Zhang, Z., Zhang, Z., Guo, S., Wang, R., Wang, X., Cheng, B., Asenov, A. and Huang, R. (2016) Investigation on the Amplitude of Random Telegraph Noise (RTN) in Nanoscale MOSFETs: Scaling Limit of "Hole in the Inversion Layer" Model. In: 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), Hangzhou, China, 25-28 Oct 2016, pp. 453-455. ISBN 9781467397193 (doi:10.1109/ICSICT.2016.7998949)

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/149202/

Deposited on: 03 October 2017

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Investigation on the Amplitude of Random Telegraph Noise (RTN) in Nanoscale MOSFETs
−*Scaling Limit of "Hole in the Inversion Layer" Model*

Zexuan Zhang[1], Zhe Zhang[1], Shaofeng Guo[1], Runsheng Wang[1*], Xingsheng Wang[2], Binjie Cheng[2], Asen Asenov[2,3], Ru Huang[1]

[1]Institute of Microelectronics, Peking University, Beijing 100871, China ([*]Email: r.wang@pku.edu.cn)
[2]Synopsys, 11 Somerset place, Glasgow G3 7JT, UK
[3]School of Engineering, University of Glasgow, Glasgow G12 8LT, UK

Abstract

In this paper, the widely adopted "hole in the inversion layer" (HIL) model for predicting the amplitude of random telegraph noise (RTN) in nanoscale MOSFETs, is theoretically revisited with focusing on its scaling limit and validation range. It is found that this simple physical model fail to apply on ultra-scaled devices with L<20nm and/or W<10nm, due to the non-negligible impact from source/drain and the failure of assumed equivalence to resistor network in ultra-scaled devices. This work provides a deeper understanding to this model and is helpful for accurate prediction of RTN amplitude in nanoscale devices and circuits.

1. Introduction

The random telegraph noise (RTN) in nanoscale MOS devices and circuits is one of the critical reliability concerns for 16/14nm node and beyond. Its amplitude increases rapidly with device scaling and is therefore arousing more and more attention nowadays. Many attempts have been made to predict RTN amplitude in nanoscale MOSFETs so far [1-12]. Among them, what is worth mentioning is the "hole in the inversion layer" (HIL) model firstly proposed by Yau *et al.* [6]. Since then, several modifications or improvements have been made on this simple physical model [7-8]. Recently, Cheung *et al.* revisited this model and proposed a quantitative method to define the "hole" size [9, 10, 12]. The application of this new version of HIL model in relatively large device (channel size much larger than the "hole" size) have been verified through both experiments and TCAD simulations [9-18].

As CMOS technology will continue to scale down beyond 10nm node in the near future, the device size might be comparable or even smaller than the "hole" size. Therefore, the applicability of this model for ultra-scaled devices should be carefully examined. In this paper, the RTN amplitudes in nanoscale MOSFETs are investigated with 3D 'atomistic' device simulations. The scaling limit of the HIL model and the underlying physics are discussed in great detail.

2. "Hole in the inversion layer" (HIL) model revisited

In this model, carrier trapped in gate oxide defect states is treated as a point charge. The charge will apparently induce a local V_{th} shift in a circle area, thus decrease the electron density and conductivity in this region. For case of simplicity, at the first-order approximation, this model assumes that the channel can be divided into two parts, as shown in Fig. 1: in one part the conductivity degrades to zero, like a "hole" in the inversion layer induced by the trapped charge; while the conductivity remains undisturbed outside the "hole".

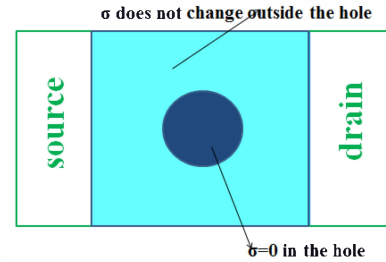

Fig. 1 Schematic diagram of the "hole in the inversion layer" (HIL) model. It is assumed the conductivity reduces to zero inside the hole and remain unaffected outside the hole.

Then the RTN amplitude ($\Delta I_d/I_d$) can be gained by solving an equivalent resistor network [4,5,9-12]:

$$\frac{\Delta \mathrm{Id}}{\mathrm{Id}} = \frac{4r^2}{(L-2r)(W-2r)+2Wr} \qquad (1)$$

where r represents the radius of the "hole", L represents channel length and W represents channel width.

3. Simulation results and analysis

The GSS 3D 'atomistic' TCAD simulator GARAND is applied here with careful calibrations [19]. For simple examine of the HIL model, single trap is located at the middle of the channel in each NMOS device with channel doping of 2×18cm^{-3} and EOT of 0.8nm. In the following simulations, V_d is kept at 0.05V.

The electron density around the mid-channel before (named "fresh device" here) and after the trapping of an electron are extracted. As shown in Fig. 2, the trapped electron induces a region where the electron density decreases drastically, consistent with the expectation from HIL model.

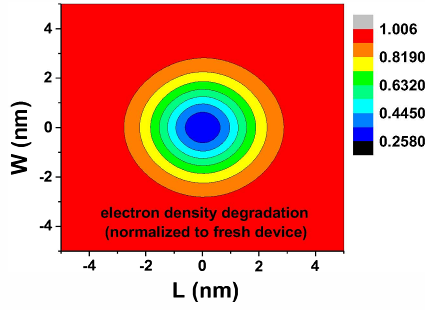

Fig. 2 The relative degradation of electron density (normalized to fresh device) around the trap location (in the middle) in a device with L=25nm and W=25nm at V_g=0.1V. Note that the L/W coordinates means the L/W direction (also in Figs. 6-9).

The RTN amplitude ($\Delta I_d/I_d$) can be extracted under different V_g and different channel length, as shown in Fig. 3. The voltage dependence is consistent with HIL model, however, interestingly RTN amplitudes in ultra-scaled devices show a non-monotonic dependence on channel length, which disagrees with the prediction of the HIL model.

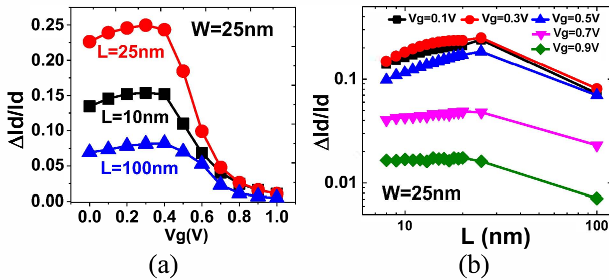

Fig. 3 RTN amplitude dependence on gate voltages (a) and channel length (b). It is worth noting that amplitude shows a non-monotonic *L* dependence in ultra-scaled devices.

Based on the value of $\Delta I_d/I_d$, one can obtain the "hole" radius using Eq. (1). For example, $\Delta I_d/I_d$=23% below threshold in the device with L=25nm and W=25nm, thus the hole size can be extracted as r=5.2nm. The extracted hole radius (@V_g=0.1V) against channel length are shown in Fig. 4. It can be seen that the hole radius remain a constant for L≥25nm, while showing almost a linear decrease with channel length when L<20nm.

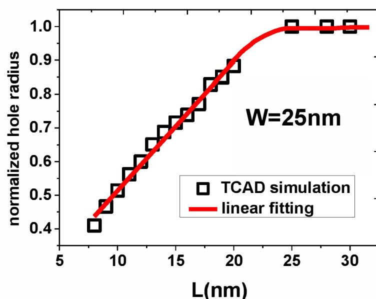

Fig. 4 The extracted hole radius (normalized) against channel length. The linear fitting in sub-20nm device is also shown.

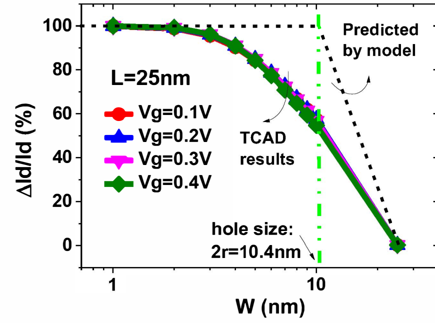

Fig. 5 RTN amplitude dependence on channel width. L is kept at 25nm. $\Delta I_d/I_d$ is far less than 100% as predicted by HIL model when W<2r, until extreme narrow W.

For devices whose width is less than 2r (r=5.2nm in large devices), the HIL model suggests a 100% RTN amplitude, which disagrees with simulation results, as shown in Fig. 5.

4. Limiting factors in HIL model

A. Impact of source/drain on "hole" size

It is believed that in sub-threshold region the electron density is very low, the trapped charge is electrostatic interacted or screened mainly by the gate and the Si body, thus will show little dependence on channel length. It is true in long devices, as shown in Fig. 4, which is consistent with [14].

However, in very short devices, the impact of the heavily doped source/drain cannot be neglected as was done in long devices. The high density of electrons in source/drain region in vicinity will have apparent impact on the electrostatics and thus affect the "hole" size. As shown in Fig. 6, the normalized electron density degradation is almost the same in longer devices with L≥25nm. However, in the device with L=10nm, the trapped charge impact on channel electron density is weaker, due to non-negligible impact from source/drain. Therefore, we cannot use a universally defined "hole" radius (from long devices) in the HIL model. Otherwise, the RTN amplitude would be largely overestimated in ultra-scaled devices.

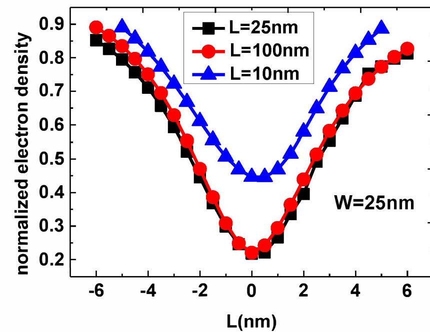

Fig. 6 Electron density degradation (normalized to fresh device) around the trap location in devices at V_g=0.1V.

B. Failure of the equivalence to resistor network

Another important issue is that, HIL model obtains the RTN amplitude by solving the equivalent network. In fact, a long channel can be divided into three parts: undisturbed region near the source, disturbed region around the trap and undisturbed region near the drain.

In the undisturbed the current transmits along the channel, which can be solved by the equivalent resistor network; while in the disturbed region, the current has component in the width direction which can hardly be expressed by the resistor network. What one can do is to equal the conductivity of the whole disturbed region to a resistive network with a cored out hole, as shown in Fig. 7.

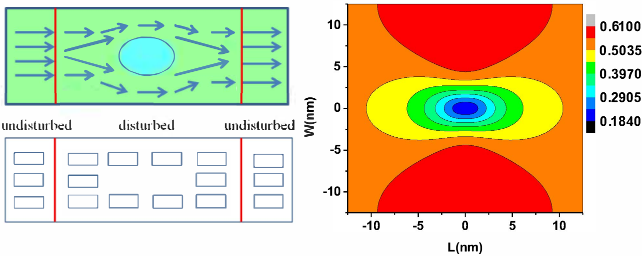

Fig. 7 (Left) Schematic picture of a long channel device and its equivalent resistor network, whose channel can be divided into three parts. (Right) Current density degradation (normalized to fresh device) in a device with L=25nm and W=25nm.

However, if the channel is not long enough to include the entire disturbed region, as is the case in a device with L=10nm shown in Fig. 8. One cannot assume the conductivity of the channel is the same as the equivalent resistor network. In such case, as the HIL model assumes the conductivity to be zero inside the hole, it will underestimate the conductivity inside the hole and thus overestimate the RTN amplitude. This factor also contributes to the decrease of the "hole" radius in very short devices.

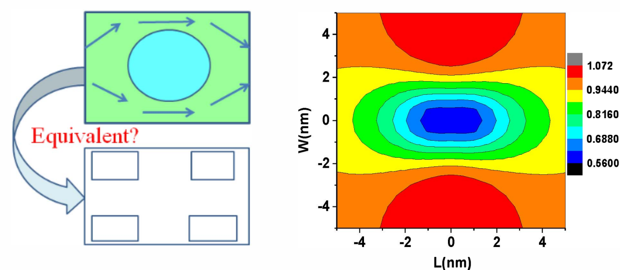

Fig. 8 (Left) Schematic picture of a short channel device and its "equivalent" resistor network, whose channel fails to include the entire disturbed region. (Right) Normalized current density degradation in a device with L=10nm and W=25nm.

5. Practical scaling limit of HIL model

As mentioned above, the channel size should be large enough to: (1) neglect the impact of source/drain (2) cover the entire disturbed region. Thus, one can estimate the practical scaling limit of channel length considering the second criterion, shown in Fig. 9. The channel should be longer than about 20nm below threshold and about 18nm in strong inversion for planar devices.

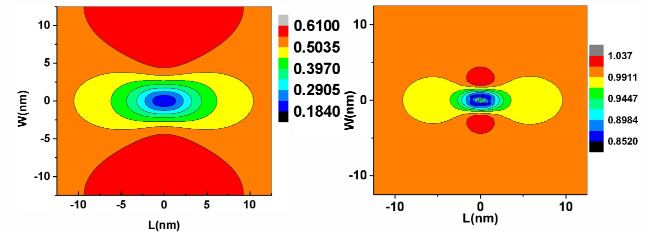

Fig. 9 Normalized current density degradation in a device at V_g=0.1V (Left) and V_g=0.8V (Right).

In addition, the second criterion also sets a limit on width direction for a planar device. From Fig. 9, it could be W>10nm below threshold and W>6nm in strong inversion, which is also consistent with Fig. 5.

6. Summary

Based on 3D 'atomistic' device simulations, RTN amplitudes in devices of different sizes is extracted. Then the widely used "hole in the inversion layer" model for predicting RTN amplitudes is examined. It is found there is a scaling limit of this simple physical HIL model, beyond which the hole radius can no longer be defined the same way as the model does in large devices. Otherwise, the RTN amplitude would be overestimated. Practical validation range for HIL model is for devices with L>20nm and W>10nm, due to the non-negligible impact from source/drain and the failure of assumed equivalence to resistor network in ultra-scaled devices.

Acknowledgement: This work was partly supported by NSFC (61522402, 61421005) and 863 Project (2015AA016601). The authors would like to thank the helpful discussion with K. P. Cheung.

References: [1] C. T. Sah, ***PRL*** p.956 (1966). [2] S. Christensson, ***SSE*** p.797 (1968). [3] G. Abowitz, ***TED*** p.775 (1967). [4] F. Berz, ***SSE*** p.631 (1970). [5] F. N. Hooge, ***Phys. Lett.*** p.139 (1969). [6] L. D. Yau, ***TED*** p.170 (1969). [7] G. Reimbold, ***TED*** p.1190 (1984). [8] A. Ohata, ***JAP*** p.200 (1990). [9] K. P. Cheung, ***ICICDT*** p.1 (2011). [10] K. P. Cheung, ***IRPS*** p.GD1 (2012). [11] R. G. Southwick III, ***SNW*** p.147 (2012). [12] K. P. Cheung, ***ICSICT*** p.531 (2012). [13] E. Simoen, ***TED*** p. 422 (1992). [14] A. Asenov, ***TED*** p.839 (2003). [15] K. Sonoda, ***TED*** p.1918 (2007). [16] H. H. Mueller, ***JAP*** p.1734 (1997). [17] C. Liu, ***IRPS*** p.XT.17 (2014). [18] M. J. Chen, ***TED*** p.2495 (2014). [19] GARAND. http://www.goldstandardsimulations.com.