



HAL
open science

“Pinch to Detect”: A Method to Increase the Number of Detectable RTN Traps in Nano-scale MOSFETs

Angeliki Tataridou, Gerard Ghibaudo, Christoforos Theodorou

► To cite this version:

Angeliki Tataridou, Gerard Ghibaudo, Christoforos Theodorou. “Pinch to Detect”: A Method to Increase the Number of Detectable RTN Traps in Nano-scale MOSFETs. 2021 IEEE International Reliability Physics Symposium (IRPS), Mar 2021, Monterey, United States. 10.1109/IRPS46558.2021.9405102 . hal-03260919

HAL Id: hal-03260919

<https://hal.science/hal-03260919>

Submitted on 25 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

“Pinch to Detect”: A Method to Increase the Number of Detectable RTN Traps in Nano-scale MOSFETs

Angeliki Tataridou, Gérard Ghibaudo, Christoforos Theodorou

IMEP-LAHC, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, Grenoble INP*, Grenoble, 38000, France

*Institute of Engineering Univ. Grenoble Alpes

phone: +33 – 04-52-94-91, email: angeliki.tataridou@grenoble-inp.fr

Abstract— This work presents a new method, called “pinch to detect”, for the characterization of Random Telegraph Noise (RTN) in nanometer length Tri-Gate Nanowire MOSFETs, allowing for the maximum number of active RTN traps to be detected. We demonstrate that the channel pinch-off effect, which takes place in saturation region, can actually be exploited for the modulation of RTN amplitudes and kinetics, and therefore disable (or enable) the appearance of RTN signals which are present (or not) in linear region, where the channel is uniform. It is shown that by combining the measurements for three different channel configurations, i.e. uniform, pinched-off near the drain and pinched-off near the source, the total number of detectable RTN traps can be maximized. This conclusion is also supported by noise spectral analysis and TCAD simulations. Furthermore, it is demonstrated through how going to longer channels with no short-channel effects does not make this effect disappear, but just limit it to traps in the pinched-off region.

Index Terms -- Random telegraph noise, Tri-Gate Nanowire MOSFETs, Characterization, Short channel effects

I. INTRODUCTION

The Random Telegraph Noise (RTN) effect [1] has become an important performance limiting factor in nano-scale circuits [2], [3], due to its random nature and high amplitude. At the same time, it is also a very powerful tool for defect characterization and reliability assessment of MOSFETs. This is thanks to its discrete signal levels, which correspond to the trap states (empty or filled), thus enabling a direct evaluation of the number of electrically active oxide traps [4]. This becomes very crucial when performing statistical measurements that aim to extract the average number of traps per device, as well as the statistical distributions of RTN amplitudes and capture/emission time constants [5]. The precise extraction of the above characteristics is of critical importance, because the estimation of a circuit’s or a system’s dynamic performance (for memories [3]) or sensitivity (for sensors [6]) is heavily based on them.

A trap’s occupancy is strongly dependent on the carrier concentration [7] and thus on the gate voltage (V_g) bias. As a result, an RTN signal is usually present only for a narrow V_g range (few kT around Fermi level E_f). Therefore, in order to achieve a reliable trap number estimation, it is necessary to scan a wide gate voltage range, from weak to strong inversion, with a small step. As far as the drain voltage (V_d) is concerned, although there can also be a V_d dependence of the trap time constants [8], [9], as well as of the relative RTN

signal amplitude, $\Delta I_d/I_d$ [10], [11], this effect has been only used for trap localization purposes [4] and not for the detection of new traps, because it is assumed that all traps are already active regardless the V_d value. This might be true for channel lengths in the micro-meter region, which do not suffer from short channel effects, nevertheless in nanometer-length devices the charge uniformity from source (S) to drain (D) is significantly influenced by V_d . We claim that this effect may result in a very different RTN behavior that could presumably help identify more traps than in the case where measurements are done under only one drain voltage bias.

In this work, we explore the validity of the above hypothesis through experiments and simulations. After describing the technical characteristics of the studied devices and the measurement equipment, we show the first experimental observations that led us assume there is an important pinch-off effect on the RTN activity. Next, we demonstrate this effect through TCAD simulations and we propose -and apply- the “pinch to detect” method: a 3-part statistical measurement process that can enhance the number of detectable RTN traps across the wafer. Finally, since this phenomenon is closely related to short-channel effects, we explore its dependence on the channel length.

II. DEVICES UNDER TEST AND MEASUREMENT SET-UP

Tri-Gate Nanowire SOI p-channel FETs, fabricated by CEA-LETI [12], were chosen to be measured as a test vehicle for this explorative study. They incorporate a buried oxide of 145 nm thickness, a H = 11 nm film thickness and a 2 nm Equivalent Oxide Thickness (EOT) HfSiON gate dielectric with TiN metal gate. The measured devices have a channel width of $W=30$ nm and a channel length ranging from $L=10$ μm down to 10 nm, where the pinch-off effect and channel length modulation (CLM) are enhanced. Both static and time domain measurements of the drain current were performed using the programmable bias point probe system NOISYS [13], combined with a semi-auto 300 mm Cascade Micro Tech probe station.

III. EXPERIMENTAL OBSERVATIONS AND DISCUSSION

An example (Device #11) of I_d - V_g characteristics in linear and saturation region of operation where the impact of pinch-off is maximized is shown in Fig. 1. Time domain measurements of I_d at linear region ($V_d = -30$ mV) are shown in Fig. 2, for three gate bias voltages, ranging from weak to strong inversion. As can be observed in both the time series

and the histograms, no RTN signals were present. Nevertheless, this is not the case when the same device was biased in saturation regime, with the drain voltage at $V_d = -0.9$ V: at least one clear two-state RTN can be detected (see Fig. 3). Therefore, it becomes evident that the channel pinch-off near the drain has some kind of effect on the trap amplitude or kinetics (capture/emission time), such that the RTN signal becomes detectable. A possible explanation is that when such short channel devices go into saturation, the length of the pinched-off region becomes comparable to the total channel length, resulting in a radically different charge distribution.

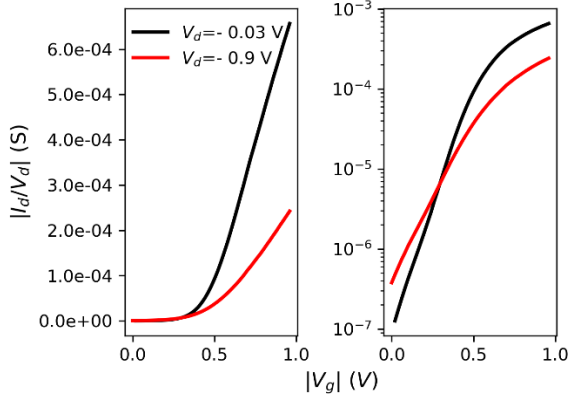


Figure 1. I_d - V_g characteristics in linear and logarithmic scale at linear and saturation region of operation.

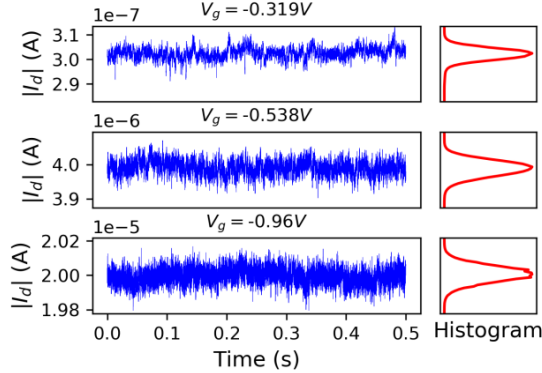


Figure 2. Device #11: Selected time-series and corresponding histograms for three gate voltage bias (from weak to strong inversion) in the linear region of operation ($V_d = -30$ mV).

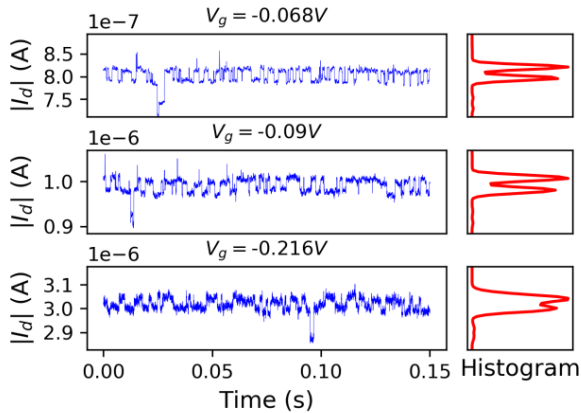


Figure 2. Device #11: Selected time-series and corresponding histograms for three gate voltage bias in the saturation region ($V_d = -0.9$ V).

On one hand, this can result in an increase of $\Delta I_d/I_d$ with V_d [6], [10], [11], which could explain how easier it is to detect the trap, since the RTN amplitude becomes more important than the background noise. On the other hand, the trap's capture time is also directly dependent on the local inversion charge density near the trap, Q_{it} , through:

$$\tau_c = \frac{q}{\sigma f_e Q_{it}} \quad (1)$$

with σ the trap cross-section and f_e the carrier escape frequency [14]. Therefore, since Q_{it} is affected by V_d in saturation, the trap occupancy can drastically change, especially if the trap is located near the drain region, resulting in the possible appearance of traps that were undetectable at low V_d (linear regime).

The validity of this idea was examined through TCAD simulations in Silvaco ATLAS with the structure's 2-D cross-section as well as its dimensions shown in Fig. 4. In Fig. 5, we demonstrate how going from linear to saturation region can dramatically affect the carrier concentration in more than half the channel length, especially close to the drain. Combining these results with (1), the impact of V_d on an interface trap's τ_c was estimated, for three different positions across the channel (see Fig. 6), at $V_g = 0.5$ V (above V_t). Clearly, the trap kinetics may be affected by even two orders of magnitude when the trap is located inside the pinched-off region. This means that in the case of a very fast trap close to the drain, measurements under high V_d would increase the capture time of the trap, allowing for its detection for a given sampling time and measurement window.

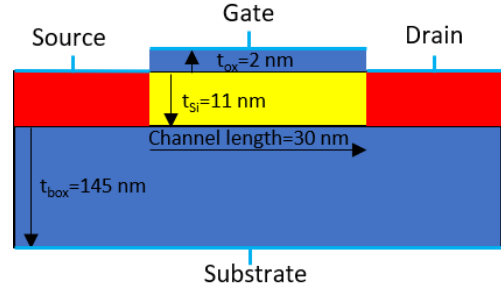


Figure 4: Simulated structure in Silvaco ATLAS TCAD.

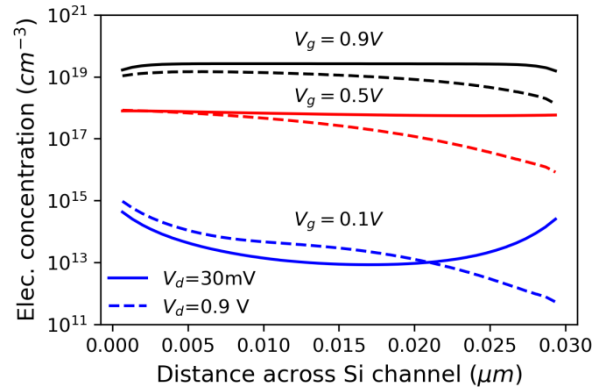


Figure 5. Simulated electron concentration across Si channel from Source ($x = 0$) to Drain ($x = 30$ nm) and 0.12 nm from the oxide/silicon interface for three gate bias and both linear ($V_d = 30$ mV) and saturation ($V_d = 0.9$ V) regimes.

On the other hand, an already slow trap in the linear region of operation close to the drain would be nearly impossible to be detected in the saturation region from the drain side. This is also true when the trap is close to the source, in which case measurements under high V_s (with grounded D) should be performed. Thus, in some cases, measurements at the saturation region of operation might lead to an easier detection of a trap.

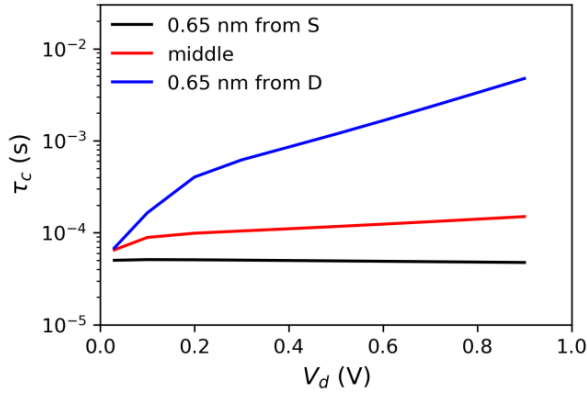


Figure 6. Calculated trap capture time using (1) and the simulation results of Fig. 5 for 3 different trap positions inside channel and 0.12 nm distance from the Si/SiO₂ interface.

Regarding the RTN signal amplitude $\Delta I_d/I_d$, its dependence on V_d is directly related to the position of the trap. In the saturation region from the drain side, the pinch-off phenomenon becomes dominant for a trap close to the drain, and it might be undetectable due to its very weak RTN amplitude. This phenomenon is shown in Fig. 7. Two different positions of the trap were considered: one at the middle of the channel and one with a 2.5 nm distance from the drain region. Moreover, two bias conditions were examined: the typical one with varying V_d and a complementary with varying V_s and grounded D. It is shown that when the trap is in the middle of the channel, the amplitude $\Delta I_d/I_d$ is the same whether the channel pinch-off takes place at the source or at the drain side. This is not true however, when the trap is close to the drain, where a very important difference in $\Delta I_d/I_d$ is observed: when increasing V_s the RTN amplitude gets amplified and reaches a maximum at high V_s , which is almost 7 times higher than when biasing with the same voltage at the drain side. Therefore, complementary measurements at high V_s (pinch-off near the S) with the drain grounded could lead to an easier detection of traps near the drain. Vice-versa, traps near the source region would be easier to detect at high V_d bias. Another interesting fact is that in the linear region of operation the RTN amplitudes $\Delta I_d/I_d$ are comparable for all 3 trap positions. This means that even if a RTN signal was detectable in terms of amplitude, it could be part of a multi-level RTN signal consisting of all these active traps. If we had not taken this issue into consideration, we would have been falsely led to the conclusion that only a single trap exists which in turn would lead to a false evaluation of the oxide/silicon interface quality. Hence, a combination of both linear and saturation regimes (at both D and S sides) is necessary for a complete trap activity monitoring.

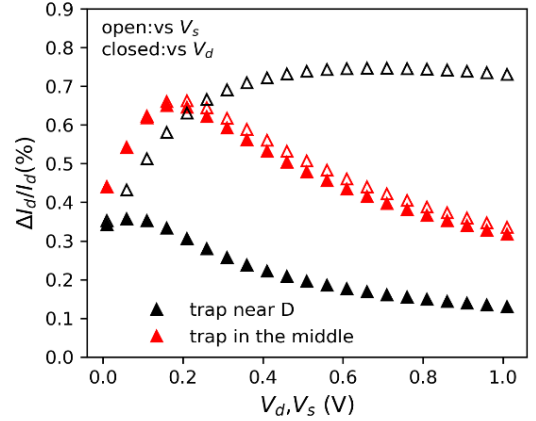


Figure 7. Simulated normalized RTN amplitude $\Delta I_d/I_d$ versus drain (closed symbols) and source (open symbols) voltage for $V_g = 0.5$ V. ($W = 1$ μ m, $L = 30$ nm).

IV. PROPOSED “PINCH TO DETECT” METHOD AND STATISTICAL RESULTS

Based on the above analysis, we propose a measurement procedure consisting of three sets of bias conditions: linear region (low V_d), saturation with pinch-off near the drain (high V_d), and finally saturation with pinch-off near the source (high V_s , $V_d = 0$ V). Each time, the gate bias varies from weak to strong inversion. To further showcase the importance of performing also source saturation measurements, we compared the behavior of Device #11 under high V_s (Fig. 8) against the one under high V_d (Fig. 3). From the obvious difference concerning the histograms -and thus the trap’s occupancy- behavior with V_g between high V_d and high V_s , it is concluded that these are two different traps.

We applied this triplet of measurements, from weak to strong inversion, on a set of 36 devices across the wafer. In Fig. 9 it is illustrated how different is the number of detected RTN traps per die for each bias condition. Combining all results, the total number of detected RTN traps can be increased even by 2. Therefore, the combination of all three sets of measurements can provide a much more reliable estimation and at the same time maximize the total number of RTN traps.

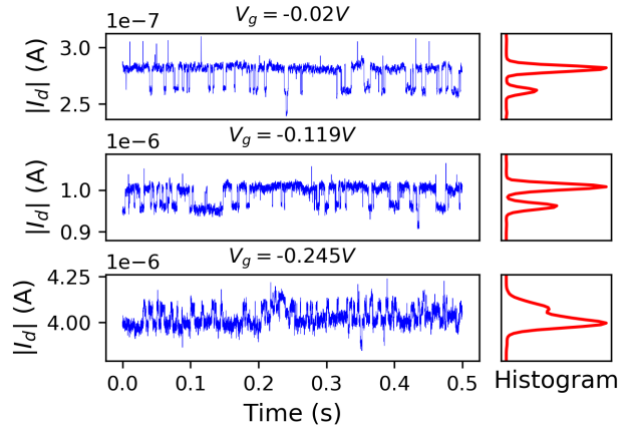


Figure 8. Device #11. Selected time-series and corresponding histograms for three gate voltage bias (from weak to strong inversion) in saturation region, but with the pinch-off at the source side ($V_s = -0.9$ V).

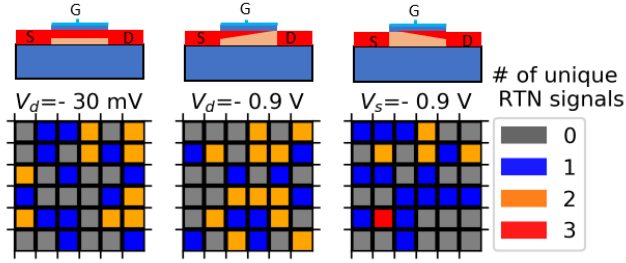


Figure 9. Colored wafer map with number of detected RTN traps for all three measurement sets of our proposed methodology.

In addition, the significant differences in the spectra of the input-referred gate voltage noise $S_{V_g} = S_{i_d}/g_m^2$ (see Fig. 10) further confirm the importance of the complementarity of these measurements. The presence of several Lorentzian ($1/f^2$ slope at high frequencies) spectra in both saturation modes but not in the linear region of operation confirms the presence of new traps. Moreover, while the mean and median spectra are both $1/f$ -like and almost coinciding for the linear regime, for both high V_d and high V_s cases the mean spectrum is non- $1/f$ and heavily influenced by the presence of additional traps, whereas the median is the same as in linear region. This shows that in saturation region there is a lack of uniformity in the trap distribution. We assume that this effect may be related to traps near the source or drain junction regions or below the spacers, that are activated or detected only in the saturation regime.

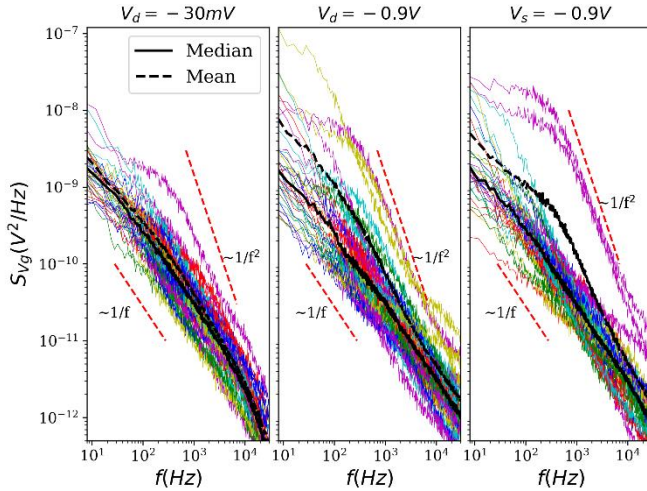


Figure 10. Input-referred gate voltage noise, $S_{V_g} = S_{i_d}/g_m^2$, for all dies and for all three measurement sets of our proposed methodology at $|V_g - V_t| = 0.4$ V. A significantly different behavior is observed in both median level and spectral characteristics.

V. INFLUENCE OF CHANNEL LENGTH

Since the pinch-off and CLM effects become less important in longer channels, one would logically expect that this dramatic change of RTN amplitudes in saturation regime is no longer true for much higher values of L . In order to investigate this influence of the channel's length on how the pinch-off affects the RTN amplitude, we simulated the $\Delta I_d/I_d$ at saturation ($V_d = 1$ V) for three channel lengths, $L = 30$ nm, $L = 0.2$ μm and $L = 0.4$ μm , and three positions of the trap across

the channel: near S, in the middle and near D. The length-normalized $\Delta I_d/I_d$ results are shown in Fig. 11. As one can surprisingly notice, for channels longer than 100 nm, the effect is not only present, but also the behavior becomes independent of the length: traps located near S and in the middle have an enhanced RTN amplitude, whereas the strength of traps near D is attenuated.

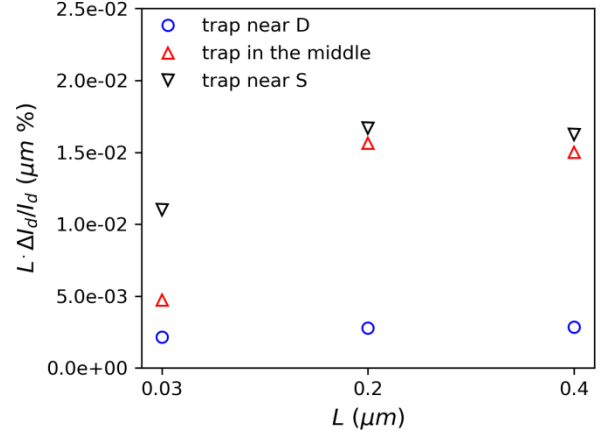


Figure 11. Simulated length-normalized RTN amplitude versus gate length, for three trap positions, biased at $V_g - V_t = 0.5$ V and $V_d = 1$ V.

Nevertheless, this is not the case for the 30 nm length device, where both “near S” and “in the middle” traps’ amplitude have decreased compared to longer devices. Regarding the former, it can actually be explained by looking at the simulation results of Fig. 5, where it becomes evident that in such a short channel, a high V_d can impact the carrier concentration all across the channel, even next to S. It can be regarded as a DIBL-like effect on RTN. It is worth noting that the normalized RTN amplitude of the trap in the middle is also attenuated a lot compared to longer lengths. In a similar way, this also seems like an ultra-short channel effect: the pinched-off region is not limited near the drain, but reaches even near the middle of the channel (also visible in Fig. 5 and Fig 7).

VI. CONCLUSIONS

It has been demonstrated through systematic measurements, as well as TCAD simulations, that in nanometer length MOSFETs, such as the Tri-Gate Nanowire transistors under study, the pinch-off effect occurring in saturation regime plays a critical role in the detection of RTN traps. It was shown that this is related to the channel uniformity across the device length and how it can affect the RTN amplitude and time constants. We have presented the “pinch to detect” method, that takes advantage of this effect in order to maximize the number of detectable RTN traps, and therefore achieve a more complete device reliability characterization in terms of electrically active defects. Finally, TCAD simulations showed that the saturation effect on the RTN amplitude does not disappear in longer channels; on the contrary, it is maximized and above a certain length becomes length-independent.

ACKNOWLEDGMENT

The authors would like to thank Dr. Sylvain Barraud from CEA-LETI for providing the samples.

REFERENCES

- [1] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFETs," *IEEE Electron Device Lett.*, vol. 11, no. 2, pp. 90–92, 1990, DOI:10.1109/55.46938.
- [2] K. Takeuchi *et al.*, "Direct observation of RTN-induced SRAM failure by accelerated testing and its application to product reliability assessment," *Dig. Tech. Pap. - Symp. VLSI Technol.*, pp. 189–190, 2010, DOI:10.1109/VLSIT.2010.5556222.
- [3] J. M. Woo, H. H. Park, S. M. Hong, J. P. Young, S. M. Hong, and H. P. Chan, "Statistical analysis of random telegraph noise in CMOS image sensors," *Int. Conf. Simul. Semicond. Process. Devices, SISPAD*, pp. 77–80, 2008, DOI:10.1109/SISPAD.2008.4648241.
- [4] C. Marquez, N. Rodriguez, F. Gamiz, and A. Ohata, "Systematic method for electrical characterization of random telegraph noise in MOSFETs," *Solid. State. Electron.*, vol. 128, pp. 115–120, 2017, DOI:10.1016/j.sse.2016.10.031.
- [5] T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi, "Statistical characterization of trap position, energy, amplitude and time constants by RTN measurement of multiple individual traps," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 628–631, 2010, DOI:10.1109/IEDM.2010.5703437.
- [6] P. Martin-Gonthier and P. Magnan, "RTS noise impact in CMOS image sensors readout circuit," in *2009 16th IEEE International Conference on Electronics, Circuits and Systems - (ICECS 2009)*, 2009, pp. 928–931.
- [7] M. J. Kirton and M. J. Uren, "Capture and emission kinetics of individual Si:SiO₂ interface states," *Appl. Phys. Lett.*, 1986, DOI:10.1063/1.97000.
- [8] C. Marquez, N. Rodriguez, F. Gamiz, R. Ruiz, and A. Ohata, "Electrical characterization of Random Telegraph Noise in Fully-Depleted Silicon-On-Insulator MOSFETs under extended temperature range and back-bias operation," *Solid. State. Electron.*, vol. 117, pp. 60–65, 2016, DOI:10.1016/j.sse.2015.11.022.
- [9] G. Ghibaudo and T. Boutchacha, "Electrical noise and RTS fluctuations in advanced CMOS devices," *Microelectron. Reliab.*, vol. 42, no. 4–5, pp. 573–582, 2002, DOI:10.1016/S0026-2714(02)00025-2.
- [10] E. Simoen, B. Dierickx, B. De Canne, F. Thoma, and C. Claeys, "On the gate- and drain-voltage dependence of the RTS amplitude in submicron MOSTs," *Appl. Phys. A Solids Surfaces*, 1994, DOI:10.1007/BF00323609.
- [11] H. C. Han, C. Theodorou, and G. Ghibaudo, "A 4-Terminal Method for Oxide and Semiconductor Trap Characterization in FDSOI MOSFETs," *25th Int. Conf. Noise Fluctuations (ICNF 2019)*, 2019, DOI:10.5075/EPFL-ICLAB-ICNF-269249.
- [12] S. Barraud *et al.*, "Opportunities and challenges of nanowire-based CMOS technologies," in *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, S3S 2015*, 2015.
- [13] J. A. Chroboczek, A. Szewczyk, and G. Piantino, "Low frequency noise point probe measurements on a wafer level using a novel programmable current amplifier," in *Noise in Physical Systems and 1/F Fluctuations*, 2001, pp. 701–704.
- [14] C. Theodorou and G. Ghibaudo, "Noise and fluctuations in fully depleted silicon-on-insulator mosfets," in *Noise in Nanoscale Semiconductor Devices*, Springer International Publishing, 2020, pp. 33–85.