresponding to the

heavily doped halo region on the drain side is higher than that of the threshold voltage corresponding to the uniform doping in device A. In other words, as the drain bias is increased, the drain end of the channel for device B pinches off earlier because it requires a higher vertical field to support inversion. Furthermore, α for the heavily doped halo region is greater than the α for the uniformly doped region. Thus the impact of α is to further reduce the V_{dsat} for device B.

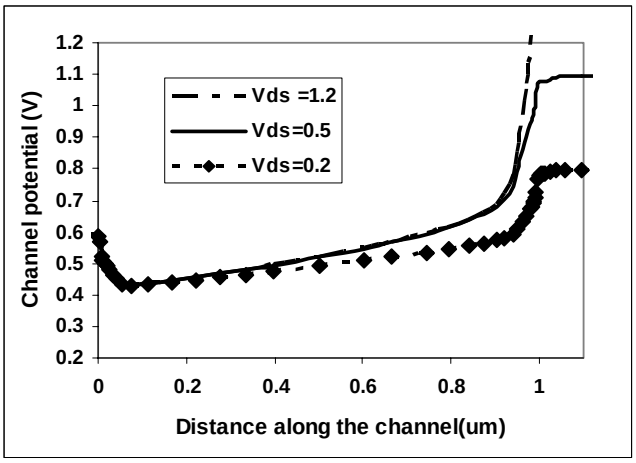


Figure 4a: Channel potential along the surface of a uniformly doped device at Vgs = 0.6V, Vbs=0.0V, Vds = 1.2V, 0.5V, 0.2V.

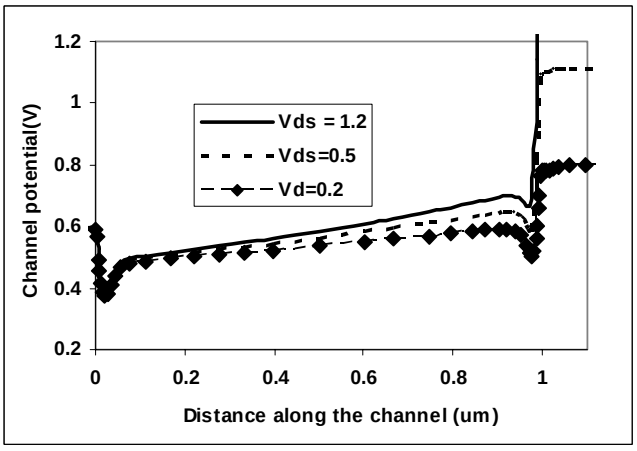


Figure 4b: Channel potential along the surface of a non-uniformly doped device at Vgs = 0.6V, Vbs=0.0V, Vds = 1.2V, 0.5V, 0.2V.

The potential along the channel is shown for both devices at multiple drain biases, in Figure 4. In Figure 4a, potential along the channel at 3 different drain biases is shown for the uniformly doped device(A). At the lowest drain bias the transistor is still in the linear region, while at the other two drain biases the transistor is in saturation. Comparing the different curves in Figure 4a, we observe that once the transistor reaches saturation there is no further drop in a significant portion of the channel and, as expected, the additional drain bias is dropped in the small pinch off region close to the drain. Thus, in saturation, the current does not increase for the uniformly doped device. In Figure 4b, the potential along the channel is shown for the non-uniformly doped device (B) at two different drain bias, both of which are in saturation. It is interesting to see that "additional" potential drop continues to occur along the entire channel length, even in saturation. Thus transistor (B) behaves almost as if it were still a resistor in the saturation region.

This anomalous behavior can be readily explained using a *two* transistor model. As previously observed [6,9] only the halo doped region on the drain side influences the Rout degradation. Hence, the channel can be split into two transistors in series, namely the lower doped transistor of length L-L_h on the source side and halo doped transistor of length L_h on the drain side as illustrated in Figure 5a. As the drain bias is increased the transistor operates in the linear region until the heavily doped halo transistor saturates. However, the lower doped transistor continues to operate in the linear region. Thus, when the halo doped transistor is operating in saturation, the following drops in potential occur in the following different regions of the transistor:

1. Vds – $Vdsat_{halo}$ is dropped in the pinch-off region of the halo transistor (Region III in Figure 5b),
2. Vdsathalo – Vd' is dropped in the linear region of the halo transistor (Region II in Figure 5), and
3. Vd' is dropped in low dopped transistor (Region I in Figure 5).

With increasing V_{ds}, the pinch off point moves closer to the lower doped region. This causes a reduction in the length of region II and thus reducing the effective resistance of region II. Since, $V_{dsatHalo}$ is dropped across both region I and region II, a larger value of Vd' must now be dropped across region I causing an increase in the drain current. This effect thus reduces the Rout of the halo doped transistor in saturation.

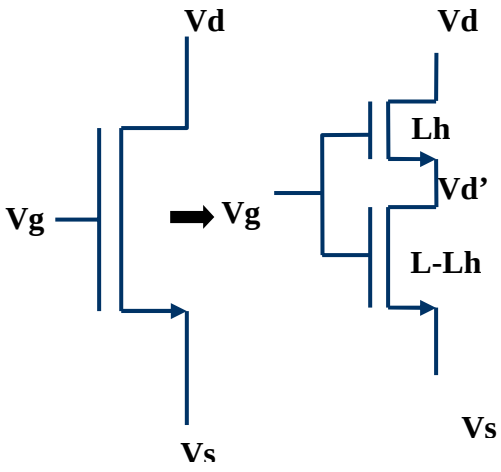


Figure 5a: Illustration of the two transistor model. The non uniform transistor can be modeled as two transistors in

series. The source side transistor is uniformly doped with Cbulk, while the drain side transistor is uniformly doped with Chalo.

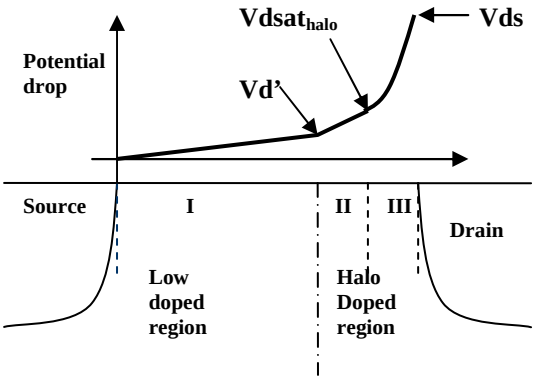
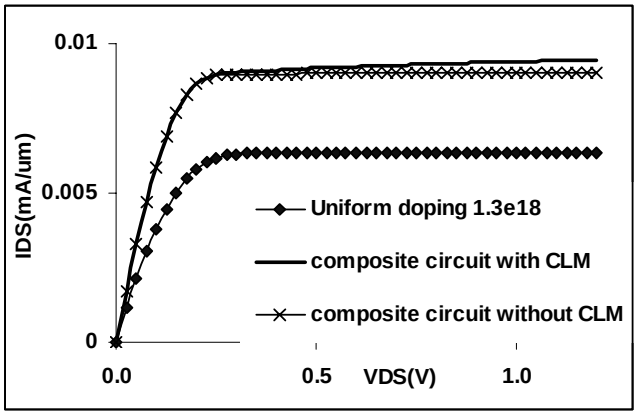


Figure 5b: Based on the operation, the halo doped transistor can be split in to 3 different regions.

4. Verification

In order to verify the output resistance degradation mechanism proposed in the previous section, circuit simulations were performed using the two–transistor composite circuit. The gate bias was fixed at 0.7V and the drain bias was swept on the composite circuit and the equivalently doped transistor. The highly doped halo transistors in the composite were simulated with the channel length modulation (CLM) model turned on and off. The results are shown in Figure 6a and 6b.



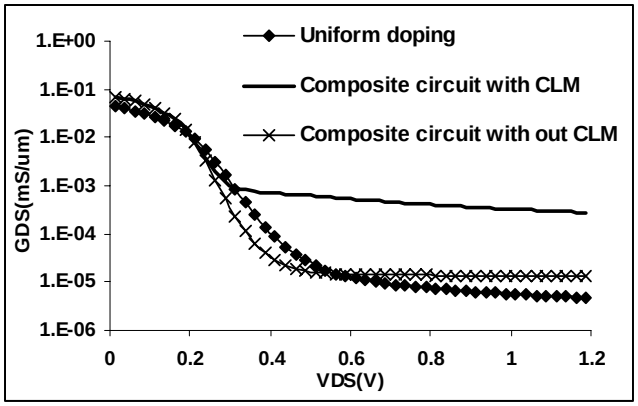


Figure 6: Inclusion of CLM in the drain-side halo transistor clearly shows the output resistance degradation in the IV and output conductance curves.

It can be observed that the composite circuit with CLM qualitatively captures the output resistance degradation observed, while the composite model without CLM reproduces does not capture the output resistance degradation. The plot in Figure 7 verifies that the node voltage Vd' is indeed low enough that the bulk transistor operates in the linear region and that the inclusion of CLM model causes Vd' to increase with increasing drain bias.

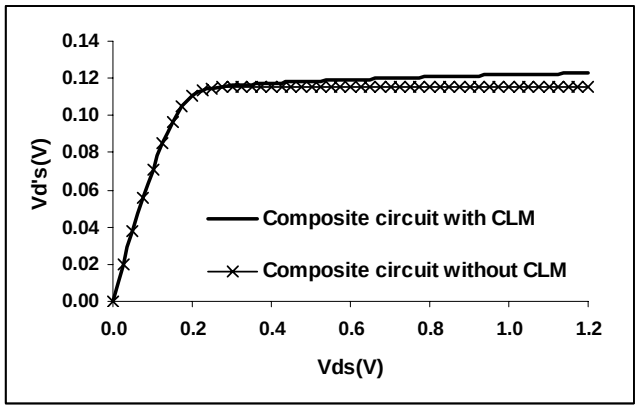


Figure 7: The voltage at the node between the low doped transistor and drain-side halo transistor increases in even in the saturation region of the transistor validating the model.

5. Conclusion

The degradation of Rout for long channel devices has been explained using 2-D numerical device simulation. A physical argument was presented to help understand how the presence of a potential barrier on the drain side impacts the channel resistance using a simple 2-transistor model. Also, the impact of halo doping on mobility values used during a compact model parameter extraction was described.

References

[1] T. Hori, IEDM Tech. Dig., pp. 75-78, 1994.
[2] Y. Okumura, et al., IEDM Tech. Dig., pp. 391 - 394, 1990.
[3]Y. Taur et al., IEDM Tech Dig., pp. 215 – 218,1997.
[4] D. Buss, IEDM Tech. Dig., pp. 423-425,1999.
[5] R. F. M. Roes et al., ESSDERC, pp. 176 – 179, 1999.
[6] K. M. Cao et al., IEDM Tech. Dig., pp. 171 – 174, 1999.
[7] A. Chatterjee et al., Proc. VLSI Symp., pp. 147, 1999.
[8] H. Ueno et al., IEEE TED, Vol. 49, No. 10, pp. 1783-1789, 2002.
[9] R. Rios et al., IEDM Tech. Dig., pp. 113 – 116, 2002.
[10] BSIM4 User's manual, www-device.eecs.berkeley.edu
[11] X. Zhou et. al, IEEE TED, vol. 47, pp. 214, January 2000.