The Amplitude of Random Telegraph Noise: Scaling Implications

Kin P. Cheung[1], J.P. Campbell[1], S. Potbhare[1], A. Oates[2]
1.National Institute of Standards & technology, Gaithersburg, USA; 2. TSMC, Hsinchu, Taiwan

Abstract: We introduce a simple and intuitive model to relate the amplitude of random telegraph noise (RTN) fluctuations to the columbic influence of single trap charges on the inversion layer. The prediction of this model is in excellent agreement with results extracted from experiment using the "hole-in-the-inversion-layer" model for RTN amplitude. This new model allows us to quantitatively examine the impact of "worst-case" RTN in future scaling nodes.

Random telegraph noise (RTN) is expected to surpass the random dopant effect as the main cause of threshold voltage (V_{TH}) variation at the 22 nm node and beyond [1]. While the physical origins of drain current RTN has recently garnered much attention [2-4], the majority of these discourses attempt to explain the time dynamics with very little regard for the size of the RTN fluctuations. However, these large RTN fluctuations actually represent the biggest threat to continued CMOS scaling. In this paper we verify the "hole-in-the-inversion-layer" model (proposed over 40 years ago) for RTN amplitude with a simple physical model. This model is then used to examine the RTN implications on future scaling nodes.

Historically, drain current 1/f noise is linked to the superposition of the RTN distributions across many frequencies and amplitudes. Thus, the number vs. or mobility fluctuation debate which permeates the 1/f literatures also applies to RTN. Amidst these complicated theoretical treatments, a simple model was proposed by Yau *et. al.* [5]. In this model a trapped charge in the gate dielectric layer (or at its interface) creates a low carrier density "hole-in-the-inversion-layer." Thus, trapping and/or de-trapping of this charge results in drain current fluctuations (RTN). Reimbold [6] extended this proposition to the case of the "hole" is completely free of inversion charges (i.e. cored out). Ohata *et. al.* [7] introduced a simple equation to link the size of this cored-out hole to the RTN amplitude. However, when applied to their experimental observations, they found that the size of the cored-out hole was much bigger than the screening length expected for the inversion charge density. Simoen *et. al.* [8] further pointed out that the model cannot account for the huge range of amplitudes often observed in experiments.

Recently, we revisited this model and showed [9, 10] that the difficulty encountered by Ohata *et. al.* lies in the comparison of experimental data collected at low gate overdrives. At low overdrives, the "effective" channel width can be significantly smaller than the drawn width due to the random dopant-induced percolation path [11]. A similar explanation clearly applies to the amplitude range problem pointed out by Simoen *et. al.* We showed that these experimental oddities can be avoided by measuring RTN amplitude at high gate overdrives. However, our results still point to cored-out hole size a few times larger than the expected screening lengths. While the screening length in a quasi-2D inversion layer is supposed to be larger than the calculated value based on a 3-D model [12], we are still left with unexplained discrepancies. In this work, we provide a simple way to estimate the size of the cored-out hole. The result is quite consistent with the "hole-in-the-inversion-layer" model for RTN. We then provide an explanation for the discrepancy between the size of the cored-out hole and the screening length. This new insight allows us to examine the RTN implications to scaling more clearly.

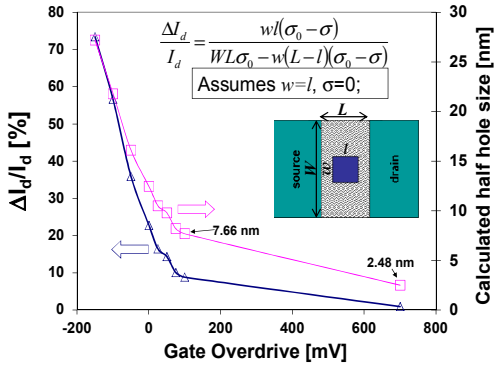

Fig. 1 Measured RTN amplitude (triangles) for this particular device is a smooth monotonic function of gate overdrive. The calculated half cored-hole size drops rapidly as gate overdrive increase from below threshold to above threshold. The as drawn size of this device is 55 nm x 85 nm. The size used for te calculation is 35 nm x 85 nm. The inserts are the cored-hole model and the associated equation of Ohata *et al*.'s.

Our discussion centers on a particularly useful set of RTN data shown in fig. 1. It is particularly useful in the sense that the RTN magnitude drops smoothly with increasing gate overdrive, indicating that the

cored-out hole is entirely within the channel not near any edges [9, 10]. The calculated half cored-out hole size (radius) for each data point is also shown. We have previously shown that the cored-out hole sizes at low gate overdrives are affected by the percolation effect [9, 10], and that the data point at 700 mV gate overdrive is almost completely free of this effect. Therefore, this point is an excellent candidate to calibrate the model.

At 700 mV gate overdrive, this device (1.2nm EOT for gate dielectric) should reach ≈10^{13}/cm^2 charge density. The calculated screening length (3D) is ~ 1nm (much smaller than the 2.48 nm extracted from the RTN data). In the following, we develop a simple, intuitive model to show that the 2.48 nm half cored-hole size is actually the correct value.

Our simple model starts with the standard way to calculate V_{TH} shift due to trapped charges Q:

$$\Delta V_{TH} = \frac{Qd_1}{C_{inv}(d_1 + d_2)} \qquad (1)$$

where d_1 and d_2 are the distances of the charge centroid from the gate and from the inversion layer respectively in equivalent oxide thickness. When the charges are at the interface, d_1 is the oxide thickness and d_2 is the dark space (region between the interface and charge centroid due to quantum confinement). Implicitly, we assume that the charges are spread out evenly across the device and the inversion layer remains uniform with reduced density (fig. 2(a)).

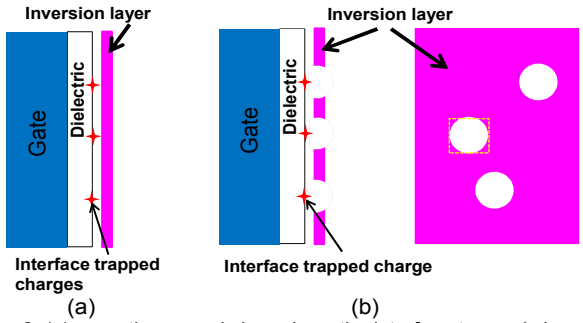

Fig. 2 (a) sows the normal view where the interface trapped charges affect the inversion layer density but not the uniformity. (b) shows the actual picture where each trapped charge affect the inversion layer density locally, leading to cored-out holes in the inversion layer. The broken square enclosing a cored-hole is the calculated size of the equare device so that the V_{TH} shift equals the gate overdrive.

In reality, each charge is highly localized with its influence on the inversion layer schematically shown in fig. 2(b). Over shorter distances, the electric field associated with the point charge is extremely high. Therefore, it is reasonable to expect that the region surrounding the trapped charge to become accumulated. This results in a continuum of inversion charge interrupted by "cored-out" regions in the inversion layer.

One can determine the size of the cored-out region by using (1) to calculate the area of a device in which a single charge generates a V_{TH} shift equal to the gate overdrive. To more easily compare this calculated result to those experimentally derived from RTN data (fig. 3), we assume a square (W=L) device geometry. Taking the dark space for silicon channel to be 0.4 nm EOT, we can calculate the size of the half core-out hole for the case of our 1.2 nm EOT. The calculated result is shown in fig. 3 (blue line). Note the excellent agreement between the calculation and the experimentally derived values. The 0.1 V gate overdrive data point is somewhat higher than the calculated value. However, this is likely due to residual percolation effects as reported earlier [9, 10]. Thus, this correspondence provides the "hole-in-the-inversion-layer" model of RTN a solid physical foundation.

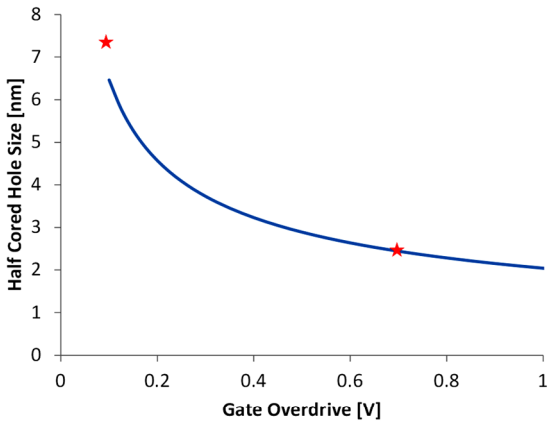

Fig. 3 Calculated half cored-hole size as a function of gate overdrive (solid curve) and the half cored-hole size extracted from RTN amplitude (star) (see fig. 1) show excellent agreement.

With the physical insight, we can now answer why the extracted half cored-out hole size is so much greater than the calculated screening length. The key is in the definition of screening length. Screening length is defined as the distance at which the potential of the charge is reduced by 63%. As we pointed out earlier, the electric field of the point charge at short distance is very large. A decrease of 63% is still very large. The cored-out hole is defined

as the point at which the local field is at threshold. This could be as little as 5% of the potential, or three times the screening length.

This new simple physical model allows us to examine RTN amplitudes more quantitatively. For example, the cored-out hole (thus RTN amplitude) for an interface charge is reduced with oxide thickness (fig. 4), as expected from screening considerations. In [9], we projected the RTN amplitude for future devices without considering the effect of thinner gate dielectric, and as a result greatly exaggerated the problem.

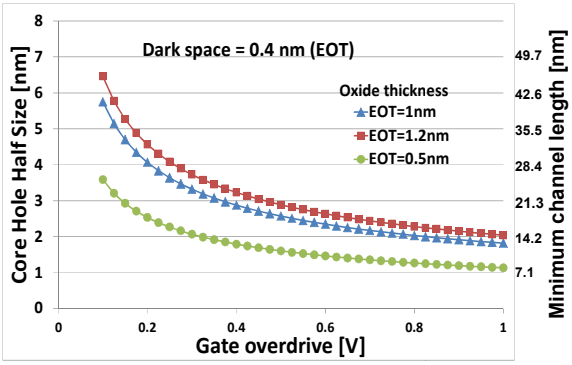

Fig. 4 Half cored-hole size as a function of gate overdrive for three different oxide thickness showing the effect of gate screening.

From the equation in the inset of fig.1, we can calculate the device size below which the RTN amplitude will be greater than 10%. Assuming W=L, we have L = 7.1 times the half cored-hole size. From fig. 4, we see that with EOT = 0.5 nm, the 8 nm channel length requires gate overdrive at least 1 V.

In [10], we reported that using a high-mobility channel to lower the gate overdrive will lead to unacceptably high RTN amplitude. These projections are less severe (fig. 5) once the gate dielectric thickness and the increased dark space (0.65 nm instead of 0.4 nm) are accounted for.

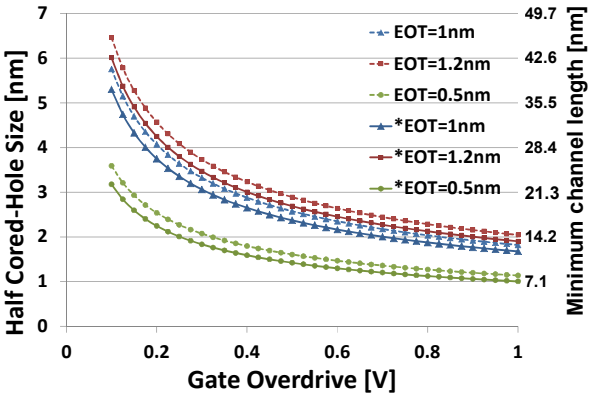

Fig. 5 Half cored-hole size as a function of gate overdrive for three different oxide thickness and for 0.4 nm dark space (broken line) as well as for 0.65 nm dark space (solid line).

In conclusion, we demonstrate a simple model that not only confirms the "hole-in-the-inversion-layer" model for RTN, but also allows for a quantitative prediction of RTN amplitude (worst case) in future devices.

References

[1] Tega, N. *et al.* Symp. VLSI Technol. 2009, p50.
[2] Campbell, J. P. *et al.* IRPS 2009, p382.
[3] Grasser, T. *et al.* IEDM 2009, p729.
[4] Veksler, D. *et al.* IRPS 2010, p73.
[5] Yau, L. D *et al.* TED 16(2): 170-177 (1969).
[6] Reimbold, G. TED 31(9): 1190-1198 (1984).
[7] Chata, A., *et. al.* (1990). J. Appl. Phys. 68(1): 200-204(1990).
[8] Simoen, E., *et al.* TED 39(2): 422-429 (1992).
[9] Cheung, K. P. *et al.* ICICDT-2011.
[10] Cheung, K. P. *et al.* ESSDERC-2011, p335.
[11] Asenov, A *et. al.* IEDM-2000, p279.
[12] Ando *et. al.* Rev. Mod. Phys. 54(2), 437 LP – 672(1982).