High Mobility Channel from the Prospective of Random Telegraph Noise

K.P. Cheung[*], J.P. Campbell
Semiconductor Electronics Division, NIST, Gaithersburg, MD 20899, 301-975-3093, *kin.cheung@nist.gov

A. Oates
TSMC Ltd., Hsin-Chu, Taiwan 300-77, R.O.C.,

***Abstract*---We experimentally verify for the first time that random telegraph noise (RTN) in ultra-scaled MOSFETs is related to the inversion charge density in the channel. We then examine the merit of high mobility channel devices from the RTN prospective. This analysis strongly suggests that RTN is a serious obstacle for high mobility channel adoption.**

I. INTRODUCTION

High mobility channel devices are under intense investigation for future CMOS technology. They offer the promise of vastly increased I_{ON} headroom and present possible solutions to future technology nodes. However, the merits of these high mobility channel materials have recently been challenged by several research groups [1-3]. This has led to much debate in the literature. In this work, we examine another unexplored aspect of the debate, namely the impact of RTN. RTN has recently been shown to produce dynamic threshold voltage (V_{TH}) variation worse than the random dopant effect in silicon devices [4]. It is, therefore, important to examine the role of RTN in high mobility channel devices.

The main motive for the adoption of high mobility channel materials is to reduce the power supply voltage. In other words, to obtain the same circuit performance at a lower gate overdrives. This is achieved via higher carrier velocities, which lowers the required carrier density and, consequently, the gate overdrive for a given current. In this work, we will show that the lower carrier density enabled by high mobility channel materials is a recipe for unacceptably high RTN.

II KEY RESULTS

Figure 1 shows the key result of this work.

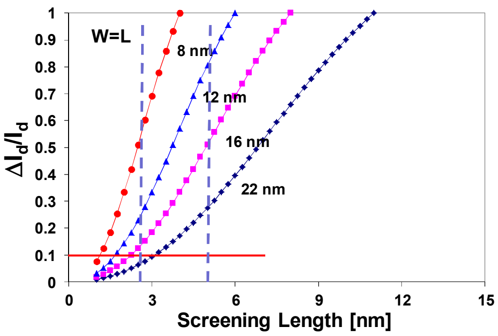

Figure 1 Worst-case RTN amplitude as a function of screening length for various technology nodes. The first broken vertical line mark the experimentally determined screening length for silicon. The second broken vertical line doubles the screening length as a result of reducing the inversion charge density by a factor of 4.

Data in figure 1 are for the worst-case RTN amplitude at strong inversion extrapolated from measured results. Worst-case refers to the tail of the RTN amplitude distribution (large number of devices). There are two main messages. One is that the worst-case RTN amplitude for the silicon MOSFET (first broken vertical line at 2.48 nm screening length) is beyond 10% at the 16 nm node. Second message is that the RTN amplitude becomes extreme when the screening length increases due to the decrease in inversion charge density, as would be the case for high mobility channels. The 2.48 nm screening length for the silicon MOSFET is experimentally extracted using a "hole-in-the-inversion-layer" model. It is an approximate model. However, even if the exact value of the screening length is off, only the values on the horizontal axis will change, and the basic conclusion of figure 1 will still hold. The key question is whether the "hole-in-the-inversion-layer" model is valid or not. We will show that it is a valid model.

III RTN and SCREENING

It is commonly believed that low frequency (1/f) drain current noise is the superposition of many RTN fluctuations. Thus, RTN is often attributed to same mechanisms responsible for the low-frequency drain current noise, namely mobility fluctuation, number fluctuation, or both. However, for very small device dimensions, a third mechanism may be more important [5, 6]. This third mechanism, first proposed by Reimbold [7], suggests that the trapped charge creates a "hole" (or cored-out section) in the inversion layer through Coulombic repulsion (figure 2).

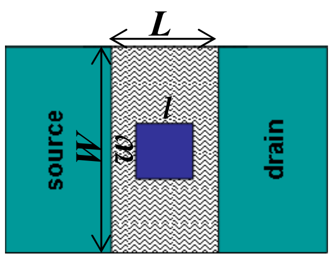

Fig. 2 Defect creates a hole (dark blue square) of the size w x l in the inversion layer.

This idea was further refined by Ohata *et al.* [8] (Equation (1)) in which the RTN amplitude is related to the hole size (approximated as a square):

$$\frac{\Delta I_d}{I_d} = \frac{wl(\sigma_0 - \sigma)}{WL(\sigma_0 - \sigma) - w(L-l)(\sigma_0 - \sigma)} \quad (1),$$

where w and l define the area of the hole, and σ_0 and σ are channel conductivities outside and inside the hole, respectively. In this construct, the screening length determines the size of the hole. Assuming a square shape hole makes the derivation of the equation simpler, but also clearly makes this only a crude approximation. The best attribute of this model is that it is highly intuitive. One could argue that this "hole-in-the-inversion-layer" model is just an extreme case of mobility fluctuation. Nevertheless, it is quite intuitive and allows one to model it in a straight forward way [8].

The intuitive nature of this "hole-in-the-inversion-layer" model allows a direct test to be carried out, and unfortunately yields (at low gate overdrive) screening lengths which seem unreasonably large [8]. At that time, it was not known that RTN amplitudes at low gate overdrives can become abnormally large due to the formation of a channel percolation path [9]. Basically, the effective channel width at lower gate overdrives is reduced by the channel's uneven potential landscape. From (1), it is clear that an over estimation of the hole size will result from measured RTN amplitude. The percolation idea resolves Ohata *et al's* difficulty and justifies a revisit of the "hole-in-the-inversion-layer" model, and its implications on high mobility channel devices.

IV PERCOLATION PATH & SCREENING LENGTH

In this study, we utilize nMOSFETs with 1.4 nm SiON gate dielectrics. RTN amplitudes from many highly-scaled devices were investigated. The observed RTN amplitude as a function of gate overdrive is typically non-monotonic (Fig. 3). This behavior is expected since it is unlikely that a single defect is physically located precisely over the "pinch-off" point in the potential landscape.

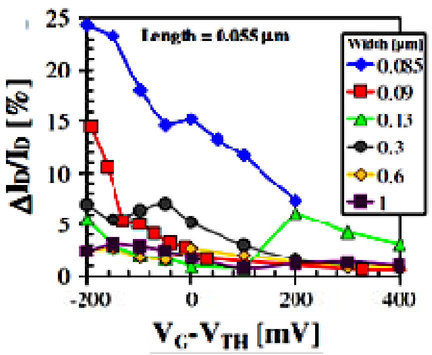

Fig. 3 RTN amplitudes as a function of gate overdrive for a number devices with different width. Typical trends are non-monotonic. Device sizes are as drawn.

However, we were able to observe one device with particularly large RTN amplitude as well as a monotonic dependence on gate overdrive (fig. 4). This device produced very clean 2-level RTN (fig. 5), implying a single defect. This behavior strongly suggests that the defect is positioned entirely within the active region of the channel, and is located right on top of the percolation path.

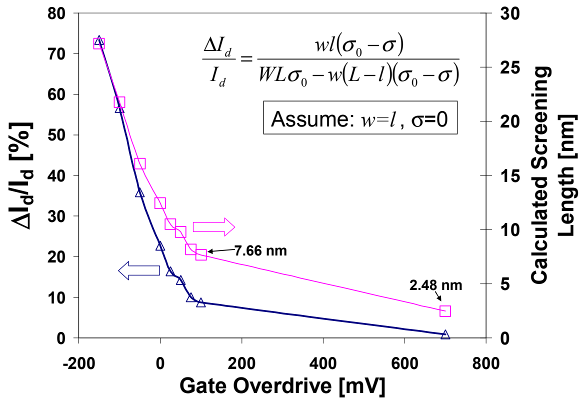

Fig. 4 RTN amplitude (triangles) for this particular device is a smooth monotonic function of gate overdrive. The calculated screening lengths drops rapidly as gate overdrive increase from below threshold to above threshold. The as drawn size of this device is 55 nm x 85 nm. The size used for the calculation is 35 nm x 85 nm.

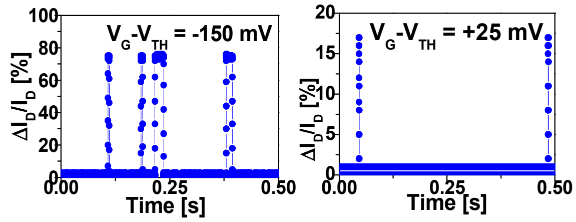

Fig. 5 The RTN signal from the device shown in fig. 3 are very large and very clean. Showing the characteristic of a single defect.

Also shown in fig. 4 are the screening lengths calculated from the observed RTN amplitudes with the assumption that the conductivity within the hole is zero (completely cored out). At low gate overdrives, the screening length increases rapidly. It was the similarly large screening lengths led Ohata *et al.* [8] to doubt the "hole-in-the-inversion–layer" model. As discussed earlier, this is the result of a percolation path induced abnormally small effective channel width, leading to an artificially large screening length.

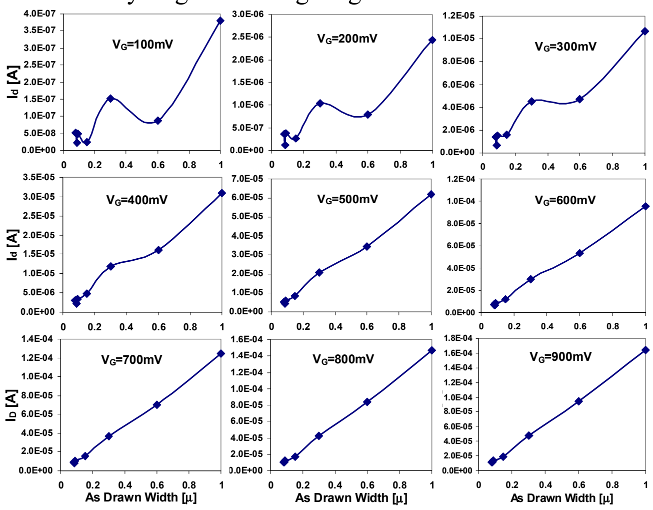

Fig. 6 Drain current as a function of device width for 55 nm (drawn) nMOSFET at various gate overdrives. Deviation from linear relationship is a sign of percolation path which makes the effective channel width smaller than the drawn width, sometimes by a larger factor. The percolation path is expected to disappear at some point above threshold. In this example, some hint of it is still there at V_G=600 mV, which is 300 mV above threshold.

As the gate overdrive increases, the percolation path widens and eventually disappears. This suggests that the screening length behavior depicted in fig. 4 is largely due to the widening of the percolation path under the hole (this includes the +150 mV gate overdrive point). To lend support to this assertion, we study the relationship between drain current and channel width (fig. 6). It is clear that a linear relationship is not established until at least +200 mV above threshold. It is also clear that the last point in fig. 4 (+700 mV) is completely free of the percolation effect. Thus the screening length of 2.45 nm for this point can be related to the channel size by (1) for the inversion charge density at the power supply of 1 V (strong inversion).

A 2.48 nm screening length at such strong inversion still seems large (Debye length (a 3-D model) ~ 1 nm). However, since the inversion layer is a 2-D system, screening length is expected to be larger than the 3-D value [10]. In addition, as we mentioned before, equation (1) is only a crude approximation. Thus we argue that the result is quite reasonable and that the model is physically correct.

V IMPLICATION to HIGH MOBILITY CHANNEL

Now we can apply the hole-in-the-inversion-layer model to ultra-scaled devices, and especially high mobility channels. Here we consider only the conditions at which percolation is absent (its presence will make matters worse). Making use of the fact that screening length is inversely proportional to the square root of the free carrier density, we let $w = l = p/\sqrt{n}$ where n is inversion charge density, and p is proportionality constant, we get

$$\frac{\Delta I_d}{I_d} = \frac{p^2}{n\left(WL - \frac{L}{p\sqrt{n}} + p^2 n\right)} \qquad (2)$$

Recall that $n \cong C_{OX,eff}(V_G - V_{TH})$ [11], where $C_{OX,eff}$ is the effective inversion capacitance. If we assume n remains more or less constant with scaling (oxide capacitance increase is offset by gate overdrive decrease), we can use equation (2) to extrapolate to smaller device sizes. Since we can anchor the equation with a known experimental data point (RTN amplitude and device size), the extrapolation should be on solid ground.

In fig. 7, we show the projected RTN amplitude for various technology nodes (for silicon MOSFET). Three different channel widths are simulated (W=L, W=3L, and W=5L). If we set the acceptable RTN amplitude to 10%, then the minimum width of $W = L$ can only be used in the 22 nm node. At 16 nm node, it is marginal. Wider channel widths do help. At $W = 5L$, the RTN amplitude is within 10% down to the 8 nm node.

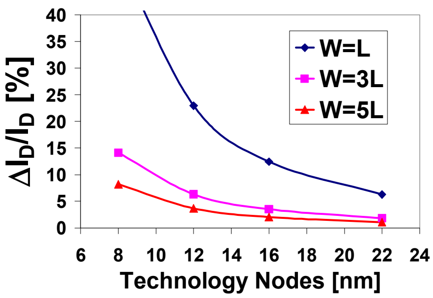

Fig. 7 RTN amplitude at different technology nodes for minimum width ($W = L$) devices as well as wider devices ($W = 3L$, and $W = 5L$).

Note that the extrapolation result in fig. 7 represents the worst case scenario – the defect is right at the center of the active area. This is very likely the tail of the RTN amplitude distribution. To experimentally observe this, the sample size must be large, or one gets lucky as in our case. Since we cannot ignore the tail of the distribution, the results presented in figure 7 are instructional.

Fig. 7 reinforces the conclusion that RTN is a new limiting factor for scaling [4]. Note that in [4], the conclusion was based on V_{TH} fluctuation that was exacerbated by the Random Dopant Effect (RDE), which causes the formation of percolation path. Because the RDE has been removed in the data, our conclusion is stronger. It suggests that while going to an undoped channel may alleviate RDE, RTN will remain a huge factor that can cause large V_{TH} or equivalently drain current fluctuations.

Turning our attention to high mobility channels, remember that the goal of high mobility channels is to reduce the inversion charge needed for the same I_{ON}. Depending on the material, this reduction can be higher than a factor of 10. Since the screening length is proportional to the square root of the dielectric constant divided by the inversion charge density n, a reduction in n can have grave consequences (Dielectric constants mostly get a little bit larger, but, the impact is relatively small).

In fig. 1, we showed the impact of screening length on RTN amplitude for various technology nodes. The two vertical dotted lines mark the silicon screening length of 2.48 nm as well as twice this value 4.96 nm which we take as the high mobility channel value. Although doubling is very conservative for many high mobility materials, the impact is clearly seen. If we set the acceptable RTN amplitude to 10% (horizontal solid line), then only silicon can use the minimum width of W=L in the 22 nm node. Doubling the screening length in this case leads to almost 30% RTN – clearly unacceptable.

These results show that if we use high mobility channels to lower the required gate overdrive to achieve the same I_{ON}, we will drastically increase the drain current noise while reducing the noise margin – a lethal combination.

VI CONCLUSIONS

We show that the “hole-in-the-inversion-layer model” together with a percolation path reduced channel width works reasonably well to explain RTN in ultra-small MOSFETs. This establishes the relationship between RTN amplitude and the inversion charge density. We argue that the physics of the model makes sense, and even though the absolute value of the extracted screening length is only a crude approximation, the relationship between RTN amplitude and inversion charge density is valid. Using this relation, we show that high mobility channel will drive up the RTN amplitude to unacceptable levels in future technology nodes.

REFERENCES

[1] Saraswat, K. C., C. O. Chui, D. Kim, *et al.* "High Mobility Materials and Novel Device Structures for High Performance Nanoscale MOSFETs." IEDM '06., pp1-4.
[2] Lubow, A., S. Ismail-Beigi and T. P. Ma, "Comparison of drive currents in metal-oxide-semiconductor field-effect transistors made of Si, Ge, GaAs, InGaAs, and InAs channels." *Appl. Phys. Lett.* **96**(12) 122105-3(2010).
[3] Skotnicki, T. and F. Boeuf, "HOW CAN HIGH MOBILITY CHANNEL MATERIALS BOOST OR DEGRADE PERFORMANCE IN ADVANCED CMOS." VLSI Technol. Symp., 2010 pp153.
[4] Tega, N., H. Miki, F. Pagette, *et al.* "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm." Symp. VLSI Technol., 2009, pp50-51.
[5] Simoen, E., B. Dierickx, C. L. Claeys and G. J. Declerck, "Explaining the amplitude of RTS noise in submicrometer MOSFETs." *IEEE Trans. Electron Dev.,* **39**(2) 422-429(1992).
[6] Vandamme, L. K. J., D. Sodini and Z. Gingl, "On the anomalous behavior of the relative amplitude of RTS noise." *Solid-State Electronics* **42**(6) 901-905(1998).
[7] Reimbold, G. "Modified 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion - Influence of interface states." *IEEE Trans. Electron Dev.,* **31**(9) 1190-1198(1984).
[8] Ohata, A., A. Toriumi, M. Iwase and K. Natori, "Observation of random telegraph signals: Anomalous nature of defects at the Si/SiO2 interface." *J. Appl. Phys.* **68**(1) 200-204(1990).
[9] Asenov, A., R. Balasubramaniam, A. R. Brown, *et al.* "Random telegraph signal amplitudes in sub 100 nm (decanano) MOSFETs: a 3D `Atomistic' simulation study." IEDM-2000, pp279-282.
[10] Ando, T., A. B. Fowler and F. Stern, "Electronic properties of two-dimensional systems." *Rev. Mod. Phys.* **54**(2) 437 LP - 672(1982).
[11] King, Y.-C., H. Fujioka, S. Kamohara, *et al.* "Dc electrical oxide thickness model for quantization of the inversion layer in MOSFETs." *Semiconductor Sci. & Technol.* **13**(8) 963(1998).